Alternative methods for calculating percentage prediction error and their implications for predicting body mass in fossil taxa

Gabriel S. Yapuncich

Department of Evolutionary Anthropology, Duke University, Durham, NC 27708, USA

ARTICLE INFO

ABSTRACT

Since body mass covaries with many ecological aspects of an animal, body mass prediction of fossil taxa is a frequent goal of paleontologists. Body mass prediction often relies on a body mass prediction equation (BMPE): a bivariate relationship between a predictor variable (e.g., molar occlusal area, femoral head breadth) and body mass as observed in extant taxa. A variety of metrics have been used to assess the reliability of BMPEs, including percentage prediction error (%PE), which involves predicting body masses of a test sample comprising individuals with associated masses. A mean %PE can be calculated in two ways: 1) as the mean %PE of multiple individual predictions (%MPE), or 2) as the %PE of mean body mass generated from the mean predictor value of multiple individuals (here termed %PEM). Differences between these two approaches have never been formally examined and no formal protocols have been recommended. Using a large sample of cercopithecoid primates (406 individuals from 50 species/subspecies) with associated body masses, body mass is predicted with six previously published interspecific BMPEs. Both %MPE and %PEM are calculated and compared. For all BMPEs, the distributions of differences between %MPE and %PEM exhibit positive skew and have medians significantly greater than zero, indicating that the examined prediction equations are more accurate at predicting mean body mass when they are applied to mean predictor values. The decreased predictive accuracy of %MPE relative to %PEM likely stems from changing the unit of analysis from mean values (in the reference sample) to individual values (in the test sample) when calculating %MPE. Empirical results are supported with a simulated dataset. Implications for body mass prediction in fossil species are discussed.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

Because an animal's body mass covaries with many aspects of its ecology, physiology, and behavior (Peters, 1983; Calder, 1984; Schmidt-Nielsen, 1984), predicting body mass is a frequent goal in comparative biology and paleontology (e.g., Gingerich et al., 1982; Conroy, 1987; Dagosto and Terranova, 1992; Delson et al., 2000; Ruff, 2003). While some studies focus exclusively on issues surrounding body mass prediction, it is often a component of broader descriptions of fossil material (e.g., Kay and Simons, 1980; Meldrum and Kay, 1997; Bloch et al., 1998; Sears et al., 2008; Boyer et al., 2013; Bertrand et al., 2016). The most common method for predicting body mass uses the observed relationship between body mass and a particular predictor variable. Popular predictor variables for paleontological application include molar dimensions (Kay and Simons, 1980; Gingerich et al., 1982; Conroy, 1987), cranial

dimensions (Aiello and Wood, 1994; Kappelman, 1996; Spocter and Manger, 2007; Silcox et al., 2009), and limb bone articular surface and diaphyseal dimensions (Ruff et al., 1991; Dagosto and Terranova, 1992; Grine et al., 1995; Egi, 2001; Ruff, 2003; Yapuncich et al., 2015). Using a robust reference sample, these relationships are typically modeled with ordinary least squares (OLS) linear regression of logarithmically transformed variables, although reduced major axis linear regression has also been used (see Warton et al., 2006; Smith, 2009).

Different predictor variables (i.e., dental versus postcranial) can have profound effects on the inferred body mass of fossil taxa. For example, molar dimensions of the Eocene omomyiform *Hemiacodon gracilis* suggest a body mass between 640 and 1150 g (Conroy, 1987; Dagosto and Terranova, 1992; Jones et al., 2014), while predictions from tarsal dimensions are much lower (250–400 g; Dagosto and Terranova, 1992). Since these values fall on either side of "Kay's threshold" of 500 g, researchers using a particular predictor variable could reach divergent conclusions regarding the primary source of protein for *Hemiacodon*. As many

E-mail address: gabriel.yapuncich@duke.edu.

paleoecological inferences rely on predicted body masses, it is crucial that researchers are able to identify which values are most reliable. These preferences are often determined by evaluating the precision and accuracy of the underlying body mass prediction equations (BMPEs) using statistical metrics such as the coefficient of determination, mean square error, and standard error of the estimate (Gingerich, 1990; Smith, 2002; Yapuncich et al., 2015). For assessing predictive accuracy, Smith (1980, 1984) suggested calculating the percentage prediction error (%PE) for each data point with the following formula:

$$([Observed - Predicted]/Predicteed)*100 = \%PE$$

A mean percentage prediction error (%MPE) can then be calculated by averaging absolute values of the %PE of each observation. Percentage prediction error has been used as an accuracy metric in many subsequent BMPE studies (Supplementary Online Material [SOM] Table S1).

Mean percentage prediction error can be calculated using the predicted values of all observations in the equation's reference sample (e.g., Aiello and Wood, 1994; Kappelman, 1996; Spocter and Manger, 2007; Squyres and Ruff, 2015; Tsubamoto et al., 2015) or a subset of the reference sample (Payseur et al., 1999). However, this procedure does not reveal the predictive accuracy of a BMPE when applied to new observations. Therefore, many BMPE studies (SOM Table S1) utilize a separate test sample—observations not used to generate the prediction equation—to calculate %MPE. Similarly, some researchers have used novel test samples to evaluate the predictive accuracy of previously published BMPEs (e.g., Ruff, 2000; Ruff et al., 2005; Elliott et al., 2014, 2016; Yapuncich et al., 2015). Since the workflow for calculating prediction error is analogous to predicting fossil body masses, evaluating BMPEs with a test sample has an intuitive and straightforward appeal.

Among those studies with test samples, two versions of %MPE have been calculated. First, the predictor value of each individual specimen has been used to predict an individual's body mass, which is then compared to that individual's observed body mass. Mean percentage prediction error is then calculated as the average of the absolute values of all individual %PEs. When test samples comprise specimens with associated body masses, this has been the preferred method for calculating %MPE (e.g., Ruff, 2000, 2003; Delson et al., 2000; Halenar, 2011; Elliott et al., 2014, 2016). Alternatively, the sample mean of the predictor values can be used to predict a sample mean body mass, and then tested against an observed sample mean body mass. This second method has been used less frequently with test samples (Dagosto and Terranova, 1992; Biknevicius, 1999; De Estaban-Trivigno et al., 2008), although studies that compute %PE from a species mean reference sample implicitly follow this approach (including many of the studies in SOM Table S1).

Van Valkenburgh (1990:197) highlighted both alternatives for calculating prediction error, but did not advocate a particular option: "Prediction errors could probably be lowered if body mass and length data were compared for the same individual, or averages of skeletal measures and mass based on large samples were used." These procedures, averaging the %PEs for all individuals (hereinafter referred to as %MPE) versus averaging predictor values before calculating a mean %PE (subsequently referred to as the percentage prediction error of the mean or %PEM), treat the same data differently and may describe the predictive accuracy of BMPEs differently. If there were a systematic difference between these methods, it would be inappropriate to compare the prediction errors from studies utilizing alternative approaches. More critically, as the application of prediction equations to test samples replicates application to fossil taxa, any systematic difference may inform the

best procedure to reduce prediction error for fossils. However, as these approaches have not been formally examined, no consistent protocol for calculating prediction error has been developed.

This study evaluates the effect of %MPE and %PEM on reported prediction error. Body masses are predicted from dental dimensions for a large sample of primates with associated body masses. Both %MPE and %PEM are calculated and compared for several published BMPEs. Results from empirical data are supported using simulated data (SOM File 1). As a null hypothesis, this study assumes that workflow differences do not result in significantly different assessments of predictive accuracy. This hypothesis generates the following predictions:

P[1]: Per-taxon differences between %MPE and %PEM will be normally distributed. Skew in the distribution of differences would indicate that one approach might modestly reduce prediction error for most taxa, while substantially increasing prediction error for one or two species.

P[2]: The mean (or median) of the distribution of per-taxon differences between %MPE and %PEM will not be significantly different from 0. A shift in the mean (or median) would indicate that one approach has less error in the majority of test cases.

If either of these predictions is not met, it should be possible to recommend a preferred method for calculating prediction error, at least in conditions similar to those of this study.

2. Methods and materials

Maximum mesiodistal length and buccolingual breadth of the first mandibular (M_1) and maxillary (M^1) molars and maximum mesiodistal length of the second mandibular molar (M_2) were measured on 406 individuals from 50 cercopithecoid primate taxa with associated body masses (SOM Table S2) at the National Museum of Natural History (Washington, DC). All taxa were represented by at least two individuals. Occlusal area (the product of the two linear dimensions) was natural log-transformed and used to predict body mass with the primate M^1 and M_1 equations of Gingerich et al. (1982), and the male primate, anthropoid, and "monkey" M_1 equations of Conroy (1987). Additionally, M_2 length was log_{10}-transformed and used to predict body mass using the primate M_2 equation of Kay and Simons (1980).

Body masses predicted with the Gingerich et al. (1982) and Conroy (1987) equations were converted from the logarithmic scale and corrected for bias using published quasi-maximum likelihood estimates (Smith, 1993). A correction factor was not available for the Kay and Simons (1980) equation. For individual %PEs, body mass was predicted for each individual and compared to that individual's associated body mass. Mean percentage prediction error was then calculated as the mean of the absolute values of all individual %PEs. For %PEM, molar occlusal area was averaged at the species/subspecies level and used to predict a mean species/subspecies body mass. An observed mean body mass was computed as the average of associated body masses for all specimens. For each taxon, the difference between %MPE and %PEM was calculated; positive values indicate lower prediction error for %PEM. Normality was evaluated using a Shapiro–Wilk test. Depending on the results of the normality test, a *t*-test or a Wilcoxon rank sum test was used to check if the mean or median of the distribution was significantly different than zero.

All individuals ($k = 406$) from all species/subspecies ($n = 50$) were input into the Gingerich et al. (1982) and Kay and Simons (1980) equations, while only male individuals ($k = 194$) from those species/subspecies with at least two male individuals ($n = 36$) were input into the Conroy (1987) equations.