10 Simulating future uncertainty to guide the selection of survey designs for long-term monitoring

Steven L. Garman, E. William Schweiger, and Daniel J. Manier

Introduction

A goal of environmental monitoring is to provide sound information on the status and trends of natural resources (Messer *et al*. 1991, Theobald *et al*. 2007, Fancy *et al*. 2009). When monitoring observations are acquired by measuring a subset of the population of interest, probability sampling as part of a well-constructed survey design provides the most reliable and legally defensible approach to achieve this goal (Cochran 1977, Olsen *et al*. 1999, Schreuder *et al*. 2004; see Chapters 2, 5, 6, 7). Previous works have described the fundamentals of sample surveys (e.g. Hansen *et al*. 1953, Kish 1965). Interest in survey designs and monitoring over the past 15 years has led to extensive evaluations and new developments of sample selection methods (Stevens and Olsen 2004), of strategies for allocating sample units in space and time (Urquhart *et al*. 1993, Overton and Stehman 1996, Urquhart and Kincaid 1999), and of estimation (Lesser and Overton 1994, Overton and Stehman 1995) and variance properties (Larsen *et al*. 1995, Stevens and Olsen 2003) of survey designs. Carefully planned, "scientific" (Chapter 5) survey designs have become a standard in contemporary monitoring of natural resources.

Based on our experience with the long-term monitoring program of the US National Park Service (NPS; Fancy *et al*. 2009; Chapters 16, 22), operational survey designs tend to be selected using the following procedures. For a monitoring indicator (i.e. variable or response), a minimum detectable trend requirement is specified, based on the minimum level of change that would result in meaningful change (e.g. degradation). A probability of detecting this trend (statistical power) and an acceptable level of uncertainty (Type I error; see Chapter 2) within a specified time frame (e.g. 10 years) are specified to ensure timely detection. Explicit statements of the minimum detectable trend, the time frame for detecting the minimum trend, power, and acceptable probability of Type I error (α) collectively form the quantitative sampling objective.

The values specified in this sampling objective affect the required sampling effort. A smaller minimum detectable trend requirement, higher power, a shorter time frame, and a lower acceptable Type I error rate generally increase the effort required to achieve the sampling objective. In addition, the spatial and temporal variability of an indicator

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.

influences sampling effort; higher variability increases the required effort. Estimates of indicator variability are acquired in pilot monitoring or related studies, and used in prospective statistical power analyses to determine the sample size and other aspects of the survey design needed to satisfy the sampling objective (Chapters 2, 7, 8). Budgetary limitations generally dictate a parsimonious annual sampling effort. In many cases rotating sampling effort among groups of sites (called panels) over time, with eventual revisitation of sites, can provide high power for trend detection while also providing suitable precision for status estimation (McDonald 2003; Chapter 7).

Determining sample sizes and re-visitation schedules that satisfy a sampling objective within the limits of allocated programmatic resources requires a concerted effort prior to selection of a final survey design and implementation (Box 10.1). In many situations, the power and precision of candidate survey designs can be assessed analytically, even with complex panel designs (Chapters 7, 8). Often, however, a simulation approach can offer higher flexibility, particularly (i) when analytical power/precision relationships are not well defined for a planned trend/status analysis approach; (ii) for examining how violation of key assumptions affects the performance of the planned trend-analysis approach; (iii) in situations where spatially explicit aspects of the monitoring scenarios need to be examined; or (iv) for examining effects of other factors not easily accounted for using analytical power/precision equations (e.g. Chapters 17, 19, 20). For example, simulation modeling has been used to determine optimal sampling designs for varying spatial patterns of an indicator (Pooler and Smith 2005, Heywood and DeBacker 2007, Morrison *et al*. 2008), to evaluate design variance and bias (McDonald *et al*. 2001), to assess sampling design sensitivity to disturbances (Edgar and Burk 2006), and to assess power to detect long-term trends and properties of sampling strategies (e.g. Eng 2004, Stevens and Olsen 2004, Field *et al*. 2005). In some cases, analytical power calculations may be feasible once suitable estimates of variance are available and sampling objectives have been specified, but simulation modeling of patterns and dynamics of the population under various levels of change may be important for helping examine variance scenarios and even defining what minimum level of trend should be of concern.

Although even basic power/precision examinations lead to more informed choices of survey designs (Chapter 8), ecologists and statisticians also should consider complications likely to arise in most monitoring situations. Prospective power analyses with estimates of historical and current variances assume that future variance components will be of similar magnitudes. In ecological systems, however, changes over time in the spatial and temporal variance of an indicator are inevitable due to stochastic processes such as disturbances and climatic fluctuations, as well as human-mediated impacts such as anthropogenic climate change. Failure to account for potential future changes in indicator variance can lead to over-optimistic assessments of expected design performance. Inflating indicator variance in an analytical evaluation can compensate for future increases, but this does not explicitly consider how agents of change may modify the spatial and temporal variance structure of an indicator. Adaptive monitoring is often invoked to safe-guard against underestimating future performance of a design, where repeated evaluation of accumulating monitoring data and the sampling objective determine when and how to modify the survey design (Ringold *et al*. 1996, Lindenmayer and

Box 10.1 Take-home messages for program managers

Quantitative evaluation of alternative monitoring survey designs is essential for choosing a design that efficiently provides useful information. "Useful" needs to be defined explicitly in the form of a carefully considered sampling objective. For example, a program focused partly on trend detection needs to define what forms and levels of change would be of concern and how quickly such a change needs to be detected to enable managers to take action. In our example in this chapter, the sampling objective is to have an 80% probability of detecting a 3.5% annual decline in grass cover in our arid-system park within 10 years if such degradation occurs, with no more than a 10% chance of falsely concluding there is a trend if degradation does not occur. A program also has to specify a general trend- or status-analysis approach that will be used once monitoring data are collected; this serves as the framework for examining power and precision.

Examination of power, precision, sample sizes, etc., also requires estimates of relevant sources of variation that will affect the accuracy of trend and status estimates once monitoring is operational. Estimates of current variability based on pilot studies and historical data are an essential starting point, but future changes in the spatial and temporal variance of a monitoring indicator are inevitable due to disturbances, climatic fluctuation, and anthropogenic stresses such as climate change. Failure to account for plausible future spatial and temporal changes in indicator variance can lead to overestimation of design performance during the planning phase of monitoring. More importantly, in newly implemented monitoring efforts, the inability of the implemented sampling design to provide the desired precision of status and trend information due to changes in spatial or temporal variance may not be recognized before an indicator has fallen below a critical level. We recommend directly considering the effects of changes in indicator variance resulting from plausible future disturbances and stressors (change agents) when evaluating survey designs for operational implementation. Simulation modeling, as demonstrated in this chapter, is a highly flexible approach for such examinations. Comparison of multiple survey designs under plausible future scenarios helps to inform the selection of a survey design most robust to near-term uncertainty.

Likens 2009; Chapter 20). Based on our experience, available programmatic resources are largely expended on the initial implementation of operational monitoring. Once operational monitoring has commenced, expanding sampling efforts becomes very difficult. Additionally, trend detection usually requires an extended period of time. In the short term, deficiencies of a sampling design may go undetected before an indicator has fallen below a critical level (see also Chapter 2).

In this chapter, we illustrate the importance of considering future uncertainty and demonstrate the use of a simulation approach in the design of monitoring surveys. We develop a simulation strategy that emulates the effects of future agents of change on the annual observations of a spatially explicit indicator population, and that has the ability to apply and compare the performance of multiple survey designs. This approach does not employ complex ecosystem modeling or perceptions of future variance structure of a population. Instead, it emphasizes modeling, in a relatively simple manner, the salient effects of key agents of change on the spatial and temporal dynamics of an indicator. Future variance of the indicator population, and ultimately of survey designs, derives from these modeled effects. We show that this simulation-based comparison of competing designs under plausible future variance helps to inform the selection of a survey design most robust to future uncertainty. To introduce this simulation strategy, we provide an overview of our recommended approach, followed by an extended example focusing on comparison of alternative survey designs for monitoring grasslands in Canyonlands National Park, Utah, USA.

Design and use of simulations of future uncertainty in assessments of survey designs

We first describe several aspects of the strategy we recommend in developing a simulation strategy for comparing monitoring designs while accounting for future uncertainty.

(i) Any evaluation of candidate survey designs for long-term monitoring should be seen as an integrated step in development and implementation of a monitoring study (Chapter 2). In our context, we assume that overall monitoring objectives, target population, general analytical approach to be used for assessing status/trend once monitoring commences, response design (measurement approach at each site), sample frame, and quantitative sampling objectives have been carefully defined, and that suitable estimates of current status and variance components have been obtained (Chapters 2, 7, 8, 9).
(ii) With our approach, the population to be monitored is represented in a spatially explicit manner. The minimum spatial unit represents the size of units in the sample frame, and therefore the size of sites (plots) to be used in the actual monitoring effort. This is not a general requirement of simulation approaches in monitoring design. However, when agents of change are expected to have disproportionately higher effects on some portions of the target population than others but these subpopulations are not partitioned in separate strata, spatially explicit simulations tailored to the population of interest will be most informative.
(iii) The dynamics of a population are simulated using estimates of initial status, annual variability, and trend. These parameters determine the future spatial pattern of an indicator population given contemporary ecological processes, and are referred to as baseline parameters. Parameters are distributed across each frame element and used to simulate annual observations at the frame-element level.
(iv) Key agents of change are selected for simulation assessments on the basis of their potential to have a large effect on an indicator, or because they are a high concern to management (Box 10.2). Simulated properties of change agents include effect, frequency, extent, and pattern. Effects are approximated by modifying the

Box 10.2 Common challenges: simulation choices

In the context of evaluating survey designs based on current and potential future levels of variability, there are three key components of simulation assessments that can be problematic. These include the specification of the initial spatial pattern of an indicator, baseline population parameters which represent the current population of an indicator, and key change agents for assessments. Detailed information on the initial status of an indicator over the target population often is lacking. Results from pilot monitoring or from relevant research studies may provide site-level estimates of an indicator, which may be spatially interpolated to the target frame with kriging or other statistical methods. In some cases, remotely sensed information may be used to spatially delineate relative condition of an indicator (e.g. low, medium, high percent cover), and relationships between these conditions and indicator values used to assign initial status values. Where data are lacking, initial status may be estimated based on biophysical properties of the sample frame given understanding of how these properties influence the indicator. Similarly, baseline population parameters may be estimated from pilot monitoring and other research efforts, and aligned with biophysical properties of the target area. Iterative evaluation of population parameters tends to be necessary. A future projection over the time interval of an assessment using initial estimates is first scrutinized to determine reasonability of spatial patterns, based on historical patterns or trends. In the absence of historical data, professional judgment alone determines reasonability. As necessary, parameters are modified and subsequent projections are used to determine further modifications.

Considering all possible future change agents and all possible permutations of their properties (effect, frequency, extent, and pattern) adds considerable time and cost to a simulation assessment. Also, results from extensive permutations can be difficult to organize, and comparisons among the different combinations can be tedious and confusing. Efficient use of a simulation approach requires a parsimonious number of simulation runs. Informed judgment is thus critical to limit assessments to change agents with the greatest potential to influence future indicator properties, and to salient properties of change agents. Monitoring programs typically use conceptual models of disturbance and stressor dynamics to synthesize current empirical understanding, to aid in understanding potential system change, to incorporate management goals and objectives, and to inform indicator selection (e.g. Britten *et al*. 2007; Chapters 2, 22). These models are a valuable source for selecting the key agents for simulation assessments. Ideally, models are developed with stakeholder input, but where this is not the case, stakeholders should have the opportunity to identify change agents of greatest concern. Also, conceptual models that illustrate causal linkages can inform the types and value ranges of change-agent properties to include in an assessment. Focusing on a limited number of change agents and varying properties facilitates interpretation of results, and expedites the survey-design assessment process.

variability of annual values of an indicator or overall trend. The effects, temporal frequency, aerial extent, and pattern (aggregated, dispersed, random) are varied over a gradient of assumed possibilities to account for future uncertainty. This gradient may represent perceptions of nominal to worst-case or severe conditions. Future projections of an indicator population are then simulated.

(v) Survey designs sample a simulated population similarly to real-world monitoring. Survey designs specify the number of sample sites, the revisitation schedule of sites, and the spatial locations of sites (relative to the sampling frame), and are used to extract annual observations from the (simulated) target population. In our example, the measurement design and effort for each site visit was fixed, but alternative measurement approaches (e.g. producing higher or lower residual variability due to measurement error) could be easily incorporated into the simulation approach. Multiple designs are applied to the same simulated population for comparisons of performance relative to the specified sampling objective.

(vi) Parsimonious summaries of major patterns observed in these comparisons are produced to facilitate discussion with managers and administrators about advantages and trade-offs of alternative designs. Again, this is a general consideration for any quantitative examination of design alternatives.

The key advantage of a simulation approach is its flexibility. In our context, this advantage is the ability to explicitly represent patterns of future possible changes based on current system understanding and to assess survey-design performance in light of these potential changes. Simulating possible changes in a spatially explicit manner provides estimates of potential future indicator variability as well as realistic assessments of the accuracy of sample variance estimates under assumed change scenarios. This aids in assessing the adequacy of designs to sample patterns of future change, and overall, to satisfy the sampling objective. Survey designs selected from this procedure are likely to be more robust to future changes relative to designs based solely on historical estimates of variance.

Simulation assessment example

In the following example, we demonstrate how simulation assessments can be used to assess which survey designs will best meet sampling objectives in the face of future uncertainty. Our case study focuses on monitoring in Canyonlands National Park (CANY), which encompasses 136 610 ha in southeast Utah (Fig. 10.1). Given the potential impacts of social trailing on soil erosion and degradation of herbaceous and shrubland ecosystems in this dryland system, the NPS Northern Colorado Plateau Inventory and Monitoring program (NCPN) selected grassland and shrubland ecological site types for monitoring (O'Dell *et al*. 2005). Our example focuses on the evaluation of survey designs for perennial grass cover in the combined Desert and Semidesert Sandy Loam (four-wing saltbush) ecological site types, which have similar soil and vegetation properties and represent the majority of native grasslands in CANY (Fig. 10.1).

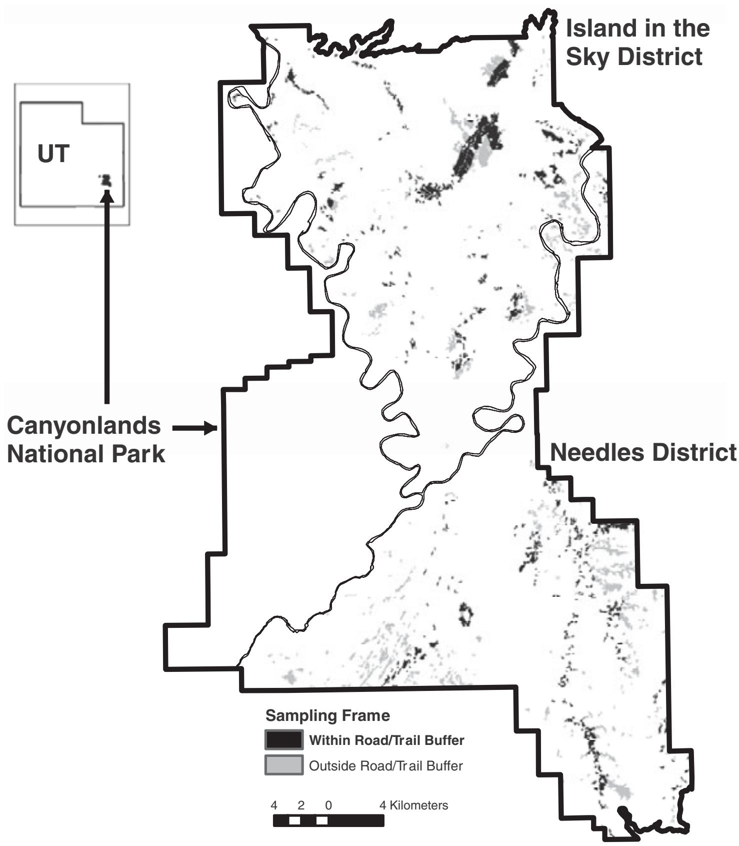

Figure 10.1 Location of Canyonlands National Park, Utah, USA, and the sample frame for the example presented in this chapter. Frame elements are additionally delineated by distance from roads and trails (≤ 500 m, > 500 m).

The target population consists of 5064 ha in these site types accessible by foot and within 4 km of a road or trail. The sample frame was derived using procedures developed by Garman *et al*. (2010). The target population was gridded using an element (i.e. cell) size of 0.25 ha, which equates to the size of an NCPN monitoring site. Using transportation network data layers, sample-frame elements within 500 m of roads and trails (ca. 48% of the sample frame) were delineated to identify regions with different potential for human disturbance.

Simulation system and initialization

We developed a customized system to simulate the annual observations of a single indicator (e.g. grass cover) across a user-provided gridded sample frame, and to simulate impacts of future change on an indicator. Annual observations are generated from estimates (mean and standard deviation) of initial status, annual fractional change (log-linear slope), and variability around a log-linear trend line (Root Mean Squared

Error – RMSE). Means and standard deviations define distributions which are then sampled to derive parameter values for each frame element.

For this study, we estimated population parameters for percent grass cover using pilot monitoring data from Schelz (2002) and Witwicki (2010), and distributed parameter values to mirror patterns generated from kriging contemporary samples of herbaceous cover. Baseline parameters were mean = 15 (SD = 8.5) for initial status (% perennial grass cover); 0.0 (0.01) for log-linear slope; and 0.30 (0.13) for RMSE. Baseline parameters of individual frame elements were then modified to emulate impacts of change agents. Any number of survey designs can be applied to extract annual observations from a simulated population. Extracted sample observations are used to determine agreement between sampled and population variance, and in analyses of status and trend.

Our system is stochastic, in that the variability of annual observations and the implementation of change-agent properties involve a random component. Multiple replicates of a simulated population are generated using different random number seeds to provide a range of possible trajectories. Similarly, a survey design is repeatedly simulated using multiple sets of locations generated with different random number seeds to assess the average performance and sample-to-sample consistency of a design.

Survey designs

Proposed design

Besides informing choice of survey design when a monitoring effort is being developed, power/precision examinations can be used to assess benefits of modifying an operational design (see *Discussion*). Our example emphasizes the use of a simulation approach to enhance a design selected using historical variances. We used the survey design proposed by the NCPN for monitoring grasslands in CANY as a starting point. This design was based on programmatic resources and power-for-trend assessments using variance estimates acquired from a pilot monitoring effort (Witwicki 2010). The proposed design includes two groups of panels. One group consists of 7 panels with 3 sites per panel, each of which are visited for 2 consecutive years then revisited after 5 years. The second group consists of 7 panels, each containing 6 sites, which are visited once and revisited after 6 years (e.g. X = 3 and Y = 6 in Fig. 10.2). Across both groups of panels, 12 sites are visited annually and the total number of sites is 63; this proposed designed is referred to as 2PS_12 (Table 10.1). This type of design, a split-panel design, performs well, compared to alternatives, for both estimation of trend (emphasized by panels 1–7 in Fig. 10.2) and status (emphasized by panels 8–14 in Fig. 10.2) (Urquhart *et al*. 1998, Breidt and Fuller 1999, Urquhart and Kincaid 1999, McDonald 2003; Chapter 7). Note that in the first group of panels, two panels are visited each year, resulting in an overlapping pattern of visits to panels. This pattern is one approach for ensuring that the revisit design is “connected” in an experimental design sense (see Chapter 7).

Alternative designs

We generated seven additional survey designs that met the following practical constraints and programmatic goals of the NCPN. First, only designs with ≤15 sample sites per

Table 10.1 Split-panel survey designs evaluated in the case study. Notation for Design Code is described in Fig. 10.2. Short-hand notation of Revisit Design (i.e. panel plan) was adapted from McDonald (2003), and is defined as follows: The sets of paired parentheses designate the temporal sub-design for each of the two or three groups of panels in each split-panel design. For each group, the first number inside the paired parentheses is the number of consecutive years a panel is sampled and the second number is the number of years between revisits; the superscript is the number of panels in that group and the subscript is the number of sites per panel.

Design code	Revisit design	Sites per year	Total unique sites
2PS_12	$[(2–5)^7_3, (1–6)^7_6]$	12	63
2PS_13	$[(2–5)^7_3,(1–6)^7_7]$	13	70
2PS_14	$[(2–5)^7_4,(1–6)^7_6]$	14	70
2PS_15	$[(2–5)^7_4,(1–6)^7_7]$	15	77
3PS_12	$[(2–5)^7_2,(1–6)^7_6, (3–3)^2_2]$	12	60
3PS_13	$[(2–5)^7_2,(1–6)^7_7, (3–3)^2_2]$	13	67
3PS_14	$[(2–5)^7_3,(1–6)^7_6, (3–3)^2_2]$	14	67
3PS_15	$[(2–5)^7_3,(1–6)^7_6, (3–3)^2_3]$	15	69

Year ⇨ Panel ⇩	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	X	X						X	X						X	X				
2		X	X						X	X						X	X			
3			X	X						X	X						X	X		
4				X	X						X	X						X	X	
5					X	X						X	X						X	X
6						X	X						X	X						X
7	X						X	X						X	X					
8	Y							Y							Y					
9		Y							Y							Y				
10			Y							Y							Y			
11				Y							Y							Y		
12					Y							Y							Y	
13						Y							Y							Y
14							Y							Y						
15	Z	Z	Z				Z	Z	Z				Z	Z	Z				Z	Z
16				Z	Z	Z				Z	Z	Z				Z	Z	Z		

Figure 10.2 Structure of the two types of split-panel designs evaluated in this study. The first type consists of seven panels (X) with sites in each panel visited two years in a row and revisited after five years and seven panels (Y) with sites in each panel visited for one year and revisited after six years. Survey designs with this panel structure are designated for the case study as 2PS_##, where the "2PS" indicates two types of panel-revisit structures, and ## is the number of plots visited annually (i.e. 12–15). The second type, designated as 3PS_##, has the same two sets of seven panels, but also has a third panel-revisit structure (Z) in which two additional panels are visited for three consecutive years and revisited after three years. More generally, adapting notation from McDonald (2003), the structure of the first design can be indicated as $[(2–5)^7, (1–6)^7]$, while that of the second design can be indicated as $[(2–5)^7, (1–6)^7, (3–3)^2]$ (Table 10.1).

year were feasible given NCPN budgetary limitations. Second, annual visitation was not desirable because plot fatigue due to constant soil disturbance by monitoring crews is a serious concern in the dryland systems of the Colorado Plateau. Third, under the scenario that population variance for percentage grass cover would remain at current estimated values, the designs had to satisfy a sampling objective defined by the NCPN to provide early detection of possibly 'important' trends (Witwicki 2010). This objective is to detect an annual decline of 3.5% within 10 years with power $\geq$ 80% and $\alpha = 0.10$.

Three of the additional designs were generated by adding one, two, and three sample sites annually to the NCPN proposed design (Table 10.1). Adding one additional site to panels 8–14 emphasized status estimation (2PS_13). Adding one additional site to the first seven panels emphasized trend estimation (2PS_14). Adding one additional site to each panel of the proposed design increased the annual sampling effort by three, and resulted in a design with the highest total number of sample sites of all designs considered (2PS_15).

The other four alternative designs were a variant of the proposed split-panel design that included a third panel-revisitation structure (Fig. 10.2). This structure consisted of sites visited in three consecutive years and revisited after three years. Two panels of this structure were added. Starting with a 12 site per year design (3PS_12, Table 10.1), additional sites were added, with a maximum of 15 sites visited per year (3PS_15). The revisitation schedule of the third panel structure was employed because increasing revisitation frequency increases power to detect trends (Urquhart *et al*. 1998, McDonald 2003). For the same annual sampling effort, these designs had fewer total number of sample sites than designs with only the two types of panel structures (Table 10.1).

All survey designs incorporated a Generalized Random Tessellation Stratified (GRTS) spatially balanced probability design (Stevens and Olsen 2004; Chapter 6). We generated samples and panel assignments with equal probability of selection, using R package "spsurvey" (Kincaid *et al.* 2009, ver. 2.1; see Chapters 6, 14). For each survey design, we used 100 replicates of simulated samples in each scenario.

Modeling future scenarios

Disturbance scenarios

Climate variability and human-mediated impacts to soil stability are key processes likely to influence future conditions of grasslands in CANY. Soil compaction from social trailing, off-road vehicles, and even occasional trespass livestock grazing promotes soil erosion and loss of native grass cover (Miller 2005). We modeled reduction in grass cover due to trampling by first randomly selecting a sample frame element within 500 m of a road or trail, then expanding to a 1-ha patch which could form beyond this distance. Disturbed elements within the 500 m buffer were assigned a log-linear slope of –0.14; otherwise elements were assigned a slope of –0.10. A higher rate of decline near roads and trails reflects a tendency for social trails to initiate along transportation networks and to be repeatedly used. These slope values represent a severe loss of grass cover in a relatively short period of time. We simulated six levels of disturbance (0.5, 1, 10, 20, 30, and 40% of the sampling frame) over a 20-year period. This gradient provided a context for evaluating design performance from nominal to extreme occurrence of disturbance. Impacted frame elements were selected, and slope values were modified in the first 5 years of a simulation run. Twenty replicates were generated for each of the six disturbance scenarios.

Climate scenarios

Impacts of climate variability over the next 20 years were modeled using Global Circulation Model (GCM) projections. We used 12-km downscaled climate projections from the NOAA Geophysical Fluid Dynamics Labs' GFDL CM2.1 model, A2 emission scenario (http://cascade.wr.usgs.gov/data/Task1-climate/index.shtm), to generate a future climate signal. We converted the downscaled daily temperature and precipitation values to annual measures, and used the WebWIMP model (The Web-based, Water-Budget, Interactive Modeling Program – http://climate.geog.udel.edu/~wimp) to generate an annual drought index (Willmott and Feddema's Moisture Index; Willmott and Feddema 1992) for each of the next 20 years for each GFDL grid cell. We derived the proportional difference between a future annual index and the historical long-term mean (i.e. yearly deviations) and input this information to our survey simulator as a geospatial data layer.

Within the simulator, the 12-km GCM grid cell covering a sample frame element was determined, and the corresponding climate time series was used in the derivation of annual observations. To emulate effects of climate extremes, we imposed coherence of observations whenever the yearly deviation in the drought index exceeded a threshold level; otherwise, annual observations varied independent of climate. When the threshold was exceeded, a scaled variant of the difference was added to the annual observation of all affected frame elements. Based on exploratory assessments, we selected two thresholds to emulate grass cover response to climate extremes. Annual observations were affected when the absolute proportional difference between the annual and long-term drought index was >0.30 and >0.20, where the lower threshold value represents high sensitivity. These thresholds correspond to about 20% and 33% of the 20 years experiencing a climate-extreme signal, respectively. Twenty replicates were simulated for each threshold level.

Changes in variance structure and accuracy of variance estimates from alternative designs

Assessing variance components

Site, year, and residual variance components influence status and trend estimation in the linear mixed-effects model used as the framework for our calculations (Larsen *et al*. 1995, 2001; Urquhart *et al*. 1998; Urquhart and Kincaid 1999; Kincaid *et al*. 2004; Chapter 7). As also specified in Chapters 7 and 9, site variance reflects inherent variation among sites that is consistent across years (i.e. persistent site effects). Coherent year to year variance reflects population-wide year effects, excluding systematic effects of a linear trend. Residual variance includes unexplained process variance (e.g. unexplained variation in trend among sites, Chapter 7) and measurement variance.

A linear mixed-effects model was used to derive population variance components of simulated populations to determine the effects of modeled change agents on the underlying variance structure, and to derive estimates from each replicated simulated survey to assess level of agreement between survey design estimates and actual population variance. The latter determines a design's ability to capture the actual population variance.

The linear mixed-effects model used for estimation of variance components mirrored the "one trend" model specified in Chapter 7 [Equation (7.7)], with "time" as a numeric fixed effect, and "site" and "year" as random effects. We used the following mixed-effects model ("PROC MIXED", SAS Institute 2001) to estimate the three variance components.

```
Proc Mixed Method=REML;
      class site year;
      Model LnCover = time/DDFM=KENWARDROGER;
      random site year;
      Run;
```

The SAS commands listed above and in Chapter 7 are very similar. However, we used Restricted Maximum Likelihood (REML) estimation (see Chapter 9), under assumed normality of random effects. Due to unbalanced designs (each sample site did not have observations in each year), the Kenward–Roger method was used to derive degrees of freedom (Spilke *et al*. 2005). Natural-log transformation of percent cover (LnCover) enhanced normality and constancy of variance. "Time" was included to de-trend observations because the coherent year variance component is tied to random year effects apart from systematic trend (see also Chapters 7, 9).

Comparison of survey designs

In disturbance scenarios, population-wide site and residual variance components increased with increasing disturbance level (Table 10.2). In general, sampled estimates of variances were within 5–8% of population values for disturbance levels ≤10%. Agreement with population site variance increased with increasing sampling effort per year. For simplicity, we present results only for designs with the lowest (12 sites/year) and highest (15 sites/year) annual efforts throughout the rest of the chapter. Using 10% disturbance as a worst case, all designs reliably sampled the population variance, with best performance from designs measuring 15 sites per year.

Under climate scenarios, the coherent year variance of simulated populations increased with increasing sensitivity of annual observations to a climate signal, while mean site variance decreased slightly with increasing sensitivity (Table 10.2). Similar to the disturbance scenarios, mean sample estimates of variance components were within 5–10% of population values. For both threshold levels of sensitivity to climate change, sample means for site and year variances decreased with increasing annual sampling effort.

Table 10.2 Mean (standard error) estimates of variance components from simulated populations (20 replicates per scenario) for the disturbance and climate scenarios. Disturbance levels are expressed as percentage of the sample frame. Climate thresholds are threshold levels to invoke a climate effect on simulated annual observations. Values were derived from a log-linear, mixed-effects model with site and year as random effects, and time as a fixed effect to de-trend observations. Each replicate contained 20 years of simulated, annual values of percent grass cover for each of the 20 256 sample frame elements.

	Variance component		
Scenarios	Site	Year	Residual
Disturbance level			
0.5	0.601 (0.001)	0.000 (0.000)	0.102 (0.000)
1	0.610 (0.001)	0.000 (0.000)	0.105 (0.000)
10	0.730 (0.012)	0.000 (0.000)	0.158 (0.001)
20	0.810 (0.020)	0.000 (0.000)	0.200 (0.001)
30	0.852 (0.028)	0.000 (0.000)	0.229 (0.001)
40	0.870 (0.036)	0.000 (0.000)	0.245 (0.002)
Climate threshold			
> 0.30	0.572 (0.002)	0.003 (0.000)	0.100 (0.000)
> 0.20	0.561 (0.001)	0.016 (0.000)	0.100 (0.000)

In both disturbance and climate scenarios, all survey designs estimated variance components fairly accurately. This was observed despite the strong additional population-level variability produced in disturbance scenarios, in which trampling disturbance tended to be aggregated along transportation networks. Without explicitly addressing this spatial pattern, the sampling approach adequately captured population variability. This may be a result of the dispersion of transportation networks throughout the target population (Fig. 10.1) as well as the efficiency of spatially balanced sampling. In addition, the total sample sizes relevant to each variance component were relatively high (based on results in Chapter 9), with relevant sample sizes of > 60 sample sites for capturing site variance, 20 years for capturing year variance, and 240–300 total site-visits for capturing residual variance.

Assessing performance in relation to the sampling objective

Assessing power and precision

In our example, we used simulations to project the current estimated population forward and to assess multiple realizations of each survey design, while we calculated power and precision of status estimates following the analytical framework of Urquhart *et al*. (1993, 1998) and Larsen *et al*. (1995, 2001). The assumed analytical model is the general linear mixed-effects model described above. Chapter 7 summarizes how the year, site, and residual variance components contribute to the variance of the estimate of the linear trend, and therefore affect power to detect trend. With this analytical model, estimates of the slope and intercept of the linear trend line can be used to calculate a model-based estimate of expected status in each year. The standard error (SE) of this status estimate is derived from the variance-covariance structure of the linear regression coefficients.

We used a set of functions developed in the R language by Tom Kincaid, US Environmental Protection Agency, Corvallis, Oregon, to calculate power and precision of status estimates for each year of simulated monitoring for each scenario, survey design, and replicate. Site, year, and residual variance components were specified as described above. Use of log-scale variance estimates was appropriate given our focus on a log-linear trend. The general variance structure for the mixed-model also includes site and year correlation components, which were set to 1 and 0, respectively [see Larsen *et al.* (1995) and Chapter 7].

To simplify comparisons among designs, we used the minimum year to achieve a power of $\geq 80\%$ for detecting an underlying annual –3.5% trend, with $\alpha = 0.10$ as a measure of performance. This measure reflects the desire to achieve at least 80% power to detect this level of change within 10 years to provide early warning of undesirable change (Witwicki 2010). We generated a mean (and standard error) for the minimum year at which 80% power was achieved by averaging values across the replicates of a design-scenario combination (100 replicates of a survey design $\times$ 20 replicates of a scenario). For each set of extracted observations, we derived the average standard error of status over the 19-year period, and then derived a mean average across the replicates of a design-scenario combination. We used an ANOVA of ranked values and Scheffe's multiple contrast for statistical comparison of mean minimum year to achieve the target power, and of mean average standard error of status among the eight survey designs.

Power and precision of alternative designs

In disturbance scenarios, the mean minimum time to achieve the target power (Fig. 10.3a) and mean average model-based standard error (Fig. 10.3b) increased with increasing population variability. Within each disturbance level, mean minimum time to detect the –3.5% trend only differed at most by ca. 1 year and mean average standard error differed by 0.017 among designs. Although these were minor differences in mean values among designs, both the minimum time and standard error means were significantly different ($P < 0.05$) among designs with different annual sampling effort, with mean values decreasing with increasing number of annual samples. That is, with higher annual effort, desired power was obtained more quickly, and status estimated more precisely. Of the designs evaluated, the 15 sites per year designs appear to be the most robust to trampling effects. Although the differences were slight, only designs with 15 sites per year consistently achieved the target power level within 10 years for $\leq 10\%$ disturbance levels (Figs. 10.3a). Above 10% disturbance, all designs failed to achieve the target power within a decade.

The survey-design assessments for disturbance scenarios illustrated the importance of the power and time-frame components of the sampling objective. We used a 10-year time frame for achieving 80% power, which was based on ecological and management considerations (Witwicki 2010). If 80% power within 12 years was deemed sufficient, all designs we evaluated would satisfy this objective for all simulated disturbance levels

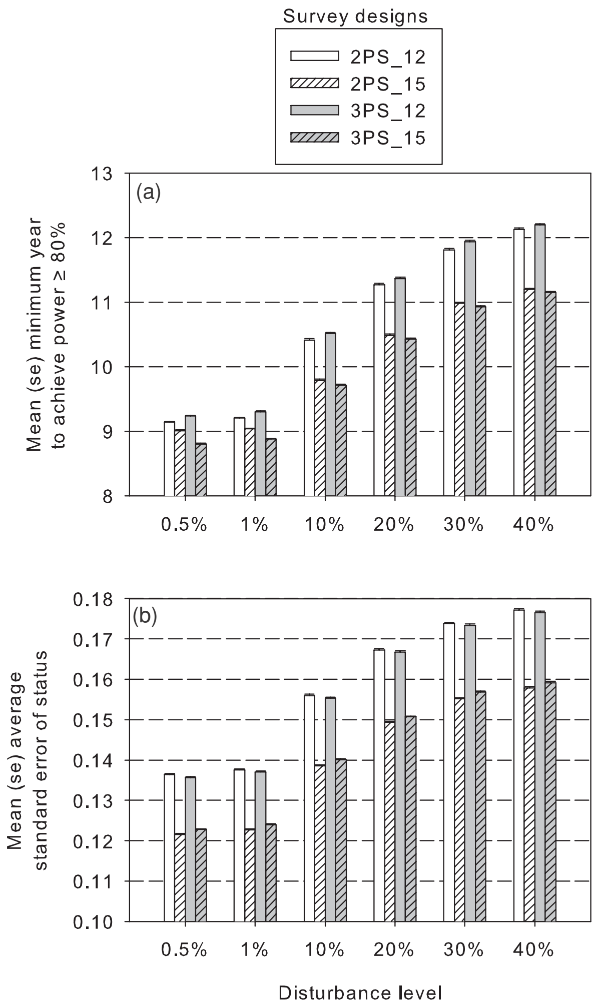

Figure 10.3 (a) Mean (standard error) minimum year to achieve a power of ≥ 80%; (b) 19-year mean (standard error) average standard error of status for survey designs for the disturbance scenarios. Values for designs were generally linear between the 12 and 15 plots per year designs. For visual clarity, only results for designs with 12 and 15 plots per year are shown. Power and standard error of status were based on a –3.5% trend and $\alpha = 0.10$. Means were derived from 100 replicates of sample locations for each survey design × 20 replicates of a disturbance scenario ($n = 2000$).

(Fig. 10.4). If 90% power was the target level, all designs would reach this target within 14 years.

Precision of status estimates responded primarily to differences in annual sampling effort among designs (Fig. 10.3b), but the type of split-panel design had minor effects. Although the 19-year mean averages were similar, the revisitation schedule of the 3PS_15 design provided higher precision of status (lower mean values for standard errors) in the first 5 years compared to the 2PS_15 design (Fig. 10.5). After year 8, the

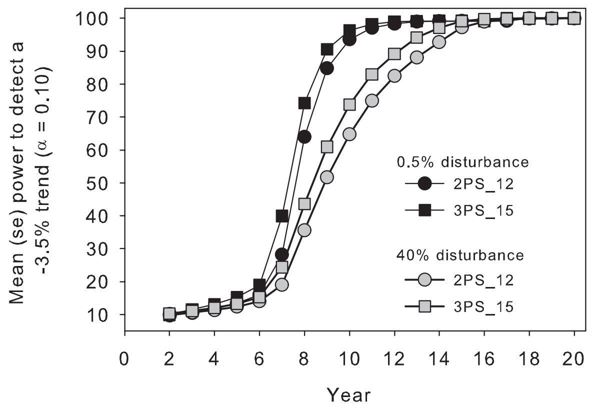

Figure 10.4 Mean power for two survey designs for the nominal and extreme disturbance scenarios. Designs bracket the minimum (2PS_12) and maximum (3PS_15) power across all designs. Means were based on $n = 2000$. Standard error bars of means are too small to be visible.

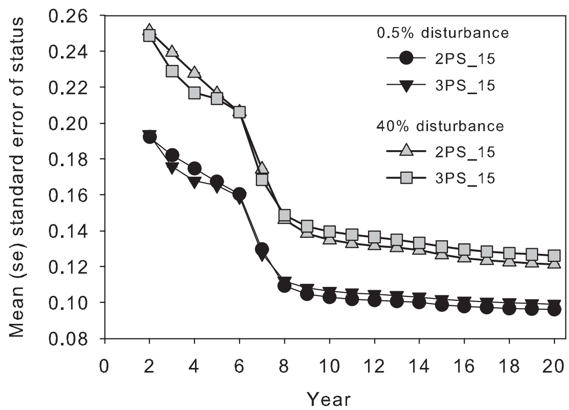

Figure 10.5 Mean standard error of status for survey designs with 15 plots per year for the nominal and extreme disturbance scenarios. Curves illustrate temporal differences between designs, where means for the 3PS designs were lower in the first 7 years then higher in the remaining years compared to 2PS designs. Means were based on $n = 2000$. Standard error bars of means are too small to be visible.

larger overall sample size of the 2PS_15 design (Table 10.1) resulted in slightly higher precision.

For climate scenarios, comparative power and precision showed patterns similar to those for the disturbance scenarios. The minimum year to achieve the target power (Fig. 10.6a) and mean average standard error of status (Fig. 10.6b) increased with increasing population variance for all survey designs. These measures for the two designs with 15 sites per year were not significantly different ($P > 0.05$), but were significantly less ($P < 0.05$) than means of other designs. With the lower sensitivity threshold, only the designs with 15 sites per year satisfied the sampling objective by year 10 (Fig. 10.6a). At the higher level of climate sensitivity, no designs satisfied the sampling objective.

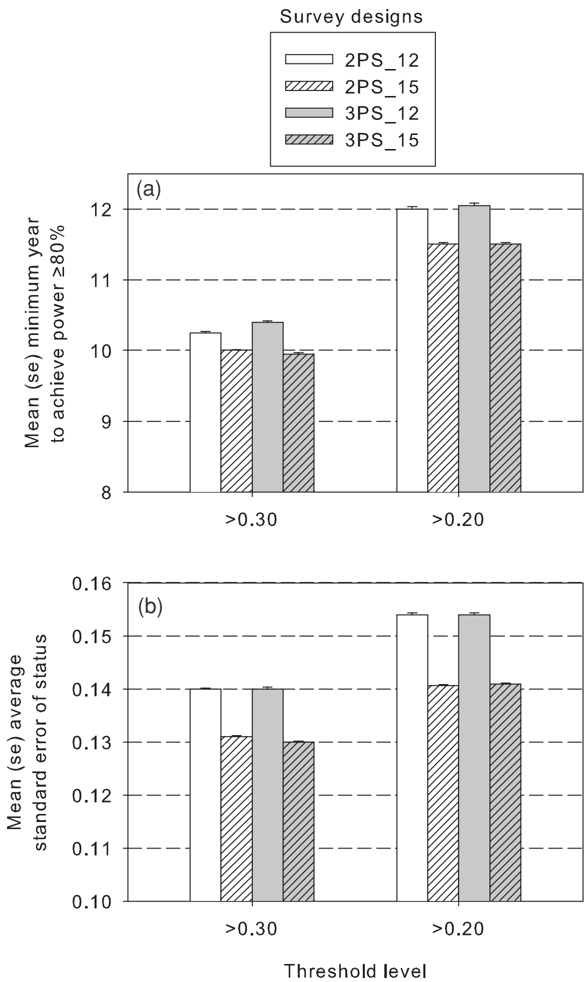

Figure 10.6 (a) Mean (standard error) minimum year to achieve a power of $\geq$ 80%, and (b) 19-year mean (standard error) average standard error of status of survey designs for the two threshold levels used to invoke a climate effect on simulated observations. Values for designs were generally linear between the 12 and 15 plots per year designs. For visual clarity, only results for designs with 12 and 15 plots per year are shown. Power and standard error of status were based on a –3.5% trend and $\alpha = 0.10$. Means were derived from 100 replicates of sample locations for a survey design $\times$ 20 replicates of a climate scenario ($n = 2000$).

In addition to examining precision of status estimates based on a linear-model framework, for all scenarios we also examined confidence-interval coverage for a design-based estimator of status. We calculated Horvitz–Thompson estimates of population means and corresponding 95% confidence intervals using the function *total.est*() in the "spsurvey" package, with confidence intervals calculated based on the local neighborhood variance estimator (Stevens and Olsen 2003; see Chapters 6, 7, and 14). Coverage of confidence intervals was calculated at years 10 and 20. For all survey designs and scenarios, >95% of the 95% sample confidence intervals contained the true population mean.

Implications for monitoring at Canyonlands National Park

In this example we limited our assessments to two key change agents, and only evaluated alternative survey designs that would be operationally viable by the NCPN, based on current programmatic limits. The designs evaluated were not substantially different in terms of annual and total sampling effort, and as expected, did not widely differ in terms of absolute differences of the power and status metrics we evaluated.

However, a key finding was that only the two designs with 15 sites per year consistently achieved the target power level within 10 years for $\leq$10% disturbance levels and with the lower climate sensitivity threshold. This increase in sampling effort of three sites per year also produced the highest precision. These results provide incentive to consider one of the two designs with 15 sites per year to ensure that the sampling objective is satisfied over plausible, near-term changes in percent grass cover. Differences in temporal precision, and plot fatigue potential of the three-year visitation scheme would factor into the choice between the 2PS_15 and 3PS_15 design for implementation. From a programmatic perspective, the added costs of these designs must be weighed against the benefit of achieving the sampling objective only ca. one year sooner than the proposed design for worst-case conditions. The time frame specified in a sampling objective is ideally selected to ensure early warning of degrading conditions. Not exceeding this time frame may be essential to avoid threshold behaviors (i.e. a rapid, nonlinear decline in an indicator) and costly restoration.

The spatially balanced GRTS design used for selecting sample sites in our example accommodates the addition of sites to the design without compromising spatial balance (Chapters 5, 6). Logistically, expanding an existing monitoring effort by three sample sites per year may appear trivial. In practice, monitoring programs are likely to monitor numerous target populations, and budgetary limitations may require convincing evidence for even nominal expansion of any sampling design. For instance, there are five target populations of terrestrial vegetation in Canyonlands NP and different survey designs for each, and Canyonlands NP is just one of 16 park units in the NCPN program. In addition to terrestrial vegetation monitoring, 12 other monitoring protocols are funded by the NCPN (see http://science.nature.nps.gov/im/units/ncpn). Simply adding three sites per year to each design may help safeguard against future uncertainty, but the additional workload and cost of even a nominal increase for each target population can be surprising large. Comparable "trivial" increases across all target populations and protocols in the NCPN would far exceed programmatic resources. Results from examinations of current designs based on results to date and from simulation assessments of future uncertainty may be a necessary incentive to expand even a few existing designs by one to three sites per year.

Discussion

Flexibility is a key advantage of a simulation approach for developing or modifying sampling designs. In our example, a simulation approach allowed us to explicitly represent patterns of future possible changes based on current system understanding and to assess survey-design performance in light of these potential changes. Adjustments of current variance estimates by professional judgment to account for future uncertainty generally are ad hoc, and lack consideration of patterns of future change. Spatially explicit simulations potentially provide more insight into potential variability under assumed change scenarios. Survey designs selected from this procedure are likely to be more robust to future changes relative to designs based solely on historical estimates of variance. A simulation approach to assess potential future variability is especially important in situations where systems are highly susceptible to known or suspected agents of change, there is high probability of future change due to these agents, and delayed detection of trends estimates can lead to substantive loss of ecological integrity or high restoration costs.

Assumptions of how future change agents may affect an indicator population can be crafted to evaluate different levels of future uncertainty. Such assumptions are necessary to account for uncertainty in the effects, frequency, extent, and pattern of future change agents. Expert opinion, professional judgment, stakeholder concerns, and objectives of an assessment determine the range of assumptions employed in an assessment. Assumptions may reflect perceptions of reasonable change based on judgment, historical rates, or proposed land-management plans with the potential to increase disturbance (e.g. increased human disturbance around proposed campgrounds, roads, and trails). They may also represent an upper level of change that may or may not be realistic, but ensure the selection of an efficient survey design for a wide spectrum of future change. For example, in our work in grasslands we have incorporated a range of potential trends to emulate trampling effects; however, the two values (10 and 14% annual decline) we currently use and reported on in our example are considered upper limits on plausible change. Trampling effects over >10% of the target population are likely unrealistic, but assessments using these extreme disturbance levels help to identify limitations of survey designs. Using 10% disturbance as an upper level for design assessments provides a stringent but not excessive criterion, and ensures that the selected design is likely to satisfy sampling objectives up to plausible levels of future change, based on current assumptions of what is "plausible". Varying each change-agent property across a range of values deemed reasonable in the future, and simulating all permutations of these values provides a gradient of plausible future outcomes. A focus on the most severe plausible impacts of change agents may be used to understand the implications of underestimating a worst case scenario, which may encourage selection of a highly robust survey design.

More generally, analytical and simulation assessments to inform selection of a monitoring survey design are ideally performed in the initial planning phase. For each target population, comprehensive assessments could employ a wide range of designs unconstrained in terms of annual and total sampling effort, but potentially reflecting other practical constraints. For instance, only revisitation schedules with low potential for plot fatigue (effects on a site's status or trend due to frequent disturbance by monitoring personnel) may be included. Varied spatial properties of candidate designs can help determine the need for stratification or the implications of using unequal probabilities in sample site selection, given current variability and future potential changes in indicator variance. For each target population, assessments can help establish optimal survey designs.

These assessments also may be instrumental in determining the feasibility of monitoring certain indicators given estimates of current and future uncertainty. Where sampling effort of an optimal design exceeds programmatic resources or requires an inordinate amount of effort relative to programmatic goals, the indicator may be dropped from consideration. Alternatively, attributes of the sampling objective (e.g. precision, power) may be modified – i.e. expectations may be lowered if this does not compromise the underlying monitoring objectives.

When operational monitoring is already in place, power/precision/effort assessments can still provide important evidence for modifying designs. As in our example, the range of alternative designs considered may be restricted to those that are moderate modifications of the current design. Broad-scale monitoring programs frequently need to monitor numerous indicators and target populations, but this need must be balanced with the need to collect meaningful information for each indicator and population. Assessments across a suite of existing designs can identify those in greatest need of enhanced sampling effort. Providing evidence for adding even a nominal number of samples to a few existing designs may be critical given budgetary limitations.

Accounting for future possible changes in indicator patterns and variance aids in the selection of a survey design that is robust to future uncertainty, but attention to actual design performance over time is still critical. Future dynamics of an indicator may result in spatial and temporal variance more extreme than anticipated or what was perceived to be plausible. Over the long term, adaptive monitoring will always be critical to ensure a sampling strategy continues to satisfy the sampling objective (Ringold *et al*. 1996, Lindenmayer and Likens 2009). Quantitative power/precision examinations can help ensure selection of a survey design that will be effective based on current variation, while simulation-based sampling of projected future populations helps to safeguard against plausible, near-term (10+ years) changes in indicator variance. In practice, these examinations ensure acquisition of reliable status and trend information while providing sufficient time to determine necessary adaptive changes to a sampling design.

Future research and development

We are not aware of an existing user-friendly program or package that can simulate spatially explicit dynamics of an indicator population, that allows a user to implement and simulate a range of change-agent properties, and that has the ability to implement a range of survey designs. Even our customized simulation system is not generalized; implementation of different change-agent properties requires modifying C code. Results from extensive simulations can be difficult to organize, and analyses and comparisons of multiple survey designs can be tedious. Automating performance analyses is thus equally important to facilitate use of simulation approaches. Monitoring programs currently seeking to use simulation assessments must develop their own customized system. This requires access to expertise in programming languages or in commercial packages with simulation capabilities. Development of a generalized and flexible package for wide-spread use would help many monitoring programs more efficiently examine candidate survey designs under a range of scenarios, to justify the selected design, and to re-examine the operational design as more information is obtained. Developing such a system in the freely available R language (R Development Core Team 2011) is perhaps the most viable option, particularly given the available packages in R for spatial sampling and analysis (e.g. Chapters 6, 14) and for spatio-temporal modeling.

Summary

A simulation approach is a highly flexible tool for assessing comparative performance of alternative sampling designs to address factors not easily incorporated into analytical power and precision examinations. We demonstrated this flexibility in the context of assessing design performance under potential future changes in population variances. Selection of an optimal sample survey design for operational monitoring tends to be based on historical or contemporary estimates of indicator variance components, with the assumption that variances will not change in the future. Even relatively near-term future changes in the spatial and temporal variance of a monitoring indicator are inevitable due to disturbances, climatic fluctuations, and human-mediated stressors such as climate change. Failure to account for future changes in indicator variance can lead to overestimation of potential survey design performance. We recommend using simulation approaches to model plausible near-term (1–2 decades) changes in indicator variance, such as those due to disturbances or climate-change impacts, when evaluating survey designs for operational implementation, as well as to enhance existing designs using historical variances. This simulation approach allows the user to assess the potential effects of change agents on the variance, status, and trend of an indicator population across a designated sample frame, and to apply multiple survey designs to sample the affected population. Comparison of these competing designs helps to inform the selection of a survey design most robust to near-term future uncertainty. Over the long term, unpredictable changes in indicator conditions may dictate more substantive changes in a sampling strategy. However, operational survey designs selected on the basis of current variation and under plausible change scenarios are likely to provide reliable status and trend information while providing sufficient time to determine the need for more significant changes to a sampling design.

Acknowledgments

Data for the delineation of grass-cover patterns were provided by the US National Park Service Northern Colorado Plateau Inventory & Monitoring Network (Moab, UT), and

Mark Miller (USGS, Flagstaff, AZ). We are indebted to Scott Urquhart for simulator-design recommendations, and assistance with the analysis and interpretation of survey-design samples. We also thank Lee Ann Starcevich for assistance with variance estimation of unbalanced designs, and Tom Kincaid for the power-for-trend R program. We thank Robert Gitzen and two anonymous reviewers for constructive comments that improved this chapter.