

RESPONSE SURFACE EXPERIMENTAL DESIGNS FOR INVESTIGATING INTERSPECIFIC COMPETITION

BRIAN D. INOUE¹

Department of Zoology, Ecology, Evolution, and Organismal Biology Group and Institute of Statistics and Decision Sciences, Duke University, Durham, North Carolina 27708 USA

Abstract. Competition is an important interaction in ecology, and many experiments have been done to examine the effects of intraspecific and interspecific competition. Unfortunately, most of these experiments have been done using either substitution or additive experimental designs. Substitution designs, and to a lesser extent additive designs, severely limit the useful inferences that ecologists can draw from the resulting data. Response surface experimental designs, which vary the densities of two competing species independently, offer the advantage of being able to compare the fits of alternative competition models and to estimate model parameters. This kind of experimental design has been relatively little used in ecology to date. I consider the use of response surface designs from two perspectives. The first is the optimal allocation of experimental effort among density treatments, with the goal of distinguishing among alternative models. Second, I compare the ability of six response surface designs to estimate known competition coefficients correctly, using simulated data. Low densities and densities near equilibrium in one generation tend to return similar densities the following generation and thus provide little or no information for comparing alternative population models. Experimental designs that span a wide range of densities, including high densities above carrying capacities, provide more accurate parameter estimates in general, but low densities are helpful for estimating population growth rates. Designs that are not fully factorial are one compromise solution that can yield accurate parameter estimates for models and yet require less experimental effort than full factorial surfaces. In general, response surfaces are a powerful class of experimental design and offer potentially stronger connections between empirical and theoretical approaches than traditional experimental designs.

Key words: *competition; experimental design; model selection; parameter estimation; response surface.*

INTRODUCTION

Interspecific competition for a limiting resource is thought to be one of the major factors limiting the distribution and abundance of species. In theory, the intensity of competition among species can influence their evolution, population dynamics (e.g., May 1974, Hassell and Comins 1976, May and Oster 1976), and coexistence (e.g., Hardin 1960, Atkinson and Shorrocks 1981, Tilman 1994, Inouye 1999a). Because of the important role of competition in ecology, hundreds of experiments have investigated intraspecific and interspecific competition (reviews in Schoener 1983, Ridsdill-Smith 1991, Gurevitch et al. 1992). Many of these experiments have used one of two basic experimental designs, substitution (replacement) series or additive designs; however, both of these experimental designs have been sharply criticized for their limitations and misinterpretations (e.g., Connolly 1986, Cousens 1991, Gibson et al. 1999). These limitations make this

large body of data less useful than it could have been, had more powerful experimental designs been used instead. The use of substitution and additive experimental designs has largely precluded generating quantitative estimates of the effects of interspecific competition on population dynamics or coexistence, beyond the inference that species do or do not compete. In this paper I discuss the benefits and uses of response surface designs, a third type of design for competition experiments that does not have the same limitations as substitution or additive experimental designs.

I concentrate on the properties of response surface experimental designs because these designs are well suited for fitting explicit competition models to data. Fitting explicit mathematical models of competition is an important objective for two reasons. First, the use of explicit models allows quantitative prediction of population dynamics, as opposed to qualitative measures of the presence or absence of competition (Peters 1991). For example, in order to rank threats to a native species that are posed by introduced competitors, we need to quantify the effects of interspecific competition on the population dynamics and persistence of the focal species. Results from experiments that use a substitution or additive design cannot be used to fit predictive

Manuscript received 24 January 2000; revised 7 August 2000; accepted 27 September 2000; final version received 6 November 2000.

¹ Present address: Center for Population Biology, Storer Hall, University of California, One Shields Avenue, Davis, California 95616 USA. E-mail: bdinouye@ucdavis.edu

models and limit our inference to the particular densities used in the experiment, even though natural populations commonly change in size from year to year. Second, fitting explicit models allows one to measure competition in the field using the same units as theoretical models, thus allowing a more direct connection between theoretical and empirical approaches and more rigorous tests of theory (Damgaard 1998) than traditional null-hypothesis tests. Despite the profusion of empirical competition studies, relatively few have calculated the interspecific competition coefficients that are the focus of most theoretical studies of competition (but see Park 1948, Vandemeer 1969, Ayala et al. 1973, Pfister 1995, Damgaard 1998, Inouye 1999b, Freckleton et al. 2000). Examples of competition experiments that used some form of response surface experimental design but did not estimate standard competition coefficients are not uncommon (e.g., Wilbur 1972, Watkinson 1981, Farmer et al. 1988, Connolly et al. 1990, Bullock et al. 1994, 1995, Juliano 1998, McPeck 1998). Indeed, few studies have fit any competition models to data, other than the unrealistically simple models implicit in some statistical tests (i.e., a global linear response to density).

Many of the advantages of response surface experimental designs for investigating competition have been mentioned before (e.g., Goldberg and Scheiner 1993, Gibson et al. 1999), but few field studies have actually used these experimental designs, and a thorough review of the ecological application and analysis of these designs is still lacking. One deterrent to the use of response surface experimental designs is that a large set of potential experimental designs are grouped in this category, with few guidelines for empiricists on how to choose among them. In this paper, I review substitution, additive, and response surfaces experimental designs, and then focus on response surface experimental designs from two different perspectives. First, given that experimental effort is always limited, it is useful to ask which treatments are most efficient for gaining information. I address the optimal allocation of experimental effort by asking whether some densities of competitors are more informative for distinguishing among competing models than others. Second, I use simulations to compare the ability of six different response surface designs with different allocations of experimental effort to estimate correctly the parameters of simple competition models under a variety of conditions.

REVIEW OF EXPERIMENTAL DESIGNS: USES AND OBJECTIVES

The first step in choosing an appropriate experimental design is to clearly state the objectives of the investigation. I consider three possible objectives for evaluating the utility of different experimental designs. The first is to quantify the effects of a competitor (Y) on the performance of a focal species (X), without

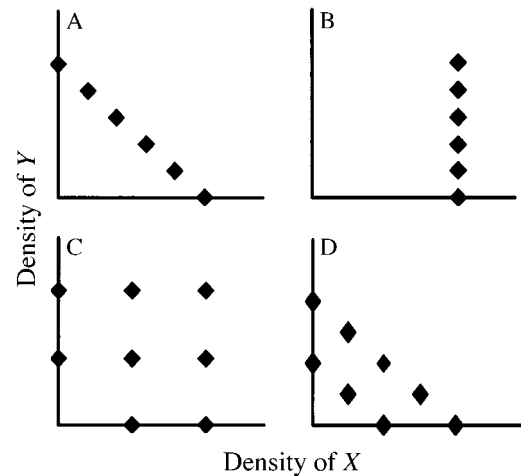


FIG. 1. Examples of the three main classes of experimental designs used for competition experiments. Each point represents a different treatment, or combination of densities, of species X and Y . Density is the number of individuals per unit area. Designs C and D are two alternative response surface designs.

regard for the reciprocal effects of X on Y . For example, in an agricultural setting, one might want to know the competitive effects of a weed on the performance of a crop, but not care about the effects of competition on the weed. A second objective might be to distinguish among models that describe the effects of both intraspecific and interspecific competition, including models that incorporate density- or frequency-dependent effects. Comparing the fit of alternative models can help to decide among competing hypotheses about competitive mechanisms, and selecting a best-fit model (or testing the fit of a particular model) is important for creating a predictive framework or verifying modeling assumptions. The third possible objective is to estimate parameter values for a particular model, once an appropriate functional form for the model has been chosen. Certainly there are other possible objectives for investigations of interspecific competition, but I have chosen to concentrate on the objectives of fitting and parameterizing models describing the effects of both intraspecific and interspecific competition. Data analysis by model selection and fitting allows close connections between empirical and theoretical work, aids ecologists in making quantitative predictions about the consequences of species interactions in natural and managed communities, and is an increasingly common tool (Hilborn and Mangel 1997, Burnham and Anderson 1998, Ellison 1999).

Examples of the three categories of experimental design discussed in this paper are shown in Fig. 1. For simplicity, I will describe the experimental designs and their analyses in the context of two competing species, denoted X and Y ; however, the designs and conclusions are equally valid for investigations of competition between genotypes of a single species, and can be ex-

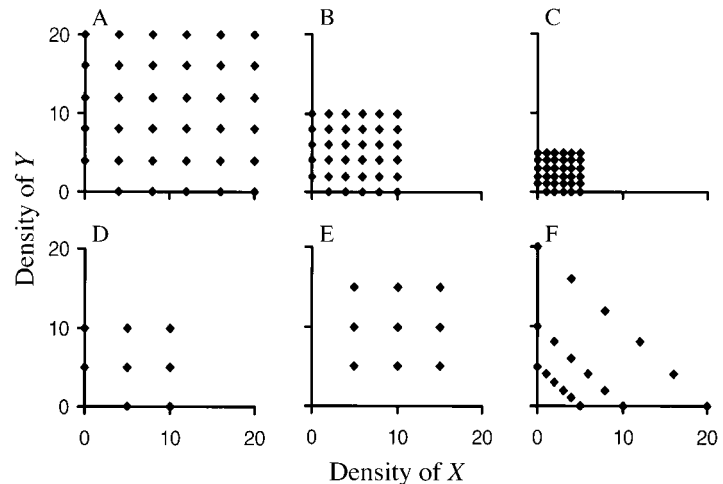


FIG. 2. Six alternative response surface designs. Density is the number of individuals per unit area.

tended to studies of more than two species or genotypes. For a substitution design (Fig. 1A), the total density of individuals in each treatment is usually held constant, while the proportion of each species is varied. In the most common form of this design, only three treatments are used: ratios of 1:0, 1:1, and 0:1 of species X and Y. Substitution designs are also commonly called "replacement series" (Connolly 1986, Snaydon and Satorre 1989, Silvertown and Doust 1993) or "de Wit series" in studies of plant competition (after de Wit 1960). In an additive design (Fig. 1B) the density of one species is held constant, and the density of the putative competitor is varied. The critical feature of these two types of designs is that all of the treatments fall along a single line through the two-dimensional plot of densities for X and Y: a diagonal line for substitution designs and a vertical (or horizontal) line for additive designs.

Substitution designs confound the effects of variation in intraspecific and interspecific density, and are poorly suited for any of the objectives outlined above. Inferences from substitution designs are limited also by the fact that the results of an experiment will depend critically upon the (arbitrary) single-species densities that are chosen as the end points of the mixtures (Inouye and Schaffer 1981, Firbank and Watkinson 1985, Connolly 1986, 1997, Snaydon 1991, 1994; but see Sackville Hamilton 1994). By controlling the density of a competitor, additive designs can detect a significant effect of interspecific competition on the performance of a focal species, thus addressing the first objective described above. Drawbacks of additive designs include the limitations that the effects of intraspecific competition cannot be estimated for comparison to interspecific effects (because the density of the focal species is held constant, Fig. 1B), and that inferences about the effect of interspecific density may depend on the density of the focal species that was chosen. Furthermore, in an additive design frequency dependent competitive effects are confounded with interspecific com-

petition. The scope of this last problem is not known, since very few studies have even attempted to look for frequency-dependent effects of competition (but see DeBenedictis 1977, Law and Watkinson 1987, Antonovics and Kareiva 1988).

Response surface experimental designs (Figs. 1C and D, 2) vary the densities of both species independently. Most often this is done by using treatments with factorial combinations of the two species at two or more densities, but any design where treatments do not all fall on a single line is a response surface design even if treatments are not factorial. Several authors have used "factorial designs" or "complete additive designs" to refer to response surface designs in general (Fig. 2A-E; e.g., Cousens 1991, Snaydon 1991, Silvertown and Doust 1993), neglecting other kinds of response surfaces that are not factorial. Response surface design F (Fig. 2) has also been called a "replacement series design" (Silvertown and Doust 1993) and an "addition series" (Cousens 1991), unfortunately confusing earlier terminology.

Response surface experimental designs are potentially suitable for a wide range of experimental objectives. Their worst shortcoming seems to be that they are overkill for some purposes (Cousens 1991). By using a range of densities for each species, response surface designs can describe intraspecific and interspecific competition without limiting inference to any particular densities. This is important for studies in natural systems, where the densities of the competitors may change over time, and is necessary for studies aiming to predict population dynamics or coexistence. Response surface designs are the only one of the three classes of experimental design considered here that can distinguish among different models and provide parameter estimates for these models. Two-species competition models describe a surface, but substitution and additive designs only provide data along a one-dimensional line under this two-dimensional surface.

Substitution, additive, and response surface designs

are not the only possibilities for investigating competition. Antonovics and Fowler (1985) used a "hexagonal fan" experimental design, in which plants of two species were placed at increasing distances from each other in two overlapping hexagonal arrays. This generates a space-efficient array where individual plants have from one to six neighbors of the competing species and grow at a range of densities, so that this experimental design could also be used for fitting and parameterizing models. Unfortunately, because each plant serves as both a focal individual and a neighbor, the treatments in this design are not independent, which means that the results of standard statistical analyses must be interpreted with caution. Another limitation is that experimental designs such as this, which create a range of competition treatments by using overlapping arrays (e.g., Veevers and Boffey 1979), can only be used for sessile organisms such as plants, and for a relatively limited range of densities.

Gibson et al. (1999) reviewed 10 yr (1984–1993) of greenhouse studies on plant competition in 11 ecological journals and found that substitution designs (replacement series) were the most common design used, representing 35% of the total 107 studies and 20% (3 of 15) of the studies in *Ecology*. Additive designs were a close second in frequency of use. I reviewed issues of *Ecology* from 1994 to 1999, and found 43 experimental studies of competition among plants, animals, or microorganisms. Only six studies (14%, mostly on plant competition) used substitution designs, so perhaps the repeated criticisms of this experimental design are beginning to take hold. The majority of the studies (23) used additive designs, increasing or decreasing the density of interspecific competitors and recording the response of a focal species. Four studies used experimental designs that were more difficult to classify, such as multigenerational chemostat experiments for competing algae (Huisman et al. 1999). The remaining ten studies used some form of response surface, usually a factorial design with two or three densities of each competitor, but only one analyzed the data from a response surface by fitting an explicit model (Juliano 1998). The remainder of the studies using response surface designs, and almost all of the studies using additive designs, used ANOVA to look for the simple presence of effects of competition, rather than estimating a strength of competition. Thus, although response surface experimental designs are perhaps more frequently used now than two decades ago, they are not being used to their full potential.

RESPONSE SURFACE DESIGNS FOR MODEL SELECTION AND PARAMETER ESTIMATION

Picking a response surface design to evaluate alternative models

In this section, I discuss general guidelines for allocating experimental effort among regions of a sur-

TABLE 1. Five competition models from the literature, where $X_{t+1} = X_t f(X_t, Y_t)$.

Model	$f(X_t, Y_t)$	Single species equilibrium
1	$1 + \lambda (1 - (X_t + \beta Y_t)/K)$	K
2	$\lambda \times \exp\{-c(X_t + \beta Y_t)\}$	$\ln(\lambda)/c$
3	$\lambda / \{1 + c(X_t + \beta Y_t)\}$	$(\lambda - 1)/c$
4	$\lambda / \{1 + (X_t + \beta Y_t)^b\}$	$(\lambda - 1)^{1/b}$
5	$\lambda / \{1 + c(X_t + \beta Y_t)\}^b$	$(\lambda^{1/b} - 1)/c$

Notes: The interspecific competition coefficient is β , which represents the effect of species Y on species X , relative to intraspecific competition. The maximum population growth rate from a low density in all models is λ . The parameters b and c are empirical constants used to set the carrying capacities; these parameters affect the responses to competition but do not have clear biological meanings. Model 5 was used with $b = 2$, in order to distinguish it from model 3.

face, so as to distinguish among competing models and select a functional form that best fits the data. The details of how to select an appropriate model for the data are beyond the scope of this paper, but can be found in many books (e.g., Dobson 1990, Edwards 1992, Christensen 1997, Hilborn and Mangel 1997, Burnham and Anderson 1998). Regardless of which models do fit the data well, a poor choice of experimental design or allocation of effort will make it difficult, if not impossible, to distinguish among the possibilities.

The resources and time available will limit the size of any experiment, and thus it is important to consider how effort should be expended efficiently to contrast a set of alternative models or to estimate model parameters (in the following section). In some cases the maximum number of organisms available at one time might limit the size of the experiment, so that there is a choice between using few replicates with a higher mean density, or more replicates with lower mean densities. In other cases the number of different treatments (and/or replicates) is limiting, because of the difficulty of constructing cages or plots where the competitors' densities can be manipulated.

Table 1 lists five different competition models that are commonly used in ecological investigations. All five models consist of a pair of recursive equations that describe the relationship between the densities of two species in consecutive generations (X_t and X_{t+1}). For simplicity I have only shown the equation for a focal species, X , competing with a species Y , and have not shown the analogous equation for species Y . These models share certain parameters to describe population dynamics and the effects of competition: the population growth rate, λ , a parameter that sets the carrying capacity, K or c , and the per capita strength of interspecific competition, β . The competition coefficient β typically measures the strength of interspecific competition relative to intraspecific competition, which is implicitly given a strength of one. Thus if β equals zero there is no interspecific competition, whereas β equal

to five means that each individual of species Y has an effect equivalent to the intraspecific competition from an additional five individuals of species X on the performance of species X (e.g., recruitment, survivorship, or yield).

These models can be broadly grouped into two families based on their basic functional forms. Models 1 (May and Oster 1976) and 2 (Ricker 1954, May 1974) predict a linear effect of density on the population growth rate, whereas models 3 (Law and Watkinson 1987), 4 (Leslie 1958), and 5 (Hassell and Comins 1976) predict nonlinear effects of density, proportional to an inverse power of the previous generations' densities. Model 5 contains one more parameter than the other models, and thus has slightly more flexibility in describing the effects of intraspecific and interspecific densities. Models 3 and 4 can be viewed as simplified versions of model 5. Models 1 and 2 are often used as discrete versions of the continuous Lotka-Volterra competition equations. Model 5 (Hassell and Comins 1976) and reduced versions of this model have been found to provide a good fit to data from several competition experiments (e.g., Bullock et al. 1994, 1995, Inouye 1999b) and have been widely used in theoretical studies of competition. Other functional forms for competition models are, of course, possible.

A plot of the density of the focal species in generation $t + 1$ (X_{t+1}) as a function of the densities of the focal species and the competitor in generation t (X_t and Y_t), describes a surface. Because the different functional forms of these equations imply different mechanisms for the way individuals gather and compete for resources, each of the models describes a different surface even when given the same parameter values (although the exact interpretation of the parameters may differ among models). For example, according to the discrete logistic and Ricker models (models 1 and 2), populations of the focal species that are above the carrying capacity in one generation will fall below the carrying capacity in the following generation, whereas the Hassell and Comins model (model 5) and its variations all predict a monotonic return to the carrying capacity. This difference in model behavior above carrying capacity will be discussed again in the section on parameter estimation.

Distinguishing among different models that describe the effects of intraspecific and interspecific density is equivalent to distinguishing among the shapes of the surfaces described by the models. Where the heights of two surfaces are identical, the models that describe those surfaces cannot be differentiated. Conversely, data from densities of X_t and Y_t , where models make different predictions are informative. Thus, in order to show where experimental effort is more informative, I calculated the geometric mean of the absolute differences among all pairs of model responses for densities of X_t and Y_t , using the five models in Table 1. The geometric mean is more appropriate than the arithmetic

TABLE 2. The experimental effort required for each of the three response surface experimental designs.

Design	Individuals	Cages	Replicates needed for similar individual effort	Replicates needed for similar cage effort
A	360	35	1	1
B	180	35	2	1
C	90	35	4	1
D	45	8	8	4
E	105	18	3	2

Note: The individual effort is the number of individuals of each species required for the complete design, and the cage effort is the number of experimental units (i.e., cages, plots, or ponds) required for the complete design.

mean for this purpose because it gives greater weight to small differences between models. Any (X , Y) treatment for which two different models predict the same response will yield a geometric mean difference of zero, regardless of the magnitudes of the differences with respect to other models. To calculate the geometric means, surfaces for the five models from Table 1 were generated using identical parameters for each model. Model 1 has the unrealistic feature that it predicts negative densities of the focal species in response to high densities of either the focal species or competitor. Therefore, when the predicted response for model 1 was negative, it was set equal to zero. For model 5 the parameter b was set equal to 2 in order to distinguish it from model 3.

Using response surface designs to estimate model parameters

To compare the efficiency of the six experimental designs from Fig. 2, I simulated data sets based on each design and then compared the parameter estimates from these simulations to the known parameters used to generate the data. I generated the simulated data using competition models 2–5 in Table 2 and all six experimental designs from Fig. 2, avoiding model 1 because of its unrealistic feature of returning negative population densities. For the most part, these simulations assume that the proper form of the competition model for estimating the parameters was known, which separates the model selection process from the parameter estimation. In practice, model selection and parameter estimation are integrated.

The simulations were repeated using either similar "individual effort" or equal "cage effort" (Table 2), and with two levels of added error. For equal individual effort, the number of replicates of each treatment in a design was adjusted so that the total number of individual competitors required for each design was equal (360 individuals of each species). To create equal cage effort the number of replicates of each design was adjusted so that the total number of treatments \times replicates was equal. For studies of plant competition, these

replicates may of course be plots or pots instead of cages. Balancing the amount of experimental effort required for each design by changing the degree of replication ensures that differences in their ability to estimate parameters are due to the allocation of treatments to different densities, and not to unequal resources for different designs. There will always be a tendency for larger experiments with more replication to yield more information, given similar experimental designs.

Data sets were simulated by first calculating the expected densities of a species X in generation 1, given the densities of X and a competitor Y in generation 0. The expected density for each replicate was then multiplied by a log-normally distributed random variable with a mean of 1 and a standard deviation of σ . Log-normally distributed error is commonly observed in real data sets (Krebs 1999), and ensures that all of the densities are positive numbers. For each simulated data set, I found the maximum-likelihood parameter estimates for the appropriate model (Edwards 1992), using log-transformed data and assuming normally distributed error (Splus 4.0, Mathsoft, Seattle, Washington, USA). The baseline simulations used $\lambda = 2$, $\beta = 1.5$, a single-species carrying capacity of 10, and $\sigma = 0.1$. To investigate the sensitivity of the results to the choice of model parameters, additional simulations were conducted with the Ricker competition model (model 2), using additional values for λ (1.5, 2, 3, 4), β (1, 1.5, 3), and σ (0.05, 0.1, 0.3). I chose to concentrate on the Ricker model because it has been widely used in theoretical and empirical investigations and its properties are relatively well understood (Hastings 1997). To investigate the robustness of parameter estimates to the choice of competition model used to fit the data when the true model is unknown, in some cases the data sets were generated using one competition model and parameters were estimated using likelihoods for a different model.

RESULTS

Using response surface designs for evaluating alternative models

Fig. 3 shows contours for the geometric mean difference among all pairs of the five models in Table 1, with $\lambda = 2.5$, carrying capacity = 10, and $\beta = 1.5$. There are two cases when all models have similar predictions, and thus data cannot distinguish among them. When the focal species is at a very low density, all of the models predict that the focal species will remain at a low density, and do not have divergent predictions regardless of the density of interspecific competitors. The second case occurs when the population growth rate of the focal species is zero, i.e., at an equilibrium density. This set of densities is described by the line $Y_i = (\text{carrying capacity} - X_i)/\beta$ (the joint equilibrium for both species is a point, not a line). Regions that

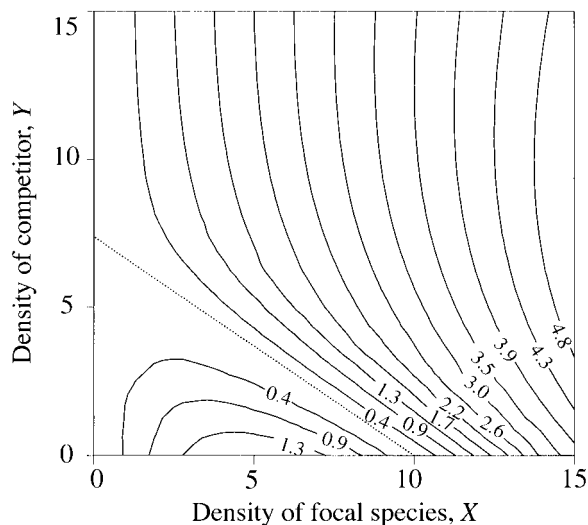


FIG. 3. Geometric mean differences among model predictions (X_{t+1}) for combinations of densities of the two competitors (X_t and Y_t), using the five different models in Table 2. For all models the growth rate $\lambda = 2.5$, the single species carrying capacity = 10, and $\beta = 1.5$. The dotted line shows equilibrium densities for species X . Small values indicate densities for which model surfaces have similar values, and thus data could not easily distinguish among models. Large values indicate densities where model surfaces differ most. Density is the number of individuals per unit area.

have high densities of both the focal species and the competitor relative to the carrying capacity, or intermediate densities of the focal species and low densities of the competitor relative to carrying capacity, are more informative for distinguishing among models (Fig. 3). Of course, to know whether a density is high or low relative to the carrying capacity it is necessary to have treatments with a range of densities, so the exact information value of a single treatment depends on the other treatments in the experimental design.

The population growth rate, λ , has no effect on the shape of the contours in Fig. 3, although higher growth rates do lead to larger absolute differences among model predictions. Changes in the competition coefficient, β , or the carrying capacity affect the location of the line describing the equilibrium densities. This alters the exact location of the contours, but the overall shape of the surface in Fig. 3 is unchanged (results not shown), and can be recovered by rescaling the axes relative to the carrying capacity or β . The range of different competition models that are commonly used in ecological studies tend to have functional forms similar to the ones used to generate Fig. 3, and thus the general shape of the contours in Fig. 3 are relatively robust to the exact choice of competition models used in the analysis (results not shown).

If the carrying capacities and competition coefficients for the competitors were known a priori, it would be possible to design an experiment to distinguish among models that allocated effort optimally; however,

TABLE 3. Parameter estimates from simulations of the six designs (A–F) shown in Fig. 2.

Noise	Parameter	True value	A	B	C
$\sigma = 0.1$	$\ln(\lambda)$	0.693	0.693 ± 0.052	0.693 ± 0.051	0.692 ± 0.052
$\sigma = 0.1$	$\ln(c)$	0.069	0.069 ± 0.003	0.069 ± 0.005	0.070 ± 0.007
$\sigma = 0.1$	β	1.5	1.504 ± 0.081	1.526 ± 0.172	1.572 ± 0.367
$\sigma = 0.3$	$\ln(\lambda)$	0.693	0.696 ± 0.155	0.681 ± 0.147	0.692 ± 0.139
$\sigma = 0.3$	$\ln(c)$	0.069	0.069 ± 0.010	0.068 ± 0.019	0.068 ± 0.036
$\sigma = 0.3$	β	1.5	1.528 ± 0.269	1.741 ± 1.318	2.698 ± 3.316
$\sigma = 0.1$	$\ln(\lambda)$	0.693	0.692 ± 0.051	0.693 ± 0.036	0.692 ± 0.026
$\sigma = 0.1$	$\ln(c)$	0.069	0.069 ± 0.003	0.069 ± 0.005	0.069 ± 0.007
$\sigma = 0.1$	β	3.0	3.015 ± 0.146	3.008 ± 0.206	3.037 ± 0.304
$\sigma = 0.1$	$\ln(\lambda)$	1.396	1.386 ± 0.051	1.388 ± 0.036	1.388 ± 0.026
$\sigma = 0.1$	$\ln(c)$	0.139	0.139 ± 0.003	0.139 ± 0.005	0.139 ± 0.006
$\sigma = 0.1$	β	1.5	1.500 ± 0.038	1.503 ± 0.058	1.502 ± 0.078

Notes: Results are shown as the mean \pm 1 SD of 500 maximum likelihood estimates. The distributions of parameter estimates are approximately normal, except for designs B, C, and D with $\sigma = 0.3$. These simulations were run using designs with similar organism effort, and used the Ricker model to generate the data and fit the maximum-likelihood estimates.

in ecological research this information is typically unknown. Fig. 3 suggests that it is important to include a relatively broad range of densities of the focal species. Cousens (1991) states that the choice of densities for treatments in a response surface design can be arbitrary, but suggests that the densities of each species be spaced according to a geometric series rather than evenly spaced. According to Fig. 3, an evenly spaced distribution of treatments may be preferable. Using a geometric series places the highest density of treatments at low densities of each species, where data are less informative. An evenly spaced distribution of points ensures that more treatments will have a high density of each species, and increases the chance that some treatments will be above the (unknown) carrying capacity of the focal species. Picking some treatments with densities of competitors that are at the upper end of naturally observed densities will both maximize the expected differences among alternative models of competition, which aids in model selection analyses, and improve the odds of detecting any significant effects of competition. On the other hand, for accurate parameter estimates of population growth rates it is helpful to use some treatments with densities below the carrying capacities (cf. designs C and F, Table 3, Fig. 4).

Clearly, for purposes of distinguishing among models of competition, using treatments with a broad range of densities is desirable, but this may not be feasible in cases where experimental effort is limited by the number of organisms available to the researcher rather than the number of experimental units (i.e., cages or plots). Given limited experimental effort, using a broad range of densities may also conflict with a desire for increased replication of each density treatment. Note that, in order to fit a model to the data from a response surface experimental design, it is not strictly necessary to replicate treatments. On the other hand, in order to estimate some common measures of lack of fit, or to analyze results with an ANOVA approach instead of a

likelihood or regression based approach, some replication is required.

Comparison of parameter estimates for six response surface designs

The mean and standard deviation of 500 maximum likelihood estimates of the competition coefficient β from the Ricker model, given the baseline parameters and using similar cage effort, are shown in Fig. 4. Note that the error bars in Fig. 4 show the standard deviation of the 500 maximum-likelihood estimates, as opposed to a measure of the confidence interval for each individual maximum-likelihood estimate. The results for estimates of the model parameters using different initial parameter values, and using equal organism effort, are given in Table 3. When the simulations contained relatively little added noise ($\sigma = 0.1$, lines 1–3, 6–12), parameter estimates from all six of the response surface designs were on average very accurate. While the mean parameter estimates were close to the true values used for the simulations, the estimates based on some experimental designs were more variable than the estimates from others. With higher levels of noise, parameter estimates were increasingly variable and less accurate, especially for designs B, C, D, and E (Table 3, lines 4–6). Increasing β led to increases in the standard deviations of the parameter estimates, whereas increases in λ led to decreases in the standard deviations of the parameter estimates (Table 3).

For all of the competition models, estimates of the competition coefficient (β) from experimental designs that included treatments with higher densities and covered a larger area had lower variance (results for the Ricker model shown in Fig. 4, Table 3). Thus estimates based on simulations of designs A and F, which cover higher densities, always had the smallest variance, though which of these two designs was best depended on the competition model being used (results not shown). For the Ricker model, with larger amounts of

TABLE 3. Extended.

D	E	F
0.697 ± 0.071	0.688 ± 0.061	0.695 ± 0.039
0.069 ± 0.006	0.069 ± 0.004	0.069 ± 0.003
1.525 ± 0.208	1.514 ± 0.105	1.506 ± 0.093
0.679 ± 0.201	0.706 ± 0.190	0.690 ± 0.110
0.068 ± 0.024	0.070 ± 0.013	0.069 ± 0.010
1.909 ± 1.668	1.545 ± 0.368	1.5204 ± 0.264
0.694 ± 0.050	0.693 ± 0.061	0.691 ± 0.030
0.069 ± 0.006	0.069 ± 0.004	0.069 ± 0.003
3.017 ± 0.262	3.014 ± 0.195	3.010 ± 0.124
1.390 ± 0.049	1.387 ± 0.061	1.385 ± 0.031
0.139 ± 0.006	0.139 ± 0.004	0.139 ± 0.003
1.499 ± 0.068	1.503 ± 0.053	1.497 ± 0.035

added noise, the mean estimate from design C, which uses the lowest densities among the experimental designs in Fig. 2, was off by a factor of almost two and the estimates were highly variable (Table 3, lines 4–6). In fact, experimental design C consistently provided the worst parameter estimates for all of the competition models used to generate and fit the data, emphasizing the utility of at least a few high-density treatments. For the Ricker model, this same pattern was observed when the simulations were run using designs with similar cage effort, but designs B, C, D, and F performed slightly less well than shown in Table 3 since they had fewer replicates compared to the simulations with similar individual effort (results not shown).

When estimating the population growth rate, λ , all six of the experimental designs provided accurate estimates, regardless of the competition model being used. Unlike the estimates of the competition coefficient, in this case estimates from design C generally had the smallest variance while those for experimental designs A, D, and E, which have fewer replicates at lower densities, returned more variable estimates. For the parameter c , used to set the carrying capacity, the ranking of parameter estimates was similar to the results for the competition coefficient estimates. All six experimental designs provided estimates that were relatively accurate, and designs that included higher densities (A and F) had smaller error bars, while parameter estimates based on design C had the largest variance and bias.

These results are robust to the choice of competition model used to simulate and analyze the data. The same qualitative ranking of the six experimental designs' performance was observed when data were simulated and analyzed using any of the competition models, with designs A and F providing the best parameter estimates. When data were simulated using models 3–5, but fit using the Ricker model, the qualitative rankings of the designs were still similar despite using different models to generate and fit the data, with designs A and F providing the best parameter estimates and design C the worst. The estimates of the competition coefficient, β ,

were quite accurate, but the estimates of the population growth rate and carrying capacity were less so. The rankings for designs B, D, and E were more variable, although in absolute terms there were no large differences among their performance. When data were simulated using the Ricker model, models 3–5 could not provide meaningful parameter estimates. This is because at densities above carrying capacity in one generation, the Ricker model returns densities below carrying capacity the following generation. Models 3–5 cannot duplicate this feature of the population dynamics, and thus fits very poorly the data from a Ricker model.

DISCUSSION

Establishing a strong connection between theoretical and empirical approaches requires considering experimental design, to ensure that experimental effort is not wasted collecting data that cannot test theoretical predictions. In order to fit explicit competition models, which can help shed light on the mechanisms of competition and estimate interspecific competition coefficients, it is necessary to use a response surface design. The results of this study provide some general guidelines for using response surface experimental designs more efficiently.

The results in the first section suggest that for distinguishing among competing models describing the effects of intraspecific and interspecific competition, some combinations of densities of the two competitors are more informative than other combinations. In particular, commonly used models differ most in their predictions for densities of a focal species above carrying capacity, and at intermediate densities of a competitor. Treatments with a focal species at low densities are still useful for estimating model parameters but are not as informative for contrasting different models. If pilot data are available to suggest the magnitude of the carrying capacity of the focal species and effects of the

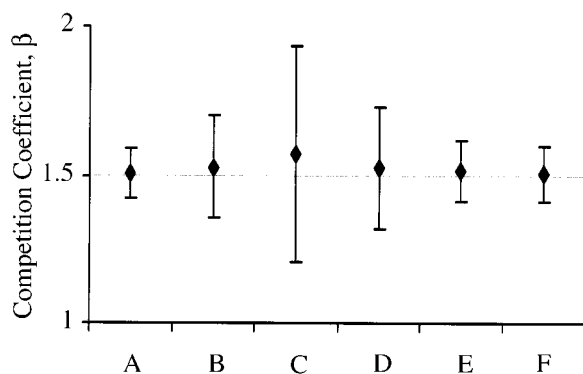


FIG. 4. Parameter estimates ± 1 SD of the maximum-likelihood estimates for the six experimental designs (A–F) in Fig. 2, using similar cage effort. The true value for $\beta = 1.5$, $\sigma = 0.1$. The distributions of parameter estimates are approximately normal.

competitor, the method outlined in the first section can be used to guide more efficient allocation of experimental effort. In the absence of data from pilot studies, these results support the recommendation to include at least a few densities near the maximum densities observed in the field, assuming this is logistically feasible. While densities above those observed in the field may be even more informative for distinguishing among alternative models, this only aids inferences if the range of models being considered actually includes at least one model that accurately describes competitive interactions at both high and low densities.

The results described in the second section also demonstrate the importance of allocating some experimental effort to treatments where the focal species is at a relatively high density, since these designs returned the most accurate estimates of the competition coefficients. Simulations such as those presented here could also be used to conduct power analyses, either for post hoc interpretation of results, or preferably to guide choices among alternative experimental designs before the actual experiment is done. When data were simulated with a larger amount of added noise, only analyses of data sets generated using response surface designs A and F (Fig. 2) were able to accurately recapture the original parameter values (Fig. 4), especially when estimating competition coefficients and carrying capacity. Design A is a factorial design, whereas design F is a set of substitution designs with different total densities. What both of these designs have in common is that the focal species is at a density relatively far above its carrying capacity in some of the treatments. To estimate the population growth rate it is useful to have treatments at low densities, where the population has the largest growth response. Overall, response surface design F appears to be a good choice for several reasons: it uses some high densities and thus returns accurate estimates of β and c , but also uses more low density treatments and thus is good at estimating λ , too. Furthermore, design F requires fewer individuals than a fully factorial design using higher densities, because it omits treatments where both species are at high densities.

While designs with a wide range of densities are better for both model selection and parameter estimation, a factorial experimental design using only two combinations of densities, one high and one low, may be a poor choice compared to an experimental design that also includes an intermediate density treatment. Competition models often describe nonlinear surfaces, and in order to capture these nonlinearities it is necessary to use experimental designs with more than two densities. Given limited experimental effort, adding more treatments to an experimental design will often require decreasing replication of each treatment. This tradeoff is not always favorable from the perspective of the current standard analysis, an ANOVA based investigation of the presence or absence of an effect of

competition, but can aid analyses that use a GLM or likelihood-based framework to fit models and estimate the magnitude of competitive effects.

The competition models discussed in this paper relate densities across consecutive generations, and have been applied successfully to both animal and plant competition (e.g., Bullock et al. 1994, 1995, Inouye 1999b). Experiments that cover multiple generations can take advantage of alternative experimental designs and analyses, which substitute many observations of a single set of populations (time series) for single observations of many populations (a replicated short-term experiment). Short-term plant competition experiments, or others that last only a fraction of a generation and relate a starting density to a final yield, can be analyzed with a range of competition models that use functional forms not considered here (e.g., Connolly et al. 1990, Connolly 1997). However, fitting these alternative models to data still requires a response surface approach. These models were not covered here because translating from short-term effects on yield or growth to long-term predictions of population dynamics can be problematic. Switching units from initial densities (counts) to final size (weight or length) can also cause problems interpreting the results of competition experiments, regardless of the experimental design being used, by introducing a size-dependent bias in the results (Connolly 1996, 1997, Gibson et al. 1999).

Conclusions

Despite the major limitations of interpreting the results of experiments that use substitution designs, such experiments continue to appear in the primary literature, and have been used in popular ecology textbooks for illustrating the effects of competition (e.g., Begon et al. 1990). The repeated criticisms of substitution designs over the past two decades (Inouye and Schaffer 1981, Firbank and Watkinson 1985, Connolly 1986, Snaydon 1991, Connolly and Wayne 1996, Gibson et al. 1999) are hopefully beginning to deter their use. Additive designs are still appropriate for a range of important basic and applied questions, but, because they cannot be used for fitting explicit models, they are less useful for linking theoretical and empirical approaches than experiments using response surface designs. Data from a response surface experimental design are well suited for evaluating the fits of alternative models and are also able to provide estimates and confidence intervals for the model parameters.

Competition experiments that use response surface experimental designs seem to remain rare for two main reasons. First, fitting models to the data from such experiments is slightly more complicated than the analysis of additive designs. However, current statistical packages now make it relatively simple to estimate model parameters from data, using either generalized linear models or maximum-likelihood methods, and to compare the fit of multiple models. Second, response

surface experiments are perceived as requiring a prohibitive amount of effort (e.g., Cousens 1991, Goldberg and Scheiner 1993). The results presented here demonstrate that even response surface experiments with a relatively low number of replicates or combinations of densities are able to provide good parameter estimates, especially if pilot data can be used to guide subsequent experiments.

ACKNOWLEDGMENTS

I thank M. Cain, P. de Valpine, A. Ellison, D. Marsh, W. Morris, N. Underwood, and an anonymous reviewer for helpful comments and suggestions that greatly improved this paper, and R. Wolpert for valuable advice.

LITERATURE CITED

- Antonovics, J., and N. L. Fowler. 1985. Analysis of frequency and density effects on growth in mixtures of *Salvia splendens* and *Linum grandifolium* using hexagonal fan designs. *Journal of Ecology* **73**:219–234.
- Antonovics, J., and P. Kareiva. 1988. Frequency-dependent selection and competition: empirical approaches. *Philosophical Transactions of the Royal Society of London B* **319**:601–613.
- Atkinson, W. D., and B. Shorrocks. 1981. Competition on a divided and ephemeral resource: a simulation model. *Journal of Animal Ecology* **50**:461–471.
- Ayala, F. J., M. E. Gilpin, and J. G. Ehrenfeld. 1973. Competition between species: theoretical models and experimental tests. *Theoretical Population Biology* **4**:331–356.
- Begon, M., J. L. Harper, and C. R. Townsend. 1990. *Ecology: individuals, populations and communities*. Blackwell Scientific, Boston, Massachusetts, USA.
- Bullock, J. M., A. M. Mortimer, and M. Begon. 1994. The effect of clipping on interclonal competition in the grass *Holcus lanatus*—a response surface analysis. *Journal of Ecology* **82**:259–270.
- Bullock, J. M., A. M. Mortimer, and M. Begon. 1995. Carryover effects on interclonal competition in the grass *Holcus lanatus*: a response surface analysis. *Oikos* **72**:411–421.
- Burnham, K. P., and D. R. Anderson. 1998. *Model selection and inference: a practical information-theoretic approach*. Springer-Verlag, New York, New York, USA.
- Christensen, R. 1997. *Log-linear models and logistic regression*. Springer-Verlag, New York, New York, USA.
- Connolly, J. 1986. On difficulties with replacement-series methodology in mixture experiments. *Journal of Applied Ecology* **23**:125–137.
- Connolly, J. 1997. Substitutive experiments and the evidence for competitive hierarchies in plant communities. *Oikos* **80**(1):179–182.
- Connolly, J., and P. Wayne. 1996. Asymmetric competition between plant species. *Oecologia* **108**:311–320.
- Connolly, J., P. Wayne, and R. Murray. 1990. Time course of plant–plant interactions in experimental mixtures of annuals: density, frequency, and nutrient effects. *Oecologia* **82**:513–526.
- Cousens, R. 1991. Aspects of the design and interpretation of competition (interference) experiments. *Weed Technology* **5**:664–673.
- Damgaard, C. 1998. Plant competition experiments: testing hypotheses and estimating the probability of coexistence. *Ecology* **79**:1760–1767.
- DeBenedictis, P. A. 1977. The meaning and measurement of frequency-dependent competition. *Ecology* **58**:158–166.
- de Wit, C. T. 1960. On competition. *Verslagen van Landbouwkundige Onderzoekingen* **66**(1):1–82.
- Dobson, A. J. 1990. *An introduction to generalized linear models*. Chapman & Hall, New York, New York, USA.
- Edwards, A. W. F. 1992. *Likelihood*, expanded edition. Johns Hopkins Press, Baltimore, Maryland, USA.
- Ellison, A. M. 1999. Abandon all *P*-values, ye who enter here! *Ecology* **80**:2129–2130.
- Farmer, R. E., D. M. Morris, K. B. Weaver, and K. Garlick. 1988. Competition effects in juvenile jack pine and aspen as influenced by density and species ratios. *Journal of Applied Ecology* **25**:1023–1032.
- Firbank, L. G., and A. R. Watkinson. 1985. On the analysis of competition within two-species mixtures of plants. *Journal of Applied Ecology* **22**:503–517.
- Freckleton, R. P., A. R. Watkinson, P. M. Dowling, and A. R. Leys. 2000. Determinants of the abundance of invasive annual weeds: community structure and non-equilibrium dynamics. *Proceedings of the Royal Society of London B* **267**:1153–1161.
- Gibson, D. J., J. Connolly, D. C. Hartnett, and J. D. Weidenhamer. 1999. Designs for greenhouse studies of interactions between plants. *Journal of Ecology* **87**(7):1–16.
- Goldberg, D. E., and S. M. Scheiner. 1993. ANOVA and ANCOVA: field competition experiments. Pages 69–93 in S. M. Scheiner and J. Gurevitch, editors. *Design and analysis of ecological experiments*. Chapman and Hall, New York, New York, USA.
- Gurevitch, J., L. L. Morrow, A. Wallace, and J. S. Walsh. 1992. A meta-analysis of competition in field experiments. *The American Naturalist* **140**(4):539–572.
- Hardin, G. 1960. The competitive exclusion principle. *Science* **131**:1292–1297.
- Hassell, M. P., and H. N. Comins. 1976. Discrete time models for two-species competition. *Theoretical Population Biology* **9**:202–221.
- Hastings, A. 1997. *Population biology: concepts and models*. Springer-Verlag, New York, New York, USA.
- Hilborn, R., and M. Mangel. 1997. *The ecological detective*. Princeton University Press, Princeton, New Jersey, USA.
- Huisman, J., R. R. Jonker, C. Zonneveld, and F. J. Weissing. 1999. Competition for light between phytoplankton species: experimental tests of mechanistic theory. *Ecology* **80**:211–222.
- Inouye, B. D. 1999a. Integrating nested spatial scales: implications for the coexistence of competitors on a patchy resource. *Journal of Animal Ecology* **68**:150–162.
- Inouye, B. D. 1999b. Estimating competition coefficients: strong competition among three species of frugivorous flies. *Oecologia* **120**:588–594.
- Inouye, R. S., and W. M. Schaffer. 1981. On the ecological meaning of ratio (De Wit) diagrams in plant ecology. *Ecology* **62**:1679–1681.
- Juliano, S. A. 1998. Species introduction and replacement among mosquitoes: interspecific resource competition or apparent competition. *Ecology* **79**:255–268.
- Krebs, C. J. 1999. *Ecological methodology*. Second edition. Addison-Wesley Press, Menlo Park, California, USA.
- Law, R., and A. R. Watkinson. 1987. Response-surface analysis of two-species competition: an experiment on *Phleum arenarium* and *Vulpia fasciculata*. *Journal of Ecology* **75**:871–886.
- Leslie, P. H. 1958. A stochastic model for studying the properties of certain biological systems by numerical methods. *Biometrika* **45**(1):16–31.
- May, R. M. 1974. Biological populations with nonoverlapping generations: stable points, stable cycles, and chaos. *Science* **186**:645–647.
- May, R. M., and G. F. Oster. 1976. Bifurcations and dynamic complexity in simple ecological models. *The American Naturalist* **110**:573–599.
- McPeck, M. A. 1988. The consequences of changing the top

- predator in a food web: a comparative experimental approach. *Ecological Monographs* **68**:1–23.
- Park, T. 1948. Experimental studies of interspecies competition. I. Competition between populations of the flour beetles, *Tribolium confusum* Duvall and *Tribolium castaneum* Herbst. *Ecological Monographs* **18**:267–307.
- Peters, R. H. 1991. A critique for ecology. Cambridge University Press, New York, New York, USA.
- Pfister, C. A. 1995. Estimating competition coefficients from census data: a test with field manipulations of tidepool fishes. *American Naturalist* **146**(3):271–291.
- Ricker, W. E. 1954. Stock and recruitment. *Journal of the Fisheries Research Board, Canada* **11**:559–623.
- Ridsdill-Smith, T. J. 1991. Competition in dung-breeding insects. Pages 264–294 in W. J. Bailey and T. J. Ridsdill-Smith, editors. *Reproductive behavior of insects*. Chapman & Hall, London, UK.
- Sackville Hamilton, N. R. 1994. Replacement and additive designs for plant competition studies. *Journal of Applied Ecology* **31**:599–603.
- Schoener, T. W. 1983. Field experiments on interspecific competition. *The American Naturalist* **122**(2):240–285.
- Silvertown, J. W., and J. L. Doust. 1993. Introduction to plant population biology. Blackwell Scientific, Cambridge, Massachusetts, USA.
- Snaydon, R. W. 1991. Replacement or additive designs for competition studies? *Journal of Applied Ecology* **28**:930–946.
- Snaydon, R. W. 1994. Replacement and additive designs revisited: comments on the review paper by N. R. Sackville Hamilton. *Journal of Applied Ecology* **31**:784–786.
- Snaydon, R. W., and E. H. Satorre. 1989. Bivariate diagrams for plant competition data: modifications and interpretation. *Journal of Applied Ecology* **26**:1043–1057.
- Tilman, D. 1994. Competition and biodiversity in spatially structured habitats. *Ecology* **75**:2–16.
- Vandemeer, J. H. 1969. The competitive structure of communities: an experimental approach with protozoa. *Ecology* **50**:362–371.
- Veevers, A., and T. B. Boffey. 1979. Designs for balanced observation of plant competition. *Journal of Statistical Planning and Inference* **3**:325–331.
- Watkinson, A. R. 1981. Interference in pure and mixed populations of *Agrostemma githago*. *Journal of Applied Ecology* **18**:967–976.
- Wilbur, H. M. 1972. Competition, predation, and the structure of the *Ambystoma–Rana sylvatica* community. *Ecology* **53**:3–21.