Index Construction for Foliar Symptoms of Air Pollution Injury

PATRICIA S. MUIR and BRUCE McCUNE, Holcomb Research Institute, Butler University, Indianapolis, IN 46208

A variety of disease agents, including air pollutants, produce visible symptoms on foliage, and these symptoms are potentially useful for identifying causal agents, for estimating doses, and, in some cases, for indicating effects on yield (5,19,27,30,44,53). In diseases induced by biotic agents, pathologists often can rely on such signs as fruiting structures and mycelia, as well as on visible symptoms, to identify the causal agents (1). Because such signs are lacking in diseases induced by air pollutants, however, visible symptoms must often be relied on to aid diagnosis, in company with information on the pollution regime of a site.

The importance of visible symptoms in studies of disease induced by air pollution hinges on the ability to relate symptoms to causal agents and biological effects. Such relationships are not always clear. The visible symptoms produced by some air pollutants are not unique to one particular pollutant, and symptoms of other diseases often mimic those produced by air pollutant exposure (25,30,32,48). Furthermore, it is not always clear that visible foliar injury relates to other biological effects such as growth. In some cases, foliar injury of trees is correlated with impaired growth (5). In other cases, injury may be "invisible" (51), and tree growth is reduced without apparent foliar injury (41,43). Finally, some crops have "tolerance" to pollutants, and visible injury occurs without corresponding growth reductions (19).

Despite these uncertainties, studies of air pollution effects on vegetation often rely in part on visible foliar injury to indicate presence and severity of effects (19,25,30,33,37,39,44,48,53). Much effort has gone into elucidating which symptoms relate least ambiguously to the pollutant(s) of concern and into methods for evaluating these symptoms in the field and in controlled experiments (15,20,25,27,30,32,33,39,47,53). Methods for quantifying foliar symptoms (and signs) induced by many pathogens have been standardized and widely accepted (22,26,27). In contrast, few standardized methods for symptom quantification have been widely adopted in the air pollution field (15), although the need for standardization has been recognized (20). For example, the methods of Phillips et al (41), Benoit et al (5), and Usher and Williams (56) have little in common even though all evaluate visible foliar injury on white pine in response to pollutants.

Here we evaluate some quantitative methods for evaluating foliar symptoms and for aggregating data on multiple visible symptoms into composite indices of injury. Although the examples presented concern symptoms induced by air pollution, the problems discussed are generally relevant to pathologists involved in disease assessment. Others have discussed methods of combining symptom data from individual plants into indices reflecting injury at the population level, as in the McKinney (36) index as described by Horsfall and Cowling (23). We are concerned here with injury indices created by combining data on several symptoms into one score to reflect injury to a given sample unit.

Present address of authors: Department of General Science, Oregon State University, Corvallis 97331.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked *"advertisement"* in accordance with 18 U.S.C. § 1734 solely to indicate this fact.

We discuss when to use indices, methods of quantifying symptom data, and techniques for standardizing, weighting, and combining symptoms into synthetic indices. Examples are drawn from our field data on white pine (*Pinus strobus* L.) in Indiana, from hypothetical data, and from the literature. Hypothetical data are used when they allow us to illustrate the effect of particular data manipulations on resultant indices more clearly than real data would. We point out strengths and weaknesses of various approaches to index construction and provide a series of recommendations. The power of indices depends to a large extent on the quality of reported data on symptomatology, and we identify specific areas from both experimental and field research where improved statistical reporting would promote the accuracy of foliar symptom analyses.

The pros and cons of indices

When injury is manifested by more than one foliar symptom (as in ozone effects on white pine foliage [10]), synthetic indices (composites of symptoms) are potentially useful. Such indices may indicate general pollution stress, levels or impacts of a particular pollutant, or growth effects. The strengths of synthetic indices relative to symptom-by-symptom analyses include: 1) reduction of "noise" associated with individual symptoms by emphasis on common signals among symptoms, 2) mathematical integration of information provided by each symptom into a smoothed and interpretable expression of each entity's (individual, population) position relative to other entities on a pollutant or effects gradient, and 3) summarization of symptom data into a single variable, facilitating analysis in relation to other variables. However, symptom-by-symptom analyses are also often useful, as they enable one to examine the relationship of each symptom to dose or to growth, which may be helpful in deciding which symptoms are the most reliable indicators. Also, the total information on symptomatology is retained; such information may be particularly useful where complex gradients involving more than one pollutant exist and where, consequently, an index to generalized "air pollution damage" may oversimplify and obscure patterns of interest. Hence, it is useful, if not essential, to collect data on individual symptoms, even when the data will be combined into a composite index of injury.

Data collection and standardization

Other workers have dealt with physiological, chemical, and visual approaches for assessing the influence of air pollutants on vegetation (19,20,25,30,33). Here we mention briefly four areas directly relevant to collecting data on the visible foliar injury component of air pollution's influence on vegetation. We assume that symptoms have been carefully chosen to represent effects of the target pollutant(s).

Level of resolution. At what level should symptoms be measured—the level of the individual leaf, the fascicle (in pines), the branch, the tree, or the population? This decision involves a compromise among replicability, time, and the goals of the research project. The level should be fine enough to allow rapid and repeatable visual assessment, yet coarse enough for efficient data collection and summarization, and at or below the level at which an estimate of variance is desired. As is true throughout this paper, the best method depends on the goal of the study. No general recommendation is possible concerning choice of level of resolution, save that the resolution level of the given hypotheses or questions should always be examined (2).

Individual symptoms or composites. Although most investigators record data on individual symptoms, some combine symptoms into a composite index at the data collecting stage, recording only the composite score (Table 1). Component ratings cannot be derived from a composite rating. Recording data on individual symptoms, however, allows maximum flexibility for subsequent analyses, and composite indices can be constructed later.

Continuous or discrete. The investigator must decide whether to collect data as continuous data or in discrete classes. Although continuous data provide maximum flexibility for subsequent manipulation, class data are collected faster and do not pretend to more accuracy than is realistically achievable. Further, analyses of classed data generally yield results similar to those of continuous data, providing the classes are not too broad (49). For most purposes, therefore, we advocate the use of carefully chosen classes.

Definition of classes. The type of classes into which data will be grouped influences the reliability and reproducibility of the data and can influence the weight given to different disease levels. Qualitative disease categories (slight to moderate to heavy), unless defined in quantitative terms as in Miller (38), suffer from subjectivity and lack of reproducibility between observers or between days for a given observer (7).

A common alternative to qualitative categories (15) is equal-sized classes of percent cover of symptoms (e.g., 0–5%, 5–10%, etc.). A drawback to this method is that accuracy of visual discrimination among disease levels is not always equal at all levels of disease (23,27,42). Discrimination is most acute at relatively high and low disease percentages and least acute in the mid ranges. (There are exceptions to this generalization, however [17].) Equal-sized classes also give equal weight to all parts of the disease scale, which may or may not be desirable, depending on the linearity of the relationship between disease levels and dose or biological effects.

In general, we advocate the use of classes approximating an arc sine–square root transformation of proportion data (Fig. 1), or scales that are logarithmic at both tails (21,22). Such classes spread values in the distribution tails, thus increasing sensitivity to high and low disease levels, and contract those in the middle, allowing for broader categories in the mid ranges where visual discrimination between disease increments is more difficult. The arc sine–square root (angular) transformation also homogenizes variances of proportion data, a necessary consideration for statistical testing (50). Zero disease should be assigned to its own class when the absence of disease deserves weight.

Constructing an index

The method used to construct indices of disease severity can influence conclusions about the severity of disease in trees or populations, and indices should be carefully designed to most faithfully represent disease severity levels. A wide variety of indices reflecting foliar injury has been used in published studies; Table 1 lists examples. Three indices of air pollution

Table 1. How some published foliar symptom indices meet the criteria of a desirable index, i.e., affirmatives for the first five characteristics, additive, and no subjective categories

Index (reference)	Defined minimum response?	Maximum score at maximum disease?	Monotonic?	Individual symptoms scored quantitatively?	Variables standardized to equal weight?	Additive, multiplicative, or intermediate?	Subjective categories?
Linzon, 1960 (31)	Yes	Yes	Yes	Yes	Yes	Additive	Some
Usher and Williams, 1982 (56)	Yes	Usually[a]	Usually[a]	Yes	Nearly[b]	Additive	Some
Rice et al, 1983 (45)	Yes	Usually[a]	Usually[a]	Yes	Yes	Additive	None
Ashmore et al, 1980 (4)	Yes	Usually[c]	Usually[c]	Yes	NA[d]	NA[d]	None
Stolte and Bennett, 1984 (52)	Yes	No	No	Yes	No	Multiplicative	Some
Costonis, 1973 (9)	Yes	Yes[e]	Yes[e]	Yes	NA[d]	NA[d]	None
Houston, 1974, index II (24)	Yes	Yes[e]	Yes[e]	Yes	NA[d]	NA[d]	None
Miller, 1973 (38)	Yes	Yes[f,g]	Yes[f]	Yes	Nearly	Additive	Some
Toole, 1949 (55)	Yes	Yes	Yes	Yes	Nearly	Additive	Some
Benoit et al, 1982 (5)	Yes	?[g,h]	?[h]	Yes	Nearly	Additive	Some
Phillips et al, 1977 (41)	Yes	Yes[e,g]	Yes[e]	No	NA[i]	NA[i]	None
Genys and Heggestad, 1983 (13)	Yes	Yes	Yes	No	NA[i]	NA[i]	?[j]
Jensen and Noble, 1984 (28)	Yes	Yes[e]	Yes[e]	No	NA[i]	NA[i]	Some
Houston, 1974, index I (24)	Yes	Yes	Yes	No	NA[i]	NA[i]	None
Saari and Prescott, 1975 (46)	Yes	Yes	Yes	No	NA[i]	NA[i]	Some
McKinney, 1923 (36)	Yes	Yes	Yes	No	NA[i]	Additive	Yes
Marsh et al, 1937 (35)	Yes	Yes	Yes	No	NA[i]	NA[i]	Yes
Todd and Arnold, 1961 (54)	Yes	Yes	Yes	No	NA[i]	Additive	?[j]
Davis and Wood, 1973 (11)	Yes	Unclear	Unclear	No	NA[i]	NA[i]	Yes
Blanchard et al, 1979 (6)	Yes	Yes	Yes	Yes	No[k]	Additive	Some
Hayes and Skelly, 1977 (16)	Yes	Unclear	Unclear	Yes	No[k]	Additive	None
Kress and Skelly, 1982 (29)	Yes	Unclear	No	Yes	Yes	Additive	None

[a] Criterion is met at low to moderate pollution levels but not at more extreme pollution levels where full expression of one or two symptoms precludes expression of other symptoms.
[b] All variables are standardized; however, needle necrosis is entered twice into index—once as length of necrosis and once as percent of needles showing necrosis.
[c] Value depends on disease level at start of study, which is subtracted from disease level at end of study.
[d] Index is based on a single symptom.
[e] Criterion is met at low to moderate pollution levels but may not discriminate successfully among higher pollution levels when symptoms are more severe than top scoring category or when top scoring category is open-ended, e.g., >25%.
[f] Value depends on number of needle whorls scored.
[g] Inverse scale.
[h] Chlorotic dwarf-type foliage is scored as maximal symptom expression.
[i] Symptoms are mentally integrated by scorer and not scored individually.
[j] Not clear from publication.
[k] Symptoms are consciously assigned differential weights.

injury to white pine foliage (5,52,56) were compared using data on white pine foliage from populations at the Indiana Dunes National Lakeshore (3). Application of the three indices to these foliar data resulted in markedly differing distributions of apparent disease severity (3) (Fig. 2).

The discussion that follows applies when injury is assessed as percent area (expressed continuously or in classes) of the study unit, such as a leaf, affected by each symptom and when the goal is to combine data on more than one symptom into a composite index. For convenience, individual leaves are considered as sample units, although the principles apply to sample units at any level of resolution. The discussion does not apply to the problems of combining data from individual leaves into a composite index of plant injury.

Standardizing symptoms. Initially, component symptoms should be scaled to have equal means and equal variances, thus giving equal weight to each symptom. If desired, differential weights can be applied later. It makes little sense to add variables such as necrosis length (0–10 cm) and percent needle retention (0–100%) into an index without rescaling them, as the variables with the larger numeric values will overpower those with the smaller.

Symptoms can be scaled to equal weight and variability by:

1. Expressing each symptom as a standard normal variate in units of standard deviations. This is done by standardizing each symptom by its mean and standard deviation in the relevant population: standardized symptom = $(x_i - \bar{x})/s$, where x_i = symptom score for individual i, $\bar{x}$ = sample mean of symptom scores, and s = sample standard deviation of symptom scores. This converts each score into a standard normal variate with $\bar{x} = 0$ and $s = 1$.

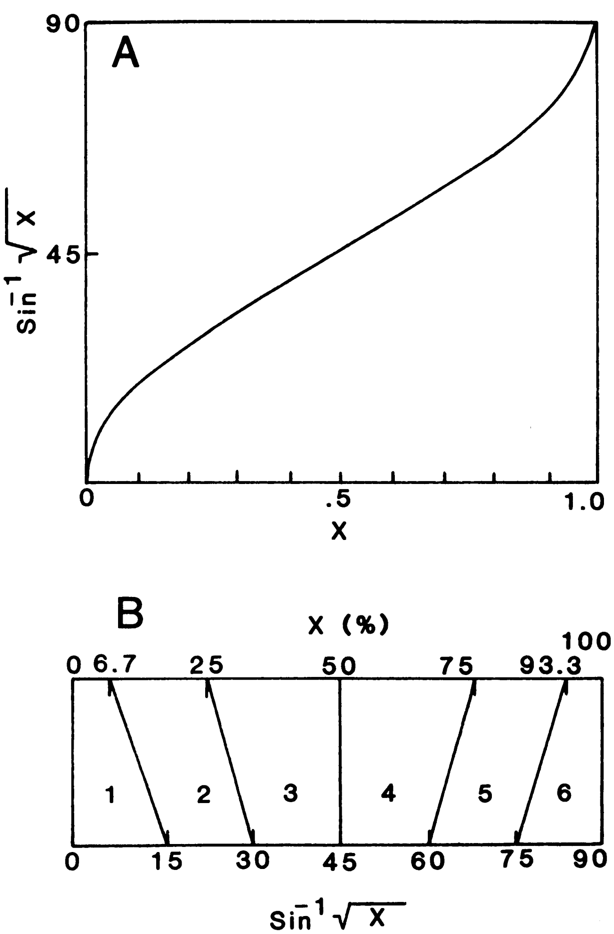


Fig. 1. Arc sine–square root function. **(A)** Relationship between observed values, x, expressed as proportions (= percent/100) and the arc sine–square root transformation of x. **(B)** Cover classes based on cover percentages transformed by arc sine–square root (percent data are converted to proportions before transformation).

2. Dividing each symptom into discrete classes covering the same range, e.g., 1–5 reflecting increasing injury. By using the same classes for each symptom (56), all symptoms are expressed in comparable units and scaled to similar ranges and variability.

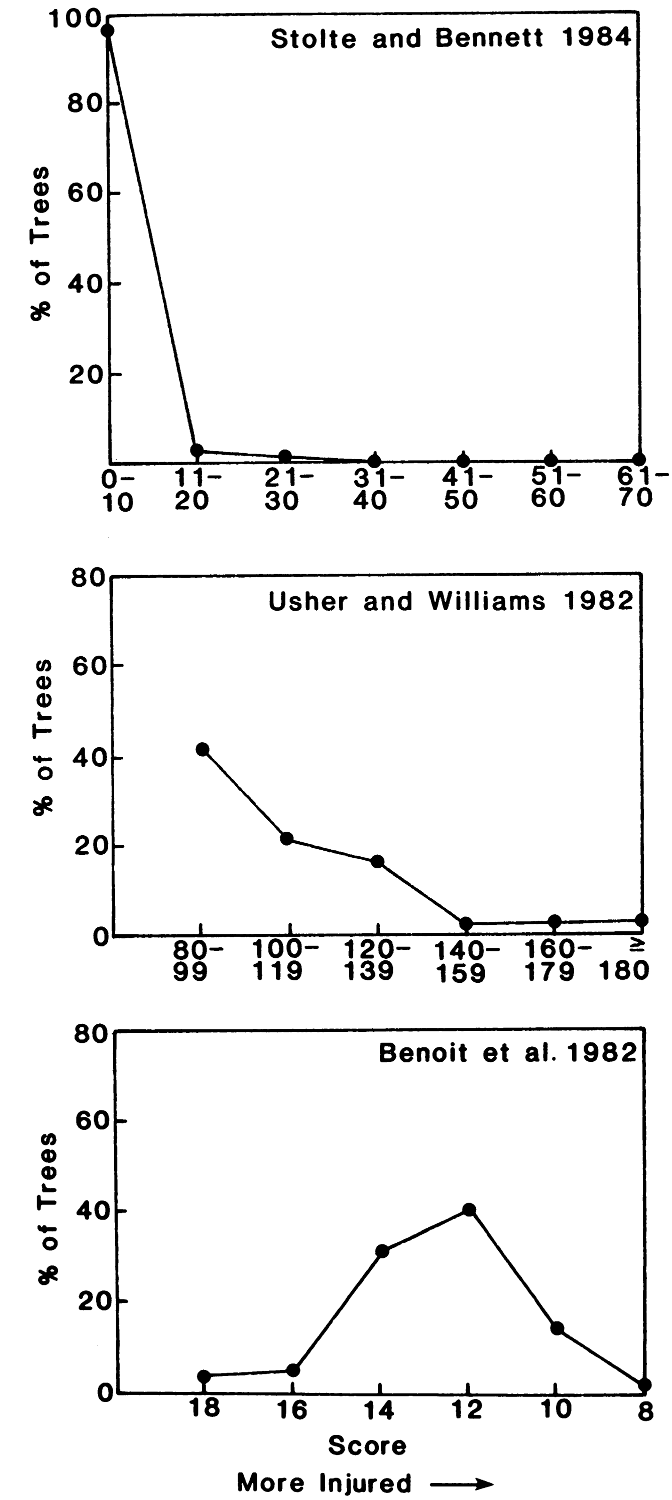


Fig. 2. Disease index scores for eastern white pine (*Pinus strobus*) at Indiana Dunes National Lakeshore (3). Symptoms of air pollution injury to needles (tip necrosis, chlorotic mottle, needle retention, and needle length) were evaluated for 96 trees. For each individual, symptoms were combined into injury indices using three methods—those of Stolte and Bennett (52), Usher and Williams (56), and Benoit et al (5). The distribution of disease severity levels among trees, as calculated by the three methods, is plotted for each injury index. The choice of symptoms and the method of combining symptoms into disease scores influenced conclusions about the severity of disease attributable to air pollution at Indiana Dunes National Lakeshore.

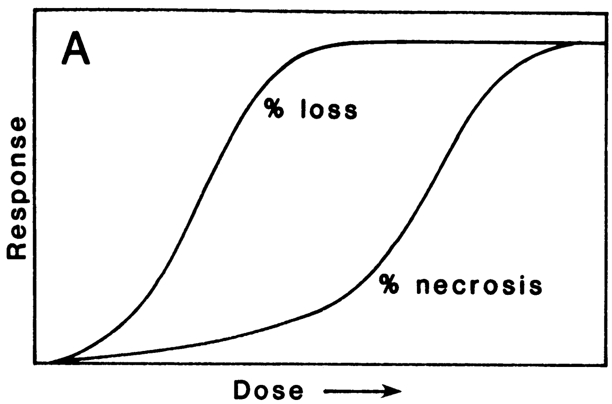
This approach is less appropriate for continuous data than transforming to a standard normal variate.

Weighting symptoms. After being standardized, symptoms may be weighted differentially, using, for example, a criterion of symptom reliability or sensitivity. If a good basis for weighting is lacking, it is best to use equal weights. When the index will consist of a large number of component symptoms and the possible differential weights are relatively small, the use of weights is unlikely to make a significant difference in the relative placements of sample units (49). In such cases, it is simplest to use equal weights.

Weighting by response to pollutants. If the necessary data are available, symptoms can be weighted in proportion to their sensitivity to pollution, their indicator value for growth, or any parameter being indexed by foliar condition. Figure 3A shows the effect of weighting symptoms differentially by response to pollutant dose; the sigmoid curve shapes and positions are hypothetical, but such responses to dose have been observed (18). In this example, needle loss is expressed at lower doses than needle necrosis and hence might be considered a more sensitive indicator of pollution presence, although not necessarily providing much information about the level of pollution, as the symptom becomes insensitive at higher doses. Figure 3B shows the indices resulting from summing these two symptoms, weighted and unweighted. In the base case, each symptom is weighted equally and the index increases more or less linearly across the dose gradient. In another case, needle loss is weighted twice as heavily as needle necrosis, as might be done if one were interested primarily in the response to low doses. This weighting exaggerates the increase at lower doses and renders the index insensitive at higher pollutant doses. In the final case, needle necrosis (which is not manifested until higher dose levels than needle retention) is weighted twice as heavily as needle loss. The resultant index is slightly insensitive at low doses but is subsequently fairly linear, even at the highest doses. The decision as to whether and how to weight symptoms by their sensitivity to pollutant dose depends on the purpose(s) of the investigation. Unless the study is designed to either emphasize or deemphasize responses to low levels of pollution, an index with a linear response to the dose gradient is generally desirable, as it is more straightforward statistically and conceptually. Since our comparisons show that differential weighting by symptom sensitivity reduces linearity, such weighting generally should not be used.

This example also demonstrates the need for accurate, quantitative data on symptom expression vs. dose for use in deriving differential weights for symptoms. We searched the literature on foliar responses of *Pinus strobus* to air pollution for data (mean and variance or regressions) on symptom expression vs. dose, hoping to use such data to derive biologically based weights. However, none of the many published studies of symptoms induced in *P. strobus* by air pollution provided the quantitative data needed to derive a sound, response-based weighting scheme. Future laboratory and field studies of the relationships of disease symptomatology to dose (biotic or abiotic causes) should report results in quantitative terms, including regressions or means and variances at different estimated doses, to increase the usefulness of study results.

In the hypothetical situation described, one symptom did not constrain the expression of another. In many cases, symptoms may be interdependent biologically and/or statistically. When constructed with groups of interrelated variables, the index is likely to have a nonlinear response to dose, as are the responses of individual symptoms. For example, assume *P. strobus* needles are being scored for percent of needle necrotic, percent chlorotic, and percent flecked (Fig. 4). The hypothetical data show that as needle necrosis and chlorosis increase, the percent of needle not necrotic or chlorotic decreases; thus, the percent of needle flecked must decrease eventually. Hence, an index constructed by summing the percent occurrences of each symptom for a given dose level is uninformative (the response curve is flat) at higher disease levels (Fig. 4). This nonlinearity in symptom expression vs. dose when symptoms are non-independent also may explain, in part, why linear correlations between individual symptoms and dose may be weak, particularly when the dose gradient is long. If response is always studied at low doses, these problems are not severe, but for maximum utility an index should increase continuously with increases in dose. Quantitative data on these types of



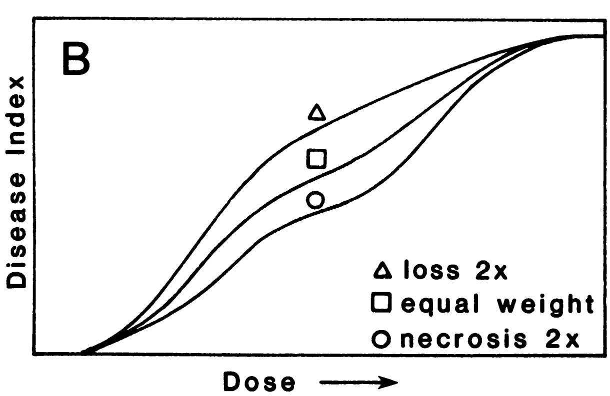


Fig. 3. (A) Hypothetical response of two symptoms—percent loss of needles and percent needle necrosis—across a pollutant dose gradient. **(B)** Disease indices as a function of pollutant dose. Indices derived by summing percent needle loss and percent needle necrosis: □ = symptoms given equal weight, Δ = needle loss given twice the weight of needle necrosis, o = needle necrosis given twice the weight of needle loss.

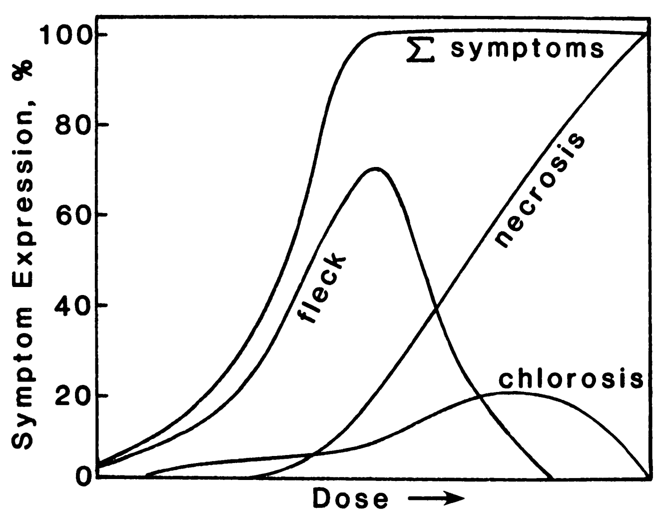


Fig. 4. Hypothetical response of three symptoms (needle necrosis, needle chlorosis, and chlorotic flecking), each expressed as percent of needle length, across a pollutant dose gradient.

interactions between symptoms are needed to choose, interpret, and weight symptoms with greater understanding.

Weighting by reliability. For most purposes, a reliable symptom may be defined as one that shows low variability at a given dose, yet varies over the range of doses. Weights may be based on symptom reliability, weighting most heavily those symptoms least variable at a given dose. Symptom variability is affected by the degree of genetic heterogeneity in the population and by the specificity of the symptom to the pollutant(s) of concern. For some studies, as in genetic variability in pollution tolerance where within-population variance in symptom expression is just as interesting as between-population variance, one may wish to weight symptoms that vary within populations if those symptoms are known to be sensitive indicators of genetic variability in pollution tolerance. Generally, however, the most useful symptoms are those that are most uniformly expressed in individual plants and consistently present at a given pollution level, and therefore least variable within populations, yet vary (with dose) among populations. Heavily weighting a symptom that is consistent within populations but varies little among populations differing in dose has little effect on differences in population level index scores because the symptom will contribute equally to the score for each population.

Symptom reliability can be estimated in at least three ways, depending on the available data:

1. When only minimal dose data are available (as is common in field studies of air pollution effects), reliability can be estimated by within-population variability in symptom expression. Rather than weighting by the within-population variance, whose size depends on the units of measurement, it is better to weight by the inverse of the coefficient of variation (C.V. = standard deviation ÷ mean), which expresses the variability as a proportion of the mean.

2. When dose data are available, reliability can be estimated by the observed coefficient of determination (r^2) for the regression of symptom scores on dose, which indicates the proportion of variance in symptom scores that is explained by the regression model.

3. Finally, weights can be derived from published reliability data (coefficients of determination or coefficients of variation) from either field or laboratory studies. Unfortunately, these results have seldom been published and are lacking even for *P. strobus*, a species whose visible symptoms in response to air pollution have been studied frequently.

Weighting with PCA. Symptoms may also be weighted using a multivariate approach, such as principal components analysis (PCA). (Useful introductions to PCA are found in Sneath and Sokal [49] and Greig-Smith [14].) In PCA, all symptoms are used simultaneously to provide information about the placement of populations (or other sample units) on a gradient of covarying symptoms. In effect, this approach weights each symptom in proportion to its shared covariance with other symptoms. Assuming that the symptoms are all responsive to pollution and no other single cause, the strongest common signal (covariance) should be pollution-induced. If so, the first principal component will represent most of the pollution-related variation in foliage condition. PCA combines the weighted symptoms additively to give an index score for each population. Unanticipated gradients beyond the known pollutant(s) of concern may influence the data; if so, PCA expresses these symptom gradients on subsequent axes. Thus, PCA can be used to study interrelationships among symptoms but, like any other method of index construction, only leads to an index indicative of pollutant dosage when the symptoms have been carefully chosen.

Weighting and collinearity. Foliar symptoms may be intercorrelated (collinear), and when collinear symptoms are included in an index, the cause of the collinearity is implicitly given weight proportional to the redundancy of the symptoms. Therefore, weighting the symptoms equally will not weight their underlying causes equally, and to avoid giving excess weight to causes other than pollution, correlations between symptoms and their possible causes must be examined. When correlations arise because the symptoms result from the same air pollutant, it is reasonable to include the collinear symptoms in an index. Indeed, part of the strength of an index in reducing noise comes from the redundancy with which the pollution signal is expressed in the symptoms. (However, including two necessarily related symptoms, such as needle length in centimeters and needle length in inches, only serves to double the weight given to factors influencing needle length.) Because collinear symptoms presumably provide a common signal emphasizing air pollution effects, their inclusion gives weight to the pollutant in constructing the index. If, on the other hand, an index is desired in which each symptom contributes only the information that it alone provides, one can remove the collinearity by using weights derived from the correlation structure of the data (40).

Combining symptoms into an index

Once standardized and weighted, component symptoms are combined into an index. (With PCA, combining and weighting are performed simultaneously.) Symptoms can be combined additively or multiplicatively. For an additive index, scores for symptoms positively correlated with pollution are added and scores for symptoms negatively correlated are subtracted (5,56). PCA is another additive approach. In a multiplicative index, the product of symptoms positively correlated with dose is divided by the product of symptoms negatively correlated (52). An intermediate method divides the sum of the components positively correlated with pollution by the sum of the components negatively correlated.

A hypothetical data set that demonstrates differences among additive, intermediate, and multiplicative methods of constructing indices is presented in Table 2. The data are idealized; two constants allow effects of manipulating the remaining variables to be seen clearly, and the remaining two variables show simple linear behavior across a hypothetical dose gradient, allowing behavior of the alternative indices to be clearly demonstrated. All three indices responded linearly to a linear change in one symptom across the dose gradient. When two symptoms changed linearly across the dose gradient, however, the responses of the indices differed (Table 2, Fig. 5). The additive index (and PCA) responded linearly to the linear change in two symptoms, the multiplicative index showed a markedly nonlinear response, and the intermediate index performed intermediately. Because an index responding linearly to linear changes in symptoms has desirable statistical properties and is more straightforward than one responding nonlinearly, we recommend using an additive index.

To be most interpretable, an index should be scaled to have a known maximum and minimum. A pollution damage index with a minimum of zero is desirable logically and can be derived

Table 2. Hypothetical data used to demonstrate effects of additive, multiplicative, and intermediate methods of index construction on behavior of resultant index across a pollutant dose gradient[w]

	Symptoms a–d[y]			
Site[x]	Chlorosis length (a)[z]	Necrosis length (b)	Needle retention (c)[z]	Needle length (d)
1	100	100	100	100
2	130	100	70	100
3	160	100	40	100
4	190	100	10	100

[w] Levels of chlorosis length and needle retention change linearly across dose gradient.
[x] Pollutant dose increases across sites 1–4.
[y] See Figure 5.
[z] All scores standardized to mean = 100, standard deviation = 30.

by solving the index equation for the zero disease case. This gives a constant that can be added to or subtracted from index scores to bring them to zero minimum. Ideally, an index should also peak at the maximum possible disease level, but this may not be possible, depending on the type of index.

A final problem concerns the level of resolution at which indices should be calculated. Assume that the data consist of symptoms scored on leaves nested within individuals within populations and that the main goal is to compare populations. We need an index score for each population. Because knowing the variance within populations is helpful, index scores should be calculated for each individual. A population-level score calculated as the average of index scores for individuals should differ little from a population-level score calculated from the average symptom scores for individual leaves. With PCA, however, the level at which the index values are calculated is more important.

To determine the effect of level of resolution and alternative symptom weightings on indices, we used data collected by Holcomb Research Institute during late May 1984 on foliar symptoms of *P. strobus*. Data were collected on needles from lower branches of trees in nine populations in Indiana differing from one another in presumed pollutant dose (primarily SO_2 and O_3; see description of populations in Usher and Williams [56]). The following symptoms of injury by these air pollutants were scored (after Costonis [8], Costonis and Sinclair [10], and Usher and Williams [56]): length of needle chlorosis (mm), length of needle necrosis (mm), percent retention of 2-year-old needles, and chlorotic mottle and flecking (1–5 = none to severe). Chlorotic mottle was scored in qualitative categories to allow comparison with an earlier study of the same populations when those categoreis were used (56). Percent needle retention was converted to proportions and was transformed by arc sine–square root before analyses. Total sample size was 234 trees, the number of trees per site ranging from 10 to 46.

Four types of indices that seemed reasonable were calculated: 1) additive using symptoms expressed as standard normal variates, as described previously; 2) additive using the same standardized symptoms but with each symptom weighted by the inverse of its within-population coefficient of variation; 3) additive with levels of symptoms assigned to numeric classes, as described previously (56); and 4) based on PCA. Each index was computed two ways for each site—at the individual level and at the population level (the population's average score for each symptom was used in the latter). Relationships between indices calculated at the individual and at the population level were examined using both parametric and nonparametric correlation procedures, as the distribution of index scores was not always normal. In all types except PCA, the correlations between the individual and population level approaches were quite high; the level at which the index was calculated did not significantly alter the outcomes (Table 3). For PCA, on the other hand, the correlation between levels, while still statistically significant ($P \leq 0.05$), was not as high, probably because PCA is a covariance-based technique while the others are mean-based. When PCA is conducted using data on individual trees, within-population covariation influences the analysis. The covariance structure among the symptoms of individuals differed from the covariance structure among population averages of symptoms. When the primary goal is to compare populations, it is normally preferable to use PCA on population average symptoms rather than on individual tree symptoms because the within-population variance is largely unrelated to differences in pollutant dose among populations. When the goal is to compare individuals within populations, however, it may be desirable to use PCA on individuals rather than on populations.

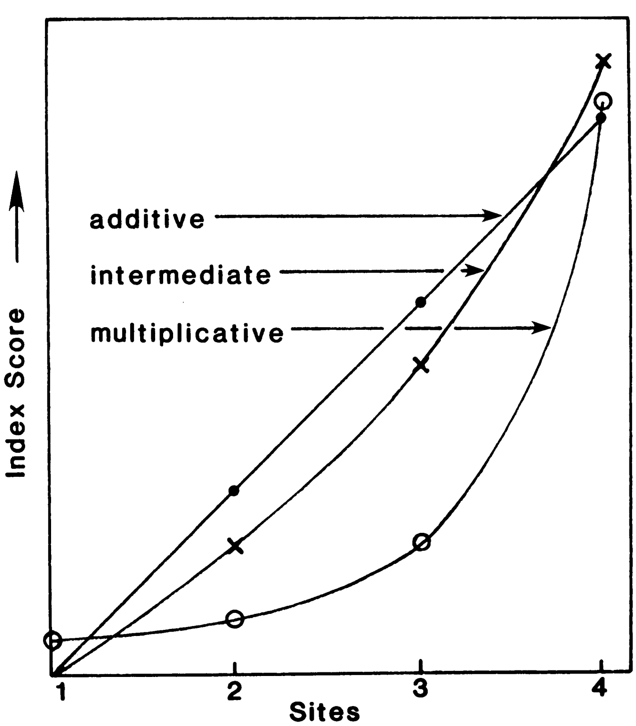


Fig. 5. Response of three hypothetical indices of foliar symptoms to increasing pollutant dose across four sites. Four symptoms (a–d) composing each index were scaled to equal means before being combined into indices. Additive index = a + b + c + d, intermediate index = (a + b)/(c + d), multiplicative index = (a · b)/(c · d).

Using the same data set, we calculated correlation coefficients among the four index types, all computed at the population level (Table 4). Again, all correlations were quite high, suggesting that none of the indices introduced significant biases or distortions into the stand patterns. We have no independent means of evaluating each index, so judging which (if any) was superior is difficult. Each was constructed by a reasonable method. PCA, however, may be preferable to the other three methods on the grounds of objectivity, since the scientist makes no decisions on weights for symptoms. PCA also handles well the collinearity likely to be found among foliar symptoms when several symptoms are produced by a given pollutant. Most packages of statistical software include routines for performing PCA, making it widely available.

Validating an index

Once an index has been constructed, its reliability should be checked. Ideally, the pattern suggested by the index should be compared with dose or such biological effects as growth reduction to determine if and how the index is related to these. If air pollution is the strongest influence on the symptoms composing the index, the relationship between injury and dose, as reflected in the index, should be strong. However, because good dose estimates are rarely available (particularly in field

Table 3. Pearson's product-moment correlations (r) and Spearman's rank correlations (r_s) between individual- and population-level indices

Index	r	r_s
1: Σ standardized[a] symptoms	0.992	0.950
2: Σ (C.V.$_i$[b] · standardized symptom$_i$)	0.990	0.967
3: Σ categorized[c] symptoms	0.979	0.975
4: PCA[d] (population level)	0.909	0.783

[a] Standardized by mean and standard deviation of symptom over all sampled populations.
[b] Average within-stand coefficient of variation of symptom i.
[c] Levels of each symptom assigned to numeric classes of equal size and range.
[d] Principal components analysis.

Table 4. Spearman's rank correlations between population positions on axes of symptoms and indices[a]

	Symptoms			Indices			
	2	3	4	1	2	3	4
Symptoms							
1: Necrosis length	0.11	0.15	−0.50	0.32	0.30	0.34	−0.31
2: Flecking	···	0.73*	−0.70*	0.83*	0.85*	0.85*	−0.83*
3: Chlorosis length	···	···	−0.33	0.82*	0.83*	0.75*	−0.70*
4: Arc $\sin^{-1}\sqrt{\text{needle retention}}$	···	···	···	−0.62*	−0.57	−0.72*	0.80*
Indices[b]							
1: Σ standardized symptoms	···	···	···	···	0.98*	0.98*	−0.95*
2: Σ ($C.V._{.i}$ · standardized $symptom_i$)	···	···	···	···	···	0.97*	−0.92*
3: Σ categorized symptoms	···	···	···	···	···	···	−0.98*
4: PCA	···	···	···	···	···	···	···

[a] N = 9; * = $P \leq 0.05$.

[b] Standardized symptoms = standardized by mean and standard deviation of symptom over all sampled populations; $C.V._{.i}$ = average within-stand coefficient of variation of symptom i; categorized symptoms = levels of each symptom assigned to numeric classes of equal size and range; PCA = principal components analysis.

studies), less direct approaches must be used. If the symptoms are reliable indicators of a single pollutant of concern, the relationships among the symptoms should be similar across years even when dose varies. Hence, the reliability of component symptoms can be checked by comparing the symptom correlation matrix for one year with that of another year. The Mantel test (12,34) can be used to compare two correlation matrices.

The reliability of indices can also be inferred by examining placement of populations relative to one another on the injury index gradient. If the sites do not change significantly relative to each other in year-to-year dose, and if the symptoms involved are faithful reflectors of air quality, positions of populations relative to one another on the injury index gradient should not change greatly.

Conclusions

An index of plant injury levels can be useful in particular applications because it allows the simultaneous use of all information on the plant's symptoms to assess the plant's condition. Further, an index mathematically integrates a mass of data that may be hard to digest at lower levels of summarization, and it can extract a common signal from symptoms that may be somewhat noisy individually. However, recording data on individual symptoms is more useful than combining symptom data at the data-gathering stage (as is sometimes done; Table 1) because individual symptoms may be helpful in certain analyses.

An ideal index: 1) uses equal weights for component symptoms or carefully chosen differential weights, 2) does not use subjective (qualitative) symptom categorization, 3) is additive, 4) has its minimum at zero and maximum at the highest possible disease level, and 5) has a monotonic, nearly linear relationship to dose across the widest possible pollution gradient. Published indices vary widely in method of construction and in the degree to which they meet these criteria (Table 1).

To promote the derivation of biologically sound indices and to make foliar scoring as useful as possible, basic data of two types are needed. First, data are needed on quantitative relationships between symptoms and pollutant dose, growth, or other parameters to which the indices are designed to relate (a truism, yet it seems to have been slighted). Estimates of the mean and variance of each symptom for each level of dose (or regression statistics for the relationship of symptoms to dose) are needed for choosing component symptoms as well as for deriving differential weights for symptoms. It is clear from published studies that much of the required data exist, yet the basic statistics are rarely included in publications. Second, data are needed on the functional dependencies between various symptoms—the extent to which expression of one symptom constrains the expression of another. These data are needed to design an index that is as linear and monotonic across a pollution gradient as possible.

In general, we recommend building indices either by using the sum of standardized symptoms weighted by their reliability or by using PCA, which in effect standardizes symptoms, weights them in proportion to their shared covariance with other symptoms, and sums them. PCA has an advantage over the other approach in that it uses the information from all symptoms to extract the strongest common signal. Another advantage is that PCA may reveal hitherto unexpected patterns in the data, e.g., it may extract a second axis of variation related to a second pollutant or to another driving factor.

In summary, indices of injury to foliage can be useful, but their power is limited by our understanding of the influence of pollutants on foliage and of the influence of foliar condition on growth. Indices should be constructed with attention to the mathematical properties, reliability, and interactions of the component symptoms.

ACKNOWLEDGMENTS

We thank T. V. Armentano, O. L. Loucks, W. L. MacDonald, E. S. Menges, J. N. Rogers, S. Will-Wolf, and three anonymous reviewers for comments on the manuscript. Field assistance was provided by J. L. Daniel, R. W. Usher, and T. V. Armentano. J. L. Daniel assisted with data entry. Figures were produced by C. Baker, and the manuscript was processed by G. Williams and E. Roach. The study was supported by the Holcomb Research Institute of Butler University, Indianapolis, IN.

LITERATURE CITED

1. Agrios, G. N. 1978. Plant Pathology. 2nd ed. Academic Press, New York. 703 pp.
2. Allen, T. F. H., and Starr, T. B. 1982. Hierarchy—Perspectives for Ecological Complexity. University of Chicago Press. 310 pp.
3. Armentano, T. V., Menges, E. S., and Molofsky, J. 1984. Analysis of air-pollution-sensitive vegetation at Indiana Dunes National Lakeshore. Air and Water Quality Division, National Park Service Pap. 33. Holcomb Research Institute, Indianapolis, IN. 110 pp.
4. Ashmore, M. R., Bell, J. N. B., and Reily, C. L. 1980. The distribution of phytotoxic ozone in the British Isles. Environ. Pollut. Ser. B 1:195-216.
5. Benoit, L. F., Skelly, J. M., Moore, L. D., and Dochinger, L. S. 1982. Radial growth reductions of *Pinus strobus* L. correlated with foliar ozone sensitivity as an indicator of ozone-induced losses in eastern forests. Can. J. For. Res. 12:673-678.
6. Blanchard, R. O., Baas, J., and van Cotter, H. 1979. Oxidant damage to eastern white pine in New Hampshire. Plant Dis. Rep. 63:177-182.
7. Chester, K. S. 1959. How sick is the plant? Pages 100-142 in: Plant Pathology. Vol. 1. The Diseased Plant. J. G. Horsfall and A. E. Dimond, eds. Academic Press, New York. 674 pp.
8. Costonis, A. C. 1970. Acute foliar injury of eastern white pine induced by sulfur dioxide and ozone. Phytopathology 60:994-999.
9. Costonis, A. C. 1973. Injury to eastern white pine by sulfur dioxide and ozone alone and in mixtures. Eur. J. For. Pathol. 3:50-55.
10. Costonis, A. C., and Sinclair, W. A. 1969. Relationships of atmospheric ozone to needle blight of eastern white pine. Phytopathology 59:1566-1574.
11. Davis, D. D., and Wood, F. A. 1973. The influence of environmental factors on the sensitivity of Virginia pine to ozone. Phytopathology 63:371-376.
12. Douglas, M. E., and Endler, J. A. 1982. Quantitative matrix comparisons in

ecological and evolutionary investigations. J. Theor. Biol. 99:777-795.
13. Genys, J. B., and Heggestad, H. E. 1983. Relative sensitivity of various types of eastern white pine, *Pinus strobus*, to sulfur dioxide. Can. J. For. Res. 13:1262-1265.
14. Greig-Smith, P. 1983. Quantitative Plant Ecology. 3rd ed. Blackwell, Oxford, England. 359 pp.
15. Gumpertz, M. L., Tingey, D. T., and Hogsett, W. E. 1982. Precision and accuracy of visual foliar injury assessments. J. Environ. Qual. 11:549-553.
16. Hayes, E. M., and Skelly, J. M. 1977. Transport of ozone from the northeast U.S. into Virginia and its effect on eastern white pines. Plant Dis. Rep. 61:778-782.
17. Hebert, T. T. 1982. The rationale for the Horsfall-Barratt plant disease assessment scale. Phytopathology 72:1269.
18. Heck, W. W., and Brandt, C. S. 1977. Effects on vegetation: Native, crops, forests. Pages 157-229 in: Air Pollution. Vol. 2. The Effects of Air Pollution. 3rd ed. A. C. Stern, ed. Academic Press, New York. 684 pp.
19. Heck, W. W., Dunning, J. A., and Hindawi, I. J. 1966. Ozone: Nonlinear relation of dose and injury in plants. Science 151:577-578.
20. Heck, W. W., Krupa, S. V., and Linzon, S. N., eds. 1979. Handbook of Methodology for the Assessment of Air Pollution Effects on Vegetation. Air Pollution Control Association, Pittsburgh. 392 pp.
21. Hofstra, G., Tonneijck, A. E. G., and Allen, O. B. 1985. Cumulative effects of low levels of SO_2 on O_3 sensitivity in bean and cucumber. Atmos. Environ. 19:195-198.
22. Horsfall, J. G., and Barratt, R. W. 1945. An improved grading system for measuring plant disease. Phytopathology 35:655.
23. Horsfall, J. G., and Cowling, E. B. 1978. Pathometry: The measurement of plant disease. Pages 119-136 in: Plant Disease. An Advanced Treatise. Vol. 2. How Disease Develops in Populations. J. G. Horsfall and E. B. Cowling, eds. Academic Press, New York. 436 pp.
24. Houston, D. B. 1974. Response of selected *Pinus strobus* L. clones to fumigations with sulfur dioxide and ozone. Can. J. For. Res. 4:65-68.
25. Jacobson, J. S., and Hill, A. C. 1970. Recognition of Air Pollution Injury to Vegetation: A Pictorial Atlas. Air Pollution Control Association, Pittsburgh. 102 pp.
26. James, W. C. 1971. An illustrated series of assessment keys for plant diseases, their preparation and usage. Can. Plant Dis. Surv. 51:39-65.
27. James, W. C. 1974. Assessment of plant diseases and losses. Annu. Rev. Phytopathol. 12:27-48.
28. Jensen, F. K., and Noble, R. D. 1984. Impact of O_3 and SO_2 on net photosynthesis of hybrid poplar cuttings. Can. J. For. Res. 14:385-388.
29. Kress, L. W., and Skelly, J. M. 1982. Relative sensitivity of 18 full-sib families of *Pinus taeda* to O_3. Can. J. For. Res. 12:203-209.
30. Lacasse, N. L., and Treshow, M., eds. 1978. Diagnosing vegetation injury caused by air pollution. U.S. Environ. Prot. Agency EPA-450/3-78-005. 249 pp.
31. Linzon, S. N. 1960. The development of foliar symptoms and the possible cause and origin of white pine needle blight. Can. J. Bot. 38:153-161.
32. Malhotra, S. S., and Blauel, R. A. 1980. Diagnosis of air pollutant and natural stress symptoms on forest vegetation in western Canada. Can. For. Serv. Inf. Rep. NOR-X-228. 84 pp.
33. Manning, W. J., and Feder, W. A. 1980. Biomonitoring Air Pollutants with Plants. Applied Science Publishers, Ltd., London. 142 pp.
34. Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. Cancer Res. 27:209-220.
35. Marsh, R. W., Martin, H., and Munson, R. G. 1937. Studies upon the copper fungicides. III. The distribution of fungicidal properties among certain copper compounds. Ann. Appl. Biol. 24:853-866.
36. McKinney, H. H. 1923. Influence of soil temperature and moisture on infection of wheat seedlings by *Helminthosporium sativum*. J. Agric. Res. 26:195-218.
37. McLaughlin, S. B. 1984. Effects of air pollution on forests. A critical review. J. Air Pollut. Control Assoc. 35:512-534.
38. Miller, P. R. 1973. Oxidant-induced community change in a mixed conifer forest. Pages 101-117 in: Air Pollution Damage to Vegetation. J. A. Neagele, ed. American Chemical Society, Washington, DC. 137 pp.
39. Mudd, J. B., and Kozlowski, T. T., eds. 1975. Responses of Plants to Air Pollutants. Academic Press, New York. 383 pp.
40. Muir, P. S., and Lotan, J. E. 1985. Disturbance history and serotiny of *Pinus contorta* in western Montana. Ecology 66:1658-1668.
41. Phillips, S. O., Skelly, J. M., and Burkhart, H. E. 1977. Eastern white pine exhibits growth retardation by fluctuating air pollutant levels: Interaction of rainfall, age, and symptom expression. Phytopathology 67:721-725.
42. Redman, C. E., and Brown, I. F., Jr. 1964. A statistical evaluation of the Barratt and Horsfall rating system. Phytopathology 54:904.
43. Reich, P. B., and Amundson, R. G. 1985. Ambient levels of ozone reduce net photosynthesis in tree and crop species. Science 230:566-570.
44. Reinert, R. A. 1984. Plant response to air pollutant mixtures. Annu. Rev. Phytopathol. 22:421-442.
45. Rice, P. M., Boldi, R. A., Carlson, C. E., Tourangeau, P. C., and Gordon, C. C. 1983. Sensitivity of *Pinus ponderosa* foliage to airborne phytoxins: Use in biomonitoring. Can. J. For. Res. 13:1083-1091.
46. Saari, E. E., and Prescott, J. M. 1975. A scale for appraising the foliar intensity of wheat diseases. Plant Dis. Rep. 59:377-380.
47. Skelly, J. M., Krupa, S. V., and Chevone, B. I. 1979. Field surveys. Pages 12.1-12.30 in: Handbook of Methodology for the Assessment of Air Pollution Effects on Vegetation. W. W. Heck, S. V. Krupa, and S. N. Linzon, eds. Air Pollution Control Association, Pittsburgh. 392 pp.
48. Smith, W. H. 1981. Air Pollution and Forests. Springer-Verlag, New York. 379 pp.
49. Sneath, P. H. A., and Sokal, R. R. 1973. Numerical Taxonomy: The Principles and Practice of Numerical Classification. W. H. Freeman and Co., San Francisco. 573 pp.
50. Snedecor, G. W., and Cochran, W. G. 1980. Statistical Methods. 7th ed. Iowa State University Press, Ames. 507 pp.
51. Stoklasa, J. 1923. Die Beschädigung der Vegetation durch Rauchgase und Fabriksexhalationen. Urban und Schwartzenberg Verlag, Berlin and Vienna. 487 pp.
52. Stolte, K. W., and Bennett, J. P. 1984. Standardized procedures for establishing and evaluating pollution injury on pines in permanent plots. National Park Service, Air and Water Quality Division, Denver. 9 pp.
53. Tingey, D. T., Wilhour, R. G., and Taylor, O. C. 1979. The measurement of plant responses. Pages 7.1-7.35 in: Handbook of Methodology for the Assessment of Air Pollution Effects on Vegetation. W. W. Heck, S. V. Krupa, and S. N. Linzon, eds. Air Pollution Control Association, Pittsburgh. 392 pp.
54. Todd, G. W., and Arnold, W. N. 1961. An evaluation of methods used to determine injury to plant leaves by air pollutants. Bot. Gaz. 123:151-154.
55. Toole, E. R. 1949. White pine blight in the Southeast. J. For. 47:378-382.
56. Usher, R. W., and Williams, W. T. 1982. Air pollution toxicity to eastern white pine in Indiana and Wisconsin. Plant Dis. 66:199-204.