VOL. 172, NO. 1 THE AMERICAN NATURALIST JULY 2008

E-ARTICLE

# Unifying Life-History Analyses for Inference of Fitness and Population Growth

**Ruth G. Shaw,[1,*] Charles J. Geyer,[2,†] Stuart Wagenius,[3,‡] Helen H. Hangelbroek,[4,§] and Julie R. Etterson[5,‖]**

1. Department of Ecology, Evolution, and Behavior, Minnesota Center for Community Genetics, University of Minnesota, St. Paul, Minnesota 55108;
2. School of Statistics, University of Minnesota, Minneapolis, Minnesota 55455;
3. Institute for Plant Conservation Biology, Chicago Botanic Garden, Glencoe, Illinois 60022;
4. Department of Ecology, Evolution, and Behavior, University of Minnesota, St. Paul, Minnesota 55108;
5. Biology Department, University of Minnesota, Duluth, Minnesota 55812

*Submitted June 24, 2007; Accepted February 8, 2008; Electronically published May 23, 2008*

ABSTRACT: The lifetime fitnesses of individuals comprising a population determine its numerical dynamics, and genetic variation in fitness results in evolutionary change. This dual importance of individual fitness is well understood, but empirical fitness records generally violate the assumptions of standard statistical approaches. This problem has undermined comprehensive study of fitness and impeded empirical synthesis of the numerical and genetic dynamics of populations. Recently developed aster models remedy this problem by explicitly modeling the dependence of later-expressed components of fitness (e.g., fecundity) on those expressed earlier (e.g., survival to reproduce). Moreover, aster models employ different sampling distributions for different components of fitness (e.g., binomial for survival over a given interval and Poisson for fecundity). Analysis is done by maximum likelihood, and the resulting distributions for lifetime fitness closely approximate observed data. We illustrate the breadth of aster models' utility with three examples demonstrating estimation of the finite rate of increase, comparison of mean fitness among genotypic groups, and analysis of phenotypic selection. Aster models offer a unified approach to addressing the breadth of questions in evolution and ecology for which life-history data are gathered.

*Keywords: Chamaecrista fasciculata*, community genetics, demography, *Echinacea angustifolia*, fitness components, *Uroleucon rudbeckiae.*

The fitness of an individual is well understood as its contribution, in offspring, to its population. Fitness has both evolutionary significance, as an individual's contribution to a population's subsequent genetic composition, and ecological significance, as the numerical contribution to a population's growth. The simplicity of these closely linked ideas belies serious complications that arise in empirical studies. Lifetime fitness comprises multiple components of fitness expressed over possibly many intervals. As a result, the distribution of fitness, even for a synchronized cohort in the absence of systematic sources of variation, is typically multimodal, with a discrete mode at 0, and highly skewed, thus corresponding to no known parametric distribution. This problem has long been acknowledged (e.g., Mitchell-Olds and Shaw 1987; Stanton and Thiede 2005) and has severely undermined efforts to link ecological and evolutionary inference, yet no single, rigorously justified approach has been available for analysis of lifetime fitness.

A new statistical approach, aster modeling, serves this role. Aster models generate the overall likelihood for a set of components of fitness expressed through the lives of individuals. Within a single analysis, aster analysis models different fitness components with different statistical distributions, as appropriate, and accounts for the dependence of fitness components expressed later in the life span on those expressed earlier. The statistical theory for aster models is given by Geyer et al. (2007). Here, we review the limitations of previous approaches to analysis of life histories, describe aster models, and present three empirical examples to illustrate the utility of aster modeling as a comprehensive approach to analysis of life-history data.

---

[*] Corresponding author; e-mail: rshaw@superb.ecology.umn.edu.

[†] E-mail: charlie@stat.umn.edu.

[‡] E-mail: swagenius@chicagobotanic.org.

[§] E-mail: helen_hangelbroek@hotmail.com.

[‖] E-mail: jetterso@d.umn.edu.

Am. Nat. 2008. Vol. 172, pp. E35–E47.
0003-0147/2008/17201-42689$15.00.
DOI: 10.1086/588063

## The Problem and Previous Efforts to Address It

Individual fitness realized over a life span typically does not conform to any well-known distribution that is amenable to parametric statistical analysis. In contrast, individual components of fitness, such as survival to age *x*, reproduction at age *x*, and the number of young produced at age *x*, generally conform much more closely to simple parametric distributions. For this reason, components of fitness are sometimes analyzed separately to obviate the problem of the distribution of lifetime fitness. For example, in a study of genetic variation in response to conspecific density of a population of the perennial plant *Salvia lyrata*, Shaw (1986) reported separate analyses of survival and size of the survivors. Implicitly, this approach considers size, or in other cases, fecundity, conditional on survival. Though the statistical assumptions underlying the analyses tend to be satisfied, they offer no way to combine the analyses to yield inferences about overall fitness.

Analyses of overall fitness commonly use total fecundity as the measure of fitness, assigning fecundity of 0 to individuals that died before reproduction. When observations are available for replicate individuals, a variant of this method treats fitness as the product of the proportion surviving and the mean fecundity of survivors (e.g., Belaoussoff and Shore 1995; Galloway and Etterson 2007). In both cases, the resulting distribution is actually a mixture of underlying discrete and (quasi)continuous distributions, yet analyses have generally treated it as a single, continuous response despite its skewness, multimodality, and discrete mode at 0 such that no transformation yields a distribution suitable for parametric statistical analysis. Authors frequently remark on the awkwardness of these distributions (e.g., Etterson 2004) but rarely publish fitness distributions. Antonovics and Ellstrand (1984), however, presented the extremely skewed distribution of lifetime reproductive output (their fig. 2) from their experimental studies of frequency-dependent selection in the perennial grass *Anthoxanthum odoratum.* Finding no transformation that yielded a normal distribution suitable for ANOVA, they assessed the robustness of their inferences by applying three distinct analyses (categorical analysis of discrete fecundity classes, ANOVA of means, and nonparametric analysis). In their study, results of the three analyses were largely consistent, but, in general, results are likely to differ.

Others have noted the importance of complete accounting of life history in inferring fitness or population growth rate, as well as evaluation of its sampling variation, and have presented methods to accomplish this. Caswell (2001) and Morris and Doak (2002) explain how to obtain population projection matrices from life-history records and, from them, to estimate population growth rate. They also describe methods for evaluating sampling variation and acknowledge statistically problematic aspects of these methods. Specifically, Caswell (2001, p. 304) notes that the delta method and other series approximations assume both that variances of the elements of a population projection matrix are small and that the population growth rate is normally distributed. It is often further assumed that all the parameters are independent (Caswell 2001, p. 302). These assumptions are likely to be violated in many cases. Consequently, Caswell (2001) recommended resampling approaches, first applied in this context by Lenski and Service (1982).

Recent efforts to evaluate the nature of selection have likewise taken a comprehensive demographic approach. McGraw and Caswell (1996) considered individual life histories but chose the maximum eigenvalue of an individual's Leslie matrix ($\lambda$) as its fitness measure. They regressed $\lambda$ on the fitness components, age at reproduction and lifetime reproductive output, to estimate selection but noted violation of the assumption of normality of residuals required for statistical testing. Van Tienderen (2000) advocated an alternative approach involving evaluation of the relationships between each component of fitness and the phenotypic traits of interest via separate multiple regression analyses to obtain the selection gradients for different episodes of selection (Lande and Arnold 1983). These selection gradients are then weighted by the elasticities (Caswell 2001) of each component of fitness obtained from analysis of the appropriate population projection matrix. Using this method, Coulson et al. (2003) also noted violation of the usual distributional assumptions. Moreover, because the method combines results from multiple analyses, it does not fully account for sampling variation. Methods targeting the problem of zero-inflated data (i.e., many observations of 0 distorting a distribution) have also been proposed (Cheng et al. 2000; Dagne 2004; Martin et al. 2005), but these methods also do not generalize readily for inference in the wide range of contexts that life-history data can, in principle, address.

## Inference of Individual Fitness with Aster Models

We present aster models (Geyer et al. 2007) for analysis of life-history records as a general, statistically sound approach to diverse questions in evolution and ecology. Two properties of life-history data are central to the statistical challenges that aster models address. First, the expression of an individual's life history at one stage depends on its status at earlier stages. For example, observation of an individual's fecundity at one stage is contingent on its survival to that stage. Second, no single parametric distribution is generally suitable for modeling all components of fitness, for example, survival and fecundity. The aster approach jointly models the components of fitness using distributions suitable for each, explicitly taking into account the dependence of each stage on previous stages. We represent the life history and, in particular, the dependence of one life-history component on another, with graphical models, as in figure 1, with arrows pointing from a predecessor variable to its successor.

The theory underlying the aster approach requires modeling the conditional distribution of each variable, given its predecessor variable, as an exponential family of distributions (Barndorff-Nielsen 1978; Lehmann and Casella 1998; Geyer et al. 2007), with the predecessor variable providing the sample size for its successor. This requirement retains considerable flexibility because many well-known distributions are exponential families, including Bernoulli, Poisson, geometric, normal, and negative binomial (Mood et al. 1974, pp. 87–103).

A predecessor variable $n$ must be nonnegative integer valued. If $n > 0$, then the successor is the sum of $n$ independent and identically distributed variables having the named distribution. If $n = 0$, then the successor is 0. This accommodates much of the dependence in life-history data. In a graph like that in figure 1*A*, where each of the variables $S_x$ models survival (0 or 1, with 1 indicating alive), a dead individual stays dead and does not reproduce.

The aster approach is generally suited to analyzing complicated life histories (e.g., fig. 1). Some widely used methods are special cases of aster models. The simplest possible aster models have graphs with only one arrow per individual ($1 \to X$). If $X$ is normal, this is a linear model (LM), as in multiple regression or ANOVA. If $X$ is Bernoulli or Poisson, this is a generalized linear model (GLM), as in logistic or Poisson regression (McCullagh and Nelder 1989). The next simplest models have graphs $1 \to X \to Y$, with $X$ Bernoulli and $Y$ zero-truncated Poisson (like fig. 1*D*); here the marginal distribution of $Y$ is zero-inflated Poisson (Martin et al. 2005). An aster model with graph $1 \to X_1 \to X_2 \to \cdots \to X_n$, with all $X_i$ Bernoulli, corresponds to survival analysis. We note that, in all these cases, multiple parameterizations arise. The parameters that are directly interpretable, the mean value parameters, are different from those that are modeled linearly, the canonical parameters. In Bernoulli (logistic) regression, the mean value parameter is the proportion $p = E(X)$, whereas the canonical parameter is $\theta = \text{logit}(p) = \log(p) - \log(1 - p)$. In Poisson regression, the mean value parameter is

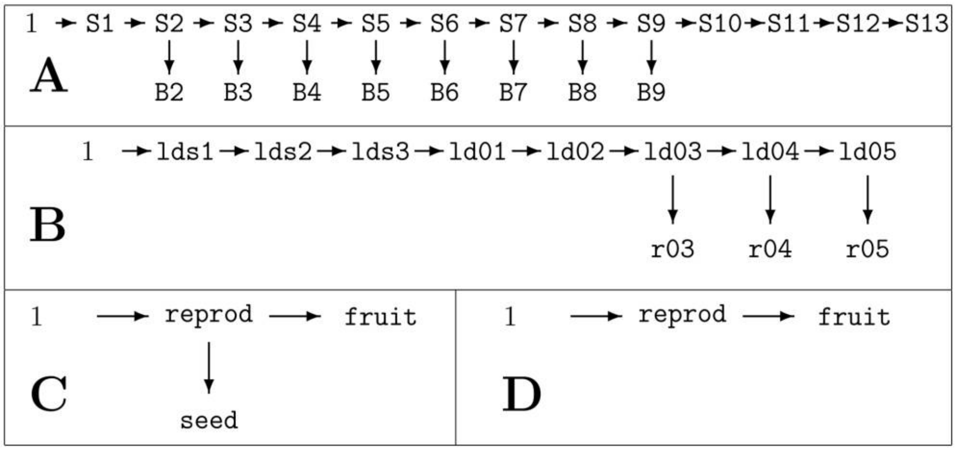

**Figure 1:** Graphical models for our three examples. Each node of a graph is associated with a variable; root nodes have the constant 1, indicating presence of an individual at the outset. Arrows lead from one life-history component to another that immediately depends on it (from predecessor node to successor node of the graph); each is associated with a particular conditional distribution of the successor, given the predecessor. If a predecessor variable is 0 for a given individual, for example, because of mortality, then its successor variables are also 0. *A*, Example 1: *Uroleucon rudbeckiae*, an aphid. An individual's fitness comprises its survival to each age, *Si*, modeled as (conditionally) Bernoulli and the number of young it produces at each age, *Bi*, modeled as (conditionally) zero-truncated Poisson. *B*, Example 2: *Echinacea angustifolia*, a perennial plant. Observations include juvenile survival at three times up to transplanting into the field (*ldsi*) and subsequent survival through 5 years (*ld0i*), as well as the plant's number of rosettes (*r0i*) in 3 years. The survival variables are modeled as (conditionally) Bernoulli (0 indicates mortality, and 1 indicates survival), and r0*i* is (conditionally) zero-truncated Poisson (i.e., a Poisson random variable conditioned on being greater than 0). *C*, *D*, Example 3: *Chamaecrista fasciculata*, an annual plant. Success or failure of reproduction (here, including survival to reproduction) is given by reprod, modeled as Bernoulli (0 indicates no seeds, and 1 indicates production of seeds). Given that a plant reproduces, the components of its fecundity are its number of fruits (*fruit*) and in *C*, also the number of seeds in a sample of three fruits (*seed*). Each is modeled as negative binomial, two truncated and zero truncated, respectively.

$\mu = E(X)$, whereas the canonical parameter is $\theta = \log(\mu)$.

In all aster models, a transition between life-history stages, that is, an arrow in the graph, corresponds to the conditional distribution of one fitness component, which contributes one term to the log likelihood

$$l(\boldsymbol{\theta}) = \sum_j (x_j\theta_j - x_{p(j)}c_j(\theta_j)), \tag{1}$$

where $x_j$ is the canonical statistic, $\theta_j$ is the canonical parameter for the $j$th conditional distribution, and $x_{p(j)}$ is the predecessor of $x_j$. Each term of equation (1) has exponential family form, but the sum does not. It can, however, be put in exponential family form,

$$l(\boldsymbol{\varphi}) = \left[\sum_j x_j\varphi_j\right] - c(\boldsymbol{\varphi}), \tag{2}$$

by a change of parameter. Either equation (1) or equation (2) is a log likelihood for the full model, with one parameter per variable, and the canonical statistic vector **x** is the same for both, but the linearly modeled canonical parameters, $\boldsymbol{\theta}$ and $\boldsymbol{\varphi}$, differ. To distinguish the two canonical parameter vectors, we call $\boldsymbol{\theta}$ conditional and $\boldsymbol{\varphi}$ unconditional.

Unconditional aster models are submodels of the full model (2) determined by the change of parameter $\boldsymbol{\varphi} = \mathbf{M}\boldsymbol{\beta}$. The submodel is also an exponential family with log likelihood

$$l(\boldsymbol{\beta}) = \left[\sum_k y_k\beta_k\right] - d(\boldsymbol{\beta}), \tag{3}$$

where $\mathbf{y} = \mathbf{M}^T\mathbf{x}$. The matrix **M** is called the model or design matrix. For this submodel, **y** is the canonical statistic, and $\boldsymbol{\beta}$ is the canonical parameter whose maximum likelihood estimate (MLE) solves the equations

$$y_k = E_{\boldsymbol{\beta}}(Y_k). \tag{4}$$

Both **y** and the MLE of $\boldsymbol{\beta}$ are minimal sufficient (contain all of the information in the data about the parameter). The expectation of the canonical statistic $E_{\boldsymbol{\beta}}(\mathbf{Y})$ is the mean value parameter. The relationship between the canonical parameter and the mean value parameter is monotone, $\partial E_{\boldsymbol{\beta}}(Y_k)/\partial\beta_k > 0$; increasing $\beta_k$ while holding the rest fixed increases the corresponding mean value parameter. Moreover, hypothesis tests and confidence intervals concerning the corresponding canonical parameters directly evaluate the statistical significance of these canonical statistics. Unconditional aster models share all of these properties with GLM.

In contrast, none of these properties is shared with conditional aster models; that is, when the conditional canonical parameter is modeled linearly, $\boldsymbol{\theta} = \mathbf{T}\boldsymbol{\gamma}$, where **T** is a model matrix. The resulting submodel itself is not an exponential family. The MLE is the value of $\boldsymbol{\gamma}$ that solves the equation

$$y_k = \sum_j E_{\boldsymbol{\gamma}}(X_j \mid x_{p(j)})t_{jk}, \tag{5}$$

where $t_{jk}$ is the component of **T** but has no simple properties. The MLE is not a sufficient statistic and has no monotone relationship with expectations.

Either kind of aster model (conditional or unconditional) is a model for the joint distribution of all the data. Either may be useful for some particular data, but we recommend unconditional models because they yield simple interpretations via unconditional mean values (of the canonical statistics), like those familiar from LM and GLM. In particular, they directly test hypotheses about lifetime fitness when it is a canonical statistic. The unconditional parameterization is not readily understood intuitively because terms in $\boldsymbol{\beta}$ that nominally refer to a single component of fitness (affecting its $\varphi_i$ only) directly influence the unconditional expectation of overall fitness by affecting not only the distribution of the specified component but also the distributions of earlier expressed components of fitness. Consequently, it is difficult (but not impossible; see our example 2) to see the role played by a single component of fitness. This is an unavoidable consequence of addressing overall fitness.

We demonstrate the value and versatility of the aster approach with three examples. In the first, we illustrate inference of population growth rate. We consider a small data set that Lenski and Service (1982) used to demonstrate their nonparametric method for inferring population growth rate from a set of individual life histories of the aphid *Uroleucon rudbeckiae*. In this case, we illustrate the use of a conditional model, though either type of model could be used. In our second example, we apply the aster approach to compare mean fitness among groups to quantify effects of inbreeding on *Echinacea angustifolia*, a long-lived plant, from observations over 5 years. Finally, we reanalyze data of Etterson (2004) to estimate the fitness surface for phenotypic traits for the annual legume *Chamaecrista fasciculata*. We show how much simpler aster analysis is when fitness is a canonical statistic and also how to proceed when, due to the experimental design, it is not. A contributed package "aster" for the R statistical language (R Development Core Team 2006) performs all calculations related to aster models, contains the data sets for our examples, and is freely available (http://www.r-project.org). Two technical reports (Shaw et al. 2007*a*, 2007*b*) provide reproducible computations of all the analyses for this article.

## Example 1: Estimation of Population Growth Rate λ

Lenski and Service (1982) recognized the need for a valid statistical approach for inferring rates of population growth λ from life-history records via the stable-age equation (Fisher 1930). They emphasized the importance of accounting for individual variation in survivorship and fecundity and of treating the full life-history record of an individual as the unit of observation. Lenski and Service (1982) presented a nonparametric approach that resamples complete records of individual life histories via the jackknife to obtain estimates and sampling variances of λ. They illustrated the approach with a small data set sampled from the aphid *Uroleucon rudbeckiae*. The survival and fecundity in each of 14 age intervals were recorded for 18 individuals (see fig. 1*A*), and these data served as the basis for estimating λ and its sampling variance.

Applying the aster approach to these data, we modeled the binomial parameter governing survival probability, logit($\sigma_x$), as a quadratic function of age $x$. Survivorship declined significantly with age ($P = .001$), with significant curvature ($P = .028$). Expected fecundities $\beta_x$, modeled according to a Poisson distribution, were estimated for each age class $x$, given survival to that age.

Interest focuses primarily on estimating λ but also on its sampling variance, as noted by Alvarez-Buylla and Slatkin (1994), because of its importance in assessing whether a population is growing or declining. The stable-age equation implicitly defines λ as a nonlinear function of the unconditional expectations $\mu_x = \sigma_x\beta_x$, which are estimated by aster models; from these, λ is determined by solving the stable-age equation, and standard errors are obtained using the delta method (Shaw et al. [2007*b*] give details). From these data, we estimated $\lambda = 1.677$, with a standard error of 0.056. Our estimate agrees closely with one of Lenski and Service (1982; 1.688), and 95% confidence intervals are also similar (aster: 1.57, 1.79; jackknife: 1.52, 1.85). We emphasize, however, that the aster approach can be used in more complicated situations where resampling methods would not be valid.

## Example 2: Comparison of Fitness among Groups

In this example, we illustrate use of aster models to compare mean fitnesses of groups. Specifically, we investigate how relatedness of parents affects progeny fitness in a perennial plant *Echinacea angustifolia* (narrow-leaved purple coneflower), a widespread species in the North American prairie and Great Plains. Following the conversion of land to agriculture that began about a century ago, the formerly extensive populations now persist in typically small patches of remnant prairie. The plant is self-incompatible, and Wagenius (2000) detected no deviation from random mating within a large population in western Minnesota. In the context of fragmented habitat, matings between close relatives in the same remnant, and perhaps also long-distance matings, may be more common.

To evaluate effects of different mating regimes on progeny fitness, formal crosses were made between pairs of plants (*a*) from different remnants, (*b*) chosen at random from the same remnant, and (*c*) known to share maternal parent. The parental plants had been growing for 3–4 years in randomized arrays in an experimental field. From the resulting seeds, 557 seedlings germinated. After 3 months in a growth chamber, the surviving 508 individuals were transplanted back into the same experimental field. Survival of each seedling was assessed in the growth chamber on three dates and, after transplanting into the field, annually from 2001 to 2005. The number of rosettes (basal leaf clusters, 1–7) per plant was also counted annually from 2003 to 2005. Here, we consider individual size as a surrogate component of fitness during the juvenile period; the typically strong positive relationship between size and eventual fecundity justifies this here, as elsewhere.

Mortality of many plants (~30%) resulted in a distribution of rosette count in 2005 having many zeros. We modeled survival through each of eight observation intervals as Bernoulli, conditional on surviving through the preceding stage; we modeled rosette count in each of three field seasons, given survival to that season, as zero-truncated Poisson (fig. 1*B*). To account for spatial and temporal heterogeneity, we also included in the models as fixed effects (*a*) year of crossing (1999 or 2000), (*b*) planting tray during the period in the growth chamber, and (*c*) spatial location (row and position within row) in the field.

Our primary focus was on evaluating the effects of mating treatments on overall progeny fitness, taken as expected rosette count in 2005 for a seed obtained in 2001. In addition, we investigated the timing and duration of the effects of mating treatment on fitness. These effects could be slight during early stages but strong during later stages; alternatively, effects of mating treatment at the early stages may largely account for differences in fitness. These scenarios differ in their implications concerning the inbreeding load expected in standing populations (Husband and Schemske 1996). To evaluate these scenarios, we developed four aster models, named "chamber," "field," "sub," and "super." Each was a joint aster analysis of all 11 fitness components (survival over eight intervals, rosette count at three times). The field model includes explicit mating treatment effects only on the final rosette count (fig. 1*B*, *r05*), but because we used unconditional aster models, these effects propagate back to earlier stages as well. The chamber model includes explicit mating treatment effects only on the final survival before transplanting (fig. 1*B*, *lds3*), but again, these effects propagate back through the two preceding bouts of survival. The remaining two models are used to test these scenarios of timing of effects; the sub model is the greatest common submodel of the chamber and field models, and the super model is their least common supermodel (i.e., the sub model includes no effects of mating treatment on any aspect of fitness, whereas the super model includes separate effects of mating treatment on survival up to transplanting and final rosette count).

The aster analysis revealed clear differences among mating treatments in mean overall fitness (field model vs. sub model, $P = 1.1 \times 10^{-5}$). The fitness disadvantage of progeny resulting from sib mating relative to the other treatments is a 35%–42% reduction in rosette count (fig. 2*B*). Because of the propagation of effects back to earlier stages, effects of mating treatment on variable r05 in the field model directly account for differences in fitness at all earlier stages. Thus, this analysis suffices for inferring the overall effects of mating treatment on fitness.

Our further investigation of the timing and duration of these effects detects differences among mating treatments in survival up to transplanting (sub model vs. chamber model, $P = .012$). Comparison of the chamber and field models with the super model shows that the field model accounts well for differences in overall fitness; the super model fits no better than the field model ($P = .34$) but does fit better than the chamber model ($P = 3.1 \times 10^{-4}$). The terms in the super model for the effects of mating treatment on survival up to transplanting are not needed to fit the data because the backpropagation of effects subsumes the effects of mating treatment in the growth chamber. This does not mean that there are no effects of mating treatment on fitness before transplanting. Comparison of the sub and chamber models confirms they exist, and figure 2 clearly shows them. The early disadvantage of progeny resulting from sib mating relative to the other treatments is clear in the 7%–10% reduced survival up to the time of transplanting, but the overall fitness disadvantage of inbreds is considerably greater (fig. 2).

### Example 3: Phenotypic Selection Analysis

Lande and Arnold (1983) proposed multiple linear and quadratic regression of fitness on a set of quantitative traits as a method for quantifying natural selection directly on

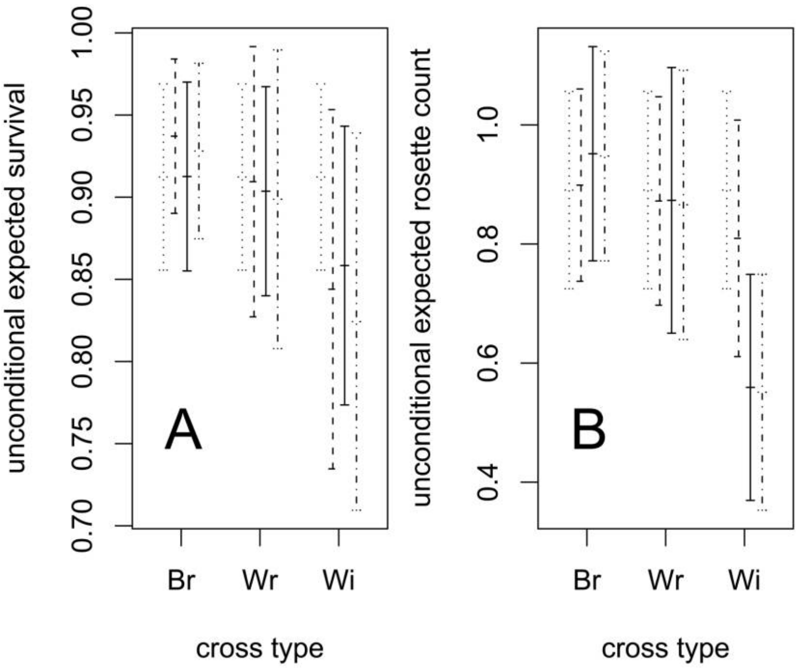

**Figure 2:** Predicted values and 95% confidence intervals for the unconditional mean value parameter for survival up to transplanting (*A*; *lds3*) and rosette count in the last year recorded (*B*; *r05*; i.e., overall fitness over the study period) for a typical individual for each cross type. The experimentally imposed crossing treatments are between remnant populations (*Br*), within remnant populations (*Wr*), and inbred within remnants (*Wi*; i.e., between sibs). Lines indicate the different models: *dotted* = sub model; *dashed* = chamber model; *solid* = field model; *dot-dashed* = super model.

each trait. In practice, these analyses have generally employed components of fitness, rather than overall fitness, as the response variable (e.g., Lande and Arnold 1983; Kingsolver et al. 2001). As a result, the estimated selection gradients (partial regression coefficients) reflect selection on a trait through a single episode of selection rather than selection over multiple episodes or over the life span, as needed for evolutionary prediction. Focusing on this limitation, Arnold and Wade (1984*b*) considered partitioning the overall selection gradient into parts attributable to distinct episodes of selection, and Arnold and Wade (1984*a*) illustrated the approach with examples. Wade and Kalisz (1989) modified this approach to allow for change in phenotypic variance among selection episodes. Whereas these developments were intended to accommodate the multiple stages of selection, they do not directly account for the dependence of later components of fitness on ones expressed earlier because they use multiple separate analyses.

Mitchell-Olds and Shaw (1987), among others, noted that statistical testing of selection gradients is often compromised by failure of the analysis to satisfy the assumption of normality of the fitness measure, given the predictors. This concern applies to McGraw and Caswell's (1996) approach, which integrates observations from the full life history. To address this problem for the case of dichotomous fitness outcomes, such as survival, Janzen and Stern (1998) recommended the use of logistic regression for testing selection on traits and showed how the resulting estimates could be transformed to obtain selection gradients. To allow for shapes of the fitness function more general than quadratic, Schluter (1988) and Schluter and Nychka (1994) suggested estimation of the relationship between fitness and traits as a cubic spline, but this also requires a parametric error distribution, whether normal, binomial, or Poisson.

Unconditional aster analysis estimates the relationship between overall fitness and traits directly in a single, unified analysis. We illustrate this use of aster models with a reanalysis of Etterson's (2004) study of phenotypic selection on three traits in three populations of the annual legume *Chamaecrista fasciculata* reciprocally transplanted into three sites. The three traits, measured in 8–9-week-old plants, are leaf number (LN; log transformed), leaf thickness (measured as specific leaf area, SLA, the ratio of a leaf's area to its dry weight; log transformed), and reproductive stage (RS; scored in six categories, with increasing values denoting greater reproductive advancement). Here, for simplicity, we consider a subset of the data for the three populations grown in the Minnesota site, comprising records of 2,235 individuals.

In this experiment, individuals were planted as seedlings, and fitness was assessed as (1) survival to flowering; (2) flowering, given survival; (3) number of fruits, given flowering; and (4) number of seeds in a sample of three fruits, also given flowering. For simplicity, we collapsed survival, flowering, and fruiting to a single component of fitness, modeled as Bernoulli ("reprod"). Plants that produced seeds have reprod $= 1$, and those that did not, regardless of the reason, have reprod $= 0$. Consequently, overall fitness was modeled jointly as reproduction, number of fruits, and number of seeds in three fruits (fig. 1*C*, *reprod*, *fruit*, *seed*). Preliminary analyses assessed the fit of truncated Poisson and truncated negative binomial distributions for both fecundity components, and the latter distribution was adopted for fruit and seed. In addition to the traits of interest, the model included spatial blocks as fixed effects.

To illustrate phenotypic selection analysis most straightforwardly, we begin by analyzing two of the fitness components, reprod and fruit (fig. 1*D*), in relation to the traits LN, SLA, and RS. We use an unconditional aster model, in which overall fitness is the number of fruits produced per individual seedling. This model detected strong dependence of fitness on all three traits such that selection is toward more leaves ($P < 10^{-6}$), thinner leaves ($P = .006$), and earlier reproductive stage ($P < 10^{-6}$).

We detected highly significant negative curvature for LN and SLA, suggestive of stabilizing selection ($P < 10^{-6}$); because RS is categorical, we did not consider models quadratic in it. The plot of the fitness function together with the observed phenotypes (fig. 3, *solid contours*) reveals that the fitness optimum lies very near the edge of the distribution of leaf number. Thus, for this trait, despite significant negative curvature, selection against both extremes of the standing variation in the trait (i.e., stabilizing selection) is not observed. The aster analysis fits the data well, as reflected by the scatterplots of Pearson residuals that show very little trend and only a few extreme outliers for fruit number (fig. 4*A*). The assumptions of the aster model appear satisfied, and the estimated fitness surface is biologically plausible and fits the data well. These points reinforce our confidence in the aster model *P* values and estimated fitness surface.

We compare the result from aster modeling with that obtained from the approach of Lande and Arnold (1983), which has become standard, ordinary least squares regression (OLS) of fruit count on traits. The bivariate fitness function inferred via OLS has positive curvature for LN, suggesting disruptive selection. This contrasts with the negative curvature obtained by aster models (fig. 3). The fitness surface fitted by aster models (fig. 3, *solid contours*) has a peak on the right side (large LN) and is fairly flat on the left (small LN). The quadratic approximation via OLS (fig. 3, *dotted contours*) cannot have flat regions; its best approximation is a saddle. Further, a quadratic function cannot have both a saddle and a peak; thus, OLS

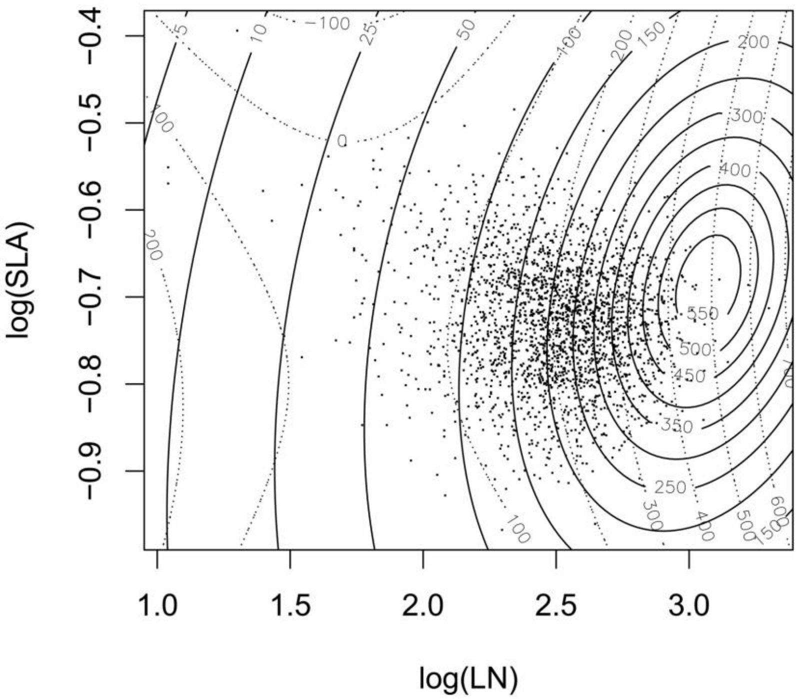

**Figure 3:** Scatterplot of specific leaf area (*SLA*; log transformed) versus leaf number (*LN*; log transformed), with contours of the fitness function (expected fruit count) estimated by aster models (*solid contours*) and those of the quadratic approximation estimated by ordinary least squares (*dotted contours*). Compare with figure A2.

misses the peak. Another problematic feature of the quadratic approximation is that it becomes negative. Thus, the main problem with OLS is the bias from the quadratic approximation of a highly nonquadratic surface. The aster model is also quadratic, but it is quadratic on the canonical parameter scale. The corresponding fitness estimates, which are mean value parameters, are necessarily positive. Transformation of LN and SLA via Box-Cox (Box and Cox 1964) to satisfy the assumption of normality (Lande and Arnold 1983) hardly affects this comparison of the shapes of the fitness functions inferred via the aster approach and OLS (fig. A1).

Nominal *P* values produced by OLS regression indicate that the positive curvature of the quadratic approximation in the LN direction is statistically significant ($P < 10^{-6}$), but the homoscedasticity and normality assumptions required for OLS regression to give meaningful *P* values are seriously violated (fig. 4*B*). Such violations of assumptions for an OLS regression analysis are expected, given that 3% of plants have fitness of 0 and that the distributions of numbers of fruits per plant is heavily skewed. These violations make the nominal *P* value from the OLS invalid.

We extend the above phenotypic selection analysis to include the additional fitness component seed (fig. 1*C*). In this case, fitness is no longer a canonical statistic; that is, there is no linear combination of the variables that corresponds to fitness. The two fecundity components fruit and seed are modeled as separately dependent on reprod. This analysis detected dependence of fruit on LN ($P < 10^{-6}$) and SLA ($P = .046$) and of seed on LN ($P < 10^{-6}$) and RS ($P < 10^{-6}$). It also found significant curvature in the relationship between fruit and both LN ($P < 10^{-6}$) and SLA ($P = .035$) and between seed and LN ($P = .0008$). Here again, we did not attempt to fit quadratic dependence on RS. Use of an unconditional aster model in these analyses yields an estimate of the relationship between each fitness component and each trait that takes into account prereproductive mortality.

The above analysis does not, however, satisfy the goal of evaluating the relationship between overall fitness and the traits; because fitness is not a canonical statistic, its expectation is not produced directly by aster methods. However, it can be approximated by averaging simulations. In detail, we use the estimated parameter values to simulate fitness records for individuals representing each trait combination, from these calculate fruit × seed/3, and average

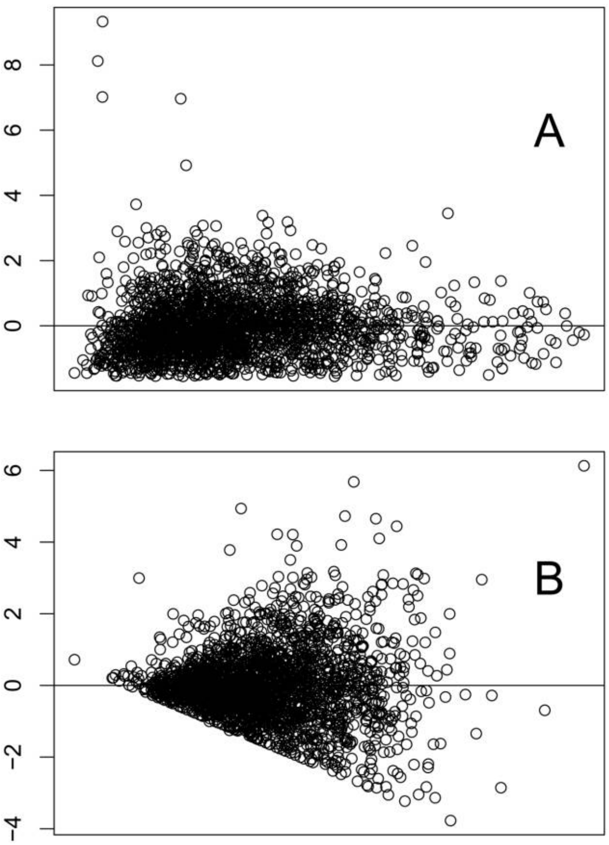

**Figure 4:** Residual plots from phenotypic selection analyses for *Chamaecrista fasciculata*. *A*, Pearson residuals for fruit count conditional on reproduction plotted against values fitted from the aster model quadratic in leaf number and specific leaf area and also containing reproductive stage and spatial block. *B*, Similar to *A*, except standardized residuals fitted by ordinary least squares (same response and predictors as in *A*).

these over the simulations to approximate the expected fitness for each trait combination. The shape of the resulting fitness surface (fig. A2, *solid contours*) resembles that estimated using only fruit as the fecundity measure, though it provides stronger evidence of true stabilizing selection on LN. The best quadratic approximation fitted by OLS has a saddle in this case too.

There are two alternative data structures that include seed counts for which a simple aster analysis, without simulation, would analyze fitness directly. The graph 1 → reprod → fruit → seed would be appropriate if all seeds (from all fruits) had been counted for each individual. Then, in an unconditional aster model, fitness would be seed, a canonical statistic, and the analysis would automatically take the contribution of reprod and fruit to fitness into account. However, it is often impractical, as in this case, to count all seeds. Subsampling is a common practice in studies of animals (e.g., Howard 1979) and plants. An alternative to exhaustive enumeration that facilitates aster analysis is to obtain for each individual the seed count for a random number of fruits, corresponding to the graph 1 → reprod → fruit → samp → seed, where samp is the number of fruits sampled for the individual, a binomial(fruit, $p$) random variable, where $p$ is fixed and known (the fraction of fruit sampled). In this sampling scheme, fitness would be proportional to seed, a canonical statistic, and aster analysis would again be simple.

For our actual data, it might seem natural to use the product of fruit count and number of seeds per fruit as one variable in an aster analysis, so the graph is 1 → reprod → fruit × seed/3, but this would not be valid because this product is not distributed according to an exponential family. Alternatively, it might seem natural to use the preceding graph with samp replaced by the constant 3, but this is also invalid because the constant 3 is not distributed according to an exponential family. Thus, the structure of this aster model precluded inference of overall selection via a simple aster analysis. Nevertheless, simulation yielded the expected fitness surface.

We have illustrated how aster models can conduct phenotypic selection analysis on complete life-history records to yield biologically interpretable estimates of the fitness surface. We emphasize that even for an annual, for which life-history analysis is commonly considered relatively straightforward, assumptions of aster analysis are more closely satisfied than are those of OLS. We have also shown that, even when available data preclude modeling total reproductive output as a canonical statistic, aster estimates the parameters of a fitness model that can be used (with simulation, if necessary) to produce a statistically sound phenotypic selection analysis.

## Discussion

Both numerical and genetic dynamics of a population depend fundamentally on individuals' contributions of offspring, their fitness. Extensive theoretical work (e.g., Fisher 1930; Charlesworth 1980) has formalized and extended this insight of Darwin, yet statistical challenges have continued to compromise empirical evaluation of fitness. Aster models address these challenges and take full advantage of available data to yield comprehensive assessments of fitness that are as precise as possible. This not only offers statistical power for tests of hypotheses but also promotes quantitative comparison of fitnesses. Most important, as a general framework for analyzing life-history data, aster models can address questions that arise in diverse evolutionary and ecological contexts. Examples presented here illustrate the breadth of aster models' applicability, including estimation of population growth rate, comparison of mean fitness among groups, and inference of phenotypic selection. Beyond analysis of life-history data, aster modeling is appropriate for any set of responses in which there are dependencies analogous to those characteristic of life histories. In a behavioral study, for example, individual subjects may forage in a given interval or not and, given that they forage, may take varying numbers of prey. We emphasize that aster models obviate the common practice of multiple separate analyses, which cannot provide valid statistical tests or sound estimates of sampling error. A single aster model can encompass the real complexities not only of multiple component responses but also of discrete and continuous predictors, thus yielding direct inferences about fitness and population growth.

Lifetime fitness rarely, if ever, conforms to any distribution amenable to parametric statistical analysis less complex than aster models, a problem that has plagued empirical studies of fitness. Resampling approaches are sometimes used, but this is not a general solution because valid resampling schemes generally are not available for complex data structures. Moreover, resampling methods sacrifice statistical precision relative to parametric analysis. As an alternative, transformations are often attempted, but the prevalence of mortality before reproduction typically results in fitness distributions with many individuals at 0, so no transformation produces a well-known distribution. Moreover, even if such a transformation could be found, analyses of fitness on an alternative scale can mislead (Stanton and Thiede 2005). Aster models address these problems by directly modeling each distinct component of fitness with a suitable parametric distribution and accounting for the dependence of each fitness component on those expressed earlier. As a consequence, it models the sampling variation appropriately and yields results on the biologically natural scale of expected number of individuals produced per individual. When records are available for only a portion of the life span, as in our example 2, joint analysis via an unconditional aster model provides comparisons based on the most complete records at hand.

Studies of variation in fitness often focus on single components of fitness likely to be conveniently distributed. These can yield insight into the nature of fitness variation during a particular episode of selection. However, the resulting understanding of fitness and its variation is fragmentary and can be misleading when the relationship between components of fitness, on the one hand, and traits or genotypes, on the other, varies over the life span (e.g., Prout 1971). Whereas Arnold and Wade (1984*b*) proposed an approach to evaluate phenotypic selection over multiple episodes (modified by Wade and Kalisz [1989]), this approach uses separate analyses of each episode, ignoring the dependence structure of fitness components. Consequently, the sampling variance of the resulting estimates of selection cannot readily be determined.

Our first example illustrates use of aster models to infer population growth rate. Lenski and Service (1982) first noted the importance of sound statistical modeling for population growth. Our use of aster models in this context builds on their work by employing parametric models for each life-history event. The resulting estimate of growth rate is similar to that obtained by Lenski and Service's (1982) method using the jackknife, as are the confidence intervals from the two approaches. The key point is that aster analysis provides a sound parametric basis for inferences about population growth, even for data structures that are not suited to resampling.

Our second example demonstrates the use of unconditional aster models to estimate and compare mean fitness for groups produced by different mating schemes and, thus, differing in genetic composition. This analysis reveals that remnant populations of *Echinacea angustifolia* are subject to severe inbreeding depression of at least 70% overall when extrapolated linearly to inbreeding arising from one generation of selfing. In a similar application of aster models, Geyer et al. (2007) have analyzed survival and annual production of flower heads jointly for samples of these remnant populations grown in the experimental field, demonstrating greater than twofold range in mean fitness among remnants ($P < .01$). In the example here, we show how the likelihood framework of aster models permits straightforward tests of several hypotheses. We show that the significant early disadvantage in size and survival of inbred plants does not adequately account for the fitness differences at the end of the period of observation. Rather, inbreeding depression in growth and survival exacerbates the fitness disparity among seedlings. Thus, in addition to its statistically rigorous comparisons of overall fitness among groups, aster models yield insights into the timing of fitness effects.

Aster models readily extend to accomplish phenotypic selection analysis, estimating the relationship between individuals' overall demographic-genetic contribution to the next generation and their traits. For this, the aster model includes the traits under consideration as predictors of cumulative fitness; inference of quadratic and correlational selection is also straightforward. Our example 3 shows aster models' estimation of the fitness surface when fitness is a linear function of the components of fitness and also demonstrates how to obtain such an estimate even when it is not. Van Tienderen (2000) presented a method with a similar goal, but it does not take into account the dependence relationships of the fitness components and is subject to the usual distributional problems (e.g., Coulson et al. 2003). Further, it cannot validly represent the statistical uncertainty of inferred parameters because it involves separate analyses to estimate selection gradients for each fitness component. In contrast, a single aster analysis using an unconditional model provides results that are interpretable as comparisons of overall fitness. Aster models report asymptotic *P* values and confidence intervals, and the aster software also easily applies the parametric bootstrap, which does not require the assumption of asymptotic normality of MLE.

The centrality of fitness to many evolutionary and ecological questions demands a statistical approach that rigorously models the inevitable, complex dependencies of life-history data. Our examples provide only a glimpse of the range of possible uses of aster models. Conceivably, all the issues in all our examples and more could arise in a single analysis, as could more extensive dependence. The aster approach addresses these challenges. Its versatility suits it to answer the full breadth of questions that life-history data can address. Aster models can play a key role linking ecological and evolutionary study of populations.

## Acknowledgments

R.G.S. and C.J.G. cordially thank J. Antonovics for his encouragement when we began work on the basic idea around 1980, for funding its development then, and for his enthusiasm about its eventual realization. Computational challenges stymied our initial efforts, and other work intervened until the richness of life-history data from recent experiments stimulated us to revisit the idea. For very helpful suggestions for clarifying the manuscript, we thank J. Antonovics, K. Mercer, M. Price, J. Travis, N. Waser, and an anonymous reviewer. Examples 2 and 3 come from research funded by the National Science Foundation (DMS-0083468, DEB-0545072, DEB-0544970) and an Environmental Protetection Agency Science to Achieve Results graduate student fellowship (U 914758-01-2), respectively, and the University of Minnesota Center for Community Genetics.

## APPENDIX

### Supplementary Analyses for Example 3

We carried out further analyses for example 3 and present here the resulting estimates of fitness surfaces. In one case, we transformed the phenotypic variables (specific leaf area and leaf number) via Box-Cox transformation (rather than log transformation) to satisfy the assumption of normality (fig. A1, drawn from Shaw et al. 2007*c*). In the second case, fitness comprised the number of fruits times the average number of seeds for three fruits rather than number of fruits only (see text; fig. A2). Both figures can be compared with figure 3.

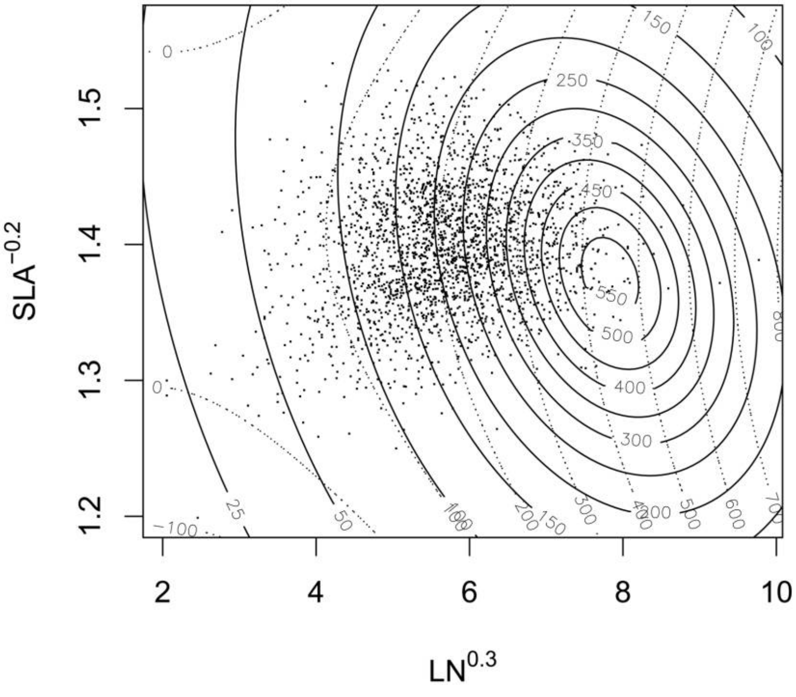

**Figure A1:** Scatterplot of specific leaf area (SLA) versus leaf number (LN; both Box-Cox transformed), with contours of the fitness function estimated by aster models (*solid contours*) and those of the quadratic approximation estimated by ordinary least squares (*dotted contours*). Fitness is number of fruits. Compare with figure 3.

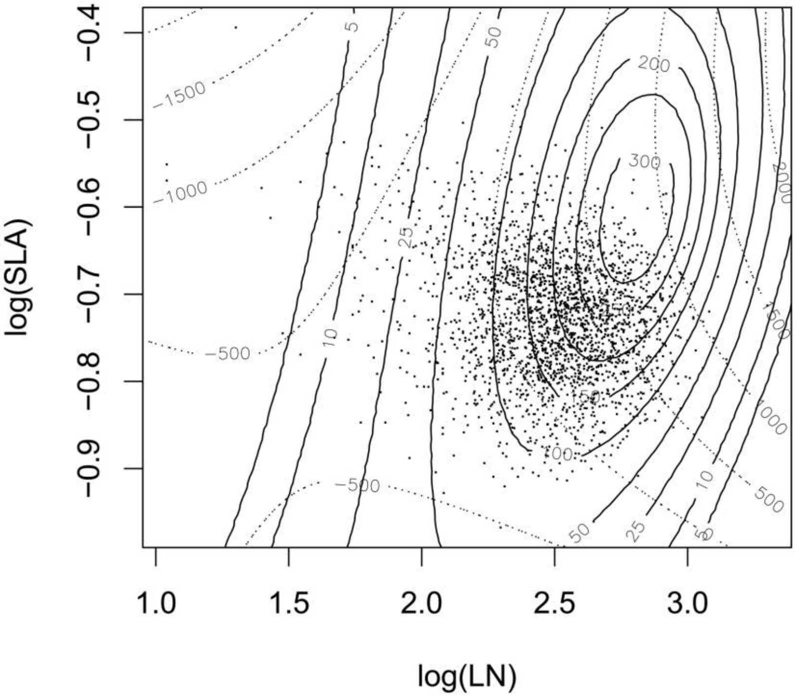

**Figure A2:** Scatterplot of specific leaf area (SLA) versus leaf number (LN; both log transformed), with contours of the fitness function estimated by aster models (*solid contours*) and those of the quadratic approximation estimated by ordinary least squares (*dotted contours*). Fitness is number of fruits times average number of seeds for three fruits. Compare with figure 3.

## Literature Cited

Alvarez-Buylla, E. R., and M. Slatkin. 1994. Finding confidence limits on population growth rates: three real examples revised. Ecology 75:255–260.

Antonovics, J., and N. C. Ellstrand. 1984. Experimental studies of the evolutionary significance of sexual reproduction. I. A test of the frequency-dependent selection hypothesis. Evolution 38:103–115.

Arnold, S. J., and M. J. Wade. 1984*a*. On the measurement of natural and sexual selection: applications. Evolution 38:720–734.

———. 1984*b*. On the measurement of natural and sexual selection: theory. Evolution 38:709–719.

Barndorff-Nielsen, O. E. 1978. Information and exponential families. Wiley, Chichester.

Belaoussoff, S., and J. S. Shore. 1995. Floral correlates and fitness consequences of mating-system variation in *Turnera ulmifolia*. Evolution 49:545–556.

Box, G. E. P., and D. R. Cox. 1964. An analysis of transformations (with discussion). Journal of the Royal Statistical Society B 26: 211–252.

Caswell, H. 2001. Matrix population models: construction, analysis, and interpretation. 2nd ed. Sinauer, Sunderland, MA.

Charlesworth, B. 1980. Evolution in age-structured populations. Cambridge University Press, Cambridge.

Cheng, S., D. Wang, and E. P. Smith. 2000. Adjusting for mortality effects in chronic toxicity testing: mixture model approach. Environmental Toxicology and Chemistry 19:204–209.

Coulson, T., L. E. B. Kruuk, F. Tevacchia, J. M. Pemberton, and T. H. Clutton-Brock. 2003. Estimating selection on neonatal traits in red deer using elasticity path analysis. Evolution 57:2879–2892.

Dagne, G. A. 2004. Hierarchical Bayesian analysis of correlated zero-inflated count data. Biometrical Journal 46:653–663.

Etterson, R. R. 2004. Evolutionary potential of *Chamaecrista fasciculata* in relation to climate change. I. Clinal patterns of selection along an environmental gradient in the great plains. Evolution 58: 1446–1458.

Fisher, R. A. 1930. The genetical theory of natural selection. Clarendon, Oxford.

Galloway, L. F., and J. R. Etterson. 2007. Inbreeding depression in an autotetraploid herb: a three cohort field study. New Phytologist 173:383–392.

Geyer, C. J., S. Wagenius, and R. G. Shaw. 2007. Aster models for life history analysis. Biometrika 94:415–426.

Howard, R. D. 1979. Estimating reproductive success in natural populations. American Naturalist 114:221–231.

Husband, B. C., and D. W. Schemske. 1996. Evolution of the magnitude and timing of inbreeding depression in plants. Evolution 50:54–70.

Janzen, F. J., and H. S. Stern. 1998. Logistic regression for empirical studies of multivariate selection. Evolution 52:1564–1571.

Kingsolver, J. G., H. E. Hoekstra, J. M. Hoekstra, D. Berrigan, S. N.

Vignieri, C. E. Hill, A. Hoang, P. Gilbert, and P. Beerli. 2001. The strength of phenotypic selection in natural populations. American Naturalist 157:245–261.

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

Lehmann, E. L., and G. Casella. 1998. Theory of point estimation. 2nd ed. Springer, New York.

Lenski, R. E., and P. M. Service. 1982. The statistical analysis of population growth rates calculated from schedules of survivorship and fecundity. Ecology 63:655–662.

Martin, T. G., B. A. Wintle, J. R. Rhodes, P. M. Kuhnert, S. A. Field, S. J. Low-Choy, A. J. Tyre, and H. P. Possingham. 2005. Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. Ecology Letters 8:1235–1246.

McCullagh, P., and J. A. Nelder. 1989. Generalized linear models. 2nd ed. Chapman & Hall, London.

McGraw, J. B., and H. Caswell. 1996. Estimation of individual fitness from life-history data. American Naturalist 147:47–64.

Mitchell-Olds, T., and R. G. Shaw. 1987. Regression analysis of natural selection: statistical inference and biological interpretation. Evolution 41:1149–1161.

Mood, A. M., F. A. Graybill, and D. C. Boes. 1974. Introduction to the theory of statistics. 3rd ed. McGraw-Hill, New York.

Morris, W. F., and D. F. Doak. 2002. Quantitative conservation biology: theory and practice of population viability analysis. Sinauer, Sunderland, MA.

Prout, T. 1971. The relation between fitness components and population prediction in *Drosophila*. I. The estimation of fitness components. Genetics 68:127–149.

R Development Core Team. 2006. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. http://www.r-project.org.

Schluter, D. 1988. Estimating the form of natural selection on a quantitative trait. Evolution 42:849–861.

Schluter, D., and D. Nychka. 1994. Exploring fitness surfaces. American Naturalist 143:597–616.

Shaw, R. G. 1986. Response to density in a wild population of the perennial herb *Salvia lyrata*: variation among families. Evolution 40:492–505.

Shaw, R. G., C. J. Geyer, S. Wagenius, H. H. Hangelbroek, and J. R. Etterson. 2007*a*. More supporting data analysis for "Unifying life history analysis for inference of fitness and population growth." Technical report 661. School of Statistics, University of Minnesota, Minneapolis. http://www.stat.umn.edu/geyer/aster.

———. 2007*b*. Supporting data analysis for "Unifying life history analysis for inference of fitness and population growth." Technical report 658. School of Statistics, University of Minnesota, Minneapolis. http://www.stat.umn.edu/geyer/aster.

———. 2007*c*. Yet more supporting data analysis for "Unifying life history analysis for inference of fitness and population growth." Technical report 666. School of Statistics, University of Minnesota, Minneapolis. http://www.stat.umn.edu/geyer/aster.

Stanton, M. L., and D. A. Thiede. 2005. Statistical convenience vs. biological insight: consequences of data transformation for the analysis of fitness variation in heterogeneous environments. New Phytologist 166:319–337.

van Tienderen, P. H. 2000. Elasticities and the link between demographic and evolutionary dynamics. Ecology 81:666–679.

Wade, M. J., and S. Kalisz. 1989. The additive partitioning of selection gradients. Evolution 43:1567–1569.

Wagenius, S. 2000. Performance of a prairie mating system in fragmented habitat: self-incompatibility and limited pollen dispersal in *Echinacea angustifolia*. PhD diss. University of Minnesota, Minneapolis.

Associate Editor: Peter D. Taylor
Editor: Donald L. DeAngelis