Journal of Ecology 1999, **87**, 1–16

Designs for greenhouse studies of interactions between plants

DAVID J. GIBSON*, JOHN CONNOLLY[†], DAVID C. HARTNETT[‡] and JEFFREY D. WEIDENHAMER§

*Department of Plant Biology, Southern Illinois University, Carbondale, IL 62901–6509, USA; †Department of Statistics, University College Dublin, Dublin 4, Ireland; ‡Division of Biology, Ackert Hall, Kansas State University, Manhattan, KS 66506, USA; and §Department of Chemistry, Geology and Physics, Ashland University, Ashland, OH 44805–3799, USA

Summary

1 Designs for greenhouse studies of interactions between plants are reviewed and recommendations for their use are provided.

2 Papers published over a 10-year period showed the replacement series design to be the most popular, especially in studying crop–weed interactions. Fifty per cent of the studies involved only two species, although studies testing the interaction between different genotypes of only a few species were also popular.

3 Limitations imposed by the choice of design, the variables measured, and the analysis used on the range of inferences that may be validly drawn from the experiment are frequently not well understood or appropriate for the questions that appear to be addressed. One example is the failure to distinguish the outcome of competition (the long-term outcome of interaction) and the effects of species on each other.

4 Studies in which only final yield is measured are severely limited as to the inferences which may be drawn. Effects due to interspecific interaction during the course of the experiment cannot then be separated from pre-existing differences, and interpretation may be biased towards species whose individuals were initially larger. In addition, measurements at several times are necessary to understand the changing dynamics of species interaction.

5 Simple pair-wise mixtures can assess the effect of treatment factors on the outcome of competition. Replacement series and related diallel designs generally produce results that may be size-biased even when initial interspecific differences are known. Additive designs (including target–neighbour designs), despite confounding density with species proportions, offer considerable scope for addressing mechanistic questions about interspecific interactions. Designs that allow response surface analysis can avoid many of the problems inherent in the other methods, but all need to be adjusted for initial interspecific differences. Designs for multiple species experiments are still largely untested, although several designs have been used. At the level of the individual plant, hexagonal fan designs permit study of the effects of varying the spatial pattern, and the densities and the relative proportions of interacting species, but suffer from lack of independence and lack of randomization.

Keywords: additive, competition, competitive hierarchy, diallel, experimental design, hexagonal fans, interspecific interaction, replacement series, response surface, size-bias, target-neighbour design

Journal of Ecology (1999) 87, 1-16

Introduction

© 1999 British Ecological Society The importance of interactions between plants in determining the structure and dynamics of plant com-

munities is widely recognized (e.g. Grime 1979; Aarssen 1983; Tilman 1988; Keddy 1989; Grace & Tilman 1990). However, demonstrating the effects of **2** Greenhouse interaction studies these interactions in the field has often proved difficult (e.g. Strong et al. 1984; Connell 1990) so attention has tended to focus on studies of artificial communities growing in the greenhouse and in the field. While studies of plant interactions in natural and semi-natural communities have been the subject of several comprehensive reviews (Connell 1983; Schoener 1983; Underwood 1986; Aarssen & Epp 1990; Goldberg & Barton 1992; Goldberg & Scheiner 1993), greenhouse experiments have not been similarly reviewed, although the different approaches used have been described both in general terms (Harper 1977; Silvertown & Lovett Doust 1993) and for particular applications (Dekker et al. 1983; Radosevich 1987; Weidenhamer et al. 1989; Weidenhamer 1996). Statistical issues and questions have also been raised about the logical validity of some of the different designs as they are commonly used (e.g. Inouve & Schaffer 1981; Jolliffe et al. 1984; Connolly 1986, 1997; Rejmánek et al. 1989; Roush et al. 1989; Firbank & Watkinson 1990; Snaydon 1991).

Having surveyed appropriate journals published over a 10-year period for such experiments, we assessed the methodology (in particular the main experimental designs that have been used in studying plant–plant interactions under greenhouse conditions) and have made some recommendations for future practice.

Why study interactions of artificial communities in the greenhouse?

The complexity of natural plant communities imposes logistic and analytical constraints on studying plant interactions. For example, large numbers of species may be present, both environmental factors and species abundance show heterogeneity in time and space, and the size and age of the plants present will vary. By contrast, specially created artificial plant communities consisting of a few species, perhaps arranged in a particular pattern, with the plants of a specified age or ontogenetic stage, and with environmental conditions quite uniform and carefully controlled, can be used to examine inter- and intraspecific interactions more precisely. Further advantages of such controlled conditions are that the effects of other factors (e.g. soil fertility, pathogens and herbivory) can be more readily evaluated (Keddy 1989), and that such studies enable mechanistic interpretation rather than simple phenomenological observation (Tilman 1987; Stiling 1992).

The high degree of experimental control, repeatability, precision and amenability to rigorous statistical design make the use of artificial communities and greenhouse experiments appealing (de Wit 1960; Harper 1983; Hairston 1989), although others (e.g. Diamond 1986) have stressed the undoubted limitations. In particular, the lack of realism restricts the ability to apply the results of such experiments to complex natural communities: long-term greenhouse experiments with perennial plants can be especially unrealistic due to the inflexible restriction of rooting volume. However, such studies do allow the separation of different components of species interaction, such as effect and response (*sensu* Goldberg 1990), and determination of relative efficiency (Connolly *et al.* 1990). In addition, the mechanisms of interaction (e.g. through root and shoot capture of resources) are more amenable to study under controlled conditions. Despite the limitations, unless plant interactions can be demonstrated under greenhouse conditions they are unlikely to be of importance in natural communities.

Framework for the review

We are conscious that several aspects of the methodology used in the study of plant interactions have engendered heated debate. Replacement series (RS) or substitutive designs (de Wit 1960), in which one species gradually replaces another in a mixture at constant overall density, have been much used but no consensus has emerged since the breakdown in general acceptance of this design. In our opinion the limitations of other approaches have been glossed over and, in general, there has been an inadequate appreciation of the limited nature of inferences that can be drawn from several such techniques that have been widely used. We do not presume to offer a resolution of all issues in this review. Indeed, we do not believe that there is currently available in the literature a full context for the resolution of the difficulties posed by the study of plant-plant interaction, but we hope that, by adopting a critical review of some of the issues, we will help to clear a path towards such a resolution. While agreeing with Cousens (1996) that 'it is illogical to condemn a group of experimental treatments for all purposes simply because of the ways in which some experimenters choose to interpret the results', we consider that a critical analysis of methods is essential if such misuse is widespread and if the area appears beset by deep confusion.

Strictly speaking, an interaction between two plants is any association between plants in a mixture that affects the net reproductive rate (\mathbf{R}_{0}) of the component species (Silvertown & Lovett Doust 1993). However, this definition may be too restrictive in practice as in many studies R_o is not measured and inferences are made on the basis of vegetative characters (Jolliffe et al. 1984; although see Benner & Bazzaz 1987; Law & Watkinson 1987). We therefore use the term in the broader sense of any effect one species has on another. There are many forms of interaction and many terms are used to specify particular facets of interaction (e.g. competitive ability, suppression, enhancement, intensity and importance of competition), and many analyses/indices have been proposed to provide a quantitative measure of them.

© 1999 British Ecological Society, *Journal of Ecology*, **87**, 1–16 Rather than a detailed review of definitions and methods, we wish to provide a more narrowly focused critique of current experimental practice with a view to demonstrating some important limitations and providing some pointers as to how they may be avoided. We regard this as a necessary starting point in creating a framework for studies of plant–plant interactions within which the definitions of different forms of interaction and the methods used to measure them will be free of the difficulties that are outlined below.

Although there may not currently be general agreement in the literature on the following key concepts, we feel that they contribute substantially to the creation of a valid framework for studies on interspecific interaction. Some of these points cause fundamental difficulties for particular approaches, while others merely limit the range of inferences that may be drawn from certain studies.

Distinction between 'outcome of competition' and 'effects of species on each other'

We deal with two main aspects of interspecific interaction. 'Outcome of competition' refers to the relative long-term success of species, i.e. the end point for the community in terms of its composition and we are concerned with what indications short-term experiments can give about this. 'Effects of species on each other' refers to the impact of each species on the other (Goldberg & Werner 1983; Goldberg 1990) and may be an important part of the process that determines the end point, but is distinct from it. So while these two aspects of interaction may often be related, they are not equivalent, and their study may require different techniques. We believe that they are regularly confounded in current practice.

Since competitive exclusion rarely occurs in shortterm experiments, the primary indicator available, however inadequate it may be, of long-term prospects is increased dominance of a species in a mixture, i.e. greater gain in terms of greater output per unit input. Although an estimate of this may be made from a single mixture, assessment of the effects of species on each other generally requires the inclusion of a range of mixtures and/or monocultures in the design.

Many experiments use methods and indices (e.g. relative crowding coefficients, coefficients of aggressivity, relative yield total) that purport to reflect the outcome of competition (Keddy & Shipley 1989) but actually address the questions of effects of species on each other or an amalgam of both. Furthermore, the indices and analytical methods used are generally susceptible to bias because they ignore initial differences between components, and therefore tend to favour larger individuals (Connolly 1986; Grace *et al.* 1993).

© 1999 British Ecological Society, *Journal of Ecology*, **87**, 1–16

The role of time, initial conditions and plant size/life history

Analysis on the basis of final yield alone may be misleading, as although final yield represents a summation of the effects of plant interaction over the course of the experiment, it may also partly reflect initial differences. Initial size differences must be discounted to assess plant interactions adequately over the experimental period and make the comparison fair to both species. The final per unit size of a species will depend on both initial size and interspecific interactions (e.g. by an asymmetric effect such as increasing its shading impact more than pro rata to its size). These effects can be included as part of the explanation of subsequent performance (e.g. Connolly & Wayne 1996). This double role and use of initial size allows the experimenter to deal with situations where species ontogeny, other life-history traits, or direct experimental manipulation (e.g. of sowing date) lead to considerable differences in size between species at the commencement of the experiment. The sole use of final yield will also miss dynamic changes in species interaction (e.g. Connolly et al. 1990; Turkington & Jolliffe 1996) and is possibly the single most neglected and important issue in current practice.

The difficulty with density

Many experimental designs (e.g. including most RS designs or some additive designs) equate species on the basis of their numbers. However, simple equivalence on the basis of density can introduce size bias (Connolly 1986, 1997; Silvertown & Dale 1991; Grace *et al.* 1993) and thus distort an assessment of interspecific relations (e.g. Connolly 1986; Snaydon 1991). Snaydon (1991) gives the extreme example of the nonsense of equating densities of oak trees and daisy plants. Size differences may of course reflect life-history traits or natural conditions, and, where present, must therefore be allowed for, for example in the double way suggested in the section above.

Competition and single mixtures

Most assessments of interspecific interactions have used a single mixture (usually 50:50) in addition to the relevant monoculture(s). Data from additional mixtures or monocultures may be useful if they allow generalizations (since interaction may depend on the proportions and densities of the components) and will increase the precision of estimation of what is observable in the single mixture. However, a problem arises when such extra data contradict the findings obtained for a particular mixture (e.g. Benner & Bazzaz 1987; an example in Connolly 1997). In other words, because of the issues of size and density equivalence raised in the second and third sections above, a monoculture may not always be the appropriate reference point for assessing the interactions in a particular mixture.

Limitations on inferences (logical limitations vs. misuse)

There are very few useless experiments (Cousens 1996). However, the inferences that can be validly drawn from a particular experiment depend on the design used, the measurements taken and the analysis of the data. If these logical limitations to inference are not fully appreciated, e.g. the second and third sections above, then biased assessments will result. We distinguish these logical limitations, which lead to pushing the inferences beyond what the design and measurements would support, from inappropriate interpretation resulting from faults with the design *per se* (i.e. misuse) (Cousens 1996).

Predictive power

Most studies of interspecific interactions are shortterm, frequently lasting less than a year, often measuring only vegetative growth rather than reproductive success, and based on one phase of the lifehistory of species, whilst largely ignoring the rest. We must not expect too much predictive power from such experiments, unless we are convinced that the phase being tested is critically important (e.g. vegetative biomass can be used as a measure of fitness in many annual plants; Goldberg & Fleetwood 1987). They will usually supply no more than indicators to the answers required by ecologists, although they may be of more direct use to the interests of agronomists (e.g. Shrefler *et al.* 1994).

Our discussion develops analyses of the relationships between (i) questions that appear to be asked in studies of interspecific interaction, (ii) variables that are measured, and (iii) the designs used, with a view to identifying the range and limits of questions that can be validly addressed using particular combinations of design and variables measured (J. Connolly, P. Wayne & F. A. Bazzaz, unpublished data). J. Connolly et al. concluded that the omission of initial information severely limits the inferences that can be drawn using several of the most common designs. In the case of RS, even with the provision of this initial information the range of inference is still quite limited, and in other designs the comparison of species may remain problematic. J. Connolly et al. also draw attention to the distinction between the outcome of competition and the effects of species on each other (see the first section above) as an issue that has led to confusion in interpretation of studies on species interaction.

© 1999 British Ecological Society, *Journal of Ecology*, **87**, 1–16

Literature survey

We surveyed studies published during 1984–93, in 11 journals (American Journal of Botany, American

Midland Naturalist, American Naturalist, Canadian Journal of Botany, Ecology, Ecological Monographs, Oikos, Journal of Applied Ecology, Journal of Ecology, Journal of Vegetation Science and Weed Science). Although the last paper in our survey was from 1993, we believe that the findings are still relevant at the time of the final revision to this paper. Ninety-nine studies contained a total of 107 experiments on plantplant interactions conducted in a greenhouse (the citations and designs used in these studies can be found in The Journal of Ecology's archive on the World Wide Web (WWW): see recent issue for address. The studies selected were limited to those with interspecific mixtures, except that intraspecific mixtures were also included when different genotypes, varieties or maternal lines were investigated. For each study, the following information was noted: experimental design, number of species studied, and identity and number of experimental treatments.

Most of the 107 experiments (35%) used RS designs, with two other designs (additive and simple pair-wise, see later) accounting for most of the rest (26% and 22%, respectively, Table 1). Clearly, RS has been the most widely used design in agricultural studies or investigations of crop-weed interactions. *Weed Science* (28%) and *Journal of Ecology* (20%) were the most commonly used journals, with an additional 23% of the studies reported in *Journal of Applied Ecology* or *Ecology*.

Fifty per cent of the studies surveyed examined interactions in mixtures involving only two species (Table 2) and fewer studies were encountered as the number of species tested increased. Only two studies in our survey examined seven species (Rabinowitz et al. 1984; Goldberg & Landa 1991) and multi-species designs (by definition) regularly studied three or more species (six in two studies: Austin et al. 1985; Thórhallsdóttir 1990). One study used three species, each of 10 genotypes, in all possible two-genotype pairs according to a diallel design (Taylor & Aarssen 1990). Gaudet & Keddy (1988) used a modified additive design to measure the relative competitive ability of 44 herbaceous plant species, but this ambitious study was not included in our survey. Twenty general topics were addressed in the 107 experiments (Table 3). Crop-weed interaction was the most frequent (21 studies), with interactions between or among genotypes, and effects of soils and nutrients, also common.

Types of design

The experimental designs that have been used in studies of plant interactions have been classified in various ways (e.g. Harper 1977; Dekker *et al.* 1983; Radosevich 1987; Austin *et al.* 1988; Rejmánek *et al.* 1989; Firbank & Watkinson 1990; Snaydon 1991; Silvertown & Lovett Doust 1993). Despite the use of different terms, three main types of design are commonly recognized: simple pair-wise (SP), additive (AD) and Table 1 Number of greenhouse experiments of plant interactions published in 99 studies in 11 leading journals from 1984 to 1993. AMN, *American Midland Naturalist*; AN, *American Naturalist*; AJB, *American Journal of Botany*; CJB, *Canadian Journal of Botany*; Ecol, *Ecology*; EM, *Ecological Monographs*; JAE, *Journal of Applied Ecology*; JE, *Journal of Ecology*; JVS, *Journal of Vegetation Science*; Wsci, *Weed Science*

Design	AMN	AN	AJB	CJB	Ecol	EM	JAE	JE	JVS	Oikos	Wsci	Total
Simple pair-wise		1	3	3	5	1		4		1	6	24
Additive		2	2	1	6		4	5		2	6	28
Replacement	3		1		3		6	7			17	37
Diallel		1	1	3								5
Fan								2				2
Multi-species mixtures				2	1			2	1	1	1	8
Other								2			1	3
Total	3	4	7	9	15	1	10	22	1	4	31	107*

*107 experiments are listed from 99 studies because some studies involved a combination of experiments and designs.

Table 2 Designs used and number of species tested in 99 greenhouse interference studies from 1984 to 1993

	Number of species								
Design	1	2	3	4	5	6	7	Total	
Simple pair-wise	5	10	5	3		1		24	
Additive	5	12	4	5	1		1	28	
Replacement series		26	8	1		1	1	37	
Diallel		2	3					5	
Fan		2						2	
Multi-species	1†		2	2	1	2		8	
Other:		2			1			3	
Total	11	53	22	11	3	4	2	107*	

*107 experiments are listed from 99 studies because some studies involved a combination of experiments and designs.

[†]A single test plant of *Ailanthus atlissima* was grown with germinating seedlings from old-field seed bank samples.

[‡]Not clearly classified as one of the designs listed above.

RS (also called substitutive designs). The differences are illustrated in Fig. 1. SP designs usually maintain a 1:1 ratio of the two competitors, whereas in the simplest case of AD experiments the density of one species is held constant while the density of the other species is varied. In RS, species are grown in varying proportions and compared to growth in monoculture, with the total density held constant across all mixtures/monocultures. A design for n species that consists of RS for all possible pair-wise combinations between the species is termed a mixture diallel design. Designs for response surfaces may consist of additive or substitutive designs at a range of densities, or may be constructed in other ways (e.g. Connolly 1987; Law & Watkinson 1987; Rejmánek et al. 1989; Roush et al. 1989; Snaydon 1991; Turkington & Jolliffe 1996). Less often used are spatially explicit designs (hexagonal fan designs) and those used to investigate multi-species interactions. Although not the focus of this review or our survey of the literature, our comments also have relevance for field experiments using these designs.

© 1999 British Ecological Society, *Journal of Ecology*, **87**, 1–16

SP DESIGNS

In SP experiments (also called additive, equal proportions; Austin *et al.* 1988), mixtures consisting of a fixed, usually 1:1, ratio of the two species are maintained (Fig. 1a). SP designs have been used to examine the role of numerous factors in plant interactions, frequently using a range of treatments applied to a particular mixture of two species (see the WWW archive for examples). Additions of monocultures at appropriate densities can convert SP designs to AD (e.g. Gurevitch *et al.* 1990) or RS (e.g. Berendse *et al.* 1992) experiments. Some studies are difficult to classify as strictly SP, diallel or AD studies (e.g. Allen & Allen 1984, where the design is a partial diallel, with pair-wise comparisons of *Salsola kali* with two other species, but not between the other two species).

SP designs at a single relative frequency and density can be used, in a limited way, to address questions about the outcome of competition between two species. Measurements over time should be included to o Greenhouse interaction studies **Table 3** Topics addressed in 107 greenhouse studies published in 11 journals from 1984 to 1993. Several studies testedmore than one factor

Topic	Number of studies
Crops and weeds	21
Genotypes	15
Soils and nutrients	15
Fungi, bacteria and diseases	9
Grazing	9
Moisture	9
Plant form and performance	8
Germination and seeds	7
Planting density	7
Spatial patterns	6
Abundance	5
Photosynthesis and light	5
Carbon dioxide	4
Modelling and data analysis	4
Roots	3
Herbicides	2
Leachates and allelopathy	2
Temperature	2
Breeding systems	1
Site of origin	1
Total	135

allow assessment of changes in relative abundance. However, SP designs do not allow assessment of the effects of species on each other, unless one or other species completely disappears. If final yield is the only parameter available then all that one can safely say is whether both species survived and which contributed most to final biomass. If an experiment includes pairwise mixtures between more than two species, then comparisons of interspecific interactions for different mixtures may be problematic (J. Connolly *et al.*, unpublished data).

SP designs therefore provide a useful, if limited, tool for screening the effects of a treatment gradient on the outcome of competition; they are efficient in that no resources are allocated to monocultures, which may not provide useful information on the question addressed. In addition, they are amenable to fairly straightforward statistical treatment, and the difficulty raised by the probable correlation between responses in a mixture can often be avoided by combining the responses to give a single per-pot measure. Perhaps SP designs are used less frequently than they should be.

RS DESIGNS

In an RS, the planting density of the two constituent species may vary but the total density is held constant



© 1999 British Ecological Society, *Journal of Ecology*, **87**, 1–16

Fig. 1 Five designs of competition experiments plotted on joint abundance diagrams denoting the density (N) of species i and species j (designs a, b, c and d) and six component species in a diallel design (e) (after Rejmánek *et al.* 1989; Silvertown & Lovett Doust 1993). In (a–d) lightly shaded symbols represent monocultures. (a) Simple pair-wise (SP) design at multiple densities and without the monocultures included by some investigators; (b) replacement series (RS) at a single total density; (c) target–neighbour or partial additive form of an additive (AD) design with a constant density of component *i*; (d) additive series; (e) diallel design including redundant intraspecific mixtures (lightly shaded symbols).

7 D.J. Gibson et al. (Fig. 1b). The effect of other factors (e.g. a soil nutrient) on interaction between the components is tested by using a replicate RS for each of several levels of the factor. Ratio and replacement diagrams (Harper 1977) offer graphical presentation of results. Of the several indices that have been proposed to present the results of RS experiments (Trenbath 1978; Connolly 1986), relative yield total (RYT; de Wit & van den Bergh 1965) is generally the most popular. The objective of some of these indices (e.g. relative crowding coefficients, de Wit 1960; competitive ratio, Willey & Rao 1980; coefficient of aggressivity, McGilchrist & Trenbath 1971) generally appears (although this is not often clearly stated) to be an attempt to assess the outcome of competition, whereas the RYT, a single value for the stand, relates to the joint capture and use of resources by the competing species (i.e. it describes a niche relationship). Although niche relationships contribute to understanding why a particular outcome occurs, they rarely predicate any particular outcome; thus the value of an index like RYT does not determine one way or another what impact niche separation will have on the outcome of competition.

Replacement designs have been widely used since they were introduced by de Wit (1960). Applications include aspects of inter- and intraspecific interactions between wild plants (e.g. Solbrig *et al.* 1988; Fone 1989), between wild plants (weeds) and crops (e.g. Ogg *et al.* 1993; Wall 1993), and between commercial cultivars of forage grasses (e.g. Frankow-Lindberg 1985). In the majority of studies, yield is the only character assessed, although other measures such as shoot/root ratios may perhaps help to illustrate the physiological basis of species' interactions (Bi & Turvey 1994).

We identified five problems with the RS design that seriously undermine its usefulness as an experimental tool (see also Cousens 1996). (i) It is generally used with final yields only, which can lead to size bias in interpretation if species differ in initial size (Connolly 1986, 1997; Grace et al. 1992; but disputed by Shipley & Keddy 1994). (ii) The validity of the RS method rests on the assumption that individuals of the competing species are exactly equivalent at the start of the experiment (Keddy 1989). If seedlings are of quite different sizes then it is both difficult to see how they can be regarded as equivalent (Connolly 1986; Snaydon 1991) and impossible to eliminate size bias (J. Connolly et al., unpublished data). (iii) The outcome of competition is frequently confused with the effects of neighbours when interpreting results of RS. Including information from monocultures in the analysis can introduce bias in the assessment of species' effects both on each other in a mixture and on the long-term outcome (Connolly 1986; J. Connolly et al., unpublished data). (iv) RS are carried out at a fixed, and often arbitrarily chosen, density (Inouye & Schaffer 1981; Taylor & Aarssen 1989; Snaydon 1991; Silver-

© 1999 British Ecological Society, *Journal of Ecology*, **87**, 1–16 town & Lovett Doust 1993) and results at that density may not generalize. Some of the density problem can be overcome by using replicate RS at different total densities derived from an additive series over a range of densities (Fig. 1d) (Firbank & Watkinson 1985; Cousens & O'Neill 1993). (v) Logistically, RS experiments necessitate tying up large numbers of experimental units (66% if only a singe mixture is used) in monocultures that may not contribute significantly to the analysis.

These problems lead to difficulty in correctly interpreting both RS diagrams and competition indices (Connolly 1986, 1988, 1997; Snaydon 1994). We are led to agree with the critics of this method (e.g. Inouye & Schaffer 1981; Jolliffe *et al.* 1984; Connolly 1986, 1988, 1997; Law & Watkinson 1987; Snaydon 1991, 1994): while RS may yield some useful information (Cousens 1996) it will be on a very limited range of questions. The tendency to misuse the method is so pervasive that its continued use should be discouraged.

AD AND TARGET-NEIGHBOUR DESIGNS

In the simplest form of AD designs (i.e. the partial additive) the density of the focal species is maintained across all mixtures and the density of the associate species is varied, usually with the goal of assessing the response of the focal species to increasing levels of the associate (Fig. 1c). More complex designs involve simultaneously varying the proportions of focal and associate species (i.e. addition series; Fig. 1d). This approach has useful applications, such as studying the impact of varying densities and distributions of weed populations on a crop sown at fixed density (Zimdahl 1980; Radosevich 1987; see the WWW archive). Additive designs have also been used to assess the role of various factors (e.g. relatedness, genotype, emergence time, initial plant size, maternal effects, herbivory) on a focal species' response to its associate in situations where comparing intra- vs. interspecific interactions and distinguishing effects of species' proportions from those of total density were less important objectives (see the WWW archive). They have been used for distinguishing allelopathic effects from resource exploitation due to densitydependent phytotoxic effects (Weidenhamer 1996; Weidenhamer et al. 1989).

A problem with this design is that the overall density and the proportions of focal and associate species can vary simultaneously and this confounding of variables makes the interpretation of results difficult (although not necessarily unrealistic compared with field situations) (Harper 1977; Silvertown & Lovett Doust 1993). Some of the problems of confounding the effects of species' proportion and density can be overcome by independently manipulating densities of both species and analysing performance of the focal species via response surface methods (Firbank & Watkinson 1985; Law & Watkinson 1987; Fredshavn 1994).

Target-neighbour designs involve growing an individual of a 'target' species with varying abundances of 'neighbours', which could be either an associate species or itself. This is essentially an AD design in which the density of the focal or target species is reduced to a single individual or to a density low enough to preclude significant intraspecific interactions. This design has been used to address a variety of mechanistic questions about plant interactions (see the WWW archive for examples). Goldberg & Landa (1991) used it to determine which plant traits are responsible for differences in the effects and responses between species, and whether these two measures of interaction are related.

The per unit (per capita or per unit biomass) effect of neighbours on individuals of a target species is measured as the slope of a regression of target plant performance against the number (or biomass) of immediate neighbours. The target-neighbour design focuses on individual plant responses rather than the mean population response and estimates the importance of interspecific interactions relative to other factors in determining the fate and performance of individuals. While these measures on the target do give information at the individual plant level, they do not allow direct assessment of the outcome of competition for the target since the factors affecting the target may also affect the neighbours to the same or greater degree. For example, increasing the density of an associate may greatly reduce the performance of the target but it may also reduce the performance of the associate. Comparison of the impact on both species is essential in assessing the outcome of competition.

This approach claims numerous additional advantages. By measuring interference on a per unit basis it incorporates asymmetries in individual plant size at harvest among competing species. The relationships can be useful in interpreting features of interspecific interactions. Comparison of the slopes of the target performance-neighbour abundance regressions can be used as a quantitative measure of the effect (sensu Goldberg & Fleetwood 1987) of different neighbour species (Goldberg & Landa 1991). Statistical comparison of these slopes under different conditions may be made using ANCOVA (e.g. Hartnett et al. 1993). However, as in all ANCOVA, care must be taken in interpretation if the covariate is estimated after the commencement of the experiment as it may also contain effects of treatments that are discounted in the comparison of slopes. For example, the method compares the effects of two associate species on the target as if they had the same final yield and, if this is not the case, may lead to an unfair comparison. The competitive 'response' can be estimated from the slopes of regression coefficients when different target species are grown with the same neighbour species (Goldberg

© 1999 British Ecological Society, *Journal of Ecology*, **87**, 1–16 & Landa 1991; Hartnett *et al.* 1993). This general approach has been described in some detail by Goldberg & Werner (1983) for use in field-based studies. Discussion of some statistical considerations for these types of additive experiments is found in Goldberg & Scheiner (1993).

An advantage claimed for target–neighbour experiments is their economy in terms of both space and plants (Thijs *et al.* 1994; compare Hartnett *et al.* 1993 with Hetrick *et al.* 1994). However, caution is necessary in claiming greater efficiency for one design over another. The statistical criterion used to compare the efficiency of different designs should in each case be the experimental resource required to achieve a particular precision in the estimation of a particular parameter(s). However, considerations other than statistical efficiency may influence the selection of design and measurement: a design that is somewhat less efficient for one particular purpose may provide a far wider basis for inference and may thus be usable to address a wider range of questions.

A variation of the target-neighbour approach incorporates measurements of the distance, as well as biomass or numbers, of neighbours, and so allows the decreasing effects of 'non-nearest neighbours' to be incorporated.

In practice, AD and target-neighbour designs often consider only final yield (but see Gibson & Skeel 1996) and so suffer from ignoring the time-course of interactions and initial differences in species' size. As well as confounding species density and relative frequency, they sometimes equate species simply on the basis of density (e.g. in comparing regression coefficients for different neighbour species where density is the independent variable in the regression). Thus conclusions may well be affected by size bias in a manner similar to RS, leading to certain species being judged more competitive simply because they were initially larger. Even if all information on initial sizes is available, a partial additive or additive series will not allow the same range of questions to be addressed as a response surface approach would (e.g. Connolly & Wayne 1996).

Despite the biases that can occur with these methods, comparisons among a range of treatments applied to the same additive series and species will give a basis for ranking treatments relative to each other, even if the absolute level of the effects may be biased. However, comparisons of treatments across species potentially suffer from difficulties unless initial size differences are measured and accounted for.

RESPONSE SURFACE METHODS

An experimental design that includes a range of densities and relative frequencies of the species under study (not necessarily including any monocultures, e.g. Connolly & Wayne 1996; Ramseier *et al.* 1996) 9 D.J. Gibson et al. may be used to generate response models for each species. Such a design allows the fitting of regressionstyle response models relating some measure of per capita performance for each species to the density of each species (e.g. Suehiro & Ogawa 1980; Spitters 1983; Connolly 1987; Law & Watkinson 1987), the initial biomass of each species (Connolly & Wayne 1996) or some other initial measure of biological potential, such as early leaf area index of each species (e.g. Kropf & Spitters 1991). The response models and their parameters are used to assess species interaction. These methods avoid some of the problems inherent in the analysis of replacement and additive designs, and in diallel analysis (e.g. Law & Watkinson 1987; Bullock et al. 1995; Connolly & Wayne 1996). As with additive designs, the inclusion of initial and intermediate measurements allows the study of species' interactions over time. Connolly et al. (1990) and Menchaca & Connolly (1990) report changes in species' interactions over time that would have been totally overlooked in an analysis of final yield only. Indeed, the conclusions drawn from a response surface analysis incorporating the time-course of plantplant interactions can be qualitatively different from, and more effectively predict the outcome of competition than, those derived from an RS (Connolly et al. 1990; Grace et al. 1993). The inclusion, for example non-destructive leaf demographic measurements, can provide a tool for time series/growth dynamics to be made.

Several ways of designing experiments for response surface models have been described. These include establishing an RS at several total densities, called an addition series (e.g. Spitters 1983; Connolly 1987; Radosevich 1987; Rejmánek et al. 1989; Rodriguez 1997) or establishing additive series (Fig. 1d, which can be regarded as either an additive design or a number of RS at different densities), similar to the bivariate factorial defined by Snaydon (1991). Any set of mixtures that allows the fitting of bivariate response models will suffice. In the absence of a statistical assessment the choice of optimal method is a moot point and may vary with the question being addressed.

Despite their definite superiority to RS and AD designs and methods, the response surface methods may also suffer from similar size bias in the estimations of species' effects and responses and the outcome of competition, unless initial differences are allowed for and the appropriate response measurements are analysed (J. Connolly et al., unpublished data). An example that corrects for initial differences is given in Connolly & Wayne (1996). Furthermore, there are several statistical issues in the fitting of some of these models (and those for AD), e.g. there is often a decrease in variance with decreasing plant size (Connolly et al. 1990) that should be allowed for. In estimating hyperbolic yield-density relationships it is preferable to use weighted regression, non-linear

© 1999 British Ecological Society, Journal of Ecology, 87.1-16

methods or the generalized linear model approach (Nelder & Wedderburn 1972) available in many statistical packages.

DIALLEL DESIGNS

Diallel designs use 'all possible combinations of nspecies', i.e. (n(n-1)/2) separate RS of two species, each represented by two pure stands and one equiproportioned mixture' (Harper 1977, p. 268; Trenbath 1978; but see also Gleeson & McGilchrist 1980 for unequal proportion extensions). Interspecific interactions are assessed using RS methods (Gurevitch et al. 1990) or by analysing a matrix of species' performance using ANOVA and/or covariance analysis (Trenbath 1978). A matrix of 'competition coefficients' (sensu Firbank & Watkinson 1985) calculated as slopes of regressions in a series of pair-wise target-neighbour experiments can also be analysed by diallel methods. Data from diallel designs carried out at more than one density may be analysed by response surface methods (Connolly 1987).

Diallel designs are derived from genetic analysis (Durrant 1965) and have been used extensively in the greenhouse and field by plant breeders and agronomists to assess interspecific interactions between cereal varieties and among forage grasses (e.g. Norrington-Davies & Hutto 1972; Rousvoal & Gallais 1973). Applications to better understand natural systems include Aarssen's (1988) study of four pasture species, Taylor & Aarssen's (1990) study of the interactions among 10 genotypes of three perennial grasses, and Aplet & Laven's (1993) study of the competitive hierarchy of four Hawaiian shrubs. The debate on competitive hierarchies (see below) relies heavily on results from experiments using diallel designs.

The diallel design at a single density is subject to the same difficulties in interpretation as RS.

HEXAGONAL FAN DESIGNS

Most experiments on interspecific interaction focus on the mean population responses of species (e.g. species yield) under varying densities and species' proportions. However, an important feature of plants and other sessile organisms is that they do not sense or respond to overall population density or frequency, but only interact with their immediate neighbours (Harper 1977). This principle argues strongly for designs, such as the target-neighbour and fan designs, that focus on the interaction between a plant and its immediate neighbours. Mead (1979) lists five spatial factors that may affect interspecific interactions between two species and hence can be included as factors in design, namely the density and the intraspecific spatial arrangement of each species and the intimacy of their interspecific arrangement. There are many approaches to the study of intraspecific interactions between individuals (Firbank & Watkinson 1987), and some of the statistical issues were reviewed by Mead (1979). Fan designs were the only ones found in our literature survey.

Hexagonal fan designs utilize a particular plant spacing pattern involving a honeycomb of overlapping hexagons such that each individual is surrounded by zero to six intraspecific neighbours and six to zero interspecific neighbours. This array of hexagons is arranged in a plant spacing gradient (fan design) with plants positioned in a particular pattern, such as a polar coordinate grid or a parallel row design (Nelder 1962; Bleasdale 1967). Thus fan designs vary density and frequency and select a particular form for intraspecific spatial arrangement and interspecific intimacy (see illustrations in the studies listed in the WWW archive).

Hexagonal fan experiments developed as a combination of the fan designs used in agronomic trials to examine the effects of plant density (Nelder 1962), and hexagonal planting designs were developed by Boffey & Veevers (1977) to study the effects of neighbour species' frequencies. Thus they have the advantages of incorporating variation in species' proportions and densities and the local spatial distribution of neighbours in assessing the response of individual plants to neighbours. In addition, they can be significantly more efficient in use of greenhouse space relative to other designs (Antonovics & Fowler 1985). Schmid & Harper (1985) used a fan design to show that interspecific interactions change in varying ways with changing density, sometimes with complete reversals of competitive outcomes between two species at different total densities. In addition, hexagonal fan experiments help in the assessment of optimal planting arrangements in mixedcropping systems and facilitate the analysis of frequency and density-dependent selection in genotype mixtures (Antonovics & Fowler 1985).

The primary advantages of hexagonal fan designs are their focus on neighbourhood interactions, their efficiency in use of space and plants, and their ability to allow assessment of interspecific interactions across a range of densities or plant spacing patterns. However, there are statistical problems associated with the analyses of such designs (Mead 1979; Antonovics & Fowler 1985): they are unrandomized and so may be biased due to underlying trends in fans, the correlated responses in neighbouring plants may require a more complex analysis, and they may have limitations in situations in which second or third nearest neighbour effects and more diffuse interactions are significant. Often the analysis of these designs assumes that 'non-nearest neighbour' effects are insignificant. In addition to these statistical difficulties, size bias may arise if initial size differences are not discounted. Like all studies of individual rather than mean response, they require a greater input of time and labour. These designs can be extended to study multispecies interactions (see below).

© 1999 British Ecological Society, *Journal of Ecology*, **87**, 1–16

DESIGNS TO ASSESS MULTI-SPECIES INTERACTIONS

Despite attempts to provide the greatest degree of realism to interaction experiments, greenhouse experiments involving interspecific interactions among mixtures of three or more species, i.e. diffuse or multispecies interactions (MacArthur 1972), have been infrequent. This is perhaps not surprising given the logistical and statistical problems inherent in the effective design and interpretation of just the multiple pair-wise interaction experiments of the diallel design (Mitchley 1987). The growth of multi-species mixtures under various treatments can be used simply to assess the outcome of competition (e.g. Grime *et al.* 1987), but we have identified five further main approaches to assessing multi-species interactions in the greenhouse.

(i) Fowler (1982) showed that, in a three-species RS design (de Wit 1960), predicted yield per plant was statistically related to the observed yield per plant. Interpretations agreed with results obtained from pair-wise RS experiments. This method is subject to the same criticisms as the RS with two species. (ii) The performance of each species in a multi-species mixture compared with its performance in monoculture is a form of multi-species AD (see the WWW archive and Ellenberg 1954; Mueller-Dombois & Sims 1966; Pickett & Bazzaz 1978; Austin 1982). (iii) Rejmánek et al. (1989) applied reciprocal yield regression models to a three-species complete additive experiment but the results did not support the interpretations drawn from previous two-species investigations of the species. (iv) Plants can be grown in hexagonal arrays (see above) in which a species has each of the different species under investigation as a neighbour, but never itself. Thórhallsdóttir (1990) and Turkington (1994) used this approach to investigate the role of interspecific interactions on the spatial dynamics of grasses. While elegant, problems with this design include those mentioned above for hexagonal designs and others discussed in Thórhallsdóttir (1990). (v) Ramseier et al. (1996) proposed a simplex design (Cornell 1990) for multi-species experiments in which all species appear in each of a number of mixtures (the minimum number of mixtures is the number of species + 1) but in different relative frequencies, each species in turn being the largest component of a sown mixture with the other species being equally represented, with an additional mixture having all species equally represented. Repeated at a number of densities and with initial sizes of species measured, this design allows a response surface analysis in which questions of outcome and effects of species on each other may be assessed. Additional design points may be added and the order of interaction terms that can be assessed in the model depends on the structure and number of design points. Advantages claimed for the particular simplex design used are that each mixture is an experimental community with all species represented, that the spread in community type allows the examination of interspecific interaction over a wide range of systems, that resource use is efficient in that there are no resources devoted to monocultures, and that it can be readily extended to larger numbers of species in a coherent manner without a major increase in experimental size. Disadvantages are the possible complexity of a full statistical treatment. The problems raised earlier with other designs must be borne in mind when using any of these multi-species approaches.

Some other issues

BACKGROUND SPECIES

Interspecific interactions are sometimes examined by establishing a spacing gradient grid of one species and overseeding the entire grid with a second species. This attempts to assess the effects of varying intensities of intraspecific interaction under the constant influence of a 'background' (Radosevich 1987), although the idea of a constant influence on all species may be illusory. Such an approach ignores the reciprocal nature of many interspecific interactions, such that the introduced individuals generally influence the background species as well as being influenced by it. Over time the background will tend to respond differentially to different species and so what started as a common influence may rapidly cease to be so. This will occur at a localized level in the vicinity of the introduced individuals. Further, an individual or a unit of initial biomass will tend to have less effect at high compared with low density. In addition, a given density of a background species will have a smaller per unit effect on a fixed density of large rather than small introduced individuals of other species, since the overall effective density with large introduced individuals is greater than with small introduced individuals and so like is not being compared with like.

COMPETITIVE HIERARCHIES

Results from using several types of design (i.e. AD, target-neighbour, RS and diallel designs) have been prominent in the search for competitive hierarchies in which a species will out-compete (in the sense of outcome of competition) all species ranked below it in the hierarchy and be out-competed by those above it (Keddy & Shipley 1989; Shipley 1993). The occurrence of reversals of rank order, indicating either a network of competitive performance (intransitivity) (Herben & Krahulec 1990; Silvertown & Dale 1991; Shipley 1993) or competitive combining ability, is controversial (Taylor & Aarssen 1990). However, these designs as usually analysed are prone to the misinterpretations and dangers of size bias (Silvertown & Dale 1991; Grace et al. 1993; Connolly 1997; J. Connolly et al., unpublished data), with factors

© 1999 British Ecological Society, *Journal of Ecology*, **87**, 1–16 such as differences in seed size or initial seedling mass providing a mechanism for that bias (although see Shipley & Keddy 1994).

Discussion/recommendations/conclusions: choosing the appropriate design

Greenhouse studies of plant interactions offer a number of practical advantages over field-based experiments, such as better control of treatments and extrinsic factors, that persuade us that they will have continued utility. Ideally, greenhouse studies should be carried out in conjunction with a field-based programme, and prior knowledge of how interactions take place in the field (e.g. densities, size differences, phenology, asymmetric effects) is necessary before planning an experiment.

Consideration of the six points proposed as a framework for this review lead us to the conclusion that experiments on interspecific interaction demand clarity in respect of the particular facet(s) of interspecific interaction that is the focus of the experiment (unambiguous terminology with precise data-based estimation of measures of those specific aspects of interaction), appropriate experimental design, measurement of appropriate variables and a correct analysis. Our survey indicated an insufficient appreciation of how these factors limit our ability to explore particular questions. While there is not yet available, in our view, a coherent approach to the difficulties posed by the study of competition, a better appreciation of some of the strengths and limitations in these areas is essential.

Inappropriate and inadequate experimental design and procedure in many studies have probably compromised our understanding of plant interactions. For example, the conclusions drawn from many RS experiments, especially those conducted at a single total density and/or based only on final yield, are unlikely to provide many meaningful ecological insights. Experiments where inappropriately sized individuals are matched against each other are similarly compromised. Although not a design issue per se, the confounding of terminology by investigators (e.g. definitions of competition vs. interference, outcome vs. effects, intensity vs. importance; Weldon & Slauson 1986) makes interpretation of experiments difficult. As an example we cite the distinction between the 'outcome of competition', the ultimate success or failure of species, and 'species' effects on each other' (possibly part of the explanation of the observed outcome) as one which rarely appears to be explicit but which has a direct impact on the design, analysis and interpretation of experiments. Additionally, a clearer realization of the limitations of short-term experiments in providing anything but simple indicators in respect of the outcome of long-term competition is desirable. Although not the focus of our review, it is also likely that inappropriate or incomplete analysis

of experimental data has limited the interpretation even of well-designed experiments (e.g. Watkinson & Freckleton 1997).

The choice of design, the variables measured and the analysis determine what questions can and cannot be answered, and should reflect the primary questions of interest. The major deficiency in this respect appeared to be the lack of recognition that many questions of interest could not be addressed adequately without introducing time as a factor. At a minimum there is the need to separate the effects of initial differences from those of subsequent interactions, which cannot be adequately done where only final harvest yield is available. Analysis based on final harvest yield alone can lead to size-bias in interpreting the results from AD, RS and response surface designs. Even when appropriate data on initial conditions are available, it may not be possible to produce unbiased information on some questions of interest for RS and AD designs (e.g. questions as to the outcome of competition).

We need to be very clear about the role of initial size in an experiment. Experiments can only measure effects from the time of establishment of the experiment to the final harvest time. If size differences between species exist at the start then it seems reasonable that the initial differences should be discounted in measures of performance over the experimental period, otherwise the measures are likely to reflect initial differences in addition to effects that arise during the course of the experiment. This is not to say that species that are initially bigger do not do better competitively on some per unit basis: they may, but the assessment of that should not be confounded with effects that simply reflect different initial sizes per se. Thus, for example, final yield per individual of a species depends both on its initial size and on its average Relative Growth Rate (RGR) through the course of the experiment. Greater initial size may lead to greater RGR due to an increased ability to compete for light, and so the final yield per individual is accordingly enhanced for larger individuals. This additional component of final yield (due to different initial size difference) must not be confounded, as it routinely is, with the mere scaling effect of initial size when comparing individuals or species that differ in initial size.

The effects of initial size differences can be allowed for by a double strategy of (i) using an initial biological measure such as total biomass of each species (Connolly & Wayne 1996) or total leaf area index for each species (e.g. Kropf & Spitters 1991) rather than density in response surface equations, and (ii) by using a per unit initial size measure of species' performance (e.g. RGR in Connolly & Wayne 1996). These approaches attempt to avoid difficulties arising from ignoring initial size differences and the use of density to equate species. They also focus attention explicitly on the influence of initial conditions and on the limited nature of inferences from this type of interaction experiment. Conclusions are valid only for the time during which the experiment was running, since what happened previously is built into the initial conditions and what happens afterwards is speculative.

Experiments based on single mixtures have been undervalued; even without information on initial conditions they can provide a simple, efficient method of addressing questions as to the changing balance between species along gradients of various kinds. When appropriate initial information is available a more powerful interpretation is possible. Single mixture experiments highlight the distinction between the outcome of competition, which can be approached within a single mixture, and interspecific effects, on which they generally provide no information.

While AD experiments, particularly those with target-neighbour designs, may need to be treated with caution if only final yield is available, they may suffice for certain objectives, e.g. to examine yield loss in crop-weed systems. However, the mechanism of this yield loss cannot be adequately addressed without allowing for initial conditions and, perhaps, taking intermediate measurements. For these same reasons, AD are inadequate and potentially misleading for some of the more evolutionary orientated concerns of ecologists. Even when information is available for several time-points comparisons of species as competitors against a range of target species may be compromised. AD do allow comparison of the rank order of treatment effects on some interspecific interactions but the absolute estimation of many interactions is beyond their scope if only final harvest data are used.

Despite much criticism in literature preceding or during the early years of the 10-year period surveyed (Inouye & Schaffer 1981; Jolliffe et al. 1984; Connolly 1986, 1988; Law & Watkinson 1987), we were surprised to find that the RS was still the most popular design. The problems with substitutive designs lead us to concur with Law & Watkinson (1987), Keddy (1989), Connolly (1986, 1988) and Snaydon (1991, 1994) that they should not be used for studies of plant interactions, except in very limited circumstances where it is clear that species are comparable in size at the start of the experiment. Even when RS are run at plant densities that closely match the range of natural abundances observed in the field, the fundamental problem remains that the same density is used for both species in monoculture (or an a priori and arbitrarily chosen x:1 ratio). Even if initial size differences are measured, the method cannot in general be corrected to produce valid results (J. Connolly et al., unpublished data). When several such potentially size-biased studies are used to address an issue such as the existence and strength of competitive hierarchies, the scope for misleading inferences is clear. In the large number of studies on crop-weed interactions, it is surprising that RS experiments have been so widely

© 1999 British Ecological Society, *Journal of Ecology*, **87**, 1–16 used compared to additive experiments (Sackville Hamilton 1994). The latter seem highly appropriate for crop loss studies since they take the form of a constant focal species (crop) density and varying associate species (weed) density (e.g. Thompson *et al.* 1994).

Response surface designs are widely seen as a generalization of AD and RS designs and as a remedy for their deficiencies. Even there, however, if only final harvest data are available the range of inferences is limited and the competition coefficients (Law & Watkinson 1987) or substitution rates (Connolly 1987) may include effects of initial differences between species as well as reflecting species' effects on each other.

There are many statistical issues beyond the scope of this review that need to be addressed in competition experiments, e.g. correlated responses, optimal design, estimation of response models and indices, but the first priority must be to ensure that the design, measurements, analyses and indices used lead to valid inferences. To make them more efficient is a secondary concern. Simple procedures have their attraction but more complex designs are necessary to address some questions. Such experimentation may have the advantage of a much wider scope for inference across a broader range of conditions.

We have taken care in this paper to focus rather narrowly on what we perceive to be some major difficulties with experimental procedures, and have not ventured into the deeper waters of deciding between competing theories (e.g. those of Grime 1979 and Tilman 1987) or the details of definition of subtle aspects of interaction, such as the distinction between the importance and intensity of competition (Weldon & Slauson 1986). We believe that a fuller appreciation of the way in which the design/variables/analysis complex determines the range of valid inferences must precede attempts to use experiments to support such theoretical positions or make such distinctions. We are led to this position by considering the confusion in the current literature, exemplified by the way in which possible size bias in RS and AD methods (Keddy & Shipley 1989; Herben & Krahulec 1990; Silvertown & Dale 1991; Grace et al. 1993; Shipley & Keddy 1994; Connolly 1997) has clouded the debate on competitive hierarchies.

We feel that perceptions of controversy and methodological turmoil have inhibited work on interspecific interactions, which is why this paper has concentrated on methodological difficulties rather than general prescriptions. The state of agreement is still not so advanced that we can move beyond the partial prescriptions of the previous few paragraphs, but at least some of the pitfalls are signposted. Once the issues are clarified the potential of these experiments on plant–plant interactions to provide reliable information will be released.

© 1999 British Ecological Society, *Journal of Ecology*, **87**, 1–16

More sophisticated designs that incorporate time and allow response surface analyses with various biotic and abiotic explanatory variables and, perhaps, deal with several species in multi-species mixtures at individual and stand level, are going to be the most informative. Inclusion of root and shoot variables in models of species' performance should help elucidate their joint role in interspecific interaction and determine which facets of species, their growth, architecture, ontogenetic stage, limiting factors, etc., are most important in interspecific relations.

Multiple species investigations can be carried out through multiple pair-wise experiments (Goldberg & Scheiner 1993) or using multi-species designs (e.g. Ramseier *et al.* 1996). While the latter can be very rich in interspecific information they also carry more analytical and interpretative complexity and it may be too early to determine in what circumstances each is to be preferred.

Of course, increasing the complexity, both temporally and spatially, of experimental designs increases the logistical problems of carrying out the experiment. Simple experimental designs are preferable when they can validly address the questions of interest without an unacceptable sacrifice of realism. The caveat is that the results of such an investigation should be followed up by more sophisticated work, ideally including field experiments (e.g. Gibson & Skeel 1996; Skeel & Gibson 1998), before conclusions regarding the performance of plants in natural settings can be made with confidence.

Acknowledgements

Roy Snaydon provided constructive comments on a draft of the manuscript. Even though he did not necessarily agree with all of the conclusions that we reached, we are grateful for the frank expression of his views. We also acknowledge the helpful comments of several anonymous referees and of Roy Turkington. Partial financial support to D. J. Gibson and D. C. Hartnett were provided by NSF grants DEB-9317976 and BSR-9020426 to Kansas State University, and to J. D. Weidenhamer by a Cottrell College Science Award of Research Corporation. Some of this work was carried out while J. Connolly was funded by a Bullard fellowship and a Forbairt International Collaboration Programme grant at Harvard University, where helpful contributions from Peter Wayne and Fakhri Bazzaz are acknowledged.

References

- Aarssen, L.W. (1983) Ecological combining ability and competitive ability in plants: toward a general evolutionary theory of coexistence in systems of competition. *American Naturalist*, **122**, 707–731.
- Aarssen, L.W. (1988) 'Pecking order' of four plant species from pastures of different ages. *Oikos*, **51**, 3–12.
- Aarssen, L.W. & Epp, G.A. (1990) Neighbor manipulations in natural vegetation: a review. *Journal of Vegetation Science*, 1, 13–30.

Greenhouse interaction studies

- Allen, E.B. & Allen, M.F. (1984) Competition between plants of different successional stages: mycorrhizae as regulators. *Canadian Journal of Botany*, 62, 2625–2629.
- Antonovics, J. & Fowler, N.L. (1985) Analysis of frequency and density effects on growth in mixtures of *Salvia splendens* and *Linum grandiflorum* using hexagonal fan designs. *Journal of Ecology*, **73**, 219–234.
- Aplet, G.H. & Laven, R.D. (1993) Relative performance of four Hawaiian shrubby plants (Asteraceae) under greenhouse conditions with implications for rarity. *Biological Conservation*, **65**, 15–21.
- Austin, M.P. (1982) Use of a relative physiological performance value in the prediction of performance in multispecies mixtures from monoculture performance. *Journal of Ecology*, **70**, 559–570.
- Austin, M.P., Fresco, L.F.M., Nicholls, A.O., Groves, R.H. & Kaye, P.E. (1988) Competition and relative yield: estimation and interpretation at different densities and under various nutrient concentrations using *Silybum marianum* and *Cirsium vulgare. Journal of Ecology*, **76**, 157–171.
- Austin, M.P., Groves, R.H., Fresco, L.M.F. & Kaye, P.E. (1985) Relative growth of six thistle species along a nutrient gradient with multispecies competition. *Journal* of Ecology, **73**, 667–684.
- Benner, B.L. & Bazzaz, F.A. (1987) Effects of timing of nutrient addition on competition within and between two annual plant species. *Journal of Ecology*, **75**, 229– 245.
- Berendse, F., Elberse, W.Th. & Geerts, R.H.M.E. (1992) Competition and nitrogen loss from plants in grassland ecosystems. *Ecology*, **73**, 46–53.
- Bi, H. & Turvey, N.D. (1994) Inter-specific competition between seedlings of *Pinus radiata*, *Eucalyptus regnans* and *Acacia melanoxylon*. *Australian Journal of Botany*, 42, 61–70.
- Bleasdale, J.K.A. (1967) Systematic designs for spacing experiments. *Experimental Agriculture*, 3, 73–85.
- Boffey, T.B. & Veevers, A. (1977) Balanced designs for two component competition experiments. *Euphytica*, 26, 481–484.
- Bullock, J.M., Mortimer, A.M. & Begon, M. (1995) Carryover effects on interclonal competition in the grass *Holcus lanatus*: a response surface analysis. *Oikos*, 72, 411–421.
- Connell, J.H. (1983) On the prevalence and relative importance of interspecific competition: evidence from field experiments. *American Naturalist*, **122**, 661–696.
- Connell, J.H. (1990) Apparent versus 'real' competition in plants. *Perspectives on Plant Competition* (eds J.B. Grace & D. Tilman), pp. 9–26. Academic Press, New York, NY.
- Connolly, J. (1986) On difficulties with replacement-series methodology in mixture experiments. *Journal of Applied Ecology*, 23, 125–137.
- Connolly, J. (1987) On the use of response models in mixture experiments. *Oecologia*, **72**, 95–103.
- Connolly, J. (1988) What is wrong with replacement series? *Trends in Ecology and Evolution*, **3**, 24–26.
- Connolly, J. (1997) Substitutive experiments and the evidence for competitive hierarchies in plant communities. *Oikos*, **80**, 179–182.
- Connolly, J. & Wayne, P. (1996) Asymmetric competition between plant species. *Oecologia (Berlin)*, **108**, 311–320.
- Connolly, J., Wayne, P. & Murray, R. (1990) Time course of plant-plant interactions in experimental mixtures of annuals: density, frequency, and nutrient effects. *Oecologia*, 82, 513–526.

© 1999 British Ecological Society, *Journal of Ecology*, **87**, 1–16

Cornell, J.A. (1990) *Experiments with Mixtures*, 2nd edn. John Wiley & Sons, New York, NY.

Cousens, R. (1996) Design and interpretation of interference

studies: are some methods totally unacceptable? *New Zealand Journal of Forestry Science*, **26**, 5–18.

- Cousens, R. & O'Neill, M. (1993) Density dependence of replacement series experiments. *Oikos*, 66, 347–352.
- Dekker, J.H., Meggitt, W.F. & Putnam, A.R. (1983) Experimental methodologies to evaluate allelopathic plant interactions. *Journal of Chemical Ecology*, 9, 945–981.
- Diamond, J.M. (1986) Overview: laboratory experiments, field experiments, and natural experiments. *Community Ecology* (eds J. Diamond & T.J. Case), pp. 3–22. Harper & Row, New York, NY.
- Durrant, A. (1965) Analysis of reciprocal differences in diallel crosses. *Heredity*, **20**, 573–607.
- Ellenberg, H. (1954) Ueber einige Fortschritte der kausalen Vegetationskunde. *Vegetatio*, **5**/**6**, 199–211.
- Firbank, L.G. & Watkinson, A.R. (1985) On the analysis of competition within two-species mixtures of plants. *Journal of Applied Ecology*, 22, 503–517.
- Firbank, L.G. & Watkinson, A.R. (1987) On the analysis of competition at the level of the individual plant. *Oecologia*, **71**, 308–317.
- Firbank, L.G. & Watkinson, A.R. (1990) On the effects of competition: from monocultures to mixtures. *Per-spectives on Plant Competition* (eds J.B. Grace & D. Tilman), pp. 165–192. Academic Press, San Diego, CA.
- Fone, A.L. (1989) Competition in mixtures of the annual *Hypochoeris glabra* and perennial *H. radicata. Journal* of Ecology, 77, 484–494.
- Fowler, N. (1982) Competition and coexistence in a north Carolina grassland. III. Mixtures of component species. *Journal of Ecology*, **70**, 77–92.
- Frankow-Lindberg, B.E. (1985) Effect of couch grass and grass cultivars on competition between timothy and red clover. *Journal of Applied Ecology*, **22**, 519–524.
- Fredshavn, J.R. (1994) Use of substitution rates to describe competition in mixed plant populations. Acta Agriculturæ Scandinavica, B44, 47–54.
- Gaudet, C.L. & Keddy, P.A. (1988) A comparative approach to predicting competitive ability from plant traits. *Nature*, **334**, 242–243.
- Gibson, D.J. & Skeel, V.A. (1996) Effects of competition on photosynthetic rate and stomatal conductance of *Sorghastrum nutans. Photosynthetica*, **32**, 503–512.
- Gleeson, A.C. & McGilchrist, C.A. (1980) Mixture diallel experiments with unequal proportions of genotypes. *Journal of Agricultural Science, Cambridge*, 95, 525–532.
- Goldberg, D.E. (1990) Components of resource competition in plant communities. *Perspectives on Plant Competition* (eds J.B. Grace & D. Tilman), pp. 27–49. Academic Press, San Diego, CA.
- Goldberg, D.E. & Barton, D.E. (1992) Patterns and consequences of interspecific competition in natural communities: a review of field experiments with plants. *American Naturalist*, **139**, 771–801.
- Goldberg, D.E. & Fleetwood, L. (1987) Competitive effect and response in four annual plants. *Journal of Ecology*, 75, 1131–1143.
- Goldberg, D.E. & Landa, K. (1991) Competitive effect and response: hierarchies and correlated traits in the early stages of competition. *Journal of Ecology*, **79**, 1013– 1030.
- Goldberg, D.E. & Scheiner, S.M. (1993) ANOVA and ANCOVA: field competition experiments. *Design and Analysis of Ecological Experiments* (eds S.S. Scheiner & J. Gurevitch), pp. 69–93. Chapman & Hall, New York, NY.
- Goldberg, D.E. & Werner, P.A. (1983) Equivalence of competitors in plant communities: a null hypothesis and an experimental approach. *American Journal of Botany*, **70**, 1098–1104.

14

- Grace, J.B. & Tilman, D. (1990) Perspectives on Plant Competition. Academic Press, New York, NY.
- Grace, J.B., Guntenspergen, G. & Keough, J. (1993) The examination of a competition matrix for transitivity and intransitive loops. *Oikos*, **68**, 91–98.
- Grace, J.B., Keough, J. & Guntenspergen, G. (1992) Size bias in traditional analyses of substitutive competition experiments. *Oecologia*, **90**, 429–434.
- Grime, J.P. (1979) *Plant Strategies and Vegetation Processes*. John Wiley & Sons, Chichester, UK.
- Grime, J.P., Mackey, J.M.L., Hillier, S.H. & Read, D.J. (1987) Floristic diversity in a model system using experimental microcosms. *Nature*, **328**, 420–422.
- Gurevitch, J., Wilson, P., Stone, J.L., Teese, P. & Stoutenburgh, R.J. (1990) Competition among old-field perennials at different levels of soil fertility and available space. *Journal of Ecology*, **78**, 727–744.
- Hairston, N.G. (1989) *Ecological Experiments: Purpose*, *Design, and Execution.* Cambridge University Press, Cambridge, UK.
- Harper, J.L. (1977) *Population Biology of Plants*. Academic Press, London, UK.
- Harper, J.L. (1983) After description. *The Plant Community* as a Working Mechanism (ed. E.I. Newman), pp. 11–26. Blackwell Scientific Publications, Oxford, UK.
- Hartnett, D.C., Hetrick, B.A.D., Wilson, G.W.T. & Gibson, D.J. (1993) Mycorrhizal influence on intra- and interspecific neighbour interactions among co-occurring prairie grasses. *Journal of Ecology*, **81**, 787–795.
- Herben, T. & Krahulec, F. (1990) Competitive hierarchies, reversals of rank order and the de Wit approach: are they compatible? *Oikos*, 58, 254–256.
- Hetrick, B.A.D., Hartnett, D.C., Wilson, G.W.T. & Gibson, D.J. (1994) Effects of mycorrhizae, phosphorus availability, and plant density on yield relationships among competing tallgrass prairie grasses. *Canadian Journal of Botany*, **72**, 168–176.
- Inouye, R.S. & Schaffer, W.M. (1981) On the ecological meaning of ratio (de Wit) diagrams in plant ecology. *Ecology*, 62, 1679–1681.
- Jolliffe, P.A., Minjas, A.N. & Runeckles, V.C. (1984) A reinterpretation of yield relationships in replacement series experiments. *Journal of Applied Ecology*, 21, 227– 243.
- Keddy, P.A. (1989) *Competition*. Chapman & Hall, New York, NY.
- Keddy, P.A. & Shipley, B. (1989) Competitive hierarchies in herbaceous plant communities. *Oikos*, 54, 234–241.
- Kropf, M.J. & Spitters, C.J.T. (1991) A simple model of crop loss by weed competition from early observations on relative leaf area of the weeds. *Weed Research*, 31, 97–105.
- Law, R. & Watkinson, A.R. (1987) Response–surface analysis of two-species competition: an experiment on *Phleum* arenarium and *Vulpia fasciculata*. Journal of Ecology, 75, 871–886.
- MacArthur, R.H. (1972) *Geographical Ecology*. Harper & Row, London, UK.
- McGilchrist, C.A. & Trenbath, B.R. (1971) A revised analysis of plant competition experiments. *Biometrics*, 27, 659–671.
- Mead, R. (1979) Competition Experiments. Biometrics, 35, 41–54.
- Menchaca, L. & Connolly, J. (1990) Species interference in white clover–ryegrass mixtures. *Journal of Ecology*, 78, 223–232.
- Mitchley, J. (1987) Diffuse competition in plant communities. Trends in Ecology and Evolution, 2, 104–106.

© 1999 British Ecological Society, *Journal of Ecology*, **87**, 1–16

Mueller-Dombois, D. & Sims, H.P. (1966) Response of three grasses to two soils and a water table depth gradient. *Ecology*, 47, 644–648.

- Nelder, J.A. (1962) New kinds of systematic designs for spacing experiments. *Biometrics*, 18, 283–307.
- Nelder, J.A. & Wedderburn, R.W.M. (1972) Generalised linear models. *Journal of the Royal Statistical Society A*, 135, 370–384.
- Norrington-Davies, J. & Hutto, J.M. (1972) Diallel analysis of competition between diploid and tetraploid genotypes of *Secale cereale* grown at two densities. *Journal of Agricultural Science, Cambridge*, **78**, 251–256.
- Ogg, A.G., Stephens, R.H. & Gealy, D.R. (1993) Growth analysis of mayweed chamomile (*Anthemis cotula*) interference in peas (*Pisum sativum*). Weed Science, **41**, 394– 402.
- Pickett, S.T.A. & Bazzaz, F.A. (1978) Organization of an assemblage of early successional species on a soil moisture gradient. *Ecology*, **59**, 1248–1255.
- Rabinowitz, D., Rapp, J.K. & Dixon, P.M. (1984) Competitive abilities of sparse grass species: means of persistence or cause of abundance. *Ecology*, 65, 1144–1154.
- Radosevich, S.R. (1987) Methods to study interactions among crops and weeds. Weed Technology, 1, 190–198.
- Ramseier, D., Connolly, J. & Bazzaz, F.A. (1996) Influence of elevated CO₂ on competition among 5 annuals. *Bulletin of the Ecological Society of America*, 77 (Suppl.) (Program and Abstracts: Part 2), 367.
- Rejmánek, M., Robinson, G.R. & Rejmánková, E. (1989) Weed–crop competition: experimental designs and models for data analysis. *Weed Science*, **37**, 276–284.
- Rodriguez, D.J. (1997) A method to study competition dynamics using de Wit replacement series experiments. *Oikos*, **78**, 411–415.
- Roush, M.L., Radosevich, S.R., Wagner, R.G., Maxwell, B.D. & Peterson, T.D. (1989) A comparison of methods for measuring effects of density and proportion in plant competition experiments. *Weed Science*, **37**, 268–275.
- Rousvoal, D. & Gallais, A. (1973) Comportement en association binaire de cinq espèces d'une prairie permanente. *Œcologia Plantarum*, 8, 279–300.
- Sackville Hamilton, N.R. (1994) Replacement and additive designs for plant competition studies. *Journal of Applied Ecology*, **31**, 599–603.
- Schmid, B. & Harper, J.L. (1985) Clonal growth in grassland perennials. *Journal of Ecology*, 73, 793–808.
- Schoener, T.W. (1983) Field experiments on interspecific competition. *American Naturalist*, **122**, 240–285.
- Shipley, B. (1993) A null model for competitive hierarchies in competition matrices. *Ecology*, 74, 1693–1699.
- Shipley, B. & Keddy, P.A. (1994) Evaluating the evidence for competitive hierarchies in plant communities. *Oikos*, 69, 340–345.
- Shrefler, J.W., Shilling, D.G., Dusky, J.A. & Brecke, B.J. (1994) Influence of phosphorus fertility on intra- and interspecific interference between lettuce (*Lactuca* sativa) and spiny amaranth (*Amaranthus spinosus*). Weed Science, 42, 574–578.
- Silvertown, J. & Dale, P. (1991) Competitive hierarchies and the structure of herbaceous plant communities. *Oikos*, 61, 441–444.
- Silvertown, J. & Lovett Doust, J. (1993) Introduction to Plant Population Biology. Blackwell Scientific Publications, Oxford, UK.
- Skeel, V.A. & Gibson, D.J. (1998) Photosynthetic rates and vegetative production of *Sorghastrum nutans* in response to competition at two strip mines and a railroad prairie. *Photosynthetica*, **35**, 139–149.
- Snaydon, R.W. (1991) Replacement or additive designs for competition studies? *Journal of Applied Ecology*, 28, 930–946.
- Snaydon, R.W. (1994) Replacement or additive designs revisited: comments on the review paper by N. R. Sackville Hamilton. *Journal of Applied Ecology*, 31, 784–786.

Greenhouse interaction studies

- Solbrig, O.T., Curtis, W.F., Kincaid, D.T. & Newell, S.J. (1988) Studies on the population biology of the genus Viola. VI. The demography of V. fimbriatula and V. lanceolata. Journal of Ecology, 76, 301-319.
- Spitters, C.J.T. (1983) An alternative approach to the analysis of mixed cropping experiments. I. Estimation of competition effects. Netherlands Journal of Agricultural Science, 31, 1–11.
- Stiling, P.D. (1992) Introductory Ecology. Prentice Hall, Englewood Cliffs, NJ.
- Strong, D.R., Simberloff, D., Abele, L. & Thistle, A.B. (1984) Ecological Communities: Conceptual Issues and the Evidence. Princeton University Press, Princeton, NJ.
- Suehiro, K. & Ogawa, H. (1980) Competition between two annual herbs, Atriplex gmelini C. A. Mey and Chenopodium album L., in mixed cultures irrigated with seawater of varying concentrations. Oecologia (Berlin), 45, 167 - 177.
- Taylor, D.R. & Aarssen, L.W. (1989) On the density dependence of replacement-series competition experiments. Journal of Ecology, 77, 975–988.
- Taylor, D.R. & Aarssen, L.W. (1990) Complex competitive relationships among genotypes of three perennial grasses: implications for species coexistence. American Naturalist, 136, 305-327.
- Thijs, H., Shann, J.R. & Weidenhamer, J.D. (1994) The effect of phytotoxins on competitive outcome in a model system. Ecology, 75, 1959-1964.
- Thompson, C.R., Thill, D.C. & Shafii, B. (1994) Growth and competitiveness of sulfonylurea-resistant and -susceptible Kochia (Kochia scoparia). Weed Science, 42, 172-179.
- Thórhallsdóttir, T.E. (1990) The dynamics of five grasses and white clover in a simulated mosaic sward. Journal of Ecology, 78, 909-923.
- Tilman, D. (1987) The importance of mechanisms of interspecific competition. American Naturalist, 129, 769-774.
- Tilman, D. (1988) Dynamics and Structure of Plant Communities. Princeton University Press, Princeton, NJ.
- Trenbath, B.R. (1978) Models and the interpretation of mixture experiments. Plant Relations in Pastures (ed. J.R. Wilson), pp. 145-162. CSIRO, East Melbourne, Australia
- Turkington, R. (1994) Effect of propagule source on com-

petitive ability of pasture grasses: spatial dynamics of six grasses in simulated swards. Canadian Journal of Botany, 72, 111-121.

- Turkington, R. & Jolliffe, P.A. (1996) Interference in Trifolium repens-Lolium perenne mixtures: short- and longterm relationships. Journal of Ecology, 84, 563-571.
- Underwood, T. (1986) The analysis of competition by field experiments. Community Ecology: Pattern and Process (ed. J. Kikkawa & D.J. Anderson), pp. 240-268. Blackwell Scientific Publications, Oxford, UK.
- Wall, D.A. (1993) Comparison of green foxtail (Setaria viridis) and wild oat (Avena fatua) growth, development, and competitiveness under three temperature regimes. Weed Science, 41, 369-378.
- Watkinson, A.R. & Freckleton, R.P. (1997) Quantifying the impact of arbuscular mycorrhiza on plant competition. Journal of Ecology, 85, 541-545.
- Weidenhamer, J.D. (1996) Distinguishing resource competition and chemical interference: overcoming the methodological impasse. Agronomy Journal, 88, 866-875.
- Weidenhamer, J.D., Hartnett, D.C. & Romeo, J.T. (1989) Density-dependent phytotoxicity: distinguishing resource competition and allelopathic interference in plants. Journal of Applied Ecology, 26, 613-624.
- Weldon, C.W. & Slauson, W.L. (1986) The intensity of competition versus its importance: an overlooked distinction and some implications. The Quarterly Review of Biology, 61 23-44
- Willey, R.W. & Rao, M.R. (1980) A competitive ratio for quantifying competition between intercrops. Experimental Agriculture, 16, 117-125.
- de Wit, C.T. (1960) On competition. Verslagen Van Landouskundige Onderzoekingen, 66, 1–82.
- de Wit, C.T. & Van den Bergh, J.P. (1965) Competition between herbage plants. Netherlands Journal of Agricultural Science, 13, 212-221.
- de Wit, C.T., Tow, P.G. & Ennik, G.C. (1966) Competition between legumes and grasses. Verslagen Van Landouskundige Onderzoekingen, 68, 3-30.
- Zimdahl, R.L. (1980) Weed Crop Competition: A Review. International Plant Protection Center, Corvalis, OR.

Received 6 October 1997 revision accepted 2 June 1998

© 1999 British Ecological Society, Journal of Ecology, 87, 1-16