ORIGINAL ARTICLE

doi:10.1111/j.1558-5646.2008.00340.x

HYPOTHESIS TESTING IN COMPARATIVE AND EXPERIMENTAL STUDIES OF FUNCTION-VALUED TRAITS

Cortland K. Griswold,[1,2,3] Richard Gomulkiewicz,[1,4,5] and Nancy Heckman[6,7]

[1]*School of Biological Sciences, Washington State University, Pullman, Washington 99164*

[2]*E-mail: ckg@email.arizona.edu*

[4]*Department of Mathematics, Washington State University, Pullman, Washington 99164*

[5]*E-mail: gomulki@wsu.edu*

[6]*Department of Statistics, University of British Columbia, Vancouver, British Columbia V6T 1Z2, Canada*

[7]*E-mail: nancy@stat.ubc.ca*

Received June 22, 2007
Accepted January 7, 2008

Many traits of evolutionary interest, when placed in their developmental, physiological, or environmental contexts, are function-valued. For instance, gene expression during development is typically a function of the age of an organism and physiological processes are often a function of environment. In comparative and experimental studies, a fundamental question is whether the function-valued trait of one group is different from another. To address this question, evolutionary biologists have several statistical methods available. These methods can be classified into one of two types: multivariate and functional. Multivariate methods, including univariate repeated-measures analysis of variance (ANOVA), treat each trait as a finite list of data. Functional methods, such as repeated-measures regression, view the data as a sample of points drawn from an underlying function. A key difference between multivariate and functional methods is that functional methods retain information about the ordering and spacing of a set of data values, information that is discarded by multivariate methods. In this study, we evaluated the importance of that discarded information in statistical analyses of function-valued traits. Our results indicate that functional methods tend to have substantially greater statistical power than multivariate approaches to detect differences in a function-valued trait between groups.

KEY WORDS: Functional data analysis, multivariate analysis, phenotype, power, repeated-measures ANOVA, repeated-measures regression.

Since the late 19th and early 20th century, evolutionary biologists have recognized that the multidimensional nature of organisms is important in the evolution of populations. Beginning with the work of Pearson (cited in Lande and Arnold [1983]) who introduced methods to understand the evolution of correlated characters and continuing with the work of Lande and coworkers (Lande 1979, 1980; Lande and Arnold 1983; Via and Lande 1985) a mathematical and statistical framework was developed to model and statistically analyze the evolution of multivariate traits.

More recently, researchers starting with Kirkpatrick and Heckman (1989) observed that a large class of evolutionarily important traits include those that are better thought of and modeled, both mathematically and statistically, as "infinite-dimensional traits." These include organismal growth trajectories, thermal performance curves, and morphological shapes among many others. Although these traits can be modeled using standard multivariate

[3]Current address: Department of Ecology and Evolutionary Biology, University of Arizona, 1041 E. Lowell St., Rm. 310 Biosciences West, Tucson, Arizona 85721

methods, the inherent unit of description of each is a function. For example, the growth trajectory of an organism describes its size as a function of its age. To highlight the functional nature of these traits, there has been a shift in terminology from "infinite-dimensional" to "function-valued" (Pletcher and Geyer 1999; Kingsolver et al. 2001; Meyer and Kirkpatrick 2005).

Methods are increasingly being designed to model and analyze function-valued traits using their inherent functional nature (Kirkpatrick et al. 1990; Meyer and Hill 1997; Wu and Lin 2006). These function-valued methods have been successfully applied in the areas of evolutionary biology and agricultural breeding (e.g., Hill 1998; Meyer and Kirkpatrick 2005). Moreover, function-valued statistical approaches provide a more general framework than multivariate methods for analyzing among- and within-species differences in that all multivariate analyses represent special cases of the function-valued approaches. Given the generality of the function-valued approach and that the basic unit of description of many traits is a function, Hill (1998) predicted that function-valued methods may eventually replace traditional multivariate methods in evolutionary biology.

Use of function-valued methods has been expanding across a wide range of disciplines, including the evolution of aging (Pletcher et al. 1998; McCarroll et al. 2004), functional morphology (Gilchrist and Huey 2004), evolutionary physiology (Kingsolver et al. 2004), life-history evolution (Dudycha and Lynch 2005; Müller and Zhang 2005), gene expression (Storey et al. 2005; Leng and Müller 2006) and quantitative genetics (Wu and Lin 2006), and statistical software implementing these methods is readily available (e.g., R [R Development Core Team 2007], SAS software [SAS Institute Inc. Cary, NC, USA], and S+ [see Ramsay and Silverman 2002; Clarkson et al. 2005]). Yet, no study has directly compared the statistical power of traditional multivariate methods relative to function-valued methods for hypothesis tests of central importance to evolutionary biologists. In comparative and experimental evolution studies, a fundamental null hypothesis is that the average phenotypes of two populations, or groups, are equal. Here we present a function-valued methodology to test this hypothesis and demonstrate that there is a significant and large gain in statistical power over multivariate methods when function-valued methods are used to detect differences.

MULTIVARIATE VERSUS FUNCTION-VALUED STATISTICAL APPROACHES

Multivariate statistical approaches that have been used to analyze functional data include univariate repeated-measures ANOVA (Milliken and Johnson 1984) and multivariate analysis (e.g., Johnson and Wichern 1988). These approaches treat functional data as a vector of phenotypic responses, but do not explicitly model the ordering and spacing of points of an individual's response. In contrast, function-valued approaches, such as repeated-measures regression (e.g., Diggle et al. 2002; Ramsay and Silverman 2005), explicitly account for both the phenotypic response of an individual as well as the ordering and spacing of points underlying the response.

Consider, for instance, growth trajectory data. A basic property of this type of data—and functional data in general—is continuity between sample points. The 40th day of an individual's life follows the completion of its 39th day. Body sizes on the 39th and 40th days of an individual's life are more likely to be similar to each other then they are to the individual's body size at day 10 or at day 60. The traditional multivariate approach ignores this continuity between sample points. Although they may estimate that body size at age of 40 days is more highly correlated with body size at age of 39 days than at age of 60 days, multivariate approaches ignore the fact that day 40 is one day from 39 and 20 days from 60. This important information is, in effect, thrown out in multivariate approaches but retained in function-valued approaches.

To further understand the difference between multivariate and function-valued approaches, consider the three examples of functional traits shown in Figure 1. For each trait a single individual's measured responses (phenotypes) are plotted at 20 different values of x, which could represent 20 ages, times, temperatures, etc. In general, an individual's responses as a function of x, $y(x)$, can be written as

$$y(x) = f(x, \vec{a}) + \varepsilon(x), \tag{1}$$

where $f(x, \vec{a})$ is the underlying function, with $\vec{a}$ being the parameters of the function that define an individual's response, and $\varepsilon(x)$ is a random measurement error at point x. The data shown were generated from a fixed underlying function for that individual with measurement error added. In the function-valued approach used in this article, a curve is fitted to the 20 observations yielding estimates of the parameters ($\vec{a}$) that define an individual's response. These parameter estimates are used for hypothesis testing, whereas in the multivariate approach, the 20 observed responses are used for hypothesis testing.

It is important to remember that the function-valued approach is built on the premise that there is a true underlying functional response for each individual and that the measured responses for each individual are representative of their true underlying functional response. A statistical procedure that more accurately estimates individual functional responses will potentially have more power to detect differences among the average functional responses of two populations. The information that is gained by retaining the ordering and spacing of points in a function-valued approach may potentially lead to more accurate estimates of individual functional responses than the estimates obtained by the multivariate approach.

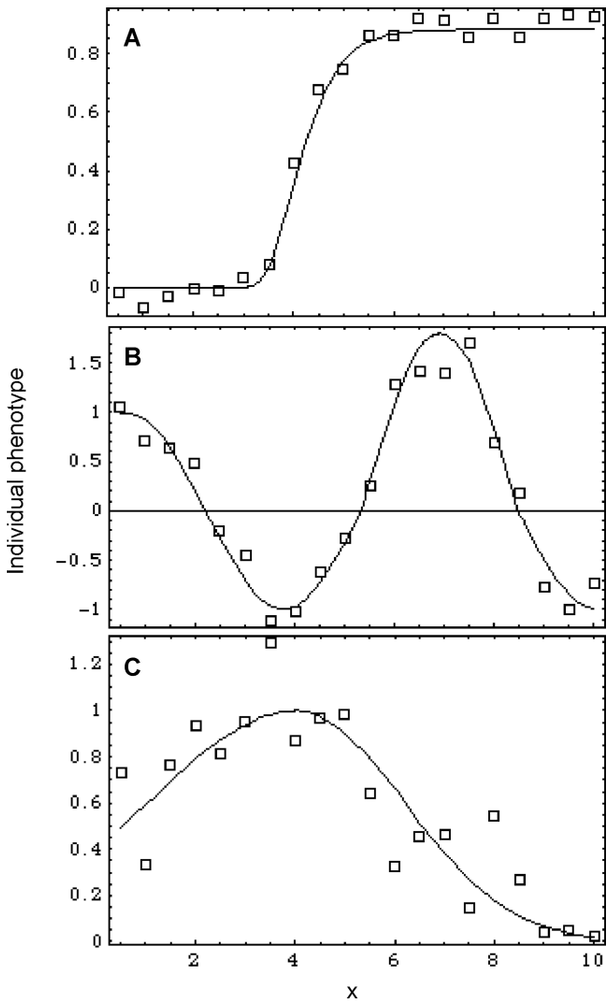

Figure 1. Plots of an individual's phenotype (open squares) for the three types of trait functions: (A) growth, (B) cyclic, and (C) performance. These plots illustrate the shapes of the curves that are modeled and the observed variability about an individual's true functional response (solid line) that is attributed to measurement error. It is assumed that, on average, there is less measurement error for growth (SD = 0.04) than for cyclic and performance responses (SD = 0.2).

Although the multivariate approach suffers from the problem of not retaining information, a potential shortcoming of the function-valued approach is that the underlying function f is generally unknown and if its form is guessed inaccurately, then all other inferences could be misleading. This is a serious problem for parametric approaches but can be avoided by using nonparametric methods to analyze the functional data.

Nonparametric methods (e.g., Ramsay and Silverman 2005) use "basis function expansions" that can approximate (to any degree of precision) any function of biological interest. Basis function expansions are constructed using families of basis functions including splines, Fourier series, and the Legendre polynomials. In this article, we use nonparametric methods to represent the functional responses of individuals, although for comparison we also explore the power of parametric approaches in a few cases.

Our main objective is to determine conditions when nonparametric function-valued statistical approaches have more power than multivariate methods for detecting phenotypic differences among populations. Using simulated data, we consider how these conditions depend on (1) the function underlying the functional data, (2) the number of observations per individual, and (3) the family of basis functions used in the analysis.

Methods

DATA

We assume throughout that the original data for an individual j in population i consist of two vectors: $\vec{y}_{ij} = \{y_{ij1}, y_{ij2}, \ldots, y_{ijn_{ij}}\}$ contains the phenotypic responses and $\vec{x}_{ij} = \{x_{ij1}, x_{ij2}, \ldots, x_{ijn_{ij}}\}$ contains the points at which those responses were measured. One particularly useful property of the function-valued statistical approach is that in practice $\vec{x}_{ij}$ can differ among individuals in both the points that are sampled and the number of points that are sampled. In the multivariate approach, the same number of points and, ideally, the same points must be sampled for each individual. Otherwise, ad hoc binning of the data must be performed. To level our comparison of function-valued and multivariate approaches, we only consider samples in our simulations for which the $\vec{x}_{ij}$'s are the same for all individuals, that is, $\vec{x}_{ij} = \{x_1, x_2, \ldots, x_n\}$ for all sampled individuals.

NONPARAMETRIC FITS

For the nonparametric function-valued approach, we fit the responses using a set of basis functions $\{B_1 \ldots B_L\}$. Specifically, we fit the phenotype of individual j in population i at sample point k using the model

$$y_{ijk} = \sum_{\ell=1}^{L} \left(\beta_{i\ell} + b_{ij\ell}\right) B_\ell \left(x_{ijk}\right) + \varepsilon_{ijk}. \qquad (2)$$

Here, the $\beta_{i\ell}$'s, or fixed effects, are attributed to individual ij's particular population. The sum $\beta_{i\ell} + b_{ij\ell}$ is the coefficient of individual ij for basis function B_ℓ and ε_{ijk} is measurement error. Equation (2) can be rewritten in matrix form as

$$\vec{y}_{ij} = \mathbf{Z}_{ij}\vec{\beta}_i + \mathbf{Z}_{ij}\vec{b}_{ij} + \vec{\varepsilon}_{ij} \qquad (3)$$

where $\vec{y}_{ij}$ is the vector of responses of individual j in population i, $\vec{\beta}_i$ is the vector of basis coefficients attributable to individual ij's population, $\vec{b}_{ij}$ is the vector of deviations from $\vec{\beta}_i$ attributable to individual ij, and $\vec{\varepsilon}_{ij}$ is a vector of random measurement errors. Because there are n_{ij} observations from individual ij, $\mathbf{Z}_{ij}$ is a n_{ij} by L

matrix consisting of evaluations of the basis functions at the components of $\vec{x}_{ij}$. For instance, the $k\ell$-th element of $\mathbf{Z}_{ij}$ has the value $B_\ell(x_{ijk})$, that is the value of the ℓ th term in the basis expansion for point x_{ijk}. The form of equation (3) is the standard mixed-model repeated measures regression with $\vec{\beta}_i$ a vector of fixed effects and $\vec{b}_{ij}$ a vector of random effects; methods used to estimate fixed and random effects for this model apply (e.g., Laird and Ware 1982), but note that our regression matrix is the same for both fixed and random effects. In a sense, the points x_{ijk} have merely been transformed into the values in the kth row of $\mathbf{Z}_{ij}$, but this is how information about the continuity of the x_{ijk}'s is preserved.

PARAMETRIC FITS

As a point of comparison, we also performed a parametric repeated measures regression analysis for a few select cases. Here the model for individual ij's response at sample point k is

$$y_{ijk} = f(x_{ijk}, \vec{p}) + \varepsilon_{ijk}, \tag{4}$$

where f is the conjectured parametric function, $\vec{p}$ contains the parameters of that function, and ε_{ijk} is measurement error. We estimated the parameters of this parametric model separately for each individual in Mathematica version 5.2 (Wolfram 2005) using the Nonlinear Fit library.

BASIS FUNCTIONS

Any appropriate family of basis functions can be used to approximate biological functions to any desired degree. The choice of basis function family may, however, affect the quality of the fit to the data. To investigate the impact of basis function choice, we implemented the nonparametric function-valued approach using three types of basis functions to fit generated data: piecewise polynomial functions, cyclic functions, and polynomials.

We use B-splines for our piecewise polynomial functions. B-splines divide a functional response up into subintervals, with each subinterval delimited by knots. Within each subinterval the data are fitted by a polynomial of degree d. These piecewise polynomial functions are connected together to fit the entire set of data for an individual. In this article we used piecewise cubic polynomials (i.e., $d = 3$), so our functions are piecewise cubic with two continuous derivatives. A recursive function can be used to calculate the terms for a B-spline basis function (De Boor 1978)

$$B_\ell^d(x) = \frac{(x - t_\ell)B_\ell^{d-1}(x)}{t_{\ell+d} - t_\ell} + \frac{(t_{\ell+d+1} - x)B_{\ell+1}^{d-1}(x)}{t_{\ell+d+1} - t_{\ell+1}}. \tag{5}$$

The parameters t_ℓ in the recursion are the knots, that is the endpoints of the subintervals. For simplicity, we used evenly spaced knots.

For cyclic basis functions, we used the cosine family $B_\ell(x) = \cos(\frac{\ell\pi x}{x_{\max}})$. This basis is a set of cosine functions with different periods standardized by the maximum value of x.

For a polynomial basis, we use the Legendre polynomials. These, like B-splines, can be defined recursively

$$B_\ell(\tilde{x}) = \begin{cases} 1 & \text{if} \quad \ell = 0 \\ \tilde{x} & \text{if} \quad \ell = 1 \\ \dfrac{(2\ell - 1)\tilde{x}B_{\ell-1}(\tilde{x}) - (\ell - 1)B_{\ell-2}(\tilde{x})}{\ell} & \text{if} \quad \ell > 1, \end{cases} \tag{6}$$

where $\tilde{x} = (2x - x_{\max} - x_{\min})/(x_{\max} - x_{\min})$ is a scaled version of x that ranges from -1 to $+1$ as x ranges from $x_{\min}$ to $x_{\max}$ (Abramowitz and Stegun 1964).

The parametric method, described in the previous section, uses a single function with a small set of parameters to fit the entire data. In contrast, the cosine and Legendre polynomials are made-up of orthogonal functions and B-splines fit small sections of the data and then piece these small sections together.

ESTIMATION

We used two different methods to estimate coefficients in the basis expansion (2). The first method, called a mixed effects model (Laird and Ware 1982), assumes that the b_{ij} vectors each follow a multivariate normal distribution with mean zero, and that the ε_{ijk}'s are independent normally distributed with mean zero. We used maximum likelihood to estimate parameters, specifically, the EM algorithm (Liu and Rubin 1994) for a mixed-model, repeated measures regression. Finding maximum-likelihood estimates of the parameters of the linear mixed effects is computationally intensive, and thus slow. To speed up convergence, we tried other algorithms, such as conjugate-gradient and Newton–Raphson methods. Subsequent analyses showed these methods were prone to finding local maxima, whereas the EM algorithm was better at finding global maxima. Because the EM algorithm is extremely slow, a thorough study of its properties by extensive simulation was not computationally feasible.

Because of the computational difficulties fitting the linear mixed effects model, we also used another method, which was far faster to implement. We treated the $\vec{b}_{ij}$'s as fixed, not random, and estimated $\vec{\beta}_i + \vec{b}_{ij}$ using standard least-squares regression, carried out independently on each individual. The least squares method is relatively simple and fast, but does not use any information provided by other individuals' data. This least-squares method yields maximum-likelihood estimates under the assumption that both $\vec{\beta}_i$ and the $\vec{b}_{ij}$'s are fixed parameters, and the measurement error is normal (Searle et al. 1992).

MODEL CHOICE

The nonparametric function-valued approach requires not only choice of the type of basis but also specification of the number of terms L in the basis function expansion (2). Expansions with more terms approximate a given set of responses more closely, but using too many terms leads to overparameterization and loss

of predictive power. Because we are comparing nonnested models, we do not use hypothesis testing procedures to choose the most appropriate model. Rather, we use the information-based AIC (Akaike 1976), a criterion that is suitable for choice among nonnested models (e.g., Burnham and Anderson 1998). We chose the number of estimated parameters P and the basis function type to minimize $AIC = -2\ln\mathcal{L} + 2P$ where $\mathcal{L}$ is the maximized likelihood of the mixed effects model and P equals L plus the number of unique terms in the covariance matrix of the $\beta_i's$ and the residual error term. In our simulation study, we present results on model choice and choice of P using a small, moderate, and large number of terms relative to the number of points sampled per individual.

DATA GENERATION

We compared the statistical performances of multivariate and function-valued approaches using sets of hypothetical data representing the spectrum of published work on function-valued traits in evolutionary biology. We generated data based on three trait function prototypes: (1) a monotonic growth function, (2) a cyclic expression function, and (3) a unimodal performance function. Each prototype function depends on a small number of parameters.

We allowed the parameters that determine the functional phenotype of an individual to vary about a fixed mean within each population. So, to construct our datasets, we drew sets of random parameters from a fixed distribution, one set for each individual in the dataset. These random parameters can be thought of as random effects.

We used a modified Gompertz model for monotonic growth

$$f(x) = \alpha \exp\{-\exp[-\kappa(x-\gamma)]\} - \exp\{-\exp[-\kappa(-\gamma)]\} \quad (7)$$

(Winsor 1932). An individual's α parameter determines where its body size reaches a maximum and κ determines the rate of increase near the point of maximum rate of growth (γ). For the power analysis we allowed the mean of either α or γ to differ or both α and γ to differ between populations.

To model cyclic expression we used the function

$$f(x) = \begin{cases} \kappa \sin(x+\gamma) & 2\pi - \gamma \le x \le 3\pi - \gamma \\ \alpha \sin(x+\gamma) & x < 2\pi - \gamma, x > 3\pi - \gamma \end{cases} \quad (8)$$

where γ determines the periodicity of the curve, and α and κ determine amplitudes at different points in the curve. In power analyses we allowed the mean of κ to differ between populations or both κ and α to differ simultaneously between populations.

Finally, we represented unimodal performance by

$$f(x) = \begin{cases} e^{-\alpha} \exp\left(-\dfrac{\alpha x(x-2\gamma)}{\gamma^2}\right) & x \le \gamma \\ e^{-\kappa} \exp\left(-\dfrac{\kappa x(x-2\gamma)}{\gamma^2}\right) & x > \gamma. \end{cases} \quad (9)$$

This function gives a maximum performance when $x = \gamma$. The decline in performance below and above γ are determined by the parameters α and κ, respectively. In power analyses we allowed the mean of either γ or κ to differ or both κ and α to differ between populations.

Each of our prototypical function-valued phenotypes involves the individual-specific parameters α, κ, and γ. For each dataset, we randomly and independently generated values of these parameters for every individual from a normal distribution. A standard deviation of 0.2 was used for each parameter. We allowed the mean of only one parameter to differ between populations for a given power analysis. The expected difference between populations in the varied parameter ranged from zero to 0.4 for the Gompertz curve (i.e., zero to two standard deviations) and from zero to 0.8 (zero to four standard deviations) for the cyclic and performance curves.

Simulation studies assumed either high or low measurement error. Figure 1 illustrates the variability about an individual's true curve due to high measurement error for the three types of functions. High measurement error corresponded to a standard deviation of 0.04 for growth curves and 0.2 for both cyclic and performance functions. Low measurement error was modeled using a standard deviation of 0.002 for all function types. Figure 2 shows some nonparametric fits to individual data generated with high measurement error. In the example data and fits provided in Figure 2, nonmonotonicity of the fits for the growth curve is the result of estimation error; likewise, the upward flip at the right-hand side of the performance curve is also due to estimation error.

Note that for a given dataset, it would be straightforward to estimate the model parameters for each individual by assuming the true model and then using parametric regression analysis. However, especially for the cyclic expression and unimodal performance, it would be difficult a priori to select an appropriate model with reasonable confidence. Empiricists using parametric functional approaches to analyze real data must grapple with this critical issue whereas those using nonparametric methods do not. Our study will thus focus on the comparative statistical performances of multivariate and nonparametric function-valued approaches.

HYPOTHESIS TESTING

Our datasets were generated to test the abilities of multivariate and function-valued statistical methods to detect among-population (or among-group) differences in mean phenotype. That is, we wished to estimate their respective capacities to reject the null hypothesis of no difference when differences existed. We also examined rates of false rejection when the null hypothesis was true.

Both the traditional multivariate approach and the function-valued approach use a version of Hotelling's T^2 statistic (Johnson and Wichern 1988) to test for a difference in means of two

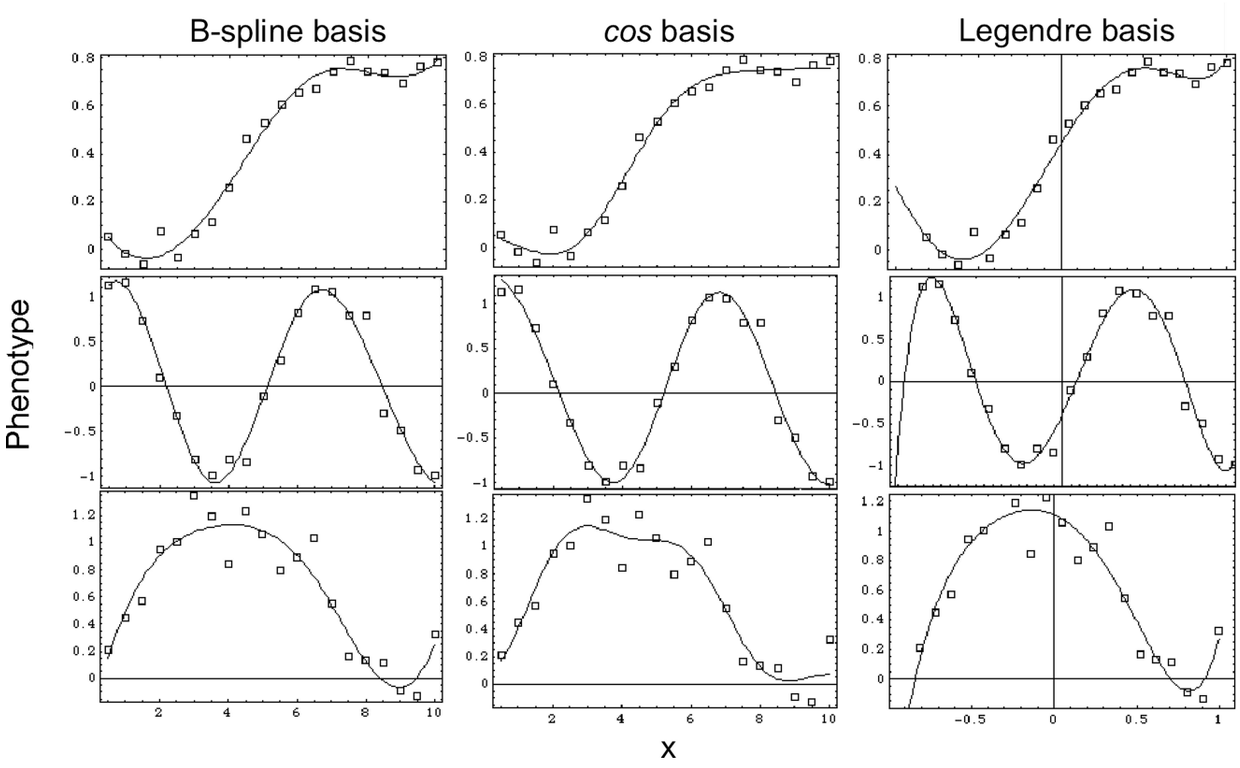

Figure 2. **Examples of fits of the three basis functions to data from an individual for each functional type. For each basis, its corresponding fits are arranged in a column. Note that for Legendre Polynomials the x-axis is rescaled from −1 to 1, as is standard for that basis.**

populations. In the case of the multivariate approach Hotelling's T^2 takes the form

$$T^2 = (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)' \left[\left(\frac{1}{N_1} + \frac{1}{N_2} \right) \mathbf{S} \right]^{-1} (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2), \qquad (10)$$

where $\bar{\mathbf{y}}_1$ and $\bar{\mathbf{y}}_2$ are the mean vectors of phenotypic responses for, and N_1 and N_2 are the sizes of samples one and two, respectively. $\mathbf{S}$ is the variance–covariance matrix of the data vectors $\vec{y}$ pooled from both populations.

Hypothesis testing in the function-valued approach treats each individual's estimated regression coefficients (or parameters) as a data vector. The function-valued version of Hotelling's T^2 statistic is thus

$$T^2 = (\bar{\hat{\beta}}_1 - \bar{\hat{\beta}}_2)' \left[\left(\frac{1}{N_1} + \frac{1}{N_2} \right) \mathbf{S}_\beta \right]^{-1} (\bar{\hat{\beta}}_1 - \bar{\hat{\beta}}_2), \qquad (11)$$

where $\bar{\hat{\beta}}_1$ and $\bar{\hat{\beta}}_2$ are the sample averages of the estimated regression coefficient (or parameter) vectors for the two populations and $\mathbf{S}_\beta$ is their pooled variance–covariance matrix.

We used permutation tests to compute P-values for the T^2 statistics under the null hypothesis of no between-population difference. This was done by randomly assigning individuals to one of two groups, recalculating the T^2 statistic for each new grouping. We repeated this 500 times thereby generating 500 "datasets" satisfying the null hypothesis of individuals being drawn from the same distribution. The P-value was estimated as the proportion of the 500 recalculated T^2 statistics greater than the observed T^2 statistic. The same collection of 500 randomized datasets was used to compute P-values for the multivariate, nonparametric, and parametric approaches for a single set of generated data.

Our hypothesis testing procedure can be extended to more than two populations. Instead of using T^2 statistics (eqs. 10, 11), a multivariate version of single-factor F-statistics (e.g., Johnson and Wichern 1988) would be used. The randomization process to calculate P-values would randomly assign individuals to groups and compute an F-statistic. This process would be repeated to generate a distribution of F-statistics under the null hypothesis of no among-group differences. A similar approach to hypothesis testing was used in a time-course study of gene expression using microarrays (Storey et al. 2005). In that study, expression data were fit with cubic splines and P-values were calculated using a residuals-based bootstrapping procedure.

COMPUTATIONAL STUDY AND POWER ANALYSIS

The ideal nonparametric function-valued procedure is to fit the data using maximum likelihood with the mixed effects model and use AIC to choose the best type and number of terms in the basis function expansion. The T^2 test statistic (11) would be

calculated using the best-fit coefficients and the *P*-value computed by a permutation test as described above.

It is possible, but computationally intensive, to analyze a single dataset using the ideal nonparametric procedure. Repeating the ideal procedure over thousands of datasets via simulation (as required in a power analysis) is, however, computationally prohibitive. We therefore separated our simulation study into two parts: a study of model choice and a study of power analysis. We created 500 datasets for each study part, trait function prototype, and parameter configuration.

For the model choice study, we used maximum-likelihood estimation with the mixed effects model. We fit each dataset with several different basis expansion types and lengths and chose the best of these expansions via AIC. Specifically, when there were $n = 10$ measurements per individual, we fit using $L = 3$, 6, and 9 terms in the basis expansion (2) for the cosine series and Legendre polynomials and $L = 5$, 6, and 9 terms for B-splines (which require a minimum of four terms). With $n = 20$ measurements per individual, we used $L = 6$, 12, and 18 terms in the expansion for all three basis types. Using AIC we determined the type and number of terms in the expansion that best fit a given form of data function (i.e., growth, cyclic, or performance) most frequently for a given n.

In our power analyses, we used these "best" basis function expansions when fitting both the original data and the randomized datasets in the permutation tests. To speed computation, all fits in the power analysis were done by least squares rather than by maximum likelihood with mixed effects. Power was estimated using 500 replicate datasets generated under a given set of assumptions with a fixed difference in mean phenotype between groups. The power was estimated as the fraction of times the null hypothesis of equal mean phenotypes was rejected at the 0.05 significance level.

Results

MODEL CHOICE

When measurement errors were large (SD = 0.04 for growth and SD = 0.2 for cyclic and performance data), AIC generally favored cosine series (Table 1). For growth curve and cyclic function data, the AIC-best cosine expansions all involved six terms**.** For performance curve data consisting of 10 measurements per individual, AIC consistently chose a three-term cosine expansion as the best model. In contrast, AIC identified Legendre polynomial expansions most often as the best model when fitting performance data with 20 measurements per individual, although only marginally so than cosine expansions (Table 1).

When measurement errors were small (standard deviation of 0.002 for all functions), AIC generally favored fits using cosine series for growth data whereas Legendre polynomials were selected for performance and cyclic data. Consult the legends of online Supplementary Figures S1–S3 for further details.

Table 1. Model choice results for large measurement errors using AIC.

	Basis function			Total number of replicates
	B-Spline	Cosine	Legendre	
Gompertz				
$n = 10$	35	391	74	500
$n = 20$	6	483	11	500
Cyclic				
$n = 10$	112	322	66	500
$n = 20$	43	437	20	500
Performance				
$n = 10$	61	436	3	500
$n = 20$	132	179	189	500

COMPARATIVE POWER

We used basis function expansions identified as "best" in the previous section to study the statistical power of the nonparametric function-valued approach. The simulations showed that the power of this function-valued approach was consistently and often substantially higher, but never lower, than the corresponding multivariate analysis of the same data when measurement error was high (Figs. 3–8). When measurement error was low the power of the function-valued approach was consistently higher but the difference in power between function-valued and multivariate analyses was less dramatic (online Supplementary Figs. S1–S3).

For data generated by the Gompertz growth curve the nonparametric function-valued approach had equal or better power to detect differences between mean functions relative to the multivariate approach when the expected values of either parameter γ or α differed between populations (Figs. 3 and 4). The contrast in power was greater with 20 than with 10 measurements per individual. The comparisons were qualitatively the same for cyclic expression data (Fig. 5) and for performance trait data (Figs. 6 and 7).

The results presented in Figures 3–7 vary the mean of a single parameter between populations. We also looked at cases in which the means of two parameters differed between populations and 20 points were sampled per individual, and so the function correspondingly differed between populations in more than one aspect. In the case of the Gompertz function, α and γ simultaneously differed causing differences in the ages of maximum growth and maximum size. Here, the function-valued approach using the cosine basis and a six-term expansion continued to have higher power than the multivariate approach. When both α and γ differed by 0.1 between populations, the power of the function-valued approach

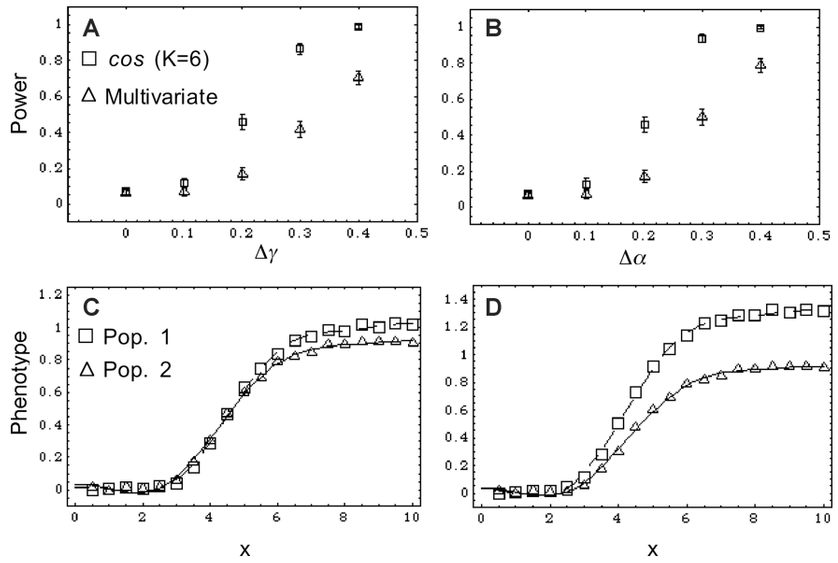

Figure 3. **The power of the function-valued and multivariate approaches: Growth curve, 20 points sampled per individual and 20 individuals sampled per population. The Type I error rate is 0.05 (parts A and B). In parts (C) and (D) examples are provided of estimated average functions (lines) and the average phenotype of individuals (symbols) for a single arbitrarily chosen replicate. In part (C) $\Delta\ \gamma = 0.3$ and in part (D) $\Delta\ \alpha = 0.3$. For power and example fits, the analyses involving cosine have six terms in the expansion because AIC consistently chose this model.**

was 0.23 {0.19,0.27} compared to 0.10 {0.07,0.13} for the multivariate approach. And when both α and γ differed by 0.2 between populations, the power of the function-valued approach was 0.85 {0.82,0.88} compared to 0.37 {0.33,0.41} for the multivariate approach. The values in brackets are 95% confidence intervals for the power, based on the simulation-based estimate of the power.

For the cyclic curve, both α and κ differed causing the functions to differ in peak height at two points along the function. Here again, the function-valued approach using the cosine series and a six-term expansion had higher power than the multivariate approach: when both parameters differed by 0.1 between populations, the power of the function-valued approach

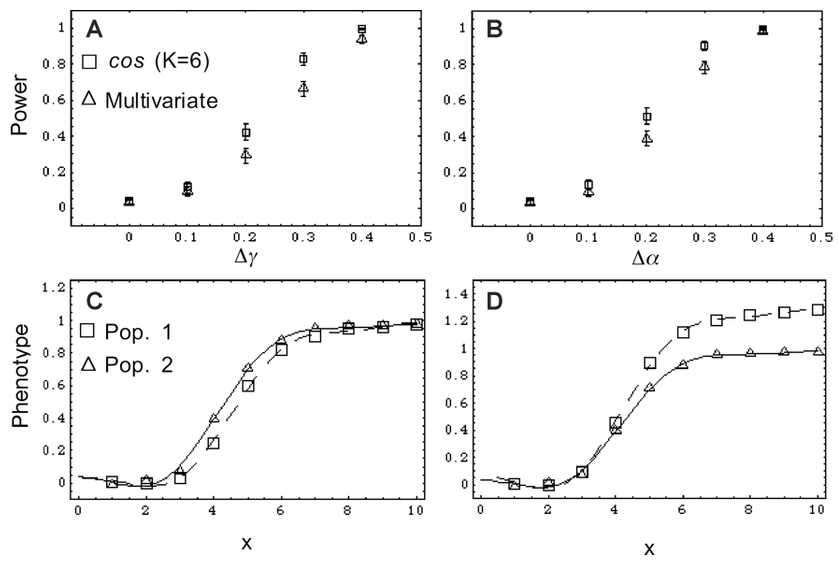

Figure 4. **The power of the function-valued and multivariate approaches: Growth curve, 10 points sampled per individual and 20 individuals sampled per population. Parts A–D as in Figure 3. In part (C) $\Delta\ \gamma = 0.3$ and in part (D) $\Delta\ \alpha = 0.3$. For power and example fits, the analyses involving cosine have six terms in the expansion because AIC consistently chose this model.**

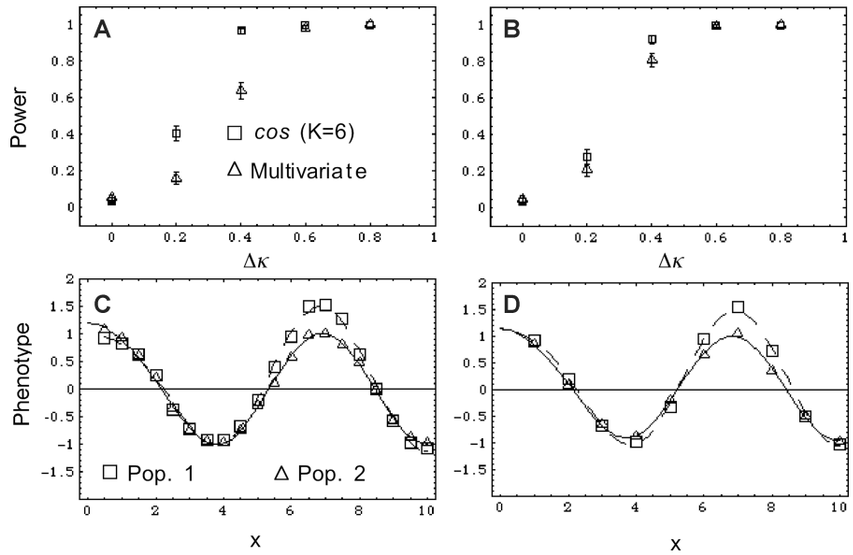

Figure 5. The power of the function-valued and multivariate approaches: Cyclic curve, (A, C) 20 points sampled per individual and 20 individuals sampled per population and (B, D) 10 points sampled per individual and 20 individuals sampled per population. In parts (C) and (D) $\Delta\ \kappa = 0.6$. For power (A, B) and example fits (C, D), the analyses involving cosine have six terms in the expansion because AIC consistently chose this model.

was 0.20 {0.16,0.24} and for the multivariate approach it is 0.12 {0.09,0.15} and when both parameters differed by 0.2 between populations, the power of the function-valued approach was 0.73 {0.70,0.76} and for the multivariate approach it is 0.33 {0.29,0.37}.

For the performance curve, both α and κ differed between populations causing differences in maximum performance and the rate of decline in performance following the maximum. The function-valued approach using Legendre polynomials and a six-term expansion had marginally higher power relative to the

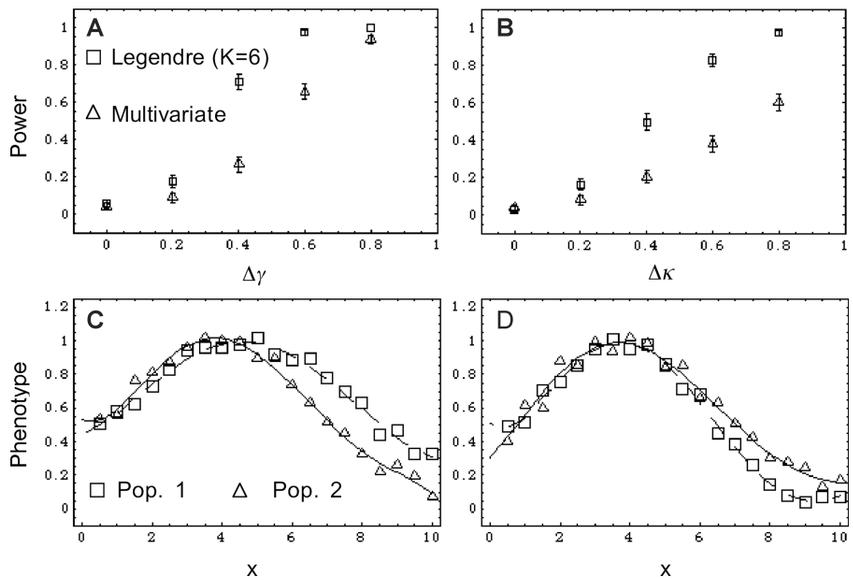

Figure 6. The power of the function-valued and multivariate approaches: Performance curve, 20 points sampled per individual and 20 individuals sampled per population. In part (C) $\Delta\ \gamma = 0.6$ and in part (D) $\Delta\ \kappa = 0.8$. For power (A, B) and example fits (C, D), the analyses involving Legendre polynomials have six terms in the expansion because AIC consistently chose this model.

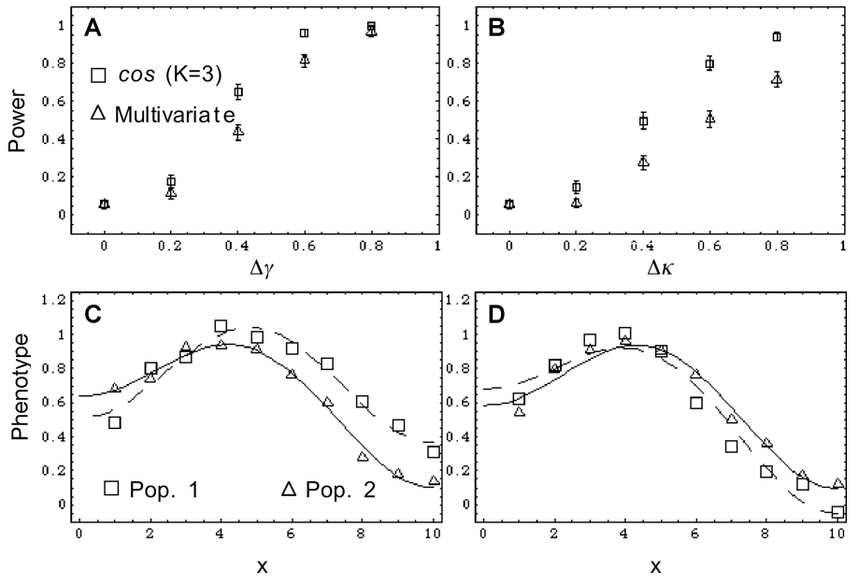

Figure 7. **The power of the function-valued and multivariate approaches: Performance curve, 10 points sampled per individual and 20 individuals sampled per population. In part (C) $\Delta\gamma = 0.8$ and in part (D) $\Delta\kappa = 0.8$. For power (A, B) and example fits (C, D), the analyses involving cosine have three terms in the expansion because AIC consistently chose this model.**

multivariate approach to detect differences in functional response between populations when both α and κ differed by 0.1: 0.09 $\{0.06,0.12\}$ for Legendre polynomials and 0.05 $\{0.05,0.07\}$ for the multivariate approach. When there was a greater difference in maximum performance and rate of decline in performance between populations (0.2), the function-valued approach fared even better relative to the multivariate approach with power of 0.32 $\{0.28,0.36\}$ compared to 0.16 $\{0.13,0.19\}$.

In all cases we examined, the gain in power of the nonparametric function-valued approach over the multivariate approach did not appear to come at the cost of an elevated Type I error rate. Whenever there was no difference between the average functions of populations, the corresponding Type I error rate did not deviate significantly from 0.05, our significance threshold (see results corresponding to $\Delta\kappa = 0$, $\Delta\gamma = 0$, or $\Delta\alpha = 0$ in the top panels of Figs. 3–7 and in online Supplementary Figs. S1–S3).

We discovered an unanticipated advantage of the function-valued approach over the multivariate approach for detecting population differences in some cases (Fig. 8). Simulations showed that as the number of measurements per individual was increased, the multivariate approach typically lost power although the power of the function-valued approach remained relatively constant or improved. This phenomenon can also be seen in Figures 3–7 and online Supplementary Figs S1–S3 on comparing the power of function-valued relative to multivariate approaches with 20 versus 10 measurements per individual for each expected difference between populations in parameter values.

As a benchmark, we compared the power of the nonparametric function-valued approach to the corresponding parametric method assuming the correct model (eq. 7 for Gompertz growth, eq. 8 for cyclic expression, and eq. 9 for performance data) for a selection of cases. For example, when the difference in expected values of γ was $\Delta\gamma = 0.3$ with 20 individuals and 20 measurements per individual, the parametric approach always rejected the null hypothesis at the 0.05 significance level. This confirms the superior power one would expect of the (correct) parametric approach to the nonparametric approach (see Fig. 3A). The parametric approach was far less efficient at rejecting the null hypothesis for cyclic expression data, assuming an expected difference in κ of $\Delta\kappa = 0.2$ and 20 individuals with 20 measurements each. In this case, the null hypothesis was rejected at an estimated rate of 0.45 ± 0.04. This lower power still exceeded, although marginally, that of the nonparametric approach under the same conditions (cf. Fig. 5A).

Unlike growth and cyclic data, our analyses of simulated performance data showed that the parametric approach could have lower power than the nonparametric approach. For instance, when the expected difference in κ was $\Delta\kappa = 0.6$, the parametric approach rejected the null hypothesis at a rate of only 0.25 ± 0.04 whereas the nonparametric approach rejected the null hypothesis over 80% of the time under the same conditions (Fig. 6B). The reason appears to be that the impact of the three parameter values on the shape of the curve is not neatly partitioned, that is, changes in the value of one parameter might appear to be attributable to changes in another parameter. With substantially

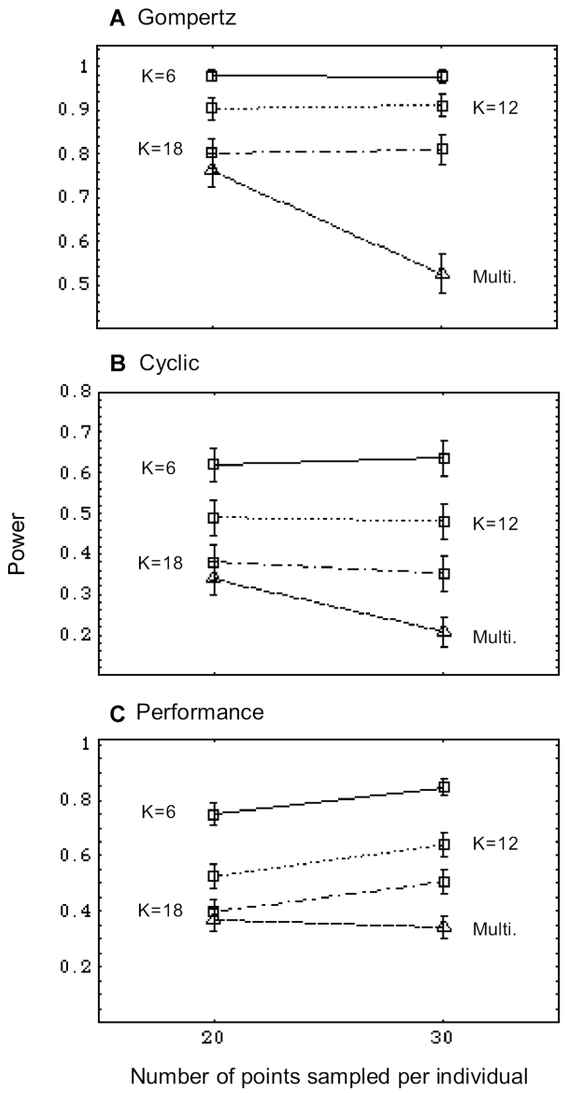

Figure 8. Power to reject the null hypothesis that the average function is the same in two populations when the number of points sampled per individual varies from 20 to 30. In all cases the number of individuals that are sampled per population is 30. Points represented with triangles are for the multivariate approach and points represented by squares are for the function-valued approach with the number of terms in the expansion noted. Based on the model choice results summarized in Table 1, the cosine series was used for the Gompertz and cyclic curves and Legendre polynomials for the performance curve. Error bars are 95% confidence intervals of the average. For the Gompertz growth curve (A), the parameter that varied, on average, between populations is γ, such that in one population it is 4.0 and in the second population it is 4.3. For the cyclic curve (B) the parameter that varied, on average, between populations is κ, such that in one population it is 1.0 and in the second it is 1.2. For the performance curve (C) the parameter that varied, on average, between populations is κ such that in one population it is 1.0 and in the second it is 1.4. High measurement error was modeled in all cases.

smaller measurement error (SD = 0.002), the parametric approach had less difficulty estimating the parameters and, once again, had better power than the nonparametric approach (data not shown).

Discussion

The capacity to address subtle comparative questions involving complex traits depends on the power of statistical methods used to detect small phenotypic differences among populations and groups. For comparative questions involving functional traits, our overall results suggest that the nonparametric function-valued approach generally had at least as good power as the multivariate approach, and often much better, for detecting differences in mean phenotypic functions between two populations. The gain in power of the nonparametric function-valued approach was consistent across the three types of functional responses that were studied.

The correct parametric function-valued approach usually had better power than the nonparametric approach but, unexpectedly, not always. In the case of the performance curve data involving high measurement error, the parametric approach failed to partition variation correctly and had reduced power relative to the nonparametric approach. This occurred despite the advantage of knowing the true parametric model from which the data were generated. In reality, the true parametric model is typically unknown although we did not evaluate the effect of assuming an incorrect parametric model.

Note that the powers of the parametric and nonparametric approaches were inferred using a T^2 statistic based on estimated function parameters. Another approach is to use each individual's estimated parameters to calculate that individual's fitted values at each measurement point and then use a T^2 statistic based on these fitted values instead of the estimated parameter values. It can be shown that the nonparametric approach using the individuals' parameter estimates is equivalent to using the individuals' fitted data in hypothesis testing and so both procedures will yield the same level of power.

In the parametric approach, T^2 tests using estimated parameters versus fitted values are not equivalent and may yield different levels of power. We performed a preliminary analysis of the parametric T^2 test using the fitted-data approach and found that in some cases the approach had better power (performance curves) and in some cases worse power (Gompertz and cyclic curves) than the parametric T^2 test based on parameter estimates. Although the fitted-data parametric approach improved power for performance curve comparisons, this power remained less than the nonparametric approach, suggesting that the nonparametric approach still yields better fits of the data in this case. The use of parameter estimates versus fits in the parametric approach needs further study.

As measurement error increases, the gain in power of the nonparametric function-valued approach relative to the multivariate approach also increases. The reason can be understood as follows. As measurement error decreases, more terms in the basis function expansion are required to best fit the data. When measurement error is small, the number of estimated coefficients in the function-valued approach is nearly the same as the number of estimated vector components in the multivariate approach, and consequently the efficiency of the function-valued approach in detecting between-group differences tends to resemble that of the multivariate approach. In contrast, the function-valued approach requires many fewer terms (and coefficients) to fit noisy data whereas the multivariate approach must estimate the same number of components whether data are noisy or not.

We found that the power of the multivariate approach typically declined with the number of measurements taken per individual whereas the power of the function-valued approach typically remained constant or improved. The reason for this is that the nonparametric function-valued approach fits datasets, whether they consist of 20 versus 30 points per individual, with about equal efficiency. As the number of points per individual increases, the number of estimated parameters remains about the same in the nonparametric function-valued analysis but increases in the multivariate approach contributing to the latter approach's declining power. The reason for this decline is that, for large dimensional data, multivariate methods tend to provide poor estimates of the variance–covariance matrix and this leads to hypothesis tests with little or no power. When data are function-valued, arising from smoothly varying processes, then a function-valued approach provides a fairly parsimonious model for the covariance structure. Thus, the function-valued approach allows for a more efficient model of the data (fewer parameters) and a better estimate of the data's covariance structure. Thus the function-valued approach can be expected to provide more powerful hypothesis tests.

If individual functional responses cannot be modeled with fewer parameters than the number of measurements per individual in the function-valued approach then there is no information gained using the ordering and spacing of points. An example of this occurs when the responses of an individual at adjacent measurement points are random and independent. Under these conditions, information about the ordering and spacing of points cannot be used predictively. Accordingly, the function-valued and multivariate approaches would estimate with equal efficiency and accuracy the underlying individual functional responses and therefore have equal power.

The relative efficiency of function-valued versus multivariate methods will also depend jointly on the inherent smoothness of the underlying function-valued traits being compared and measurement error. If our data arise from erratically changing function-valued traits, we doubt the nonparametric function-valued approach would work much better than the multivariate approach because, as noted in the previous paragraph, the number of basis terms required to fit the data would approach the number of observations per individual. If the data arise from smoothly varying function-valued traits and measurement error is very small, then the nonparametric function-valued approach will try to fit the data exactly, requiring a large number of basis functions to do so. Thus we expect that the nonparametric function-valued approach and the multivariate approach will have similar power. However, if data arise from smoothly varying function-valued traits and measurement error is relatively large, then the within-individual correlation has a structure that is well modeled by our basis function approach, with just a few parameters. Under these conditions, we expect the nonparametric function-valued approach to have higher power than the multivariate approach.

Another important issue in the analysis of function-valued data concerns the optimal design of an experiment with respect to the number of measurements per individual and the number of individuals sampled. Zhang and Zhong (2006) addressed this issue when assessing the power to detect genetic linkage among loci contributing to variation of a function-valued trait. In that study they found only a weak relationship between power and the number of individuals sampled and measurements taken per individual. Furthermore, the relationship was dependent on the genetic model and assumed sizes of genetic effects. They suggest that prior to an experiment, a power analysis be done for the particular type of problem being addressed.

In this article we have not thoroughly addressed optimal design strategies. Our results suggest, however, that at least for the types of functions, the extent of measurement error, and the variation in underlying parameters simulated here, there is often little to no gain in power when increasing the number of measurements from 10 to 20 points per individual. An exception is the performance curve, for 20 versus 30 measurements per individual; here there was a clear advantage to increasing the number of measurements per individual. Analyses of other functions or lower levels of measurement error may benefit from sampling more points per individual. As a rough guide, our results suggest that samples consisting of 10 or more measurements per individual should suffice when using function-valued methods for analysis. Although implementing a function-valued approach highlights the issues concerning measurements per individual, these issues must also be addressed in applications of multivariate approaches.

We found that the cosine basis provided, most often, optimal fits to the data. This result may depend on the prototype functions and various assumptions used in our simulations. It is conceivable that under different conditions, B-splines or Legendre polynomials may prove to be the better basis for statistical analysis. Furthermore, the computational demands of our power study precluded use of sophisticated B-spline implementations that are available

to choose the optimal number of knots and knot placements. If these methods had been used, then B-splines might sometimes have been identified as the optimal basis.

This study did not consider semiparametric statistical approaches (e.g., Pletcher and Geyer 1999; Izem and Kingsolver 2005), which can also be used to analyze function-valued data. These approaches hypothesize models or modes of variation thought to be important based on, for instance, physiological considerations. Parametric functions are derived for the modes of variation and then pieced together to model a complete functional response. An appealing aspect of the semiparametric approach is that biologists can attribute variation in functional responses to biologically intuitive modes or models of variation. It is unclear, however, if semiparametric approaches would lead to better statistical discrimination than nonparametric methods. Moreover, like parametric methods, semiparametric methods are vulnerable to mistaken intuition.

The function-valued approach can be directly applied in comparative phylogenetic studies. For instance, Dudycha and Lynch (2005) used the function-valued approach in combination with comparative phylogenetic methods to study trade-offs between growth and reproduction in species of Daphniidae. Our simulations suggest that nonparametric function-valued methods can substantially improve statistical discrimination in phylogenetic comparisons. Generally, more comparative phylogenetic studies of function-valued traits are needed to determine which aspects of functional phenotypes are conserved or radiate rapidly across species.

The nonparametric function-valued approach presented here has all of the useful properties of a parametric approach plus the added benefit that it does not require a priori assumptions about the nature of the underlying curve. Basis function expansions are continuous functions, enabling analyses of time-dependent rates of change, such as done in the work of Badyaev et al. (2001). With the multivariate approach, interesting questions related to rates of change can only be addressed in an ad hoc manner. Generally, compelling biological questions can be asked when data are modeled functionally that cannot be asked when data are modeled as a set of distinct values (see Gomulkiewicz and Kingsolver 2007 for discussion).

This article suggests that even for basic questions about differences in mean functions, the function-valued approach never has lower, and often has substantially better, power than the multivariate approach. It seems reasonable to expect that the statistical advantages of function-valued methods will extend to detecting differences among genotypes and in doing so enable superior estimates of the patterns of heritable variation and covariation within populations that are necessary to predict evolutionary change (Kirkpatrick and Heckman 1989; Kingsolver et al. 2001). These statistical advantages, as well as the capacity to address compelling questions related to functional phenotypes and the availability of accessible software packages implementing these powerful methods (such as R [R Development Core Team 2007], SAS software [SAS Institute Inc. Cary, NC, USA] and S+ [e.g. Ramsay and Silverman 2002; Clarkson et al. 2005]) build a strong case that function-valued approaches are a highly productive way to study functional data in evolutionary biology.

ACKNOWLEDGMENTS

We thank the function-valued trait group for comments on an earlier version of this article. The comments by J. Kingsolver and two anonymous reviewers were very helpful and appreciated. Funding was provided by National Science Foundation grant EF 0328594 to RG and NH. The Initiative for Bioinformatics and Evolutionary Studies program at the University of Idaho kindly provided computer resources and they are supported by NSF grant EPS0080935 and National Institutes of Health grants P20 RR16454 and P20 RR16448 from the Center for Biomedical Research Excellence and Idea Networks of Biomedical Research Excellence programs of the National Center for Research Resources.

LITERATURE CITED

Abramowitz, M., and I. A. Stegun. 1964. Handbook of mathematical functions. Dover, New York.

Akaike, H. 1976. An information criterion (AIC). Math. Sci. 14:5–9.

Badyaev, A. V., G. E. Hill, and L. A. Whittingham. 2001. The evolution of sexual size dimorphism in the house finch. IV. Population divergence in ontogeny. Evolution 55: 2534–2549.

Burnham, K. P., and D. R. Anderson. 1998. Model selection and inference, a practical information-theoretic approach. Springer, New York.

Clarkson, D. B., C. Fraley, C. C. Gu, and J. O. Ramsay. 2005. S+ functional data analysis: user's manual for windows. Springer, New York.

De Boor, C. 1978. A practical guide to splines. Springer-Verlag, New York.

Diggle, P., K. Y. Liang, and S. L. Zeger. 2002. Analysis of longitudinal data. Oxford Univ. Press, New York.

Dudycha J. L., and M. Lynch. 2005. Conserved ontogeny and allometric scaling of resource acquisition and allocation in the Daphniidae. Evolution 59: 565–576.

Gilchrist, G. W., and R. B. Huey. 2004. Plastic and genetic variation in wing loading as a function of temperature within and among parallel clines in *Drosophila subobscura*. Integr. Comp. Biol. 44: 461–470.

Gomulkiewicz, R., and J. G. Kingsolver. 2007. A fable of four functions: function-valued approaches in evolutionary biology. J. Evol. Biol. 20:20–21.

Hill, W. G. 1998. Inferences from evolutionary biology to livestock breeding. Proc. Sixth World Congress on Genetics Applied to Livestock Production 23: 32–39.

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

Johnson, R. A., and D. W. Wichern. 1988. Applied multivariate statistical analysis. Prentice Hall, Englewood Cliffs, NJ.

Kingsolver, J. C., R. Gomulkiewicz, and P. A. Carter. 2001. Variation, selection and evolution of function-valued traits. Genetica 112–113: 87–104.

Kingsolver, J. G., R. Izem, and G. Ragland. 2004. Plasiticity of size and growth in fluctuating thermal environments: comparing reaction norms and performance curves. Integr. Comp. Biol. 44: 450–460.

Kirkpatrick, M. and N. Heckman. 1989. A quantitative genetic model for growth, shape and other infinite-dimensional characters. J. Math. Biol. 27: 429–450.

Kirkpatrick, M., D. Lofsvold, and M. Bulmer. 1990. Analysis of inheritance, selection and evolution of growth trajectories. Genetics 124: 979–993.

Laird, N. M., and J. H. Ware. 1982. Random-effects models for longitudinal data. Biometrics 38: 963–974.

Lande, R. 1979. Quantitative genetic analysis of multivariate evolution, applied to brain:body size allometry. Evolution 33: 402–416.

———. 1980. The genetic covariance between characters maintained by pleiotropic mutations. Genetics 94: 203–215.

Lande, R., and S. J. Arnold. 1983. The measurement of selection on correlated characters. Evolution 37: 1210–1226.

Leng, X., and H. G. Müller. 2006. Classification using functional data analysis for temporal gene expression data. Bioinformatics 22: 68–76.

Liu C., and D. B. Rubin. 1994. The ECME algorithm: a simple extension of EM and ECM with faster monotone convergence. Biometrika 81: 633–648.

McCarroll, S. A., C. T. Murphy, S. Zou, S. D. Pletcher, C. S. Chin, Y. N. Jan, C. Kenyon, C. I. Bargmann, and H. Li. 2004. Comparing genomic expression patterns across species identifies shared transcriptional profile in aging. Nat. Genet. 36: 197–204.

Meyer, K., and W. G. Hill. 1997. Estimation of genetic and phenotypic covariances for longitudinal or repeated records by restricted maximum likelihood. Livestock Product. Sci. 47: 185–200.

Meyer, K., and M. Kirkpatrick. 2005. Up hill, down dale: quantitative genetics of curvaceous traits. Philos. Trans. R. Soc. Lond. B 360: 1442–1455.

Milliken, G. A., and D. E. Johnson. 1984. Analysis of messy data Vol 1: designed experiments. Lifetime Learning, Belmont.

Müller, H. G., and Y. Zhang. 2005. Time-varying functional regression for predicting remaining lifetime distributions from longitudinal trajectories. Biometrics 61: 1064–1075.

Pletcher, S. D., and C. J. Geyer. 1999. The genetic analysis of age-dependent traits: modeling the character process. Genetics 153: 825–835.

Pletcher S. D., D. Houle, and J. W. Curtsinger. 1998. Age-specific properties of spontaneous mutations affecting mortality in *Drosophila melanogaster*. Genetics 148: 287–304.

R Development Core Team. 2007. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.

Ramsay, J. O., and B. W. Silverman. 2002. Applied functional data analysis: methods and case studies. Springer, New York.

———. 2005. Functional data analysis. Springer, New York.

Searle, S. R., G. Casella, and C. E. McCulloch. 1992. Variance components. John Wiley & Sons, New York.

Storey, J. D., W. Xiao, J. T. Leek, R. G. Tompkins, and R. W. Davis. 2005. Significance analysis of time course microarray experiments. Proc. Natl. Acad. Sci. USA 102: 12837–12842.

Via, S. and R. Lande. 1985. Genotype-environment interaction and the evolution of phenotypic plasticity. Evolution 39: 505–522.

Winsor, C. P. 1932. The Gompertz curve as a growth curve. Proc. Natl. Acad. Sci. USA 18: 1–8.

Wolfram, S. 2005. The Mathematica book, 5th edn. Wolfram Media, Champaign, IL.

Wu, R. L. and M. Lin. 2006. Functional mapping—how to study the genetic architecture of dynamic complex traits. Nat. Rev. Genet. 7: 229–237.

Zhang, H. and X. Zhong. 2006. Linkage analysis of longitudinal data and design consideration. BMC Genetics 7: 37.

Associate Editor: D. Promislow

Supplementary Material

The following supplementary material is available for this article:

Figure S1. The power to reject the null hypothesis that the average function is the same for two populations when measurement error is small (standard deviation = 0.002) and the Gompertz growth curve is modeled.

Figure S2. The power to reject the null hypothesis that the average function is the same for two samples when measurement error is small (standard deviation = 0.002) and the cyclic growth curve is modeled.

Figure S3. The power to reject the null hypothesis that the average function is the same for two samples when measurement error is small (standard deviation = 0.002) and the performance growth curve is modeled.

This material is available as part of the online article from:
http://www.blackwell-synergy.com/doi/abs/10.1111/j.1558-5646.2008.00340.x
(This link will take you to the article abstract).

Please note: Blackwell Publishing is not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.