

Designing Long-Term, Large-Scale Forestry Experiments with Research Objectives at Multiple Scales

Lisa M. Ganio
Klaus J. Puettmann

ABSTRACT. A number of large-scale manipulative studies (LSMEs) have been installed in recent decades. They were designed to test operationally practical silvicultural treatments on large tracts of forest land and over long periods of time. The interdisciplinary nature of LSMEs and the associated variety of research objectives provide special challenges in study design and implementation that usually do not occur in the setup of traditionally smaller-scaled research studies. We present and discuss these issues, including the development of a prioritized list of objectives with explicit spatial and temporal scales and clear definitions of the scope of inference for each objective. In this context we discuss the variation within large experimental units; the choice of replications; treatment definitions, including multiple manipulations over time; and the choice, scale, and timing of measurements. Above all, it appears that agreeing on a clear hierarchy of study objectives will prevent future conflicts between different study components and will provide guidance for the evaluation of treatment and measurement choices. doi:10.1300/J091v26n01_01 [Article copies available for a fee from The Haworth Document Delivery Service: 1-800-HAWORTH. E-mail address: <docdelivery@haworthpress.com> Website: <<http://www.HaworthPress.com>> © 2008 by The Haworth Press. All rights reserved.]

KEYWORDS. Experimental design, large-scale silviculture studies, long-term study objectives, multiple scales

Lisa M. Ganio (E-mail: lisa.ganio@oregonstate.edu) and Klaus J. Puettmann (E-mail: Klaus.puettman@oregonstate.edu) are Associate Professors, Department of Forest Science, 321 Richardson Hall, Oregon State University, Corvallis, Oregon, USA 97333.

Journal of Sustainable Forestry, Vol. 26(1) 2008
Available online at <http://jsf.haworthpress.com>
© 2008 by The Haworth Press. All rights reserved.
doi:10.1300/J091v26n01_01

INTRODUCTION

Recent interest in ecosystem response to forest management activities has led to the implementation of “Large-Scale Management Experiments” (LSMEs: *sensu* Monserud, 2002). As defined by Monserud (2002), the primary objective of LSMEs is to examine the effects of silvicultural practices at operational (broad) scales over long periods of time. LSMEs do not necessarily tackle novel questions but they are designed to address management and policy issues at scales at which management occurs, practically eliminating the challenge to scale up research results to operational activities (Bruce, 1977). The large-scale objectives addressed by LSMEs may have been the subject of studies at smaller scales; however, results from smaller-scale studies may not be considered adequate to fully answer large-scale questions.

At the same time, many of these experiments aim to integrate investigations of social, ecological, and economic aspects within a single study. Consequently, interdisciplinary teams commonly develop a diverse list of objectives and design treatment and measurement schedules to meet those objectives. For example, the Demonstration of Ecosystem Management Opportunities study was designed to address management effects on wildlife, including birds and amphibians; community dynamics of ectomycorrhizal fungi; snow melt and rain-on-snow dynamics with respect to canopy densities; and hydrology, as well as economics (Halpern et al., 1999). Participating organizations justify the large expenditures of LSMEs by the need to address the complex management issues through coordinated, interdisciplinary research. Allowing researchers from different disciplines to take advantage of a single experimental setup is an efficient use of resources and at the same time facilitates interdisciplinary cooperation and team building. However, integrating the work of multiple investigators in a single study provides special challenges not typically encountered in short-term, smaller-scaled research studies.

LSMEs generally provide information for multiple objectives, often at different scales within a single study (e.g., Halpern et al., 1999). Ensuring that all objectives are met creates novel challenges in study design. Complications arise because the components of the study design are linked and changes to any one component will likely affect others. Design challenges are also a direct result of combining multiple objectives in a single study setup, with each objective potentially having different “optimal” treatment and measurement scales, that is, scales that provide useful data in an efficient manner. We propose that dealing

with a diverse set of objectives in a single experiment, especially at large scales, not only highlights the importance of proper experimental design, but brings up issues for planning and implementation that are not intrinsic in single-objective studies. Our goal is to provide an overview of challenging opportunities and potential pitfalls when dealing with LSMEs. We discuss the challenge of linking the scope of inference, replications, and treatment and measurement definitions in the face of multiple spatial and temporal scales. We provide suggestions for forging successful linkages. Specifically, we highlight that development of a clear hierarchy of objectives can help guide many important decisions.

METHODS

Hypothetical Example of a Large-Scale Management Experiment

To facilitate this discussion we will refer to the following hypothetical example of an LSME. Suppose that the study is motivated by a desire to evaluate the effects of thinning treatments intended to produce late successional structure and the goal is to investigate responses to thinning and subsequent underplanting of Douglas-fir and western red cedar in mature Douglas-fir forests in the Coast Range of the western United States. The goal can be broken into the following specific research objectives.

Primary Objectives

- *Forestry*: How do responses of overstory trees, understory shrubs, herbs, mosses, and lichens differ among thinning treatments?
- *Economics*: How do the harvesting costs compare between the thinning treatments?

Secondary Objectives

- *Forestry*: How do underplanted Douglas-fir and western red cedar seedlings compare in terms of growth and survival under the different thinning regimes?
- *Wildlife*: How do populations of small mammals and migrant songbirds respond to thinning treatments? If they respond, what component of stand structure determines their response?

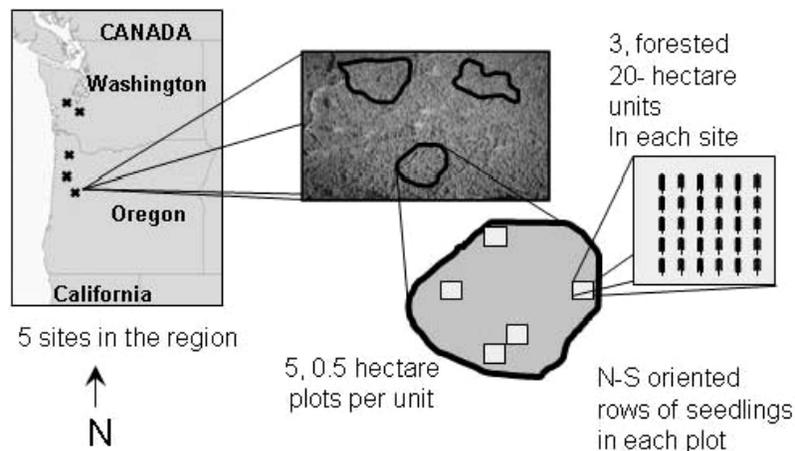
Ancillary Objective

- *Hydrology*: Obtain continuous weather and stream-flow data from an automated on-site weather station and stream gauging station

The hypothetical field study design is described in Figure 1. The stands of interest are approximately 60 years old and about 70 hectares in size, with more than 80% of basal area in Douglas-fir. Land management agencies with holdings in western Oregon and Washington were surveyed for stands that met the selection criteria. Two forested areas in western Washington state and three in the Oregon Coast Range that meet the criteria are randomly selected as study sites. Three 20-hectare units are identified within each forested area. Units are delineated so as to maximize straight-line edges and minimize edge-to-interior ratios, approximately circular. Within each 20-hectare unit, five 0.5-hectare square plots are randomly selected for the underplanting study and plot edges are aligned along north-south and east-west axes.

One of three thinning treatments is randomly assigned to each unit within each forested area. The treatments are (1) no thinning control; (2) thinning to 30 tpha, even residual spacing; and (3) thinning to 30 tpha, clumped residual spacing with gaps. Immediately after thinning,

FIGURE 1. Schematic of hypothetical study. Five sites in the Pacific northwest region of the US. Each site consists of three forested units. Within each forested unit are 5, 0.5 hectare plots. Within each plot are a set of north-south oriented rows of alternating Douglas-fir and western red cedar.



Douglas-fir and western red cedar are underplanted in the noncontrol units in alternating north-south oriented rows.

Prioritizing Objectives

Priorities for study outcomes are based on criteria that differ among the stakeholders. Priorities may be driven by available resources, by a need for data to inform forest management, or by a need for broadly applicable results. During the design phase, all types of competing priorities, statistically or otherwise, should be identified and coordinated among all interested parties. We will discuss specific challenges of constructing priorities for LSMEs based on the statistical inference and precision, with the understanding that, in large-scale studies, it is unlikely that all competing preferences of the cooperating researchers can be met.

Prioritizing the multiple objectives in an LSME provides an opportunity to prioritize resources and facilitates decision making throughout the study design and implementation process. In that context, three types of objectives can be distinguished (see hypothetical example). Primary objectives determine the general study setup and should receive resources necessary to answer their questions with adequate precision. Secondary objectives are fit within the framework established for the primary objectives. They may be funded at lower levels and their implementation may be compromised whether it interferes with the primary objectives, resulting in less precise estimates or answers. Ancillary objectives do not influence or alter the study setup and opportunistically take advantage of the experimental setup.

For the example, the response of vegetation to thinning is a primary objective in the hypothetical example. Adequate sample sizes, relevant scopes of interest, and measurement regimes will be developed to address this primary objective. The effects of thinning on underplanting (a secondary objective) will be addressed only after ensuring that resources are adequate for addressing the primary objectives. The ancillary objective was added when equipment and funds were made available and it was determined that the monitoring station could be included without significantly affecting the primary and secondary objectives.

Although these distinctions seem obvious, formally developing a clear hierarchy of objectives provides guidance for numerous decisions throughout the life of the experiment. Several scenarios involving conflicting choices, as well as the benefits of a clearly defined hierarchy of objectives, are discussed in the Results and Discussion section.

RESULTS AND DISCUSSION

Defining the Scope of Inference

The scope of inference is the set of situations represented by a study; it defines the range of variation to be represented by the replicate units of the study (Hurlbert, 1984; Ford, 2000). After the scope and appropriate replicates are identified for each objective, the totality of resources required to achieve the specified levels of precision of the LSME is defined. When objectives have different spatial or temporal scales, a critical and detailed discussion of potential scopes of inference for each objective quickly identifies realistic or unrealistic goals. Spatial scope determines whether study implications are limited to a single tree or a stand or apply to multiple geographic regions. The temporal scope determines whether study findings are representative of days, seasons, or years. Temporal scope may need further clarification as well. For example, the scope may be identified as “similar spring seasons.” But since each season is relatively unique, hallmark features of the seasons should be identified, that is, are the seasons unusually dry springs, or especially long springs?

Replicate units are randomly selected from the range of variation in the scope of inference to ensure that results can be generalized to the scope of inference (Hurlbert, 1984). The large areas and long-time scales over which treatments are applied in LSMEs mean that the range of variation is quite large and many replications may be required to statistically detect treatment effects. But, even if replications are available (and sometimes they are not), the cost of adequate replication may be prohibitive (Monserud, 2002). Thus despite a large expense, LSMEs have a heightened risk of being insufficiently precise to meet the objectives. Recognizing this, an assessment of replication is crucial (Steidl, Hayes, and Schaubert, 1997). Researchers have several options whether the level of replication is considered inadequate for one or more objectives. For example, if replication requirements cannot be met for an objective, the objective could be omitted from the study. Alternatively, the scope of inference and associated range of variation could be reduced. In our hypothetical example, study sites could be limited to north-facing slopes to minimize the influence of different solar radiation levels. This would limit potential replicates to stands on north-facing slopes, but fewer replicates would be required to statistically detect treatment effects. However, this would limit the scope of inference, as the study results would only be applicable to north-facing slopes.

In LSMEs with multiple objectives, it is important to understand that limiting the scope for one objective can affect the scope of inference for other objectives as well. Although this may be an acceptable compromise for some objectives, it may limit others. In the hypothetical example, the scope of inference is mature Douglas-fir forests in the Coast Range of the western United States and the dispersion of study sites throughout the Coast Range reflects this scope well. If the 0.5-hectare underplanting plots are well distributed throughout each 20-hectare unit, the scope of inference for the secondary objective will match that for the primary objective. As discussed already, however, limiting treatment units to north-facing slopes would prevent investigations comparing responses on south- and north-facing aspects. Using only a subsection of each 20-hectare unit (e.g., north-facing slopes) limits the range of environmental variation for underplanting studies and reduces the scope of inference for this objective. This highlights how scopes may differ among objectives and how the relative ranking of research objectives clarifies the implications of design choices.

Considerations for Spatial Extent and Resolution of Plot Size

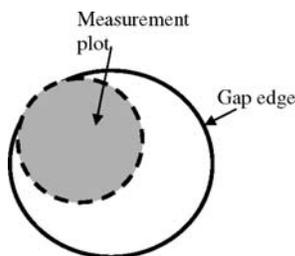
An organized sampling plan for responses within and among treatment units is required to link ecological responses and to accommodate limited resources over the multiple and large spatial and temporal scales. Multiple definitions of replicates may be required, each linked to the spatial scale of its associated objective and scope of inference. Frequently in LSMEs, smaller-scale treatments are superimposed on large-scale treatments to address objectives at smaller than operational scales. In our hypothetical study, the comparison of seedling growth among different tree species is embedded within large-scale thinning units. These smaller-scaled treatments may require different sample sizes than the large-scale treatments in order to statistically detect differences. Ideally, when large-scale objectives require the most replications, small-scale objectives are also adequately replicated. However, if small-scale objectives require more replications, large-scale experimental units may have to be split to accommodate the small-scale treatments, effectively employing a split plot design (Schabenberger and Pierce, 2002). Alternatively, a study may add neighboring stands as replicates for small-scale treatments. Since the definition of a replicate is linked to the intended scope of inference, the revision of the definition should trigger an assessment of whether the intended scopes of inference are retained for all objectives.

Avoiding Bias

Simple random sampling or systematic sampling is commonly used to select representative subsamples of large-scale units. Random sampling ensures that the average response is accurate, precise, and representative. However, when uncommon characteristics exist within larger homogeneous treatment units, it may be tempting to investigate the effects of rare or unique features on the value of the average. In the hypothetical example, the few treeless gaps within an otherwise homogeneous forested unit may not be sampled at all when simple random sampling is used. In this case, the lack of subsamples within gaps may be perceived as a biased response, underestimating the true variability in the unit. Although gaps were implemented to provide a range of conditions within the treatment units, failing to sample them provides no information about the role of gaps. To circumvent this potential dilemma, a larger sample size will increase the probability that gaps will be sampled, but will result in higher sampling costs. Forcing subsample plots to fall in the gaps can overestimate uncommon gap characteristics relative to the more common forest vegetation. Gaps can be legitimately sampled using stratified random sampling to account for the proportion of area in each stratum (e.g., thinned area and gaps). Samples are located randomly within each stratum (but see discussion of random plot location in gaps in the following section) and a weighted average (weighted by area in each condition class) will ensure that measured values are representative of the large treatment unit.

When subsample plots are selected within treatment units by choosing plot centers, a (usually small) bias is incurred. With simple random sampling, any potential subplot within the unit should have the same chance of selection as any other subplot. But randomly selecting a subsample by randomly selecting a plot center so that the entire plot will be located within the unit has the unintended consequence that the probability that a subsample is included in a plot decreases with decreasing distance to the unit edge (Figure 2). This sampling scheme weights the conditions near unit centers more heavily and biases the representation of the entire unit. When treatment units are large, the edge area of the unit is a small proportion of the total area and the negative effects of this subsampling bias are likely to be negligible. But if units are small, for example, small gaps, the bias from subsampling using randomly located plot centers will be non-negligible. One solution is to divide each gap into polygons of equal area and then randomly select polygons. This may be time consuming. If the conditions within the gap are not of

FIGURE 2. Example of a plot layout highlighting sampling concerns in gap studies. The solid line represents the edge of a half-acre gap (83.3 feet radius) and the dashed line the border of a quarter-acre measurement plot (gray). If measurement plots are required to be completely inside a gap, randomly placed plots will always include the gap center and the probability of a point being inside the measurement plots decreases towards the gap edge.



interest, it may be preferable to ignore any distinction between gaps or forested areas and rely on simple random sampling to represent the full range of conditions in the treatment unit. A prioritized hierarchy of objectives can facilitate choices when gaps are to be examined separately, but complications of the sampling scheme potentially compromise objectives. This sampling issue is not unique to LSMs, but becomes more prominent when a study involves multiple disciplines and objectives and when nonuniform areas of interest are present within large-scale treatment plots.

Measurement Plots within Large-Scale Units

Since the variation associated with treatment effects changes as the sizes of the observational units change (Mercer and Hall, 1911; Home and Schnieder, 1995), the design process should include an evaluation of the minimum spatial area over which the treatment effect will be measured. At least for primary objectives, the response of the entire large-scale treatment unit is of interest. But for many responses, such as understory vegetation in our hypothetical example, it is only practical to collect data from smaller plots within the large-scale unit. In this case, multiple subsamples are taken within a single large-scale unit and the responses from all subplots are synthesized to calculate a single response measurement that represents the entire large-scale unit. Alternatively, the response of interest can be a smaller-scaled process that is most appropriately measured on small plots. A synthesis of subsamples

represents the response on a small plot within a single treated unit. The hypothetical experiment has a large-scale secondary objective, investigating effects of thinning on migrant songbird populations and a smaller-scale secondary objective of comparing with the ability of underplanted Douglas-fir and western red cedar seedlings to grow and survive under operationally thinned and unthinned stands. The 20-hectare treatment unit is necessary and appropriate for assessing songbird population dynamics, but is larger than necessary for assessing the underplanting objective. In the example, Douglas-fir and western red cedar could be underplanted and measured in 0.5 hectare patches within each thinning unit, but songbird responses should be assessed over the entire 20-hectare unit.

The hypothetical example shows that using small plots to obtain small-scale responses for large-scale treatments may be cost-effective; however, three implications must be kept in mind. First, multiple underplanted patches in each operationally thinned treatment unit do not increase the replication of thinning treatments (or the degrees of freedom for statistical testing differences). The seedling survival objective is restricted to the comparison of thinned versus unthinned stands, and the units to which the thinning treatments were randomly assigned define the “true” replication of these treatments. Using multiple underplanted patches within large-scale units as replications is pseudoreplication (Hurlbert, 1984). Second, the scope of inference for effects of the small-scaled treatment is limited to the small scale: In the hypothetical example, the 0.5-hectare underplanted patches, rather than 20-hectare stands. Some responses to treatments measured on the small scale may be inferred to larger scales (e.g., the influence of overstory density or seedling growth), but other effects may not scale up (Brandeis, Newton, and Cole, 2002). For example, small clumps of underplanted seedlings may attract the attention of herbivores, and one visit of hungry deer (e.g., one 30-minute feeding) will have a greater effect on a small patch than it would have on large-scale underplantings (Brandeis, Newton, and Cole, 2002). Third, gains in short-term efficiency from placing small-scale measurement units within an operationally defined larger-scaled experimental unit may be counteracted by long-term complications. For example, if songbird responses after 30 years are of interest, the use of underplanted patches may eventually affect songbird habitat, that is, being efficient with the study design by using small-scale underplantings for short-term objectives could make it impossible to isolate long-term effects of thinning from those of patch underplanting.

Since both the songbird and underplanting objectives are secondary, the two objectives have equal priority and thus the hierarchy of objectives is not helpful here. Within levels of the hierarchy there still may be priorities to set and hard choices to make, since choices for one objective may still impact other objectives. As with earlier issues, complexities of this nature are inherent in LSMs and careful consideration of the potential effects during the planning phase may be the safest way to avoid future conflicts. In our example, the problem can be solved by defining a time frame for measuring seedling performance and scheduling a removal of these seedlings after the objective has been accomplished.

Temporal Variation

The diversity of objectives in LSMs incorporates a range of temporal scales that induce temporal variation not present in shorter-term studies. Schedules are constrained by the need to coordinate among disciplines. The inability to apply all treatments or take all measurements at the same time also adds temporal variation. Again, the prioritization of objectives is helpful to define the role of temporal variation and to coordinate and integrate timelines and treatment schedules (Walters and Holling, 1990).

Each objective and associated ecosystem response may have a unique, temporal, and dynamic and is best assessed at its own temporal scale. The temporal issue is further complicated as responses may have multiple time scales of interest. Several LSMs in the Pacific northwest (e.g., Coates et al., 1997; Carey, Thysell, and Brodie, 1999; McClellan et al., 2000; Aubry, Halpern, and Maguire, 2004; Curtis, Marshall, and DeBell, 2004; Cissel et al., 2006) and other regions (e.g., Saunders and Wagner, 2005) have implicit objectives to investigate whether silvicultural treatments enhance development of old-growth or mature forest habitat characteristics over relatively long time frames (50-100 years). At the same time, these experiments were initiated because of interest in diversifying large homogeneous landscapes in the short term, that is, to determine whether silvicultural treatments can lead to heterogeneous stand structures in young stands over a 5- to 10-year time frame. Another common goal of LSMs is to correlate responses of various ecosystem components at the same point in time in order to document "overall" ecosystem responses. Thus the optimal time to measure a response and the need to draw associations among responses has to be balanced. To do this, optimal timing of treatment and measurement schedules for each objective should be determined. Then, using

the prioritization of objectives, schedules can be developed to emphasize primary over secondary objectives and strike a balance acceptable to all participants.

For example, understanding the dynamics of understory vegetation and seedling establishment may require annual or biannual measurements, whereas the dynamics of large downed wood as a habitat component for small mammals may be sufficiently represented with a 10-year measurement cycle (assuming no major windstorm or similar disturbances). A coordinated sampling schedule for the prioritized list of responses helps assist in planning and ensures that resources are efficiently used. The sampling plan can help determine the synthesis stages, that is, times when all study components can be put together to investigate “integrative ecosystem responses.” A common approach uses the most dynamic response to determine the minimum measurement interval, and other responses are measured in multiples of that minimum measurement interval, for example, collecting understory vegetation data every 2 years, tree regeneration every 4 years, and downed wood every 12 years.

Note that temporal scales and measurement procedures are intrinsically linked to objectives, the scope of inference, and other components of the study. If temporal scales change during the study, either intentionally or unintentionally, an assessment of all implications is warranted (e.g., see the description of effect of patch underplanting on songbird habitat in the *Measurement Plots* section).

Considerations for Describing Temporal Trends

A common objective in LSMEs is to assess time trends and determine how these trends differ among treatments. Prioritizing the importance of this objective is crucial, as data analysis methods for detection and estimation of trends can easily be influenced by the study design. When measurements are made repeatedly on the same experimental unit and the measurements are used to estimate trends over time, the correlation among the measurements should be accounted for in the analysis, for example, a repeated-measure, analysis-of-variance-type analysis (Schabenberger and Pierce, 2002).

Repeated-measures analyses require that all experimental units for all treatments are assessed at the same point in time and that sufficient numbers of experimental units relative to the number of time points when measurements are made. This is necessary to estimate correlations among repeated measurements on the same experimental units.

A minimal requirement is to install more replications than the number of time periods of interest. A regression approach may be used when replication of treatments at each point in time is minimal. However, since many LSMEs are designed as randomized block designs (Monserud, 2002), trend detection needs to account for blocking factors, further complicating the statistical analysis. Therefore, during the planning stage of an LSME it is prudent to identify responses for which time trends are of interest and the associated analysis that will be used. This exemplifies how the number of replications is determined by the need for a particular analysis, as well as by the need to represent variation in the scope of inference.

Effects of Temporal Processes on Study Design

In long-term studies the meaning of a treatment may be affected by time-dependent processes. The treatments may not be completely determined before the study installation because they are a series of manipulations applied sequentially. The timing, type, and intensity of subsequent manipulations may depend on assessments of conditions within treatment units, and this has ramifications for treatment definitions and scopes of inference. Delaying a treatment application by a few months (i.e., the time it takes for an operational thinning operation in a large stand to be completed) may not add much variation to investigations of long-term growth responses or long-term songbird response (e.g., Cissel et al., 2006). For other responses it may be highly influential (e.g., Walters and Holling, 1990). For example, a 3-month delay in harvest, over which trees shift from dormancy to actively growing, may influence the extent and impact of harvesting damage to residual trees (Moore et al., 2002).

Suppose thinning treatments in our hypothetical example were applied at the same point in time to all treatment units in all five stands. All units are subjected to the same environmental conditions at the same points in time relative to the treatment. In this case, the scope of inference, replication, treatments, and experimental units are defined and allow direct comparisons among treatments to address the study objectives. Next, suppose that the second thinning will occur in a treatment unit when the average basal area within that unit reaches a threshold. Is this treatment protocol well defined? The definition allows thinning to take place in different years for different stands. When replicate units are thinned in different years, each replicate is subjected to a different set of environmental conditions immediately following thinning. Extra

temporal variation may be present in the data or replicates may no longer represent comparable conditions. Alternatively, treatment timing may be determined by average responses for all replicate stands of one treatment. Both scenarios confound year-specific effects (e.g., a good growing season) with treatments or sites, making it difficult to determine whether responses are due to the thinning treatments (Walters and Holling, 1990).

A potential solution is to thin all replicate units within a stand (a block) in the same year, but entire units maybe thinned in different years. Annual variation is, therefore, confounded with unit-to-unit or block variation. Variation captured by blocks may be removed in the usual randomized block design analysis (Steel, Torrie, and Dickey, 1997).

Operational limitations may constrain the portion of an LSME that can be treated at the same time and may cause interpretations of “same time” to differ. “Same time” may mean same calendar year and season, or the “same relative time,” that is, years since treatment. In small-scale experiments, these two interpretations are usually the same. In LSMEs, the potential discrepancy may be resolved in several ways. In studies where differences in timing of treatment application were not large, all measurements were made in the same calendar year and differences in the time since treatment were ignored in the analysis, but stated in any documentation (e.g., Beggs, 2005). In other studies, where treatment application varied over longer time periods, measurements were scheduled at fixed intervals after treatment, for example, 5 years after treatment application (Cissel et al., 2006). Viewing areas with different “time since treatment” as if they were treated at the same time is equivalent to accepting a measurement error, that is, time since treatment is not measured correctly (Kmenta, 1997). The influence of this type of measurement error may decrease over time (Davis and Hutton, 1975). Consequently, in LSMEs where multiple objectives cover a range of temporal scales, the priority of the objectives (and their associated scale) may be helpful when deciding to use the same chronological or relative time for measurements.

Measurement schedules are further complicated when treatments consist of multiple manipulations over time. Effects of individual manipulations and responses to cumulative sets of manipulations are usually of interest. In this case, conditions right before each manipulation must be documented (similar to “Eingriffsinventuren” sensu von Gadow and Schmidt, 1998). They provide the final conditions necessary to quantify changes owing to the earlier treatments. Additional

measurements immediately following a manipulation and detailed documentation of treatments provide baselines for assessing response to current treatments. In this case, timing of measurements will be determined by the timing of manipulations.

In addition to the definition and timing of treatments, LSMEs have special issues when the ecosystem response to a treatment results in unacceptable stand conditions. The objectives determine what conditions will be unacceptable and whether temporal alterations of treatments or schedules are required. During the implementation of our hypothetical example, it may become clear that a thinning treatment will result in closed canopy conditions that eventually lead to 100% understory seedling mortality. Once it is apparent that this treatment will never result in multiple canopy layers, which are an important part of late successional stand structures, should an unplanned follow-up treatment be implemented to return the stands to a trajectory more in keeping with the motivation of the study? The answer lies once more in the prioritization of objectives. If the primary objective is to evaluate the effects of the original treatments over time, then it is not appropriate to apply an unplanned treatment. If the primary objective is to evaluate the effects of various possible treatments designed to produce late successional structure, modification of the original treatment definition is necessary (Cissel et al., 2006).

The objectives of LSMEs can include responses to stochastic elements that cannot be described by prescheduled, regular sampling schemes. In these cases, timing of measurement may have to be determined by a trigger. Triggers may be external or internal to the experiment, the distinction being whether the trigger assessment is limited to the experiments. An example of an external trigger is the occurrence of a regional bumper seed crop, which may trigger measurements of seed banks or tree regeneration. An assessment when tree regeneration overtops shrubs in all (or a subset of) treatments is an internal trigger. Both sets of triggers require additional sampling or informal reconnaissance to determine if the trigger condition has been met.

Internal triggers can be used to define timing, but also the definition of treatments. The Level-of-Growing-Stock study (LOGS) (Marshall and Curtis, 2002) investigates growth-growing stock relations over a range of site conditions by comparing different thinning regimes applied on each site. It has a narrower focus than LSMEs, but provides an example of this approach. The timing of treatment application is based on height growth. Treatment definitions are based on the growing stock increment of the control, unthinned plots. Treatments are defined as the

proportions of retained growing stocks as a percentage of the gross basal area of the control. The range of predetermined proportions to be retained ensures that treatments cover a range of tree densities. The growth increment and height of trees in an unthinned control site reflects site productivity for all treatment units within that site. The scope of inference is not to an absolute thinned value (e.g., 300 trees per hectare) but to a proportion of the original stand density (% removed). This assumes that site-specific environmental conditions affect all treatments within a site equally and that differences are accounted for by blocking on-site comparisons among treatments.

CONCLUSIONS

Large-scale management experiments offer an opportunity to investigate a range of questions pertinent to forest management. The first installations initiated in the 1990s are currently being evaluated for their usefulness and assessed for continued funding. Interactions among conflicting study design elements for study objectives at multiple scales can produce unintended consequences in time and space. But forethought and careful planning for a prioritized list of objectives, combined with lessons learned from current LSMs, provides guidance for handling specific sources of spatial and temporal variation in future studies. Accounting for sources of variation through better designs will ultimately reduce the background variation and improve the assessment of treatment effects. But the study design needs to acknowledge and balance statistical and nonstatistical priorities as well. It is crucial that discussions and coordination should continue throughout the life of studies to elucidate consequences of particular choices, as well as suggest potential solutions for possible setbacks.

REFERENCES

- Aubry, K.B., C.B. Halpern, and D.A. Maguire. 2004. Ecological effects of variable-retention harvests in the northwestern United States: The DEMO Study. *Forest Snow and Landscape Research* 78:119-138.
- Beggs, L. 2005. Vegetation response following thinning in young Douglas-fir forests of western Oregon: Can thinning accelerate development of late-seral structure and composition? MS Thesis. Oregon State University, Corvallis, OR.

- Brandeis, T.J., M. Newton, and E.C. Cole. 2002. Biotic injuries on conifer seedlings planted in forest understory environments. *New Forests* 24:1-14.
- Bruce, D. 1977. Yield differences between research plots and managed forests. *Journal of Forestry* 75(1):14-17.
- Carey, A.B., D.R. Thysell, and A.W. Brodie. 1999. The forest ecosystem study: Background, rationale, implementation, baseline conditions, and silvicultural assessment. *Gen. Tech. Rep. PNW-GTR-457*. Portland, OR: USDA Forest Service Pacific Northwest Forest and Range Experiment Station. 129 p.
- Cissel, J.H., P. Anderson, S. Berryman, S. Chan, D. Olson, K. Puettmann, and C. Thompson. 2006. BLM density management and riparian buffer study: Establishment report and study plan. US Geological Survey Scientific Investigations Report 2006-5087. 151 p.
- Coates, K.D., A. Banner, J.D. Steventon, P. LePage, and P. Bartemucci. 1997. The Date Creek silvicultural systems in the interior cedar-hemlock forests of Northwestern British Columbia: Overview and treatment summaries. *Land Management Handbook* 38. British Columbia Ministry of Forests Research Program. 129 p.
- Curtis, R., D.D. Marshall, and D.S. DeBell. 2004. Silvicultural options for young-growth Douglas-fir forests: The Capitol Forest Study—establishment and first results. *Gen. Tech. Rep. PNW-GTR-598*. Portland, OR: USDA Forest Service Pacific Northwest Forest and Range Experiment Station. 110 p.
- Davis, R.B. and B. Hutton. 1975. The effect of errors in the independent variables in linear regression. *Biometrika* 62:383-391.
- Ford, E.D. 2000. *Scientific Method for Ecological Research*. Cambridge University Press, Cambridge UK. 564 p.
- Gadow, K.V. and M. Schmidt. 1998. Periodische Inventuren und Eingriffsinventuren. *Forst u. Holz* 53(22):667-671.
- Halpern, C.B., S.A. Evans, C.R. Nelson, D. McKenzie, D. Liguori, D.E. Hibbs, and M.G. Halaj. 1999. Response of forest vegetation to varying levels and patterns of green-tree retention: An overview of a long-term experiment. *Northwest Science* 73:27-44.
- Home, J.K. and D.C. Schneider. 1995. Spatial variance in ecology. *Oikos* 74:18-26.
- Hurlbert, S.H. 1984. Pseudoreplication and the design of ecological field experiments. *Ecological Monographs* 54:187-211.
- Kmenta, J. 1997. *Elements of Econometrics*. University of Michigan Press, Ann Arbor, MI.
- Marshall, D.D. and R.O. Curtis. 2002. Levels-of-growing-stock cooperative study in Douglas-fir: *Report no. 15-Hoskins: 1963-1998*. USDA Forest Service Pacific Northwest Research Station Research Paper RP-537.
- McClellan, M.H., P.E. Hennon, R.L. Deal, T.L. De Santo, and M.S. Wipfli. 2000. Alternatives to clearcutting in the old-growth forests of southeast Alaska: Study plan and establishment report. *Gen. Tech. Rep. PNW-GTR-494*. Portland, OR: USDA Forest Service Pacific Northwest Forest and Range Experiment Station. 40 p.
- Mercer, W.B. and A.D. Hall. 1911. The experimental error of field trials. *Journal of Agricultural Science (Cambridge)* 4: 107-132.
- Monserud, R.A. 2002. Large-scale management experiments in the moist maritime forests of the Pacific Northwest. *Landscape and Urban Planning* 59:159-180.

- Moore, J.R., D.A. Maguire, D.L. Philips, and C. Halpern. 2002. Effects of varying levels and patterns of green-tree retention on amount of harvesting damage. *Western Journal of Applied Forestry* 17(4):202-206.
- Saunders, M.R. and R.G. Wagner. 2005. Ten-year results of the Forest Ecosystem Research Program (FERP)—Success and Challenges. In: Peterson, C.E. and Maguire, D.A. (Eds.), *Balancing Ecosystem Values: Innovative Experiments for Sustainable Forestry*. Gen. Tech. Rep. PNW-GTR-635. U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station, Portland, OR.
- Schabenberger, O. and F.J. Pierce. 2002. *Contemporary Statistical Models for the Plant and Soil Sciences*. CRC Press. Boca Raton, Florida. 738 p.
- Steel, R.G., J.H. Torrie, and D.A. Dickey. 1997. *Principles and procedures of statistics, a biometrical approach*. McGraw-Hill, New York. 666 p.
- Steidl, R.J., J.P. Hayes, and E. Schaubert. 1997. Statistical power analysis in wildlife research. *Journal of Wildlife Management* 61:270-279.
- Walters, C.J., J.S. Collie, and T. Webb. 1988. Experimental designs for estimating transient response to management disturbances. *Canadian Journal of Fisheries and Aquatic Sciences* 45:530-538.
- Walters, C.J. and C.S. Holling. 1990. Large-scale management experiments and learning by doing. *Ecology* 71(6):2060-2068.

RECEIVED: 11/28/05

REVISED: 05/10/06

ACCEPTED: 01/09/07

doi:10.1300/J091v26n01_01