Estimating and comparing thermal performance curves

Michael J. Angilletta Jr*

Department of Ecology and Organismal Biology, Indiana State University, Terre Haute, IN 47809 USA

Received 3 March 2006; accepted 10 June 2006

Abstract

I show how one can estimate the shape of a thermal performance curve using information theory. This approach ranks plausible models by their Akaike information criterion (AIC), which is a measure of a model's ability to describe the data discounted by the model's complexity. I analyze previously published data to demonstrate how one applies this approach to describe a thermal performance curve. This exemplary analysis produced two interesting results. First, a model with a very high r^2 (a modified Gaussian function) appeared to overfit the data. Second, the model favored by information theory (a Gaussian function) has been used widely in optimality studies of thermal performance curves. Finally, I discuss the choice between regression and ANOVA when comparing thermal performance curves and highlight a superior method called template mode of variation. Much progress can be made by abandoning traditional methods for a method that combines information theory with template mode of variation.
© 2006 Elsevier Ltd. All rights reserved.

1. Introduction

In a recent paper, Bulté and Blouin-Demers (2006) raised two points concerning the estimation and analysis of thermal performance curves. First, they noted thermal performance curves have a natural shape, which one should heed when seeking a statistical description. Second, they argued the estimation of the thermal optimum by ANOVA ignores important variation among individuals within populations. These points merit serious consideration by thermal biologists because they should influence the manner in which we estimate and compare thermal performance curves. In fact, these issues extend far beyond the study of thermal performance curves and have played a prominent role in previous debates among biologists. Although the outcomes of these debates bear directly on the problems of estimating and comparing thermal performance curves, the approaches used by thermal biologists have not changed much in several decades. Here, I prescribe an approach that will enable thermal

biologists to avoid the pitfalls described by Bulté and Blouin-Demers.

2. What is the true shape of a thermal performance curve?

Bulté and Blouin-Demers (2006) noted thermal performance curves have a characteristic shape. Indeed, other thermal biologists (including myself) have suggested a general form can describe these curves (Huey and Stevenson, 1979; Huey and Kingsolver, 1989; Angilletta et al., 2002b). Typically, performance curves are bounded at extreme temperatures and possess a single intermediate mode. As Bulté and Blouin-Demers suggest, we should avoid linear (or other) approximations of thermal performance curves that differ fundamentally from the characteristic shape. As an example, they compared the fit of linear and nonlinear functions describing the relationship between body temperature and food intake by lizards (data from McConnachie and Alexander, 2004). Although a linear function fit well over the range of 20–32 °C, a nonlinear function described the data better over a broader range of temperatures (20–35 °C). Bulté and Blouin-Demers criticized McConnachie and Alexander (2004) for omitting the data recorded at 35 °C and using a linear model to describe

*Fax: + 1 812 237 2526.
E-mail address: mangilletta@indstate.edu.*

the relationship. I feel certain that no one, including McConnachie or Alexander, would seriously expect the linear function to describe performance outside the range of values used to fit the model; indeed, this assumption would be unrealistic for any linear model. Still, Bulté and Blouin-Demers raised an important issue for thermal biologists to consider. How should we decide which statistical model best describes a thermal performance curve?

Naively, we might fit several functions to our data and choose the one that describes the greatest amount of variation (i.e., the function that has the smallest residual sum of squares (RSS) or the highest r^2). Although this function would certainly describe our data well, it may not reflect the mean performance curve of the individual or population that we sampled. Consequently, any conclusions drawn or predictions made from this function could be erroneous. Why? The problem stems from over-fitting the data (Burnham and Anderson, 2002). A model with more parameters typically describes more variation than a model with fewer parameters. Most likely, however, temperature is not the only source of variation in performance; non-random sampling and measurement error also cause variation in performance. Therefore, a function that describes 100% of the variation does not actually describe the true performance curve. What we desire is a function that describes the curve without fitting the noise in the data.

More sophisticated approaches exist for selecting the appropriate statistical model when we do not know the true model. Two interrelated approaches are information theory and Bayesian theory (Burnham and Anderson, 2004; Ellison, 2004; Johnson and Omland, 2004). Both approaches enable one to select the best function from a set of plausible candidates without over-fitting the data (Ellison, 1996; Johnson, 1999; Anderson et al., 2000). These approaches differ in an important way: the Bayesian method enables one to consider prior information as well as the new data when selecting a model. Prior information might be available from similar studies of closely related species or pilot studies of the same species. When no prior information is used, certain Bayesian and information-theoretic approaches will yield identical results (Ellison, 2004). For this reason, I shall focus on the information-theoretic approach.

To find the best model, we must calculate the Akaike information criterion (AIC) for each model under consideration (Burnham and Anderson, 2002). The AIC is calculated as follows:

$$AIC = -2L + 2K + \frac{2K(K+1)}{N-K-1}, \qquad (1)$$

where L is the maximized log-likelihood value of the model, K is the number of parameters (including the error term), and N is the sample size. The maximized log-likelihood value of a model can be computed easily from the model's RSS:

$$L = \log\left(\frac{RSS}{N}\right)\frac{-N}{2}, \qquad (2)$$

The AIC estimates the information lost when using a particular model to describe the data (Burnham and Anderson, 2002). Hence, we seek the model with the lowest AIC, which is the function that minimizes our loss of information (or best describes the data). Note the AIC differs from a simple measure of fit, such as the r^2, because it also depends on the model's complexity (i.e., the number of parameters). The second and third terms in Eq. (1) eliminate the bias in fit associated with more complex models. By correcting this bias, we avoid choosing a model that overfits the data.

To illustrate the value of information theory for modeling thermal performance curves, I reanalyzed data from a previous study in which my colleagues and I measured the locomotor performance of lizards over a wide range of body temperatures (Angilletta et al., 2002a). In our original analysis, we avoided the problem of model selection by adopting the minimum convex polygon approach (van Berkum, 1986). Here, I fit these data to five functions, each of which could be considered a plausible model of the data (Fig. 1). Three of these functions—the Gaussian, Quadratic, and Weibull functions—have been used to theoretically or empirically describe thermal performance curves (e.g., see Huey and Kingsolver, 1993; Huang and Yang, 1995; Palaima and Spitze, 2004). The remaining functions—the modified Gaussian and the exponentially modified Gaussian functions—were chosen because their complex structure should provide a better fit to nonlinear data. First, I fit each model using Table Curve (version 5.01; Systat Software, Inc., 2002). Then, I calculated the AIC and related statistics for each model (Table 1). Interestingly, each of the two complex models was extremely unlikely to be the best model in the set, despite an extremely high value of r^2. The modified Gaussian function provides a great example of
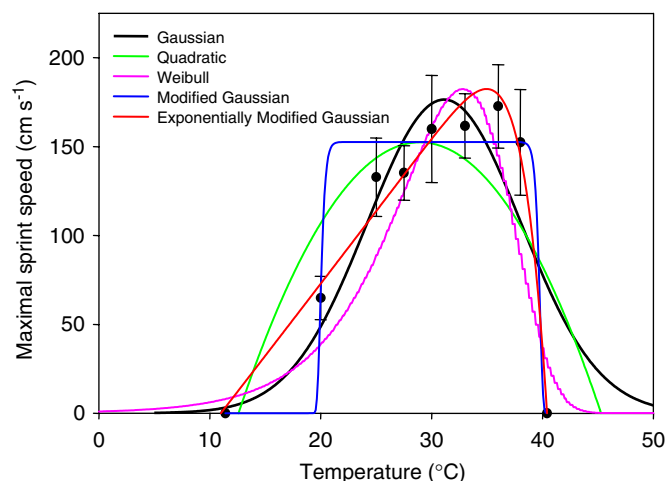

Fig. 1. The relationship between body temperature and sprint speed was described using five functions. A comparison of AIC values indicated the Gaussian function provides the best fit of the data (see Table 1). Each datum is the mean of the sprint speeds of 12 individuals, whose performance was measured at all temperatures (i.e., a repeated measures design); data were taken from Angilletta et al. (2002a).

Table 1
A comparison of plausible functions to describe the relationship between temperature (T) and performance (P)

Function	K	AIC	Δ_i	w_i	r^2
Gaussian: $P = a\mathrm{e}^{[-0.5(|T-b|/c)^2]}$	4	45.20	0	0.62	0.70
Quadratic: $P = aT^2 + bT + c$	4	46.63	1.43	0.30	0.64
Modified Gaussian: $P = a\mathrm{e}^{[-0.5(|T-b|/c)^d]}$	5	49.23	4.03	0.08	0.95
Weibull: $P = a\left(\frac{d-1}{d}\right)^{1-d/d}\left[\frac{T-b}{c} + \left(\frac{d-1}{d}\right)^{1/d}\right]^{d-1} e^{-[(T-b/c)+(d-1/d)^{1/d}]^d} + \frac{d-1}{d}$	5	56.96	11.96	0.00	0.77
Exponentially modified Gaussian: $P = \frac{ac\sqrt{2\pi}}{2d}e^{(b-T/d)+(c^2/2d^2)}\left[\frac{d}{|d|} - \mathrm{erf}\left(\frac{b-x}{\sqrt{2}c} + \frac{c}{\sqrt{2}d}\right)\right]$	6	71.41	26.21	0.00	0.98

For each model, I report both the AIC and the differential AIC (Δ_i), which is the difference between a given model's AIC and the lowest AIC. I also report the Akaike weight (w_i), which is the normalized likelihood that the model is the best one in the set; the Akaike weights help us decide whether we feel confident that the function with the lowest AIC is actually the best function in the set.

overfitting the data; this model described 95% of the variation in performance, but I doubt the performance breadth is really as wide as the model suggests. My gut instinct leads me to favor the exponentially modified Gaussian function, with its familiar left-sided skew and its superior r^2. Despite my feeling, this model does not provide a sufficiently better fit to justify its complexity. Because the calculation of AIC depends on sample size (see Eq. (1)), additional data could cause us to rank this model higher in the future. For now, however, the best model appears to be a simple Gaussian function. This result should be somewhat satisfying because theorists often use a Gaussian function to model the evolution of thermal performance curves (Lynch and Gabriel, 1987; Gabriel and Lynch, 1992; Huey and Kingsolver, 1993).

The superior likelihood of the Gaussian model should trouble astute thermal biologists because thermal performance curves typically appear skewed. Indeed, a theoretical basis for this skewness even exists (Sharpe and DeMichele, 1977; Schoolfield et al., 1981). Given these facts, we should endeavor to collect sufficient data to fit a complex model accurately. In my example, the small sample ($N = 9$) imposed a severe bias correction because smaller samples more likely lead to overfitting. A much larger sample might have not only made the exponentially modified Gaussian model more likely but might have also made the modified Gaussian model less likely (assuming additional data would indicate a narrower performance breadth). If we are unable to collect a large sample, we might decide to adopt a Bayesian approach, which would enable us to use prior information about the shape of a performance curve (Ellison, 2004). Regardless of which approach we adopt, we should avoid using the data to specify constants, such as the thermal optimum or critical thermal limits (e.g., see Hertz et al., 1983). This procedure poses a serious risk of overfitting because it forces the function through a certain point, which probably differs from the least-squares or maximum likelihood estimate of the parameter. Conceivably, we could correct this bias by adding an extra parameter for each "fudged" constant when calculating AIC. Nevertheless, we would do better to fit all parameters using a least-squares or maximum

likelihood procedure to avoid biasing the fit of the function.

3. How should we compare thermal performance curves? Regression versus ANOVA

Thermal biologists routinely employ one of two methods to compare thermal performance curves: regression and ANOVA (Bulté and Blouin-Demers, 2006). When taking the regression approach, we would fit a function to the data for each individual and estimate a thermal optimum from the resulting model. Given a thermal optimum for each individual, we could compute a mean and confidence interval for the population. Furthermore, we could use a linear or nonlinear regression analysis to compare thermal optima (or other parameter values) between populations; Peek and his colleagues (2002) described a nonlinear regression analysis designed for repeated measures, which we commonly deal with in studies of thermal performance curves. When taking the ANOVA approach, we would conduct a repeated measures ANOVA and post hoc comparisons to identify the temperature at which performance was maximal; often, performance at two or more temperatures cannot be considered significantly different and these temperatures would define an optimal temperature range (Huey and Stevenson, 1979). This range likely contains the true mean of the thermal optimum for the population. To compare thermal optima between populations, we would examine the interaction between population and temperature and make post hoc comparisons of performance between populations at each temperature. Of course, the confidence intervals of the mean performance at each temperature will determine the outcome of these comparisons.

Bulté and Blouin-Demers criticized the ANOVA approach for being sensitive to variation in the thermal performance curve among individuals. They provided a simple example in which ANOVA was unable to distinguish between a population whose individuals exhibited different thermal optima and a population whose individuals shared a single thermal optimum. Based on this example, they argued that the regression approach better

describes variation in thermal performance curves than the ANOVA approach does. In fact, biologists already debated the relative merits of these approaches in the 1990's when they began to quantify genetic variation in phenotypic plasticity (reviewed by Via et al., 1995). In these debates, the regression approach and the ANOVA approach were labeled the polynomial approach and the character-state approach, respectively. Ironically, this debate was put to rest in 1995, when de Jong proved the two approaches are mathematically equivalent (de Jong, 1995). In other words, one should obtain identical results when comparing thermal performance curves by polynomial regression or by ANOVA. Since a nonlinear function cannot easily be transformed into character states, nonlinear regression does differ from ANOVA.

Why does the example provided by Bulté and Blouin-Demers appear to favor the use of regression? Possibly, this result stems from their choice of a nonlinear function. But more likely, the answer lies in the false assumption that regression estimates the thermal optima of individuals without error. In fact, error arises from two sources. First, one generally has a finite sample of observations recorded under imperfectly controlled conditions. Second, one must describe the data with a function that probably differs from the true form of the performance curve. Both the choice of the function and the estimation of its parameters produce error in the thermal optimum (intra-individual error). A third source of error arises from sub-sampling individuals from a population (inter-individual error). Both regression approach and ANOVA involve intra- and inter-individual sources of error. In the regression approach, this error is reflected in the confidence interval of the mean thermal optimum for the population. In the ANOVA approach, this error is reflected in the confidence intervals of performance at each temperature. In the example provided by Bulté and Blouin-Demers, the ANOVA approach appears less powerful than the regression approach because large confidence intervals at each temperature prevent one from pinpointing the mean thermal optimum with precision. But their confidence interval from the regression approach was artificially deflated because their data was generated from a known function without introducing error. If they had introduced intra-individual error, I suspect the regression approach and the ANOVA approach would have performed similarly.

For real data, one might favor regression or ANOVA depending on the situation. When performance can be measured at only a few temperatures for each individual, a repeated measures ANOVA might be advantageous. For small samples, estimates of the thermal optimum could be very sensitive to the choice of regression model, particularly when the performance curve is rather broad (e.g., see the modified Gaussian curve in Fig. 1). In contrast, ANOVA assumes nothing about the form of the relationship between independent and dependent variables (Steury et al., 2002); inferences would be restricted to performance at the specific temperatures included in the study. When performance can be measured at many temperatures for each individual, regression seems advantageous because an information-theoretic approach could be used more effectively to choose the best model. In fact, one could determine the best model for each individual, assuming two or more models fit better or worse for certain individuals (see Hertz et al., 1983).

Recently, a superior alternative to both ordinary regression and ANOVA was developed. This method, called Template Mode of Variation, uses a polynomial function to decompose variation in thermal performance curves (Izem and Kingsolver, 2005). Variation in the thermal optimum (horizontal position of the curve) can be isolated from variation in the mean performance over all temperatures (vertical position of the curve). Consequently, the method eliminates potential artifacts caused by variation in mean performance among individuals. This method should be particularly powerful when combined with information theory to identify the best function for separating modes of variation.

Acknowledgments

References

Anderson, D.R., Burnham, K.P., Thompson, W.L., 2000. Null hypothesis testing: problems, prevalence, and an alternative. J. Wildlife Manage. 64, 912–923.

Angilletta, M.J., Hill, T., Robson, M.A., 2002a. Is physiological performance optimized by thermoregulatory behavior?: a case study of the eastern fence lizard, *Sceloporus undulatus*. J. Therm. Biol. 27, 199–204.

Angilletta, M.J., Niewiarowski, P.H., Navas, C.A., 2002b. The evolution of thermal physiology in ectotherms. J. Therm. Biol. 27, 249–268.

Bulté, G., Blouin-Demers, G., 2006. Cautionary notes on the descriptive analysis of performance curves in reptiles. J. Therm. Biol. 31, 287–291.

Burnham, K.P., Anderson, D.R., 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. Springer, New York.

Burnham, K.P., Anderson, D.R., 2004. Multimodel inference: understanding AIC and BIC in model selection. Sociol. Methods Res. 33, 261–304.

de Jong, G., 1995. Phenotypic plasticity as a product of selection in a variable environment. Am. Nat. 145, 493–512.

Ellison, A.M., 1996. An introduction to Bayesian inference for ecological research and environmental decision-making. Ecol. Appl. 6, 1036–1046.

Ellison, A.M., 2004. Bayesian inference in ecology. Ecol. Lett. 7, 509–520.

Gabriel, W., Lynch, M., 1992. The selective advantage of reaction norms for environmental tolerance. Issue Series Title: J. Evol. Biol. 5, 41–59.

Hertz, P.E., Huey, R.B., Nevo, E., 1983. Homage to Santa Anita: thermal sensitivity of sprint speed in agamid lizards. Evolution 37, 1075–1084.

Huang, C.-H., Yang, C.-M., 1995. Use of Weibull function to quantify temperature effect on soybean germination. Chin. Agron 5, 25–34.

Huey, R.B., Kingsolver, J.G., 1989. Evolution of thermal sensitivity of ectotherm performance. Trends Ecol. Evol. 4, 131–135.

Huey, R.B., Kingsolver, J.G., 1993. Evolution of resistance to high temperature in ectotherms. Am. Nat. 142, S21–S46.

Huey, R.B., Stevenson, R.D., 1979. Integrating thermal physiology and ecology of ectotherms: discussion of approaches. Am. Zool. 19, 357–366.

Izem, R., Kingsolver, J.G., 2005. Variation in continuous reaction norms: quantifying directions of biological interest. Am. Nat. 166, 277–289.

Johnson, D.H., 1999. The insignificance of statistical significance testing. J. Wildlife Manage. 63, 763–772.

Johnson, J.B., Omland, K.S., 2004. Model selection in ecology and evolution. Trends Ecol. Evol. 19, 101–108.

Lynch, M., Gabriel, W., 1987. Environmental tolerance. Am. Nat. 129, 283–303.

McConnachie, S., Alexander, G.J., 2004. The effect of temperature on digestive and assimilation efficiency, gut passage time and appetite in an ambush foraging lizard, *Cordylus melanotus melanotus*. J. Comp. Physiol. B 174, 99–105.

Palaima, A., Spitze, K., 2004. Is a jack-of-all-temperatures a master of none? An experimental test with *Daphnia pulicaria* (Crustacea: Cladocera). Evol. Ecol. Res. 6, 215–225.

Peek, M.S., Russek-Cohen, E., Wait, D.A., Forseth, I.N., 2002. Physiological response curve analysis using nonlinear mixed models. Oecologia 132, 175–180.

Schoolfield, R.M., Sharpe, P.J.H., Magnuson, C.E., 1981. Non-linear regression of biological temperature-dependent rate models based on absolute reaction-rate theory. J. Theor. Biol. 88, 719–731.

Sharpe, P.J.H., DeMichele, D.W., 1977. Reaction kinetics of poikilotherm development. J. Theor. Biol. 64, 649–670.

Steury, T.D., Wirsing, A.J., Murray, D.L., 2002. Using multiple treatment levels as a means of improving inference in wildlife research. J. Wildlife Manage. 66, 292–299.

van Berkum, F.H., 1986. Evolutionary patterns of the thermal sensitivity of sprint speed in *Anolis* lizards. Evolution 40, 594–604.

Via, S., Gomulkiewicz, R., de Jong, G., Scheiner, S.M., Schlichting, C.D., van Tienderen, P.H., 1995. Adaptive phenotypic plasticity: consensus and controversy. Trends Ecol. Evol. 10, 212–217.