Part I

Foundations

Chapter 1

METHODOLOGICAL TOOLS

Ethan P. White, Xiao Xiao, Nick J. B. Isaac, and Richard M. Sibly

SUMMARY

1 In this chapter we discuss the best methodological tools for visually and statistically comparing predictions of the metabolic theory of ecology to data.

2 Visualizing empirical data to determine whether it is of roughly the correct general form is accomplished by log-transforming both axes for size-related patterns, and log-transforming the y-axis and plotting it against the inverse of temperature for temperature-based patterns. Visualizing these relationships while controlling for the influence of other variables can be accomplished by plotting the partial residuals of multiple regressions.

3 Fitting relationships of the same general form as the theory is generally best accomplished using ordinary least-squares-based regression on log-transformed data while accounting for phylogenetic non-independence of species using phylogenetic general linear models. When multiple factors are included this should be done using multiple regression, not by fitting relationships to residuals. Maximum likelihood methods should be used for fitting frequency distributions.

4 Fitted parameters can be compared to theoretical predictions using confidence intervals or likelihood-based comparisons.

5 Whether or not empirical data are consistent with the general functional form of the model can be assessed using goodness-of-fit tests and comparisons to the fit of alternative models with different functional forms.

6 Care should be taken when interpreting statistical analyses of general theories to remember that the goal of science is to develop models of reality that can both capture the general underlying patterns or processes and also incorporate the important biological details. Excessive emphasis on rejecting existing models without providing alternatives is of limited use.

1.1 INTRODUCTION

Two major functional relationships characterize the current form of the metabolic theory of ecology (MTE). Power-law relationships, of the form $y = cM^{\alpha}$ (Fig. 1.1A,B), describe the relationship between body size and morphological, physiological, and ecological traits of individuals and species (West et al. 1997; Brown et al. 2004). The Arrhenius equation, of the general form, $y = ce^{-E/kT}$ (Fig. 1.1C,D), characterizes the

Metabolic Ecology: A Scaling Approach, First Edition. Edited by Richard M. Sibly, James H. Brown, Astrid Kodric-Brown.

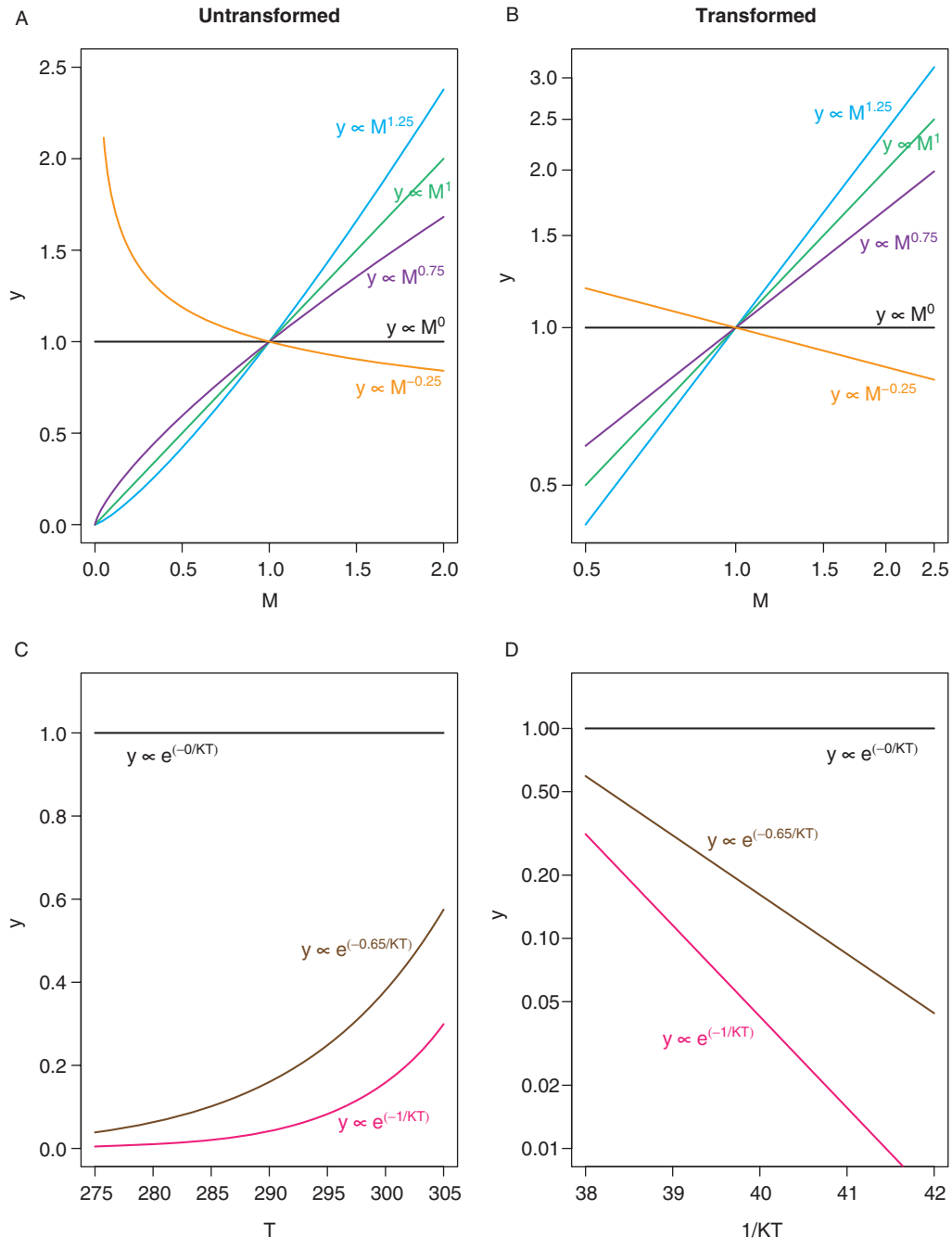

Figure 1.1 Examples of power-law relationships and exponential temperature relationships. Several power-law relationships are shown on untransformed (A) and logarithmically scaled (B) axes. Power-law relationships with exponents equal to one characterize direct proportionalities, which are linear relationships with intercepts of zero. Several temperature relationships are shown on untransformed (C) and Arrhenius plot axes ($1/T$ vs. logarithmically scaled y) (D). Power laws with exponents greater than 1 are described as superlinear because their slope is increasing in linear space and power laws with exponents less than 1 are described as sublinear because their slope is decreasing. Relationships that have exponents equal to zero do not change with the variable of interest and are therefore described as invariant with respect to mass or temperature. Note that in the Arrhenius plots different coefficients are used to allow for clear presentation.

relationships between temperature and physiological and ecological rates (Gillooly et al. 2001; Brown et al. 2004). In addition to being central to metabolic theory, these empirical relationships are utilized broadly to characterize patterns and understand processes in areas of study as diverse as animal movement (Viswanathan et al. 1996), plant function (Wright et al. 2004), and biogeography (Arrhenius 1920; Martin and Goldenfeld 2006).

Methodological approaches for comparing metabolic theory predictions to empirical data fall into two general categories: (1) determining whether the general functional form of a relationship predicted by the theory is valid; and (2) determining whether the observed values of the parameters match the specific quantitative predictions made by the theory. Both of these categories of analysis rely on being able to accurately determine the best fitting form of a model with the same general functional form as that of MTE, so we will begin by discussing how this has typically been done using ordinary least-squares (OLS) regression on appropriately transformed data. Potential improvements to these approaches that account for statistical complexities of the data will then be considered. We will discuss methods for comparing the fitted parameters to theoretical values and how to determine whether the general functional form predicted by the theory is supported by data. This will require some discussion of the philosophy of how to test theoretical models. So we will end with a general discussion of the technical and philosophical challenges of testing and developing general ecological theories.

1.2 VISUALIZING MTE RELATIONSHIPS

Before conducting any formal statistical analysis it is always best to visualize the data to determine whether the model is reasonable for the data and to identify any potential problems or complexities with the data.

1.2.1 Visualizing functional relationships

The primary model of metabolic theory describes the relationship between size, temperature, and metabolic rate; combining a power function scaling of mass and metabolic rate with the Arrhenius relationship describing the exponential influence of temperature on biochemical kinetics.

$$I = i_0 M^{\alpha} e^{-E/kT} \tag{1.1}$$

See Brown and Sibly (Chapter 2) or Brown et al. (2004) for details.

Most analyses of this central equation focus on either size or temperature in isolation, or attempt to remove the influence of the other variable before proceeding. As such, the most common analyses focus on either power-law relationships, $y = cM^b$, or exponential relationships, $y = ce^{-E/kT}$, both of which can be log-transformed to yield linear relationships (Fig. 1.1).

$$y = cM^{\alpha} \Rightarrow \log(y) = \log(c) + \alpha \log(M) \tag{1.2a}$$

$$y = ce^{-E/kT} \Rightarrow \log(y) = \log(c) - (E/kT) \tag{1.2b}$$

The linear forms of these relationships form the basis for the most common approaches to plotting these data and graphically assessing the validity of the general form of the equations. Plots of these linearized forms are obtained either by log-transforming the appropriate variables or by logarithmically scaling the axes so that the linear values remain on the axes, but the distance between values is adjusted to be equivalent to log-transformed data. In this book all linearized plots will used log-scaled, rather than log-transformed, axes. Relationships between size and morphological, physiological, and ecological factors are typically plotted on log-log axes and relationships between temperature and these factors are displayed using Arrhenius plots with the log-scaled y variable plotted against the inverse of temperature (Fig. 1.2A,B). If the relationships displayed on plots of these forms are approximately linear then they are at least roughly consistent with the general form predicted by metabolic theory.

When information on both size and temperature are included in an analysis to understand their combined impacts on a biological factor, this has been displayed graphically by removing the effect of one factor and then plotting the relationship for the other factor (Fig. 1.2C,D). The basic idea is to rewrite the combined size–temperature equation so that only one of the two variables of interest appears on the right-hand side.

$$\frac{y}{ce^{-\frac{E}{kT}}} = M^{\alpha} \Rightarrow \log(y) + \frac{E}{kT} - \log(c) = \alpha \log(M) \tag{1.3a}$$

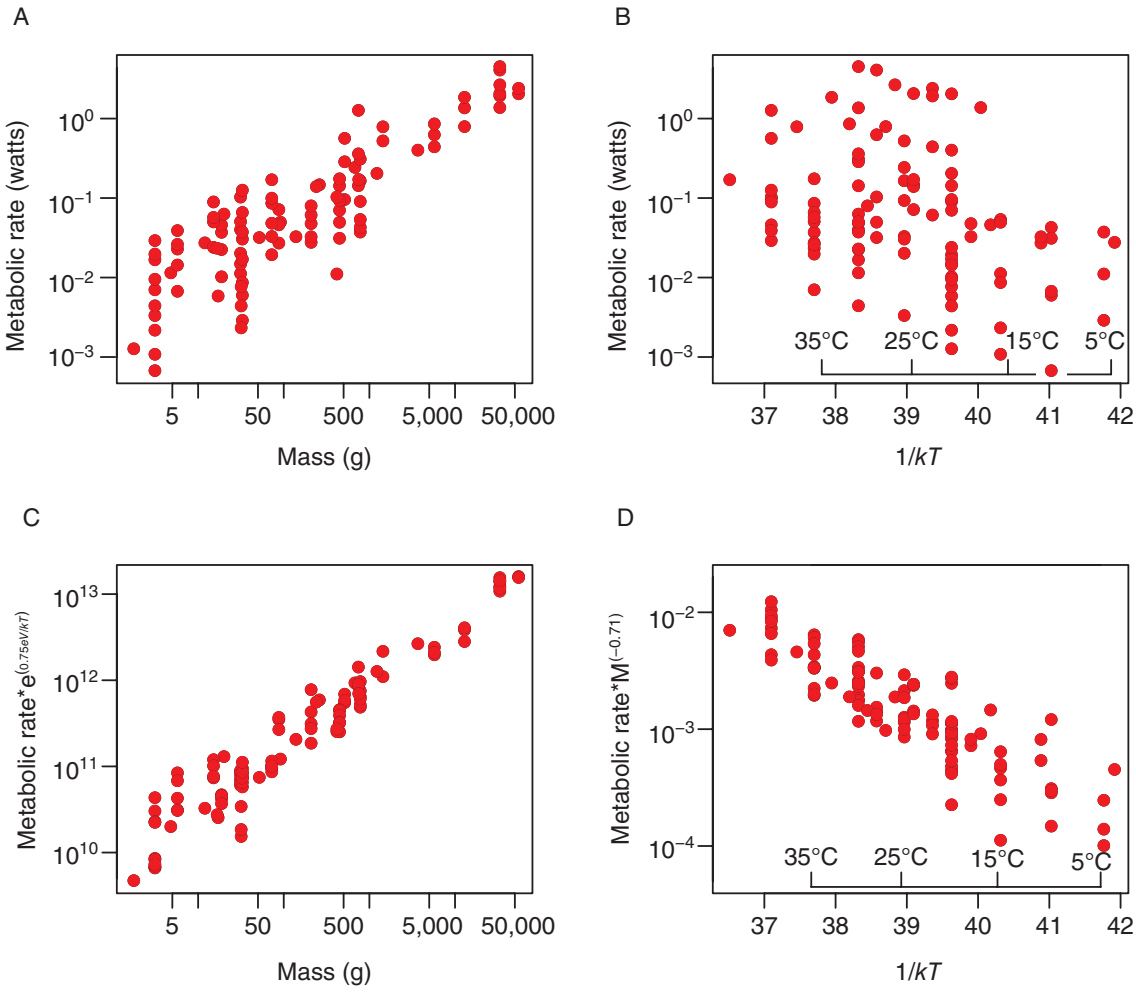

Figure 1.2 Plots of metabolic rate as a function of mass and temperature. (A) Log-log plot of mass vs. metabolic rate not accounting for temperature. (B) Arrhenius plot of temperature vs. metabolic rate not accounting for mass. (C) Log-log plot of mass vs. metabolic rate accounting for temperature. (D) Arrhenius plot of temperature vs. metabolic rate accounting for mass. Data is for reptiles from Gillooly et al. (2001).

$$\frac{y}{cM^{\alpha}} = e^{-E/kT} \Rightarrow \log(y) - \alpha\log(M) - \log(c) = -\frac{E}{kT} \tag{1.3b}$$

The value for the dependent variable (i.e., the value plotted for each point on the vertical axis) is then determined by dividing the observed value of y by the appropriate transformation of temperature or mass for the observation and log-transforming the resulting value. This is equivalent to the standard approach of plotting the partial residuals to visualize the relationship with a single predictor variable in multiple regression. Often in the MTE literature the theoretical forms of the relationships ($\alpha = 0.75$, $E = -0.65$) have been used rather than the fitted forms based on multiple regression. For reasons discussed below we recommend using the fitted values of the parameters, or simply using the partial residuals functions in most statistical packages, to provide the best visualization of the relationship with the variable of interest.

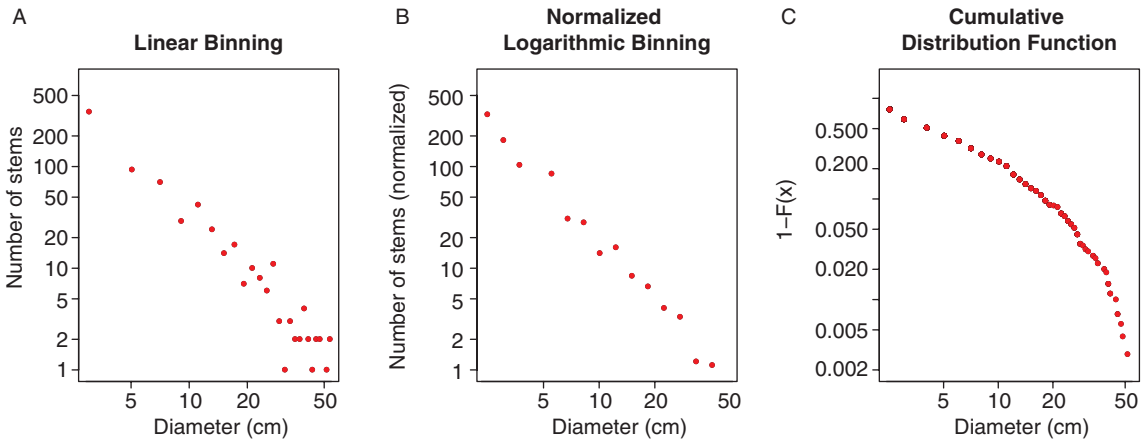

Figure 1.3 Examples of visualizations of frequency distributions. Methods include (A) linear binning, (B) normalized-logarithmic binning, and (C) linearizing the cumulative distribution function. Data are from the Nosy Mangabe, Madagascar, site of Alwyn Gentry's tree transect data (site 201; Phillips and Miller 2002).

1.2.2 Frequency distributions

In addition to making predictions for the relationships between pairs of variables – e.g., size, temperature, and metabolic rate – metabolic ecology models have been used to make predictions for the form of frequency distributions (i.e., histograms) of biological properties such as the number of trees of different sizes in a stand (Fig. 1.3; West et al. 2009). The predicted forms of these distributions are typically power laws and have often been plotted by making histograms of the variable of interest, log-transforming both the counts and the bin centers and then plotting the counts on the *y*-axis and the bin centers on the *x*-axis (Fig. 1.3A; e.g., Enquist and Niklas 2001; Enquist et al. 2009). This is a reasonable way to visualize these data, but it suffers from the fact that bins with zero individuals must be excluded from the analysis due to the log-transformation. These bins will occur commonly in low probability regions of the distribution (e.g., at large diameters), thus impacting the visual perception of the form of the distribution. To address this problem we recommend using normalized logarithmic binning (*sensu* White et al. 2008), the method typically used for visualizing this type of distribution in the aquatic literature (e.g., Kerr and Dickie 2001). This approach involves binning the data into equal logarithmic width bins (either by log-transforming the data prior to constructing the histogram or by choosing the bin edges to be equal logarithmic distances apart) and then dividing the counts in each bin by the linear width of the bin prior to graphing (Fig. 1.3B). The logarithmic scaling of the bin sizes decreases the number of bins with zero counts (often to zero) and the division by the linear width of the bin preserves the underlying shape of the relationship. Another, equally valid approach is to visualize the relationship using appropriate transformations of the cumulative distribution function (Fig. 1.3C; see White et al. 2008 for details), but we have found that it is often more difficult to intuit the underlying form of the distribution from this type of visualization and therefore recommend normalized logarithmic binning in most cases.

1.3 FITTING MTE MODELS TO DATA

1.3.1 Basic fitting

Since the two basic functional relationships of metabolic theory can be readily written as linear relationships by log-transforming one or both axes, most analyses use linear regression of these transformed variables to estimate exponents, compare the fitted values to those predicted by the theory, and characterize the overall quality of fit of the metabolic models to the data. Given the most basic set of statistical assumptions, this is the correct approach.

Specifically, if the data points are independent, the error about the relationship is normally distributed when the relationship is properly transformed (i.e., it is multiplicative log-normal error on the untransformed data):

$$\log(y) = \log(c) + b\log(M) + \varepsilon, \varepsilon \sim \mathrm{N}(0, \sigma^2) \quad (1.4a)$$

$$y = cM^b e^{\varepsilon}, \varepsilon \sim \mathrm{N}(0, \sigma^2) \quad (1.4b)$$

and there is error (i.e., stochasticity) only in the y-variable, then the correct approach to analyzing the component relationships is ordinary least-squares regression.

Given the same basic statistical assumptions, analyzing the full relationship including both size and temperature should be conducted using multiple regression with the logarithm of mass and the inverse of temperature as the predictor variables. This approach is superior to the common practice of using simple regression after correcting for the influence of the other variable (see, e.g., Gillooly et al. 2001; Brown et al. 2004) because it appropriately allows for correlation between the predictor variables, thus yielding the best simultaneous estimates of the parameters for each variable and the appropriate estimates of the confidence intervals for those parameters (Freckleton 2002; Downs et al. 2008).

In many cases the assumptions underlying these basic statistical analyses may be reasonable, and these methods are often robust to some violations of the assumptions. However, there are also a number of instances in common MTE analyses where substantial violations of assumptions related to the independence of data points, and even the basic form of the error about the relationship, may necessitate the use of more complex methods to obtain the most rigorous results.

1.3.2 Log-transformation vs. nonlinear regression

While most analyses utilize the fact that log-transforming one or both sides of the equation yields a linear relationship, allowing appropriately transformed data to be modeled using linear regression (log-linear regression), it has recently been suggested that analysis on logarithmic scales is flawed and that, instead, analysis should be carried out on the original scale of measurement using nonlinear regression (e.g., Packard and Birchard 2008; Packard and Boardman 2008, 2009a, 2009b; Packard 2009; Packard et al. 2009, 2010).

One fundamental difference between log-linear regression and nonlinear regression on untransformed data lies in the assumptions that the two approaches make about the nature of unexplained variation. In nonlinear regression the error term (i.e., residuals) is assumed to be normally distributed and additive, $y = \alpha x^b + \varepsilon$, $\varepsilon \sim \mathrm{N}(0, \sigma^2)$, while log-linear regression assumes the error term is log-normally distributed and multiplicative (equation 1.1). The form of the error distribution in the empirical data determines which method performs better, with the method that assumes the appropriate error form (i.e., nonlinear regression with additive error, and log-linear regression with multiplicative error) yielding the best results (Xiao et al. 2011).

Throughout this chapter we recommend that the form of the error distribution be explicitly considered, when possible, in deciding which methods to use (Cawley and Janacek 2010; Xiao et al. 2011). However, log-normal error is substantially more common than normal error in physiological and morphological data (Fig. 1.4: Xiao et al. 2011; see also Gingerich

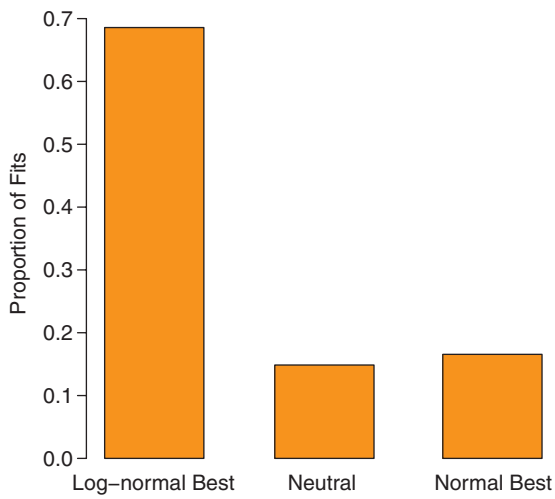

Figure 1.4 Likelihood analysis comparing the fits of normal vs. log-normal error to 471 biological power laws shows that most morphological and physiological relationships are better characterized by multiplicative log-normal error and therefore that traditional log-transformed regression is better in most cases than nonlinear regression (Xiao et al. 2011).

2000; Kerkhoff and Enquist 2009; Cawley and Janacek 2010), which implies that for most metabolic theory analyses log-linear regression is appropriate. This is good news because log-linearity allows the implementation of some approaches discussed below which cannot readily be implemented in a nonlinear context.

1.3.3 Alternatives to ordinary least-squares regression

The ordinary least-squares (OLS) approach is just one of several available choices for fitting a linear relationship between X and Y variables, with each method making different assumptions about the variation around the regression line. Understanding which of these methods to use can seem complicated because these choices depend on information about the sources and magnitude of variability around the regression line, the nature of the relationship between X and Y, and the goal of the analysis. In addition, there is conflicting advice in the literature regarding when to use which method, and uncertainty about best practice has led to many studies reporting regression slopes determined using more than one approach (e.g., Coomes et al. 2011).

The main alternative to OLS regression is commonly known as reduced major axis (RMA) regression. Whereas OLS assumes that residual variation occurs only in the vertical direction, RMA allows for variation also in the horizontal direction by minimizing the sum of the products of deviations in the vertical and horizontal directions. For most datasets, slopes estimated by RMA are steeper than those estimated by regression (Smith 2009). Other alternatives include major axis (MA), which generates estimates of the slope that are intermediate between RMA and OLS regression, and the OLS bisector, which determines the average of the slope of X on Y and the slope of Y on X (Isobe et al. 1990). OLS, RMA, and MA are all special cases of a general model in which the ratio of the error variances in X and Y can take on any value (Harvey and Pagel 1991; M.P. O'Connor et al. 2007a).

A common argument for the use of alternatives to OLS in allometric studies is that it is inappropriate to assume that X is measured without error, as implied in OLS regression (e.g., Legendre and Legendre 1998). However, this argument relies on the assumption that all of the variation about the regression line is due to measurement error, which is unlikely to be the case in biological systems (e.g., Sokal and Rohlf 1995; Smith 2009) and even then this argument is not valid in most situations (Warton et al. 2006; Smith 2009). Recent advice regarding when it is appropriate to use RMA (or a related alternative) vs. OLS is based on a combination of the goal of the analysis and the causal relationship between the variables (Warton et al. 2006; O'Connor et al. 2007a; Smith 2009). For an excellent treatment of the logic behind RMA vs. OLS see Smith (2009). All line-fitting techniques discussed can be implemented using the SMATR package in R (http://www.bio.mq.edu.au/ecology/SMATR/).

1.3.4 Which method(s) should I use?

Our interpretation of the recent discussion on which method to use is that, for the majority of cases in metabolic theory, OLS regression on log-transformed data is the correct approach. Most analyses in metabolic theory are causal in nature – the hypothesis is that the size and temperature of an organism determine a broad suite of dependent variables. In the case of hypothesized causal relationships we are logically assigning all equation error (i.e., variability about the line not explained by measurement error; Fuller 1987; McArdle 2003) to the Y variable and therefore should be estimating the form of the relationship using OLS (Warton et al. 2006; Smith 2009). In addition to causal relationships, OLS regression is also most appropriate in cases where one wants to predict unknown values of Y based on X (Sokal and Rohlf 1995; Warton et al. 2006; Smith 2009). Metabolic theory is often used in this context to estimate the metabolic rate of individuals based on body size (e.g., Ernest and Brown 2001; White et al. 2004; Ernest et al. 2009). The fact that OLS is appropriate for many metabolic theory predictions is convenient because variants on simple bivariate relationships (e.g., phylogenetic correction, mixed effects models) are typically based on OLS.

There are some cases where directional causality between the two variables being analyzed is not implied by metabolic models. For example, predictions for the relationships between different measures of size (e.g., height and basal stem diameter in trees) do not imply a direct causal relationship between the variables but an "emergent" outcome of a process affected by two interdependent variables. In this case, the choice of

which variable to place on the *x*-axis is arbitrary. In this case (and in many similar cases in other areas of allometry; e.g., the leaf economics spectrum) RMA or a related approach is more appropriate for analysis because we want to partition equation error between X and Y, rather than assigning it all to Y.

1.3.5 Phylogenetic methods

A common goal of analysis in metabolic ecology is to understand the relationship between two morphological, physiological, or ecological properties, across species. The data points in these analyses are typically average values of the two properties for each species, which leads to a potential complication. Because there are limits to how quickly traits can evolve, closely related species may not be statistically independent due to their shared evolutionary history. This lack of independence among data points violates a key assumption of ordinary least-squares regression (and general linear models more broadly).

The problem of phylogenetic non-independence is well known in evolutionary biology, and a method known as independent contrasts (Felsenstein 1985) remains popular for correcting for the phylogenetic signal in comparative data. Independent contrasts have been recently superseded by phylogenetic general linear models (PGLMs), which allow a wide range of evolutionary scenarios to be modeled (Garland and Ives 2000).

The current implementation of PGLMs was devised by Mark Pagel (Pagel 1997, 1999). There are three parameters, λ, κ, and δ, each of which can be specified a priori or estimated from the data. The most important of these is λ, which is a measure of the strength of the phylogenetic signal in the data. Suppose some trait(s) have been measured in five species for which an evolutionary tree, i.e., phylogeny, is available, as shown in Figure 1.5A. If the pattern of trait variation among these species is consistent with random evolutionary change along the branches of the phylogeny, then λ is said to be 1. At the other extreme it is possible that close relatives are no more similar to each other than distantly related species. It is then as if all species were completely independent, equally distant phylogenetically from their common ancestor, as shown in Figure 1.5B. In this case λ is said to be 0. Most analyzed cases fall in between these two extremes and find that some proportion λ of the variation is accounted for by the phylogeny, the rest being attributable to recent independent evolution, as in Figure 1.5C. Parameters κ and δ provide a way of scaling the rates of evolutionary change along the branches of the phylogeny. For example, $\kappa = 1$ corresponds to gradual evolution, and $\kappa = 0$ is a model in which evolution is concentrated at speciation events. Parameter δ, which is rarely used, measures whether the rates of evolution have increased, decreased, or stayed constant over time. The best mathematical account of the method is provided by Garland and Ives (2000, p. 349) where it is referred to as the generalized least-squares approach. A recent guide to the use and misuse of PGLMs is given in Freckleton (2009).

The traits of interest in metabolic scaling analyses tend to show strong phylogenetic signals. For example, in mammals, $\lambda = 0.984$, 1.0, and 0.84 for basal metabolic rate, mass, and body temperature, respectively

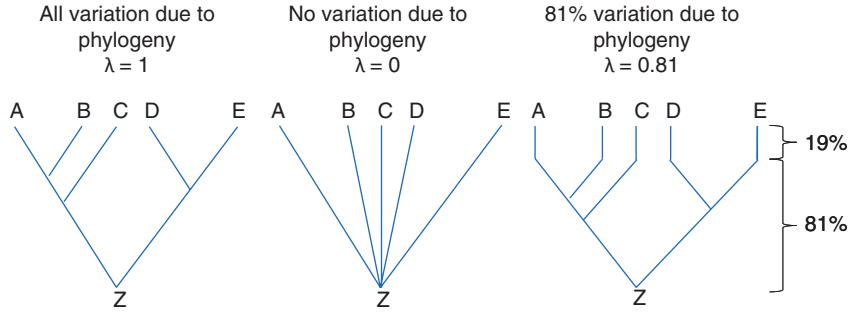

Figure 1.5 The Pagel λ approach to modeling the evolution process. (A) shows the phylogeny of five species A–E, which are descended from a common ancestor Z ; (B) shows how evolution is modeled if the species appear to be independent; (C) shows the type of intermediate model currently used (the Pagel λ model).

(Clarke et al. 2010). However, estimates of scaling parameters from PGLMs and conventional GLMs tend to be similar, converging on the same answer when the explanatory power (R^2) approaches 1.

Despite their promise, PGLMs are currently difficult to use. They require that a phylogeny, ideally with branch lengths, be available or assembled for the species of interest. They also assume that the form of the phylogeny and the assumed models of evolution are accurate. However, little analysis has been done to determine the impacts of error in either of these inputs on the outcome of the analysis. In addition, while software is available for conducting PGLM analyses, including BayesTraits (http://www.evolution.rdg.ac.uk/BayesTraits.html) and several packages in R including ape (http://cran.r-project.org/web/packages/ape/) and caper (http://r-forge.r-project.org/projects/caper/; Orme et al. 2011), the documentation is fragmentary and utilizing these packages can be difficult for new users.

In general we recommend that PGLMs be used when quality phylogenies are available. However, in cases where the relationship between two variables is strong this is unlikely to have a demonstrable influence on the results. If no phylogeny is available, an alternative is to use taxonomy as a proxy for phylogeny in a hierarchical (mixed effects) model (e.g., McGill 2008; Isaac and Carbone 2010). We also caution that factors other than phylogenetic relationship, such as similar body size or environment, can potentially be additional causes of non-independence of data in species-level analyses.

1.3.6 Methods for fitting frequency distributions

The predicted form of MTE frequency distributions is typically power law, $f(x) = cx^{\theta}$ (also known as the Pareto distribution in the probability and statistics literature), and the fit of these predictions to empirical data has typically been evaluated by fitting a regression through the data generated using histograms for visualization (i.e., binning the values of the variable of interest, counting how many values occur in each bin, log-transforming the counts and the position of the bin, and then fitting a relationship to those data points using linear regression). An example of this would be fitting a regression through the points in Figure 1.3A or 1.3B. While this seems like a reasonable approach to this problem, it fails to properly account for the structure of the data, which can result in inaccurate parameter estimates (Clark et al. 1999; Edwards 2008; White et al. 2008) and incorrect estimates of the quality of fit of the model to the data (Newman 2005; Edwards et al. 2007; Clauset et al. 2009).

The correct approach for fitting frequency distributions in metabolic theory to data is based on likelihood (Edwards et al. 2007; White et al. 2008). Maximum likelihood estimation determines the values of the parameters that maximize the likelihood of the model, given the data. In the case of the metabolic theory this is typically finding the best-fitting exponent of a power-law frequency distribution. Determining the best parameters using maximum likelihood estimation for power laws is straightforward in most cases, requiring only a simple calculation. In the most common case where there is a meaningful lower bound (e.g., trees <1 cm are not measured) and the upper bound is assumed to be infinite, the exponent is determined simply by

$$\theta = -1 - \left[\frac{1}{n}\sum\left(\log\left(\frac{x}{x_{\min}}\right)\right)\right]^{-1} \tag{1.5}$$

where the summation is over all values of x. In other cases the calculations may be different, so care is required to confirm that assumptions being used to determine the MLE of the parameters are consistent with the data to which the calculation is being applied. In the case of the power-law frequency distributions predicted by metabolic theory, MLEs for all possible detailed forms are available in White et al. (2008; see Johnson et al. 1994, 2005, for more technical treatments).

1.4 ARE THE FITTED PARAMETERS CONSISTENT WITH THEORETICAL PREDICTIONS?

Having fit a relationship of the same general form as the MTE predictions using the methods above, the next step in evaluating the MTE is to determine whether the fitted parameters are consistent with the specific quantitative predictions of the theory.

In regression-based analyses this is typically done by determining whether or not the 95% confidence

interval (CI) about the best-fitting parameter includes the theoretical prediction. This is a well-established practice and easy to apply (most statistical software that will generate parameter estimates will also generate confidence intervals for those estimates). However, hypothesis testing of this kind is not intended to determine whether two values of a parameter are similar. The appropriate interpretation of a CI containing the theoretical value is that we cannot reject the model, but this is not the same as supporting it. Alternatives that focus on determining whether or not two values are meaningfully similar are available (i.e., equivalence testing; Dixon and Pechmann 2005) but have never been applied to metabolic theory and are only rarely used in ecology in general.

Comparing the parameters of frequency distributions to those predicted by theory can also be done using confidence intervals, which can be determined accurately for all forms of power-law distribution when the number of data points is large (see appendix in White et al. 2008) and for small sample sizes for the most common form of the distribution (the Pareto; Johnson et al. 1994; Newman 2005; Clauset et al. 2009). Confidence intervals can also be calculated using bootstrap or jackknife techniques if necessary (Newman 2005). An alternative approach is to explicitly test whether a distribution with a fitted value provides a meaningfully better fit to the data than one with the theoretical value. This can be done using likelihood ratio tests (Vuong 1989; Clauset et al. 2009).

1.5 IS THE SHAPE OF THE RELATIONSHIP CONSISTENT WITH THEORETICAL PREDICTIONS?

1.5.1 Goodness-of-fit tests

For frequency distributions it is possible to directly ask whether or not the observed form of the distribution is consistent with (i.e., not significantly different from) the form predicted by the theory. This is done using goodness-of-fit tests, where the null hypothesis is that observed data are drawn from the theoretical distribution. A number of goodness-of-fit tests are available that entail different detailed assumptions including the chi-square test, the Kolmogorov–Smirnov test, and the G-test. If the sample size is sufficiently large and data are continuously distributed, all of these tests should give similar answers.

Determining the goodness of fit of regressions is more complicated and therefore simple tests are not available. Instead, it is standard to evaluate several assumptions of regression to determine whether the regression should be used or to compare the fits of linear regressions to more complex models (see below). Evaluating the assumptions of regression is good general practice, and failure to satisfy these assumptions can indicate that the model is not sufficient for characterizing the pattern in the data. Specifically the two most relevant tests are to determine: (1) whether or not the residuals about the regression are normally distributed (which can be done using standard goodness-of-fit tests); and (2) whether the variance of the residuals does not change as a function of the value of the predictor variable (i.e., the residuals are homoskedastic).

1.5.2 Comparison to alternative models

The other approach to determining whether or not the observed relationship has the same shape as the predictions of MTE is to compare the fit of the relationship or distribution to alternative models. The most common example of this is the use of polynomial regression to determine whether or not a simple linear relationship (among log-transformed variables) is an appropriate fit to the data. The standard approach is to fit polynomial regressions that include one or more higher-order terms (x^2, x^3, etc.) and determine whether or not those terms are significant in the regression. If they are, this is typically considered to be an indication that a different, or more complex, model than the simple linear relationship (on log-transformed data) is necessary. This polynomial approach has only rarely been used in MTE analyses, perhaps for reasons discussed below (see section 1.6), but it has been successfully utilized to indicate that the current metabolic theory predictions for the relationship between temperature and species richness are not sufficient to fully characterize the observed patterns (Algar et al. 2007; Hawkins et al. 2007; but see Gillooly and Allen 2007).

A more general approach is to use likelihood and information criteria-based methods. These methods determine which of a set of models is most consistent with the empirical data and whether that model provides a meaningfully better fit than alternative models

(Hilborn and Mangel 1997; Burnham and Anderson 2002). A full introduction to this area is beyond the scope of this chapter, but the basic approach is to calculate the likelihoods of all the candidate models and then compare those likelihoods to one another, taking into account that some models have more parameters than others and are therefore more likely to provide good fits to empirical data (for ecological examples see Muller-Landau et al. 2006a and Coomes and Allen 2007). We strongly recommend Hilborn and Mangel (1997) to those looking for an accessible introduction to this area of statistics. Equivalent Bayesian methods are also available, but have rarely been applied in the context of metabolic ecology. Good examples are available in Dietze et al. (2008) and Price et al. (2009) for those interested in this approach.

In addition to testing the basic shape of the predicted relationship and the specific parameter values, these methods can be used to assess the form of the error distribution to allow for decisions to be made about whether to use log-linear or nonlinear regression (Xiao et al. 2011; see above) and to determine the degree of phylogenetic non-independence among data points that needs to be accounted for (Freckleton 2009).

1.6 THOUGHTS ON TESTING ECOLOGICAL THEORIES

It is useful and informative to compare the fits of metabolic theory models (and ecological models in general) to alternative models to see if a better characterization of the empirical data is possible. If an alternative model provides a better fit to the data there are two different conclusions that can be drawn: (1) the model is not useful and should be abandoned; or (2) the model is incomplete and requires further development. In ecology we have tended to prefer the language of rejection – any model for which data deviates from the prediction using a goodness-of-fit test, or for which an alternative model is found to provide a superior fit, is rejected. This attitude likely has its origins in an emphasis on Plattian inference (Platt 1964) and an, arguably improper (Hurlbert and Lombardi 2009), emphasis on the arbitrary definition of $p < 0.05$ as being "significant." Further discussion of how a rejected model may be improved is rarely undertaken. However, in cases where a model is based on reasonable starting assumptions and makes reasonable predictions, it may be better to modify and improve that model than to abandon it. This iterative process of hypothesis refinement is considered essential for the development of ecology (Mentis 1988), and several recent attempts to refine models from metabolic ecology make valuable contributions to this process (Banavar et al. 2010; Savage et al. 2010).

The goal of theory is to provide simplified characterizations of reality; so rejecting models is only useful if it leads to better models. Testing models and identifying their flaws is a necessary, but not sufficient, part of the process. This raises questions about the merits of comparing process-based models to purely phenomenological models that lack a biological mechanism. Consistent, directional, deviations from a general theory indicate that the theory is either incomplete or simply wrong. However, studies that only demonstrate the superior performance of phenomenological over mechanistic models often yield little direct progress towards acceptable theories. In contrast, comparing theoretical predictions to mechanistic models that include either additional or alternative processes has the potential to yield improved characterizations of biological systems. An illustrative example is Fisher's sex ratio theory, which predicts a canonical ratio of 1 : 1. When sample sizes are large, significant deviations are almost always observed. This does not mean the theory is wrong. Indeed, considering the direction and magnitude of the deviations (large in eusocial hymenoptera, small in humans) leads to progress in understanding the additional processes that affect sex ratios in real populations.

It is important to consider the goal of a model when determining whether it should be replaced or modified (Martinez del Rio 2008). For example, in many cases related to MTE the goal is to understand the fundamental processes that produce the first-order relationship between body size and metabolic rate. MTE is successful at characterizing the relevant empirical pattern, because a 3/4-power allometric relationship is the best-supported pattern, both when analyzing large numbers of species and when the average form of the model across taxonomic groups is determined (Savage et al. 2004b; Isaac and Carbone 2010). As such, MTE may provide information about the underlying process. However, if the goal is to accurately predict the metabolic rate of species for which data is not available then it is necessary to consider the empirical evidence of variation among taxonomic groups (e.g., Nagy et al.

1999; Isaac and Carbone 2010). In this case models that incorporate taxonomic variation are an important improvement over the more general MTE (Isaac and Carbone 2010).

Evaluating models is further complicated by the fact that general ecological theories (including MTE) typically make predictions for multiple empirical patterns (see Brown et al. 2004). This generality is desirable because it makes metabolic theory applicable in a broad range of situations, but it also makes MTE easier to reject since rejection of any prediction implies rejection of the entire theory. However, it is unreasonable to compare a model that makes a large number of predictions to a model that makes one or a few specific predictions without penalizing the more specific model for its lack of generality and resultantly larger number of parameters per prediction (Price et al. 2009). Unfortunately there are no general approaches for dealing with this type of difference among models, and the one example that we are aware of (Price et al. 2009) represents a first attempt rather than a general solution to the challenge of evaluating models that make multiple predictions.

In conclusion, the goal of science is to develop models of reality that both capture general underlying patterns and processes, and incorporate important biological details. Developing general ecological theories allows us to understand how ecological systems operate and make predictions for how they will respond to global change and other major perturbations. Rigorous statistical approaches and proper testing of theories are necessary to accomplish this result. Efforts to improve methodological approaches and to use these approaches to test existing theories should always be undertaken with the goal of improving our understanding of ecological systems.

ACKNOWLEDGMENTS

We thank Jamie Gillooly for providing the data for the metabolic rate graphs, and Alwyn Gentry and the Missouri Botanical Garden for making the data for the frequency distribution graphs publicly available. James Brown, Astrid-Kodric Brown, and Alison Boyer provided valuable comments that improved the chapter.