Vertebrata PalAsiatica

2026-08-12

ChinaRxiv

Abstract

Body mass is a fundamental biological parameter for inferring the flight performance, physiological constraints, and ecological niches of early birds. However, estimating the body mass of Mesozoic birds remains challenging because fossils are often affected by taphonomic compression, deformation, and incomplete skeletal preservation. Based on a dataset of specimens from 765 extant bird species, this study established univariate and multivariate regression models to identify optimal predictors and evaluate model performance. The results show that midshaft humeral width (HW) has the strongest predictive power among the univariate models; after excluding outliers (N=745), the HW-based univariate model reached an R^2 of 0.941. Multivariate regression further improved model fit (adjusted R^2=0.967, N=745) and reduced both mean prediction bias (%MPE from 7.5% to 4.1%) and the magnitude of prediction error [mean (|%PE|) from 31.6% to 22.0%]. By applying these models to Mesozoic avian taxa, we generated more precise body mass estimates, compared patterns across different groups with model outputs, and benchmarked the results against existing models. Overall, the analysis supports multivariate regression as a more robust framework for paleobiological body mass estimation and provides a stronger foundation for future studies of body size evolution in early birds.

Full Text

Vertebrata PalAsiatica

DOI: 10.19615/j.cnki.2096-9899.260731

CSTR: 32090.14.j.cnki.2096-9899.260731

Body mass estimation in Mesozoic birds: evaluating the performance of univariate and multivariate regression approaches

LIU Di[1] ZHAO Tao[2] ZHOU Zhong-He[3] ZHANG Yu-Guang[1] QIU Rui[1] WANG Bao-Peng[1]

(1 *Natural History Museum of China* Beijing 100050 liudi@nnhm.org.cn)
(2 *Institute of Palaeontology, Yunnan University* Kunming 650500)
(3 *Key Laboratory of Vertebrate Evolution and Human Origins, Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences* Beijing 100044)

Abstract Body mass is a fundamental biological parameter for inferring flight performance, physiological constraints, and ecological niches in early birds. However, estimating body mass in Mesozoic birds remains challenging because

fossils are often affected by taphonomic compression, deformation, and incomplete skeletal preservation. Here, we developed both univariate and multivariate regression models based on a dataset of 765 extant bird specimens to identify optimal predictors and evaluate model performance. Results show that humeral midshaft width (HW) provides the strongest predictive power among univariate models; after outlier screening of the dataset ($N = 745$), the HW-based univariate model achieved an R^2 of 0.941. Multivariate regression further improves model fit (adjusted $R^2 = 0.967$, $N = 745$) and reduces both mean prediction bias (%MPE from 7.5% to 4.1%) and prediction error magnitude [mean(|%PE|) from 31.6% to 22.0%]. Applying these models to Mesozoic bird taxa, we generated refined body-mass estimates, compared clade-level patterns and model outputs, and benchmarked our results against published models. Together, these analyses support multivariate regression as a more robust framework for paleobiological mass estimation and provide a firmer basis for future studies of early avian body-size evolution.

Key words Mesozoic birds, body mass estimation, univariate regression, multivariate regression, allometry

Citation Liu D, Zhao T, Zhou Z H et al., in press. Body mass estimation in Mesozoic birds: evaluating the performance of univariate and multivariate regression approaches. Vertebrata PalAsiatica. DOI: 10.19615/j.cnki.2096-9899.260731

1 Introduction

Body mass (BM) serves as a fundamental biological parameter influencing physiological performance, ecological interactions, and locomotor adaptations across vertebrates (Peters,

Supported by the Beijing Municipal Finance Project (No.: 11000026T000003728324).
Date of receipt:

1983; Schmidt-Nielsen, 1983; West et al., 1997; Gillooly et al., 2001). In paleobiological studies, BM estimation provides an important quantitative framework for reconstructing body-size evolution (Butler and Goswami, 2008; Hone et al., 2008). Two approaches are commonly used: volumetric reconstruction and allometric regression models based on skeletal dimensions (Henderson, 1999; Campione and Evans, 2012). Volumetric methods are broadly applicable across taxa, but their precision is often limited by incomplete fossil preservation, uncertain soft-tissue reconstructions, and challenges in accounting for internal air spaces (Campione and Evans, 2012; Serrano et al., 2015). In contrast, allometric regression models, particularly those using limb bone metrics, offer a more objective and reproducible alternative when homologous skeletal measurements are available, particularly because they can also be applied to fragmentary specimens (Campione and Evans, 2012; Campione et al., 2014). This approach is

especially valuable for Mesozoic birds, whose fossils are frequently preserved in flattened, two-dimensional states due to taphonomic compression (Liu et al., 2012; Serrano et al., 2015). The Early Cretaceous Jehol Biota of northeastern China has yielded exceptionally well-preserved avian fossils, accounting for a substantial proportion of known Mesozoic bird specimens representing major early avian clades (Zhou and Zhang, 2006). Reliable BM estimates for these taxa are crucial to addressing key evolutionary questions, including: 1) the emergence of flight-related adaptations in early avialans (e.g., Enantiornithes and Ornithuromorpha), 2) body size trends across major clades, and 3) ecological niche partitioning during the early avian radiation (Clarke et al., 2005; Butler and Goswami, 2008; Chiappe et al., 2020). Given this critical need, this study aims to: 1) develop robust univariate and multivariate regression models for estimating BM in Mesozoic birds based on extant avian skeletal scaling relationships, and 2) benchmark model outputs against previously published regressions and evaluate clade-specific differences among fossil BM estimates.

Abbreviations BM, body mass; FL, femur length; FW, femur midshaft width; HAF, humeral articular facet diameter; HL, humerus length; HW, humeral midshaft width; OLS, ordinary least squares; SEE, standard error of the estimate; SPSS, IBM SPSS Statistics; TAL, tarsometatarsus length; TAW, tarsometatarsus midshaft width; TL, tibiotarsus length; TW, tibiotarsus midshaft width; UL, ulna length; UR, univariate regression; UW, ulna midshaft width; VIF, variance inflation factor; %MPE, mean percent prediction error; %PE, percent prediction error; %SEE, percent standard error of the estimate.

2 Materials and methods

2.1 Dataset construction

To develop robust body mass (BM) estimation models for Mesozoic birds, we assembled a reference dataset of 765 extant avian specimens with reliably recorded BM values from three museum and research collections: the Natural History Museum of China (NHMC), the

Natural History Museum of Los Angeles County (LACM), and the Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences (IVPP). This dataset encompasses 401 species (235 genera, 84 families, 25 orders) and spans a body mass range from 3.9 g (*Phylloscopus proregulus*) to 6995 g (*Diomedea epomophora*), representing an approximately 1800-fold range in BM (Appendix 1). We selected taxa from a broad range of ecological guilds, including terrestrial, aquatic, raptorial, and arboreal/perching taxa, ensuring comprehensive representation of both morphological and ecological diversity. Flightless paleognaths (e.g., *Struthio camelus*) were excluded due to their highly divergent morphology and distinct allometric scaling patterns, which are outside the allometric scope of the volant avian taxa targeted here (Dunning, 2007).

Our dataset exhibits a degree of taxonomic skew: Passeriformes account for approximately 44% of specimens, whereas some less speciose orders such as

Coraciiformes comprise only ~1%. Such uneven sampling may, in principle, skew regression outputs toward overrepresented clades (Smith, 2002). To evaluate the impact of sampling imbalance, we conducted a sensitivity analysis following Serrano et al. (2015) by comparing family-weighted versus unweighted regressions. Weighting did not significantly improve model fit or error metrics for the key predictors, and regression coefficients remained broadly similar; therefore, we retained the unweighted dataset for subsequent analyses.

2.2 Osteometric data collection

Body mass and osteological measurements were obtained from adult extant bird specimens, with skeletal maturity confirmed by complete fusion of long-bone epiphyses. Specimens encompassed a broad range of geographic origins, collection dates, and physiological states, and included both wild-caught and captive (zoo-housed) individuals. Body mass values were taken from specimen records and recorded in grams.

When selecting skeletal predictors for body mass estimation in Mesozoic birds, we considered the limitations of the humeral articular facet (HAF) diameter (Field et al., 2013). Although HAF exhibits strong predictive performance in crown-group birds, its transferability to stem avialans may be reduced by differences in shoulder joint morphology (Baier et al., 2006; Mayr, 2007), which could alter scaling relationships between articular features and body mass.

To improve transferability across clades and reduce sensitivity to proportional specialization, we selected ten linear osteometric variables from both the forelimb and hindlimb. Variable selection emphasized traits characterized by 1) evolutionary conservatism (e.g., midshaft widths related to biomechanical load-bearing) (Currey, 2002); 2) preservation compatibility (i.e., measurable in flattened Mesozoic fossils, thus excluding perimeter-based metrics); and 3) functional relevance (associated with locomotor energetics and body mass scaling) (Campione and Evans, 2012). The following parameters were measured: length

(L) and midshaft width (W) for the humerus (HL, HW), ulna (UL, UW), femur (FL, FW), tibiotarsus (TL, TW), and tarsometatarsus (TAL, TAW) (Fig. 1). All measurements followed standardized avian osteology protocols (Baumel et al., 1993). Linear dimensions were taken along the proximo-distal axis between articular surfaces. Midshaft widths were measured at the mediolateral diameter of the mid-diaphysis, perpendicular to the bone' s long axis. Measurements from extant specimens were acquired using Mitutoyo digital calipers (precision: (±)0.01 mm). For fossil taxa, dimensions were either taken directly from original specimens or estimated from calibrated published images using ImageJ when direct measurement was not feasible, using scale bars provided in the original publications.

Fig. 1 Illustration of limb bone measurements in *Phasianus colchicus* (NHMC 214941), shown in left lateral view
Left: complete skeletal mount with numbered labels indicating the measured

Illustration of limb bone measurements in *Phasianus colchicus*.

Figure 1: Illustration of limb bone measurements in *Phasianus colchicus*.

limb elements: 1. humerus, 2. ulna, 3. femur, 4. tibiotarsus, 5. tarsometatarsus; right: isolated limb bones showing measurement landmarks for each variable used in the regression analyses

2.3 Regression analyses

In this study, body mass from extant birds served as the dependent variable (y), and skeletal measurements were treated as independent variables (x). We used Ordinary Least Squares (OLS) regression to establish linear relationships between body mass and skeletal metrics. Both univariate and multivariate regression models were derived for estimating body mass. Although alternatives such as Reduced Major Axis (RMA) and Standardized Major Axis (SMA) regression are sometimes used in allometric studies, OLS is widely

preferred for predictive modeling and parameter estimation when the assumptions of linearity, homoscedasticity, and normally distributed residuals are met (Sokal and Rohlf, 1995; Smith, 2009; Campione and Evans, 2012). Because both body mass and skeletal measurements were positively skewed, we applied base-10 logarithmic transformations to stabilize variance and linearize the allometric relationships (Packard and Boardman, 2008).

While univariate regression models have traditionally dominated body mass estimation in fossil vertebrates, their reliance on single skeletal proxies may inadequately capture complex organismal integration. Multivariate regression (MR) addresses this limitation by incorporating multiple predictors, improving estimation accuracy and accommodating interspecific morphological and ecological variation (Campione and Evans, 2012; Serrano et al., 2015). To construct the MR model, we followed the 10:1 sample-to-predictor ratio guideline (Darlington, 1990), ensuring statistical robustness with 765 specimens and 10 skeletal metrics. Variable selection was conducted using stepwise regression in SPSS, allowing predictors to enter or be removed based on their partial contribution to the model (entry criterion $p = 0.05$; removal criterion $p = 0.10$), consistent with Serrano et al. (2015) and Mendoza et al. (2006). All statistical analyses were conducted in IBM SPSS Statistics 26 (IBM Corp., Armonk, NY).

2.4 Predictive performance of regression models

To evaluate model performance, we report the coefficient of determination (R^2) and adjusted R^2 for multivariate regression (MR) models, together with error-based metrics that quantify both error magnitude and directional bias. Because models were fitted in $\log_{10}$ space, the percent standard error of the estimate (%SEE) was calculated as $(10^{SEE} - 1) \times 100$, where SEE is the standard error of the estimate in $\log_{10}$ units. Under the assumption of approximately normal

residuals, %SEE estimates the typical multiplicative error and approximates the 68% prediction interval (±1 SEE) around fitted values on the original scale (Smith, 1984). The percent prediction error (%PE) for individual specimens is calculated as %PE = [(actual BM − predicted BM) / predicted BM] × 100. Under this definition, negative %PE indicates overestimation of body mass, whereas positive %PE indicates underestimation (Smith, 1984). We assessed systematic prediction bias—specifically directional bias—using the mean percent prediction error (%MPE). This metric quantifies the model' s tendency to consistently overestimate or underestimate body mass. For model comparisons, we primarily emphasize error magnitude using the mean absolute percent prediction error [mean(|%PE|)], i.e., the mean of the absolute value of %PE; this quantity was denoted as |%MPE| by Serrano et al. (2015).

3 Results

3.1 Univariate regression analyses

Univariate allometric models initially derived from the full extant dataset (N = 765) revealed moderate to strong linear correlations between limb-bone metrics and body mass,

with coefficients of determination (R^2) ranging from 0.627 to 0.924 (Table 1). Among all variables, humeral midshaft width (HW) exhibited the strongest fit (R^2 = 0.924, $p < 0.001$), likely reflecting the biomechanical significance of humeral midshaft robustness in supporting the mechanical loads associated with powered flight (Campione and Evans, 2012; Field et al., 2013). In contrast, the length of the tarsometatarsus (TAL) yielded the weakest correlation (R^2 = 0.627), consistent with greater ecological and locomotor variability in distal hindlimb proportions across avian taxa (Prange et al., 1979).

Table 1 Ordinary least squares (OLS) regression parameters for body-mass estimation in extant birds

Predictor (x)	Slope (b)	Intercept ($\log_{10} a$)	R^2	%SEE	%MPE	mean (|%PE|)
HL	1.925	−1.054	0.894	66.2429	16.7454	45.6988
HW	2.422	0.808	0.924	53.8216	10.4371	36.1420
UL	1.967	−1.209	0.842	86.2351	31.4313	66.2696
UW	2.518	0.953	0.92	55.6699	11.4244	36.5500
FL	2.664	−1.907	0.861	79.2102	19.9828	54.6784
FW	2.45	1.067	0.899	64.2173	14.2851	44.3983
TL	2.551	−2.328	0.827	91.5584	255.0917	290.2129
TW	2.345	1.161	0.91	59.8291	12.4667	41.1441
TAL	2.226	−1.319	0.627	160.0129	86.6287	136.9528
TAW	2.223	1.297	0.86	79.4502	18.8754	51.1437

N = 765; $\log_{10}$-transformed variables: linear regression statistics for $\log_{10}$-transformed body mass against ten individual skeletal measurements. For each predictor, regression slope, intercept, coefficient of determination (R^2), %SEE, %MPE (mean %PE), and mean(|%PE|) are reported.

3.1.1 Outlier screening and model refinement

To assess model robustness and minimize the influence of extreme values, we screened each univariate regression separately using SPSS casewise diagnostics, flagging observations with an absolute standardized residual > 3 following the guideline of Tabachnick and Fidell (2013). The number of flagged and excluded cases varied among predictors, and thus the final sample size varied by model (e.g., N = 742–763 across the ten univariate regressions; see Table 2; Fig. 2; Appendix 2). For example, the HW model excluded 20 cases (N = 745) and the coefficient of determination increased from R^2 = 0.924 to 0.941. Because the flagged cases

Table 2 Ordinary least squares (OLS) regression parameters for body mass estimation following predictor-specific outlier screening

Predictor (x)	N	Slope (b)	Intercept ($\log_{10} a$)	R^2	%SEE	%MPE	mean (|%PE|)
HL	742	1.905	−1.04	0.922	53.0363	9.7279	35.9969
HW	745	2.44	0.794	0.941	45.7702	7.5181	31.6368
UL	752	1.966	−1.223	0.875	72.2008	18.1825	50.5066
UW	751	2.561	0.929	0.941	45.6155	7.6023	31.2386
FL	763	2.67	−1.916	0.864	77.8974	19.4175	53.9470
FW	760	2.472	1.058	0.906	61.2234	13.0425	42.5818
TL	761	2.619	−2.452	0.862	78.4439	19.6705	53.2902
TW	761	2.344	1.159	0.915	57.7281	11.5802	39.8766
TAL	749	2.34	−1.518	0.705	131.4462	45.8314	92.3034
TAW	759	2.256	1.288	0.87	75.5502	17.3807	49.3090

were not identical across predictors, outlier screening was performed within each regression rather than by applying a single global “screened dataset” . The overall improvement in model fit following predictor-specific outlier screening is summarized in Fig. 3, which illustrates the relative increase in R^2 across all skeletal predictors.

Following refinement, residual Q-Q plots and influence diagnostics (e.g., Cook’ s distance) indicated no major departures from model assumptions and no single observation dominated the fit. Despite these refinements, tarsometatarsus length (TAL) remained the

Fig. 2 Univariate regression analyses of limb-bone dimensions versus body mass in extant birds following predictor-specific outlier screening

Fig. 2. Univariate regression analyses of limb-bone dimensions versus body mass in extant birds following predictor-specific outlier screening. The plots show log10-transformed relationships between limb-bone measurements and body mass after outlier screening. Solid black lines indicate best-fit regressions, and red lines indicate the 95% prediction intervals. Each panel reports the coefficient of determination (R^2). Humeral midshaft width (HW) and ulna midshaft width (UW) exhibit the strongest correlations (R^2 = 0.941), whereas tarsometatarsus length (TAL) remains the weakest predictor (R^2 = 0.705).

Figure 2: Fig. 2. Univariate regression analyses of limb-bone dimensions versus body mass in extant birds following predictor-specific outlier screening. The plots show log10-transformed relationships between limb-bone measurements and body mass after outlier screening. Solid black lines indicate best-fit regressions, and red lines indicate the 95% prediction intervals. Each panel reports the coefficient of determination (R^2). Humeral midshaft width (HW) and ulna midshaft width (UW) exhibit the strongest correlations (R^2 = 0.941), whereas tarsometatarsus length (TAL) remains the weakest predictor (R^2 = 0.705).

Fig. 3 Improvement in R^2 after outlier removal across skeletal indicators

Figure 3: Fig. 3 Improvement in R^2 after outlier removal across skeletal indicators

The plots show $\log_{10}$-transformed relationships between limb-bone measurements and body mass after outlier screening. Solid black lines indicate best-fit regressions, and red lines indicate the 95% prediction intervals. Each panel reports the coefficient of determination (R^2). Humeral midshaft width (HW) and ulna midshaft width (UW) exhibit the strongest correlations ($R^2 = 0.941$), whereas tarsometatarsus length (TAL) remains the weakest predictor ($R^2 = 0.705$)

weakest predictor ($R^2 = 0.705$), consistent with the expectation that distal limb length can vary with locomotor and ecological specialization and may be less tightly constrained by body mass than width-based load-bearing proxies.

Fig. 3 Improvement in R^2 after outlier removal across skeletal indicators
Blue bars show R^2 from the full dataset ($N = 765$), whereas orange bars show R^2 after outlier removal within each regression (sample size varies by predictor; see Table 2). Forelimb metrics (HW, UW) show a notable improvement, whereas TAL remains the weakest predictor

3.1.2 Predictive performance of univariate regression models

In the initial dataset ($N = 765$), humeral midshaft width (HW) emerged as the most reliable predictor, exhibiting the lowest error magnitude and directional bias, with a %SEE of 53.8% (approximately a 68% prediction interval under normality) and a %MPE of 10.4%, indicating relatively low directional bias. By con-

trast, hindlimb metrics—particularly tibiotarsus length (TL; %MPE = 255.1%) and tarsometatarsus length (TAL; %MPE = 86.6%)—exhibited markedly greater predictive bias (Table 1). After excluding outliers, both humeral midshaft width (HW) and ulna midshaft width (UW) showed the strongest univariate performance ($R^2 = 0.941$; Table 2). While UW exhibited slightly lower error magnitude, as reflected by mean(|%PE|) and %SEE values, HW showed slightly lower directional bias (%MPE); however, the overall performance of the two predictors was broadly comparable. For consistency with subsequent analyses and because HW is widely preserved in fossil bird specimens, we selected HW as the primary univariate proxy. These results are consistent with biomechanical expectations that forelimb elements subject to flight-induced loading may be more tightly coupled with body mass than hindlimb elements (Field et al., 2013).

When benchmarked against published models, our HW-based univariate regression shows comparatively low systematic bias. For example, Campione and Evans (2012) reported a vertebrate-wide stylopodial circumference model with %MPE = 25.6%, and Field et al. (2013) reported a humeral articular facet (HAF) diameter model with %MPE 13.0%. By comparison, our HW model exhibits smaller directional bias (%MPE = 7.5%). Within

our dataset, it also shows relatively low error magnitude [mean(|%PE|) = 31.6%; Table 3], supporting its utility for avian body mass estimation. Direct comparability may be limited by differences in taxonomic scope and metric definitions.

Table 3 Summary of published univariate equations for estimating BM in birds, including sample sizes (N), coefficients of determination (R^2), and error metrics (%MPE, mean (|%PE|))

Source	$y = ax^b$ (BM)	N	R^2	%MPE	mean(
This study	$6.427\ HW^{2.422}$	765	0.924	10.4371	36.1
This study	$6.223\ HW^{2.44}$	745	0.941	7.5181	31.6
Yalden, 1984	$13.25\ dFW^{2.353}$	~ 100	-	-	36.1
Yalden, 1984	$8.398\ dHW^{2.5}$	~ 100	-	-	39.0
Yalden, 1984	$0.143\ HL^{2.189}$	~ 100	-	-	71.3
Yalden, 1984	$0.047\ UL^{2.118}$	~ 100	-	-	53.7
Liu et al., 2012	$0.117\ HL^{1.733}$	422	0.851	-	66.4

Source	$y = ax^b$ (BM)	N	R^2	%MPE	mean(
Campione and Evans, 2012	0.0788 $(CH+F)^{2.749}$	245	0.988	25.6	-
Field et al., 2013	7.398 $HAF^{2.44}$	863	0.98	12.95	-
Serrano et al., 2015	0.0001 $HL^{1.974}$	486	0.904	-	40.1
Serrano et al., 2015	0.008 $dUW^{2.715}$	489	0.940	-	30.6
Serrano et al., 2015	0.016 $dFWcc^{2.33}$	480	0.953	-	25.8

3.2 Multiple regression analyses

While univariate models have traditionally dominated body mass estimation in fossil vertebrates, their reliance on single skeletal proxies may inadequately capture organismal integration and ecological diversity. In contrast, multiple regression (MR) incorporates multiple morphometric variables, improving predictive accuracy and accounting for the complex allometric relationships observed in avian taxa (Christiansen and Fariña, 2004; Tabachnick and Fidell, 2013; Serrano et al., 2015).

Following a commonly used heuristic of a 10:1 sample-to-predictor ratio (Darlington, 1990), our dataset ($N = 765$; 10 candidate variables) provided adequate support for multivariate modeling. We performed variable selection using the automatic stepwise procedure in SPSS 26, allowing predictors to enter or be removed at each step based on their partial contribution to the model, with entry and removal criteria of $p = 0.05$ and $p = 0.10$, respectively. Among the resulting models, we selected parsimonious solutions based on adjusted R^2 and biological interpretability.

The preferred model, derived from the full dataset, retained six key predictors –humeral midshaft width (HW) and length (HL), ulna midshaft width (UW) and length (UL), tibiotarsus width (TW), and femur length (FL)–yielding an adjusted R^2 of 0.952, accounting for 95.2% of the variance in $\log_{10}$-transformed body mass. After excluding outliers, the refined model ($N = 745$) retained the same six variables and demonstrated improved performance, with an adjusted R^2 of 0.967 (Table 4). The full list of specimens excluded during multivariate outlier screening, together with justifications, is documented in Appendix 2.

Notably, the dominance of forelimb metrics (HW, HL, UW, UL) in the final model aligns with flight-driven biomechanical constraints, while the inclusion of hindlimb variables (TW, FL) likely reflects scaling related to

terrestrial locomotion. The corresponding regression equations for both the initial and outlier-adjusted datasets are presented in Table 5.

Table 4 Multivariate regression model statistics for the initial dataset ($N = 765$) and the outlier-screened dataset ($N = 745$)

	Adjusted		Intercept			mean	
N=765 Variable	R^2	b	($\log_{10} a$)	%SEE	%MPE	(	%PE
N=765 HW	0.952	0.592	0.212	41.0402	6.6619	26.3074	39.642
N=765 TW		0.477					22.475
N=765 HL		1.088					46.61
N=765 UL		−0.75					33.209
N=765 UW		0.641					32.744
N=765 FL		0.38					13.719
N=745 Variable	Adjusted	b	Intercept	%SEE	%MPE	mean	%PE
	R^2		($\log_{10} a$)			(	
N=745 UW	0.967	0.768	0.175	32.9765	4.1038	21.9593	37.578
N=745 TW		0.403					22.446
N=745 HL		1.177					46.866
N=745 UL		−0.85					33.117
N=745 FL		0.414					13.647
N=745 HW		0.537					42.808

Table 5 Multivariate regression equations for the initial dataset (a) and the outlier-screened dataset (b)

Multiple regression functions	$\log_{10}(\mathrm{BM}) = \log_{10}(a) + b_1 \log_{10}(x_1) + b_2 \log_{10}(x_2) \ldots + b_p \log_{10}(x_p)$
$N = 765$ (a)	$\log_{10}(\mathrm{BM}) = 0.212 + 0.592 \times \log_{10}(\mathrm{HW}) + 0.477 \times \log_{10}(\mathrm{TW}) + 1.088 \times \log_{10}(\mathrm{HL}) - 0.75 \times \log_{10}(\mathrm{UL}) + 0.641 \times \log_{10}(\mathrm{UW}) + 0.38 \times \log_{10}(\mathrm{FL})$
Back-transformed form	$\mathrm{BM} = 1.63 \times \mathrm{HW}^{0.592} \times \mathrm{TW}^{0.477} \times \mathrm{HL}^{1.088} \times \mathrm{UW}^{0.641} \times \mathrm{FL}^{0.38} \div \mathrm{UL}^{0.75}$
$N = 745$ (b)	$\log_{10}(\mathrm{BM}) = 0.175 + 0.768 \times \log_{10}(\mathrm{UW}) + 0.403 \times \log_{10}(\mathrm{TW}) + 1.177 \times \log_{10}(\mathrm{HL}) - 0.85 \times \log_{10}(\mathrm{UL}) + 0.414 \times \log_{10}(\mathrm{FL}) + 0.537 \times \log_{10}(\mathrm{HW})$
Back-transformed form	$\mathrm{BM} = 1.496 \times \mathrm{UW}^{0.768} \times \mathrm{TW}^{0.403} \times \mathrm{HL}^{1.177} \times \mathrm{FL}^{0.414} \times \mathrm{HW}^{0.537} \div \mathrm{UL}^{0.85}$

Note: equations are presented in both log10-transformed and back-transformed (power-law) forms. BM is expressed in g and linear measurements in mm.

The predictive performance of the multivariate regression (MR) model was evaluated using the outlier-screened dataset ($N = 745$). Compared with the best univariate model (HW; $R^2 = 0.941$), MR improved model fit (adjusted $R^2 = 0.967$) and reduced both directional bias (%MPE = 4.1% vs. 7.5%) and error magnitude [mean(|%PE|): 22.0% vs. 31.6%; %SEE: 33.0% vs. 45.8%] (Tables 2, 4). Similar patterns were observed in the full dataset ($N = 765$).

In the final MR model, variance inflation factors (VIFs) were high (13.6–46.8), indicating substantial collinearity among predictors. However, because our primary goal is prediction rather than inference on individual coefficients, multicollinearity is less problematic in this context. Consistent with this, MR shows lower mean(|%PE|) and %SEE values, indicating better predictive performance. This accords with prior work noting that, in linear regression, collinearity primarily inflates coefficient standard errors and complicates interpretation of individual predictor coefficients, but it does not necessarily degrade predictive performance when the focus is on overall estimation error rather than inference on individual coefficients (Serrano et al., 2015).

Moreover, integrating multiple biomechanically linked traits (e.g., humerus–ulna scaling) allows MR to capture covariation that single-proxy models cannot capture and reduces reliance

on any single measurement. As Christiansen and Fariña (2004) noted, correlations among osteological variables are intrinsic to organismal biology and are not inherently problematic for prediction-focused frameworks. Overall, our results demonstrate that MR reduces directional bias and improves predictive performance in the extant calibration dataset, and we recommend it for fossil avian body-mass estimation when multiple skeletal elements are available.

4 Body mass estimation for Mesozoic birds

We applied our two top-performing models—the humeral midshaft width (HW) univariate regression (UR; $N = 745$) and the multivariate regression (MR; $N = 745$)—to a dataset of 92 Mesozoic bird specimens representing 54 species. Due to incomplete preservation of the limb elements required for the multivariate model (e.g., ulna, tibiotarsus), MR could be applied to only 84 specimens representing 48 species (see Appendix 3). The dataset encompasses basal avialans, enantiornithines, and ornithuromorphs, enabling a broad-scale comparative assessment of body mass evolution from the Late Jurassic to the Late Cretaceous.

4.1 Body mass estimates from univariate regression

The HW-based univariate model reveals substantial intraspecific variability in Mesozoic bird body mass (BM). Among basal avialans, *Archaeopteryx* ranges from 315.9 g (BSP 1999 I 50) to 717.9 g (BMMS 500), and *Jeholornis prima*

spans 533.9–1886.1 g (approximately a 3.5-fold range; Appendix 3). Other basal taxa, such as *Sapeornis chaoyangensis* and *Confuciusornis sanctus*, also show substantial variation in estimated BM, ranging from 926.0 to 1571.3 g ($N = 9$) and from 285.9 to 743.1 g ($N = 11$), respectively. The wide range in *Confuciusornis* may partly reflect ontogenetic variation, as the smallest specimens likely represent immature individuals.

Estimated BM values range from 33.8 g in the smallest taxa of our sample (e.g., *Eoalulavis hoyasi*, *Iberomesornis romerali*, and *Rapaxavis pani*) to 533.9 g in *Pengornis houi*, with intermediate values including 52.7–76.1 g in *Protopteryx fengningensis* and 161.7 g in *Longipteryx chaoyangensis* (Appendix 3). Additionally, bohaiornithid specimens (including *Bohaiornis guoi*) ranged from 150.5 to 244.3 g, highlighting a relatively robust body plan within this clade. Notably, the Late Cretaceous *Elsornis keni* reached 1325.4 g, representing one of the largest enantiornithines in our sample. In contrast, Ornithuromorpha displays a broader size spectrum. BM estimates are 161.7–257.7 g for *Archaeorhynchus spathula*, 844.0 g for *Jianchangornis microdonta*, 743.1 g for *Patagopteryx deferrariisi*, and 315.6 g for *Vegavis iaai*.

These findings underscore both substantial intraspecific variation—especially in *Sapeornis chaoyangensis* and *Confuciusornis sanctus*—and broader clade-level trends. Enantiornithines are generally small-bodied (<500 g), though a few taxa (e.g., *Elsornis keni*)

approach the larger sizes more typical of ornithuromorphs. Conversely, ornithuromorphs exhibit a wider range of BM values, from moderately sized forms (e.g., *Archaeorhynchus*) to large-bodied taxa such as *Jianchangornis* and *Patagopteryx*. The observed variation across and within clades may reflect ontogenetic trajectories, sexual dimorphism, and ecological niche differentiation (Bell and Chiappe, 2011).

4.2 Comparative analysis of univariate and multivariate regression models

To compare model outputs and examine clade-specific differences, we compared BM estimates for 84 Mesozoic specimens for which both models could be applied, using the HW-based UR and MR equations calibrated on the outlier-screened extant dataset ($N = 745$). Overall, the two models yielded broadly comparable estimates, but clade-specific differences were evident. Given MR' s superior performance in the extant calibration dataset, these patterns likely reflect the greater sensitivity of the single-proxy UR to proportional specialization, whereas MR integrates multiple traits and is less dependent on any single proxy.

Among basal avialans (e.g., *Archaeopteryx*, *Sapeornis*), MR estimates exceeded UR estimates in 54% of specimens (19/35), with a mean discrepancy (ΔBM = mean [MR−UR]) of 5.8 g, indicating that UR tends to yield slightly lower estimates than MR in this group. For instance, BM estimates for *Archaeopteryx* were 16%–39% higher under MR (438.4–830.0 g) than under UR (315.9–717.9

g). Similarly, for *Sapeornis*, MR estimates ranged from 1132.5 to 1529.4 g, with 7 out of 9 specimens showing higher values than the univariate model. However, for *Confuciusornis*, there was more variation, with only 3 out of 11 specimens showing higher multivariate estimates than the univariate results. In enantiornithines, MR estimates exceeded UR estimates in 56% of specimens (15/27), yet the mean ΔBM was −3.0 g. This apparent discrepancy arises because a few specimens show large negative ΔBM values (UR > MR) that dominate the mean despite MR being higher in a slight majority of cases (e.g., *Sulcavis*: UR = 484.8 g vs MR = 310.4 g, ΔBM = −174.4 g). Such cases illustrate that reliance on a single proxy (HW) can yield directionally sensitive results in proportionally specialized taxa, whereas MR integrates both forelimb (HW, HL, UW, UL) and hindlimb (TW, FL) metrics and can better accommodate departures from typical scaling relationships. Among ornithuromorphs, MR estimates were lower than UR estimates in 73% of specimens (16/22), with a substantial mean ΔBM of −44.4 g, indicating that UR commonly yields higher estimates than MR in this clade. For example, *Jianchangornis microdonta* (IVPP V16708) was estimated at 720.6 g by MR versus 844.0 g by UR (ΔBM = −123.4 g).

A paired t-test on the 84 overlapping specimens revealed no statistically significant difference in the mean signed UR–MR discrepancy ($t = 0.85$, $p = 0.401$). The non-significant result likely reflects clade-dependent discrepancy in sign and magnitude that cancel when averaged across the pooled sample. Boxplot comparisons (Fig. 4) illustrate this heterogeneity and show that MR generally yields narrower interquartile ranges than UR, indicating reduced

dispersion (greater consistency) of estimates. An exception occurs in basal avialans, where MR distributions are slightly wider than UR, likely reflecting genuine morphological heterogeneity in early avialans which is more sensitively captured by MR due to its integration of multiple skeletal metrics. In contrast, in Enantiornithes and Ornithuromorpha, MR consistently reduces dispersion, reflecting more convergent morphologies and more predictable scaling relationships in these derived clades.

In summary, both regression approaches are methodologically sound for fossil BM estimation. However, MR is preferred whenever multiple skeletal elements are preserved, because it shows lower prediction error in the extant calibration dataset and yields more internally consistent estimates across fossil clades—particularly by reducing the tendency of the single-proxy UR to produce higher BM estimates than MR in Ornithuromorpha. We therefore recommend using MR whenever preservation permits, while UR remains a practical alternative when only a single measurement is available.

Fig. 4 Boxplots comparing univariate regression (UR) and multivariate regression (MR) body mass estimates for basal avialans (B, $N = 35$), Enantiornithes (E, $N = 27$), and Ornithuromorpha (O, $N = 22$), as well as the pooled sample (All, $N = 84$)

Boxplots comparing univariate regression (UR) and multivariate regression (MR) body mass estimates for basal avialans, Enantiornithes, Ornithuromorpha, and pooled sample.

Figure 4: Boxplots comparing univariate regression (UR) and multivariate regression (MR) body mass estimates for basal avialans, Enantiornithes, Ornithuromorpha, and pooled sample.

5 Discussion

To benchmark our body mass (BM) estimation framework, we compared predicted BMs for 28 Mesozoic bird specimens that have published estimates from previous regressions, including the univariate equations of Liu et al. (2012) and the univariate and multivariate equations of Serrano et al. (2015) (Appendix 4). Because independent "true" BM values are unavailable for fossil specimens, these comparisons quantify between-model differences rather than absolute accuracy. Throughout, discrepancies are reported as published model − benchmark model; negative values indicate that the published model yields lower BM estimates than the benchmark for the same specimen. Using our outlier-screened HW univariate regression as the benchmark for univariate comparison, Liu et al.'s (2012) HL equation yields substantially smaller estimates on average (mean discrepancy = −225.5 g),

whereas their HW equation is closer (mean signed difference = −68.2 g), consistent with generally stronger transferability of width-based proxies in this fossil sample. Serrano et al.'s (2015) HL equation likewise yields smaller BM estimates relative to our HW benchmark (mean discrepancy = −118.3 g), illustrating that length-based proxies can produce larger directional offsets in early birds.

To assess whether multivariate formulations reduce these cross-model offsets, we further compared Serrano et al.' s HL- and MR-based predictions against our outlier-screened multivariate model ($N = 745$). A paired t-test confirms that the offset between Serrano et al.' s HL model and our MR benchmark is statistically significant ($t = -4.10$, $p < 0.001$), while the discrepancy between Serrano et al.' s MR and our MR is non-significant ($t = -0.88$, $p = 0.388$). Under this common MR benchmark, the mean signed difference decreases from −112.6 g (Serrano HL − our MR) to −23.9 g (Serrano MR − our MR), indicating substantially closer agreement in the average directional offset between multivariate formulations (Appendix 4). Within our own framework, the mean signed difference between our HW univariate model and our MR model across the same 28 specimens is small (mean ΔBM = −5.7 g), although specimen-level differences can still be substantial, particularly in proportionally specialized taxa. Importantly, a near-zero mean signed difference does not imply that two models are interchangeable or that their outputs are uniformly close, because opposing deviations across taxa or clades can cancel in the mean. Accordingly, these fossil comparisons are best interpreted as describing directional offsets among models, whereas our preference for MR is primarily supported by its lower prediction error and

directional bias in the extant calibration dataset, including %SEE, %MPE, and mean(|%PE|), as reported in the Results.

In addition to differences in predictor choice and extant calibration samples, differences among model outputs can also arise from where a fossil falls within the predictor (calibration) space spanned by extant birds. Our regressions—and those of Serrano et al. (2015) and Liu et al. (2012)—are parameterized using a limited set of linear limb-bone variables (lengths and widths) rather than overall morphological similarity. Thus, even fossils that appear "non-modern" in gross anatomy may yield convergent BM predictions if their input measurements lie near the core of the extant calibration range. Conversely, fossils can show larger divergence among model predictions when their measurement combinations fall near the edge of calibration space or exhibit atypical proportional relationships (e.g., relatively robust forelimb widths vs. slender hindlimb widths, or vice versa).

The ternary analyses presented in Liu (2022) support the general plausibility of applying extant-based regressions to most Mesozoic avialans by demonstrating broad overlap in overall limb-proportion morphospace. However, such overlap should not be interpreted as evidence that extinct taxa necessarily share identical covariance structure and scaling relationships with extant crown birds.

Although multicollinearity does not necessarily compromise predictive performance within the extant calibration dataset, it may become more consequential when the model is extrapolated to extinct taxa with limb proportions that differ substantially from those of crown birds. Some Mesozoic avialans, particularly morphologically specialized basal forms or certain enantiornithines, may occupy regions of morphospace that are only sparsely sampled among extant taxa. Under such circumstances, correlated predictors may amplify uncertainty rather than simply contribute shared signal, especially when forelimb–hindlimb proportional relationships fall outside the range captured by the training sample. Accordingly, the superior overall performance of the multivariate model in extant birds should not be interpreted as implying uniformly high reliability for all fossil taxa. Instead, estimates for proportionally unusual extinct species should be interpreted cautiously and evaluated in light of their anatomical specialization, phylogenetic position, and preservational context.

Importantly, because BM is recovered on the original scale via a power-law relationship after back-transformation from $\log_{10}$ regressions (Packard and Boardman, 2008), relatively small differences in input measurements—or in how different multivariate models weight correlated predictors—can translate into comparatively large discrepancies in predicted BM for a subset of specimens. Taphonomic compression may further amplify this effect in flattened fossils, particularly for width-based variables such as humeral midshaft width (HW). In one enantiornithine specimen examined using CT reconstruction, one humerus is preserved three-dimensionally whereas the contralateral humerus is visibly flattened; the compressed humerus exhibits a midshaft width of approximately 2.2 mm compared to 1.8 mm in the undeformed side, representing an increase of

roughly 22%. As illustrated by Supplementary Figs. 1–2, such compression-related inflation in width measurements may be disproportionately magnified by allometric regressions, leading to substantially larger differences in estimated body mass. These observations suggest that univariate models relying heavily on width-based proxies may systematically overestimate BM in some flattened specimens. However, because the degree and direction of compression vary among taxa and specimens depending on original bone geometry, burial orientation, degree of compaction, and subsequent diagenetic deformation, a universal correction factor is currently impractical. Since most Jehol avialan specimens are preserved as two-dimensionally flattened skeletons, this issue cannot always be avoided in fossil BM estimation. Under such circumstances, measurements should preferentially be taken from the least deformed specimens available whenever possible. One potential advantage of the multivariate model is that it integrates both length and width information; because length variables are generally less directly affected by transverse flattening than breadth measures, their inclusion may partially buffer the influence of distortion in any single taphonomically sensitive proxy.

Accordingly, large signed differences may reflect directional shifts between models,

whereas large absolute differences may reflect divergence in magnitude regardless of direction. These effects should be interpreted as reflecting sensitivity to extrapolation, proxy choice, and preservational distortion, rather than as simple "right vs. wrong" outcomes.

6 Conclusion

This study establishes an integrative framework that improves body mass estimation in Mesozoic birds through methodological refinement and performance evaluation. We developed high-performing univariate regressions—most notably for humeral midshaft width (HW; $R^2 = 0.941$ after outlier screening)—that shows reduced directional bias relative to previously used single-proxy models. To further enhance predictive performance, we constructed a multivariate regression model (Model VI; adjusted $R^2 = 0.967$) incorporating both forelimb and hindlimb metrics. Relative to the best univariate model, this multivariate model reduces directional bias (%MPE from 7.5% to 4.1%) and lowers error magnitude (mean(|%PE|) from 31.6% to 22.0%; %SEE from 45.8% to 33.0%). Benchmarking against published models (Liu et al., 2012; Serrano et al., 2015) suggests that multivariate formulations are generally less sensitive to proxy choice and proportional specialization, particularly in morphologically specialized taxa where single proxies can be directionally biased. While univariate models remain useful when only a single skeletal element is preserved, we recommend multivariate regressions as the preferred approach when multiple measurements are available.

Nevertheless, because some Mesozoic avialans may possess limb proportions or covariance structures that differ from those of extant crown birds, body-mass

estimates for highly specialized fossil taxa should still be interpreted cautiously and evaluated in conjunction with anatomical specialization and preservational context. More robust body-mass estimates provide an important quantitative framework for investigating growth rates, ecological niche partitioning, and macroevolutionary patterns of body-size evolution in early birds.

Acknowledgements We thank Dr. Kenneth E. Campbell and Kimball Garrett of the Natural History Museum of Los Angeles County for their generous assistance with avian skeletal measurements. We also thank Li Dongsheng and Hou Yemao of the Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences, for providing access to comparative avian skeletal specimens for osteological measurements and for assistance with the micro-CT scanning of fossil bird specimens, respectively. Finally, we are grateful to Fan Shanshan for her assistance in preparing the figures.

Supplementary material can be found on the website of Vertebrate PalAsiatica (http://www.vertpala.ac.cn/CN/10.19615/j.cnki.2096-9899.260731).

Body Mass Estimation in Mesozoic Birds: Performance Evaluation of Univariate and Multivariate Regression Methods

Liu Di[1] Zhao Tao[2] Zhou Zhonghe[3] Zhang Yuguang[1] Pei Rui[1] Wang Baopeng[1]

(1 National Natural History Museum of China, Beijing 100050)

(2 Institute of Paleontology, Yunnan University, Kunming 650500)

(3 Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences; Key Laboratory of Vertebrate Evolution and Human Origins, Beijing 100044)

Abstract: Body mass is an important parameter for inferring flight ability, physiological constraints, and ecological niches in early birds. Because Mesozoic bird fossils are commonly affected by taphonomic factors such as two-dimensional compression, skeletal deformation, and the loss of key bones, body-mass estimation still involves considerable uncertainty. Based on a dataset of 765 extant bird specimens, this study established and compared univariate and multivariate regression models. By analyzing model fit and predictive stability, the best skeletal proxy variables were selected, and the predictive performance of different models was systematically evaluated. The results show that, among univariate models, humeral midshaft width (HW) has the strongest predictive ability (after screening out outliers, $R^2 = 0.941$, $N = 745$). Multivariate models further improve model fit (adjusted $R^2 = 0.967$, $N = 745$), and effectively reduce the mean percentage error (%MPE: 7.5% $\rightarrow$ 4.1%) and mean absolute percentage error [mean(|%PE|): 31.6% $\rightarrow$ 22.0%]. Applying the models to

Mesozoic birds yields more robust body-mass estimates, providing a reliable quantitative basis for studies of body-size evolution, ecological differentiation, and flight capability in early birds.

Keywords: Mesozoic birds; body-mass estimation; univariate regression; multivariate regression; allometric growth

References

Baier D B, Gatesy S M, Dial K P, 2006. Three-dimensional skeletal kinematics of the shoulder girdle and forelimb during guinea fowl locomotion. J Exp Biol, 209: 1969–1983

Baumel J J, King A S, Breazile J E et al., 1993. Handbook of Avian Anatomy: Nomina Anatomica Avium. Cambridge: Nuttall Ornithological Club. 1–779

Bell A, Chiappe L M, 2011. Statistical approach for inferring ecology of Mesozoic birds. J Syst Palaeontol, 9: 119–133

Butler R J, Goswami A, 2008. Body size evolution in Mesozoic birds: little evidence for Cope' s rule. J Evol Biol, 21: 1673–1682

Campione N E, Evans D C, 2012. A universal scaling relationship between body mass and proximal limb bone dimensions in quadrupedal terrestrial tetrapods. BMC Biol, 10: 60

Campione N E, Evans D C, Brown C M et al., 2014. Body mass estimation in non-avian bipeds using a theoretical conversion to quadruped stylopodial proportions. Method Ecol Evol, 5: 913–923

Chiappe L M, Li D, Serrano F J et al., 2020. Anatomy and flight performance of the early enantiornithine bird *Protopteryx fengningensis*: information from new specimens of the Early Cretaceous Huajiying Formation of China. Anat Rec (Hoboken), 303: 716–731

Christiansen P, Fariña R A, 2004. Mass prediction in theropod dinosaurs. Hist Biol, 16: 85–92

Clarke J A, Tambussi C P, Noriega J I et al., 2005. Definitive fossil evidence for the extant avian radiation in the Cretaceous. Nature, 433: 305–308

Currey J D, 2002. Bones: Structure and Mechanics. Princeton: Princeton University Press. 1–436

Darlington R B, 1990. Regression and Linear Models. New York: McGraw-Hill. 1–448

Dunning J B Jr, 2007. CRC Handbook of Avian Body Masses. 2nd ed. Boca Raton: CRC Press. 1–672

Field D J, Lynner C, Brown C et al., 2013. Skeletal correlates for body mass estimation in modern and fossil flying birds. PLoS One, 8(11): e82000

Gillooly J F, Brown J H, West G B et al., 2001. Effects of size and temperature on metabolic rate. Science, 293: 2248–2251

Henderson D M, 1999. Estimating the masses and centers of mass of extinct animals by 3-D mathematical slicing. Paleobiology, 25: 88–106

Hone D W, Dyke G J, Haden M et al., 2008. Body size evolution in Mesozoic birds. J Evol Biol, 21: 618–624

Liu D, 2022. Study on body mass estimation of Mesozoic birds and relevant biological issues. Ph. D dissertation. Beijing: University of Chinese Academy of Sciences. 1–186

Liu D, Zhou Z H, Zhang Y G, 2012. Mass estimate and evolutionary trend in Chinese Mesozoic fossil birds. Vert PalAsiat, 50: 39–52

Mayr G, 2007. The tenth skeletal specimen of *Archaeopteryx*. Zool J Linn Soc, 149(1): 97–116

Mendoza M, Janis C M, Palmqvist P, 2006. Estimating the body mass of extinct ungulates: a study on the use of multiple regression. J Zool, 270(1): 90–101

Packard G C, Boardman T J, 2008. Model selection and logarithmic transformation in allometric analysis. Physiol Biochem Zool, 81: 496–507

Peters R H, 1983. The Ecological Implications of Body Size. Cambridge: Cambridge University Press. 1–329

Prange H D, Anderson J F, Rahn H, 1979. Scaling of skeletal mass to body mass in birds and mammals. Am Nat, 113: 103–122

Schmidt-Nielsen K, 1983. Animal Physiology: Adaptation and Environment. New York: Cambridge University Press. 1–619

Serrano F J, Palmqvist P, Sanz J L, 2015. Multivariate analysis of neognath skeletal measurements: implications for body mass estimation in Mesozoic birds. Zool J Linn Soc, 173(4): 929–955

Smith R J, 1984. Allometric scaling in comparative biology: problems of concept and method. Am J Physiol, 246: R152–R160

Smith R J, 2002. Estimation of body mass in paleontology. J Hum Evol, 43: 271–287

Smith R J, 2009. Use and misuse of the reduced major axis for line-fitting. Syst Biol, 58: 476–490

Sokal R R, Rohlf F J, 1995. Biometry: The Principles and Practice of Statistics in Biological Research. 3rd ed. New York: W.H. Freeman. 1–887

Tabachnick B.G., Fidell L.S., 2013. Using Multivariate Statistics. 6th ed. Boston: Pearson. 1–983

West G B, Brown J H, Enquist B J, 1997. A general model for the origin of allometric scaling laws in biology. Science, 276: 122–126

Yalden D W, 1984. What size was *Archaeopteryx*? Zool J Linn Soc, 82(1-2): 177–188

Note: Figure translations are in progress. See original paper for figures.