Cambridge University Press
978-0-521-19154-8 - Design and Analysis of Long-term Ecological Monitoring Studies
Edited by Robert A. Gitzen, Joshua J. Millspaugh, Andrew B. Cooper and Daniel S. Licht
Excerpt
More information

# Section I

## Overview

# 1 Ecological monitoring

## The heart of the matter

Robert A. Gitzen and Joshua J. Millspaugh

### Introduction

Most environmental scientists and natural resource managers view long-term monitoring as a good thing. To fulfill their mandates and meet their goals, natural resource management and conservation organizations need information on the current status and patterns of change in high-priority resources, critical ecological processes, and stressors of high concern. To inform policy makers and the general public as well as management organizations, government agencies may have mandates for assessing the condition of selected natural resources in perpetuity because of the special importance of these resources to society and because of the possibility of both suspected and unforeseen changes to these resources. More directly, these organizations need information for assessing whether current management is effectively maintaining the state of managed systems and populations within the desired range of conditions, and for reducing uncertainty affecting management performance.

Moreover, there is incredibly high scientific value in many long-term data sets collected for management and conservation purposes, as well as data collected as part of long-term ecological research programs. Natural systems usually are characterized by complex temporal dynamics and interactions often not evident from a collection of short-term research studies. Data from long-term monitoring can be invaluable for empirical examination of hypotheses about spatial and temporal dynamics – for example, about community dynamics, population growth and density dependence, and influences of infrequent extreme climatic events. With the possibility of rapid and major changes to the Earth's climate, there is unprecedented recognition of the importance of existing long-term data sets and demand for new monitoring studies (e.g. see Chapter 22).

In light of these multiple potential uses of information from monitoring, countless monitoring programs have been implemented or are being developed by national, regional, and local agencies (see numerous examples in subsequent chapters); private entities such as forest management companies; conservation groups and other non-government organizations and citizen groups (e.g. Curtin 2002, Topp-Jørgensen *et al.* 2005, Fernandez-Gimenez *et al.* 2008; www.monitoringmatters.org; Chapter 21); and individuals (e.g. Bradley *et al.* 1999). There are few if any organizations in the field of

*Design and Analysis of Long-term Ecological Monitoring Studies*, ed. R.A. Gitzen, J.J. Millspaugh, A.B. Cooper, and D.S. Licht. Published by Cambridge University Press. 

natural resources that do not profess an interest in long-term monitoring, and most are planning or conducting monitoring in some form.

Although interest in monitoring is not new, recently there has been an explosion of discussion in the scientific literature about its design and analysis. Much of this discussion has focused on (i) the purpose and role of monitoring, including discussion of objectively assessing costs: benefits of monitoring in relation to other uses of funding (Nichols and Williams 2006, Hauser *et al*. 2006, Lindenmayer and Likens 2010b, McDonald-Madden *et al*. 2010, Wintle *et al*. 2010); and (ii) statistical aspects of monitoring. The latter focus includes guidance and case studies discussing how monitoring can be designed properly as well as continued development and application of new approaches for design and analysis (e.g. Legg and Nagy 2006, Field *et al*. 2007, Chandler and Scott 2011, MacKenzie *et al*. 2011, Reynolds *et al*. 2011). Availability of broad qualitative and quantitative guidance about how to do things "right" in monitoring continues to increase rapidly (e.g. Thompson *et al*. 1998, Elzinga *et al*. 2001, de Gruijter *et al*. 2006, Lindenmayer and Likens 2010a, McComb *et al*. 2010). Yet, despite this high interest in monitoring and increased understanding of the importance of design and analytical considerations, there have been no broad syntheses addressing these key methodological issues. In particular, there has been a lack of comprehensive guidance providing diverse perspectives and recommendations from monitoring experts, incorporating some design and analytical approaches that are or are rapidly becoming standard tools in ecological monitoring (e.g. spatially balanced sampling, hierarchical modeling, structural equation modeling, modeling threshold changes; see subsequent chapters), and providing advanced and extensive coverage of both design and analysis of complex monitoring studies.

This volume focuses on key questions and some important methods related to the design and analysis of long-term ecological monitoring studies. The choices of survey designs and analytical approaches for monitoring are questions with inherent statistical aspects. However, choices among design and analysis alternatives are not made in a statistical vacuum; they are intricately tied to the purpose and objectives of monitoring. Therefore, from the standpoint of this volume, "design" of monitoring includes not only specification of how, where, and when to collect data, but also specification of the specific objectives for monitoring, which in turn is driven by the general rationale for monitoring. Another reason why such a broad definition is appropriate for this quantitative text is because decisions about why and what we should monitor – and even whether to start or continue monitoring – increasingly are being placed in a quantitative framework (e.g. Wintle *et al*. 2010; Chapter 23).

In this chapter, we provide our perspectives on the context for this volume, and discuss challenges and recommendations related to the design and analysis of monitoring. We start by briefly emphasizing the importance and role of quantitative considerations in monitoring (Box 1.1); these issues are discussed further in Chapters 2 and 22, and emphasized, implicitly or explicitly, by every other chapter. We next discuss some potential reasons why inadequate attention to qualitative and quantitative design issues has been reported to be such a common problem in monitoring programs, and provide our perspectives on current risks from these causal factors. We then highlight several indicators

**Box 1.1** Take-home messages for program managers

Quantitative issues are not the only critical aspects of an environmental monitoring program. However, if monitoring is intended to produce useful and conclusive information, none of the other components of the program can compensate for inadequate attention to design and analytical issues. Long-term monitoring involves complex statistical and subject-matter considerations. An organization can either invest in the expertise needed to carefully address these issues during design and analysis (see Chapter 2), or it can take short cuts and rely on blind luck in the hope that money spent on monitoring produces something useful, eventually. We recommend the former.

An important issue during development of a monitoring study is the identification of specific quantitative sampling objectives (desired statistical power, precision) related to primary variables or processes of interest, and of specific high-priority statistical analyses that are planned once monitoring is operational. These are essential inputs into quantitative study design. For example, a power or sample size analysis has to be based on a specific analytical method expected to be appropriate for analyzing the monitoring data. The fact that these features have been defined, and the process by which they are identified, are also critical indicators of whether a program knows why it is planning to conduct monitoring. It is not a good sign if they are selected in an offhand fashion. If a study is intended to be relevant to natural resource managers, it is worrisome if managers do not take the lead or collaborate closely in specifying these quantitative objectives. This issue should be addressed in the same way as all quantitative issues in monitoring: with integrated consideration of statistical, ecological, and practical aspects of the problem, based on clearly defined reasons for monitoring.

of whether a monitoring program is thoroughly addressing quantitative issues. Finally, we suggest selected general steps for continued progress in ensuring high quantitative standards for monitoring.

## The role of quantitative considerations in monitoring studies

For a monitoring program to succeed, quantitative issues must be addressed adequately unless the program is focused primarily on education, making monitoring clients feel good in the short term, or allowing an organization to check a box indicating that "monitoring" occurred. The critical importance of design and analytical issues is the rationale for this volume. It may seem self-evident to many readers, yet many reviews suggest that inadequate attention to quantitative issues is one of the common causes for failure or ineffectiveness of monitoring programs (Noon 2003, Legg and Nagy 2006, Lindenmayer and Likens 2010a, and multiple other reviews).

If a major purpose of a monitoring study is to produce useful and relevant information for managers, policy makers and scientists, quantitative issues have to be addressed adequately (Box 1.2). The meaning of "useful and relevant" must be defined for each individual study in the process of specifying clear objectives. To be useful, information also needs to be timely, based on sufficiently accurate data and analyses, and able to stand up to objective peer scrutiny as well as agenda-driven criticisms. Adequate consideration of quantitative issues includes assessing whether reliable data can be produced with feasible survey effort, determining how to allocate available effort into an effective, efficient, and defensible survey design, and determing how to produce useful information from the resulting data; see Chapter 2 for further discussion.

**Box 1.2** Common challenges: how low can you go?

In practice, the selection of meaningful quantitative objectives usually involves some degree of iteration: defining a target level of information quality, conducting an examination that often shows there is no realistic way of obtaining this quality, and lowering expectations, such as accepting a lower expected precision or restricting the target population to focus available resources more effectively (Chapter 10). Also, in practice, there is a strong temptation to take the following approach: set the budget for a particular study, determine what level of power or precision can be obtained with the feasible effort, and set this level as the sampling objective. The problem is that this requires no thought about what quality of information the study needs to provide to be successful. The first, iterative approach of progressively lowering expectations can also simply end up at the same point as the second approach – "we don't know what we want, but we do know what we can afford" – despite the best of initial intentions. Money is critical – but shouldn't there be some safeguards against simply spending money because it is available?

Because of this, we recommend defining "quality" and "usefulness" by also defining "worthlessness". In addition to setting the desired level of power or precision needed by the program, also define some hard-line threshold or "floor" for minimum acceptable expected performance. Prior to running any quantitative examination, what is the highest reasonable threshold below which you would say, with confidence, that "if this is the best we can do, we should not even bother collecting data"? This is a useful but uncomfortable question. If you can't define such a threshold, then you essentially are assuming one of the following: (a) if we collect data, we have faith that some useful information will materialize; (b) our primary focus is not on producing useful information; or (c) any data are better than no data. The first two possibilities relate to organizational priorities and efficiency, and are not inherently wrong or right. "Any data are better than no data" is the most worrisome. From an exploratory standpoint, there is value in what may turn out to be a 15-year pilot study. However, for addressing specific objectives, too little information or low quality information can produce the wrong conclusion and perhaps a wasted or harmful management response, or lack of response (Chapters 2, 18).

A sticky issue is that terms such as "suitable accuracy", "adequately", and "defensibility" are partially linked to subjective professional judgment. Accuracy should be assessed in reference to quantified sampling objectives, but at some level professional judgment and consensus usually are used to set desired precision, power, and other criteria. The closest we may come to a priori "objective" sampling objectives may be the precision benchmarks of Robson and Regier (1964) in an ecological context and perhaps the effect sizes of Cohen (1988) in a human-behavioral context, and these authors offer such standards simply as helpful but arbitrary conventions. Yet, professional judgments about desired accuracy are necessary, appropriate, and well-informed if based on hard thought about some intended uses of the monitoring data. Adequate consideration of quantitative issues does not rule out the need for some professional judgments, but rather leads to reduced unnecessary use of subjective judgments and greater use of structured scientific thinking, numerical evidence, and available statistical guidance. Later in the chapter we suggest some general criteria for assessing whether a monitoring study is likely to have adequately addressed quantitative issues.

Our emphasis in this chapter is on quantitative issues, but by no means are we implying that quantitative issues on their own determine the success or failure of a program, or that other factors have not been just as important in driving many past programs to failure (Noon 2003, Lindenmayer and Likens 2010a). Chapter 22 provides a broad context for this volume by discussing other critical factors which must be addressed in addition to those related to statistical planning and implementation, and illustrating how these factors have been addressed in the US National Park Service's long-term monitoring program. For example, monitoring programs must have efficient administration, strong support by intended clients of the program, effective data management, and reporting that is regular, timely, and useful, such that the program is widely viewed as productive and relevant (Elzinga *et al*. 2001, Fancy *et al*. 2009, Lindenmayer and Likens 2010a, McComb *et al*. 2010; see also Chapter 3). Ultimately, programs must have money; high-quality, productive, well-supported programs still face the budget axe. Programs try to maximize their survival probability by simultaneously and successfully meeting all of these requirements, quantitative and qualitative.

## Flawed designs: their causes and avoidance

As noted above, monitoring programs have been highly susceptible to inefficiency and failure as a result of poor designs. If this problem was over-stated, we doubt that monitoring experts and biometricians would feel the need to continually emphasize some basic design issues, such as the critical importance of probability sampling in surveys when inference about a target population of interest, such as a study area, is to be based on measurement of only a subset of the population (e.g. Anderson 2001; Chapters 2, 5, 6, 18, 22). Why have inadequate designs been too frequently developed, particularly when simultaneously there have been many successful "status and change" surveys with rigorous sampling designs in place, at national and even cross-boundary scales (e.g. see examples in subsequent chapters and in McComb *et al*. 2010)? We suggest the problems can be traced at least partially to inadequate knowledge and attitudes of scientists and managers involved in monitoring.

Frequent use of non-defensible survey designs in ecology (Anderson 2001) suggests that many scientists and managers either do not understand or have tended to gloss over the importance of carefully, sometimes painfully developing rigorous survey designs to meet defined quantitative objectives. Interest in monitoring has long been high, but this often has been paired with limited knowledge and experience in carefully designing surveys. In many ecological disciplines, mandatory statistical training for students typically has focused primarily on a cookbook of standard analyses, with some coverage of experimental designs, often little coverage of survey design (except perhaps in forestry programs), and even less emphasis on conceptual thinking about objectives, important sources of bias and variance, and integrated choice of design and analytical approaches. Arguably, an enthusiasm to get started on monitoring combined with insufficient backgrounds in study designs often led to ineffective efforts, especially prior to the availability of rigorous guidance accessible to field personnel that emphasized broader design issues (e.g. Wilson *et al*. 1996, Elzinga *et al*. 1998, 2001, Thompson *et al*. 1998) as much as choice of measurement techniques. For example, in an important reference for habitat monitoring (Cooperrider *et al*. 1986), sample selection received very little attention, perhaps under the assumption that most field biologists and managers already had a firm understanding of the importance of proper sampling. This does not seem to have been a valid assumption.

Engagement of academic researchers to help design monitoring often probably has helped avoid some monitoring pitfalls, but not all. The default approach of such researchers when dealing with observational studies is to set up comparisons that mirror standard experimental designs *sans* randomization. In ecological disciplines, such comparative studies are an essential complement to controlled experiments, although proper use of these studies requires explicitly acknowledging limits to the inference possible (Shaffer and Johnson 2008). On the one hand, the researcher's natural tendency to incorporate careful comparisons into monitoring has important advantages. When monitoring is viewed as data collection free of specific scientific questions and hypotheses other than about presence/absence of a population-wide trend, the potential management and scientific value of the study is short-changed (Noon 2003, Lindenmayer and Likens 2010a).

However, a monitoring study usually is not just a research study with a multi-year component. Often the goals of surveys, such as in monitoring programs, include estimation of population-wide parameters as well as comparison of subpopulations of interest and examination of relationships (Cochran 1977). However, many ecological researchers either did not have much experience viewing observational studies with a survey-sampling perspective rather than a purely quasi-experimental design perspective, or perhaps viewed their role partly as shifting descriptive surveys into the realm of research. Either way, this probably has contributed to the traditional overemphasis of null hypothesis testing vs. parameter estimation and other approaches to inference in ecological studies (Johnson 1999, Burnham and Anderson 2002). Conversely, unless framed as research comparisons, observational studies and surveys, including monitoring, often have been seen simply as the purview of management, not science (Noon 2003, Lindenmayer and Likens 2010b). The view that monitoring is distinct from "science" may have led both ecological scientists and, ironically, managers to have fairly low standards for the quality of data expected from monitoring. In addition, the attitude that "we are simply trying to detect changes" often has led practitioners to move forward with poorly defined objectives, a bane of many programs.

For comparative studies viewed as non-randomized experiments, judgment or convenience sampling (Chapters 2, 5) of "representative" sites in different groups being compared has been widespread, rather than random selection from a pool of available sites. Such an approach, problematic enough for research comparisons, is simply a fatal flaw for monitoring studies seeking to make population-wide inferences based on a sample of sites. However, in monitoring programs these types of critical flaws often have been seen as necessary concessions to the reality of working "in the real world."

In truth, time and money always are limited. Perhaps we are underestimating the effect of financial limitations (Noon 2003). When budget cuts doom a monitoring program, this may have nothing to do with its statistical design. In other situations, financial limitations may be the proximate reason for inadequate design, but underlying attitudes and philosophies are part of the problem. For example, although money always has limited sample sizes and access to expertise, this often was and is combined with attitudes that "whatever we can afford has to be good enough" and "getting some data is always better than having no data." Later in the chapter we will discuss the importance of avoiding these dangerous attitudes, but their prevalence is demonstrated by the relatively low use of quantitative evaluations (power/precision examinations) by many past monitoring efforts to justify their designs (Legg and Nagy 2006, Marsh and Trenham 2008). To some extent, this low use is fairly sensible, if we simplistically assume for a moment that the only goal of these examinations is to assess whether there is any chance of meeting quantitative objectives with available effort, and less naively assume that the planned design was going to be implemented with only minor tweaks regardless of the results of the examination. Also, such examinations are based on specific intended uses of monitoring data. When the first step in quantitative design examinations indicates that that objectives have not yet been specified clearly enough to support meaningful examinations, there is a strong temptation simply to skip that part of the process. A rushed, non-careful design phase is also likely when the impact of quantitative considerations on the quality of information produced by monitoring is underestimated.

We have argued that frequent monitoring-design problems are often caused by insufficient knowledge and insufficiently careful approaches to monitoring design. These factors often lead ecologists (students, scientists, and managers) to underestimate their need for additional input from statistical experts (Millspaugh and Gitzen 2010), in monitoring as well as other contexts. Typically, ecological statisticians have been most commonly viewed as quantitative plumbers, to be consulted after data are collected if the analysis becomes clogged. Still, even when monitoring practitioners recognized their need for additional quantitative help, they may not have had feasible access to such experts, a deficiency needing continued attention. Overall, if suitable statisticians or other personnel with suitable training and expertise weren't involved from the beginning, there is a good chance that a monitoring study would have an inadequate design that either doomed it to failure, led it to produce information of much lower value than could have been obtained, or, in the worst-case scenario, led it to survive and produce misleading data. Are those days in the past?

## What has and hasn't changed?

### Progress

Although historically it was mainly a few regional- or national-scale programs that had high standards regarding quantitative aspects of monitoring, high quantitative standards are becoming the norm at many levels of monitoring. Partly this may be an outcome of learning by experience, given frequent commentaries on the prevalence and problems of inadequate designs. Moreover, there is greater awareness of the success of well-established programs with probability survey designs that have an ongoing history of supporting flexible inference at local to national scales (e.g. Nusser *et al*. 1998, Bechtold and Patterson 2005). Hopefully, the number of good examples is on its way to exceeding the number of bad examples. For example, in the USA, ecologists with the National Park Service (NPS) and their collaborators from other agencies and organizations have implemented or are in the process of designing hundreds of monitoring protocols with clear objectives and defensible survey designs (Chapters 10, 16, 22), even for very small target populations (e.g. vegetation monitoring at Fort Union Trading Post National Historic Site in North Dakota/Montana, USA, uses spatially balanced sampling from a 133-ha sample frame; Symstad *et al*. 2011). This type of national-scale program with a local-scale focus broadly expands awareness of available methods, and pulls in many professional ecologists and statisticians to further develop and apply diverse design and analysis approaches. It also leads many managers and biologists to think about the importance and requirements of sound statistical design. Managers and environmental scientists who learn about and successfully implement sound designs motivate and raise the bar for other programs, and build more widespread availability of quantitatively skilled people with experience in monitoring design and analysis.

There also has been a rapid decrease in the tendency to view monitoring as less interesting, and requiring less rigor, than "real science." This may be driven partly by the blurring of any past divisions between basic ecology and highly applied, management/conservation-focused science; an increased focus on estimation rather than rote hypothesis testing in ecological science (e.g. Burnham and Anderson 2002); and by increased realization about the potential for monitoring programs – and management/policy-focused data collection in general – to provide data highly useful for addressing ecological questions. As a result, there is broader recognition of the need for strong scientific partnerships in the design and analysis of monitoring (e.g. Lindenmayer and Likens 2010a; Chapter 22), and in turn ecological scientists are highly interested in the why and how of monitoring (e.g. Nichols and Williams 2006, Lindenmayer and Likens 2010a, b). In addition to being able to help monitoring