NISS

Combining Environmental Information: Environmetric Research in Ecological Monitoring, Epidemiology, Toxicology, and Environmental Data Reporting

Lawrence H. Cox and Walter W. Piegorsch

Technical Report Number 12
April, 1994

National Institute of Statistical Sciences
19 T. W. Alexander Drive
PO Box 14006
Research Triangle Park, NC 27709-4006
www.niss.org

COMBINING ENVIRONMENTAL INFORMATION: ENVIRONMETRIC RESEARCH IN ECOLOGICAL MONITORING, EPIDEMIOLOGY, TOXICOLOGY, AND ENVIRONMENTAL DATA REPORTING

Lawrence H. Cox[1] **and** ***Walter W. Piegorsch***[2]

ABSTRACT

An increasingly important concern in environmental studies is the need to combine information from diverse sources that relate to a common endpoint or effect and to combine environmental monitoring and assessment data. These are statistical problems, and statistical techniques are integral to analyses that combine environmental information. These techniques are still under development, however, as modern statistical methods for combining environmental information usually require subject-specific formulations. Herein, we discuss opportunities for statistical research in the area of combining environmental information, based on information presented during a 1993 Workshop on Statistical Methods for Combining Environmental Information, organized by the National Institute of Statistical Sciences and the U.S. Environmental Protection Agency.

KEY WORDS: Combining information; data reporting; data aggregation; environmental epidemiology; meta-analysis; environmental monitoring and assessment.

---

[1] Senior Statistician, US Environmental Protection Agency, AREAL (MD-75), Research Triangle Park, NC 27711, USA

[2] Department of Statistics, University of South Carolina, Columbia, SC 29208, USA

1. INTRODUCTION

Combining information is a growing issue of importance in environmental science, for which statistical methods play a central role. To explore statistical problems in combining information as posed by environmetric applications, the National Institute of Statistical Sciences (NISS) and the U.S. Environmental Protection Agency (US EPA) organized a Workshop on Statistical Methods for Combining Environmental Information that was held September 27-28, 1993 in Chapel Hill, NC. The Workshop was organized around a selection of environmental problems and studies that demonstrate multiple needs for combining environmental information.

The program covered the following environmental data problems. (1) Combining environmental data from multiple and diverse sources: statistical reporting on environmental conditions and trends in aquatic, terrestrial, and atmospheric settings, and combining design-based ecological data and observed data for environmental assessment purposes. (2) Combining environmental epidemiologic studies for hazard identification and risk assessment. Associated public health issues include assessing exposures to environmental tobacco smoke, dioxin, and nitrogen dioxide, assessing acute inhalation risk, assessing effectiveness of lead abatement strategies, and using prior information and hierarchical Bayes methods to model uncertainty in effect estimation. (3) Forming environmental indicators and indexes, including issues of aggregation, combined mapping procedures, and multiple data source conformance.

In this paper, we report on presentations and discussions from the Workshop, and expand upon some of the questions and open areas of environmetric research that were identified. Our goal is to identify and disseminate methods and research themes leading to solutions for these problems, and to stimulate further interdisciplinary research in this important area.

2. COMBINING ENVIRONMENTAL DATA FOR ENVIRONMENTAL MONITORING AND ECOLOGICAL ASSESSMENT

2.1. EXAMPLE: THE U.S. ENVIRONMENTAL MONITORING AND ASSESSMENT PROGRAM (EMAP)

The Environmental Monitoring and Assessment Program (EMAP) is a multi-agency program of the U.S. government being organized by the U.S. Environmental Protection Agency. The objectives of EMAP are: to determine the extent of the nation's ecological resources; to define status and trends in the condition of these resources; to identify possible causes of the current and changing condition of the resources; and to report regularly on the findings. EMAP poses several challenging problems in the area of statistical methods for combining environmental information.

The scope of EMAP comprises all ecological resources, categorized within eight resource groups: landscapes, forests, agroecosystems, arid ecosystems, estuaries, the Great Lakes, lakes and streams, and wetlands. Information on status and trends in ecological condition will be developed by geographic region for each resource. This requires ecological monitoring, assessment and reporting across time and space by region, resource category and individual resources, necessitating the development of ecologically meaningful environmental indicators with suitable statistical properties. An important EMAP goal is to combine environmental indicators within and across resources into ecological indexes designed to measure overall ecological well-being. As with the development of any statistical index, the first issue of statistical importance here is the proper delineation and aggregation of the data (Ott 1978). This is only the first step, however, as EMAP is based on an approach to ecological assessment which stresses extension beyond simple data aggregation to the integration and synthesis of information.

2.2. COMBINING PROBABILITY BASED AND NON-PROBABILITY BASED MONITORING DATA

EMAP programs will collect data on ecological resources using multi-stage probability based spatial sampling designs; see Stevens (1994) for the EMAP sampling strategy. There is in existence a large quantity and variety of environmental monitoring data from non-EMAP sources from either other forms of probability designs or, most typically, not based on a probability design at all. Such data will continue to be collected by researchers in large volume into the foreseeable future, and these data are likely to be based on different forms of sampling designs, to cover different but overlapping ecological populations and time periods, and to measure different characteristics or measure EMAP characteristics differently. The processes of identifying and collecting the data, performing data quality assurance, analyzing and reporting on the data, and the policy actions resulting therefrom are quite complex as well as time and cost intensive. Within this context, the EMAP strategy is to make maximum appropriate use of relevant data from all sources.

In addition to complexity and cost, ecological monitoring programs share the following characteristics: limited coordination of design and measurements across monitoring programs; different variables or indicators measured; different time frequency of sampling with a year; different number of years sampled; and, different objectives of monitoring programs (e.g., emphasis on monitoring predetermined sites versus emphasis on measuring predetermined ecological resource characteristics).

Studies focusing on ecological regions typically fall into one of the following three categories: probability based sample studies of well-defined ecological resource populations designed to estimate status and trends of these characteristics; sites selected via **purposive** (i.e., non-random) **sampling** from multiple resource populations used to represent or model surrounding regions (e.g., regions defined based on the sites using **variography** (see Cressie 1991); and, predetermined regions (e.g., regions based on geo-political boundaries) from which representative sites are purposively selected.

The EMAP objectives pose a suite of statistical problems involving the combination of probability based monitoring data (**P-sample data**) with other probability based data and with non-probability based data (**NP-sample data**). Candidates for data combination include environmental monitoring data among the following types and sources: compliance monitoring data from designated sites; data from point source impact assessment sites; data from ecological resource research sites and studies; state monitoring data of special interest sites; local region probability based data and studies of ecological resources; and national and regional probability based studies of ecological resources.

Examples of environmental data sets, studies, or other information, within the estuary resource category include: EMAP estuarine monitoring data; the U.S. National Oceanographic and Atmospheric Administration (NOAA) Status and Trends Program data involving sites representative of large coastal areas (but avoiding known sites exhibiting extreme characteristics); the US EPA National Estuaries Program involving approximately 20 designated estuaries monitored independently; and various state estuary monitoring programs or studies, typically designed for compliance or regulatory purposes with sites selected to meet established criteria. Examples from the lake resource category include: EMAP Surface Waters Lake Monitoring Program based on a probability sample of a well-defined lake population; State 305(b) lake monitoring programs which exhibit different selection criteria and both fixed and (annually) changing sample populations of lakes; and, U.S. Geological Survey (USGS) surface water monitoring networks involving purposive sampling of sites based on scientific judgment and multiple annual samples. In addition, national probability based ecological monitoring programs include: EMAP; NOAA's National Wetlands Status and Trends; the U.S. Forest Service's Forest Inventory and Analysis monitoring programs; the National Resource Inventory; and, various agroecosystems monitoring programs of the National Agricultural Statistics Service.

The kinds of questions ecological monitoring programs are designed to answer are numerous and varied. For example, what is the status (amount or extent) of an ecological resource in a region? What is the trend in the status of the resource? Has the spatial pattern/geographic distribution of the resource changed over the reference time period? What is the status of the ecological condition of the resource in the region? And, what is the trend in the condition of the resource? Typical information needed to combine monitoring data to address these questions involves knowledge of site selection procedures; knowledge of exact definition of attributes measured; knowledge of sampling, response and reporting periods for attributes; knowledge of the quality of the data; and, knowledge of the ecological resource population of interest in each study or program, if any.

The standard data combination paradigms in ecological monitoring are: combining data or estimates from two or more P-samples; combining P-sample data with NP-sample site-based data from one or more monitoring programs; and combining NP-sample site-based data from two or more monitoring programs. Statistical techniques are available for simplified cases of the first two paradigms (see below); however, rigorous statistical work is needed for the third.

Two standard statistical approaches are available for combining estimates from two probability samples (a P_1-sample and a P_2-sample). The first is to generate estimates independently from the P_1- and P_2-samples and then combine the estimates using weighting. Strata can be defined based on partitions of the two overlapping populations, enabling the construction of combined subpopulation estimates that can be used to post-stratify the samples. The multiple frame methodology of Hartley (1974) then can be used to produce combined estimates. Alternatively, the P_1- and P_2-samples can be combined into a single probability sample (Overton 1990). This requires computing combined-sample inclusion probabilities for all sample units, which can be done if both the first and second order inclusion probabilities for the P_1- and P_2-samples are available.

Overton (1990) and Overton, Young and Overton (1993) address the problem of combining a P-sample with a NP-sample. The NP-sample must be a subset of the population represented by the P-sample. The population is then partitioned such that one of the partitions is the NP-sample. The crucial assumption is that the NP-sample is a probability based (P_2-) sample of some (larger) subset of the population. In general, this is difficult to verify and is liable to introduce systematic bias into combined sample estimates. Overton suggests validating this assumption by demonstrating similarity of P_2-frame attributes of the NP-sample with population subsets. This might be accomplished using **statistical record linkage**. However, little is known about record linkage in the environmental context; see Cox and Boruch (1988), Rodgers (1984) and references therein for discussion of statistical record linkage in the socio-economic context. An alternative is to treat the NP-sample as a separate stratum of **self-representing units**, but this is unlikely to improve precision. Once a probability basis for the NP-sample is created, the problem has been reduced to that of combining two probability samples, as above. Strata can be based on similarity clusters of frame attributes, from which combined sample inclusion probabilities can be computed.

A conceptual data model for combining data from a P-sample with data from an NP-sample (developed by A. Olsen, US EPA), is based on the following matrix:

$$Y_t = \begin{pmatrix} L_P & A_{PF} & Y_{PM} & Y_{PD} \\ L_N & A_{NF} & Y_{NM} & Y_{ND} \end{pmatrix}$$

where

L_P = spatial location of P-sample units
A_{PF} = frame attributes of P-sample units
Y_{PM} = variables directly measured on P-sample units
Y_{PD} = variables derived from external data and associated with P-sample units
L_N = spatial location of NP-sample sites
A_{NF} = frame attributes of NP-sample sites
Y_{NM} = variables directly measured on NP-sample sites
Y_{ND} = variables derived from external data and associated with NP-sample sites.

The conceptual data model $\mathbf{Y_t}$ defines the data available at time **t** for use in statistical estimation of a population characteristic. The data model recognizes specific types of data attributes that are important in environmental applications, probability-based sampling, and spatial statistics. The upper portion of the matrix addresses the P-sample attributes; the lower addresses NP-sample attributes. $\mathbf{L_P}$ and $\mathbf{L_N}$ define the spatial location of sample units. Typically, location is given as latitude and longitude coordinates; or at least can be easily converted to them, as it is not sufficient to have location coordinates given in an arbitrary, non-geo-referenced coordinate system. $\mathbf{A_{PF}}$ and $\mathbf{A_{NF}}$ contain frame attributes as usually defined by survey design literature. For example, lake area and stream order are two attributes that are used in the sample design for lakes and streams in EMAP. Other attributes could be acquired that may not be used directly in structuring the sampling design; but could be used in a ratio or regression estimation procedure (e.g. lake elevation). $\mathbf{Y_{PM}}$ are attributes measured directly on sampling units in the P-sample. Examples are crown transparency in forests, index of biotic integrity in streams, and net primary productivity for agricultural land. These attributes are collected under well-defined protocols and field methods. In some situations, $\mathbf{Y_{NM}}$ attributes are available for NP-sample units.

A critical issue is knowing if the NP-sample attributes were collected using same procedures as the P-sample, or if they are similar attributes that have been, or must be, calibrated prior to using in joint estimation procedure. $\mathbf{Y_{PD}}$ and $\mathbf{Y_{ND}}$ are variables derived from sources other than direct measurements on a sampling unit in a monitoring program. Examples would be soil characteristics as determined from Soil Conservation Service soils databases or point source discharge into stream from an EPA database. A second critical issue is establishing the linkage between the external database elements and the sampling unit from the monitoring program(s). Spatial location is an important linkage function; however, nearest neighbor is only one approach for spatial linkage.

Probability design-based and some model-based inference of status can be based on A_{PF} and Y_{PM}. Model-based inference of status that incorporates spatial location can be based on L_P, A_{PF} and Y_{PM}. In some cases, variables Y_{PD} derived from external data can be associated with a P-sample unit. Estimates of status for these variables and associations of these variables with directly measured variables Y_{PM} may be of interest. Often, directly measured variables Y_{NM} from NP-sample sites that measure the same variables Y_{PM} from the P-sample are available.

Overton (1990) suggests that (Y_{NM}, Y_{ND}) relationships can be combined with Y_{PM} data to construct derived variables Y_{PD}, as follows. Model (Y_{NM}, Y_{ND}) relationships via regression and apply the regression equations to Y_{PM} to estimate Y_{PD}. Problems with this procedure are the potential presence of undetected selection bias in the NP-sample and failure of the regression to account for true variation in the Y_{PD}. An important related question is whether A_{NF} and Y_{NM} information can be combined or associated with A_{PF} and Y_{PM}. Clearly, much statistical research remains in the process of developing valid and useful statistical inferences for these data combination scenarios.

2.3. ISSUES IN COMBINING SPATIALLY REFERENCED MONITORING AND ASSESSMENT DATA

Environmental monitoring data are inherently spatial, and spatial statistics has an important role to play in modeling and understanding environmental and ecological problems. Eco-environmental data bases typically exhibit the following characteristics: continuous data from monitoring measurements or laboratory analysis combined with typically fewer categorical and classificatory variables; data on some variables such as vegetation collected and possibly stored in image form; long collection periods needed to detect trends; and, presence of important spatial variables that are not necessarily well-defined (e.g., determining boundaries of geo-political units from satellite images). Analyses of eco-environmental data share the following features: lack of well-developed methods of aggregation; interest in heterogeneous, "combined" phenomena rather than homogeneous properties of variables; and, greater interest in spatial rather than population distributions. Differences in statistical approaches to spatial problems exist between environmental and ecological problems, viz., geostatistical models have traditionally been used in spatial environmental statistics, while lattice models and point processes are more commonly used in spatial ecological statistics. These approaches need to be combined for eco-environmental applications, and the study of eco-environmental relationships and dynamics demands further development and use of spatio-temporal statistics.

Application of spatial statistics is based upon the intuitively appealing notion that nearby data or objects are more likely to be more alike than those far apart. This observation goes back at least as far as Fisher (1935), who noted "after choosing the area, we usually have no guidance beyond the widely verifiable fact that patches in close proximity are commonly more alike, as judged by the yield of crops, than those which are far apart." Thus, natural spatial correlation causes one to lose access to most statistical theory which is predicated on independent, identically distributed (**i.i.d.**) errors. Attempts to transfer time series theory to the spatial context has been of some, but limited value. Spatial dependence needs to be recognized and modelled in modern eco-environmental data analysis.

In addition to gaps in the theory of spatial statistics, there are practical obstacles to analyzing and combining spatial data. Two such obstacles are: difficulties in referencing spatial data in terms of assigning and manipulating spatial labels for statistical analysis; and, dealing with data designated as "nondetects" in laboratory studies (where information can be ignored or deleted).

The emergence of so-called geographic information systems (**GIS**) is beginning to address the problem of spatial labels. GIS is an integrated computer hardware and software system designed to collect (input), manage (store, retrieve), analyze (aggregate, estimate, optimize, simulate), and display (map, graph, tabulate) spatially referenced data. Application of GIS is a powerful computing tool with desirable capabilities for spatial labeling, but unfortunately it has not been applied in a consistent statistical manner, and has not yet become a statistical tool (Cressie and ver Hoef 1993). Limited attempts to piece together GIS systems (e.g., ARC/INFO, GRASS) with statistical software systems (e.g., SAS, S-PLUS) provide inadequate solutions, since, in current systems, statistical analysis and computations such as variance estimation are based predominately on classical i.i.d. assumptions which are often inappropriate in the spatial context.

Similarly, the **nondetect problem** can be viewed simply, as essentially a censored data problem, and is an important area where statistical theory and practice has substantial application; see Lambert, Peterson, and Terpenning (1991). Certain laboratory measurements are reported as nondetects (**ND**), either because the value falls below a threshold limit of detection (**LOD**) value of the instrumentation or because the presence of the compound of interest cannot be verified reliably. This inability to quantify the concentration of the chemical may be due to limitations in instrumentation, laboratory protocol, or both. In the former case, nondetect values may vary from instrument to instrument or among laboratory technicians. In the latter case, protocol definitions may cause otherwise useful measurement data to be discarded. In both cases, it is desirable to estimate missing values and their effect on estimates of dispersion in the data.

To the statistician, laboratory procedures involving nondetects may appear to be an overly rigid interpretation of Neyman-Pearson hypothesis testing, in effect forcing an accept/reject philosophy for data quality assurance. Imputing 0, LOD or LOD/2 for ND, or the use of nonparametric ranking or continuous models for censored data, often perform poorly. This is not uncommon when dealing with spatial data. Lambert, Peterson and Terpenning (1991) address the nondetect problem by using local logistic regression to model the probabilities of detection for both measured and actual concentrations. Another approach is to use the geostatistical method known as **kriging**, which involves the use of optimal spatial averaging; see Cressie (1990; 1991) and Myers (1991).

A spatial statistic of frequent interest is the **volume average**

$$z(B) = (1/|B|) \int_B z(\underline{u})\,d\underline{u}$$

where B is the **support** of z(B). Support is a key concept in spatial statistics with particular importance in environmental applications. A generic problem in volume average estimation is to predict z(B) or a function g(z(B)), based on monitoring data. The geostatistical method of choice here is kriging, and it solves this problem in a careful way. An advantage of kriging and other averaging methods is the reduction of measurement error through the modeling process. Through kriging, nondetect values are estimable. In addition, kriging can be employed to combine P-sample and NP-sample based data, as described below.

An environmetric application of the use of kriging to combine P-sample and NP-sample environmental data is the snow water equivalent (SWE) problem faced by the National Weather Service. The SWE predicts the amount of water in streams during the spring thaw based on data from snow course sites (P-sample data) and airborne surveys (NP-sample data). This problem is of particular importance in the western United States. A spatial statistical model for the SWE problem is constructed as follows. Let **z(s)** denote the SWE at location **s**. A geostatistical model is assumed of the general form:

$$\begin{aligned} E(z(\underline{s})) &= X(\underline{s})'\underline{\beta} \\ \operatorname{cov}(z(\underline{s}), z(\underline{u})) &= C(\underline{s}, \underline{u}) \quad \text{or} \\ \operatorname{var}(z(\underline{s}) - z(\underline{u})) &= 2\gamma(\underline{s}, \underline{u}) \end{aligned}$$

where **z(B)** denotes the volume average of SWE with support **B**. The geostatistical model is estimated from P-sample data located at snow course sites $\{s_i: i=1,...,n\}$ and airborne survey sites $\{B_i: i=n+1,...,m\}$. Although SWE could be estimated (using kriging) separately from the snow course sites and the airborne data, a combined kriging estimate would combine the strengths and compensate for the weaknesses of both data sets: a combined estimate would be based on a larger data set, and it would take advantage of the generally more representative airborne estimates. However, model complexities arise from factors such as irregular spacing of sites and differences in types of support, viz., point support for snow course sites and block support for airborne surveys, necessitating care in the modelling and further research. See Carroll, Day, Cressie, and Carroll (1994) for discussion of these issues and the SWE problem.

Many similar useful applications are envisioned in the area of combination of spatial environmental data.

2.4 CHESAPEAKE BAY POLLUTION STUDY

Another environmetric application of combining environmental data occurs with a joint US EPA/State of Maryland study of nutrient reduction due to pollution in Chesapeake Bay. The study is in the planning stages, and design issues are being considered of where and how to sample certain Bay locations. Of interest is how nutrient abundance affects the plant and animal communities of the sediment (the **benthos**) of the Bay. A nutrient index will be employed to assess conditions of the benthic assemblages at particular locations within the Bay. The index consists of a weighted sum of various species quantifiers:

$$\mathbf{Z = .011\upsilon + .671\pi + .817\alpha + .577\omega + .465\gamma}$$

where for each location sampled, υ is the salinity-adjusted number of species observed, π is the percent of total benthic bivalve abundance, α is the number of amphipods (crustaceans with multi-purpose feet), ω is the average weight per polychaete (a type of segmented marine worm), and γ is the number of capitellids (a special form of polychaete) sampled. The index is taken over 31 different locations throughout the Bay, representing four different forms of aquatic ecosystems: tidal fresh/oligohaline; low mesohaline; shallow high mesohaline; and, deep high mesohaline.

An initial goal of the study is to represent the "benthic quality" of the Chesapeake Bay in the form of a map based on the benthic invertebrate summary index **Z**. From this, other assessments can be made, such as using the mapped summary index to estimate the percentage of Bay acreage exhibiting degraded biotic conditions, or to estimate the portion of acreage that might be restored by reducing pollutants that affect benthic invertebrates. Of course, any map of the Bay developed from **Z** will be affected by spatial variation across the 31 different sample strata; hence, methods for spatial data combination are important for the construction and use of the index.

A major concern in this data combination is to make the mapping as robust as possible to effects of unadjusted spatial variation within and across strata. To do so, the mapping procedure is based on a two-part model: large-scale variation such as spatial trends and other covariates that model the effects along large spatial scales; and, small-scale variation due to measurement error, spatial correlation, and other spatial small-scale effects. The large-scale variation may be estimated using regression methods such as least squares or **L_1 regression** (Narula and Wellington 1982): from the sample strata in the Bay, the observed index values are combined using a linear regression model incorporating location-specific covariates such as stratum depth, latitude, and longitude, to estimate a predicted value, $\hat{\boldsymbol{Z}}(\boldsymbol{s})$, for the large-scale variation at each stratum $\boldsymbol{s}$.

This predicted value does not correct for small-scale effects such as spatial correlation, however. For this, the small-scale variation is estimated via more specialized methods such as kriging, wherein the spatial correlations are modelled after removing the large-scale trends. This leads to a kriged predicted value, and from that, a kriged residual, $\hat{\varepsilon}(\boldsymbol{s})$. Then, the regression predictor and the kriged residual are added to form a hybrid predictor, $\hat{\boldsymbol{Z}}(\boldsymbol{s}) + \hat{\varepsilon}(\boldsymbol{s})$, for any spatial stratum $\boldsymbol{s}$ within the Bay. By construction, this hybrid calculation corrects for both small-scale spatial effects and large-scale trends. It is anticipated that the hybrid analysis will enhance scientists' ability to report ecological data for public use.

2.5. DEVELOPING AND COMBINING ECOLOGICAL INDICATORS AND INDEXES

EMAP will measure and report ecological condition by **ecological indicators** such as crown height in forests, eutrophication of lakes, or acidity in soil. Indicator values will be constructed on scales useful for assessment purposes. At a minimum, values will be partitioned into nominal/subnominal categories, which can be used for reporting, comparison and, potentially, regulatory purposes. Estimates of trend in indicator values will be used to identify and assess changes in ecological condition.

Many environmental policy questions will require answers on large geographic or ecological scales. For instance, what is the condition of lakes in the Northeast? Or, how has the ecological condition of the Louisianian Province changed over the past ten years? From the policy perspective, it is desirable to aggregate or in some other way combine ecological indicators across geographic regions, ecological resources and resource groups. For example, what can be said quantitatively about the condition of streams in the West given indicator data on individual watersheds? How reliable are such statements statistically? And, how can this information be used over time to measure the progress of clean-up and remediation programs? From the reporting and regulatory perspective, it is desirable to be able to combine different indicators into **eco-environmental indexes** that would convey broad-gauge measures of environmental health to the public and policy makers. For example, was the condition of forests in the Northwest during the past year good or poor? And, how does this year compare with the last (ten) year(s)?

Much is known and is in use for the quantity- and monetary-based indicators and indexes familiar in economic statistics, stemming from the work of German economists Etienne Laspeyres and Hermann Paasche during the mid-nineteenth century. The (price) indexes bearing their names illustrate some of the elementary statistical problems encountered in index construction. Both indexes measure change in price ($\mathbf{p}$) over time for a fixed market basket of quantities ($\mathbf{q}$) using the arithmetic average: $\mathbf{C_t = \Sigma p_{it} q_i}$ and define the index as $\mathbf{I_t = C_t / C_0}$ relative to a reference time $T = 0$. They differ in terms of the time at which the market basket quantities $\mathbf{q_i}$ are selected. Indexes also can be constructed using geometric means or other measures of centrality including, more recently, the use of regression. Early work on the statistical properties of indexes was performed by Fisher (1922). ISI (1956) provides an historical bibliography. All of this work deserves consideration as problems--old and new--arise in the development and combination of environmental indexes.

Some work has been done on index construction elsewhere in the environmental context which can be brought to bear in the ecological context. The general problem was considered from the policy perspective by the National Academy of Sciences (NAS 1975), which concluded that environmental indexes had an important role in: assisting policy formulation; providing a means for evaluating the effectiveness of environmental protection programs; assisting in designing such programs; and, facilitating communication with the public on environmental conditions and progress towards environmental enhancement. Subsequently, Ott (1978) offered a comprehensive approach towards developing environmental indexes, including detailed examination of air and water quality indexes.

NAS (1975) recommended that a uniform national system of air quality indexes be developed and adopted. This was undertaken by US EPA, resulting in the national Pollutant Standards Index (**PSI**). The PSI takes daily air pollution information for five pollutants for which exist short term air quality standards or significant harm levels have been established (CO, SO_2, PM_{10}, O_3 and NO_2) and computes a single daily air quality index value. The air quality indicator information for each pollutant is reported on a common scale (0 - 500) comprising five increasing health effects (risk) categories ranging from Good to Hazardous. The value of the daily index is the maximum of the five indicator values, ensuring that the index value does not **eclipse** (mask) the value of any constituent pollutant. Hunt (1991) provides a comprehensive discussion of the issues and methodology surrounding the development of the PSI. This experience should provide a baseline for the present problem of developing broadly useful ecological indicators and indexes.

These problems will require advanced, practical formulations and solutions from the statistical and environmental research communities. Key ecological issues are: to identify meaningful indicators and indexes and measurement scales, to develop robust definitions of nominal/subnominal and other meaningful descriptors of condition, and to determine what data need to be combined at what levels for policy, reporting and research purposes. Key statistical issues will be in aggregation, reliability and other statistical properties of the indicators and indexes, and in the development of methods for presenting and analyzing this information, including graphical methods.

A significant statistical issue in this and other environmental contexts is how to adjust indicator values for local conditions such as weather. The related environmental policy issue is whether and when to do so. Interdisciplinary issues include determining what information is lost or masked through data combination, resolving trade-offs between statistical and ecological properties of the construction of indicators and indexes, and developing data sources and methods for evaluating and recalibrating (**benchmarking**) indicators and indexes periodically.

3. COMBINING INFORMATION IN ENVIRONMENTAL EPIDEMIOLOGY

An important problem in combining environmental information is application of meta-analytic methods (Hedges and Olkin 1985; Wolf 1986) to epidemiology and environmental medicine. **Meta-analysis** is the rubric used to describe quantitative methods for combining evidence across studies (Hedges and Olkin 1985, p. 13).

The results of clinical or environmetric studies are often reported as **p-values** which measure the statistical significance of test results: the p-value is the probability of observing the test results or more extreme results under the **null hypothesis** (i.e., the hypothesis that the association or effect of interest is not present). A standard meta-analytic technique which is often very powerful (Koziol and Perlman 1978) is a simple method due to Fisher (1948) for combining p-values from independent studies. To combine the results of **K** independent studies that test the same hypothesis, the corresponding p-values $\mathbf{p_k}$ from the individual studies are combined to form:

$$-2\sum_{k=1}^{K} \ln(p_k)$$

which is known to exhibit a χ^2 distribution with 2K df; see Koziol and Perlman (1978) or Elston (1991). Fisher's method is also known as the **inverse χ^2 method**. Another method, the **inverse normal method**, involves the standard normal variate:

$$Z = \frac{\sum_{1}^{K} \Phi^{-1}(p_i)}{\sqrt{K}}$$

where $\Phi(z)$ denotes the cumulative distribution function along either tail of the standard normal distribution. The combined p-value is $p = \Phi(Z)$. Weighted versions of both methods are available; see Hedges and Olkin (1985) for details and a full discussion of meta-analytic methods. The formulation of the inverse normal method above illustrates the power of meta-analysis: among K (one-tailed) studies, significance in $\sqrt{K}$ studies is sufficient to ensure significance in the combined analysis.

In environmental applications, meta-analysis is often used for quantitative combination and synthesis of data over multiple studies on a specific endpoint. It is in increasing use in the biomedical sciences and clinical trials (Chalmers 1991), since it is uncommon for a single, well-designed biological or clinical experiment or study to evaluate completely hazardous materials and assess definitively population exposure risk(s). Rather, many small epidemiologic and biomedical studies are carried out on environmental stimuli under different conditions, on different populations, using different exposure regimens or different chemical metabolites, etc. These smaller studies provide limited amounts of information about environmental phenomena or effects, especially those that are more complex than perhaps originally perceived, or that cannot be examined fully under limited study designs or budgets. In some cases, the effects of interest are small and therefore hard to detect with limited sample sizes; or, data on too many different endpoints may mask or divert attention from small or highly-localized effects.

Methods such as meta-analysis for combining information are intended to synthesize disparate data into a single, conclusive statement about the phenomenon under study, from which policy decisions and other issues of scientific and social importance may be based. In this section, we discuss a number of environmental and biomedical studies where combining information has surfaced in important ways and has played an crucial role in the synthesis of understanding about an environmental/biomedical issue, and also note cases where more work needs to be done.

3.1. COMBINING EPIDEMIOLOGICAL STUDIES: ENVIRONMENTAL TOBACCO SMOKE

An important example of the potential of combined analyses concerns the recent US EPA study on health effects of environmental tobacco smoke (**ETS**) (US EPA 1992). Thirty epidemiologic studies from eight countries were considered as part of the EPA analysis. In all cases, the measured effect of interest was an increase in risk for lung cancer mortality over risk for non-exposed controls; the associated **relative risk** is simply the ratio of probability of exposure death to unexposed death: $Pr\{D|E_+\}/Pr\{D|E_-\}$. When the disease prevalence in the population is small, say, less than 5%, relative risk may be estimated via the **odds ratio**: the ratio of odds of exposure for cancer deaths (**cases**) to odds of exposure for **controls** (Breslow and Day 1980). Each set of odds simplifies to the number of exposed deaths--cases or control--divided by the number of corresponding unexposed deaths. The odds ratio is then simply "cancer odds" divided by "control odds." Since estimates of odds ratios and their large-sample variances are more easily and reliably calculated than relative risks, statistical methods are employed for estimating and testing relative risks via odds ratios (Breslow and Day 1980). When an odds ratio equals 1.0, no increased risk is evidenced. Various statistical analyses can identify when the odds ratio is significantly greater (less) than 1.0, corresponding to a significant increase (decrease) in risk (Breslow and day 1980; Wallenstein and Bodian 1987). More complex sampling scenarios that include, for example, studies of an entire cohort of individuals over an extended period, yield more complex estimators of the relative risk (Breslow and Day 1987). In the ETS study, these were employed whenever necessary.

The results from the 30 individual ETS studies yielded relative risk estimates ranging from 0.88 to above 2.0, giving contradictory information. Indeed, when analyzed separately, only one of the eleven U.S.-based studies showed a significant increase in the risk of lung cancer mortality after ETS exposure. Thus, at issue in the EPA analysis was whether proper combination of the individual estimated risks could identify an overall increased risk of lung cancer death due to ETS exposure. The data combination involved a simple averaging method to approximate exposure levels from the various forms of ETS, including sidestream smoke and exhaled mainstream smoke, and required construction of relative risk models that adjusted for background exposures and for potential systematic downward bias due to control group exposures to ETS. For example, to correct for background exposures, the adjusted relative risk was modelled as

$$RR_{adj} = \frac{\textbf{risk with ETS exposure}}{\textbf{risk without ETS exposure}} = 1 + z\beta d_N$$

where $\mathbf{z}$ is the ratio between the mean dose level in the exposed group and the mean dose level in the unexposed group, β is the increased risk per unit dose, and $\mathbf{d_N}$ is the mean dose level in the unexposed group. The estimates are then pooled, so that, on a logarithmic scale:

$$\log(\hat{RR}_{Pooled}) = \frac{\sum w_i \log(\hat{RR}_i)}{\sum w_i}$$

where the study weights are simply the inverse estimated variances of the log-relative risk estimates:

$$w_i = \frac{1}{var(\log(\hat{RR}_i)} \quad .$$

Exponentiating the log-RR estimate yields an estimate of the combined relative risk; see US EPA (1992) for greater detail, including some more complex models employed for the adjusted relative risk.

When completed, combination over all U.S. studies produced a statistically significant (p = .02) estimate of relative risk of 1.19. Adjusted for background exposures, the risk estimate rose to 1.59, i.e., an estimated 59% increase in lung cancer mortality in U.S. non-smokers when exposed to ETS. Extensions of these approaches, leading to more complex models of exposure risk, are under investigation by EPA for the potent toxin dioxin.

3.2. COMBINING EPIDEMIOLOGICAL STUDIES: NITROGEN DIOXIDE STUDIES

As suggested above, environmental epidemiologic studies can become fairly complex when considering human exposures to environmental stimuli. Another example of this feature involves a US EPA meta-analysis of respiratory damage after indoor exposure to nitrogen dioxide, NO_2. The study was undertaken in an attempt to integrate information on the relationship of NO_2 exposure to respiratory illness, since, in similar fashion to the ETS study, many previous studies had given mixed results on the health effects of NO_2 exposure.

Using as an outcome variable the presence of adverse lower respiratory symptoms in children aged 5 to 12 years, odds ratios were employed as estimates of increased relative risk in the exposed population. The meta-analysis combined a set of nine North American and western European studies, wherein estimated odds ratios for increased lower respiratory distress ranged from 0.75 to 1.49. Based on separate analyses, only four of the nine odds ratios suggested a significant increase in respiratory distress due to NO_2 exposure. Data combination and synthesis of these results involved standard meta-analytic techniques for odds ratio combination (Hasselblad 1994, Sec. 5): likelihood functions for each separate odds ratio are constructed, and, under an independence assumption, multiplied to yield a likelihood for the combined odds ratio, which, when maximized, yields an estimate of the combined odds ratio. The results showed a combined odds ratio of 1.17, with an associated 95% confidence interval from 1.11 to 1.23. That is, the meta-analysis suggested that an increase in NO_2 exposure can lead to an increased risk of respiratory illness of about 10-20% over unexposed controls. In this case, the meta-analysis gave good evidence of an important lower respiratory effect due to increased NO_2 exposure.

Further adjustments in the NO_2 study for additional sources of variability, such as socio-economic status, smoking and gender difference, led to similar values for the increased odds of respiratory distress. In these cases, measurement error concerns regarding proper correction for inaccurate or mis-measured exposures are important issues, and a number of associated statistical problems remain for further study; see Hasselblad, Eddy and Kotchmar (1992).

A specific alternative for combined data analyses that may be useful here involves **Bayesian methods**, where a prior probability distribution is incorporated into the analysis to help account for uncertainty in various unknown parameters (for a good introduction to Bayesian methods, see Box and Tiao 1973). Specific restrictions on these parameters can be incorporated into this hierarchy of distributions, making the Bayesian approach quite flexible. Combination of the prior distribution with the information in the data leads to calculation of a posterior distribution for the unknown parameters, in which belief in the value of the parameters is "updated" by the information in the data. Upper and lower percentiles from these posterior distributions can serve as fairly accurate confidence limits on the unknown parameters (e.g., confidence limits may be calculated for dose response parameters that represent effects of environmental stimuli such as NO_2 or other airborne toxins; see DuMouchel and Harris 1983). Applied to the nine studies in the US EPA NO_2 analysis, **hierarchical Bayes** computations similar to those performed by DuMouchel and Harris (1983) show that five of the nine studies exhibit significant increases in relative risk, yielding a pooled log-odds ratio with pooling weights $\mathbf{w}_i$ equal to the reciprocal posterior variances (see above), of approximately 0.1567. The corresponding odds ratio of $e^{0.1567} = 1.1697$ mimics the US EPA estimate of combined risk to NO_2 exposure.

Of course, a number of complementary statistical approaches are also available for modeling parameter heterogeneity or hierarchical effects, including **weighted linear regressions** (Carroll and Ruppert 1988) that mimic the data combination effect by viewing it as a heterogeneous variance setting, or **random effects models** (Rubin 1992) that incorporate differential effects of the data combination over multiple studies. Indeed, hierarchical Bayes models are quite similar in nature to random effects models for many common statistical models (see, e.g., Reinsel (1985), Schall (1991), and references therein).

3.3. DEVELOPING METHODOLOGY FOR ACUTE INHALATION ASSESSMENT

A third perspective on the need for new statistical formulations for combining data is available by considering the US EPA's more general development of methodology for acute inhalation risk assessment. This project involves combination of data from studies of inhalation damage from various airborne toxins in order to estimate human health risk. The studies vary greatly in their endpoints: short- and long-term exposures in laboratory animals, acute exposures to humans in chemical and/or community accidents, chronic exposure studies in urban areas, etc.

Current research in statistical combination of such data focuses on analysis of categorical data. The research goal is to develop methodology for data combination that incorporates the range of endpoint severity, exposure concentrations, and exposure durations. Particular emphasis is directed at acute exposures, since these are thought to be more common than chronic, long-term exposures in many human situations. The paradigm is based on **severity modeling**, wherein concentration, duration, and response are integrated to determine potential risks to humans after acute inhalation exposure to some environmental toxin. The method groups the response data into ranked severity categories, and assumes that duration and concentration are independent explanatory variables for predicting response. This is essentially an **ordinal**

regression, using a logistic or another discrete-data regression model for the concentration-duration-response (Greenland 1985; Tutz 1989). From the regression, one wishes to estimate the level at which an exposed subject will respond with a small probability, say 10%. This is the 10% **effective dose,** or **ED_{10}**, for which a lower bound is calculated, at, say, 95% confidence. For risk assessment, this lower bound is then divided by an arbitrary "safety factor" (Kaplan, Hoel, Portier and Hogan 1987; Johnson 1988). The resulting value is instituted as a regulatory upper limit on acute human inhalation exposure to the environmental stimulus.

Logistic regression models are useful for such constructions, as described, for example, by Simpson, et al. (1994): as a first approach, consider an ordinal regression model that accepts censored data when animal or human subjects die prior to study termination, models uncontrolled variables via random effects terms, stratifies the analysis to examine systematic differences across multiple studies, and combines the resulting information for ED_{10} estimation. Notice in particular that the stratification and random effects features provide familiar data analytic tools for studying the potential bias incurred by modelling error.

$\mathbf{Y}_{ij}$ denotes an ordered response in exposure group **j** for study **i** ($j=1,\ldots,n_i$; $i=1,\ldots,M$). If $Y_{ij} = 0$ indicates no toxic response, a logit-linear model is

$$Pr[y_{ij}\geq 1 \mid b_i,\ x_{ij},\ z_{ij},\ u_{ij}] = \frac{1}{[1+\exp\{-(\alpha+\beta' x_{ij}+\gamma' z_{ij}+b_i' u_{ij}\}]}$$

where $\mathbf{x}_{ij}$ are study/exposure-specific concentration and duration variables, $\mathbf{z}_{ij}$ are study/exposure-specific covariates and/or stratum indicators, and random effects are modelled via study/exposure-specific controlled values in $\mathbf{u}_{ij}$ and via study-specific parameters $\mathbf{b}_i$: $b_i \sim N(0,\ \sigma^2 I)$. The simplest random effects case is where u_{ij} = 1 for all **j**, and where $b_i \sim N(0,\ \sigma^2)$ (i.e., a stratified, random effects, ordinal logistic regression). If, for example, M = 1 and $\mathbf{x}_1$ denotes $\log_{10}$(concentration), then ED_{100q} estimation is achieved by solving $q = Pr[Y\geq 1 \mid x_1,\ z]$ for $\mathbf{x}_1$ after integrating out the random effect (i.e., a marginal ED_{100q}), if necessary. Other functional forms and constructions are possible, and further development of these models is an important area of continuing environmetric research.

3.4. SYNTHESIS OF ENVIRONMENTAL STUDIES: LEAD ABATEMENT

There are, of course, settings where combining information may still not achieve data synthesis, and where more data collection and finer model determination is required to yield powerful statistical inferences. An interesting example occurs with studies of three U.S. cities' (Boston, Baltimore, and Cincinnati) attempts to reduce childhood exposure to lead. The studies are focused on intervention efforts to reduce lead exposure and presumed consequent high blood lead levels in at-risk, inner-city children.

Individually, each city exhibited a different response to its lead abatement efforts: Boston reported slight declines in potential for environmental lead exposures, Baltimore reported no significant impact, and Cincinnati reported mixed results. Could data combination be used here to improve the sensitivity and power of the statistical analyses?

To estimate changes in blood levels during the abatement period, and to incorporate site differences encountered as each city addressed and implemented its abatement strategy, city-specific **structural equations** (Austin and Wolfle 1991) were modelled to account for different lead pathways into the bloodstream. These equations modelled multiple metabolic body pathways and interactions, in an attempt to mimic the processes and metabolic activities lead encounters as it passes through the body. A meta-analysis then combined outcomes from hypothesis tests of whether there was a significant reduction in post-abatement blood lead levels based on the structural equations. Unfortunately, the combined results suggested only minimal reduction in exterior lead exposure and consequent lowering of blood lead levels. There was, however, evidence that some relationship existed between exterior and interior lead dust and blood lead levels, and that this relationship--approximately a 1 μg/dL drop in blood lead per 1000 ppm lead dust decrease--was essentially uniform across all three study sites. This suggested that further efforts in lead abatement may achieve significant reductions in blood lead levels, but that the current data were unable to identify significant improvements in lead abatement.

The data combination highlighted problems in meta-analyses of this sort: unwelcome sensitivities to structural equation specifications and difficulty in developing proper lead pathway models were encountered, possibly due to unadjusted heterogeneity in measurements of important predictor variables; the structural equation models relied upon highly non-normal random variable assumptions and statistical error structures that made analysis difficult; and, multivariate analyses to account for repeated measures on study subjects (Diggle and Donnelly 1989; Carr and Chi 1992) were necessary to adjust for seasonal cycles and long-term time trends in blood lead concentrations. In effect, the complexity of the different models overwhelmed any gains in sensitivity that data combination was able to provide.

4. OTHER APPLICATIONS

4.1 SITE CHARACTERIZATION

Meta-analytic methods such as the Fisher method (Sec. 3) have been used to combine environmental information from samples taken at different locations within a single site, such as a hazardous waste site. Data collected at the individual locations is combined to identify whether clean-up of the site has been successful, or if more clean-up effort is required, etc. An extension of the Fisher method is useful under conditions where the samples are of material having multiple (and correlated) toxic features. For K = 2, this involves computing the two-sample correlation coefficient **R**. Comparison of R with a known reference distribution yields a corresponding p-value $\mathbf{p_R}$. The test statistic $\mathbf{-2(\ln(p_1) + \ln(p_2) + \ln(p_R))}$ may be compared to a χ^2 reference distribution with 6 df, and yields a test that, based on simulations, is more powerful than the Fisher test with K = 2. See Mathew, Sinha and Zhou (1993) for details.

4.2 BAYESIAN APPLICATION TO DOSE RESPONSE IN NON-CANCER TOXICITY

In other instances, such as when different studies provide data on a non-cancer dose-response to an environmental stimulus in different organisms, Bayesian methods can serve as useful vehicles for data synthesis, via their ability to represent comparable information on unknown parameters in posterior distributions. A posterior distribution on, say, a toxicity dose-response parameter in rats after exposure to a toxin can be compared directly with a posterior on an ED_{100q} parameter for toxicity in human cells after exposure to the same toxin, in order to estimate safe dose levels for occupational or environmental exposure to the toxin.

Jarabek and Hasselblad (1991) provide an example of these sorts of calculations for the airborne neurotoxin **n-hexane**. They combine cross-sectional epidemiologic cohort study data (Sanagi et al. 1980) with sub-chronic laboratory rodent toxicity data (Dunnick et al. 1990), in similar fashion to other studies using both field and laboratory data (Wolpert and Warren-Hicks 1992). As part of their analysis, Jarabek and Hasselblad show that posterior distributions for the concentration of the toxin that produce a detrimental health effect (such as an ED_{10} for nasal issue damage) could be combined into a single posterior distribution for a risk-free concentration (**RfC**) exposure level, yielding an estimated RfC of 0.2 mg/m^3. Their use of a posterior distribution offers the advantage of a visual display for the distribution of the RfC, along with incorporation of uncertainty and safety factors in calculating the combined risk estimate. In general, posterior distributions and associated hierarchical models (Morris and Normand 1992) may be combined in fairly straightforward manners, allowing for more flexible data synthesis, visualization, and interpretation of toxic effects after environmental exposures (Hasselblad 1994). Graphical displays of dose responses, confidence limits, etc. are possible using these methods, providing useful illustrations of statistical uncertainty, such as in an estimated ED_{100q} and its associated population risk estimates.

4.3 DATA REPORTING AND CONFORMANCE

Other important applications of statistical methodology in environmental science include hazard identification and how one reports statistical results from field studies. Combining information from multiple field studies becomes especially important when results from previous studies are employed to determine if or how further study of an environmental hazard is to be undertaken. Proper statistical reporting can obviate the need for further studies, saving scientific and taxpayer resources for other projects. In this area, different agencies, countries, etc., may have different definitions and ideals for similar data structures. Data conformance, particularly at the international level, becomes an important consideration. Development of such conformance is now ongoing, with the goal of setting unambiguous standards for naming, defining, and documenting data elements. To date, however, the fundamental principles of international data conformance are barely in place. Guidance is called for from subject-matter scientists working in appropriate environmental areas, especially regarding issues such as quality assurance to check on protocol compliance, and to enhance basic data quality.

Similar issues are critical in statistical reporting of environmental data, particularly when the data are combined. In reporting combined environmental data, basic needs include emphasis on standard forms for presenting and displaying data, and on guidelines for recognizing and analyzing spatial effects and identifying correct statistical features of spatial variability. As above, further collaborative research is necessary to establish unambiguous, easily-reported standards for these environmental reporting issues.

5. COMBINING ENVIRONMENTAL INFORMATION: A RICH AREA OF MULTI-DISCIPLINARY RESEARCH

This exposition has provided summaries and discussion of some current examples and issues in combining environmental information, and suggested potential research directions. Applications and improvements in associated areas of statistical research include survey sampling, geostatistics, biological effect modeling, use of hierarchical Bayes models, and statistical index theory.

There are two areas where new statistical theory will make important contributions to ecological monitoring and assessment. The first is that of combining NP-sample data with P-sample data and other NP-sample data. P-sample based programs such as EMAP provide a rigorous framework to study and assess ecological resources, the value of which can be magnified enormously in the presence of methods for combining P-sample based data such as EMAP data with data from other sources. An important research question is how to use NP-data to improve the efficiency of P-sample designs and P-sample based estimation strategies. The need for these and related methods extends to other environmental areas, such as Superfund: for example, to design multi-stage samples for site characterization and remediation (Englund and Heravi 1994), or to draw combined inferences from preliminary (NP-sample) site measurements and formal (P-sample) site clean-up monitoring data. Similarly, new and extended results in spatial statistics, particularly on multivariate and robust kriging, would have immediate application to both ecological and compliance/remediation monitoring problems.

The second ecological monitoring and assessment area is that of developing environmental indicators and indexes that are meaningful in terms of ecological and environmental science but are rigorous statistically. The issues requiring attention here are interdisciplinary and involve statistics. They include: defining appropriate reference populations and systems, determining appropriate units of measure, adjusting for local conditions such as climate, avoiding spurious correlations caused by inappropriate or over-aggregation, calibration/recalibration to monitor operating characteristics of statistical measures, and developing and using baseline data to benchmark indicators and indexes periodically.

In the areas of environmental epidemiology and biomedical research, combining information has an important role to play in terms of synthesizing disparate conclusions from studies of weak or hard-to-identify effects. Particularly in cases where sample sizes are too low to identify subtle or minimal effects, the application of meta-analytic methods or other forms of combining information can synthesize equivocal results, and lead to proper identification of true environmental effects. Routine use by researchers of methods to examine the **sensitivity** of meta-analyses to individual studies (e.g., systematic removal of studies of concern, performing meta-analyses on the reduced sets of studies, and comparing the results) and the development of new sensitivity analysis techniques for meta-analysis would be beneficial. In all these areas, more research and investigation is necessary to recognize when, where, and how information combination can improve upon standard and perhaps outmoded forms of data analysis and scientific interpretation of study results.

ACKNOWLEDGEMENTS

The authors are indebted to the 75 participants at the Workshop on Statistical Methods for Combining Environmental Information, and especially to the presenters and discussion leaders without whom this synthesis of this important topic would not have been possible. The presenters and discussion leaders are: B. Bargmeyer, S. Bayard, D. Carr, R. Carroll, N. Cressie, W. DuMouchel, D. Gaver, D. Guth, V. Hasselblad, D. Hoag, E. Hyatt, A. Jarabek, D. Kotchmar, R. Linthurst, A. Marcus, B. Nussbaum, A. Olsen, J. Rawlings, K. Reckhow, R. Shafer, C. Shoaf, D. Simpson, and B. Sinha. Special thanks go to discussion leader I. Olkin.

This paper is intended to be a synthesis and examination of the issues raised at the Workshop, and not a proceedings, and the authors are solely responsible for any errors of omission, commission or emphasis. The opinions expressed are those of the authors and are not intended to represent the policies or practices of the US EPA or NISS.

The authors thank the U.S. Environmental Protection Agency and the National Institute of Statistical Sciences for organizing the Workshop; partial support was provided under US EPA Cooperative Agreement CR819638 with NISS. Thanks go to D. Holland, project officer of the US EPA/NISS cooperative agreement. Partial support by NISS of Piegorsch in developing this paper is acknowledged. Special thanks go to J. Sacks for his efforts in co-organizing the Workshop and for his constructive suggestions in the development of this paper.

REFERENCES

Austin, J.T. and L.M. Wolfle (1991). "Annotated bibliography of structural equation modeling: Technical work". *British Journal of Mathematical and Statistical Psychology* 44, 93-152.

Box, G.E.P. and G.C. Tiao (1973). **Bayesian Inference in Statistical Analysis**. Addison-Wesley, Reading, Mass.

Breslow, N. and N. Day (1980). **Statistical Methods in Cancer Research. I. The Analysis of Case-Control Studies**, Vol. 32. IARC Scientific Publications, Lyon, France.

Breslow, N. and N. Day (1987). **Statistical Methods in Cancer Research. II. The Analysis of Cohort Studies**, Vol. 82. IARC Scientific Publications, Lyon, France.

Carr, G.J. and E.M. Chi (1992). "Analysis of variance for repeated measures data: A generalized estimating equations approach". *Statistics in Medicine* 11, 1033-1040.

Carroll, R.J. and D. Ruppert (1988). **Transformation and Weighting in Regression**. Chapman and Hall, New York.

Carroll, S.S., G.N. Day, N.A.C. Cressie and T.R. Carroll (1994). "Spatial modeling of snow water equivalent using airborne and ground-based snow data". *Environmetrics*, to appear.

Chalmers, T.C. (1991). "Problems induced by meta-analysis". *Statistics in Medicine* 10, 971-980.

Cox, L.H. and R.F. Boruch (1988). "Record linkage, privacy and statistical policy". *Journal of Official Statistics* 4, 3-16.

Cressie, N.A.C. (1990). "The origins of kriging". *Mathematical Geology* 22, 239-252.

Cressie, N.A.C. (1991). **Statistics for Spatial Data**. Wiley, New York.

Cressie, N.A.C. and J.M. var Hoef (1993). "Spatial analysis of environmental and ecological data". In **Environmental Modeling with GIS**. eds: M.F. Goodchild, B.O. Parks and L.T. Steyaert, Oxford Univ. Press, New York.

Diggle, P.J. and J.B. Donnelly (1989). "A selected bibliography on the analysis of repeated measurements and related areas". *Australian Journal of Statistics* 31, 183-193.

DuMouchel, W.M. and J.E. Harris (1983). "Bayes methods for combining the results of cancer studies in humans and other species" (with discussion). *Journal of the American Statistical Association* 77, 293-313.

Dunnick, J.K., D.G. Graham, R.S. Yang, S.B. Haber and H.R. Brown (1990). "Thirteen-week toxicity study of n-hexane in $B_6C_3F_1$ mice after inhalation exposure". *Toxicology* 57, 163-172.

Elston, R.C. (1991). "On Fisher's method of combining p-values". *Biometrical Journal* 33, 339-345.

Englund, E.J. and N. Heravi (1994). "Multi-Phase Sampling for Soil Remediation". Manuscript: US EPA, Environmental Monitoring systems Laboratory, Las Vegas, 16pp.

Fisher, I. (1922). **The Making of Index Numbers: A Study of Their Varieties, Tests, and Reliability**. Houghton Mifflin, Boston.

Fisher, R.A. (1935). **The Design of Experiments**. Oliver & Boyd, New York.

Fisher, R.A. (1948). "Combining independent tests of significance". *American Statistician* 2, 30.

Greenland, S. and A. Salvan (1990). "Bias in the one-step method for pooling study results". *Statistics in Medicine* 9, 247-252.

Hartley, H.O. (1974). "Multiple frame methodologies and selected applications". *Sankhya*, series B 36, 99-118.

Hasselblad, V. (1994). "Meta-analysis in environmental statistics". In *Handbook of Statistics Volume 12: Environmental Statistics*, eds G.P. Patil and C.R. Rao, North Holland/Elsevier, New York, in press.

Hasselblad, V., D.M. Eddy and D.J. Kotchmar (1992). "Synthesis of environmental evidence: nitrogen dioxide epidemiology studies". *Journal of the Air & Waste Management Association* 42, 662-671.

Hedges, L.V. and I. Olkin (1985). **Statistical Methods for Meta-Analysis**. Academic Press, Orlando.

Hunt, W.F. (1991). "An examination of alternative air quality indicators to better understand global pollution". Workshop on Air Quality Guidelines for Air Pollution Control Strategies in Western and Northern Europe, Paris, July 2-5, 1991.

International Statistical Institute (ISI) (1956). **Bibliography on Index Numbers**. Mouton, The Hague.

Jarabek, A. and V. Hasselblad (1991). "Inhalation reference concentration methodology: Impact of dosimetric adjustments and future directions using the confidence profile method". From: 84th Annual Meeting of the Air and Waste Management Association, Vancouver, Canada.

Johnson, E.M. (1988). "Cross-species extrapolations and the biological basis for safety factor determinations in developmental toxicology". *Regulatory Toxicology and Pharmacology* 8, 22-36.

Kaplan, N., D.G. Hoel, C.J. Portier and M.D. Hogan (1987). "An evaluation of the safety factor approach in risk assessment". In *Developmental Toxicity: Mechanisms and Risks*, eds J.A. McLachlan, R.M. Pratt and C.L. Markert, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 335-344.

Koziol, J.A. and M.D. Perlman (1978). "Combining independent chi-squared tests". *Journal of the American Statistical Association* 73, 753-763.

Lambert, D., B. Peterson and I. Terpenning (1991). "Nondetects, detection limits, and the probability of detection". *Journal of the American Statistical Association*, 86, 266-277.

Mastala, Z., K.V. Balogh and J. Salanki (1992). "Reliability of heavy metal pollution monitoring utilizing aquatic animals versus statistical evaluation methods". *Archives of Environmental Contamination and Toxicology* 23, 476-483.

Mathew, T., B.K. Sinha and L. Zhou (1993). "Some statistical procedures for combining independent tests". *Journal of the American Statistical Association* 88, 912-919.

Morris, C.N. and S.L. Normand (1992). "Hierarchical models for combining information and for meta-analysis". In **Bayesian Statistics 4**, Bernardo, J.M., J.O. Berger, A.P. Dawid and A.F.M. Smith (eds), 321-344. Amsterdam: Elsevier Science Publishers B.V. (North Holland).

Myers, D.E. (1991). "Interpolation and estimation with spatially located data". *Chemometrics and Intelligent Laboratory Systems* 11, 209-228.

Narula, S.C. and J.F. Wellington (1982). "The minimum sum of absolute errors regression: A state of the art survey". *International Statistical Review* 50, 317-326.

National Academy of Sciences (NAS) (1975). **Planning for Environmental Indices**. National Academy Press, Washington, DC.

Ott, W.R. (1978). **Environmental Indices Theory and Practice**. Ann Arbor Science Publishers, Ann Arbor.

Overton, J.M., T.C. Young and W.S. Overton (1993). "Using found data to augment a probability sample: procedure and a case study". *Environmental Monitoring and Assessment* 26, 65-83.

Overton, W.S. (1990). "A strategy for use of found samples in a rigorous monitoring design". Technical Report number 139. Department of Statistics, Oregon State University, Corvallis.

Reinsel, G.C. (1985). "Mean squared error properties of empirical Bayes estimators in a multivariate random effects general linear model". *Journal of the American Statistical Association* 80, 642-650.

Rodgers, W.L. (1984). "An evaluation of statistical matching". *Journal of Business & Economic Statistics* 2, 91-102.

Rubin, D.B. (1992). "Computational aspects of analyzing random effects/longitudinal models". *Statistics in Medicine* 11, 1809-1821.

Sangi, S., Y. Seki, K. Sugimoto and M. Hirata (1980). "Peripheral nervous system functions of workers exposed to n-hexane at a low level". *International Archives of Occupational Health* 47, 69-70.

Schall, R. (1991). "Estimation in generalized linear models with random effects". *Biometrika* 78, 719-727.

Simpson, D.G., R.J. Carroll, D. Guth and H. Zhou (1994). "Using information from disparate studies in toxicological risk assessment". Technical Report. University of Illinois, Champaign-Urbana.

Stevens, D.L. (1994). "Implementation of a national monitoring program". *Journal of Environmental Management*, to appear.

Tutz, G. (1989). "Compound regression models for ordered categorical data". *Biometrical Journal* 31, 259-272.

US EPA (1992). **Respiratory Health Effects of Passive Smoking: Lung Cancer and Other Disorders**. US EPA/600/6-90/006F, Washington, DC, December 1992.

Wallenstein, S. and C. Bodian (1987). "Inferences on odds ratios, relative risks, and risk differences based on standard regression programs". *American Journal of Epidemiology* 126, 346-355.

Wolf, F.M. (1986). **Meta-Analysis: Quantitative Methods for Research Synthesis**, Vol. 59. Sage Publications, Newbury Park.

Wolpert, R.L. and W.J. Warren-Hicks (1992). "Bayesian hierarchical logistic models for combining field and laboratory data". In **Bayesian Statistics 4**, Bernardo, J.M., J.O. Berger, A.P. Dawid and A.F.M. Smith (eds.), 525-546. Amsterdam: Elsevier Science Publishers B.V. (North Holland).