Considerations for Evaluating Controlled Exposure Studies of Tree Seedlings

Charles E. Peterson, Jr. and Robert A. Mickler

Reprinted from the *Journal of Environmental Quality*
Vol. 23, no. 2, March–April 1994, Copyright © 1994. ASA, CSSA, SSSA
677 South Segoe Road, Madison, WI 53711 USA

Considerations for Evaluating Controlled Exposure Studies of Tree Seedlings

Charles E. Peterson, 1r. and Robert A. Mickler*

ABSTRACT

Tree seedling exposure studies, covering a wide range of experimental conditions in pollutant treatments, species, facilities, and exposure regimes, have been conducted during the past several years to determine acute effects and relative sensitivity of tree species in response to simulated acid precipitation and gaseous pollutants. Because of tile difficulties inherent in conducting controlled exposures with mature trees (e.g., size, variability among experimental units, and costs associated with replication of treatments), seedling exposure studies have been initiated as the quickest way to address these issues. However, sufficient consideration has not been given to either the comparabillty, of seedling studies or to their appropriate inference. The statistical power of any given analysis is rarely discussed when the outcomes are published. Appropriate and documented statistics of experimenter bias are often not reported, and variability in the exposure regime (ie., treatment target levels) and the measurement of experimental variables is assumed to be zero, ratites than quantified. Finally, the populations of seedlings for which seedling experiments have inference, the extent to which seedling responses are applicable to mature trees and forest condition, and the limitations in national or regional generalisations are crucial issues often left to an individual reader's interpretation without the benefit of adequate quantitative information presented by the authors.

IN THE LAST DECADE, there has been a great deal of research dedicated to answering pollutant related questions on trees and forest ecosystems. Much of that research has involved tree seedlings in some type of exposure chamber. The interest in forest health tends to focus on effects on mature (i.e., reproductive stage) trees, with the hope that seedling studies might identify important physiological mechanisms or processes such as photosynthesis, C allocation, nutrient uptake, gas exchange, and water relations which, if affected under various air pollutant scenarios, might affect tree growth. Because seedlings represent the future forest, studies of seedlings are useful as a direct contribution

---

C.E. Peterson, Jr., USDA Forest Service, Pacific Northwest Research Station, P.O. Box 3890, Portland, OR 97208; and R.A. Mickler, ManTech Environmental Technology, Southern Global Change Program, 1509 Varsity Drive, Raleigh, NC 27606. Received 8 Sept. 1992. *Corresponding author.

Published in J. Environ. Qual. 23:257–267 (1994).

Abbreviations: NAPAP, National Acid Precipitation Assessment Program; CSTRs, continuously stirred tank reactors: *D*, stem diameter; *H*, total height.

to risk assessment for seedling populations in stand establishment and for regeneration. These studies have often been employed to screen species (e.g., Hogsett et al., 1989) and screen families (open-pollinated plant material) within a species (e.g., McLaughlin et al., 1988) for relative sensitivities to pollutant exposures, to build exposure-response models, or to identify mechanisms of physiological response for subsequent studies of mature trees. Choosing the levels of treatment exposure often involves the balance between biological objectives that may be in conflict with each other. For example, exposure levels needed to adequately describe a change in average yields or total biomass may be too high or otherwise inappropriate for studying mechanisms of physiological response. While the variety of experimental approaches assures a wide breadth of scientific information from seedling studies, it also poses some challenges for a collective synthesis of the results.

The studies reviewed were funded within six research cooperatives under the Forest Response Program, which was established in 1986 by the National Acid Precipitation Assessment Program (NAPAP). The studies were designed to quantify seedling responses to simulated acid deposition, sulfur dioxide, and ozone in 12 conifer and 12 hardwood species. The results are synthesized elsewhere (Peterson et al., 1989; Mattson et al., 1990; Shriner et al., 1990). This study focuses on major considerations for interpreting those results, in order to realize the potential strengths of these designs and enhance the inherent value of tree seedlings in forest research. Our objective is to identify and discuss common problems among the NAPAP studies for their relevance to the planning, implementa tion, and interpretation of pollutant exposure plant response studies and to make recommendations to resolve those problems in future studies.

METHODS

Although a more detailed description of methods can be found elsewhere for exposure studies in general (e.g., Guderian, 1985) and individual NAPAP studies (see Peterson et al., 1989), the following section is intended to provide a basis for discussing some of the difficulties in interpreting final outcomes. In order to provide some reference for discussing seedling exposure experiments, we have attempted to categorize the experimental methods in Table 1 after Guderian (1985), recognizing that Guderian's approach works better for levels of biological organization than for a continuum of facilities.

Experimental Material

In the context of seedling exposure studies, the term *seedling* broadly refers to trees small enough in size or stature (e.g., initial height <1 meter) to be used in standard open top chambers. Ages of seedlings at the time of treatments ranged from 12 wk to 4 yr; the majority were 2 yr old or younger at the beginning of the studies. In most of the seedling exposure studies reviewed for this study, the experimental material was

Table 1. Considerations for evaluating controlled exposure studies of tree seedlings.

Exposure characteristics	Various environmental conditions			
	Controlled laboratory	Controlled greenhouse	Controlled field	Natural
Ambient	Ambient pollutant concentrations and ambient environmental conditions difficult to achieve.	Ambient pollutant concentrations and ambient environmental conditions difficult to achieve.	The pollutant exposure duration is controlled in open-top chambers with and without air filtration; tree seedlings and sapling grown in soil or pot culture; experiment duration limited by tree size; not reproducible; environmental conditions approach natural.	Forest health and productivity field studies sited across natural pollutant gradients; tree seedling, sapling, and mature trees grown in soil; long-term experiments on mature trees and stands; environmental conditions natural; experiments not reproducible.
Episodic	Pollutant exposures in CSTRs and growth chambers; controlled environments rooms with controlled pollutant concentration and exposure duration; seedlings cultured in pots; reproducible short-term exposures up to 1 year; environmental conditions kept constant.	CSTR exposures in partial climate control greenhouse conditions; pollutant exposure and duration controlled; seedling cultured in pots; reproducible short-term exposures up to 1 year; environmental conditions vary diurnally.	Pollutant exposures in open-top chambers and branch chambers with and without filtration; seedlings or saplings grown in soil or pots; experimental duration limited by tree size. Reproducible limited by near natural environmental conditions.	Nonchambered pollutant exposures of field plots with variable gradient of concentrations; long-term experiments on seedlings, saplings, and mature trees grown in soil; experiments not reproducible.
Square wave	Constant pollutant concentrations and exposure duration in CSTRs and growth chambers; seedlings cultured in pots; reproducible short-term experiments up to 1 year in duration; environmental conditions kept constant.	CSTR exposures in partial climate control greenhouse conditions; pollutant exposure and duration controlled; seedling cultured in pots; reproducible short-term exposures up to 1 year; environmental conditions vary diurnally.	Uniform concentration exposures difficult to achieve.	Uniform concentration exposures difficult to achieve.

← Increasing difficulty in extrapolating experimental results to the field

↑ Increasing difficulty in establishing cause-effect and quantifying exposure response

† CSTRs, continuously stirred tank reactors.

germinated from known seed sources: (i) specific regions of forest occurrence for spruce (*Picea* sp.) and fir (*Abies* sp.); (ii) tree nurseries for the eastern hardwoods; (iii) commercial and research seed orchards of loblolly pine (*Pinus taeda* L.); and (iv) regions of forest occurrence and finest tree nurseries for western conifers. Most seedlings were planted in individual containers where rooting media were typically composed of commercial mixtures (e.g., peat, vermiculite, and perlite). A homogenized soil prepared from surface soil horizons taken from a field site was sometimes used to represent a forest soil series thought to be typical of a region. Many of the studies conducting short-term exposures (e.g., 12 wk) primarily used tree seedlings in containers throughout the experiment; however, for studies of 1 yr or longer duration, seedlings were started in containers and then replanted in prepared field plots. In addition to trees germinated from seeds, some studies using slow growing species planted bare root nursery stock into containers. Seedlings were grown under nonstressed conditions with adequate nutrients, water, and light. In some cases, seedlings were screened prior to treatments and seedlings of atypical growth form were rejected.

Chambers

Controlled environment chambers were used for a large number of studies. Chambers are designed to apply precise treatments while reducing or controlling the variation in response due to fluctuations in ambient conditions. They offer the advantage of providing reproducible experimental conditions when using several separate chambers for an tent. Growth chambers (Heck et al., 1968), continuously stirred tank reactors (CSTRs) (Heck et al., 1978), and open top chambers (Heagle et al.,1973,1989); represent seedling exposure technologies that provided for delivery of simulated precipitation and gaseous pollutants. The choice of chamber type usually involved tradeoffs in experimental design, precision in the application of treatments, and approximating an ambient environment (e.g., see Guderian, 1985). The CSTRs located in greenhouses or laboratories allowed for higher precision in the application of gaseous treatments for short term (e.g., 12 wk) experiments. In contrast, opentop chambers located at field sites were used for studies of longer duration where some intrusion of ambient air is accepted in exchange for near ambient exposure to sunlight, humidity, and air temperatures. Open top chambers may or may not have used rainfall exclusion devices, depending upon experimental objectives. Although significant differences exist between the chamber and ambient environments, open-top chambers are considered the best technology for controlled field exposure of seedlings, where the field is usually a site external to the laboratory or greenhouse facilities.

Growth chambers, CSTRs, and open top chambers are exposure technologies that have inherent changes to ambient conditions. Air temperature, light intensity, and relative humidity inside and outside tile chambers were monitored at the majority of study sites to document these differences and, where possible, to maintain these important determinants of plant growth within predetermined ranges to simulate ambient conditions.

Treatments

Treatments were applied to the seedlings during periods of active growth during intervals varying from 10 wk to 28 wk for most of the earlier studies. In more recent studies, treatments spanned both tile growth and dormant seasons during a single year (Hogsett et al.,1989; Leininger et al.,1991) and over multiple years (Stow et al., 1992; Byers et al., 1992). Simulated precipitation was generally applied as rain or mist, and in a few

instances as fog (Hogsett et al., 1989), usually to foliage and rooting medium in quantities and patterns reflecting historical trends for a specific region. Ozone was applied over regulated time intervals, usually during daylight hours. The applications varied among studies, but were of two general types. Static exposures generally use a square-wave regime where a constant concentration of ozone is applied over a defined time interval during the day. In more complex designs, dynamic ozone exposures followed the monitored ambient concentrations for the region during a 24-h period, where ozone concentrations typically increased to a mid afternoon peak then decreased until dusk. Most of the current technology and research in exposure dynamics is oriented towards ozone, and patterned after past crop studies; however, the perennial nature of tree seedlings requires extended lengths of exposure, post-exposure measurements (Hogsett et al., 1989), multi-year exposures and larger chambers (for a comprehensive review on the subject, see Hogsett et al., 1987).

Response Variables

Numerous variables were measured in the seedling exposure studies. These included individual tree and stand measurements; organic and inorganic analysis of foliage, wood, and soils; and physiological measurements. Measurement frequency ranged from several times during the treatments, to initial and final measurements, to measurements only at the termination of treatments. Visible effects included foliage discoloration chlorosis and/or necrosis) and foliage loss (senescence). Determination of growth effects involved some measure of seedling biomass (linear measures of stems, branches or roots; diameter of stem; mass of various components). Carbon allocation involved measures of photosynthetic rates, respiration rates, tissue chemistry (sugars, starch and nonstructural carbohydrats, photosynthetic per, or enzymes), and root/shoot ratios. Winter injury was examined as an interacting stress. In these cases, seedling responses were measured after treated seedlings were allowed to over winter at ambient temperatures or after tissues were exposed to simulated frosts. Foliar leaching involved some measure of solution chemistry of throughfall or of solutions in which treated tissues were leached. Details of objectives, methods, and measurements taken at each study site are summarized elsewhere (Peterson et al., 1989).

Statistical Methods and Data Quality

Building on exposure studies of crops in the National Crop Loss Assessment Network (e.g., see Heagle et al., 1983; Heck et al., 1988), the experimental designs were generally a variation of split plot or randomized blocks, with chambers rather than seedlings as the units of replication because chambers are the level at which the pollutant treatment can be applied.

Most seedling studies incorporated repeated measurements of growth and growth process variables. Data were analyzed via analysis of variance and regression techniques with the intention of statistically testing hypotheses. Information on statistical power and data quality ordinarily clarify the interpretation of these statistical tests, but the lack of this information can leave the interpretation of experimental outcomes in doubt.

The goal of data quality assessment in biological research is to provide procedures that document random and systematic error in treatment application and response measurements within defined (statistically supported) limits. Data quality is assessed in terms of accuracy, precision, comparability, completeness, and representativeness. These measures traditionally have been applied to analytical measurements and generally have been limited to one plant response variable, namely, final crop yield (Coffey et al., 1988). For the NAPAP nee seedling studies, procedures were developed to apply statistical quality control to dozens of biological and physical variables, including measurements and monitoring of microclimate factors such as temperature and humidity (Mickler and Medlarz, 1987; Cline and Burkman, 1989).

DISCUSSION

The primary objective of exposure research is to deter mine various ways in which plant condition is affected by pollutants and apply that knowledge to the population of plants in the field that may be affected by pollutants. It is important that the experimental designs offer a reasonable chance for the objective to be fulfilled (Table 1). Although an individual study may achieve its own objective, the confidence in the outcome of that study is increased when the same outcome is observed in other studies. If the responses of plants are observed among treatments in a way that is consistent (e.g., in magnitude, direction, and duration) across studies, we may be able to assess air pollutant elects on a variety of species, irrespective of experimental methods (e.g., location, chamber type, or exposure dynamics). Conversely, if we have studies that are different in experimental approaches and do not demonstrate consistent patterns of plant response, it is extremely difficult if not impossible to sort out causal relations.

After synthesizing results from recent NAPAP seedling studies (Peterson et al., 1989), we conclude that the most important considerations for evaluating experimental outcomes can be grouped under four categories: selection of response variables, level of statistical power, documentation of data quality, and scope of inference. In the discussions that follow, we offer examples to illustrate the importance of each of these major considerations and make recommendations for future research.

Response Variables

All measurements or variables known to be important or of high priority should ultimately relate to plant growth. In order to estimate accuracy and precision for growth and physiological measurements, it is best to separate variables measured by traditional analytical techniques from those requiring more subjective interpretation. For example, foliar injury has not correlated well, if at all, with changes in growth. It is poorly defined as to scale or standard (i.e., very subjective), and quite variable with high measurement error. Therefore, while an estimate of foliar injury may have specific use at a given site, in the absence of an assessment of data comparability between studies it does not appear practical for use in a comprehensive analysis which might combine results from experiments conducted by different investigators (Innes et al., 1993). In contrast, dry weight or N content of plant tissues can be determined by analytical techniques that follow well established standard operating procedures. An estimate of accuracy and precision can be determined by incorporating National Institute of Standards Technology certified weights and tissue samples. The use of control chart techniques allows the investigator to track the prescribed bounds for which the measurement process is in statistical control.

Direct tree growth measurements such as stem diameter (*D*) and total height (*H*), are commonly taken on seedlings. The repeated measures of these variables afford a nondestructive means of detecting incremental changes in growth or growth patterns, in contrast to harvesting subsamples of seedlings to measure biomass. Combining variables, however, such as D^2H, commonly employed as a surrogate for biomass, should be done with caution. Responses in *D* and *H* to the same treatment can differ in both magnitude and direction, a good example of which can be found in McLaughlin et al. (1988). And yet, differential changes in *D* and *H* can result in the same incremental change in the variable D^2H. Considering that changes in *H* are generally more important than changes in *D* for early seedling competitiveness, changes in D^2H from pollutant treatments may be misleading or at least mask effects on seedling *H*.

Height and diameter determinations are objective measurements. But, natural irregularities in the geometry of seedling stems and different seedling growth forms makes interpretation of these determinations difficult or subjective. Measurements of height have greater precision than diameter measurements for seedlings (Fig. 1). This is to be expected since the variance associated with replicate measurements of plant diameter is greater than that for plant height. In fact, analysis of repeated measurement data indicated that the precision of diameter measurements generally fell outside the specified tolerance established in the measurement quality objectives for this variable.

Instantaneous measurements of processes such as photosynthetic rate do not usually permit repeated measurements to estimate precision since these processes are both inherently dynamic and subject to change from the measurement procedures. However, information on processes such as photosynthesis, C allocation, nutrient uptake, and water relations should be provided in order to better understand plant response. Interpretation of plant response is enhanced when methods information is provided on the measurement process, such as time of day, season, and location on the plant.

Future study and publication of forest research should consider the lessons learned from recent quality assessments of short- and long-term data sets as applied to the selection and reporting of experimental variables. Researchers need to report sufficient quantitative information in publications to assist in decision making under conditions of uncertainty and to allow the journal reader the ability to make their own inferences within the context of their research experience. Variables must

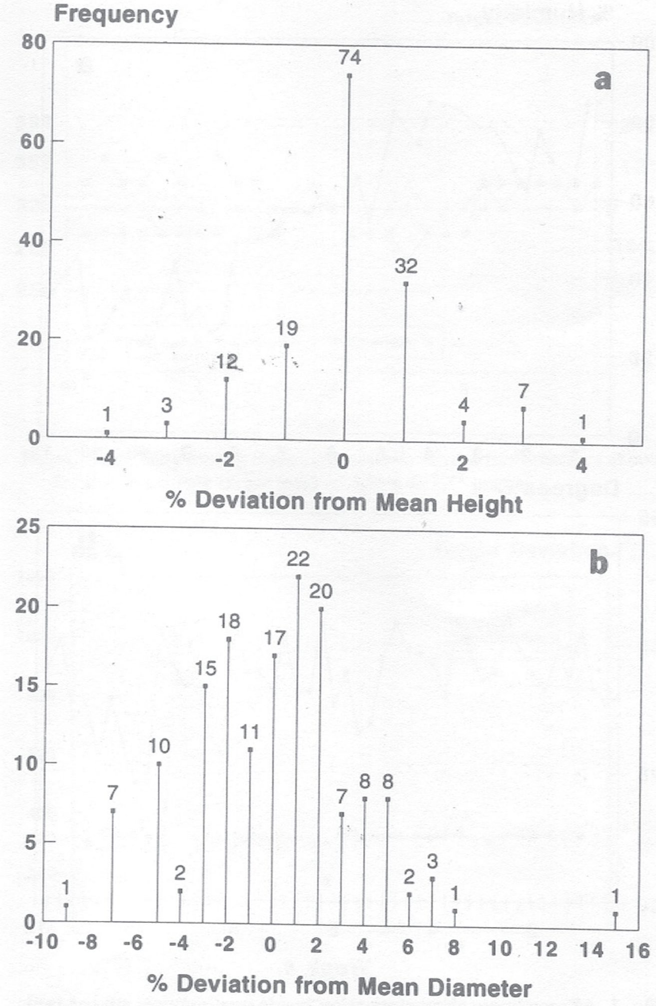

Fig. 1. Variability in height (*a*) and diameter (*b*) in one study from repeated measurement data collected weekly for 12 wk.

be well defined and the methods of measurement, frequency of measurement, and reporting units documented in archived study plans and summarized in the published literature. Ideally, data associated with all variables should be summarized in tables and the data referenced to facilitate retrieval from archived, well documented, and accessible data bases.

Our analyses of the uneven quality of reported forestry data indicates that shortcomings in the data could be significantly reduced by objective reporting of variables and associated data. Clearly, the ability to interpret research is dependent upon the presentation of experimental variables within the larger context of experimental design and the other four components of an experiment: hypothesis, experimental execution, statistical analysis, and interpretation (Hurlbert, 1984).

Statistical Power

The probability of detecting some consequential change in condition is of major importance and fundamental to the success of controlled experiments. Researchers want the power of the test (probability of correctly rejecting the null hypothesis) to be high. The consideration that should be given by policymakers and decision making managers to statistical power for testing research hypotheses has received renewed emphasis (e.g., see Cohen, 1988; McCaughran, 1977; Millard, 1987; Barnthouse et al., 1983), particularly for controlled exposure studies of annual crops (e.g., Rawlings, 1986) and tree seedlings (Peterson et al., 1989).

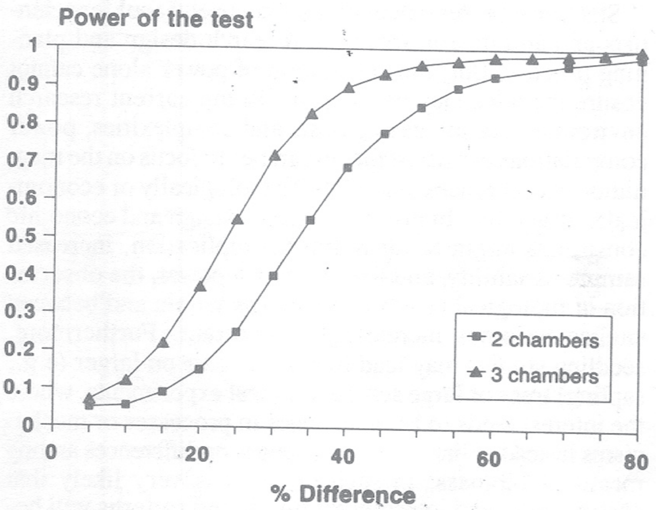

Fig. 2. Probability of detecting relative differences in biomass due to treatment was improved (i.e., power was increased) when replications were increased from two chambers per treatment to three chambers per treatment.

Although power is a function of the true parameter value (e.g., true mean difference between two treatments), sample size, sample variability, and parameter estimates can be used to estimate statistical power during the design stage of experiments. Scientists should address whether or not experimental accuracy and precision were a consideration in determining sample size. In seedling exposure studies, the number of replicates (i.e., chambers) usually influence power more than the numbers of seedlings used per treatment combination. Chamber replication, however, is considerably more costly than seedling replication and beyond a certain magnitude may provide only a small increase in the power of a statistical test. This can be determined by setting both the a level and power desired for a statistical test, and calculating the number of plants per replicate (varying the number of replicates) that are needed to detect various differences. One should expect the power of an experiment to be not only species specific, but to vary with the considered response variable, such as diameter, height, bud elongation, and biomass. A power curve for detecting relative differences for a biomass component in ponderosa pine (*Pinus ponderosa* Laws.) is given in Fig. 2, computed from a well replicated study[1] that used two and three chambers (treatment replicates) and >30 seedlings per treatment. Power was computed with Student's *t*-statistic for an a level of 0.05. As might be expected, there is a high probability ($P > 0.70$) of detecting a treatment effect when the differences in means are quite large (e.g., >40 %). Conversely, the power of the experiment to detect differences of <20% was quite low, particularly when limited to only two replicates. Because many seedling exposure studies are generally constrained to two chambers per treatment, we believe they are unlikely to detect treatment effects when treatment means

---

[1] Data on file with the USEPA, Environmental Research Lab, 200 S.W. 35th St., Corvallis, OR.

(e.g., for measures such as biomass) differ by <20 % . On the other hand, it is reasonable to ask whether or not detecting differences <20% is important. Our intention here is not to qualify the usefulness of any one study as a function of its power, but rather to put expectations for an experimental outcome into perspective. For example, scientists should address whether or not experimental accuracy and precision were considerations in determining sample size.

Statistical power calculations are a useful tool for scientists and administrators in the research design and planning process. But, the calculation of power alone cannot ensure the adequacy of a design. In the current research environment of increasing costs and complexities, power computations can assist the researcher to focus on the magnitude of differences that might be biologically or economically attainable. In instances where design and economic constraints might result in limited replication, increased sample variability, and low statistical power, the observation of biological trends and patterns within and between studies becomes increasingly important. Furthermore, seedling studies may lead to experiments on larger (e.g., sapling) trees or large scale ecological experiments, where the interest tends to be on changes in processes or mechanisms instead of linear growth changes or differences among means of biomass. In such cases, it is very likely that documenting and interpreting trends and patterns will become increasingly important in interpreting treatment effects.

Data Quality

Although treatment and response are equal partners in any exposure experiment, reporting results all too often focuses on measuring, quantifying, and interpreting response, without assuring that target levels of treatment are indeed achieved. Traditionally, variation in the response variables has received some consideration when accounting for measurement error as a component of experimental error, while the inherent variability associated with treatment application has been ignored or assumed to be zero. In the previous discussions we reviewed the variability in plant measurements (Fig. 1) and the importance of statistical power in designing the experiment (Fig. 2). Of equal concern is the variability in treatment application and measurements that are a major source of experimental error in some exposure studies.

Growth chambers are designed for maximum environmental control with some trade offs (e.g., chamber size limits the number of seedlings). Laminar air flow characteristics of chambers and associated pollutant delivery systems may produce significant horizontal and vertical gas concentration gradients above and within the tree seedling canopy. The result is that not all seedlings are exposed to the same level of treatment within the same chamber.

The subsequent development of CSTRs and the use of a mixing impeller improved the internal mixing of gases and tended to eliminate gas concentration gradients within the chambers; thus reducing the variability within treatments. The tighter control of the gas treatments is often achieved at the cost of controlling environmental variables. There are usually no environmental control systems associated with CSTRs. When placed in a conventional greenhouse, the CSTR is affected by the seasonal variation in environmental factors such as temperature and humidity throughout an experiment (Fig. 3), which in turn can affect seedling response to treatments. On the other hand, CSTRs enclosed in laboratories or climate controlled greenhouses offer reduced environmental variability.

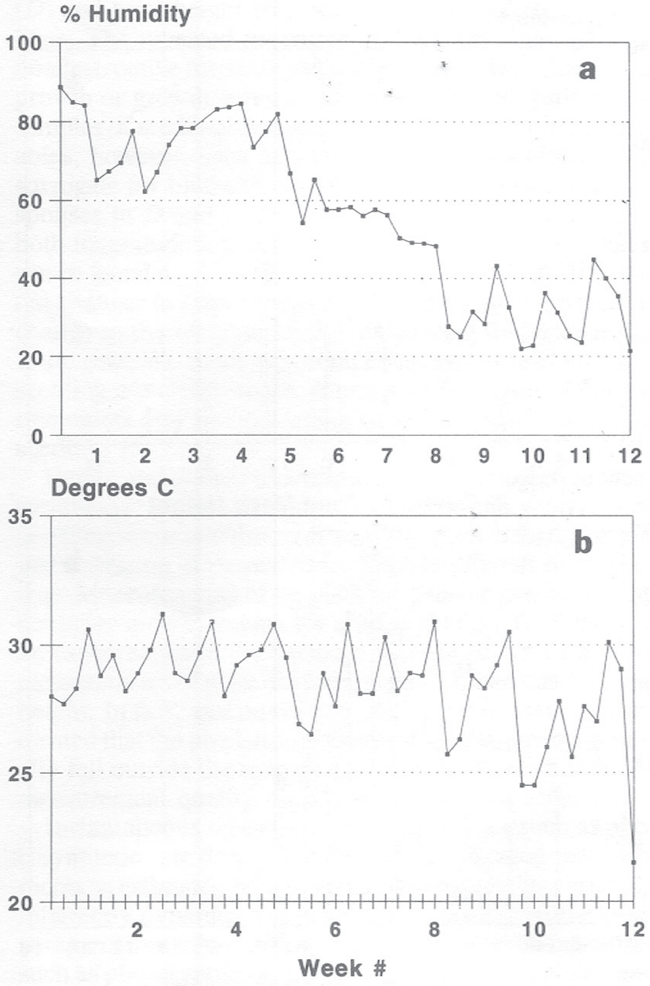

Fig. 3. Microclimate characterization inside continuously stirred tank reactors in a greenhouse, where percent humidity (*a*) decreased substantially over exposure period, and average weekly temperature (*b*) decreased with increasing variability over the same period.

Although CSTRs are designed for thorough mixing of gases throughout the chamber, there is still likely to be variability about the targeted exposure level (Fig. 4) within and among chambers. In projects reviewed for this study, many sites using CSTRs had varied amounts of manual control over the gas delivery systems. Although manual adjustments (i.e., based on operator judgement) to the gas concentrations within the CSTR contribute to this variability, the manual system can achieve uniform control of the treatments that meet the data quality objective (Fig. 4). As seen in Fig. 5, equipment malfunctions and/or operator error can produce even greater departures from exposure targets. In contrast, computer operated systems generally stop treatments when actual levels exceed preset limits bounding the target level. Clearly the po-

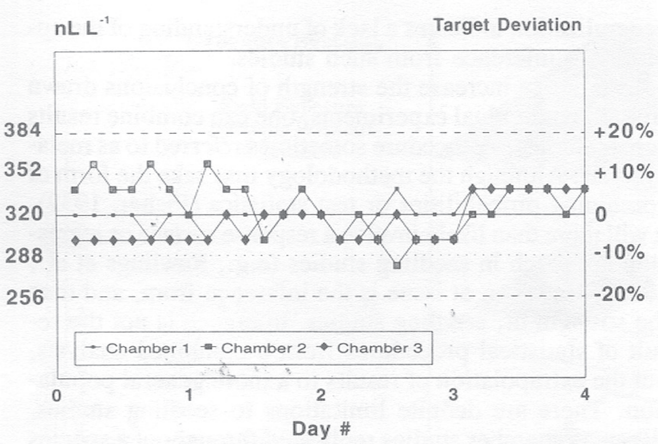

Fig. 4. Variability among three continuously stirred tank reactors at one site for treatment target level of 320 n L L^{-1} ozone.

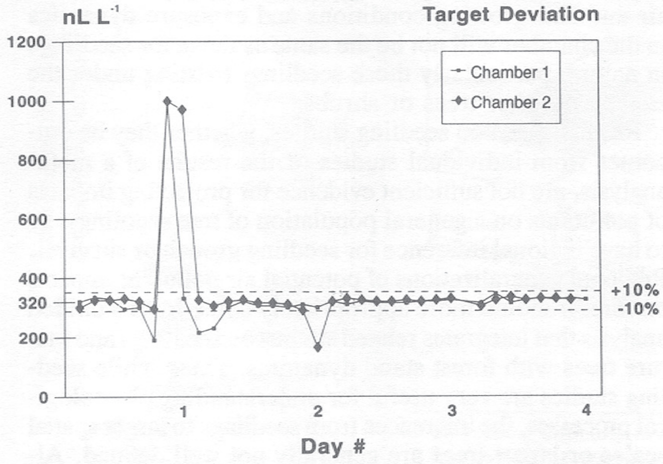

Fig. 5. Example of equipment failure during treatment application for treatment target level of 320 n L L^{-1} ozone.

tential exists for large departures from targets to affect plant responses.

The large number of open top chambers now used at many research sites requires computer controls for treatment application, and thus reduces the opportunity for operator error in correcting system malfunctions. Time-sharing of air quality monitors and custom computer software programs has significant impacts on the accuracy and precision of treatment application. Although a computerized technology is capable of applying fairly uniform treatments over time, environmental factors such as wind speed and direction can affect the incursion of ambient air, which in turn might increase variability in treatment (Fig. 6). Analyses in progress[2] indicate that rain exclusion covers lessened the incursion of ambient air through the top of chambers and improved the potential for maintaining target treatment levels. It is likely that factors such as the study type, exposure regime, and facility also contributed to the experimental error associated with the application of treatments, although these variance components have not been estimated.

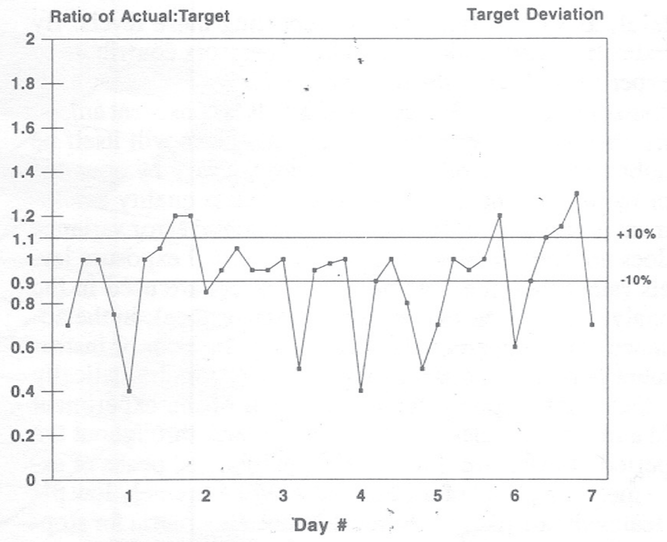

Fig. 6. Example from one study where actual daily ozone exposures during 1 wk consistently fell short of target value (2.25 × Ambient).

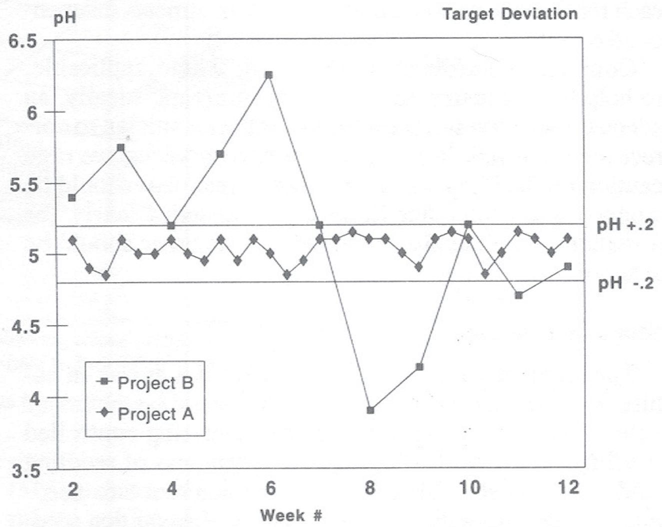

Fig. 7. Example from two studies with an acidic rain treatment target of pH = 5.0, where the variability in treatment levels from operator error (Project B) confound any comparison of that project's plant responses with those from the other study (Project A).

Acidic precipitation treatments, rain or mist, were used at most sites. Treatments were formulated in tank batches and distributed to individual chamber application systems. Although batch mixing allows for high precision in treatments being applied to large numbers of plots, operator error can account for fluctuations around target concentrations (Fig. 7). This type of experimenter error may make it difficult to statistically differentiate between treatments within projects and confound comparisons between projects.

The reduction and reporting of measurement and treatment error should be a major focus of experimental design and execution. Procedures should be identified that result in high precision, low detection limits, and low bias so that measure-

---

[2] Data on file with USDA Forest Service, 1509 Varsity Drive, Raleigh, NC 27606.

ment and treatment errors become statistically insignificant when compared with the population variance. This is an especially important issue when bias is not consistent across treatments within a study or between studies, and when operator bias associated measurement instrumentation is inconsistent.

We highly recommend establishing tolerance limits associated with target (nominal) values for each treatment level, as well as tracking and reporting these levels. By indicating how equipment and/or operators contribute to experimental error, the results offer opportunities for midcourse corrections. For example, if bias is consistent among treatments, the estimate of treatment effects will itself be unbiased. On the other hand, bias may vary by operator or by instrument, in which case, the data quality assessment is very important, since experimental error variance does not capture bias. As long as the actual exposure levels rather than the nominal target values are. used in the analysis (e.g., regression or response surface), in the absence of any apparent bias, large operator error or instrument failure (as shown is Fig. 5 and 7) can dramatically affect plant response for the duration of the experiment in a manner unrelated to the average level throughout the period of exposure. For example, unplanned peaks of extremely high level of ozone exposure or extremely low pH treatments are likely to be more important criteria for stopping a treatment than a consistent failure to meet target treatment values. Ideally, treatment errors should be avoided, but if such error can occur, we recommend that each replicate of each treatment receive treatments independently

Computer controls should be used, where applicable, to help reduce instrument or operator errors. Ideally, an adequate warning system would alert the technician to correct the treatment in a timely manner; however, we urge caution on deciding whether or not a treatment should be stopped when tolerance limits are exceeded. Clearly, the direction as well as the magnitude of the error should be a factor.

Scope of Inference

The ambient air quality, soil, and species at risk in nature (i.e., natural or artificial stands) should be the major determinants for designing and characterizing controlled seedling experiments. Clearly, the usefulness of seedling studies is to establish a baseline for inference to seedling populations of the particular species. Insofar that seedlings represent the future forest, treatment effects on seedlings may give some indication as to how those same seedlings might later develop into mature trees. However, while seedling studies might indicate where to look for impacts in mature trees and how to design appropriate studies, they should not be used to infer impacts directly to current forests of mature trees or to extrapolate to regional scales.

The discussions of Brennan and Harkw (1987) with Wang et al. (1987) on Wang et al. (1986), provide a recent example of the problems in misstating or misunderstanding inference from the outcomes of seedling experiments. Wang et al. (1986) found no foliar injury on seedlings that exhibited growth decreases. From this, the authors concluded that since foliar injury to seedlings did not correlate well, if at all, with seedling growth changes, than widespread or regional growth decreases are likely to be present in forests for which there is no visible injury. Their conclusions, based on the unassociated response between the growth and foliar injury, are questionable enough for application to other seedlings. Inferring that those same conclusions might be extended to mature trees with a regional generalization indicates a lack of understanding of the appropriate inference from such studies.

In order to increase the strength of conclusions drawn from the individual experiments, one can combine results across studies, a procedure sometimes referred to as metaanalysis. Although the methodology may take the form of combining probabilities or test statistics (Fisher, 1932), it will more than likely involve a response surface or regression approach in seedling studies (e.g., Rawlings et al., 1988). However, at issue is the inference from, and thus the strength of, seedling studies. Inference is not the result of statistical procedures from a combined analysis, but the extrapolation of results to a more general population. There are definite limitations to seedling studies. Whereas chamber studies replicated throughout a species distribution might be combined to better characterize dose response patterns, the inference is still to the subpopulation of seedlings in experimental studies, since the air and soil growing conditions ‘and exposure dynamics in the chamber will not be the same as those for seedlings in nature, particularly those seedlings existing under the canopy of other trees or shrubs.

Results based on seedling studies, whether they be outcomes from individual studies or the results of a metanalysis, not sufficient evidence for projecting impacts of pollutants on a general population of tree seedlings, or to have regional inference for seedling growth or survival. Regional generalizations of potential air pollutant impacts on forests should more appropriately come from a unified analysis that integrates related studies on seedlings and mature trees with forest stand dynamics. Thus, while seedling studies are very useful for understanding physiological processes, the inferences from seedlings to larger spatial scales or larger trees are generally not well defined. Although these inferences may seemingly provide criteria to assist policymakers under conditions of uncertainty, seedling studies alone cannot address their concerns.

Policy questions are growing increasingly complex, as any one answer seems to generate a half dozen new questions to answer or scenarios to evaluate. This suggests that investigators will have to consider increased complexity in experimental design in making their research more relevant to the field. Therefore, in addition to a Guderian type description of details in Table 1, we believe it is also critical to have in place some understanding of trade offs for decision making and planning of future exposure experiments. With the hope of provoking some thought in this direction, we are providing a framework in Fig. 8 with our own expectations as to how realistically one might expect some combinations of treatment systems and exposure dynamics to satisfy objectives such as establishing cause effect relationships or making inferences to populations in nature.

RECOMMENDATIONS

The first recommendation for future and ongoing exposure studies is to recognize that treatments applied to tree seedlings and other long lived perennials are an intervention experiment. Unlike annual crop exposure studies, in which growth

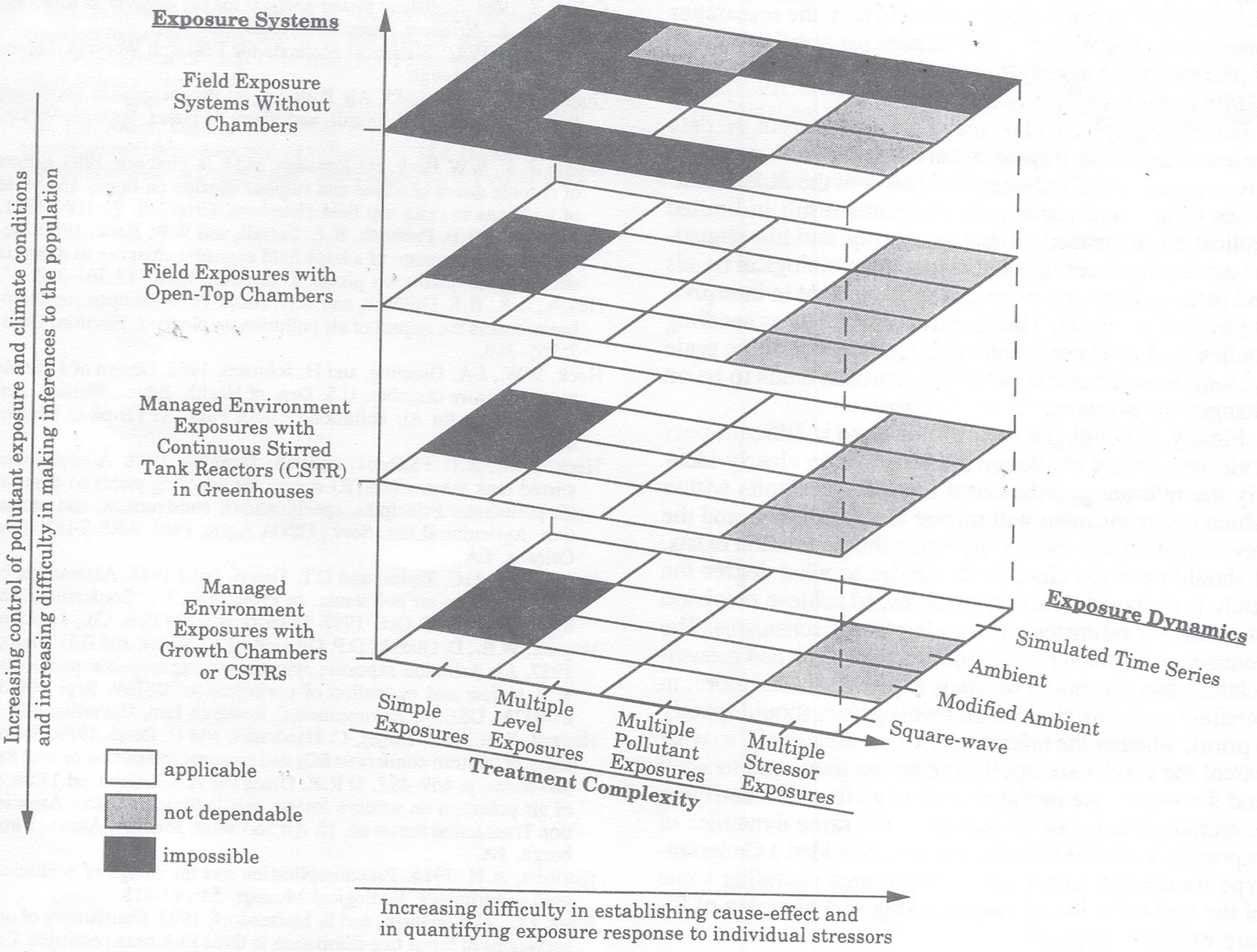

Fig. 8. Expected tradeoffs for assessing data and for designing experiments from various exposure dynamics, systems, and treatments.

is assessed as a measurement of final yield at the end of the plants life cycle, the final growth of a tree as a measure of yield is not possible. During the life-cycle of a tree and within the time frame of the experiment, morphological and physiological changes, such as flowering and C allocation, occur that may either buffer the response to treatments, cause the response to be delayed, or allow for a recovery from pollutant stress. Therefore we need to recognize that tree seedling experiments need multi year observations before, during, and following the period of treatment application. In addition, we need to recognize that data interpretation and inferences made from tree seedling experiments are representative of a small window of time in the growth and mortality stages of a tree.

Secondly, there is the issue of how to make use of data quality as an objective criteria for analyzing data, interpreting results, and making decisions under conditions of uncertainty. Variation in the response variables has received some consideration when accounting for experimental measurement error, but this information is rarely quantified by researchers or reported in the literature. The inherent variability associated with treatment application as a component of measurement error has traditionally been ignored or assumed to be zero. Researchers who ignore data quality may underestimate the overall variance of experimental data that can invalidate the results of individual experiments, confound meta analyses, and increase uncertainty to such an extent that the results become useless to policymakers. If investigators can identify measurement errors before or during the conduct of the experiment, and adjust the source, the experimental error will be reduced, estimates will be more precise, and tests will be more powerful. Researchers should report data quality information for treatments and response variables so the reader can make his or her own inferences on the basis of the data and their quality.

Thirdly, statistical power calculations are useful in aiding decisions with regard to the funding, managing, and design of ecological research. Although our discussions of statistical power focused on testing hypotheses of treatment differences, power should also be considered by any investigator interested in estimating the magnitude of treatment effects, or describing a response curve across various levels of treatment, in which case the standard error should accompany all reported parameter estimates. We recognize that the calculation of power alone cannot ensure the adequacy of a design; however, given the significant investment of scarce resources required for any ecological study, the power computations should help the researcher focus on the magnitude of differences (or standard errors of parameter estimates) that are biologically meaningful and/or economically attainable.

Fourthly, graphic and pictorial presentations of the data are essential in portraying experimental outcomes of interest, regardless of the estimated power of the experiment. When design and economic constraints result in limited replication, increased sample variability, and low statistical power, the observing and reporting of biological trends and patterns becomes increasingly important in interpreting treatment effects. This is particularly true as seedling studies lead to experiments on larger trees or large scale ecological experiments, where the interest tends to be on changes in processes or mechanisms.

Finally, although the issue of inference is difficult, peer reviewers should challenge investigators to clearly identify the relevant population of interest, the limits within which the experiment will mirror the population, and the environment (soil and air) in which the population exists. It should be made clear to the reader to what degree the study is designed to test hypotheses, to achieve precision of parameter estimates, or to understand mechanisms. The connection between process mechanism work and growth related changes should be made explicit. Furthermore, in seedling exposure studies, all investigators should specify a priori, whether the inference is for seedlings only, to what extent the results are applicable to the field, and to what end the results are useful in understanding or identifying potential mature tree responses to the same dynamics of exposure. With this in mind, we have provided a Guderian type framework as a place to begin understanding some of the trade-offs for decision-making end planning of future exposure experiments.

ACKNOWLEDGMENTS

This paper is based on research conducted under the Forest Response Progrpm, a cooperative program of the US. Environmental Protection Agency and the USDA Forest Service, with support from the National Council of the Paper Industry for Air and Stream Improvement, (NCASI). This paper has not been subject to USEPA or Forest Service policy review and should not be construed to represent policies of either agency or of NCASI. We thank Drs. Henry Lee, John Laurence, and C. Jeffrey Brandt for helpful comments.

REFERENCES

Barnthouse, L W , W. Van Winkle, and US. Vaughan. 1983. Impingement losses of white perch at Hudson River power plants: magnitude and biological significance. Environ. Manage. 7(4):355-364.

Brennan, E., and R.S. Harkov. 1987. Comment on Wang et al. (1986). Environ. Sci. Technol. 21:606-607.

Byers, D.P., TA Dam, and J.D. Johnson. 1992. Long term effects of ozone and acid rain on foliage dynamics of slash pine (Pines elliottii vat elliottii Engehm.), New Pbytol. 120:61-67.

Cline, S.P, and W.G. Burkman. 1989. The role of quality assurance in ecological research programs. p. 361-365. In LB. Bucher and 1. Bucher Wallin (ed.) Proc. of the Fourteenth Int. Meeting for Specialists in Air Pollution Effects on Forest Ecosystems. 2-8 Oct. 1988 Birmensdorf, Switzerland. lat. Union of Forestry Res. Organization, Project Group P2.05, Interlaken, Switzerland.

Coffey, US., J.C. Sprenger, UT. Tingey, G.E. Neely and J. McCarty. 1988 National Crop Loss Assessment Network: Quality Assurance Program. Environ. Pollut. 53:89-99

Cohen, J. 1988 Statistical power analysis for the behavioral sciences. 2nd oil. L. Erlbaum Assoc., Hillsdale, NJ.

Fisher, R.A. 1932: Statistical Methods for Research Workers. Oliver & Boyd, Edinburgh.

Guderian, R. (ed.). 1985. Air Pollution by Photochemical Oxidants: Formation, transport, control, and on plants. Springer Verlag, New York.

Heagle, A.S., W.W. Heck, J.0. Rawlings, and R.B. Philbeck. 1983. Effects of chronic doses of ozone and sulphur dioxide on injury and yield of soybeans in open top field chambers. Crop Sci. 23:1184-1191.

Heagle, A.S., R.B. Philbeck, R.E. Ferrell, and W.W. Heck. 1989. Design and performance of a large field exposure chamber to measure its of air quality on plants. J. Environ. Qual. 18:361-368

Heagle, A.S., R.B. Philbeck, and W .W. Heck. 1973. An open top chamber to assess the impact of air pollution on plants. J. Environ. Qual. 2:365-368

Heck, W.W., LA. Dunning, and H. Johnson. 1988 Design of a simple plant exposure chamber. US. Dep. of Health, Educ., Welfare. National Center for Air Pollution Council Publ. APTD-68-6 Cincinnati, OH.

Heck, W.W., R.B. Philbeck, and J.A. Dunning. 1978. A continuous stirred tank reactor (CSTR) system for exposing plants to bus air pollutants: Principles, specifications, construction, and operation. Agricultural Res. Serv., USDA Agric. Publ. ARS-S-181. New Orleans, LA.

Heck, W.W., O.C. Taylor, and D.T. Tingey. (ed.) 1988. Assessment of crop loss from air pollutants. ***In*** Proc. of an Int. Conference, Raleigh, NC. 25-29 Oct. 1987. Elsevier Science Pub. Co., London.

Hogsett, W.E., D. Olszyk, D.P. Ormrod, G.E. Taylor, and D.T. Tingey. 1987. Air pollution exposure systems and experimental protocols: 1. A review and evaluation of performance. USEPA Rep. 600/387/037A. USEPA, Environmental Research Lab, Corvallis, OR.

Hogsett, W.E., D.T Tingey, C. Hendricks, and D. Rossi. 1989. Sensitivity of western conifers to SO_2 and seasonal interaction of acid fog and ozone, p. 469-492. ***In*** R.K. Olson and A.S. Lefohn (ed.) Effects of air pollution on western forests. Air Pollution Control Association Transaction Series no. 16. Air and Waste Manage. Assoc., Pittsburgh, PA.

Hurlbert, A.H. 1984. Pseudoreplication and the design of ecological field experiments. Ecological Monogr. 54:187-211.

Innes, J.L., G. Landman, and B. Mettendorf. 1993. Consistency of observations of forest tree defoliation in three European countries. Environ. Monit. Assess. 25:29-40.

Leininger, T.D., P.R. Miller, S.L. Schilling, and P.H. Dunn. 1991. Seedling responses of five species of western conifers to simulated ambient sulfur dioxide exposures. For. Sci. 37(6):1538-1549.

Mattson, K.G., L.Y. Arnaut, G.A. Reams, S.P. Cline, C.E. Peterson, and R.J Vong. 1990. Response of forest trees to sulfur, nitrogen, and associated pollutants. USEPA Rep. N104/F/06/M66/7756A. USEPA Environmental Research Laboratory Corvallis, OR.

McCaughnn, D.A. 1977. The quality of inferences concerning the effects of nuclear power plants on the environment. p. 229-742. *In* W. Van Winkle (ed.) Proc. of the conference on assessing the effects of powerplant-induced mortality on fish populations. Gatlingburg, TN. 3-6 May 1977. Pergamon Press, New York.

McLaughlin, S.B., M.B. Adams, N.T Edwards, P.J. Hanson, P.A. Layton, E.G. O'Neill, and W.K. Roy. 1988. Comparative sensitivity, mechanisms, and whole plain physiological implications of responses of loblolly pine genotypes to ozone and acid deposition. ORNL/TM10777, Environmental Sciences Division Public. no. 3105. National Technical Information Service, Springfield, VA.

Mickler, R.A., and S. Medlarz. 1987. The role of quality assurance in national acid rain research in the U.S. Environ. Tech. Let. 8:459-466.

Millard, S.P. 1987. Environmental monitoring, statistics, and the law: Room for improvement. Am. Stan. 41:249-253.

Peterson, C.E., K.G. Mattson, and R.A. Mickler. 1989. Seedling response to sulfur, nitrogen, and associated pollutants. USEPA Rep. 600/3-89/081. USEPA Environmental Research Laboratory, Corvallis, OR.

Rawlings, J.O. 1986. Design of experiments for controlled exposure studies. p. 36-47. *In* Proc. of workshop on controlled exposure techniques and evaluation of tree responses to airborne chemicals. Atlanta, GA. 4-5 Feb. 1986. Tech. Bull. 500. National Council of the Paper Industry for Air and Stream Improvement, New York.

Rawlings, J.O., V.M. Lesser, and K.A. Dassel. 1988. Statistical approaches to assessing crop losses. p. 389-416. In W.W. Heck, et al. (ed.) Assessment of Crop Loss from Air Pollutants. Elsevier Applied Science, New York.

Shriner, D.S., W .W. Heck, S.B. McLaughlin, D.W. Johnson, P.M. Irving, J.D. Joslin, and C.E. Peterson. 1990. Response of vegetation to atmospheric deposition and air pollution. p. 18-1-18-206. *In* Acidic Deposition: State of Science and Technology. Vol. 3. NAPAP SOS/T Rep. 18 National Acid Precipitation Assessment Program, Washington, DC.

Stow, T.K., H.L. Allen, and L.W. Kress. 1992. Ozone impacts on seasonal foliage dynamics of young loblolly pine. For Sci. 38:102-119.

Wang, D., F.H. Bormann, and D.F. Karnosky. 1986. Regional tree growth reductions due to ambient ozone: evidence from field experiments. Environ. Sci. Technol. 20:1122-1125.

Wang, D., F.H. Bormann, and D.E. Karnosky. 1987. Comment on Brennan and Harkov. Environ. Sci. Technol. 21:607-608.