

EXPLORING FITNESS SURFACES

DOLPH SCHLUTER* AND DOUGLAS NYCHKA†

*Ecology Group, Department of Zoology, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada; †Department of Statistics, North Carolina State University, Raleigh, North Carolina 27695

Submitted January 8, 1993; Revised June 16, 1993; Accepted July 7, 1993

Abstract.—We present a nonparametric method to estimate the form of multivariate selection on a suite of quantitative traits. Its advantages are threefold. First, the procedure is flexible and does not force estimates of the surface to conform to a specific mathematical shape. The need for a flexible method is illustrated by an example using quadratic regression. Second, estimates of multidimensional surfaces can be visualized in two or three dimensions. This simplification is accomplished by making cross sections of the surface in the few most interesting directions. Finally, the method is designed to handle survival and other nonnormal fitness components. We apply the procedure to two data sets. In song sparrows, the survival surface is approximated by a ridge favoring an allometric relation between body mass and wing length. Survival in human infants rises steeply with increasing birth mass and maternal gestation period to a broad flat dome. Our results emphasize the ubiquity of “correlational” selection and illustrate how traits jointly determine fitness.

The “surface of selective value” is one of the most enduring representations of the evolutionary process. Wright (1932) likened natural selection to a multidimensional landscape of peaks and valleys, whose height at any point indicated fitness of the corresponding genotype. (In later discussions he focused on a related surface, the “adaptive landscape,” describing population mean fitness as a function of genotype frequency; see Provine 1986 for the history of these two concepts.) The concept of a selection surface can be extended to the continuously varying phenotype (Pearson 1903; Simpson 1953; Lande 1979; Lande and Arnold 1983; Schluter 1988; Phillips and Arnold 1989). In this case the selection surface or fitness function f relates survival or reproductive success W of individuals to the m phenotypic traits z_1, z_2, \dots, z_m under natural selection

$$W = f(z_1, z_2, \dots, z_m) + \text{random error}. \quad (1)$$

For example, survival of individuals over some part of the life history may depend on three traits: z_1 , body mass; z_2 , coloration intensity; and z_3 , running speed. In this case W is survival (one or zero), and $f(z_1, z_2, z_3)$ is the probability that an individual survives as a function of its measurements for the three traits.

The function f is a complete description of natural selection pressures on phenotypic traits (Schluter 1988). An estimate from field data would thus be invaluable. Knowledge of f would allow one to make quantitative predictions of survival or

reproductive success of individuals and to determine whether one or more fitness peaks, valleys, or ridges occur within the range of phenotypes present in the population. It would also indicate whether selection favors correlations between traits (Lande and Arnold 1983). Additionally, f can be used to determine which traits are under direct natural selection and which become modified only through their correlation with selected traits (Lande and Arnold 1983). Finally, the surface f can be regarded as a feature of the external environment that determines fitness of alternative phenotypes but exists independently of them (Schluter 1988; this view must be modified when selection is frequency-dependent). If the ecological importance of traits is known a priori, then f provides a description of the ecological environment in relevant units of fitness (Schluter and Grant 1984; Schluter 1988).

A second goal when studying natural selection on several variables is to measure coefficients of directional and quadratic selection intensity (Lande and Arnold 1983; Phillips and Arnold 1989). Such coefficients are necessary to predict evolutionary response to selection (Lande 1979). However, these coefficients do not correspond in any straightforward way to the form of the fitness surface f , particularly when the phenotype distribution is not multivariate normal (Mitchell-Olds and Shaw 1987; Schluter 1988). For example, whether fitness reaches a peak within the range of phenotypes in the population cannot usually be ascertained from the selection coefficients. For this reason, if the features of the selective surface are of interest (e.g., peaks, ridges, or valleys), then the most reliable way to detect them is to estimate the fitness surface directly. Here we suggest a novel way to accomplish this objective.

In an earlier article (Schluter 1988), a nonparametric method was used to estimate f from field data in the special case of a single phenotypic trait (i.e., $m = 1$). Below, we extend the method to estimating the form of selection on a multidimensional phenotype. We use the method to reanalyze two data sets previously handled in other ways (song sparrow data: Schluter and Smith 1986; infant mortality data: Karn and Penrose 1951). New interpretations of these data reveal the power of the approach.

For most applications, the resulting estimate of the selection surface is sufficiently interesting on its own that no further analysis is necessary. However, the estimate can also be used as a preliminary step to finding a suitable mathematical description of the surface, such as exponential or Gaussian.

THE FORM OF NATURAL SELECTION

Different terms are used by different authors to describe the form of natural selection, and to avoid confusion we clarify our definitions here. To be consistent with earlier definitions (Schluter 1988), we define fitness in terms of the phenotype: fitness is the mean survival or reproductive success of all individuals having the same phenotype z_1, z_2, \dots, z_m . The realized survival or reproductive success W of an individual is therefore the fitness of its phenotype plus a random error term (eq. [1]). Survival and reproductive success are actually components of true fitness, but for simplicity we use the general term *fitness* throughout.

Directional selection occurs on a single trait when average fitness is higher on

one side of the population mean than on the other side. In pure directional selection, f is nonincreasing or nondecreasing over the specified range of phenotypes. Stabilizing (or disruptive) selection occurs on a single trait when the fitness function f has a peak (or valley) within the range of phenotypes in the population. These definitions are the traditional ones (Endler 1986; Schluter 1988). Lande and Arnold (1983) provided new definitions for directional, stabilizing, and disruptive selection; but as they refer to coefficients rather than to features of f , the traditional terminology is preferred (Schluter 1988; Phillips and Arnold 1989).

Possible forms of selection in two and higher dimensions are so varied that a detailed classification would be laborious. To simplify, we classify multivariate selection by the form of the fitness surface f in specific cross sections. Each cross section j defines a new compound trait x_j that is a linear combination of the original traits z_i :

$$x_j = a_{j1}z_1 + a_{j2}z_2 + a_{j3}z_3 + \dots + a_{jm}z_m = \sum_{i=1}^m a_{ji}z_i,$$

which in vector notation is

$$x_j = \mathbf{a}'_j \mathbf{z},$$

where ' indicates transpose. The a_{ji} are constants representing the contributions of each trait z_i to the compound trait x_j . Since x_j is a single variable, selection on it may be directional, stabilizing, or disruptive as defined above. We define correlational selection to be the presence of stabilizing or disruptive selection on any linear combination x_j of the original traits, other than the trivial combinations (i.e., at least two of the $|a_{ji}| > 0$, $i = 1, \dots, m$). This definition departs from the original one (Lande and Arnold 1983; Phillips and Arnold 1989). We preferred it because we felt that correlational selection, like other forms of selection, should be based on features of the fitness surface rather than on parameters of a specific equation (e.g., the quadratic). The two definitions may conflict: for example, "correlational selection" by the original definition (Lande and Arnold 1983) encompasses forms of selection that we would classify instead as purely directional (fig. 1).

ESTIMATING THE FORM OF NATURAL SELECTION

One possible way to estimate the true selection surface is to use quadratic regression (Karn and Penrose 1951; Lande and Arnold 1983; Phillips and Arnold 1989). This method is effective if the surface is indeed quadratic (Phillips and Arnold 1989) but may mislead if the true surface has another mathematical form (e.g., truncation [fig. 1], exponential, asymptotic, asymmetric, bell-shaped, or polymodal selection; Schluter 1988). Hence, a more flexible method to explore the surface is desired. We propose the use of nonparametric regression to achieve this flexibility.

Another difficulty of estimating fitness surfaces is an abundance of variables (traits). For example, the number of parameters in any mathematical description of the surface becomes prohibitive when the number of traits increases beyond

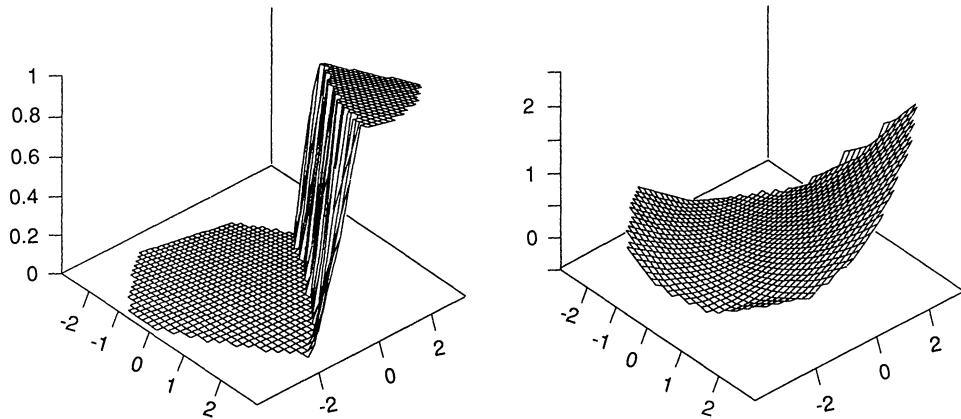


FIG. 1.—An example of pure directional selection on the multivariate phenotype. *Left*, truncation selection applied to a linear combination of two uncorrelated standard normal traits (mean = 0, variance = 1). Individuals survived if the sum of their two traits was $\geq \sqrt{2}$; otherwise they died. *Right*, the best-fit quadratic approximation to the surface on the left. Quadratic regression coefficients suggest “disruptive” ($\gamma_1 = \gamma_2 = 0.76$) and correlational ($\gamma_{12} = 0.76$) selection on the traits by the Lande and Arnold (1983) definitions; $n = \infty$.

three or four. Additionally, the multidimensional surface cannot be visualized easily in two or three dimensions. An ad hoc solution to the problem of many traits is to carry out a principal components analysis to locate the most variable phenotype dimensions (see, e.g., Lande and Arnold 1983; Schluter and Smith 1986). Estimates of selection are then applied to the first one or two principal components rather than to the larger number of original traits. Yet, there is no reason to believe that selection is confined to, or acts most strongly on, the one or two components chosen. Indeed, the approach must be regarded as arbitrary because the procedure chooses the components while disregarding the relationship between the multidimensional phenotype and fitness. For example, a component with large variability may have little influence on fitness.

Below, we describe a flexible nonparametric technique that overcomes the problems of many traits. We first describe the case of selection on a single trait. We then extend the method to explore surfaces of many dimensions.

Estimating Functions Using the Cubic Spline

A nonparametric procedure, the cubic spline, could be used to estimate the form of selection on a single trait (Schluter 1988). The cubic spline is a regression curve named for its resemblance to the mechanical spline, a pliable stick used for interpolation by draftspenople. Mathematically, the cubic spline is a collection of cubic polynomial pieces joined seamlessly end to end (Wahba 1990). The curve allows one to estimate a wide variety of possible fitness functions including the straight line, the bell-shaped curve, and even bimodal forms of selection (Schluter 1988; Smith 1990).

The smoothness of the spline curve fit to any set of data is controlled by a

nonnegative constant, λ (Craven and Wahba 1979). If λ is nearly zero, then the curve will be extremely rough, rising and falling many times as it passes through all the individual data points. Conversely, when λ is large, the curve is a simple linear regression. Other curves lie between these two extremes. Often termed the "smoothing parameter," λ acts similarly to the sliding interval of k points used in calculations of moving averages. The effect of alternative values of λ on estimates of a fitness function has been demonstrated (Schluter 1988).

The best λ for the data at hand is determined objectively by the predictive power of the corresponding function using the method of cross validation. The procedure involves deleting observations one at a time from the data set and predicting survival or reproductive success of the missing individual from the cubic spline curve fit with the remaining points. The generalized cross validation (GCV) score is a measure of the overall prediction error associated with a given value of λ (Craven and Wahba 1979). The best λ is that value minimizing the GCV score (see Schluter 1988 for examples).

A multivariate extension of the cubic spline can be used to estimate a selection surface involving many traits (Wahba 1990). O'Sullivan et al. (1986) present some biological examples involving two variables. The approach is flexible but has several disadvantages (Friedman and Stuetzle 1981). For our purposes, its main drawbacks are the large sample size needed and the difficulty of visualizing the surface when the number of traits exceeds two. We overcome these problems with a simpler method that combines a small number of univariate splines.

Projection Pursuit Regression

Definition.—The procedure we recommend is founded on the premise that selection does not act strongly on all phenotype dimensions, and hence the entire fitness surface is not worth looking at. We take cross sections of the surface in the directions of strongest selection using the method of projection pursuit regression (Friedman and Stuetzle 1981). We are then able to visualize the surface in reduced form.

The projection pursuit approximation of the true selection surface f is

$$f(\mathbf{z}) \approx f_1(\mathbf{a}'_1\mathbf{z}) + f_2(\mathbf{a}'_2\mathbf{z}) + \dots + f_p(\mathbf{a}'_p\mathbf{z}) = f_1(x_1) + f_2(x_2) + \dots + f_p(x_p). \quad (2)$$

Here, f_1, f_2, \dots, f_p are each single-variable regressions referred to as *ridge functions*, and \mathbf{z} is the vector of original phenotypic traits with means set to zero. Each x_j is a single variable computed as a linear combination of the original traits. By *projections* we mean the vectors of constants \mathbf{a}_j that identify the directions of each cross section; *pursuit* refers to the computational steps needed to find them. Surfaces of general form can be represented with this method, including those involving interactions between traits (Friedman and Stuetzle 1981). To standardize the directions we fix $\mathbf{a}'_j\mathbf{a}_j = 1$.

Simple example.—The essence of projection pursuit can be made clear using the simplest case: Imagine that selection acts mainly in one direction \mathbf{a}_1 , on the linear combination of traits $x_1 = \mathbf{a}'_1\mathbf{z}$:

$$f(\mathbf{z}) \approx f_1(\mathbf{a}'_1\mathbf{z}) = f_1(x_1).$$

This implies that fitness is unvarying along directions orthogonal to \mathbf{a}_1 . In this case all the important features of the surface may be seen by making a cross section in the single direction \mathbf{a}_1 . Assume further that the form of selection on this direction is linear

$$W - \bar{W} \approx f_1(x_1) + e = bx_1 + e.$$

Here b is the slope of the line, \bar{W} is mean survival or reproductive success, and e is random error.

The problem of estimating the selection surface is now reduced to finding the single direction \mathbf{a}_1 and the constant b . This search may be carried out by iterating between two steps: first, search over possible values for \mathbf{a}_1 subject to the constraint $\mathbf{a}'_1\mathbf{a}_1 = 1$; then for each choice of \mathbf{a}_1 , find b to minimize the weighted residual sum of squares:

$$\text{WRSS} = \sum_{k=1}^n \frac{(W_k - b\mathbf{a}'_1\mathbf{z}_k)^2}{\text{var}(W_k)},$$

where n is the number of observations, \mathbf{z}_k is the vector of measurements for the k th individual, and $\text{var}(W_k)$ is the variance of the residuals at $\mathbf{z} = \mathbf{z}_k$. It can be shown that, when the estimates converge, the resulting vector of coefficients $\hat{\mathbf{a}}_1$, when multiplied by the estimate of b , is identical with the coefficients β obtained from a multiple linear regression of $W - \bar{W}$ on the traits \mathbf{z} . That is, multiple regression is analogous to identifying the single direction on which linear selection uniquely acts (Lande 1979) and is therefore a special case of projection pursuit regression. The iterative algorithm is simply a very inefficient way to find β .

The single direction \mathbf{a}_1 under selection may also be interpreted as representing a "latent variable." Such an interpretation would be analogous to viewing the first principal component of variation in morphology as representing an underlying latent variable called "size" (Crespi and Bookstein 1989). The advantage of projection pursuit is that it does not restrict attention a priori to size; all potential directions are considered.

General case.—True projection pursuit generalizes the above simple case in two ways. First, we let the functions f_j be of any form, including linear, exponential, bell shaped, bimodal, or other. We use nonparametric regression (the cubic spline) to estimate these functions. Second, we allow for selection in more than one direction. In practice we will assume that the surface can be approximated using only one or a small number of directions.

METHODS

Projection pursuit approximates the true surface f (eq. [2]) using p directions:

$$f(\mathbf{z}) \approx \sum_{j=1}^p f_j(\mathbf{a}'_j\mathbf{z}) = \sum_{j=1}^p f_j(x_j).$$

To approximate f we must solve two computational problems. First, the direc-

tions \mathbf{a}_j must be found. Second, the functions f_j must be estimated. Below, we outline how these problems were solved.

Directions.—We found the directions \mathbf{a}_j using the iterative steps given earlier, minimizing the weighted residual sum of squares:

$$\text{WRSS} = \sum_{k=1}^n \frac{(W_k - \hat{f}(\mathbf{z}_k))^2}{\text{var}(W_k)}, \quad (3)$$

where $\hat{f}(\mathbf{z}_k)$ is the predicted survival or reproductive success of individual k , based on its phenotype \mathbf{z}_k . If the residuals $W_k - \hat{f}(\mathbf{z}_k)$ have a normal distribution with equal variance, then $\text{var}(W_k)$ is a constant. If W_k is survival (i.e., is zero or one), $\text{var}(W_k)$ is estimated by the binomial variance $\hat{f}(\mathbf{z}_k)(1 - \hat{f}(\mathbf{z}_k))$. Similarly, if W_k is Poisson distributed (e.g., number of mates), we estimate $\text{var}(W_k)$ by $\hat{f}(\mathbf{z}_k)$.

Functions.—For each fixed direction \mathbf{a}_j we estimated f_j from the data using the cubic spline regression. If survival or reproductive success W_k corresponding to a given phenotype \mathbf{z}_k is normally distributed, then for each projection \mathbf{a}_j the estimate \hat{f}_j is calculated as a univariate regression of $W - (\sum_{l \neq j} \hat{f}_l)$ on $x_j = \mathbf{a}_j' \mathbf{z}$. This procedure is modified slightly if the fitness measurement is binomial (e.g., survival) or Poisson (e.g., number of mates) using the tools of generalized linear models (see Appendix; McCullagh and Nelder 1983; O'Sullivan et al. 1986). For simplicity, we applied the same value of the smoothing parameter λ to each function f_j in f .

Algorithm.—Unfortunately, there was no analytical route to estimating the selection surface in the general case, and a computer-intensive search for the directions \mathbf{a}_j and functions f_j was required. Our search was done efficiently by employing an iterative algorithm known as “backfitting,” which involves two steps. For a predetermined number of directions p , loop through the directions \mathbf{a}_j , $j = 1, \dots, p$. Then, holding all \mathbf{a}_l fixed, $l \neq j$, estimate \mathbf{a}_j by minimizing WRSS of equation (3). These steps are repeated until the estimates converge. For each direction j the first part of the minimization inspects a large number of possible values for \mathbf{a}_j created using random numbers. Subsequent iterations refine this direction (i.e., to find the one minimizing WRSS) using a simplex algorithm (Nelder and Mead 1965). Note that at any stage the WRSS is only being minimized over the j th direction. However, if this procedure converges on a solution, it must be a local minimum of the WRSS. This strategy is called backfitting because, having found a candidate \mathbf{a}_j for a given direction j , one cycles back and reestimates the other directions.

An interactive computer program to carry out projection pursuit regression can be obtained from us on request. The program was written in FORTRAN77 and is available as source code or a compiled program for the Sun workstation or PC-compatible microcomputer.

Data

We analyzed two data sets. The first consisted of measurements of overwinter survival (recruitment) of 152 juvenile male song sparrows (*Melospiza melodia*) over 5 yr on Mandarte Island, British Columbia (Schluter and Smith 1986). Six

external morphological traits were measured: mass (grams); wing length; tarsus length; and length, depth, and width of the beak (all millimeters).

The second data set provided measurements of survival to 28 d of 7,037 newborn male human infants (Karn and Penrose 1951). Two traits were available, infant mass at birth (pounds) and maternal gestation period (days). We obtained virtually identical results when the data were analyzed with and without a single outlying point at large birth mass. Results presented here exclude the point.

Karn and Penrose (1951) visualized the fitness surface for the human infant data using quadratic regression. The results suggested stabilizing selection on birth weight and correlational selection on weight and gestation period. These data are a classic example of selection in our own species, and they suggest how a maternal trait (gestation) and an offspring trait (mass) interact to determine offspring fitness.

In each data set we hunted for the projection(s) using a series of fixed values of the smoothing parameter λ . We then chose the value of λ such that the corresponding estimate minimized the GCV score (Craven and Wahba 1979). In the case of a single projection (i.e., $p = 1$), we used the standard formula for this score. When hunting for two dimensions (i.e., $p = 2$), we calculated the GCV score using a modification of the standard formula (see Appendix). In all cases we also checked the repeatability of estimated directions a_j by carrying out the analysis several times using different random number seeds for the initial search (see first step described earlier).

All phenotypic traits were standardized to mean zero and variance one prior to analysis. In both data sets, individual measurements of survival were either zero or one, and we used the cubic spline method appropriate for binary data (see Appendix; O'Sullivan et al. 1986; Schluter 1988). Our computer program for projection pursuit regression provides estimates of fitness $f(\mathbf{z}_k)$ for each individual k in the sample. These values were then imported into a graphics program that drew surfaces by interpolation.

RESULTS

Song Sparrows

Because of small sample size ($n = 152$), we restricted the search to a single cross section of the surface. That is, we assumed that selection was strong in one direction only and hence that the true selection surface in seven dimensions could be approximated by a simple univariate function. The best single direction, that explaining most of the variation in survival among individuals, is given in table 1. The estimated form of selection in this single direction is illustrated in figure 2.

The contributions of the six traits to the best direction were not equal (table 1). Body mass and wing length contributed most of all, and we tentatively conclude that variation among individuals in probability of survival was determined chiefly by these two traits. Their effects, though similar in magnitude, were opposite in sign. This result has interesting biological significance and is interpreted below.

TABLE 1

THE SINGLE DIRECTION a_1 OF STRONGEST NATURAL
SELECTION ON JUVENILE MALE SONG SPARROWS

Trait	Direction a_1
Mass	.61
Wing length	-.73
Tarsus length	.10
Beak length	.19
Beak depth	.07
Beak width	.22

NOTE.—Coefficients measure the contributions of each standardized trait to the direction. The smoothing parameter used was that minimizing the GCV score: $\ln(\lambda) = -6$; $n = 152$.

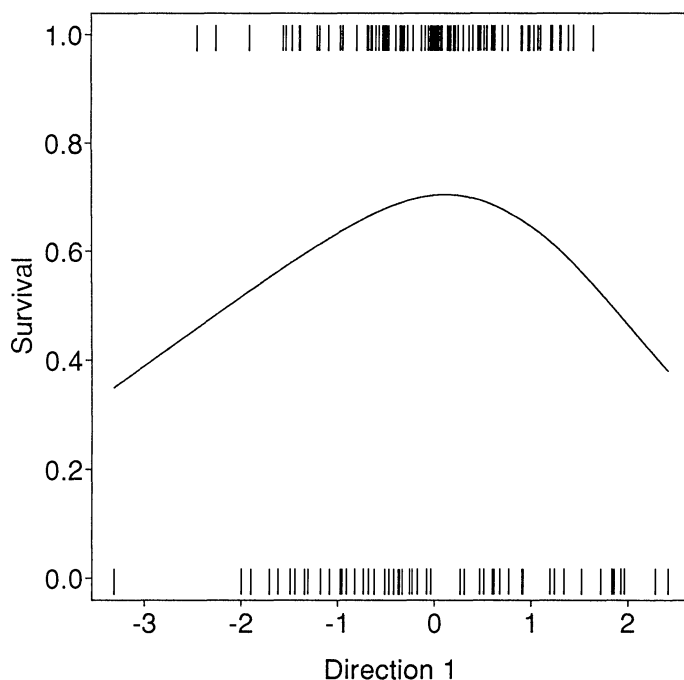


FIG. 2.—Survival (recruitment) of juvenile male song sparrows as a function of morphology. Direction 1 is the linear combination a_1 of the six original traits (standardized) described in table 1. This combination is the direction estimated to be experiencing strongest selection. Vertical marks are the original observations. $\ln(\lambda) = -6$; $n = 152$.

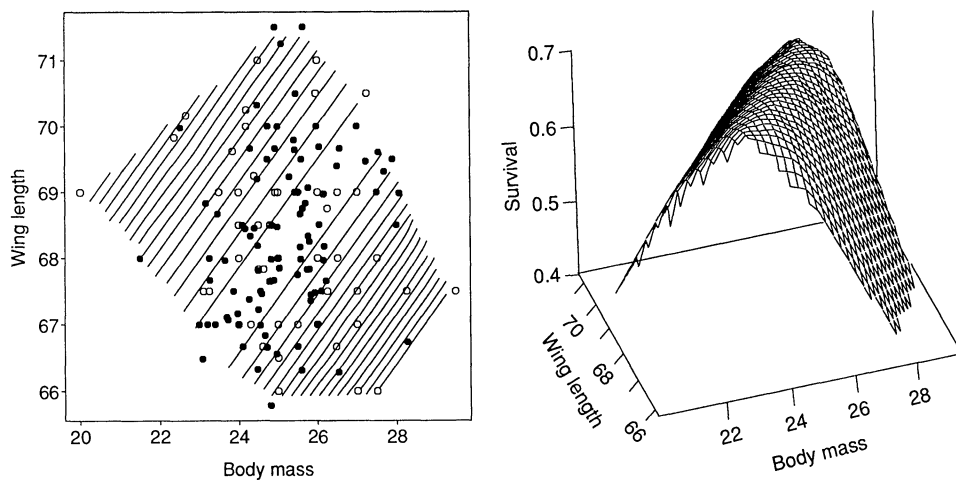


FIG. 3.—Survival (recruitment) of juvenile male song sparrows in relation to wing length and body mass. Symbols on *left panel* indicate measurements of individuals and whether they survived (*filled*) or disappeared (*open*). Fitness contours describe a ridge oriented from lower left to upper right, with survival decreasing to either side. *Right panel* gives a three-dimensional perspective of the surface. $\ln(\lambda) = -6$; $n = 152$.

Natural selection was stabilizing in this single best direction (fig. 2). Males with intermediate morphology had highest mean survival, whereas extreme males had lowest survival. Because at least two of the original traits are represented in the linear combination x_1 (i.e., wing length and body mass; table 1), stabilizing selection on it implies correlational selection on the multivariate phenotype. Males with lowest survival were those of large mass and short wings and of low mass and long wings. In effect, stabilizing selection acted on a measure related to “wing loading.”

Correlational selection on song sparrow morphology is most easily visualized by plotting survival against the two most important traits, wing length and body mass. The resulting fitness surface is a ridge over the distribution of individual measurements (fig. 3). The ridge is oriented at an angle to both traits, such that selection worked against individuals whose bodies were too heavy or too light for a given wing length. Selection thus favored a positive correlation between wing length and body mass. We emphasize that the simple structure of the fitness surface (fig. 3) results from our allowing only a single cross section. More complex surface estimates may result when more than one direction is allowed, as in the larger data set analyzed next.

Human Infants

We began the analysis of human infants by hunting for the single phenotype dimension on which selection was strongest (table 2). Both traits, maternal gestation period and infant mass, contributed to the best direction, but the latter trait was most influential. Selection in this direction was chiefly directional (fig. 4):

TABLE 2
 THE SINGLE DIRECTION a_1 BEST EXPLAINING
 VARIATION IN SURVIVAL AMONG MALE HUMAN
 INFANTS

Trait	Direction a_1 (SE)
Maternal gestation period	.44 (.12)
Infant birth mass	.90 (.07)

NOTE.—Coefficients measure the contributions of each standardized trait to the direction. The smoothing parameter was the value minimizing the GCV score: $\ln(\lambda) = -10$. Standard errors were based on 100 bootstrap replicates; $n = 7,036$.

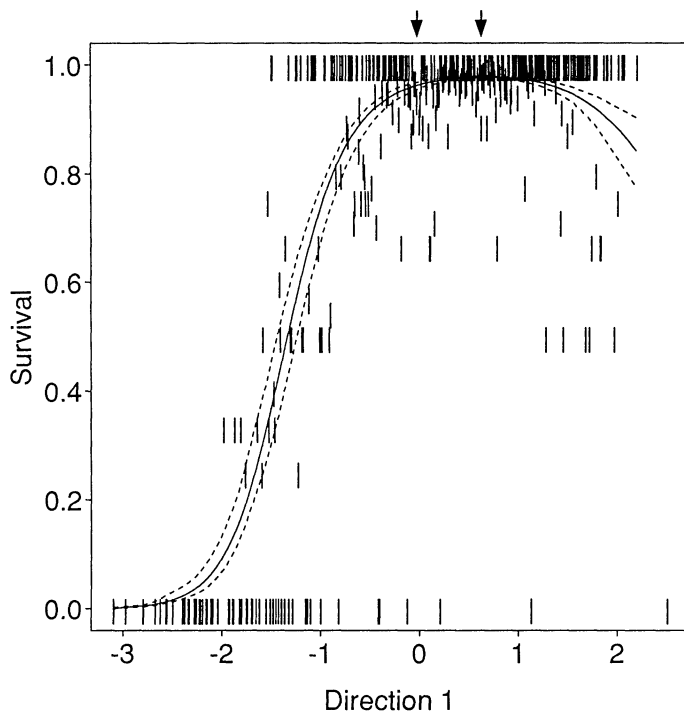


FIG. 4.—Survival probability of male human infants as a function of variation in the direction a_1 experiencing strongest selection (*solid curve*). The direction is determined chiefly by infant mass but also duration of maternal gestation (table 2). *Dashed curves* are ± 1 SE of the predicted probability of survival. *Vertical marks* indicate the average survival of infants in each mass and gestation class. *Arrows* indicate mean (*left*) and optimum (*right*) values for the x variable. The most extreme observation at large size was excluded from the analysis. $\ln(\lambda) = -10$; $n = 7,036$.

TABLE 3
THE TWO DIRECTIONS \mathbf{a}_1 AND \mathbf{a}_2 THAT JOINTLY BEST EXPLAIN
VARIATION IN SURVIVAL AMONG MALE HUMAN INFANTS

Trait	Direction \mathbf{a}_1	Direction \mathbf{a}_2
Maternal gestation period	.59	.70
Infant birth mass	.80	-.71

NOTE.—Elements (coefficients) of \mathbf{a}_1 and \mathbf{a}_2 measure the contributions of each standardized trait to the two directions. The smoothing parameter was that minimizing the GCV score: $\ln(\lambda) = -10$; $n = 7,036$.

survival probability increased steeply with increasing x_1 , rising to a broad plateau before dipping slightly at the upper extreme. The decline in survival at the upper end of the distribution indicates that stabilizing selection is also present. Correlational selection on mass and gestation is therefore also implied, since x_1 is a linear combination of them.

We carried the analysis one step further by repeating it and searching for two cross sections of the surface rather than just one. That is, we investigated the relationship between male infant survival and its mass and gestation period by hunting for the pair of directions, \mathbf{a}_1 and \mathbf{a}_2 , under strongest selection. The coefficients of the first of these two directions, \mathbf{a}_1 (table 3), were similar to those from the previous analysis (table 2), although they were not identical. This is to be expected because the backfitting algorithm essentially estimates the two directions simultaneously rather than sequentially. Hence, the two directions that together best explain variation in survival may not include the direction that is best when used alone.

Both mass and gestation also contributed to the second direction \mathbf{a}_2 . This second direction was also determined by both traits, mass and gestation, but the coefficients were of opposite sign (table 3). This axis thus distinguishes individuals with large mass and low gestation from those with low mass and long gestations. As shown in figure 4, selection on this axis is weakly stabilizing; hence, it is a second source of correlational selection. Note that the two directions are not orthogonal (perpendicular), as we have not so constrained them.

The estimated bivariate selection surface is a relatively flat dome with a steep descent at low mass and short gestation (fig. 5). Selection clearly acted jointly on both mass and gestation, rather than on each trait separately, such that an individual with low mass but moderate gestation had the same fitness as one with greater mass but shorter gestation. Survival declined slightly to the right of the dome at the greatest masses (fig. 5). Survival also declined above the dome at long gestation, although more weakly (a weak decline at the edges of the dome occurred whether or not the single observation at large mass was included). The environment thus appeared to be relatively tolerant of long gestation periods unless accompanied by large mass.

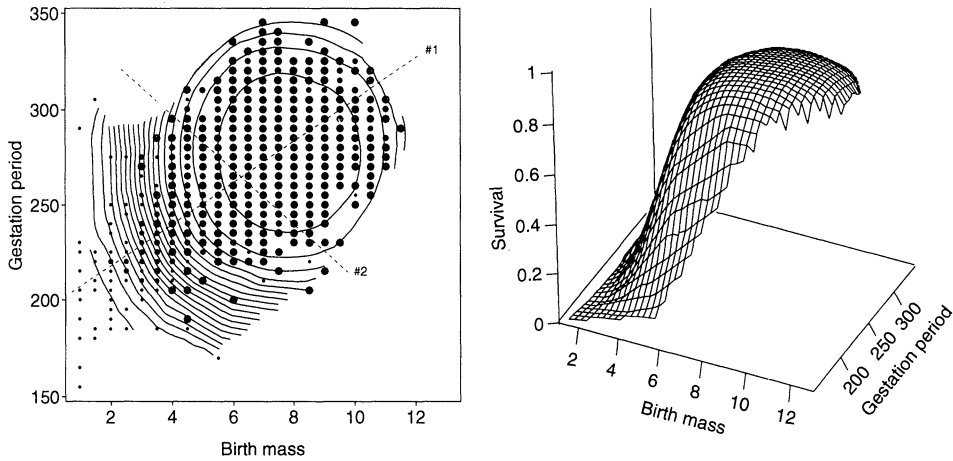


FIG. 5.—Survival probability of male human infants as a function of birth mass (pounds) and gestation period (days). The surface was approximated using two directions (table 3) indicated by the *dashed lines*. Fitness contours in the *left panel* are in increments of 0.5. They describe a dome rising steeply from lower left to a broad plateau above and right of center, declining slightly along its outer edges. *Right panel* gives three-dimensional perspective. Symbol size increases with increasing average survival of individuals in each mass (rounded to nearest 0.5 lb.) and gestation class (rounded to nearest 5 d); differences in symbol size are exaggerated in the range 0.90–1.00 to highlight slight survival differences. The most extreme observation at large mass was excluded from the analysis. $\ln(\lambda) = -10$; $n = 7,036$.

Sampling Variability and Significance Tests

The sampling distributions of fitness surface estimates are unknown, but one may calculate approximate standard errors and significance levels using the bootstrap (Efron and Tibshirani 1986). Here we illustrate the procedure using the human infant data. We have not carried out simulations to test the validity of the bootstrap for inference of this kind or to assess its power. Results should thus be interpreted cautiously.

We calculated standard errors for the single direction \mathbf{a}_1 using 100 bootstrap replicates (table 2). Each replicate was created by assigning every individual k a new survival measurement W_k (zero or one) generated randomly, where the probability that $W_k = 1$ was its predicted survival $\hat{f}(z_k)$ in the original estimate of the fitness surface (fig. 4). Projection pursuit regression was then carried out on each bootstrap replicate using a constant value for the smoothing parameter, $\ln(\lambda) = -10$. The standard error was computed as the standard deviation of the 100 replicate directions. These errors were relatively low (table 2). For each of the coefficients of \mathbf{a}_1 , we also counted the fraction of bootstrap replicates in which the coefficient was of different sign than the original estimate. This fraction provides an approximate P value for a test of the null hypothesis that the true

coefficient is zero. None of the bootstrap coefficients was of different sign than the original estimates. Hence, both coefficients of \mathbf{a}_1 (table 2) are significant at level $P < .01$.

We additionally recorded the predicted survival probabilities of every individual in all 100 of the above bootstrap replicates. The standard deviation of the replicate values for a given individual is the standard error of its survival probability. These standard errors are plotted against the original direction \mathbf{a}_1 in figure 4. Standard errors are small, which indicates that predicted survival probabilities are relatively accurate.

The bootstrap may also be used to test the significance of added directions. That is, one can ask whether an estimate of the surface that employs $p + 1$ cross sections fits the data significantly better than an estimate based on only p directions. We used the following steps to address the question with male human infants. First, we estimated the "null" surface using the original data and p directions and calculated the resulting goodness of fit $L(\hat{f}, p)$ (see below). We then generated 50 replicate data sets by resampling. Each replicate involved randomly assigning new values of survival W (zero or one) to every individual in the sample, where the probability that $W = 1$ was the predicted probability of survival from the surface based on p directions. For each bootstrap replicate we calculated the fitness surface using $p + 1$ directions and computed the resulting goodness of fit. This step measures the improvement in fit resulting when $p + 1$ directions are used on data generated randomly from a surface having only p directions. Finally, we estimated the fitness surface from the original data using $p + 1$ directions and the corresponding fit $L(\hat{f}, p + 1)$. The fraction of bootstrap replicates whose fit exceeded $L(\hat{f}, p + 1)$ yielded an approximate P value for a test of the null hypothesis that the true surface has only p directions. As a measure of goodness of fit we used the binomial log-likelihood

$$L(\hat{f}) = \sum_{k=1}^n W_k \ln(\hat{f}(z_k)) + (1 - W_k) \ln(1 - \hat{f}(z_k)).$$

Our tests employed a constant value of the smoothing parameter, $\ln(\lambda = -10)$.

To begin, we set the number of directions $p = 0$ and asked whether adding a single direction \mathbf{a}_1 led to a significant improvement in fit (table 2; fig. 4). The null surface was estimated by a constant (the mean survival of all individuals in the original sample), which yielded $L(\hat{f}, 0) = -1,358$. This fit was substantially worse than that of the estimate based on $p = 1$ added direction ($L(\hat{f}, 1) = -1,038$). The latter quantity also exceeded the log-likelihoods of all 50 bootstrap replicates (range = $-1,211$ to $-1,457$). Hence, the first direction \mathbf{a}_1 was significant at $P < 1/50$.

Next, we tested whether two directions (fig. 5) yielded a significantly better fit to the original data than only one direction. Adding the second direction improved the fit marginally ($L(\hat{f}, 2) = -1,028$), an increase that was not significant ($P = 20/50$).

DISCUSSION

Visualizing multivariate selection is a complex endeavor, but we recommend that the task is worth the effort. Our reason is simply that natural selection acts on suites of traits, as our examples demonstrate, and not separately on individual traits. The most vivid example of this principle is correlational selection (figs. 2–5; Lande and Arnold 1983; Phillips and Arnold 1989). Correlational selection is expected whenever a vital function of an organism depends jointly on two or more traits. One example is flight performance, which should depend on mass relative to wing form as well as to other aspects of the phenotype. Our suspicion is that correlational selection is the most prevalent form of selection in nature. However, univariate methods will never allow us to test this.

Our goal was to provide a compact representation of a selection surface. Rather than attempting to visualize the full surface, our approach estimates the most critical directions and takes cross sections, which thereby allows the surface to be viewed in reduced form. We have presented a method that accomplishes this reduction and have demonstrated its use. A further advantage of our method is that it is nonparametric: our functions are not restricted to straight lines or parabolas but may assume any form. Finally, we outlined statistical procedures to assess the accuracy of estimated surfaces and to test the significance of added directions.

In song sparrows, we detected correlational selection on wing length and body mass (fig. 3). Individuals of highest fitness were those lying along a line of allometry extending from low mass and short wings to high mass and long wings. Selection thus favored a positive correlation between the two traits. The pattern suggests selection on wing loading, possibly through its effects on flight performance, although the link with performance needs to be tested.

There are few other examples of correlational selection on a pair of traits. A preliminary analysis of the same song sparrow data showed stabilizing selection on a shape principal component (Schluter and Smith 1986). Brodie (1992) demonstrated disruptive selection in snakes along a composite axis of color pattern and escape behavior. Swain (1992) revealed stabilizing selective predation on the ratio of caudal to precaudal vertebrae number in sticklebacks. Simms and Rausher (1993) demonstrated stabilizing and disruptive selection on ratios of parasite resistance traits in morning glory. Jayne and Bennett (1990) indirectly suggested correlational selection in snakes by demonstrating stabilizing selection on residuals from a regression of mass on body length. Similarly, Lindén et al. (1992) showed stabilizing selection on residuals from a regression of mass on tarsus length in collared flycatchers and great tits. Stabilizing or disruptive selection on size-corrected residuals may result from correlational selection between the trait and size.

The song sparrow example gives one indication of how a univariate analysis of selection can be misleading. The estimated surface was a ridge oriented at an angle to both of the individual traits (fig. 3). When each of the two traits is inspected separately, selection appears to be stabilizing. For example, when aver-

aged over wing lengths, males of intermediate mass had higher survival than males at the upper or lower extremes. On its own the pattern would suggest that mass is at an evolutionary optimum, and the population is near equilibrium for this trait. However, if the true function in higher dimensions is a ridge as illustrated in figure 3, then there is no unique optimum phenotype or equilibrium. Instead, there is a line all of whose infinite number of points have equal fitness and between which the population is free to drift. Endler's (1986) survey of the literature revealed many examples of apparent stabilizing selection on single traits; yet, many might represent ridges in higher dimensions.

Note, however, that the simplicity of the fitness surface in our first example (fig. 3) is imposed by a projection pursuit model in which only one principal direction was sought. Further wrinkles of the surface are expected to appear when more directions are added. For example, adding a second direction to the bivariate surface in figure 3 revealed that the ridge was slightly higher at low mass and short wings than at the other extreme.

Selection on human infants was mainly directional, but weak stabilizing and correlational selection was also detected (fig. 5). This example of multivariate selection is particularly interesting because a child's fitness appeared to be determined jointly by its own phenotype (mass) and that of its mother (gestation; Lande and Kirkpatrick 1989). Also interesting is the large difference between mean and optimum (fig. 4), which produces the persistent directional selection. Two mechanisms explain how this difference might be evolutionarily stable. Zhivotovsky and Feldman's (1992) polygenic model leads to a stable difference between mean and optimum under two assumptions: that the optimum differs from the midpoint of all possible genotype values and that heterozygotes have smaller nongenetic components of variance than homozygotes. While their proposal may be part of the explanation, it is also likely that apparent directional selection on mass results from the positive effects of (nonheritable) offspring nutrition on both mass and survival (Price et al. 1988).

The infant data are from postwar England 1945–1950 (Karn and Penrose 1951). Ullizzi and Terrenato (1987) have followed the declining intensities of selection on infant mass over subsequent decades in association with improved medical practices (see also Terrenato et al. 1981; Ullizzi and Manzotti 1988). It would be interesting to document further the changing joint effects of mass and gestation on survival.

Despite its advantages, our approach does not eradicate all problems of multidimensional analysis. The main problem is that unmeasured variables may also affect fitness. Their absence from the analysis is likely to bias the estimate of the surface as well as estimates of the direct contributions of different traits to fitness (Lande and Arnold 1983; Mitchell-Olds and Shaw 1987). For example, the apparent effect of maternal gestation on human infant survival might be mediated by unmeasured aspects of infant phenotype such as stage of organ development (e.g., the lungs). If so, then the observed effect of gestation period would disappear once these offspring traits are included in the analysis.

Another problem is that traits may be strongly correlated with one another, which would lead to unstable estimates of the contributions of individual traits to fitness (Mitchell-Olds and Shaw 1987; Crespi and Bookstein 1989). This result

occurs because apparently very different linear combinations of highly correlated traits may fit the data nearly equally well. Even in such cases, however, our method will give a good estimate of the form of the selection surface, but it will not necessarily identify the traits on which selection principally acts. Finally, if selection acts equally strongly on many different combinations of traits, then the method will not be able to estimate all of them. Nevertheless, we expect that in such cases a few cross sections will give some insight into the modes of selection.

These problems emphasize that, when data are merely observational, as in the present examples, multivariate surface estimation is ultimately an exploratory procedure, whose results would require confirmation by experimental tests and by investigation of trait function. The method would also be useful to test predictions about the shape of the adaptive topography generated from a priori functional considerations. For example, an earlier article (Schluter and Grant 1984) predicted the adaptive landscape for beak size in Darwin's finches from seed abundance and known relations between seed size/hardness and beak morphology. A subsequent article (Schluter et al. 1985) reported a preliminary test of the shape of these functions using a field study of natural selection. Good recent examples of this approach with bivariate data are Brodie's (1992) prediction of joint selection on stripedness and escape behavior in snakes and Swain's (1992) prediction of the form of selective predation on sticklebacks. The next step in such studies would be to manipulate the phenotype experimentally, modifying traits singly and together and observing the fitness consequences. In addition to exploration, the present methods will be useful in analyzing such experimental results.

ACKNOWLEDGMENTS

Thanks again to J. Smith for allowing us to use his song sparrow data and for making useful comments on the manuscript. Thanks also to T. Hatfield, M. Kirkpatrick, M. Lindén, P. Phillips, T. Smith, D. Ward, and an anonymous reviewer for comments. Our work was supported by the Natural Sciences and Engineering Research Council (Canada) and the National Science Foundation (U.S.A.).

APPENDIX

DETAILS OF COMPUTATION

This appendix gives some details on computation of the projection pursuit estimate of the fitness surface.

ADDITIVE MODEL FOR FITNESS

To begin, assume that fitness is a continuous random variable (e.g., is normally distributed). The projection pursuit model for the fitness surface is then

$$W_k = \sum_{j=1}^p f_j(\mathbf{a}'_j \mathbf{z}_k) + e_k, \quad (\text{A1})$$

where ' indicates transpose, W_k is survival or reproductive success of the k th individual, e_k is a random error term with expectation $E(e_k) = 0$, and variance $\text{var}(e_k) = \text{var}(W_k)$ is

a known function of $E(W_k)$. The vectors, $\mathbf{a}_1, \dots, \mathbf{a}_p$ denote the p directions in the projection pursuit model and f_1, \dots, f_p the corresponding ridge functions.

For a fixed set of directions $\mathbf{a}_1, \dots, \mathbf{a}_p$, the ridge functions are estimated by minimizing

$$\frac{1}{n} \sum_{k=1}^n (W_k - \sum_{j=1}^p f_j(\mathbf{a}'_j \mathbf{z}_k))^2 / \text{var}(W_k) + \lambda \sum_{j=1}^p \int (f''_j(x))^2 dx \tag{A2}$$

over all possible ridge functions. Note λ is a constant and $\int (f''_j(x))^2 dx$ measures the roughness of the ridge function f_j . Technically, the estimates f_j satisfying this minimization are cubic splines governed by the smoothing parameter λ .

Now let $\hat{W}_k = \sum_{j=1}^p f_j(\mathbf{a}'_j \mathbf{z}_k)$ denote the predicted value for W_k based on the set of directions $\mathbf{a}_1, \dots, \mathbf{a}_p$ and on those ridge functions satisfying the minimization given above. The directions are estimated by minimizing the weighted residual sum of squares:

$$\sum_{k=1}^n \frac{(W_k - \hat{W}_k)^2}{\text{var}(W_k)} \tag{A3}$$

over all possible choices of directions $\mathbf{a}_1, \dots, \mathbf{a}_p$. As mentioned in the text, this minimization is approached using the backfitting algorithm: we attack one direction at a time and iterate among the directions until a global solution is achieved. The advantage of this iterative approach is that at any step only a single ridge function is being estimated. This method is much simpler than attempting to minimize equation (A3) across all the directions simultaneously, yet it achieves the same goal.

EFFECTIVE NUMBER OF PARAMETERS

The individual estimates of the ridge functions f_j also give some approximate statistical information about the surface. In particular they can be used to calculate the "effective" number of parameters used to define the fitness surface. We describe this quantity next.

Suppose that in the backfitting algorithm the l th direction is to be refined. For a new choice of \mathbf{a}_l , f_j (where $j \neq l$) is reestimated by minimizing equation (A2) holding the projections and the other ridge functions constant. This is the well-known univariate smoothing problem and is equivalent to estimating f_j in the model

$$Y_k = f_j(x_{jk}) + e_k, \tag{A4}$$

where $Y_k = W_k - \sum_{j \neq l} f_j(\mathbf{a}'_j \mathbf{z}_k)$ and $x_{jk} = \mathbf{a}'_j \mathbf{z}_k$. Associated with the cubic spline estimate of f_j is an $n \times n$ "hat" matrix $\mathbf{A}_j(\lambda)$ that maps the adjusted fitness measurements Y_k onto the predicted values based on the estimated ridge function $\hat{Y}_k = \hat{f}_j(x_{jk})$. Then in vector notation $\hat{\mathbf{Y}} = \mathbf{A}_j(\lambda) \mathbf{Y}$.

Theory for spline estimates suggests that the trace of $\mathbf{A}_j(\lambda)$ measures the effective number of parameters needed to reproduce the estimated curve. This index is useful in interpreting the flexibility of the spline estimate, which is fundamentally a nonparametric curve. For example, if λ is large and the estimate is approximately linear, then $\text{trace}(\mathbf{A}_j(\lambda)) = 2$, exactly as in linear regression. If $\nu_j = \text{trace}(\mathbf{A}_j(\lambda))$ then a good measure of the total effective number of parameters of the fitness surface is $\sum_{j=1}^p \nu_j$. This quantity depends on the smoothing parameter.

GENERALIZED ADDITIVE MODELS

Some modifications need to be made to the fitness surface estimates when the fitness measurements W_k do not have a continuous distribution. For example, survival is measured as zero or one. The general principle is to use an appropriate transformation of the additive relationship. For example, if $W_k = 0$ or 1 , one could use the logistic transformation $\phi(u) = 1/(1 + \exp(u))$ and the model

$$W_k = \phi \left[\sum_{j=1}^p f(\mathbf{a}'_j \mathbf{z}_k) \right] + e_k, \tag{A5}$$

where $E(e_k) = 0$ and $\text{var}(e_k) = E(W_k)(1 - E[W_k])$. The choice of the logistic is natural on statistical grounds and is also a convenient way to ensure that the predicted fitnesses lie between zero and one.

The estimates of the ridge functions must now reflect the presence of the transformation in equation (A5), and the sum of squares in equation (A3) must be modified to $\sum_{k=1}^n (W_k - \phi[\sum_{j=1}^p f\{a_j' z_k\}])^2 / \text{var}(W_k)$. In the backfitting algorithm the univariate estimates obtained by considering each direction separately are more complicated to compute. However, there is still an approximate $A_j(\lambda)$ matrix that links the adjusted fitness measurements to the predicted values from the model. The trace of this matrix remains a valid measure of the effective number of parameters.

SELECTING THE SMOOTHING PARAMETER

The preceding discussion has all been under the assumption that the smoothing parameter λ is fixed. We chose among different values for λ by an approximate form of cross validation. This criterion has the form

$$\text{GCV}(\lambda) = \frac{1}{n} \frac{\sum (W_k - \hat{W}_k)^2 / \text{var}(W_k)}{1 - \left(\sum_{j=1}^m v_j \right) / n} \quad (\text{A6})$$

This criterion estimates how well the estimated fitness surface can predict the observed measurements of survival or reproductive success W as a function of the smoothing parameter. Thus, in order to make the prediction error as small as possible, one would choose λ to minimize $\text{GCV}(\lambda)$. Some justification for this method can be found in chapter 9 of Hastie and Tibshirani (1990).

LITERATURE CITED

- Brodie, E. D., III. 1992. Correlational selection for color pattern and antipredator behaviour in the garter snake *Thamnophis ordinoides*. *Evolution* 46:1284–1298.
- Craven, P., and G. Wahba. 1979. Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation. *Numerische Mathematik* 31:377–403.
- Crespi, B. J., and F. L. Bookstein. 1989. A path-analytic model for the measurement of selection on morphology. *Evolution* 43:18–28.
- Efron, B., and R. Tibshirani. 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science* 1:54–77.
- Endler, J. A. 1986. *Natural selection in the wild*. Princeton University Press, Princeton, N.J.
- Friedman, J. A., and W. Stuetzle. 1981. Projection pursuit regression. *Journal of the American Statistical Association* 76:817–823.
- Hastie, T., and R. Tibshirani. 1990. *Generalized additive models*. Chapman & Hall, London.
- Jayne, B. C., and A. F. Bennett. 1990. Selection on locomotory performance capacity in a natural population of garter snakes. *Evolution* 44:1204–1229.
- Karn, M. L., and L. S. Penrose. 1951. Birth weight and gestation time in relation to maternal age, parity, and infant survival. *Annals of Eugenics* 16:147–164.
- Kirkpatrick, M., and R. Lande. 1989. The evolution of maternal characters. *Evolution* 43:485–503.
- Lande, R. 1979. Quantitative genetic analysis of multivariate evolution, applied to brain:body size allometry. *Evolution* 33:402–416.
- Lande, R., and S. J. Arnold. 1983. The measurement of selection on correlated characters. *Evolution* 37:1210–1226.
- Lindén, M., L. Gustafsson, and T. Pärt. 1992. Selection on fledgling mass in the collared flycatcher and the great tit. *Ecology* 73:336–343.
- McCullagh, P., and J. A. Nelder. 1983. *Generalized linear models*. Chapman & Hall, New York.
- Mitchell-Olds, T., and R. G. Shaw. 1987. Regression analysis of natural selection: statistical inference and biological interpretation. *Evolution* 41:1149–1161.

- Nelder, J. A., and R. Mead. 1965. A simplex minimization for function minimization. *Computer Journal* 7:308–313.
- O'Sullivan, F. B., S. Yandell, and W. J. Raynor. 1986. Automatic smoothing of regression functions in generalized linear models. *Journal of the American Statistical Association* 81:96–103.
- Pearson, K. 1903. Mathematical contributions to the theory of evolution. XI. On the influence of natural selection on the variability and correlation of organs. *Philosophical Transactions of the Royal Society of London A, Mathematical and Physical Sciences* 200:1–66.
- Phillips, P. C., and S. J. Arnold. 1989. Visualizing multivariate selection. *Evolution* 43:1209–1222.
- Price, T., M. Kirkpatrick, and S. J. Arnold. 1988. Directional selection and the evolution of breeding date in birds. *Science (Washington, D.C.)* 240:798–799.
- Provine, W. B. 1986. *Sewall Wright and evolutionary biology*. University of Chicago Press, Chicago.
- Schluter, D. 1988. Estimating the form of natural selection on a quantitative trait. *Evolution* 42:849–861.
- Schluter, D., and P. R. Grant. 1984. Determinants of morphological patterns in communities of Darwin's finches. *American Naturalist* 123:175–196.
- Schluter, D., and J. N. M. Smith. 1986. Natural selection on beak and body size in the song sparrow. *Evolution* 40:221–231.
- Schluter, D., T. D. Price, and P. R. Grant. 1985. Ecological character displacement in Darwin's finches. *Science (Washington, D.C.)* 227:1056–1059.
- Simms, E. L., and M. D. Rausher. 1993. Patterns of selection on parasite resistance in *Ipomoea purpurea*. *Evolution* 47:970–976.
- Simpson, G. G. 1953. *The major features of evolution*. Columbia University Press, New York.
- Smith, T. B. 1990. Natural selection on bill characters in the two morphs of the African finch *Pyrenestes ostrinus*. *Evolution* 44:832–842.
- Swain, D. P. 1992. Selective predation for vertebral number in *Gasterosteus aculeatus*: reversals in the direction of selection at different larval sizes. *Evolution* 46:998–1013.
- Terrenato, L., M. F. Gravina, A. San Martini, and L. Ullizi. 1981. Natural selection associated with birth weight. III. Changes over the last twenty years. *Annals of Human Genetics* 45:267–278.
- Ullizi, L., and C. Manzotti. 1988. Birth weight and natural selection: an example of selection relaxation in man. *Human Heredity* 38:129–135.
- Ullizi, L., and L. Terrenato. 1987. Natural selection associated with birth weight. V. The secular relaxation of the stabilizing component. *Annals of Human Genetics* 51:205–210.
- Wahba, G. 1990. *Spline models for observational data*. Society for Industrial and Applied Mathematics, Philadelphia.
- Wright, S. J. 1932. The roles of mutation, inbreeding, crossbreeding, and selection in evolution. *Proceedings of the VI International Congress of Genetics* 1:356–366.
- Zhivotovsky, L. A., and M. W. Feldman. 1992. On the difference between mean and optimum of quantitative characters under selection. *Evolution* 46:1574–1578.

Associate Editor: Mark Kirkpatrick