2

Experimental design

Multivariate methods are no longer restricted to the exploration of data and to the generation of new hypotheses. In particular, constrained ordination is a powerful tool for analysing data from manipulative experiments. In this chapter, we review the basic types of experimental design, with an emphasis on manipulative field experiments. Generally, we expect that the aim of the experiment is to compare the response of studied objects (e.g. an ecological community) to several treatments (treatment levels). Note that one of the treatment levels is usually a control treatment (although in real ecological studies, it might be difficult to decide what is the control; for example, when we compare several types of grassland management, which of the management types is the control one?). Detailed treatment of the topics handled in this chapter can be found for example in Underwood (1997).

If the response is univariate (e.g. number of species, total biomass), then the most common analytical tools are ANOVA, general linear models (which include both ANOVA, linear regression and their combinations), or generalized linear models. Generalized linear models are an extension of general linear models for the cases where the distribution of the response variable cannot be approximated by the normal distribution.

2.1. Completely randomized design

The simplest design is the completely randomized one (Figure 2-1). We first select the plots, and then randomly assign treatment levels to individual plots. This design is correct, but not always the best, as it does not control for environmental heterogeneity. This heterogeneity is always present as an unexplained variability. If the heterogeneity is large, use of this design might decrease the power of the tests.

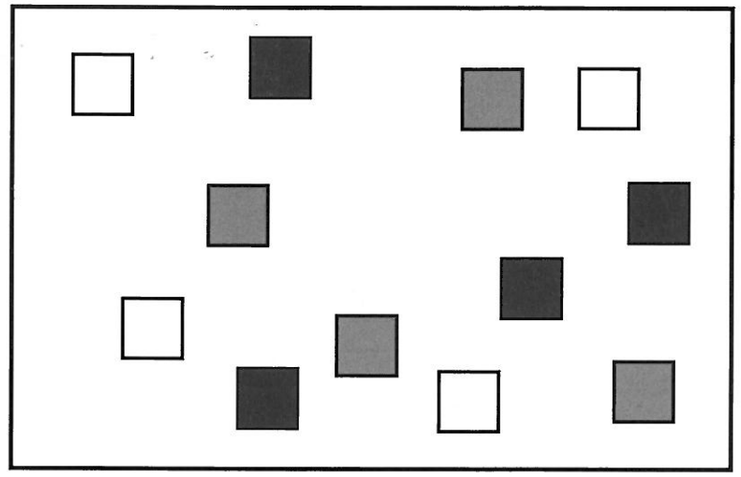

Figure 2-1. Completely randomized design, with three treatment levels and four replicates (or replications).

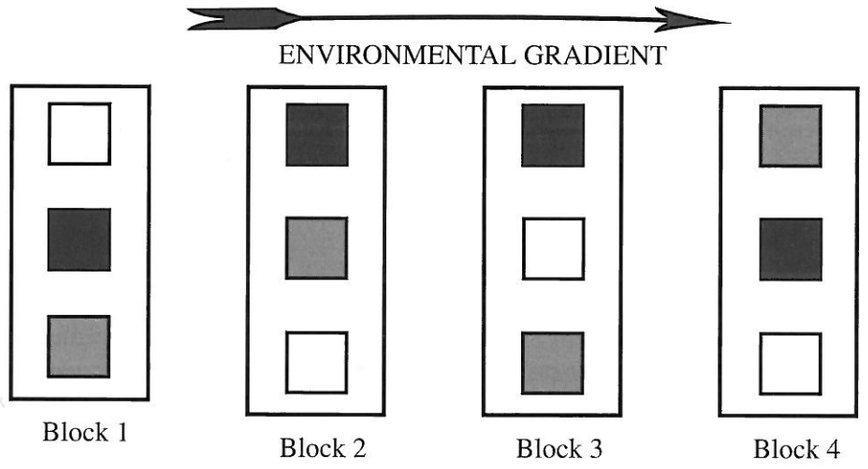

Figure 2-2. The randomized complete blocks design.

2.2. Randomized complete blocks

There are several ways to control for environmental heterogeneity. Probably the most popular one in ecology is the randomized complete blocks design. Here, we first select the blocks so that they are internally as homogeneous as possible (e.g. rectangles with the longer side perpendicular to the environmental gradient, Figure 2-2). The number of blocks is equal to the number

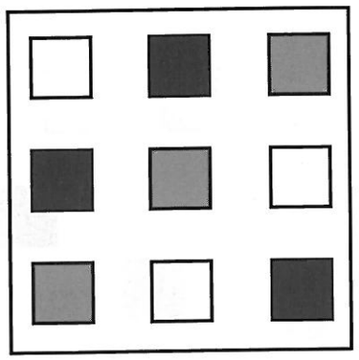

Figure 2-3. Latin square design.

of replications. Each block contains just one plot for each treatment level, and their spatial position within a block is randomized.

If there are differences among the blocks,* this design provides a more powerful test than the completely randomized design. On the other hand, when applied in situations where there are no differences among blocks, the power of the test will be lower in comparison to the completely randomized design, because the number of degrees of freedom is reduced. This is particularly true for designs with a low number of replications and/or a low number of levels of the experimental treatment. There is no consensus among statisticians about when the block structure can be ignored if it appears that it does not explain anything.

2.3. Latin square design

Latin square design (see Figure 2-3) assumes that there are gradients, both in the direction of the rows and the columns of a square. The square is constructed in such a way that each column and each row contains just one of the levels of the treatment. Consequently, the number of replications is equal to the number of treatments. This might be an unwanted restriction. However, more than one Latin square can be used. Latin squares are more popular in agricultural research than in ecology. As with randomized complete blocks, the environmental variability is powerfully controlled. However, when there is no such variability (i.e. that explainable by the columns and rows), then the test is weaker than a test for a completely randomized design, because of the loss of degrees of freedom.

* And often there are: the spatial proximity alone usually implies that the plots within a block are more similar to each other than to the plots from different blocks.

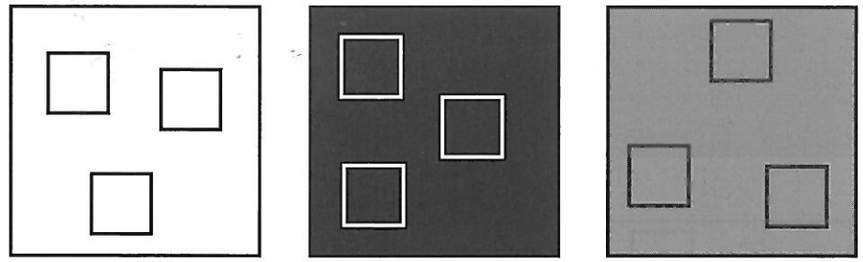

Figure 2-4. Pseudoreplications: the plots are not replicated, but within each plot several subsamples are taken.

2.4. Most frequent errors – pseudoreplications

Pseudoreplications (Figure 2-4) are among the most frequent errors in ecological research (Hurlbert 1984). A possible test, performed on the data collected using such a design, evaluates differences among plot means, not the differences among treatments. In most cases, it can be reasonably expected that the means of contiguous plots are different: just a proximity of subplots within a plot compared with distances between subplots from different main plots suggests that there are some differences between plots, regardless of the treatment. Consequently, the significant result of the statistical test does not prove (in a statistical sense) that the treatment has any effect on the measured response.

2.5. Combining more than one factor

Often, we want to test several factors (treatments) in a single experiment. For example, in ecological research, we might want to test the effects of fertilization and mowing.

Factorial designs

The most common way to combine two factors is through factorial design. This means that each level of one factor is combined with each level of the second factor. If we consider fertilization (with two levels, fertilized and non-fertilized) and mowing (mown and non-mown), we get four possible combinations. Those four combinations have to be distributed in space either randomly, or they can be arranged in randomized complete blocks or in a Latin square design.

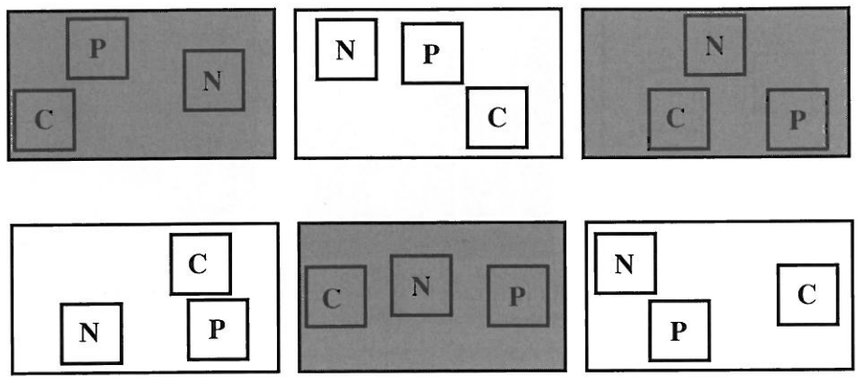

Figure 2-5. The split-plot design. In this example, the effect of fertilization was studied on six plots, three of them on limestone (shaded) and three of them on granite (empty). The following treatments were established in each plot: control (C), fertilized by nitrogen (N), and fertilized by phosphorus (P).

Hierarchical designs

In hierarchical designs, each main plot contains several subplots. For example, we can study the effect of fertilization on soil organisms. For practical reasons (edge effect), the plots should have, say, a minimum size of 5 m × 5 m. This clearly limits the number of replications, given the space available for the experiment. Nevertheless, the soil organisms are sampled using soil cores of diameter 3 cm. Common sense suggests that more than one core can be taken from each of the basic plots. This is correct. However, the individual cores are not independent observations. Here we have one more level of variability – the plots (in ANOVA terminology, the plot is a random factor).

The plots are said to be nested within the treatment levels (sometimes, instead of hierarchical designs, they are called **nested design**). In such a design, a treatment effect is generally tested against the variability on the nearest lower hierarchical level. In the example above, the effect of fertilization is tested against the variability among plots (within a fertilization treatment level), **not** against the variability among soil cores. By taking more soil cores from a plot we do not increase the error degrees of freedom; however, we decrease the variability among the plots within the treatment level, and increase the power of the test in this way.

Sometimes, this design is also called a **split-plot design**. More often, the split-plot design is reserved for another hierarchical design, where two (or more) experimental factors are combined as shown in Figure 2-5.

In this case, we have the **main plots** (with three replications for each level) and the **split-plots**. The main plots are alternatively called whole plots.

The effect of the bedrock must be tested against the variability among the main plots within the bedrock type. Note that this design is different and consequently has to be analysed differently from the factorial design with two factors, bedrock and fertilization – there is one additional level of variability, that of the plots, which is nested within the bedrock type.

2.6. Following the development of objects in time – repeated observations

The biological/ecological objects (whether individual organisms or communities studied as permanent plots) develop in time. We are usually interested not only in their static state, but also in the dynamics of their development, and we can investigate the effect of experimental manipulations on their dynamics. In all cases, the inference is much stronger if the **baseline data** (the data recorded on the experimental objects before the manipulation was imposed) are available. In this case, we can apply a BACI (before after control impact) design (Green 1979).

In designed manipulative experiments, the objects should be part of a correct statistical design; we then speak about replicated BACI designs.

As has been pointed out by Hurlbert (1984), the non-replicated BACI is not a statistically correct design, but it is often the best possibility in environmental impact studies where the 'experiment' is not designed in order to enable testing of the effect. An example is given in Figure 2-6. In this case, we want to assess the effect of a newly built factory on the quality of water in a river (and, consequently, we do not have replicated sites). We can reasonably expect that upstream of the factory, the water should not be affected, so we need a control site above the factory and an impact site below the factory (nevertheless, there might be differences in water quality along the river course). We also need to know the state before the factory starts operating and after the start (but we should be aware that there might be temporal changes independent of the factory's presence). We then consider the changes that happen on the impact site but not on the control site* to be a proof of the factory's impact.

Nevertheless, the observations within a single cell (e.g. on the control site before the impact) are just pseudoreplications. Even so, the test performed in this way is much more reliable than a simple demonstration of either temporal changes on the impact site itself or the differences between the impact and control sites in the time after the impact was imposed.†

* In ANOVA terminology, this is the interaction between time and the site factors.

† The only possibility of having a replicated BACI design here would be to build several factories of the same kind on several rivers, which is clearly not a workable suggestion!

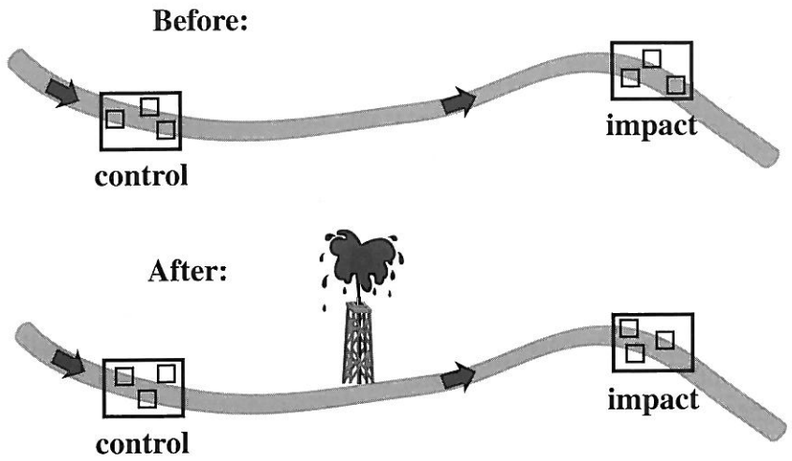

Figure 2-6. The non-replicated BACI (before after control impact) design.

Even if we miss the 'before' situation, we can usually demonstrate a lot by a combination of common sense and some design. For example, in our example case, we can have a series of sites along the river and an abrupt change just between the two sites closest to the factory (one above and one below it), with the differences between the other sites being gradual.* This is often a suggestive demonstration of the factory effect (see Reckhow 1990 for additional discussion).

Another possibility is to repeat the sampling over time and use time points as (possibly correlated) replication (Stewart-Oaten et al. 1986; Ter Braak & Šmilauer 2002).

In designed experiments, the replicated BACI is probably the best solution. In this design, the individual sampling units are arranged in some correct design (e.g. completely randomized or in randomized complete blocks). The first measurement is done before the experimental treatment is imposed (the baseline measurement), and then the development (dynamics) of the objects is followed. In such a design, there should be no differences among the experimental groups in the baseline measurement. Again, the interaction of treatment and time is of greatest interest, and usually provides a stronger test than a comparison of the state at the end of experiment. Nevertheless, even here we have various possible analyses, with differing emphases on the various aspects that we are interested in (see Lindsey 1993). As we control the environmental

* Or similarly, time series from the impacted site, with an abrupt change temporarily coincident with the impact-causing event.

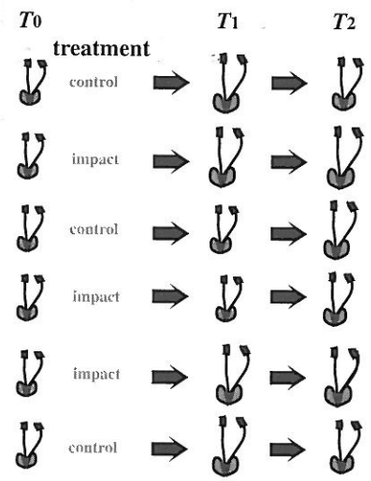

Figure 2-7. Replicated BACI design. The objects (individual plants) are completely randomized and measured at the first time T_0. Thereafter, the impact (fertilization) is imposed and the development is followed (with two sampling times T_1 and T_2 in our example). The null hypothesis we are interested in is: *The growth is the same in the control and impact groups.*

heterogeneity by knowing the initial state, the block structure is usually not as efficient as in the static comparisons. An example is given in Figure 2-7.

One should be aware that, with 5% probability, there will be significant differences between treatments in the baseline data (when tested at $\alpha = 0.05$). It is therefore advisable to test the baseline data before imposing the treatments, and when significant differences are found re-assign the treatments to the units (with the same random process as used originally).

With the continuing development of statistical methods, the designs are becoming more and more flexible. Often the design of our experiment is a compromise between the statistical requirements and the feasibility of the implied manipulations in the field. The use of pseudoreplication in designed experiments is, nevertheless, one of the most common mistakes and can lead to serious errors in our conclusions. Each researcher is advised to have a clear idea about how the data will be analysed before performing the experiment. This will prevent some disappointment later on.

2.7. Experimental and observational data

Manipulative experiments in ecology are limited both in space and time (Diamond 1986). For larger spatial or temporal scales we have to rely on observational data. Also, each manipulation has inevitably some side-effects

and, consequently, the combination of observational data and manipulative experiments is necessary. Generally, the observational data are more often used in hypotheses generation, whereas the data from manipulative experiments are used for hypotheses testing. However, this need not always be so. In manipulations on the community level, we usually affect a whole chain of causally interconnected variables (e.g. in the fertilization experiment in Case Study 3 (Chapter 13), fertilization affected the cover of the crop, which, together with the direct effect of the nutrients, affected the weed community). In replicated manipulative experiments on a community level in space and time, there are many factors that we are not able to control, but still affect the results (weather in a year, bed rock of a site). When the response is species community composition, we often know the biological traits of some species (which we are not able to manipulate). We primarily test the hypothesis concerning change of species composition, but the results very often suggest new hypotheses on the relationship between species traits and species response to the manipulation. All those situations call for careful exploration of data, even in the analysis of manipulative experiments (Hallgren et al., 1999).

Also in non-manipulative studies, we often collect data to test a specific hypothesis. It is useful to collect data in a design that resembles the manipulative experiment. Diamond (1986) even coined the term 'natural experiment', for the situations in nature that resemble experimental designs and can be analysed accordingly (for example, to compare the abundance of spiders on islands with and without lizards). According to Diamond, such situations are under-utilized: 'While field experimentalists are laboriously manipulating species abundances on tiny plots, analogous ongoing manipulations on a gigantic scale are receiving little attention (e.g. expansion of parasitic cowbirds in North America and the West Indies, decimation of native fish by introduced piscivores, and on-going elimination of American elms by disease).' In non-manipulative situations, however, one can never exclude the possibility of confounding effects – consequently, special attention should be paid to sampling design that minimizes this danger.

3

Basics of gradient analysis

The methods for analysing species composition are usually divided into gradient analysis and classification. The term **gradient analysis** is used here in the broad sense, for any method attempting to relate species composition to the (measured or hypothetical) environmental gradients.

Traditionally, the classification methods, when used in plant community ecology, were connected with the **discontinuum approach** (or vegetation unit approach) or sometimes even with the Clementsian superorganismal approach, whereas the methods of gradient analysis were connected with the **continuum concept** or with the Gleasonian individualistic concept of (plant) communities (Whittaker 1975). While this might reflect the history of the methods, this distinction is no longer valid. The methods are complementary and their choice depends mainly on the purpose of a study. For example, in vegetation mapping some classification is necessary. Even if there are no distinct boundaries between adjacent vegetation types, we have to cut the continuum and create distinct vegetation units for mapping purposes. Ordination methods can help find repeatable vegetation patterns and discontinuities in species composition, and show any transitional types, etc. These methods are now accepted even in phytosociology. Also, the methods are no longer restricted to plant community ecology. They became widespread in most studies of ecological communities with major emphasis on species composition and its relationship with the underlying factors. In fact, it seems to us that the advanced applications of gradient analysis are nowadays found outside the vegetation sciences, for example in hydrobiological studies (see the bibliographies by Birks et al. 1996, 1998).