

Statistical Considerations for Plot Design, Sampling Procedures, Analysis, and Quality Assurance of Ozone Injury Studies

Michael Arbaugh and Larry Bednar¹

The sampling methods used to monitor ozone injury to ponderosa and Jeffrey pines depend on the objectives of the study, geographic and genetic composition of the forest, and the source and composition of air pollutant emissions. By using a standardized sampling methodology, it may be possible to compare conditions within local areas more accurately, and to apply the results to larger geographical contexts, such as pine populations within and among watershed areas.

We present recommendations in this chapter to develop plot design and sampling strategies that follow established principles of observational studies, and we discuss the usefulness of indices and need for quality assurance for long-term monitoring of ozone injury to ponderosa and Jeffrey pine. The recommendations pertain only to the goals and conditions unique to ozone pollution impacts on mixed conifer, ponderosa and Jeffrey pine forests of the Western United States. Differences in climate, physiology, and air pollutant composition and exposure patterns will likely alter sample design and analysis needs for other forest types and locations.

One of the principal objectives of long-term sampling by repeated observations is to document tree and stand changes through time. In this paper sampling methods will be chosen to optimize long-term, repeated observations of tree populations within a watershed area, and to compare damage between watersheds. The specific questions will be addressed about plot design, the method used to locate plots, and the optimal sample size of branches per tree and trees per plot needed to estimate summary statistics and indices.

Cochrane (1983) defines observational studies as those in which the objective is to determine the causal effects of certain agents, procedures, treatments or programs that the investigator cannot subject to controlled experimentation. The lack of ability to impose treatments or procedures distinguishes observational studies from experimental studies. Nearly all long-term or regional surveys are observational because of the uniqueness of history, climate, soil and topography of each area. These data are also the most valuable, since most issues of forest health involve large areas and long periods (50 to 250 years).

Statistical principles are well-established for the analysis of observational studies and include:

- Comparison of “quasi-treatment” and “quasi-control” groups to resemble a designed experiment.
- Comparison of “treatment” group(s) with more than one “control” group to develop different contrasts with the “treatment” group.
- Comparison of “treatment” and “control” groups with important exogenous variables if feasible; if infeasible, groups should be adjusted for differences using covariates in the analysis.
- Use of a variety of measures and comparisons to reduce the dependence of study results on single aspects of the data and on

Considerations of Observational Studies

¹Statistician, Pacific Southwest Research Station, 4955 Canyon Crest Drive, Riverside, CA 92507; and Mathematical Statistician, Pacific Northwest Research Station, P.O. Box 3890, Portland, OR, 97208

assumptions inherent for single methods of analysis (Strauss 1990).

The underlying theme behind observational studies is that the system cannot be completely defined or manipulated, and large sources of unknown variation may be present that violate the assumptions of a single analysis approach. Thus, repetition of the comparisons, measures and approaches is used to reduce the dependence of conclusions on incorrect assumptions associated with single measures.

Stratification of Target Populations

The technique of dividing a target population with high variation into smaller non-overlapping populations (strata) that are more homogenous is used whenever possible as a part of a sampling scheme. This division allows separate estimates to be developed for each stratum, and usually allows estimates for the entire population to be developed with a higher degree of precision for a given level of expense. To attain these benefits, strata are designed so that within-stratum variance is less than the variation in the target population. Ideally, strata will have meaningful interpretations, such as division by species composition, age groups, or density. In addition a long-term study should also be included to define strata stable through time.

The absence of reliable information is the most limiting factor for describing strata for visible injury surveys. Few studies provide information about the effects of density, topography, and stand structure on visible injury levels. Strata, therefore, need to be defined by coarse criteria using known information.

Drainage basins are commonly used as geographical stratifications because they often contain similar soils, history, air flow, and hydrological dynamics. Thus, one possible stratification scheme for visible ozone injury would be to divide populations by watersheds, and then subdivide watersheds into three strata: Jeffrey pine, ponderosa pine and a stratum with mixes of these two species.

Plot Considerations

Design

Plots are generally designed to yield a constant amount of information. This design simplifies aspects of statistical comparisons. For long-term monitoring an important consideration is whether plots should have constant area or a constant number of trees. The choice of approach used depends on the intended uses of the survey information. Defining plots with constant area facilitates extrapolation of survey estimates to the entire area of the target population. Defining plots to contain equal numbers of trees allows survey results to be used in making statements about the injury found on the tree level. Equal numbers give equal resolution of information about the average tree response within a plot.

Because visible injury monitoring is oriented to average tree response within watersheds rather than to area oriented watershed-level assessments, plots consisting of a constant number of trees are appropriate. This requirement is especially important because severe visible injury is rare. Often only a few trees in a stand will be severely affected by ozone injury. Reducing the number of trees per plot would reduce the ability to detect these effects.

Consistency with prior studies is also desirable. Information that is comparable to historical work allows a temporal picture to be developed that is

important for evaluating trends. Previous pine monitoring has used both circular and rectangular plots. Circular plots, such as those used in the Eridanus injury index (EII) procedures, have several disadvantages. First, if plots containing a fixed number of trees are needed, then the radius of a circular plot must be uniquely selected for each plot. Several trial values for plot radius may be needed to arrive at the correct plot size. This process may be quite time-consuming if the number of trees needed is not small. Second, the selection of trees falling within a circular plot is identical to selecting nearest neighbors to a point. This results in biased estimates of stand density if trees occur in a clustered spatial distribution, as ponderosa and Jeffrey pine often do. Circular plots also increase the chance of including spatial correlations that may confound results by using the most compact possible plot layout.

Other types of pine monitoring include fixed area belt transects, as used in the USDI National Park Service Air Quality Division Procedure (Stolte and Bennet 1985), and variable area belt transects used in the San Bernardino National Forest (Miller 1983). Belt transects avoid disadvantages associated with circular plots. The longer plots sample across clusters of similar age, size, and genotypes and reduce spatial correlation effects associated with nearest neighbors. Thus, for our design belt transects are more appropriate than circular plots.

Number of Plots per Stratum

The results of a monitoring, or observational, study are estimates of representative conditions in the target population(s). This type of result contrasts with experimental design, which results in hypothesis testing and statements concerning treatment differences. Monitoring designs are less useful for comparisons, but yield estimates of the general target population (for which experimental data are sometimes poor). Sampling design cannot optimize for experimental and monitoring goals simultaneously, but our long-term monitoring approach can be used to develop population estimates with accompanying confidence levels.

Estimates derived from a survey consist of point estimates of the population value, with stated range of values for those estimates expressed as a confidence interval. Higher levels of confidence result in conservative (wide) intervals, while lesser degrees of confidence result in narrower intervals. Although other confidence levels are valid, it is customary to use 95 percent confidence levels.

Sample Size Determination

If sensitivity to a change from one level of a population characteristic to another is of interest, the survey can be designed to provide a confidence interval width that will reveal a change of that size with a stated reliability. Increased sample size (number of plots) will result in narrower confidence intervals, despite the confidence level chosen. It is necessary to postulate a difference of interest (the precision) before sample sizes can be determined. The level of precision should be based upon the desired size of the difference to be detected, and should be linked to a biologically significant indicator, such as loss of foliage and productivity.

Given a choice of confidence level, desired precision level, and some preliminary information about the variance, a sample size (number of plots) expected to give a confidence interval of the desired width can be estimated. The resulting sample size should be considered a minimum number, the actual number being 10 to 25 percent larger for long term surveys. This will insure the design against loss of plots due to fire or other causes.

Specification of the Desired Precision Level

Index Construction

Index Characteristics

The use of indexes to summarize information has several advantages: it reduces complex results to an understandable form, combines several diverse measures into a single measure, and enables policy-makers to have readily available numbers for justifying legislation. Indexes also have several negative features when used as scientific findings. Information is lost when measures are summarized, especially when several measures are combined, so that interactions occurring between the differing measures are missed. Indexes also involve some arbitrary decisions based on investigators expert knowledge. Although these estimates incorporate all knowledge available at the time of formulation, new information may change index construction decisions.

Inclusion of arbitrary, or rough estimates of tree biology, based on expert knowledge cannot be avoided, especially for visible injury assessments. In observational studies this weakness is compensated by using several measures, each depending on differing assumptions to increase the reliability of study results.

Thus, we recommend that several measures be developed, each using different assumptions, instead of a single index. A primary measure can be designated, but several other measures should be developed as supporting information. These measures may be other indexes, or simple summaries of measures used to construct the primary index. For example, the Forest Health Monitoring Program includes crown indices for crown transparency, and crown structure. It is important that the indices used depend on different assumptions because this information makes each measure useful. In addition, data used for index development should also be retained for future index refinement. Retention of the data used to construct indices will also allow alternatives to be calculated retroactively in the future if only one index is presently available.

Primary Index

The OII seems to satisfy the five 'requirements' stated by Muir and McCune (1987). The most important primary index component is the presence or absence of chlorotic mottle since it is the most direct evidence of ozone effects on pine foliage. Additional measures, multivariate graphs, principal components of the measures, and alternative indices can be also be developed. They should be alternate presentations of the data not dependent on the assumptions used to construct the OII index. The type of summarizations needed should be decided by experts familiar with the effects of ozone injury on sensitive tree species. Statisticians can help by enumerating the choices of summary presentation.

Within-Plot Sampling Considerations

Tree Sample Size

Previous studies considered all trees with DBH > 10 cm (Miller 1973, Muir and McCune 1987, Stolte and Bennett 1985). The OII method considers only trees with prunable crowns, possibly excluding large co-dominant trees. This exclusion may bias stand estimates if visible injury differs for this group. Trees with DBH < 10 cm are less important for stand estimates because they usually comprise a small portion of the stand, are highly stressed, and ozone exposure conditions differ from larger DBH classes.

Because the analysis goal is average tree damage, the number of trees included in the sample should be large enough to allow confident application of

the Central Limit Theorem. Sufficiently large numbers will guarantee that averages will be distributed according to the normal (Gaussian) distribution, simplifying later analysis. The choice of any particular number is arbitrary, but inclusion of 30 or more trees per plot is likely to be sufficient for monitoring purposes.

In a general sense, the plots should be large enough to include sufficient sample trees for the categories of interest. This requirement is important if small numbers of sensitive trees are present. If these trees are randomly distributed, the probability that a group of n trees contain no sensitive trees is approximated by the binomial distribution, so

$$\text{Prob}\{\text{no sensitive trees}\} = (1-p)^n,$$

in which p is the proportion of the general population that is sensitive. The number of sample trees should be chosen to obtain an acceptably low probability of exclusion for these trees if sufficient knowledge is obtainable. If this knowledge is not available, then standard sample size formulas should be used.

During long-term monitoring, trees will be lost at times, so the initial plot layout must include enough trees to reasonably ensure that future mortality will not reduce the tree numbers to below that needed. The number of trees per plot should be chosen to match or exceed the foreseeable maximum required. Estimates of the sample size needed should be obtainable from the information in Duriscoe (1988). The standard sample size determination formula can be applied using several precision levels, and the standard deviation estimates obtained from graphs of the trees.

Branchlet Sample Size

Muir and Armentano (1987) determined optimal number of branchlets removed from a sampled tree, estimating that 5 branchlets were needed. Their methods involved application of a sequential likelihood ratio test (SLRT). Strict application of a SLRT will generally underestimate the required sample size for a fixed sample scheme. However, they used several SLRT's and a conservative value of N was chosen that exceeds the N from individual tests. This application of SLRT techniques probably avoids underestimation. Thus their suggested sample size is likely adequate to estimate the true number needed.

Summary

Sufficient information exists to design a simple robust sampling design for long-term visible injury monitoring. Past studies and expert knowledge are available for critical sampling decisions. More complex designs are possible, but they increasingly become more dependent on assumptions and knowledge that are not presently available. More refined issues include spatial autocorrelation effects, optimal stratification criteria, optimal sample size determinations and long-term divergence of plots from surrounding populations. Future studies are needed to examine the impact of these factors on the design recommendations.

Design of observational studies involves approaches that compensate for the inability to experimentally manipulate trees and stands. Because of this limitation sample sizes should be estimated conservatively. In this design comparisons between sites at the same time, and between time periods for the same sites represent our "quasi-controls" and "quasi-treatments." If some sites are placed near air monitoring stations the link visible damage to pollutant exposure can be better quantified.

Reporting multiple measures is important to reduce the reliance of conclusions on assumptions that are approximate. Changes in forest species

and stand structure, landscape features, and exposure may effect the assumptions used for a single index in unpredictable ways.

We also recommend additional surveys that do not require establishing permanent plots to complement monitoring efforts. Permanent plots may become less representative of surrounding forest conditions through time, either by different management applications, or by stochastic processes associated with local site conditions. Periodic examination of the representativeness of plots is needed to insure that remeasured plots continue to reflect surrounding forest conditions.

Quality Assurance/ Quality Control

General Considerations

Ideally a work plan/quality assurance plan should be completed at the beginning of the project (Cline and Burkman 1989). An ideal quality assurance (QA) plan is designed to support an overall program, rather than just the methods used in collecting information. The QA project plan is a comprehensive description of research procedures and methods and associated internal QA/quality control (QC) activities (Zedaker and Nicholas 1990). The QA plan is written by the principal investigators of the project and includes project objectives; experimental design, sampling procedures, and statistical methods and analysis; project management and personnel; research facilities and equipment; measurement and analytical procedures; data quality objectives and data quality assessment procedures; sample and data custody and archive procedures; data validation and analysis; and mechanisms for implementing changes in research or QA.

Quality Assurance at the Measurement Level

Within any project an important QA question is: what is the measurement error associated with each recorded variable? Determining the cause and reducing the amount of measurement error is the traditional emphasis of QA programs. QA for the measurement process includes operations and procedures in which the data produced are of the specified quality within a stated level of uncertainty. Quality is acceptable when data are consistent and have a small uncertainty when compared to the stated requirements. Precision and bias are the traditional quantitative indicators of the quality of measurement data. Typically, precision is estimated by repeated measures of reference materials or actual samples (duplicates). Estimation of measurement bias is best done by systematic use of reference materials. Estimates are not as feasible for certain biological data since reference materials change with time.

Quality Assurance Recommendations

The following specific procedures are recommended for a suitable quality assurance program:

- Provide an annual training session for workers gathering plot/tree data.
- Include the collection and analysis of quality control data in the overall statistical design.
- Collect and track remeasurement data throughout the measurement period.
- Include both internal and external components in the QA program.
- Implement a systemic approach to data handling and database management.
- Document QA and survey methods throughout the program.