



Designing an Ecological Study

from G.W. Cox, General Ecology Laboratory Manual

- ☞ What kinds of questions do ecologists study?
- ☞ How can one identify an ecological topic for investigation?
- ☞ What steps do ecologists follow in testing hypotheses about ecological systems?

Ecology can be defined as the study of ecological systems. A *system* is any set of components, living or nonliving, that are tied together by regular interactions. An *ecological system* is made up of one or more organisms, together with the nonliving environment with which they interact. Ecological systems exist at several different levels of organization. An ecological system can be a single organism and its surroundings, a population or set of interacting populations in a certain habitat, or the entire community together with the abiotic environment with which these species interact, a unit termed an *ecosystem*.

Ecologists are interested both in the structure and function of ecological systems. *Structure* refers to measurable conditions of the system at one point in time. These include biotic attributes such as the body mass of an organism, the density of individuals in a population, the ratio of predator numbers to those of a prey species, or the biomass of all species that share some basic feature, such as photosynthesis. Structure also includes the physical and chemical conditions that prevail in the space occupied by organisms.

Function refers to the processes that create structure at a given instant, and how these processes are affected as structure changes. Function includes, for example, the relationships that determine the growth rates of organisms, the rates of survival and reproduction of individuals in populations, the rates of predation by one species on another, and the intensity of other interactions, such as competition and mutualism. These processes determine the distribution and abundance of species, and thus the makeup of communities and ecosystems. Function also includes system-level processes, such as the cycling of nutrients and flow of energy among the components of an ecosystem.

A major goal of modern ecology is to understand how ecological systems function, so that their behavior can be predicted, and so that they can be managed for long-term human benefit. To this end, ecologists seek to be rigorous in their techniques of investigation.

What is the hypothesis-testing method?

Most ecologists use the hypothesis-testing or hypothetico-deductive approach in their studies. This approach, based largely on the ideas of Karl Popper (1968), is the statement of explicit hypotheses about ecological relationships, followed by collection of data that lead to their acceptance or rejection. It views activities that cannot lead to rejection of certain hypotheses and the acceptance of others as wasted efforts, which do not advance ecological knowledge. The need for a rigorous hypothesis-testing approach has been stressed by many ecologists (Quinn and Dunham 1983, Connor and Simberloff 1986). The effort to make ecological study more rigorous in this way is one of the key features of the current phase of ecology (Peters 1991, but also see Schrader-Frechette and McCoy 1993), Anderson et al. 2000.

Where do ecologists get their hypotheses?

The stimulus for almost all ecological research, whether in the field or the laboratory, comes initially from the observation of some distinctive pattern in nature. Usually, an initial observation is of some difference between two or more ecological situations. Usually, as well, this observation is tentative, involving a few organisms, a single population, or conditions at one time. Sometimes ecologists make their own initial observations, but often they base studies on patterns that have been observed and documented by others.

How does an initial observation lead to a comprehensive study?

Beginning with an initial observation, an idealized study consists of two general stages: (1) a descriptive stage, concerned with whether or not a distinctive structural pattern truly exists; and (2) a functional stage, in which the cause or effect of this pattern is explored (figure 1.1).

What does the descriptive phase of study involve?

The initial observation, in effect, is a *hypothesis* that a difference in structure exists. This hypothesis therefore must be tested by collecting and analyzing data to determine if the difference exists with a probability greater than expected by chance. Such a test usually requires the collection of unbiased, quantitative data that can be analyzed statistically. These steps form the descriptive stage of an ecological study, and they determine, with a specified degree of probability, whether or not the postulated structural pattern exists.

What does the functional phase of study involve?

If a distinctive structural pattern does exist, the stage is set for studies of function. These might relate either to the cause of the structural difference or to its consequences for the ecological system. To examine causes or consequences, one must first identify the possible cause-effect relationships that might be operating. Controlled experiments or controlled observations can then be designed to distinguish among these possibilities. These activities also must be designed to furnish quantitative data that can be tested statistically.

How can one make a fruitful initial observation?

In selecting a topic, one of the most difficult steps is identifying a problem that can be carried through to the functional stage. Although intuition and experience are valuable, a systematic approach to observations can be helpful. Deliberate comparison of the different components of a single ecological system (such as the species that coexist in the same habitat) or of one component in different systems (such as the same species in different habitats) is one way to do this. Such comparisons will almost always reveal differences, some of which may provide the basis for a productive study. When such comparisons are being made, one should think about the kind of sampling and analysis that will be needed to determine if an apparent difference is real, and about the sorts of testable functional hypotheses that might be made if a structural difference is shown to exist.

What kinds of ecological systems should be compared?

In making comparisons one should remember that good problems need not only deal with single species of plants or animals, but also can concern interacting populations of different species, or even relationships at the ecosystem level. Although patterns at these more complex levels of organization are often difficult to recognize, they involve some of the most challenging and important questions in ecology.

How specific should comparisons be?

To make a study practical and manageable, ecological questions should be as specific as possible. In general, a good initial observation is one that clearly implies the type of measurement required to test it. Ideally, the situations compared should differ in only a few physical or biotic features, thus making it easier to postulate the causes of any differences that are noted.

DESCRIPTIVE STAGE

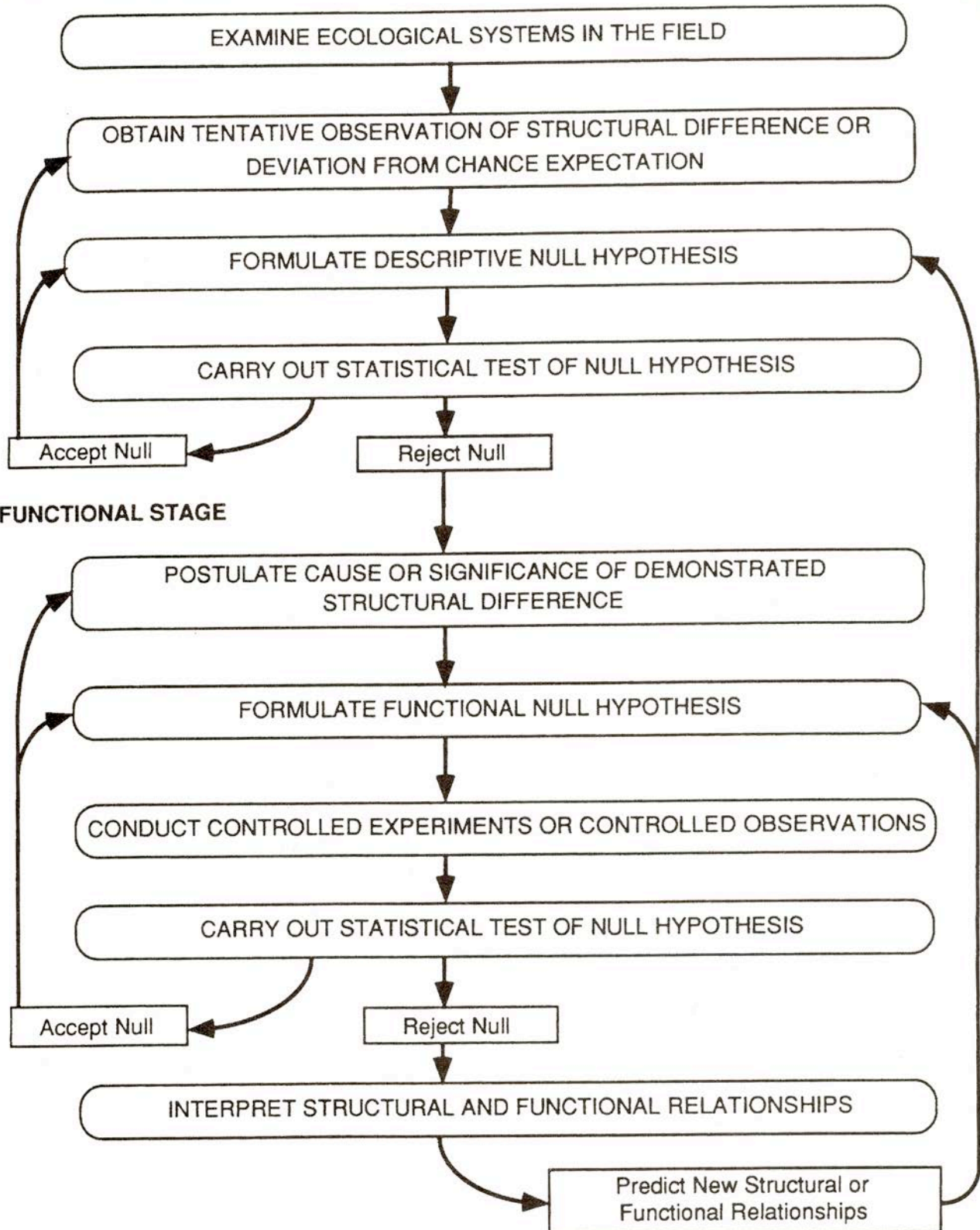


Figure 1.1.

Idealized sequence of steps leading from a tentative initial observation in the field through descriptive and functional stages of ecological study.

Making Initial Observations

Some of the comparisons that can be made are outlined here. Suggestions about how structural differences might be related to function also are given. These comparisons cover all levels of organization from the organism to the ecosystem.

Compare one species in different situations.

Compare individuals or populations of one species in different geographical areas, in different habitats, or at different times. Couple these observations with notes on conditions of the physical and biotic environment. Specific features to compare:

Abundance. Presence versus absence of a species or a difference in population density might be due to habitat selection behavior, to availability of specific resources, to limits of tolerance for conditions of the physical environment, or to beneficial or detrimental interactions with other species.

Morphology. Individuals may differ in size, shape, structure, or color due either to genetic differences or to the direct action of the environment during development. Whether genetic or not, such features may have adaptive value, by increasing the growth, survival, or reproduction under the respective physical or biotic regimes (e.g., climate, herbivores, predators, competitors).

Behavior. Daily or seasonal activity times, foraging behaviors, nest site selection, frequency of aggressive displays, and many other behaviors may differ in various situations. These may reflect differences in physical or biotic conditions of the habitat, the population density of the species itself, or the abundance of competitors or predators.

Population dispersion and structure. How individuals are dispersed (randomly, clumped, or uniformly) may reflect the heterogeneity of habitat conditions, positive or negative interactions among individuals, or mode of reproduction. Conditions that affect reproductive success, intensity of predation, or risk of mortality may influence age structure and sex ratio.

Compare characteristics of different species in the same situation.

Coexisting species may show features that prevent detrimental interactions among them or that make their interaction favorable to one or more species.

Specific features to compare:

Morphology or behavior. Differences may serve to partition the use of resources such as space, food, or nest sites, thus reducing interspecific competition. They may also provide protection against predators (herbivores, in the case of plants).

Within-habitat distribution. Differences in the exact sites occupied may reduce interspecific competition for resources, or may reflect different patterns of tolerance of conditions of the physical environment. *Patterns of co-occurrence.* Individuals of different species may tend either to associate closely or to avoid each other, reflecting symbiotic, commensal, competitive, or feeding relationships. Such patterns also may result from actions of one species that modify the physical environment favorably or unfavorably for others.

Compare characteristics of ecologically similar species in different situations.

Similar species may be restricted to different habitats, or to different daily or seasonal activity times in the same habitat, by competition with each other. Such restriction also may reflect adaptation to different physical or biotic environments.

Specific features to compare:

Morphology and behavior. Species similar in resource use may have adaptations that make each competitively superior in its own situation. Other differences may be adaptations to use resources or to tolerate environmental conditions that differ between the habitats or times.

Population dispersion and structure. Differences in the heterogeneity of habitat conditions may affect dispersion patterns. Differences in physical or biotic factors that affect survival and reproduction may be reflected in population density, age structure, or sex ratio.

Compare community or ecosystem characteristics in different situations.

The abundance of various ecological groups of organisms (e.g., annual or perennial plants, herbivores, carnivores) may vary with climate, nutrient and energy supply, or disturbing factors such as fire or flooding. Abiotic conditions such as microclimate, soil structure, and the profile of physical and chemical conditions in lakes may also differ due to such influences. Biotic or abiotic features of communities and ecosystems may differ because of processes of biotic succession, or because of differences in the ability of organisms to colonize the sites.

Specific features to compare:

Characteristics of species. The patterns of morphology, behavior, or population structure of species may differ (e.g., annuals versus perennials, grazers versus browsers, generalists versus specialists). These might reflect different physical conditions, different mechanisms of resource use, or different stages of biotic succession.

Number of species. Habitat heterogeneity, stability of physical conditions, constancy and level of primary production, degree of geographic isolation, abundance of predators, and stage in succession may influence the diversity of species.

Distribution of individuals among species. Differences in the frequency of rare and common species might relate to the mechanism of division of resources among community members or to the influence of disturbance.

Numbers and biomass of different trophic groups. The abundance of producers, herbivores, carnivores, and decomposers might be related to the rate of energy or nutrient input to the system, export or import of organic matter, body size-metabolism relationships of community members, patterns of nutrient cycling, stage in biotic succession, and many other factors.

Formulating and Testing Null Hypotheses

Hypotheses about structure or function must be tested to determine if they have a high probability of being correct. Subjective impressions about such relationships often turn out to be wrong, due either to the inadequacy of the initial observations or to bias of the observer. A formal test of a hypothesis is carried out by stating a null hypothesis (H_0), collecting unbiased observations or experimental data, and performing a statistical test of the null hypothesis (but see Krebs 2000). The null hypothesis has the following form:

H_0 : The difference between two or more sets of observational or experimental data is not greater than expected by chance.

or

H_0 : The values obtained in an experiment or set of observations do not differ from a theoretical expectation more than expected by chance.

In the latter case the comparison is between a real data set and a theoretical data set in which the values are those predicted by a *model*. If the model assumes the operation of random processes, it is termed a *null model*. For example, if a hypothesis deals with the spacing of individual plants in a population, the theoretical expectation could be one of random spacing. If the model contains mechanistic interactions that can be expressed as equations (e.g., the exponential or logistic equation), it is termed a *mathematical model*. For example, if a growth of a population is measured over time, it might be compared to predictions of the logistic equation of population growth. A model of relationships within a complex system, using a series of equations or other modeling techniques, is termed a *simulation model*. Simulation models can yield predictions of many measurable features, to which data from a real system can be compared.

A statistical test of the null hypothesis leads either to its acceptance or rejection. Accepting the null hypothesis means that either no difference exists or that the data are inadequate to demonstrate it. Rejection means that a difference exists with a probability corresponding to that of the significance level used in the statistical test. Rejection of the null hypothesis results in acceptance of a second, or alternate hypothesis (H_a), which has the following form:

H_a : The difference between two (or more) observational or experimental situations is so great that it is very unlikely to have occurred by chance.

or

H_a : The set of observed values differs so greatly from the theoretical expectation that it is very unlikely to have occurred by chance.

A good null hypothesis is simple, suggests a characteristic that can be measured accurately, and concerns a relationship for which adequate sampling or experimental data can be obtained.

Example

The following example illustrates the sequence leading from an initial observation through an analysis of one aspect of functional significance.

Initial observation.

A plant species was observed in two locations, one on an ocean bluff, the other on a ridge 1 kilometer inland. The ocean bluff plants appeared to have thicker leaves.

Descriptive null hypothesis.

H_0 : There is no difference in mean thickness of leaves taken from plants in the two areas.

Descriptive sampling and testing program.

In each area, 30 plants are sampled. One leaf per plant, between 3.5 and 4.0 centimeters in length, is taken. Leaf thickness is measured to the nearest 0.1 millimeter with calipers. Mean values for plants from the two areas are compared by a t test (see exercise 4), and the null hypothesis is rejected.

Functional null hypothesis.

The difference might be due, proximately, to genetic differences in the plants rather than to direct effects of habitat factors during individual growth (many other functional hypotheses can be made). The following null hypothesis is then stated:

H_0 : Plants grown side-by-side from seed taken from the two populations will show no difference in mean thickness of leaves.

Functional experimentation and testing program.

Seeds are collected in the two areas and grown until the plants reach some minimum size. Leaves are sampled and measured, and the data summarized and tested as described earlier. If this test rejects the functional null hypothesis, the conclusion follows that the populations, with a certain probability, differ genetically in factors related to leaf thickness. If the functional null hypothesis is not rejected, the conclusion is that the data are not consistent with a genetic difference in leaf thickness, not that genetic differences are absent.



Experimental Design

- ☞ How can one take random variability of observations or experimental data into account?
- ☞ How can one assure that the data collected can be used to test a stated hypothesis?
- ☞ How can one evaluate the impact of major environmental actions that cannot be replicated?

The conclusions that can be gained from an experiment depend primarily on its design. Unfortunately, many experiments are conducted with inadequate attention to design, so that enormous amounts of hard work often yield data that cannot be used to test the hypotheses that led to the experiment.

This exercise describes basic features of manipulative experiments—that is, laboratory or field experiments in which events affecting the experimental units are controlled by the investigator.

Experimental Design

Manipulative experiments involve one or more *treatments* that modify conditions in some or all of a set of *experimental units*. The effects of treatments are evaluated by measuring one or more *response variables*. In testing the influence of a fertilizer on productivity, for example, some plots—the experimental units—might be treated with a certain amount of fertilizer while others receive no treatment. Or, different sets of plots might be given different amounts of fertilizer. Response variables of the experimental units, such as the producer biomass or rate of primary production, are then measured.

The *design* of an experiment refers to the characteristics of experimental units, the types of treatments applied to them, the numbers of units receiving each treatment, and how treatments are applied in space and time. Manipulative experiments are subject to several sources of *variability* that can make the results difficult or impossible to interpret. The confusion caused by these sources of variability (table 2.1) can be reduced by several design features: treatment characteristics, replication, controls, randomization, and interspersion.

Treatments

A *treatment* is a feature of the experimental system that is manipulated by the investigator to determine how the system responds to its alteration. In designing an experiment, care must be taken to assure that the treatment does not have multiple components, some of which may not be obvious (Huston 1997). For example, in a grassland productivity experiment in which plots are fenced to exclude grazing animals, fencing may not only reduce animal herbivory, but may also modify microclimatic conditions and eliminate nutrient inputs in animal urine and feces. Such "hidden treatments" may influence the experimental results and lead to their incorrect interpretation.

Replication

In manipulative experiments, units chosen to be as similar as possible may exhibit differences in the response variables because the units were not truly identical. Also, differences can appear due to random errors in the measurement process itself. Of course, care in choosing the experimental units and in making measurements can minimize these sources of variability, but they cannot be eliminated completely. This variability, usually termed experimental error, can be measured if experimental units are *replicated*—that is, if several units are assigned to each treatment or condition.

Experimental units must be systems of the type to which the investigator intends to apply the conclusions of the experiment. For example, if one is interested in testing the hypothesis that addition of phosphorus fertilizer increases the primary productivity of ponds in general, the appropriate experimental units are *entire ponds*. Such an experiment, therefore, must consist of two or more ponds treated with fertilizer and compared to two or more untreated ponds of similar type. If only a single pond is fertilized, productivity measurements taken at

several locations or at several times do not constitute replicate experimental units. The analysis of such measurements as replicate experimental units, equivalent to different ponds, is termed *pseudoreplication* (Hurlbert 1984, Hurlbert and White 1993, Heffner et al. 1996), and is a common error in ecological studies. Avoiding pseudoreplication thus involves increasing the number of ponds, or accepting the fact that results apply not to ponds in general, but only to the ponds utilized (Bart et al. 1998).



Table 2.1. Sources of variability in experiments, and procedures for reducing confusion caused by such variability

Source of Variability	Reduction by Experimental Design
1. Variability among experimental units	Replication, interspersed, and simultaneous measurement
2. Random error in measurement of response variables	Replication
3. Change in conditions through time	Controls
4. Unsuspected side effects of treatment procedures	Controls
5. Bias of investigator	Randomized assignment of treatments to experimental units
6. Chance influences on experiment in progress	Replication and interspersed

Several types of pseudoreplication have been recognized. *Simple pseudo replication* involves taking multiple measurements of response variables within individual experimental units and analyzing these measurements as if they came from different experimental units. This is usually evidenced by the use in statistical tests of a number of degrees of freedom based on the total number of measurements rather than on the number of experimental units. For example, suppose that one pond is treated with fertilizer and another is untreated, and that 10 productivity measurements are subsequently made in each pond. A statistical test that purports to test whether fertilization affects the productivity of ponds in general, and involves error degrees of freedom based on the 20 measurements (e.g., a *t* test of means with 20 - 2 degrees of freedom) is an example of simple pseudoreplication.

Temporal pseudo replication occurs when measurements taken within the same experimental unit at different times are treated as coming from separate experimental units. If the 10 productivity measurements in the pond example are taken on successive days, and 18 (20 - 2) error degrees of freedom are used in a comparison of the effect of fertilization on ponds, temporal pseudoreplication has been committed.

Sacrificial pseudo replication occurs when the opportunity to distinguish variability due to measurements within and between experimental units exists, but is ignored or "sacrificed" by combining all data in a single analysis. Suppose, for example, that three different ponds were fertilized and three were not fertilized, and that five productivity measurements subsequently were taken in each (a total of 30 measurements). A simple comparison of the effect of fertilization on pond productivity, using 28 (30 - 2) error degrees of freedom in a statistical test, is an example of sacrificial pseudoreplication. In this case, a nested analysis of variance (see exercise 4) could be used to partition variability in productivity values within ponds, among ponds of similar treatment, and between fertilized and unfertilized ponds. The appropriate test of the influence of fertilization on pond productivity would have 4 (6 - 2) error degrees of freedom.

Controls

In manipulative experiments, controls are experimental units that are identical to those receiving manipulative treatments except in the critical treatment factor. For example, if fertilizer is applied to a plot of grassland by spraying a solution of fertilizer in water, control plots should receive an identical spraying treatment, with water only. In this case, the control plot serves to reveal whether change in the plots receiving fertilizer is indeed due

to the fertilizer, rather than to an accessory feature of the fertilizer treatment, such as addition of water or trampling by the experimenters. Controls can also reveal whether or not some change through time is tending to occur in the plots because of factors the experimenter cannot hold constant, such as seasonal change in day length. Controls are essential for ecological field experiments, because it can rarely be assumed that conditions in nature will remain constant for any substantial time, and because almost any measurement or manipulation involves incidental impacts of the investigator.

Randomization and Interspersion

In manipulative experiments, the investigator may create replicated experimental units, some serving as controls and some as treatments. These must be interspersed in space or time so that, on average, the different sets of units experience the same environmental conditions. Treatments and controls can be assigned to experimental units randomly. If the number of replicates is large, a random assignment procedure serves well, since it is very unlikely that, for instance, all controls would end up clustered in one location and all treatment units clustered in another.

In ecological field experiments, however, the number of replicates is often small, out of necessity. If an investigator sets up trapping grids to determine the effect of supplementary food on populations of rodents, the effort and large area required for each experimental unit probably limit the number of experimental units per treatment to three or four. If control or supplementary food status were assigned to grids randomly, it is very possible that control grids would by chance be clustered in one part of the study site and treatment grids in another. Such an arrangement opens the door to a chance systematic influence operating differentially on one part of the area in which the grids are located. Hurlbert (1984) terms such an influence a "non-demonic intrusion." For example, predators might enter the grid area more frequently from one side than from the other, or vegetation density might change gradually from one side to the other. Without interspersion, such differences might exert a bias on responses of control or treatment plots.

Where such a possibility exists, semisystematic or systematic interspersion of experimental units is desirable. Control and treatment units can be alternated, or arranged in checkerboard fashion. A *randomized block design* (figure 2.1), which combines randomization and systematic interspersion, is a frequently used arrangement. In this procedure, the total number of experimental units is divided into sets known as blocks. Each block consists of a number of experimental units equal to the number of different types of treatments (including control). Within each block, treatments are then assigned randomly to the experimental units. A *Latin square design* (figure 2.1) is even more systematic in arrangement, having treatments assigned so that a given treatment occurs only once in each row or column.

RANDOMIZED BLOCK DESIGN

D	A	B	C	Block 1
B	C	D	A	Block 2
A	B	D	C	Block 3
C	B	A	D	Block 4

LATIN SQUARE DESIGN

B	D	A	C
A	C	B	D
D	A	C	B
C	B	D	A

Figure 2.1.

Schematic representation of plot layout for experiments with a randomized block design and latin square design. The letters A, B, C, and D indicate four different treatments.

Unreplicated Perturbations

In some cases, replication is impossible or impractical, as in the case of assessing the ecological impacts of a power plant, dam, or other large development on an aquatic ecosystem (Schindler 1998). Although these are, in a sense, giant experiments, in that they are planned ahead of time and their location selected, they are single "treatments" that do not permit an experimental assessment of general treatment effects. It may be possible, however, to test whether or not the specific project causes a significant change in conditions that prevailed before it was constructed.

Suggested designs for evaluating such a perturbation are the *Before-After Control-Impact Pairs (BACIP) Method* (Stewart-Oaten et al. 1992) and a modification of BACIP involving the use of multiple control sites (Underwood 1994). These techniques involve simultaneously measuring response variables at impact (treatment) and control sites on a series of occasions before the structure is constructed and on a series of occasions afterwards. The control site or sites must be nearby and similar to the impact site, so that it can reasonably be assumed that without the impact the sites would change through time in parallel fashion (the validity of the technique depends on this assumption being fulfilled!). Intervals between measurements must also be long enough that the values obtained are statistically independent. The differences between measurements at control and impact sites can then be compared for the before and after periods by an appropriate statistical test.