# Towards Monitoring of Long-term Trends in Terrestrial Ecosystems

by

W. TED HINDS, M.S., Ph.D.(Washington)

Senior Research Scientist, Applied Ecology Section, Environmental Sciences Department, Battelle Pacific Northwest Laboratories, P.O Box 999, Richland, Washington 99352, USA.

# INTRODUCTION

An accurate judgement of environmental consequences resulting from technological development requires that appropriate designs for the collection of environmental data be available and consistently followed. This is usually called 'environmental monitoring' (Holdgate, 1979), or 'surveillance', if corrective action is not contemplated (see Cairns, 1980). Environmental monitoring consumed about seven or eight hundred million dollars in the USA alone in 1977 (Buffington, 1978). Apparently little, if any, of this effort went into determining whether earlier environmental assessments and predictions had been accurate or not. But, even if some research had been aimed at estimating the effectiveness of environmental impact assessments, the feedback would have been only partially fruitful-if only because environmental monitoring usually involves only air and water quality, ignoring biota.

One must examine biotic assemblages of one sort or another before one can tell whether ecological problems exist. This examination I will call 'ecological effectsdetection', or ecological monitoring (Gray, 1980); it is the purposeful and repeated examination of the state or condition of specifically-defined biotic groups in relation to external stress. It differs from most environmental monitoring (as described e.g. in National Research Council, 1977; Anon., 1978; Luepke, 1978) in that ecological monitoring emphasizes changes in living organisms, not merely in the physical environment. Furthermore, it differs from most biological monitoring ('biomonitoring') because ecological monitoring has its focus-its very raison d'être-in the biotic community itself; in contrast, biomonitoring activities typically use the biota as a surrogate 'filter' to be analysed to indicate environmental quality (cf. Goldberg et al., 1978; Manning & Feder, 1980; National Research Council, 1980; Worf, 1980). Finally, this approach to ecological effects-detection differs from ecosystem effects-detection, because its focus is on ecologically interacting taxa within the system rather than on system-level processes per se (that is, aggregated processes such as total nutrient export; see e.g. O'Neill et al., 1977).

Ecological monitoring designs have at least the following three identifiable difficulties that must be overcome if the design is to be successful: (1) The major ecological difficulty is selecting and quantifying specific biotic conditions or activities within the continuous spatial and temporal flux that characterizes life; (2) The major statistical difficulty is specifying appropriate replicationstandards in a world that is full of unique places; and (3) The major difficulty with monitoring in general is that it is expensive. This tripartite requirement for ecologically relevant, statistically credible, and cost-effetive, monitoring methods is very stringent, and failure to meet one or more of these characteristics is at the root of many problems that are found in ecological monitoring. A particularly clear example surfaced in the Organization for Economic Cooperation and Development (OECD) monitoring programmes, where a reviewer (Holden, 1978) remarked that none of the national programmes seemed to have had statistical analysis in mind.

A different problem can be found in the outcome of a recent workshop on Long-term Ecological Measurements, sponsored by the US National Science Foundation (National Science Foundation, 1977). The ecological, chemical, and environmental, methods espoused there were so extensive that cost alone prevented replication. Yet variance among experimental-units-treated-alike, is the appropriate variance for effects-detection (Eberhardt, 1976; Holden, 1978). Inappropriate replication appears to be a serious flaw in many ecological monitoring efforts (Eberhardt, 1976; Skalski & McKenzie, 1982).

Some of the difficulties involved in detecting long-term ecological change stem from the amount and kind of information that may be required to produce a valid design. For example, an adequate long-term method usually cannot evolve from mere repetition of an arbitrarily-chosen field method, although this has been suggested by experienced and indeed eminent scientists (*see*, for example, Holdgate, 1979 p. 195). Problems with interpretation, replication, variability, or cost, might be expected to frustrate such straightforward approaches.

To help to prevent expensive failures, I suggest a stepwise and systematic development of long-term methodologies, as outlined in Fig. 1. Beginning with ecological judgements about processes that are sensitive to the objective, a selection of ecological targets can be made. Both temporal and replication variances will be needed to evaluate changes, so considerable statistical input will be required to minimize expensive field-work. After the method has been shown to be capable of providing adequate sensitivity and power (in the statistical sense), the cost of the design can be minimized, consonant with the objectives. Failure at any point in this sequence suggests that the monitoring method itself will fail.

11

Environmental Conservation, Vol. 11, No. 1, Spring 1984-C 1984 The Foundation for Environmental Conservation-Printed in Switzerland.



FIG. 1. A flow-chart of recommended considerations for development of ecological effects-detection methodologies.

In this paper, I outline some important ecological and statistical considerations. These suggestions are the result of several years of research aimed at developing costeffective methods that are in principle capable of detecting slowly-paced ecological change.

#### **ECOLOGICAL JUDGEMENTS**

Ecosystems are highly complex (Knox & Polunin, in press), posing a variety of choices for ecological monitoring measurements. One straightforward way to study ecosystems is to focus on parts that are linked by energy or mass transfers, usually through food-webs or trophic position-including transfers to and from decomposer organisms. However, no single selection is correct for all purposes, and consequently a variety of approaches may be considered. Each level of organizational complexity, from individuals to total systems, has strengths and weaknesses for detection of long-term ecological effects, as suggested in Fig. 2. Furthermore, ecological interactions influence the choice: for example, effects from bioaccumulated or biomagnified toxicants, such as DDT, will be more obvious at high trophic levels, such as carnivors (Newton, 1979), than at low ones, while effects from relatively immobile materials, such as heavy-metals, should be clearest in scavengers or decomposers (Goldberg et al., 1978; Inman & Parker, 1978).

| LEVEL OF ORGANIZATION                          |       | ECOLOGICAL<br>EFFECTS     | CURRENT<br>UNDERSTANDING | DESIGN<br>DEVELOPMENT     | REPLICATION             | PROBABLE          |
|--|-------|---------------------------|--------------------------|---------------------------|-------------------------|-------------------|
|  |       | INAPPROPRIATE<br>MODERATE | GREAT<br>GREAT           | EASY                      | EASY                    | SMALL<br>MODERATE |
|  |       |                           |                          |                           |                         |                   |
| STRUCTURAL<br>GROUPS<br>(TROPHIC<br>TRANSFERS) | GREAT | MODERATE                  | MODERATE                 | HARD                      | LARGE?                  |                   |
| SYSTEM-LEVEL PERFORMANCE<br>OR MEASUREMENT     |       | VALUABLE                  | SMALL                    | UNEXPLORED;<br>DIFFICULT? | NOT CLEARLY<br>POSSIBLE | UNCLEAR           |

FIG. 2. Some characteristics of various levels of biological organization in relation to the detection of ecological effects. The 'MULTI-SPECIES GROUPS' portions (double-ruled around), and especially the 'functional groups' and 'structural groups' boxes, indicate the principal focus of the discussion in this paper.

Additionally, the repeated examination of the same place or process puts significant constraints on the kinds of measurements that might be useful in the long run. An ideal monitoring method should be non-destructive and passive, while yielding replicable samples for chemical analysis if necessary (cf. Holdgate, 1979). These in turn suggest that, currently, a sub-system focus might be more practical than a total-system approach—particularly with regard to replicability (a facet that will be discussed in more detail in the following section).

In monitoring contexts, current thought about the systems-level approach is ambiguous, being mildly both optimistic (cf. O'Neill *et al.*, 1977; National Research Council, 1981) and pessimistic (cf. Bourdeau & Treshow, 1978; Cowell, 1978). Perhaps the usefulness of a systems-level approach is system-specific (and perhaps most applicable to marine ecosystems; *see* Gray, 1980). The systems-level approach is conceptually correct but, depending upon the measurement strategy, it may be very expensive. Furthermore, the possibility of replication at the system level in field conditions is almost totally unexplored.

Similarly, ecosystem-level simulation models pose fundamental problems for projecting or interpreting long-term effects. A major obstacle is that the model builders should have both species-specific and systemwide stress/response data in hand during the development of the model (Dyne, 1981; Kendall, 1982), so that the model would reflect known behaviour under stress rather than mere extrapolations (National Research Council, 1981). Unfortunately, ecosystem-wide toxicology data are rarely available, so extrapolations have of necessity been built into models. This can lead to disconcerting, and sometimes serious, errors. This problem suggests that much work remains to be done to integrate extensive field-work and sophisticated modelling (Keulen *et al.*, 1980–81).

Depending upon the objective of monitoring, three sub-system levels of organization may be useful: (1) single species; (2) guilds (functional groups [e.g. Jacsic, 1981], namely groups of taxonomically unrelated species that make their living in a similar fashion); and (3) trophically linked species (predator-prey-decomposer linkages). Many specific approaches are possible, such as relative abundance of rare species (Cairns, 1974; Gray, 1980), guild theory (Severinghaus, 1981), net nutrient output from a system (O'Neill et al., 1977), tree-rings (Symeonides, 1979), or microbial communities (Colwell, 1978). Currently, ecological science cannot predict the long-term usefulness of these and many other approaches (Barrett & Rosenberg, 1981) without more experience. Variety in approach and design is to be expected, and encouraged.

## STATISTICAL JUDGEMENTS

When once the specific processes or activities to be monitored have been selected and proven appropriate, attention must shift from qualitative assessments of ecological pertinence to quantitative assessments and inferences. Two considerations are paramount: error rates and replication. Unless these are specifically addressed, the technique will probably yield only descriptive data that are inadequate for quantitative assessments of relatively small changes.

The need to detect changes before they are obvious brings to attention the classic statistical tradeoff between Type I and Type II errors that is discussed in detail in standard texts and references (e.g. Snedecor & Cochran, 1967). In a monitoring context, a Type I error means that an impact is proclaimed when no impact exists, and a Type II error means that no impact is declared when an impact does in fact exist. The ability to avoid Type I errors is described in terms of significance or sensitivity of the design—its ability to distinguish real differences from random error. The ability to avoid Type II errors is described in terms of the power of the design—its ability to declare that a given amount of change is real not random error.

Traditionally, the two types of error are not treated equally; for example, a 5% chance of Type I error and a 20% chance of Type II error may be accepted in relation to a given amount of change (Snedecor & Cochran, 1967; Skalski & McKenzie, 1982). Sometimes Type II error and power are ignored altogether, while sensitivity of the design is maximized. But ignoring power leads to problems in field ecology, where resources, and thus sampling effort, can be limiting; in such cases

'All too often, either by default or lack of understanding, the real rate [of Type II error] is about 50%, which can be likened to settling the issue by flipping a coin and doing no field study whatsoever' (Eberhardt, 1978).

The implications of using 'powerless' designs are particularly severe in a monitoring context, where many years of effort and expense may be involved. If a biologically significant change cannot be determined to be real, the monitoring effort is a failure.

It is irrational to set power and Type I error simultaneously and a priori, while lacking quantitative estimates of penalties. I suggest an unconventional approach: equalize the rates for Types I and II error (Vanderhorst & Wilkinson, 1979). This implies that delta  $(\Delta = \text{minimum detectable change, which is one of the})$ parameters needed to calculate power) would be fixed a posteriori by the affordable (or maximum available) sample-size and the observed variance (see the discussion in Skalski & McKenzie, 1982). If the affordable samplesize and observed variance lead to a  $\Delta$  that is too large to be acceptable, the method is inadequate. This approach, where sample-size is a limiting factor, differs somewhat from the traditional approach, wherein  $\Delta$  is set a priori and sample-size is left to be determined (see, for example, the extended discussion in Snedecor & Cochran, 1967). Whichever approach may be thought appropriate, setting the error rates as equal presumes that, in a monitoring context, the overall penalty for error is independent of the error-type. This may not be accurate, but I know of no analysis that suggests one way rather than the other.

In addition, good reason exists to set both of the errorrates substantially higher than the traditional 1% or 5% for Type I error—a tradition which developed in relatively benign experimental circumstances. Compared with ecological field-studies, those circumstances included controlled conditions, relatively easy sampling, smaller coefficients of variation, greater certainty about underlying distributions, and quantifiable penalties for error. I suggest that equal error-rates of 10% (perhaps even 15%!) will allow practical and credible monitoring or assessment studies in undomesticated situations.

The number of replications, or, better yet, replicated pairs (Skalski & McKenzie, 1982) in a design, influences error-rates decisively; but two opposing viewpoints cause some confusion about replication in field conditions. A fine-grained view of the landscape will suggest that every place is different than every other place. A coarse-grained view will suggest that many areas are practically identical. These contrasting stances often lead to disagreements about appropriate observational (or experimental) units, and therefore about the validity of making inferences regarding other places in the landscape. I will now discuss possible ways of avoiding this controversy.

## Some Ecological Aspects of Replication

Significant differences of opinion surround the selection of replicated units. In an extended analysis of the difficulties of assessing environmental impact, Eberhardt (1976) pointed out that field ecologists generally expect that no two places, populations, or processes, will be exactly alike. But, he noted, a great resemblance among sites is nevertheless expected, due to correlations induced by macroclimates, restricted genetic bases, and other similarities. Although Eberhardt (1976) did not draw attention to the point, it may be that this aspect of ecological analysis is a 'judgement call'-that is, a 'landscape' can justifiably be analysed at several places along a continuum ranging from coarse-grained aggregations of a few relatively large units to fine-grained aggregations of many relatively small units. Such a continuum complicates the statistical idea of 'independence' as a function of space.

It is clear that our judgements need not be arbitrary; reasonable approaches to estimating or describing similarity among many units do indeed exist. Gauch (1982) devotes an entire chapter to classification, noting several criteria about clustering algorithms that would lead to excellent determinations of replicate units. An appropriate cluster analysis thus can provide a good way to build a suite of observational units that can be considered 'equally suitable sites' for research. Selection of an appropriate number of sites at random from this suite of sites then produces a set of independent replications that can be defined to satisfy most observers. But, of course, a drawback exists: the initial data-collection to support the cluster analysis will be open to a variety of criticisms, depending upon the judgements made in collecting those initial data.

I am not aware of specific monitoring applications of clustering techniques. However, a recent example of a similar approach has been published (Laven, 1982). Laven's objective was to explore the utility of multivariate analysis for defining homogeneity among forested sites. He found that a simple clustering technique was useful, but that significant advantages would be gained by adding principal-component analysis. However, in reference to the initial selection of his suite of plots, Laven (1982) recorded only that 'Forty-five 10 m  $\times$  30 m studyplots were located in Jeffrey pine regions...'. Later, in reference to his study objectives, he noted that the plots were situated where at least two trees had wood sections that reflected a similar fire-history, and that plots were located where soil-type differences were not evident.

I suggest that Laven's (1982) procedure can be strengthened into a reasonably reproducible method by attention to a couple of details. First, as he suggests, the sites should meet a set of minimal qualifications as 'equally suitable sites' for the research. The specifications can be brief or detailed, as necessary, but they must be articulated before they can be defended as anything other than arbitrary. If the specifications can be outlined as a coherent set of criteria in the Methods sections of any resulting papers, matters are much-improved. Second, specifically in a monitoring context, the ideal number of candidate sites should be much greater than the eventual number of replicate sites to be selected. This very arduous effort defines the population about which statistical inferences can validly be made. Without this step, statistical inferences about the wider implications of monitoring results found at specific sites will be criticized, either by scientists who perceive a different 'grain size' in the continuum of landscape units, or by those whose interests may conflict with the results.

Ideally, of course, ecologists and biometricians should agree on what is replication in field conditions. Currently, such agreement is not near; what some call 'replication', others call 'subsampling' (Eberhardt, 1976; Skalski & McKenzie, 1982). Practical and conceptual matters converge at this point, because replication determines the cost, as well as the statistical power, of any effort to detect change.

## **COST-EFFECTIVE DESIGNS**

Cost is an important part of any monitoring effort, but cost data are rarely used in the design-development effort. The usual approach to cost treats it as a fixed constraint rather than a variable that can be manipulated or controlled. But it is clear that the cost of doing routine tasks in monitoring will be substantially less than the cost of doing the prior research to develop the proper routine. This difference implies that one of the tasks inherent in monitoring research must be that of separating cost data into appropriate categories, some reflecting the actual costs associated with routine data-collection and others reflecting the costs associated with research—including any erroneous decisions. It is the cost of the routine work that defines a cost-effective monitoring design.

Unfortunately, as a practical matter, cost data are generally not useful for rigorous scientific analysis; they are sometimes only marginally useful even as accounting tools (Thomas *et al.*, 1981). For example, cost-accounting systems for a research grant or contract usually need not trace the hours of labour by specific individuals engaged in each of several small research sub-tasks. But accurate cost-data to be associated with routine data-collection in developing a monitoring method, require accurate timedata. If *research* time were to be confounded with the time for doing the *routine* work, it would be impossible to make an accurate estimation of the probable cost of doing the *monitoring*: all that could be given would be an accurate cost-accounting of having done the research.

But personnel time is not the only cost-source that must be allocated properly to either routine work or research effort. Investigator time often amounts to only one-half or two-thirds of the total expense of monitoring work; several papers considered by Luepke (1978) gave this estimate, although Holden (1978) reports a much lower figure (6 to 17%). Other appropriate costs include purchases, subcontracts, travel, and chemical analyses. Tracking these specific cost data in a way that can separate research expense from routine expense, suggests that detailed cost analyses may be required. Furthermore, the task of compiling, printing, and distributing, such detailed records in a timely fashion for the investigator, implies a system with many helpers-a significant indirect-cost burden. This is a price that many research workers and institutions hesitate to recommend as part of research cost (Brown, 1981). But these cost data seem necessary if the costs inherent in the monitoring designsuch as field work, chemical analysis, interpretation, and statistics-are to be separated from research costs (such as proposing, planning, reporting, and publication).

However, even if appropriate cost-data were available, the puzzles in allocating funds among tasks within the monitoring design would remain. One problematical choice is the allocation to field collection *versus* the allocation for chemical analysis. In some cases, the major part supports chemical analysis: OECD experience (Holden, 1978), for example, was that 83% to 94% went to analysis. Misallocation here may lead to significant error:

'In many instances, sampling may account for the greatest source of error. Indeed, sampling procedures have seldom received the same critical study as have methods of analysis' (Loon, 1975).

Critical analysis of sampling techniques and designs are now becoming available (Eberhardt, 1976; National Science Foundation, 1977; Thomas *et al.*, 1981), but the truth remains that field-sampling variability is both large and poorly understood in relation to the rest of any longterm-effects detection effort. Here, as in statistical design, practical and conceptual concerns intersect, while developing field methods that are capable of detecting longterm changes is difficult.

## SPECIFIC EXAMPLES

When we, at the Pacific Northwest Laboratories, began our research into ways to detect slowly-paced or subtle changes in ecosystems of the western parts of the USA, we focused on two contrasting ecological situations: closedcanopy conifer forests, and toxic materials in food-webs leading into avian predators. Given the desirability of observing the same place and process over and over again (as discussed earlier), our strategy was to collect organic matter produced by or associated with 'target' organisms (see Holdgate, 1979), as it was transferred to decomposers. This approach met all the constraints noted earlier for repeated and passive sampling. Furthermore, the field-work was simple enough for replication to be feasible. In the next several paragraphs, I will describe briefly our applications of the judgements discussed in the preceding sections—first in an ecological context of selection of the process to be monitored, and then in a statistical context of design development. Finally I will note our attempts to estimate the cost associated with using each design, and our approach to optimizing the effort.

Trees subjected to pollutant stress lose their older needles etc. sooner than do unstressed trees (Axelrod *et al.*, 1980; Smith, 1981; McLaughlin *et al.*, 1982; Williams, 1983). Therefore, our approach to forest sampling design was to collect litterfall. We did not include large litter, such as boles and branches; rather we concentrated on small litter such as needles, bits of bark, cones, bud-scales, and epiphytic matter. These were brought to the laboratory where they were dried, separated into a few ecologically meaningful categories (needles, reproductive parts, epiphytes, and miscellaneous debris), weighed, and all pertinent details recorded.

Litterfall collection has a long history in forest ecology. A typical sampling technique (reviewed by McShane et al., in press) might employ from four to six collectors (0.2 to 1 m<sup>2</sup> in area) arranged systematically across a study plot; unfortunately, large coefficients of variation (CV) were common with this approach (W.A. McKee, Oregon State University, pers. comm.; G.C. Carrol, University of Oregon, pers. comm.). For our preliminary research, I elected to use 24 small samplers (435 cm<sup>2</sup> in cross-section) placed at random in a one-hectare plot in each of four Research Natural Areas (Franklin et al., 1972) in US Pacific Northwest forest types (Waring & Franklin, 1979). I also elected to emphasize estimation of sampling variance (within-plot variance) before beginning the more expensive estimation of replication variance. Although other approaches were possible, this sequential approach seemed advisable, because I expected the among-plots variance to be much larger than within-plot variance. If, as was suggested by the large CVs noted above, the within-plot variance was by itself too large to allow costeffective sampling designs, then the among-plots variance was certain to thwart monitoring efforts.

Sampling statistics have indicated that, at three of the four sites, mean monthly needlefall could be estimated with 90% confidence intervals of  $\pm 10\%$ , using about twenty optimally-sized collectors (200 to 600 cm<sup>2</sup> cross-section) (McShane *et al.*, in press).

These sampling statistics indicate that collector size is crucial for cost-efficient data, but not for accurate data. The smaller collectors provided as good an estimate of total needlefall as the largest collector, but at about one-tenth the cost (Fig. 3) (McShane *et al.*, in press). Using about 20 collectors of 435 cm<sup>2</sup> area, about two hours per month was required to estimate total (not species-specific) needlefall per month per site to within  $\pm 10\%$  with 90% confidence. That is, optimal sampling reduced both error and effort far enough that one may contemplate using many replicated plots in a widespread design.

For food-chain contaminant monitoring, we followed a biomonitoring developmental scheme focused on the 'quality' of the environment (Goldberg *et al.*, 1978; Worf, 1980) by using suitable biota to sample for the presence of toxic materials in the environment. Specifically, food scraps, faeces and prey parts, have been shown to reflect both the presence and abundance of toxic materials in food-webs (Rickard *et al.*, 1978; Newton, 1979), so we collected such debris that had been ejected from nests of wide-ranging carnivorous birds that predictably (unfortunately not invariably) return to previous nest-sites.

To develop data on any ecological effects of such contaminants on these populations, we noted fledging success. In arid habitats, we collected regurgitated pellets and noted fledging success at Swainson's and Ferruginous Hawks' nests (*Buteo swainsoni* and *B. regalis*, respectively). The pellets were dissected as to hair, feathers, and bones, for later chemical analysis. In the moister habitats, we collected faeces and debris, as well as noting fledging success, in Great Blue Heron (*Ardea herodias*) rookeries. Collectors with a screen-bottom above a watertight bottom were placed under nest trees or control trees. These collectors sampled solid debris separately from the leachate caused by rainfall during the sampling (nesting) period.

Sampling to date has shown that differences in toxic elements in collectors under different Heron colonies were easily detectable; furthermore, sources of toxic elements could be differentiated as to airborne contamination or food-chain contamination (Fitzner *et al.*, 1982). Replication variances are currently being developed as a function of sampler design and number, and of geographic location of the colonies.

We found that appropriate cost considerations focused on one distinctive category: technician-time required to handle each sample from the time it came in from the field until it was 'archived', and its corresponding numerical data were ready for computer entry. For example, we found from three years of weekly time-records in our forest monitoring work that the annual average time required to make species-specific needle separations was  $1.82\pm0.02$  (standard error over

#### COST OF ANNUAL NEEDLEFALL ESTIMATES VS SAMPLER SIZE



FIG. 3. Graph showing the relation between size of needlefall collectors and cost of obtaining estimates of needlefall (to within  $\pm 10\%$  with 90% confidence).

n = 3 years) hours per sample. Other, less detailed, separations required much less effort. For example, merely sorting the needles as a whole from other litter, then weighing the samples, took about 0.33 hour. A third approach, which is to treat litterfall samples as a whole, by drying and weighing them but without any sorting, required about 0.12 hour per sample.

We then sought to optimize the design by reducing costs in relation to variance and sample-handling time. Specifically, the required number of samplers (n) for a desired level of sampling precision is a function of the coefficient of variation ( $CV = s/\bar{x}$ ). Assume that n is a linear function of variance,  $n = k(CV)^2$ . As noted above, handling time to determine needlefall was three times as great as that required for litterfall. Then, the cost of sampling needlefall (NF) will be less than that for sampling the total litterfall (LF) to the same level of precision whenever:

$$\frac{\text{CV}_{\text{NF}}}{\text{CV}_{\text{LF}}} < (\frac{1}{3})^{\frac{1}{3}} = 0.58$$

This 'critical ratio' lies well within the observed range of ratios, which was 0.45 to 1.15. Thus, the most costeffective approach appears not to be uniquely determined: in some areas, needlefall will be most efficient, in others, total litterfall. However, the penalty for using slightly sub-optimal sampling can be shown to be small (McShane *et al.*, in press); our estimate is that only one or two hours per month per site will usually make the difference between the more informative needlefall separations and the easier total litterfall estimations. Consequently, our current recommendation is to focus on needlefall, which is the direct indicator of response to stress.

### CONCLUSIONS

To date, our research suggests that optimized ecological sampling methods may be achieved more or less predictably in tightly-delineated observational units. However, it should be clear that ecological monitoring of 'landscape-sized' areas is in its infancy. Ecological monitoring has a long way to go to be considered an 'early-warning system' (Cairns & Schalie, 1980).

A few-years-ago review (Bourdeau & Treshow, 1978) discussed ecosystem responses to various kinds of pollutant stress, but noted with misgivings that most data refer to single species. This is one important reason why our research currently emphasizes what we consider to be well-chosen but restricted groups of species, rather than systems. The proper selection of species, or perhaps of guilds, currently seems essential to simplify the field-work required to obtain precise data.

Probably the best-developed aspect of ecological monitoring is found in statistical-design development; many years of work in short-term effects-assessment and environmental sampling are now beginning to bear fruit (Eberhardt, 1976; Suter, 1981; Thomas *et al.*, 1981; Skalski & McKenzie, 1982). However, long-term changes or trends are not reliably detectable if we restrict ourselves to the use of inconsistent or short-term data. Success in long-term assessment will probably depend upon creating a variety of efficient methods that, like climatological methods, are both robust and widely useful.

# ACKNOWLEDGEMENTS

My thanks are due to many colleagues for their critical commentary on this manuscript: Dr James B. States and Dr Peter Van Voris, of Pacific Northwest Laboratories, were particularly helpful in the earlier stages of the several-years' gestation which this paper required. Others who have read and commented on the paper in this form or in earlier drafts were Dr Anne S. Fege and Dr Jerry F. Franklin, US Forest Service; Dr Robert A. Lewis, University of Saarbrucken; Dr William S. Osburn, US Department of Energy; and anonymous referees, some of whom have not agreed with all that I have said. Much of any credit for clarity and precision of thought is, however, due to these critics.

Dr Jerry F. Franklin and Dr Arthur McKee provided important assistance in site selection, and Dr McKee was instrumental in providing the long-term field support required in the methods optimization effort.

This research was performed for the US Department of Energy under contract number DE-AC06-76RLO 1830.

## SUMMARY

Ecological monitoring is the purposeful observation, over time, of ecological processes in relation to stress. It differs from biological monitoring in that ecological monitoring does not consider the biota to be a surrogate filter to be analysed for contaminants, but rather has changes in the biotic processes as its focal point for observation of response to stress. Ecological monitoring methods aimed at detecting subtle or slow changes in ecological structure or function usually cannot be based on simple repetition of an arbitrarily chosen field measurement. An optimum method should be deliberately designed to be ecologically appropriate, statistically credible, and cost-efficient.

Ecologically appropriate methods should consider the ecological processes that are most likely to respond to the stress of concern, so that relatively simple and well-defined measurements can be used. Statistical credibility requires that both Type I and Type II errors be addressed; Type I error (a false declaration of impact when none exists) and Type II error (a false declaration that no change has taken place or that an observed change is random) are about equally important in a monitoring context. Therefore, these error rates should probably be equal. Furthermore, the error rates should reflect the large inherent variability in undomesticated situations; the optimum may be 10%, rather than the traditional 5% or 1% for controlled experiments and observations.

Application of these principles in two contrasting situations, closed-canopy forests and contaminated food-'chains, is under way. In forests an increase of stress induces premature abscission of senescent leaves. Using

needlefall as an indictor of stress conditions in coniferous forests of western North America, our research suggests that site-specific needlefall rate can be estimated within  $\pm 10\%$  with 90% confidence, using an optimum collector design that requires only 2 man-hours per site per month to handle the sample from field to computer-compatible data. Our examination or food-chains leading into Herons (Ardea herodias) suggests that contaminant levels can be readily estimated by collecting food-scraps and faeces ejected from nests. Collection during the nesting season is simple, minimizes injury and disturbance of the birds, and allows sampling and replication variances sufficiently to suggest that only a few samples per site need to be analysed. Depending on the contaminant under study, the number of independent Heron nesting-sites required to sample a region adequately varies from ten upwards.

These applications of optimal sampling approaches to large-scale monitoring design are not exhaustive, and many other kinds of ecological measurements are desirable. However, long-term trends in ecological structure or function are impossible to detect by the use of poorlydesigned methods or intermittently-collected data. Therefore, work must continue towards the development of long-term measurements that, in the manner of temperature in climatology, reflect widely useful and robust measurements.

#### REFERENCES

- ANON. (1978). Federal environmental monitoring; will the bubble burst? *Environmental Science and Technology*, 12(12), pp. 1264–9.
- AXELROD, M.C., COYNE, P.I., BINGHAM, G.E., KERCHNER, J.R., MILLER, P.R. & HUNG, R.C. (1980). Canopy analysis of pollutant-injured Ponderosa Pine in the San Bernardino National Forest. P. 227 in Proceedings of Symposium on Effects of Air Pollutants on Mediterranean and Temperate Forests (Ed. P.R. MILLER). Pacific Southwest Forest and Range Experiment Station, Berkeley, California, USA: 256 pp., illustr.
- BARRETT, G.M. & ROSENBERG, R. (Ed.) (1981). Stress Effects on Natural Ecosystems. (Environmental Monographs & Symposia, Convener & Gen. Ed. N. POLUNIN.) John Wiley & Sons, Chichester-New York-Brisbane-Toronto-Singapore: xviii + 305 pp., illustr.
- BOURDEAU, P. & TRESHOW, M. (1978). Ecosystem response to pollution. Pp. 313–30 in *Principles of Ecotoxicology* (Ed. G.C. BUTLER). John Wiley & Sons, New York, NY, USA: xxii + 350 pp., illustr.
- BROWN, K.T. (1981). Indirect costs of federally supported research. Science, 212, pp. 411-8.
- BUFFINGTON, J.D. (1978). Statement. Pp. 146-86 in Hearings before the Subcommittee on the Environment and the Atmosphere of the Committee on Science and Technology, US House of Representatives, June 27-29, July 21, 1978. Government Printing Office, Washington, DC, USA: iii + 321 pp.
- CAIRNS, J., jr (1974). Indicator species vs the concept of community structure as an index of pollution. Water Resources Bulletin, 10, pp. 338–47.
- CAIRNS, J., jr (1980). Estimating hazard. *BioScience*, **30**, pp. 101-7.

- CAIRNS, J., jr & SCHALIE, W.H. van der (1980). Biological monitoring Part 1–Early warning systems. Water Research, 14, pp. 1179–96.
- COLWELL, R.R. (1978). Toxic effects of pollutants on microorganisms. Pp. 275–94 in *Principles of Ecotoxicology* (Ed. G.C. BUTLER). John Wiley & Sons, New York, NY, USA: xxii + 350 pp., illustr.
- Cowell, E.B. (1978). Ecological monitoring as a tool in industry. Ocean Management, 4, pp. 273-85.
- DYNE, G.M. Van (1981). Response of shortgrass prairie to maninduced stresses as determined from modelling experiments. Pp. 57–70 in BARRETT & ROSENBURG (q.v.).
- EBERHARDT, L.L. (1976). Quantitative ecology and impact assessment. J. Environmental Management, 4, pp. 27-70.
- EBERHARDT, L.L. (1978). Appraising variability in population studies. J. Wildlife Management, 42, pp. 209-38.
- FITZNER, R.E., RICKARD, W.H. & HINDS, W.T. (1982). Excrement from Heron colonies for environmental assessment of toxic elements. *Environmental Monitoring and Assessment*, 1, pp. 383-6.
- FRANKLIN, J.F., JENKINS, R.E. & ROMANCIER, R.M. (1972). Research Natural Areas: Contributors to environmental quality programs. J. Environmental Quality, 1, pp. 133–9.
- GAUCH, H.G., jr (1982). Multivariate Analysis in Community Ecology. Cambridge University Press, Cambridge, England, UK: x + 298 pp., illustr.
- GOLDBERG, E.D., BOWEN, V.T., FARRINGTON, J.W., HARVEY, G., MARTIN, J.H., PARKER, P.L., RISEBROUGH, R.W., ROBERTSON, W., SCHNEIDER, E. & GAMBLE, E. (1978). The Mussel Watch. *Environmental Conservation*, 5(2), pp. 101-25, illustr.
- GRAY, J.S. (1980). Why do ecological monitoring? Marine Pollution Bulletin, 11, pp. 62-5.
- HOLDEN, A.G. (1978). Monitoring environmental materials and specimen banking for organohalogenated compounds in aquatic ecosystems. Pp. 320–41 in LUEPKE (q.v.).
- HOLDGATE, M.W. (1979). A Perspective of Environmental Pollution. Cambridge University Press, Cambridge–London– New York–Melbourne: x + 278 pp., illustr.
- INMAN, J.C. & PARKER, G.R. (1978). Decomposition and heavymetal dynamics of forest litter in northwestern Indiana. *Environmental Pollution*, 17, pp. 39–51.
- JACSIC, F.M. (1981). Abuse and misuse of the term 'guild' in ecological studies. Oikos, 37, pp. 397-400.
- KENDALL, R.J. (1982). Wildlife toxicology. Environmental Science and Technology, 16(8), pp. 448A-53A.
- KEULEN, A. Van, Seligman, N.G. & Benjamin, R.W. (1980-81). Simulation of water use and herbage in arid regions—a re-evaluation and further development of the model ARIDCROP. Agricultural Systems, 6, pp. 159-93.
- KNOX, G.A. & POLUNIN, N. (Ed.) (in press). *Ecosystem Theory* and Application. (Environmental Monographs & Symposia, Convener & Gen. Ed. N. POLUNIN). John Wiley & Sons, Chichester-New York-Brisbane-Toronto-Singapore.
- LAVEN, R.D. (1982). Establishing homogeneity in studies of forest succession. Forest Ecology and Management, 4, pp. 161-77.
- LOON, J.C. Van (1975). How useful are environmental chemicaldata? Pp. 349–55 in International Conference on Heavy Metals in the Environment: Symposium Proceedings Volume 1. (27–31 October 1975.) Institute for Environmental Studies, University of Toronto, Ontario, Canada: MSSIAI, vii + 371 pp., illustr.

- LUEPKE, N.P. (Ed.) (1978). Monitoring Environmental Materials and Specimen Banking. Martinus Nijhoff Publishers, The Hague, Netherlands: xiii + 591 pp., illustr.
- McLAUGHLIN, S.B., McCONATHY, R.K., DUVICK, D. & MANN, L.K. (1982). Effects of chronic air pollution stress on photosynthesis, carbon allocation, and growth of White Pine Trees. *Forest Science*, **28**, pp. 60–70.
- MCSHANE, M.C., CARLILE, D.W. & HINDS, W.T. (in press). The effect of sampler size on forest litterfall collection and analysis. *Canadian Journal of Forest Research*.
- MANNING, W.J. & FEDER, W.A. (1980). *Biomonitoring Air Pollutants with Plants*. Applied Science Publishers, London, England, UK: x + 142 pp., illustr.
- NATIONAL RESEARCH COUNCIL (1977). Analytical Studies for the US Environmental Protection Agency, Volume IV. National Academy of Sciences, Washington, DC, USA: xiii + 181 pp., illustr.
- NATIONAL RESEARCH COUNCIL (1980). The International Mussel Watch. National Academy of Sciences, Washington, DC, USA: xvi + 248 pp., illustr.
- NATIONAL RESEARCH COUNCIL (1981). Testing for Effects of Chemicals on Ecosystems. Committee to Review Methods for Ecotoxicology, National Research Council. National Academy Press, Washington, DC, USA: xv + 103 pp., illustr.
- NATIONAL SCIENCE FOUNDATION (1977). Long Term Ecological Measurements. National Science Foundation, Washington, DC, USA: 26 pp.
- NEWTON, I. (1979). *Population Ecology of Raptors*. Buteo Books, Vermillion, South Dakota, USA: 399 pp., illustr.
- O'NEILL, R.V., AUSMUS, B.S., JACKSON, D.R., HOOK, R.I. Van, VORIS, P. Van, Washburne, C. & Watson, A.P. (1977). Monitoring terrestrial ecosystems by analysis of nutrient export. *Water, Air, and Soil Pollution*, 8, pp. 271-7.

- RICKARD, W.H., HEDLUND, J.D. & SCHRECKHISE, R.G. (1978). Rejecta cast from Heron nests as an indicator of food-chain contamination. *The Auk*, **95**, pp. 425–7.
- SEVERINGHAUS, W.D. (1981). Guild theory development as a mechanism for assessing environmental impact. *Environ*mental Management, 5, pp. 187–90.
- SKALSKI, J.R., & MCKENZIE, D.H., (1982). A design for aquatic monitoring programs. J. Environmental Management, 14, pp. 237–51.
- SMITH, W.H. (1981). *Air Pollution and Forests*. Springer Verlag, New York, NY, USA: xv + 379 pp., illustr.
- SNEDECOR, G.W. & COCHRAN, W.G. (1967). Statistical Methods. Iowa State University Press, Ames, Iowa, USA: xiv + 593 pp., illustr.
- SUTER, G.W. (1981). Ecosystem theory and NEPA assessment. Bulletin of the Ecological Society of America, 62, pp. 186–92.
- SYMEONIDES, C. (1979). Tree-ring analysis for tracing the history of pollution: application to a study in Northern Sweden. J. Environmental Quality, 8, pp. 482–6.
- THOMAS, J.M., MCKENZIE, D.H. & EBERHARDT, L.L. (1981). Some limitations of biological monitoring. *Environment International*, **5**, pp. 3–10.
- VANDERHORST, J.R. & WILKINSON, P. (1979). The Littleneck Clam, *Protothaca staminea*, as a tool for potential oil pollution assessment, Part 1: Density of stock. *Marine Environmental Research*, **2**, pp. 223–37.
- WARING, R.H. & FRANKLIN, J.F. (1979). Evergreen coniferous forests of the Pacific Northwest. *Science*, **204**, pp. 1380–6.
- WILLIAMS, W.I. (1983). Tree growth and smog disease in the forests of California: case history, Ponderosa Pine in the southern Sierra Nevada. *Environmental Pollution* (Series A), 30, pp. 59–75.
- WORF, D.L. (Ed.) 1980. Biological Monitoring for Environmental Effects. Lexington Books, D.C. Heath and Company, Lexington, Massachusetts, USA: xii + 227 pp., illustr.