NISS

Statistical Advances in Environmental Science

Walter W. Piegorsch, Eric P. Smith, Don Edwards,
and Richard L. Smith

Technical Report Number 73
November, 1997

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

Statistical Advances in Environmental Science

Walter W. Piegorsch, Eric P. Smith, Don Edwards, and Richard L. Smith

Abstract. We discuss selected applications of statistical theory and practice as motivated by and applied to environmental sciences. Included in the presentation are illustrations on how the interaction between environmental scientists and quantitative researchers has been used to enhance and further learning in both areas, and how this interaction provides a source of further challenges and growth for the statistical community.

Key words and Phrases: Biometry, Combining environmental information, Detection limits, Ecological statistics, Environmetrics, Environmental health, Environmental statistics, Exceedance over threshold, Extreme value analysis, Hierarchical Bayesian methods, Hot spots, Inter-disciplinary collaboration, Meta-analysis, Model selection, Quantitative risk assessment, Risk Analysis, Rare events, Space/time sampling, Spatial statistics, Stochastic modeling, Time series, Trend analysis.

1. INTRODUCTION

"Measuring the environment is an awesome challenge, there are so many things to measure, and at so many different times and places." (Hunter, 1994, p.6)

As a discipline, statistics has been affected greatly by the other sciences with which it interacts. In the environmental sciences, this is particularly true. Links to problems in the atmospheric, ecological, geological, toxicological, biomedical, and economic sciences, and concerns in public health, risk management, and social policy, have provided rich data for quantitative analyses, collectively called *environmetrics*. Advances in these subject-matter sciences have produced data analytic challenges that often motivate new statistical developments (Olkin *et al.*, 1990). In turn, these developments prompt the subject-matter researchers to re-evaluate their design and analytic goals, leading to further data gathering. If communication is good between the disciplines, the cycle repeats, with each scientific discipline growing and improving. Rather than view this as simply a circular interaction, we prefer to call it a spiral, an *upward spiral*, where science and society continue to benefit as the disciplines interact.

Walter W. Piegorsch and Don Edwards are Professors, Department of Statistics, University of South Carolina, Columbia, South Carolina 29208. Eric P. Smith is Professor, Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061-0439. Richard L. Smith is Professor, Department of Statistics, University of North Carolina, Chapel Hill, North Carolina 27599-3260.

These events are perhaps best illustrated by the number of both edited and sole-authored books appearing in the past few years that deal with quantitative methods and applications in the environmental sciences; examples include Cressie (1993), Günther (1997), Helsel and Hirsch (1992), Hewitt (1992), Ott (1995), Pearson and Turton (1993), Rivoirard (1994), and Scheiner and Gurevitch (1993). Herein, we survey recent progress in the field, with particular reference to a series of edited compilations on environmetrics that have appeared since 1992: Walden and Guttorp (1992), Barnett and Turkman (1993; 1994; 1997), Cothern and Ross (1994), and Patil and Rao (1994). We use examples from these collections and many other fine works to help illustrate modern interactions between the statistical and environmental sciences. These include studies of exceedance modeling, atmospheric pollution and mortality analysis, space-time modeling of acid rain, trend analysis, ecological monitoring and assessment, low-dose risk extrapolation, and environmental effects on animal populations. The examples are intended not only to illustrate important advances in modern environmental statistics, but also to emphasize that many important problems in these areas remain unsolved.

Before continuing, we acknowledge that it is impossible to survey all the excellent theoretical and applied environmetric research presented in the sources noted above. Indeed, we were struck by the breadth of environmetric problems illustrated in these works, and by how different editors could portray so successfully a wide variety of viewpoints on modern environmental statistics. Within this context, our goal will be to highlight a selection of subject-matter problems (noted above), and to accent the diversities of views that have motivated statistical/environmetric advances for them.

2. Extreme values and rare events, with an application in modeling atmospheric ozone

Many environmental problems concern extreme values of some measured environmental variable. Traditional examples for which there is a long history of statistical work include floods and sea levels; see, e.g., Tawn (1993). For example, one might seek to build a sea wall to withstand a "N-year flood," i.e., a water level which is exceeded in any one year with probability $1/N$, where N is some large number such as 50, 100 or 250. Associated applications occur when

studying reservoir levels (Anderson and Nadarajah, 1993) and rainfall data (Buishand, 1993). Other, modern applications include:

- *Extreme levels of tropospheric ozone.* Ozone standards are based typically on the number of exceedances of some measure of ozone severity (e.g., hourly maximum value or eight-hour average) about some specified threshold such as 120 or 80 parts per billion (ppb). A question of statistical interest involves monitoring whether the rate of exceedance of this level is increasing or decreasing with time. This could have a completely different answer from the question of whether the mean ozone level is increasing or decreasing with time, but it is widely recognized that the frequency of extreme ozone events is a more relevant indicator of health effects than the average level of daily ozone.
- *Extremes in meteorology.* Apart from the intrinsic interest in extreme weather events, there is also a wide interest in whether projected global warming will affect extreme (high or low) temperatures and rainfalls to a greater or lesser extent than the mean levels. This again leads to consideration of trends in extreme values.
- *Insurance.* Part of the concern over environmental catastrophes such as Hurricane Andrew or Mississippi River flooding has to do with the effect of very large claims on the financial solvency of insurance companies. This has focused attention on the statistical properties of insurance claims to environmentally-caused damage.

The traditional method of extreme value analysis popularized by Gumbel (1958) was the *annual maximum* method, in which one of the three classical types of extreme value distributions was fitted to, say, the annual maxima of a river or sea level series. For many modern environmental applications, however, traditional methods are too restrictive. For example, a typical ozone data series consists of between 10 or 15 years of daily readings, and to base the statistical analysis on annual maxima, or even monthly maxima, wastes valuable data. Moreover, many of the issues associated with ozone have to do with the effects of daily meteorological variation on the observed ozone levels, and this cannot be captured by studying only maxima over extended time periods.

The second major approach to extreme value analysis is based on exceedances over a high threshold. Originally popularized by hydrologists in the 1970s as the *peaks over threshold* (POT)

method, this was subsequently refined to deal with covariates and trends in the data, and also to cope with time-series dependence (Davison and Smith, 1990; Gomes, 1993). A variant is to model the two-dimensional point process of exceedance times and excess values over a high threshold, using probabilistic characterizations of the limiting process. Statistical inferences based on this approach were developed by Smith (1989), with an application to trends in tropospheric ozone. Even this refinement does not cope very well with dependence in the data, however. For instance, time-series dependence affecting the joint distribution of neighboring values in a single series, or the dependence between different series such as wind and wave extremes, can introduce dependencies that exceed the capabilities of these models. In recent years a number of further improvements have been developed, based on the joint distributions of dependent extremes (Coles and Tawn, 1994; Smith *et al.*, 1997b), but much additional work is needed.

To go into more details about these techniques, consider the simplest form of a threshold problem: observe a sequence $Y_1, Y_2, \ldots$ of independent identically distributed (i.i.d.) data sampled from some unknown distribution function F(y), and center interest on the upper tail behavior of F. Let $\omega_F = \sup\{y: F(y)<1\}$. (Notice that ω_F can be infinite; the resulting theory holds whether the endpoint of the distribution is finite or infinite.) An argument originally given by Pickands (1975) suggests that exceedances over a high threshold u can be modeled approximately by the *generalized Pareto distribution* (GPD)

$$\Pr[Y_1 \le u + x \mid Y_1 > u] \approx 1 - \left(1 + \frac{\xi x}{\sigma}\right)_+^{-1/\xi}, \qquad x \ge 0,$$

for suitable values of σ (depending on u) and ξ. The simplest form of threshold-based analysis fits the GPD to all exceedances of some high level u via, e.g., numerical maximum likelihood (Davison and Smith, 1990). An important practical question in this case is selection of the threshold u. This has been the focus of much theoretical discussion over the years — see, e.g., Smith (1987) — but in practice the selection is usually handled by a mixture of graphical methods of assessing the GPD fit and *ad hoc* judgment (Davison and Smith, 1990).

One practical difficulty in applying this to, say, daily river levels, is that such series are always correlated. The traditional method of dealing with correlation identifies clusters of dependent

extreme values, and applies the GPD analysis to the cluster maxima or peaks, rather than to all the threshold exceedances (hence the name, peaks over threshold). Asymptotic theory of extremes in stationary sequences (Leadbetter *et al.*, 1983; Hsing *et al.*, 1988) justifies this approach by showing that high-level exceedances do tend to form clusters and that the cluster maxima also follow a GPD. An additional parameter, important for the asymptotic distribution of extreme values from a stationary sequence, is the *extremal index* θ (Leadbetter, 1983), which is most simply defined by setting θ^{-1} as the limiting mean value of the number of exceedances per cluster. Thus if one can identify appropriate clusters, the POT method is justified by asymptotic theory. Difficulties with this approach do exist, however. For instance, it is unclear how to identify the clusters; see Smith and Weissman (1994) for a discussion of this problem in connection with estimating the extremal index. Or, often one needs to know the joint distribution of extreme values within a cluster rather than just the distribution of the cluster maximum. For example, one such quantity is the cumulative excess by all exceedances over the threshold (Anderson, 1994).

A second major concern in POT modeling is the treatment of covariates. Suppose Y_i is the observation taken on day i, and let the associated GPD parameters be denoted by σ_i and ξ_i. In principle, one could let σ_i and ξ_i depend in quite general ways on covariates, but a convenient representation in practice is to assume ξ_i is some constant value ξ, and that

$$\log(\sigma_i) = \sum_{j=1}^{p} x_{ij}\gamma_j$$

where $x_{i1},\ldots,x_{ip}$ are measured values of p covariates and $\gamma_1,\ldots,\gamma_p$ are unknown parameters to be estimated.

A complete model must also take account of the frequency of crossing the threshold. One approach, which in effect assumes we are observing the process in continuous time, models the point process of times when the threshold is exceeded by a non-homogeneous Poisson process of intensity λ_t (here, t is time); $\log(\lambda_t)$ may depend linearly on covariates. An alternative discrete-time approach is to let ϕ_i denote the probability of crossing the threshold on day i and to apply a logistic regression model:

$$\log\left\{\frac{\phi_i}{1-\phi_i}\right\} = \sum_{j=1}^{p} x_{ij}\beta_j,$$

where $\beta_1,...,\beta_p$ is another set of unknown parameters. In practice, some of $\beta_1,...,\beta_p$ or $\gamma_1,...,\gamma_p$ may be offset to zero, thus allowing for different covariates in the models for σ_i and ϕ_i. All these parameters may be estimated by joint maximum likelihood.

An application of these methods is to tropospheric ozone measurements. Interest here is in determining whether the rate of threshold exceedances has increased or decreased in response to government regulatory policies, correcting for the confounding effects of meteorology. One can apply the models above using both time and meteorological variables as covariates, and separate studies by Smith and Huang (1994) and by Smith and Shively (1995) have indicated a clear decrease in ozone exceedances when meteorological effects are removed. This mirrors an earlier conclusion by Smith (1989) based on similar statistical methodology, though without taking account of meteorology.

In the case of ozone it is widely assumed that day-to-day values are conditionally independent given the meteorology. This is based on physical models for ozone formation which show that ozone drops to very low levels during the night, and re-forms from fresh emissions the next day. Unfortunately, it is unclear if the conditional independence assumption is valid here, and methods have been devised based instead on bivariate extreme value theory. Classical bivariate extreme value theory is concerned with asymptotic expressions for the joint distribution of maxima from two dependent random variables (Resnick, 1987). By analogy with the univariate theory, it is possible to develop threshold models for extreme values from bivariate distributions; e.g., use bivariate extreme value theory to characterize the joint distribution of successive values in a time series (Smith *et al.*, 1997b). Via such a model, Smith and Huang (1994) found that ozone values on successive days were in fact *dependent* even after adjusting for meteorology, and this has a significant effect on the goodness of fit of the model for tail probabilities. (We give some other illustrations of environmental modeling with ozone data in §6, below.)

The first two volumes by Barnett and Turkman (1993; 1994) include many examples of extreme value methodology applied in many different areas: hydrology (Anderson and Nadarajah, 1993; Gomes, 1993; Klemes, 1994), rainfall modeling (Buishand, 1993; Coles, 1994; Reed and Stewart, 1994), air pollution (Lindgren *et al.*, 1993), and sea levels (Tawn, 1993; Dixon and

Tawn, 1994; Tawn *et al.*, 1994; Vrijling, 1994). These cover many more techniques than have been reviewed here, and they provide an excellent overview of the broad range of modern theory and applications of this methodology.

3. ENVIRONMENTAL EPIDEMIOLOGY AND THE DEBATE OVER PARTICULATE MATTER

A major concern in modern environmental health science is the study of associations between environmental pollutants and adverse health outcomes. As might be expected, there are many statistical difficulties associated with detection of such associations, and in this section we review some of these. For illustrative purposes, we also direct attention to a particular area of current controversy: the health effects of airborne particulate matter.

3.1 Disease clustering

Elliott *et al.* (1995) reviewed some general issues concerned with making inferences about adverse health effects from observed associations. One of the earliest successes of environmental statistics was of this form: John Snow's studies of cholera in London in 1854 led him to suggest that the cause of the problem was a particular water pump, years before there was an established medical link between cholera and infected water. Elliott *et al.* remarked, however, that there are very few modern instances where an observed "cluster" of disease cases has been followed by the identification of a causal mechanism associating the cluster with a particular environmental health hazard. They cite many reasons why such studies are difficult, including

- availability of data — often only aggregated health effects data are available whereas a precise analysis would require individual data;
- difficulties of measuring a population in small regions;
- migration, i.e., when individuals who become infected in one location move to another;
- confounding, e.g., by socioeconomic status; and
- the *post hoc* nature of many of the studies.

Diggle (1990) and Diggle and Rowlingson (1994) have considered some of the more methodological issues associated with detecting clusters in spatial point processes of disease. Diggle *et al.* (1997) extended the approach to the modeling of spatially aggregated data. Earlier

Stone (1988) proposed a nonparametric test for identifying disease clusters which has been widely cited.

In monitoring the effects of very widely spread pollutants, such as ozone or sulfur dioxide in the atmosphere, it is usually impossible to associate the effect with specific sources. Hence, procedures based on cluster detection are not appropriate. In such cases it is generally assumed that comparisons of mortality or disease incidence with levels of pollution *across* different spatial regions is subject to so much confounding with other environmental effects that no meaningful conclusions can be drawn. Nevertheless some of the key studies currently cited in support of the adverse health effects of particulate matter are of precisely this form; we return to this point below.

In a wide-ranging review, Zidek (1997) cited these confounding difficulties as the reason for concentrating on *longitudinal* studies — i.e., studies based on detecting associations in temporal fluctuations of both the health effects and pollutant data at a single site. Nevertheless, the information available at any one location is usually limited so there is still a need for statistical methods of combining data from different locations (cf. §10).

Zidek formally modeled the errors in variables problem in terms of a triple $(Y,\mathbf{X},\mathbf{X}_g)$ where Y is a set of outcome variables, $\mathbf{X}$ the set of "true" covariates, and $\mathbf{X}_g$ a set of measured or "gauged" covariates. Zidek highlighted the well-known result that if Y is regressed linearly on $\mathbf{X}_g$ rather than $\mathbf{X}$, the regression coefficient is underestimated by an amount corresponding to the linear regression coefficient of $\mathbf{X}_g$ on $\mathbf{X}$. Rather than try to apply a correction factor based on the latter regression coefficient, Zidek advocated regressing Y on $E[\mathbf{X} \mid \mathbf{X}_g]$. The distinction between the two methods is especially relevant when extended to nonlinear regression functions, such as $E[Y \mid \mathbf{X}] = e^{\beta'\mathbf{X}}$. Written as $\log\{E[Y \mid \mathbf{X}]\} = \beta'\mathbf{X}$, this *log-linear model* is very commonly adopted for count data in Poisson regression.

In pursuit of concrete statistical methodology based on these ideas, Zidek assumed outcome measures $\{Y_{kt}\}$ where, for instance, Y_{kt} is the number of deaths in the kth region or location on day t. Each Y_{kt} depends on a vector of covariates $\mathbf{X}_{kt}$, some components of which may be measured in error. The latter feature is captured by assuming certain functions for the means and

covariances of $\mathbf{X}_{kt}$, where components which are measured exactly have all variances and covariances equal to zero. Random location effects also are allowed.

Based on these associations, Zidek was able to compute approximate means and covariances of Y, conditional on the observed $\mathbf{X}$, and hence to apply quasi-likelihood methods to estimate the unknown parameters of the model. The methodology assumes, of course, that it is possible to characterize the joint distribution of the measured and unmeasured components of $\mathbf{X}$. This may be difficult in the presence of spatial heterogeneity in the measured covariates. To estimate the required means and covariances, Zidek adapted a number of ideas from the Bayesian approach of Le and Zidek (1992), Brown *et al.* (1994), and Le *et al.* (1997) on spatial prediction of a multidimensional variable.

As an illustration, Zidek reported ongoing work from a study of hospital admissions due to respiratory problems in Ontario. He fitted a model of the form

$$E[Y_{kt} \mid \mathbf{X}_{kt}] = m_{kt} \exp\{\boldsymbol{\beta}'\mathbf{X}_{kt}\}$$

where m_{kt} is a multiplier (assumed known) which accounts for the effects of seasonality, day of week, and variable population size between regions. The emphasis was on effects of the pollutants ozone (O_3) and nitrogen dioxide (NO_2), but with maximum daily temperature and average daily humidity also incorporated into the analysis as possible climatological confounders. Some questions raised by the analysis, which also recur in other problems of this nature, were: which measure of the pollutant variable to adopt, and how to deal with long-term trends? The first question essentially boiled down to whether the current day's value or that with a one- or two-day lag should be taken as the best predictor. After considering all three possibilities for both O_3 and NO_2, the two-day lagged variable was adopted. The analysis was also performed separately for each year to look for long-term trends. In this case the results appeared fairly consistent for O_3 but with wide year-to-year variations in the effect due to NO_2.

In another paper from the same volume, McCready *et al.* (1997) studied the influence of road traffic pollution on asthma. An ingenious feature of their approach was the use of a geographical information system (GIS) to compute a "road traffic pollution index" (RTPI) based on all roads within a fixed distance of a subject's residence. They then correlated both (a) whether a subject

has ever had asthma, and (b) whether a subject is currently suffering from asthma, with the RTPI and other factors such as cigarette smoking and gender. They found a strong association between "ever had asthma" and RTPI, but not between "current asthma" and RTPI. However, there seemed to be some inconsistencies in the data. For example, the number of "ever had asthma" patients was smaller than the number of "current asthma" patients. Another curious result was that, apparently, smoking was not a significant risk factor for asthma. Whatever the inconsistencies in these particular results, the idea of using GIS in this way is undoubtedly an excellent one and could be of considerable value in future studies of this nature.

3.2 PM_{10}

It has long been recognized that airborne particulate matter can have major public health impacts. One of the most famous air pollution events of history, the December 1952 'London smog' which resulted in thousands of deaths, was caused primarily by very high levels of particulate matter. This and similar incidents in other European countries and in the U.S. were a major stimulus for new air pollution legislation, including the (U.S.) Clean Air Act of 1970. During the 1990s, however, the debate has shifted, with claims that even low levels of particulates are responsible for thousands of deaths. In the U.S. this has led to a highly charged political debate, with stringent new air pollution standards being proposed despite opposition from industry and from some sections of the scientific community.

An example of the kind of work which underlies this controversy is given by Schwartz (1993). Schwartz collected four years' of daily mortality data as well as related data on meteorology from Birmingham, AL. To this was added data on PM_{10}, i.e., particulate matter of aerodynamic diameter less than 10 μm. Schwartz performed a Poisson regression analysis of deaths against seasonal and long-term trend effects, meteorology, and three-day averages of PM_{10}. He also included corrections for overdispersion and serial correlation, though neither of these was a significant factor in this particular data set. On the basis of these analyses, Schwartz concluded that a unit $\mu g/m^3$ rise in PM_{10} would increase the rate of deaths in the elderly population by about 0.08%. Although this seems a modest enough effect, when combined with similar results from

studies in other cities, and converted to deaths across the whole nation, it has led to claims that up to 60,000 deaths per year in the U.S. may be attributed to elevated levels of particulate matter.

Subsequent studies on the same or similar data sets have been brought forth; see Samet *et al.* (1995; 1997), or a number of analyses performed originally at the U.S. National Institute of Statistical Sciences (Styer *et al.*, 1995; Smith *et al.*, 1997a; Smith *et al.*, 1998). These various studies have identified a number of important quantitative concerns, including:

- *Seasonal variation and trends.* Whatever effects may be attributed to either meteorology or air pollution, there always remains a substantial seasonal component of variation. In addition, there are irregular trends. As an example, Fig. 1 shows weekly total deaths in Birmingham, AL for four years, together with a smoothed trend. The seasonal effect is very strong but also irregular; e.g., during each of the winters of 1985-1986 and 1987-1988 the peak deaths occurred in late February, but in 1986-1987 they occurred at the end of December. A possible approach for modeling the trend component involves some form of spline or LOESS smoother employed in a nonparametric or generalized additive model.
- *Choice of meteorological variables.* Meteorology plays an important role in studies of air pollution because of its role as a possible confounding factor. However, the interpretation of different meteorological variables is open to question. For example, wind speed is sometimes found to be correlated with increased death rates, but it is open to question whether this is a surrogate for particulates or *vice versa* (high winds tend to keep particles suspended). Also, most studies use temperature and humidity as the main meteorological variables of interest, the latter measured either by dewpoint or specific humidity. Implementation varies, however. For example, Schwartz' (1993) main model included temperature but not humidity, whereas Smith *et al.* (1997a) found for the same data set that humidity was an important factor. Sensitivity of the estimated PM effect to the assumed choice of meteorological variables remains one of the key questions in this area.
- *Choice of exposure measure.* Various combinations of current and lagged days of particulate matter have been used to define an appropriate measure of exposure. For example, Schwartz (1993) used three-day averages of PM_{10} *excluding* the current day in his study of Birmingham,

AL, whereas Styer *et al.* (1995) in a similar study of data from Chicago, IL used three-day averages *including* the current day. Smith *et al.* (1998) pointed out that the two measures cannot be interchanged without losing statistical significance. Other studies have used anything from single-day values to five-day averages. The selection effect created by such differing exposure measures has never been quantified, but nonetheless must be taken into account in the assessment of any study claiming a significant particulates-mortality relationship.

• *Existence of a threshold.* Schwartz and Marcus (1990), in a re-analysis of historical data from London in the 1960s, questioned the existence of a particulate threshold — in other words, a level below which there is no discernible effect — to determine whether current standards provide sufficient public health protection. Subsequent papers have generally supported the claim that there is no such threshold, but in most cases without any formal test. Smith *et al.* (1998) proposed a very simple test based on a particulate matter effect of the form $\beta(P - P_0)_+$ where P is the level of particulate matter calculated as a three-day average (or whatever exposure measure is under study), P_0 is a threshold level, and β the regression coefficient. By fitting a linear model including this term for a sequence of values of P_0, it is possible to compute a profile likelihood function for P_0. When this was applied to the Birmingham, AL data, it was found that there was little evidence to discriminate between any two values of P_0 below about 80 $\mu g/m^3$; the bulk of the evidence for a PM_{10} effect comes from data above this value. On the other hand, a similar analysis for Chicago, IL led to the conclusion that any threshold must be close to zero, in other words, supporting the lack of any measurable threshold. There appears to be a need to conduct systematic tests of this nature with other data series.

• *Mortality displacement (harvesting).* Some studies have correlated daily deaths with PM_{10} levels in order to identify an association between the two. However, such efforts do not resolve the question of whether the individuals dying are those who were already very sick and would have died anyway, or whether they were otherwise healthy. The first scenario is known as mortality displacement, or alternatively, the harvesting effect. Harvesting is one of the major uncertainties associated with the interpretation of air pollution mortality data. In fact, the evidence for the existence of a harvesting effect is indirect, and such results that have been obtained must be

regarded as extremely tentative. To aid with this problem, Smith *et al.* (1998) proposed a compartment-type model, dividing the population into "healthy" and "frail" subsets and assuming that most of the deaths occur among the latter group. In principle this may be treated as a latent variable problem, where the size of the frail population may be estimated using Markov chain Monte Carlo methods. The results indicated that the frail population size may be finite and indeed quite small. This would point to a strong harvesting effect, but there is great uncertainty about this conclusion.

• *Influence of different pollutants.* In their analysis of particulate data from Philadelphia, PA, Samet *et al.* (1997) considered the effect of total suspended particulates (in place of PM_{10}) along with other major airborne pollutants. These were: ozone, sulfur dioxide (SO_2), NO_2 and carbon dioxide (CO). In one model with five covariates representing the five pollutants, all were statistically significant. Curiously, one coefficient, that of NO_2, was negative. Samet *et al.* attributed this to multicollinearity among the covariates rather than the implausible conclusion that NO_2 has a protective effect. In a similar but more limited study of data from Chicago, Smith *et al.* (1998) included PM_{10}, ozone, and SO_2 in the same equation, and achieved similar results: all three pollutants contributed significantly, but now the coefficient of SO_2 was negative. Such ambiguities are hardly surprising, since it is known that there is substantial chemical coupling between the different pollutants (Meng *et al.*, 1997). Our own conclusion from these analyses is that while there is indeed evidence that air pollution in general has adverse health impacts, it can be very difficult to separate out a specific effect due to particulate matter.

4. ADAPTIVE SAMPLING FOR POLLUTION "HOT SPOTS"

The issue of clustering in environmetric analyses extends beyond the epidemiological studies noted in §3. Other examples where clustering plays a role include: endangered animal or plant populations, geophysical investigations in which mineral ores cluster unevenly, or fisheries research where schools of fish often cluster together tightly. A particularly important concern in environmental pollution studies results when chemical contamination occurs in "hot spots"

separated by uncontaminated expanses. In this section, we adopt the pollutant setting as a backdrop to illustrate briefly this issue of environmental "hot-spots."

Suppose the primary interest in such a setting lies in estimating the population mean concentration of the chemical pollutant; we will also be interested in locating as many of the "hot spots" as possible, and perhaps quantifying these individually. A traditional approach to sampling might select a simple random or (random-start) systematic sample of sites; the sample average then provides an unbiased estimate of the population average, and the sample could provide a contour map that identifies some pollution peaks. A majority of the measurements will record a zero level of pollution, however, and there is a good chance that some clusters will be overlooked. Although the sample mean is unbiased as an estimator of the population mean, it will have a large variance; in addition, the contour map will have little accuracy in the areas of high concentration, which are of particular interest.

Adaptive sampling provides a sensible alternative to this situation. In adaptive sampling, the direction taken by the sampling procedure at any stage is determined at least in part by the information that has been obtained in the previous sampling. Such a strategy might involve the following: first, take a random sample of a given size from the region of interest; if any of the selected units show contamination, then return and sample every unit neighboring a contaminated unit. If any neighboring units show contamination, sample their neighboring units, etc., until a clean boundary is established for each discovered cluster. Figure 2 illustrates this strategy in the case of a 400-unit population and a simple random sample of size 20. After initial sampling (Figure 2a), the 4 nearest neighbors of any contaminated sample unit are also inspected (Figure 2b). An obvious shortcoming to such a procedure is that the final sample size is not known in advance; the advantages in terms of greater accuracy for estimating the hot spots can far outweigh this drawback, though.

This method of sampling will produce biased estimates of population parameters if the resulting data are naively analyzed. To avoid this, Seber and Thompson (1994) outline a sampling theory and estimation methodology applicable to a wide range of variants on the general adaptive sampling scheme. The classical Horvitz-Thompson or Hansen-Hurwitz estimators (see Stehman and

Overton, 1994) can be modified to obtain unbiased estimators of the unknown population mean. These estimators, along with the mean of the initial sample, are unbiased, but do not necessarily possess minimum variance. To improve them, the Rao-Blackwell theorem may be applied. Seber and Thompson (1994) provide further details, and also outline useful strategies for selecting the initial sample in clusters, stratification, and alternative choices of the criterion for further sampling.

5. TREND ANALYSIS

An effect often studied in environmental science is the analysis of trend in some environmental phenomenon over time. This often leads to adjustments for autoregressive effects or other spatial/temporal correlations in the data, and this is another important area of environmetric trend analysis (Esterby, 1996). We noted some examples of this in §2, above. We give here two additional illustrations: assessing global warming and monitoring ecological systems.

5.1 Estimation of global and regional trends

An important, ongoing concern that incorporates trend analysis is assessing whether global warming is occurring in our environment. Specifically, has average ambient temperature increased in the past 50-200 years as the pace of industrial and economic development has increased worldwide? Data representing such an effect possess some form of time-dependence and possible autocorrelation, calling for a time series analysis. Since the effect may involve very small relative increases over an extended period, however, complementary issues of long-range dependence enter into the analysis. For example, continuing on works by Bloomfield and Nychka (Bloomfield, 1992; Bloomfield and Nychka, 1992), Smith (1993) illustrated a method of analysis that incorporates long-range dependence into the trend assessment using simple linear forms. Specifically, let $\{Y_1,...,Y_N\}$ denote a sample of N observations from a stationary time series with mean zero and autocovariance $\gamma_k = E[Y_n Y_{n+k}]$. Take the spectral density as

$$f(\omega) = \frac{\sum_{n=-\infty}^{\infty} \gamma_n \exp\{-in\omega\}}{2\pi}$$

over $-\pi < \omega < \pi$. This is estimable at any frequency ω via the periodigram

$$I_N(\omega) = \frac{2}{N}\left\{ \sum_{n=1}^{N} Y_n^2 + 2\sum_{k=1}^{N-1}\sum_{m=1}^{N-k} Y_m Y_{m+k}\cos(k\omega) \right\}.$$

Long-range dependence is incorporated via the relationship $f(\omega) \approx b\omega^{1-2H}$ as $\omega \downarrow 0$, for constants $b > 0$ and $H \in (\frac{1}{2}, 1)$. To test for a linear trend, use the trend variable $x_n = n - \frac{1}{2}(N+1)$, where x_n is given zero mean for simplicity. Then, the ordinary least squares estimator of rate of increase takes the well-known form $\hat{\beta} = \sum_{n=1}^{N} Y_n x_n / \sum_{n=1}^{N} x_n$. A simple approximation for the variance of $\hat{\beta}$ is found as

$$\mathrm{Var}[\hat{\beta}] \approx \frac{36\, b\pi\,(1-H)}{H(1+H)\,\Gamma(2H)\,\sin(\pi-\pi H)}\, N^{2H-4}.$$

If b and H are unknown, these must be estimated from the data. Applied to environmental warming data over a variety of sites in central England and the continental United States, Smith (1993) estimated the temperature increases to be between 0.27°C and 0.35°C per year. Standard errors of these estimators ranged between about 0.19 and 0.31, however, suggesting a marginal, but not strongly-significant increase in long-range temperatures. Smith noted that uncertainties with the assumption of a simple linear trend, effects of estimating b and H on the standard errors, and other sensitivities with selected model parameters make these inferences at best preliminary; nonetheless, methods taking into account long-range dependence can provide improvements over simpler autoregressive analyses, and further research into their use is called for, perhaps incorporating more complex polynomials in the trend (Yajima, 1991), including a comprehensive spatial model to account for differences across geographic areas (Solow, 1994), centering on the extremal properties of the distributions (Smith, 1989), or joint estimation of the long range and regression features (Smith and Chen, 1996).

Of course, this application has focused largely on the detection of a linear trend, and this need not be the main question of concern to climatologists. It is an agreed, empirical fact that global temperatures have been rising over the past 150 years, and some climatologists debate the usefulness of testing its statistical significance. Those scientists are much more concerned with distinguishing among different *causes* of global warming. In this connection, there has been much research into the relative effects of greenhouse gases, sulfate aerosols, variations in solar flux, and other influences. Numerical models for the earth's climate produce conjectured "signals" for each of these effects, and the current task is to determine to what extent each of these signals is present in the observed temperature record. Several specific methods have been developed, including the

pattern correlation statistics of Santer *et al.* (1996) and the optimal fingerprinting technique of Hegerl *et al.* (1996). Both of these methods ultimately depend on testing for the presence of a trend of known functional form in a multivariate time series, which from a statistical point of view is a generalization of the problem of detecting a linear trend in the univariate case.

5.2 Monitoring status and trends: EMAP

Estimation of environmental trends also raises important questions on proper technique and bias reduction when sampling environmental data. Driven in part by the need to assess national legislation (such as the U.S. Clean Water Act), there has been considerable emphasis on the spatial status of and trends in pollutant concentrations. One effort which has received much statistical attention is the Environmental Monitoring and Assessment Program (EMAP) of the U.S. Environmental Protection Agency. Designed to describe status and trends of ecological indicators, the sampling program is based on a sophisticated, systematic, hexagonal grid system (Stehman and Overton, 1994). Within this context, an important focus is on the ability to evaluate both status and trends under the EMAP sampling scheme.

For the purposes of optimal sampling, the status and trend outcomes conflict for resources. In evaluating status, the emphasis is on sampling the regions over which an inference is to be made. Samples over time would use different sites. Trend testing and estimation involve the use of the same sites rather than different sites. A practical compromise involves designs that vary some of the sites for estimating status but fix some sites for the estimation of trends. One design uses a rotating panel (Duncan and Kalton, 1987) in which a series of sites are monitored for several consecutive years, then replaced by another randomly selected site. These are augmented by another set of sites that is monitored at all times. Another approach is to use an augmented, serially alternating design in which some sites are monitored on each sampling occasion — e.g., each year — while others are monitored using a wider interval — e.g., every four years. Urquhardt *et al.* (1993) studied the power of tests for trend and status under these two sampling schemes. They proposed the model

$$Y_{ijk} = S_{ik} + T_j + \varepsilon_{ijk}$$

where Y_{ijk} is the measurement made at the ith set of sites (i=1,2,...,s) for year j (j=1,2,...,t) at the kth site within the ith set of sites (k=1,2,..., n_i), S_{ik} is the site effect, T_j is the year effect, and ε_{ijk} is a random error term. Assuming simple temporal autocorrelation, Urquhardt *et al.* evaluated the designs in terms of the power of test of trend/status and the precision of the designs. Here, the augmented design appeared useful for the initial years of the study but became less important over time. Otherwise, the two designs exhibited fairly similar power and precision. Bringing into consideration additional characteristics such as the total number of sites visited suggested, however, that the serially alternating design may in fact be more useful in practice.

In principle, this model has the capacity to account for other, important statistical features, such as substantial gaps in the data (i.e., missing data), or random effects in S_{ik} and/or T_j. For example, van Leeuwen *et al.* (1996) discussed the problem of testing for trend when the trend is viewed as a fixed effect, but time is viewed as random. Their results led to exact tests for testing various hypotheses, under a variety of spatial and temporal correlation structures. Further study is needed, however, to determine how such complications affect the model's ability to make comparisons between trends at different sites.

6. SPACE-TIME MODELING, WITH APPLICATIONS TO ATMOSPHERIC POLLUTION AND ACID RAIN

As noted throughout the preceding sections, the analysis of environmental time series data at a fixed point in space has received generous attention in the statistics literature; likewise, the analysis of spatial patterns at any fixed point in time has undergone extensive, albeit more recent, development. However, developing useful methods for *joint* spatio-temporal analysis remains one of the great challenges facing statistical researchers (Cressie, 1993, Ch. 1). Driven in part by the demand for quantitative methods for large-scale environmental monitoring, spatio-temporal modeling has seen considerable activity in the past decade, and the approaches put forth are as diverse as the data they address. Solow (1994) and Guttorp and Sampson (1994) discuss early work in this and related areas. In this section, we borrow from their reviews, and supplement with discussions of additional work published thereafter to discuss these issues in more depth. A number of treatments deserve more discussion than space considerations allow, however, and we

refer the reader to the early approaches of Eynon and Switzer (1983), Egbert and Lettenmeier (1986), and Eynon (1988). For more recent works, the interested reader may study also the Bayesian approach of Handcock and Wallis (1994), the separate mean and scale approach of Høst *et al.* (1995), and the moving-cylinder models of Haas (1995).

Notation for spatio-temporal modeling is difficult to coordinate over the array of different models seen in this area; one common aspect we employ is that the response of interest at time t and location $\mathbf{s}$ will be denoted $Z(t,\mathbf{s})$. In most cases, a single $Z(t_i,\mathbf{s}_j) = Z_{ij}$ is obtained (perhaps after some imputation or averaging) at each combination of n regularly spaced time points t_i, i=1,2,...n, and p locations $\mathbf{s}_j$, j=1,2,...p.

For example, an early work by Bloomfield *et al.* (1983) studied global trends in total ozone (measured in a column extending from the earth's surface to the top of the atmosphere) via a frequency domain approach. Total ozone has been measured monthly at 36 locations in 7 regions worldwide since the late 1950s. If Z_{ijk} denotes the total ozone at location k in region j at time i, the basic model takes again a log-linear form:

$$\log(Z_{ijk}) = \mu_{ijk} + \alpha_i + \eta_{ij} + \varepsilon_{ijk}$$

where μ_{ijk} is the mean, α_i is a random component common to all stations, η_{ij} is a random component common to all stations within region j, and ε_{ijk} is a station-specific random error. The random components are assumed to be mean-zero, stationary, with no cross-correlations among them. The regional effects η_{ij} are assumed to have the same spectra, as are the errors ε_{ijk} within a region. The mean term is modeled as $\mu_{ijk} = h_i + \nu_{jk}$, where h_i is a temporal trend component and ν_{jk} is a spatial component independent of time. The data were deseasonalized by subtracting the monthly location averages, allowing the terms ν_{jk} to vanish in the subsequent analysis. The global temporal trend h_i was assumed initially to satisfy $h_i = \beta m_i$, where β is an unknown constant to be estimated and m_i is a depletion curve predicted by a photochemical model; testing for β=1 assesses the validity of this model. Bloomfield *et al.* (1983) also considered extensions to include solar effects and effects due to atmospheric nuclear testing.

Under the Bloomfield *et al.* model the discrete Fourier transform of the site-specific series, denoted by $d_{ijk}(f)$ ($-1/2 < f < 1/2$), decomposed into terms as follows (using obvious notation):

$$d_{ijk}(f) = d_h(f) + d_\alpha(f) + d_\eta(f) + d_\varepsilon(f).$$

Bloomfield *et al.* used variance terms within regions and between regions to estimate d_ε and d_η; using structural model(s) for h_i and two kinds of assumptions on the α_i (a first-order autoregressive process and a process with a self-similar spectrum) they were able to estimate d_α. They fit their models via maximum likelihood in the frequency domain. The results showed no strong evidence of a trend in global ozone in the 1970s: the trend coefficient was estimated as $\hat{\beta}$ = 0.1% with a standard error of 0.55%. The associated 95% confidence interval for β did contain the value β=1.

More recently, Niu and Tiao (1995) used extensive data from the total ozone mapping spectrometer (TOMS) aboard the NIMBUS-7 satellite, which has been collecting daily data since 1978. They reduced the data to monthly averages for the period 1979-1989 on 1° latitude by 1.25° longitude pixels spanning the globe, and took advantage of the regular spatial grid of sample points to model the average ozone observation at each latitude at time t_i and longitude j as

$$Z_{ij} = \mu + \beta_j t_i + T_j(t_i) + \varepsilon_{ij}.$$

Here, $T_j(\cdot)$ is a seasonal cyclic term — a weighted sum of sine functions with 12- and 6-month periods — and ε_{ij} is a space-time autoregressive moving-average (STARMA) process (Cliff *et al.*, 1975). Niu and Tiao settled on a STAR(2,1) model for most latitudes:

$$\varepsilon_{ij} = \alpha_1\varepsilon_{i,j-1} + \theta_1\varepsilon_{i,j+1} + \alpha_2\varepsilon_{i,j-2} + \theta_2\varepsilon_{i,j+2} + \phi\varepsilon_{i-1,j} + u_{ij}\ . \qquad (6.1)$$

In this model the α- and θ-unknowns are spatial autoregressive parameters modeling correlations to the west and east of longitude j (respectively), and ϕ specifies a first-order autoregressive parameter for correlation in time. The u_{ij} were assumed uncorrelated, with variances allowed to depend on month. Niu and Tiao used these models to estimate the trend in ozone by latitude and longitude. Their results included a contour plot showing negligible ozone trends in equatorial latitudes, but with increasingly negative and statistically significant trends in total ozone depletion moving towards the poles. Their contours showed a decline for the decade on the order of 5% at the north pole and 10% at the south pole.

A very different atmospheric contaminant model was given by Loader and Switzer (1992), who analyzed the logarithms of sulfate concentration in rainfall at 19 sites in the eastern and midwestern United States for 24 monthly observations, 1982-1983:

$$Z(t,\mathbf{s}) = \mu + T(t) + S(\mathbf{s}) + \varepsilon(t,\mathbf{s})$$

where $T(\cdot)$ and $S(\cdot)$ are smooth unknown functions, and $\varepsilon(t,\mathbf{s})$ is a zero-mean noise, uncorrelated in time at any given site s, but with a possibly nonstationary spatial covariance structure $\boldsymbol{\Sigma}$ at any fixed time. The functions $T(\cdot)$ and $S(\cdot)$ were estimated using a LOESS smoother on the marginal means of the Zs. Loader and Switzer derived the variance of their estimate of E(Z) in terms of the covariance structure of ε and used it to improve the estimated variance of predictions. The sample spatial covariance matrix $\mathbf{S}$ from the residuals of the fit were smoothed via an empirical Bayes approach: An inverted-Wishart density was assumed as a prior for the true covariance matrix $\boldsymbol{\Sigma}$, with prior covariance matrix $\mathbf{C}$ and degrees of freedom m. $\mathbf{C}$ was chosen by fitting an isotropic exponential semivariogram model to the residuals. The parameter m was chosen via empirical Bayes estimation, using information in the marginal density of $\mathbf{S}$. As noted by Guttorp and Sampson (1994), the empirical Bayes enhancement does not affect point predictions for $Z(t,\mathbf{s})$, which remain what they would be under the isotropic exponential semivariogram model used to estimate $\mathbf{C}$. The estimated variances of the predicted values are affected by the empirical Bayes approach, however.

Loader and Switzer examined their models via cross validation. Their analysis showed that the prediction variance formulas performed well. They noted that the estimator for spatial signal $S(\mathbf{s})$ was most likely oversmoothed, but that it might be unwise to reduce the smoothing parameter with so few sites.

Oehlert (1993) also modeled log sulfate concentrations in eastern North America, combining information from 4 modeling networks (the APIOS-C, MAP3S, NADP/NTN, and UAPSP networks), for a total of 94 stations over the five-year span 1982-1986. Most of his analysis was based on yearly precipitation-weighted means. He first estimated the five-year mean and linear trend in log concentrations for all stations by ordinary least squares. He reported results only for the five-year overall means, however, citing a need for longer series to adequately address trends.

His spatial approach tiled eastern North America into 1° latitude by 1.5° longitude cells. If a_j denotes the estimate of five-year mean for station *j*, this assumes

$$E(a_j) = \alpha_{k(j)} + \nu_u$$

where $\alpha_{k(j)}$ denotes the true mean of the cell in which station *j* resides and ν_u is an adjustment term for network u, u=1,2,3 (relative to the NADP/NTN network). Oehlert incorporated the assumption of similarity between neighboring cells via a partially improper prior on $\boldsymbol{\alpha}$, with mean vector $\mathbf{0}$ and inverse covariance matrix $\lambda_\alpha(\mathbf{A}'\mathbf{A})$. $\mathbf{A}$ has a row for every pair of adjacent cells and a column for every cell; it is all zeros except that each row has entries of 1 and –1 for the associated cell pair coefficients. (This formulation could also be viewed as a discrete, two-dimensional first-order smoothing spline.) Oehlert placed independent, zero-mean normal priors on the network bias terms. The prior variances were set to $1/\lambda_\nu$. The resulting posterior distribution for $(\boldsymbol{\alpha},\boldsymbol{\nu})$ was Normal, with mean vector

$$\left(\mathbf{W}'\boldsymbol{\Sigma}_\mathbf{a}^{-1}\mathbf{W} + \begin{bmatrix} \lambda_\alpha\mathbf{A}'\mathbf{A} & 0 \\ 0 & \lambda_\nu\mathbf{I} \end{bmatrix}\right)^{-1} \mathbf{W}'\boldsymbol{\Sigma}_\mathbf{a}^{-1}\mathbf{a}$$

and covariance matrix

$$\left(\mathbf{W}'\boldsymbol{\Sigma}_\mathbf{a}^{-1}\mathbf{W} + \begin{bmatrix} \lambda_\alpha\mathbf{A}'\mathbf{A} & 0 \\ 0 & \lambda_\nu\mathbf{I} \end{bmatrix}\right)^{-1},$$

where $\mathbf{W}$ is the matrix relating each a_j to its expected value in terms of the parameters $\alpha_{k(j)}$ and ν_u, and $\boldsymbol{\Sigma}_\mathbf{a}$ is the covariance matrix (to be determined) of $\mathbf{a}$.

Oehlert used a combination of historical information, indirect generalized cross-validation (IGCV, see Altman, 1990), and sensitivity analysis to determine values of unknown parameters such as λ_α, λ_ν, and unknowns in $\boldsymbol{\Sigma}_\mathbf{a}$. The terms in this latter matrix were assumed to have a component due to site-specific effects (which could be estimated since several cells had multiple sites), a first-order moving-average correlation structure across years within sites (with correlation on the order of 0.01, citing historical sulfate studies), a long-term, large-scale temporal correlation structure similar to ARMA(1,1) models reported for precipitation, and a spatial covariance modeled in one of three ways: (1) an equal variance/equal covariance model; (2) a kernel-smoothing

approach with variances and covariances taken as a function of distances between sites, and smoothing controlled by a parameter λ_K; or (3) an isotropic, exponential semivariogram structure.

The IGCV analysis showed essentially identical fits for the first two models, with somewhat less satisfactory fits for the isotropic exponential semivariogram. Due to its simplicity, Oehlert adopted the equal variance/equal covariance model for interpretational purposes. It is worth noting, however, that the kernel covariances showed distinct but weak anisotropy with peak correlations at angle approximately 0.3π, roughly parallel to dominant high-altitude wind vectors in the region; anisotropy of this sort was also reported for this region (using hydrogen ions and winter data) by Guttorp and Sampson (1994), employing a deformation approach (described below) to modeling heterogeneous covariance functions. Under the equal variance/equal covariance model, the station-specific variance was clearly the largest important component, but this effect would be expected to decrease with regional averaging. There were few, if any, network differences; a suggestion of an effect was evidenced for the APIOS-C network, but this could have been attributable to partial confounding with spatial structure since the APIOS-C sites had little overlap in space with sites from the other networks.

The approaches of Loader and Switzer (1992) and Oehlert (1993) feature attempts to model anisotropy and nonstationarity in the covariance structure of the data. An important general approach to this problem, known as the *deformation approach*, has been the subject of considerable development and application to atmospheric data. The summary provided here is taken from Meiring et al. (1997); in recent years, the methodology has been applied in analyses of solar radiation (Sampson and Guttorp, 1992), acid precipitation (Guttorp *et al.*, 1992; Guttorp and Sampson, 1994), and tropospheric ozone (Guttorp *et al.*, 1994), among others.

The fundamental idea underlying the deformation approach is to compute a deformation of the geographic plane so that the spatial covariance structure can be considered stationary and isotropic in terms of a new spatial coordinate system. For simplicity of presentation, assume independence in time. The spatial dispersion function, defined as $D(\mathbf{s},\mathbf{u}) = \mathrm{Var}[Z(t,\mathbf{s}) - Z(t,\mathbf{u})]$ for each pair of spatial locations $(\mathbf{s},\mathbf{u})$, is modeled as

$$D(\mathbf{s},\mathbf{u}) = \gamma_\theta \, || f(\mathbf{s}) - f(\mathbf{u}) ||$$

where $f(\cdot)$ represents a smooth (bijective) transformation of the original geographic coordinate system (the "G-space") to the new coordinate system (the "D-space", which has been of dimension 2 and referred to as the "D-plane" in applications to date), and γ_θ represents an isotropic variogram function with parameters θ. The transformation is accomplished via a pair of thin-plate splines (Wahba, 1990), with transformation and semivariogram parameters chosen to minimize the objective criterion

$$C_{\theta,f,\lambda} = \sum\sum_{i \neq j} \left[\frac{d_{ij} - \hat{d}_{ij}}{\hat{d}_{ij}} \right]^2 + \lambda(\text{BEP})$$

where d_{ij} and $\hat{d}_{ij}$ denote the empirical and fitted dispersions, respectively, between sample sites s_i and s_j, λ is a smoothing parameter, and BEP denotes a bending energy penalty for the transformation. This penalty is a quadratic form in the D-plane coordinates. The second term above controls the smoothness of the transformation; small λ may result in a "folded" D-plane representation which is generally uninterpretable, while very large λ results in a stationary or homogeneous model with elliptical anisotropy. Meiring et al. (1997) demonstrated effects of various choices of λ using both simulated and real data. Visual interpretation of the deformation mapping is accomplished via biorthogonal grids, as discussed by Guttorp and Sampson (1994). Amongst other applications, these spatial correlation models may then be used in the estimation of the values of a spatio-temporal process at unmonitored locations.

7. ENVIRONMENTAL RISK ASSESSMENT VIA LABORATORY EXPERIMENTATION

Another important area of environmetric research is that of quantitative risk assessment. In its simplest environmental characterization, risk assessment concerns the identification of potential risks to public health from hazardous chemicals, radiation, etc. (Portier, 1989). Data often come from bioassays on small mammals or other biological systems, or epidemiological analyses of human populations at risk. (We illustrated an example of this issue with human populations in the discussion on atmospheric particulates — see §3.) A major quantitative component of such studies is statistical characterization of the stimulus/dose-response of the biological organisms to the hazardous agent, and from this, estimation of possible human risks based on low dose extrapolation from the dose-response data.

A full overview of the many quantitative problems in environmental risk assessment could (and has!) covered entire journal issues and even full textbooks; hence our discussion here is necessarily incomplete. We hope, however, to highlight some open areas of research; for more in-depth surveys, see the early review by Krewski and Brown (1981), special journal issues as introduced, e.g., by Redmond (1991) or Bailar (1988), or some of the more recent works in the literature, such as Bailer and Portier (1994) or Hallenbeck (1993), among many others.

When conducting predictive or environmental toxicity studies that generate data based on a dose-response, it is common for the dose levels to be taken at fairly high values. This is true primarily for laboratory animal experiments conducted as screens for certain toxic effects (Haseman, 1984) due to the relatively short time span available for the animals to exhibit the toxicity. It is a long-recognized concern in quantitative risk assessment, however, that the lack of observed low-dose information often results in suspect inferences (Crump and Howe, 1985)

Historically, to estimate low-dose effects regulators have used the lowest or least potent exposure to a chemical at which toxicity is observed. This is called the *lowest-observed-effect level* (LOEL). Lying ostensibly below the LOEL is the highest concentration where *no* toxicity is observed: the *no-observed-effect level* (NOEL). Extensions include no-observed-*adverse*-effect levels (NOAELs), lowest-observed-adverse-effect levels (LOAELs), etc. These quantities are determined statistically by comparing each concentration's observed response with the zero-concentration control group: e.g., the NOAEL can be estimated as the highest concentration at which no significant increase in response is seen over the control, after adjusting for the multiple comparisons. Unfortunately, observed effect level estimation is tied critically to the spacing of the doses chosen for each study. If the dose grid is not fine enough, the resulting observed effect level may be only a crude estimate.

To illustrate, consider the following data, which are proportions of mice exhibiting bladder tumors after exposure to sodium saccharin, as discussed by Kodell and Park (1995). At saccharin exposures of 0.01, 0.10, 1.0, 5.0, and 7.5% of diet, 0/25, 0/27, 0/27, 1/25, and 7/29 mice exhibited tumors, respectively. To determine the NOAEL, we perform a series of one-sided Fisher exact tests comparing each exposure group with the control, and applying a Bonferroni adjustment

to the *P*-values. (For the saccharin data, we use the response at $d = 0.01\%$ to approximate the control response.) At $\alpha = 0.05$, and adjusting for multiplicity, the NOAEL for these data is the exposure at 5.0%. Clearly, however, this is a relatively crude measure of the chemical's potential toxic risk, and many questions regarding it remain. For instance, is the observed response at this exposure truly different from the control (or is the lack of significance a function of the small sample size), would higher exposures of 6 or 7% produce essentially the same response, or, for that matter, is the underlying response at 1% also essentially similar to that at 5%? If the risk analyst were interested in specifying some form of "safe dose" estimate for use in public health management — e.g., the dose at which a very small response rate of, say, 10^{-6} is incurred — is the NOAEL even useful in this case? Here, since the NOAEL corresponds to a response rate well above 10^{-6}, it is not particularly useful for "safe dose" estimation.

Common modifications of the NOAEL include downward adjustment by an "uncertainty factor" — say, dividing by 10 — to account for incertitude in the estimation process. The result is known as a *reference dose*, or RfD (Cicmanec *et al.*, 1996), but clearly suffers from the same limitations as the NOEAL on which it is based. Some of these questions could be addressed with more advanced statistical tools, such as confidence intervals for NOAEL or NOEL (Schoenfeld, 1986), but these may not overcome the fundamental austerity of information the measure presents. In general, observed effect levels are viewed as poor summary statistics for modern risk analysis (van der Hoeven, 1997).

Unfortunately, numerous gaps remain in methodological understanding and implementation of low-dose extrapolations when used to support quantitative risk assessments (Freedman and Zeisel, 1988). In order to develop more realistic risk assessment formulations, complex models are under continuing development. For example, in cancer risk assessment multi-stage modeling of the carcinogenic process has received extensive interest, going back to an original formulation of Armitage and Doll (1954). The basic multi-stage model assumes that after time t of exposure to a dose d of some hazardous agent, a group of normal cells acquire a mutation that leads to unregulated growth (cancer) over a series of k progressive stages. This leads to a probability function for tumor development of the form $P(d,t) = 1 - \exp\{-Ct^k(q_0 + q_1 d + \cdots + q_k d^k)\}$,

where C and $q_i \geq 0$ are unknown constants. (Other link functions are possible, such as the logistic form $P_L(d,t) = [1+\exp\{-\beta_0 - \beta_1 d - \cdots - \beta_k d^k)\}]^{-1}$. These do not share a mechanistic motivation available with the multi-stage model, however.) At k=1, the multi-stage form corresponds to a simple *one-hit model* of carcinogenesis (Hoel, 1985). Notice that the one-hit form $1 - \exp\{-Ct(q_0 + q_1 d)\}$ is approximately linear in d as $d \rightarrow 0$; this feature makes it a popular first approximation for many low-dose problems.

For example, consider the saccharin data discussed above (for simplicity, set $t = C = 1$). Applying the one-hit model, we find the maximum likelihood prediction equation to be $\hat{P}(d,1) = 1 - \exp\{0.1946 - 0.0512d\}$. From this, greater flexibility in, e.g., safe dose estimation is available than, say, the NOAEL. For example, setting $\hat{P} = 10^{-6}$ and solving the prediction equation for d yields a safe dose estimate of $d_{SD} = 3.80\%$. This is clearly below the NOAEL of 5% seen above. Improvements in the estimation process are also possible; e.g., higher-order linear predictors or low-dose linear approximations can improve precision in the model fit, or confidence limits on the slope parameters can add a conservative property to the safe dose estimates (Kodell and Park, 1995).

Further development of the multi-stage model has incorporated multi-step/multi-stage biological processes such as cellular proliferation and transformation, *in vivo* pharmacokinetics and pharmacodynamics of the hazardous agent, and complex features of dose-related mutagenesis. A popular variation is due to Moolgavkar and colleagues (Moolgavkar and Venzon, 1979; Luebeck and Moolgavkar, 1996), where two mutations are assumed necessary for a normal cell to transform into a cancerous one. Various levels of initiation and promotion of the cancer are accommodated in this model, although further study of the two-mutation assumption, and stability/interpretation of statistical estimates for the model parameters remain open areas of study (Portier and Kopp-Schneider, 1991; Little, 1995).

8. ECOLOGICAL MODELING: HOW ENVIRONMENTAL FACTORS AFFECT ANIMAL POPULATIONS

8.1 Modeling salmon populations in the San Joaquin River

With many environmental data sets, statistical analyses may be developed from complex models of the phenomena being studied. (The Moolgavkar two-stage model noted at the end of §7 is a good example.) This can be particularly useful in cases where a standard regression or time series analysis would ignore key features of the data structures. For instance, in the area of ecological population dynamics, Speed (1993) discussed the problem of modeling the number of Chinook salmon in the San Joaquin River. There, the salmon population has declined over the last few decades, and great interest exists in understanding possible causes of the decline, especially those related to environmental factors. Data are available on Spring river flow, escapement (abundance of fish available for spawning), and catch of fish. The data form a time series, and Speed indicates that standard approaches based on classical regression analysis cannot approximate the complexity of the problem. Conversely, basing the analysis on complex age-class models produces far too many parameters and components, making them difficult to fit and evaluate. A compromise is found in a model which includes both the age-class components and a stochastic component. The model is described through the (unobservable) number of fish that survive to a particular age — called 'recruits' — and the number of these which spawn. These then lead to models for the observable number of fish which are caught (C_t) and the escapement (E_t). The model equations begin with

$$R^2_{t+1} = \alpha Q_t S_{t-1} \exp\{-\beta S_{t-1}\} + e_{t+1}$$

$$R^3_{t+1} = (1-\mu)(1-\omega)(1-\rho_2) R^2_t$$

$$R^4_{t+1} = (1-\mu)(1-\gamma)(1-\rho_3) R^3_t$$

and

$$S_t = (1-\mu)(1-\omega)\rho_2 R^2_t + (1-\mu)(1-\gamma)\left(\rho_3 R^3_t + R^4_t\right),$$

where R^k_t is the recruitment in year t for age fish of age k (k = 2,3,4), and α is a recruitment parameter. Q_t is the flow, S_t is the abundance of spawners in year t, μ is the fraction lost to ocean mortality, ω the fraction lost to fishing for two-year old fish, γ the fraction caught for three- and

four-year old fish, and ρ_i is the fraction of year class i returning to spawn ($i = 2,3$). The observable quantities are then described via the equations

$$E_t = S_t + \delta_{1t}$$

and

$$C_t = (1-\mu)\gamma\left(R_t^3 + R_t^4\right) + \delta_{2t},$$

where the δs are additive error terms. By reducing the number of model parameters to only a select few, the age-class model is fit using Kalman filtering. The resulting model provides a reasonably good fit to the San Joaquin data, although it raises almost as many questions as it answers regarding fishery population management in the face of environmental disruptions. Speed's (1993) discussion touches on some of these.

Speed's article illustrates several important aspects of stochastic environmental modeling. First, it is often foolish to apply statistical methods blindly to solve complex environmental problems. Second, understanding of the problem and of its fundamental components are essential. Third, models and data must match the questions that are of interest.

8.2 Modeling animal abundance for assessing ecological risk

Statistics has a long, fruitful relationship with fisheries and wildlife sciences in developing stochastic models of vertebrate populations. Some of these methods are finding use in the assessment of environmental impacts. For instance, Anderson *et al.* (1995) review the use of animal abundance models for assessing ecological risks to vertebrate populations. They apply their methodology to analyze survival of the northern spotted owl after the animal experiences habitat loss, employing the well-known Leslie-Lefkovitch model (Leslie, 1945). The model uses information about survival and fecundity in a matrix framework to predict future age structure from past age structure information. For spotted owls, the model is applied to a post-birth population with four age classes. The model sets

$$\mathbf{N}_{t+1} = \mathbf{A}\mathbf{N}_t$$

where $\mathbf{N}_t$ is a vector containing population sizes of the four age classes,

$$\mathbf{A} = \begin{bmatrix} 0 & b_1\phi_1 & b_2\phi_2 & b_3\phi_3 \\ \phi_1 & 0 & 0 & 0 \\ 0 & \phi_1 & 0 & 0 \\ 0 & 0 & \phi_a & \phi_a \end{bmatrix},$$

b_i is an age-specific fecundity for category i, and ϕ_i is an age-specific annual survival probability, any of which may depend on environmental factors. This model has the attractive feature that the dominant root of the characteristic equation of **A** determines the rate of increase (or decrease) of the population (assuming **A** is fixed). By estimating this parameter and its uncertainty from estimates of the parameters of **A**, the effects of important environmental factors can be assessed. With female spotted owls, for example, Anderson *et al.* determined that the characteristic root was significantly less than zero, suggesting a decline in female owl populations due perhaps to habitat loss. However, other parameters, including many vital rates, did not exhibit significant negative trends, and the issue remains open for further study.

Anderson *et al.* (1995) and the references therein provide other interesting examples of the application of statistics to ecological risk assessment. An important point of their article is that use of appropriate variance models is critical in stochastic modeling. For example, misleading results can occur with an age structured model when variance estimates for projections are computed based on data from a single sampling season. A better estimate is obtained by considering the process variance, which reflects changes in the parameters over a number of sampling occasions. Further work into how best to incorporate and apply process variance information is still needed, however.

9. DETECTION LIMITS

Environmental monitoring as described, e.g., in §5.2 often requires determination of a chemical pollutant's existence and measurement of its concentration. Detection limits and related quantities arise in these settings primarily from the need to deal with instrument measurement error. When presented with an environmental sample, say of water, laboratory instruments for measuring for the presence of a pollutant, a chemical analyte, or biological microorganism may not have the

sensitivity to detect small amounts of the (bio-)chemical component of interest. Or, they may incorrectly give non-zero responses when presented with material not containing the chemical.

The problem of assessing environmental effects with observations that may fall below a detection limit has a long history (see Akritas *et al.*, 1994). In most applications, the detection limit paradigm is based on the evaluation of machine error and the processing of blank (no chemical) signals. This has led to definitions such as those presented by the American Chemical Society (ACS) Committee on Environmental Improvement (1980), in which the detection limit is given as "the lowest concentration level that can be determined to be statistically different from a blank." Computationally, the ACS suggested using three times the standard deviation of blank responses. The underlying foundation for this (and other) thinking about detection limits comes from work by Currie (1968), who took essentially a decision theoretic view. A detected observation was simply one that led to rejection of the (null) hypothesis that a chemical or analyte was absent.

Specifically, let ξ be the true concentration which results in measured response Y. Y is viewed as a random variable whose distribution depends on ξ. A calibration curve also depending on ξ, $F(\xi)$, is employed to estimate the underlying concentration, via $\hat{\xi} = F^{-1}(Y)$. Currie (1968) delineated three possible limits based on estimating Y or $\hat{\xi}$. The first such is a "decision" limit, as presented in Davis (1994): "the signal level (response) above which an observed signal may be reliably recognized as being detected." This may be interpreted statistically as the critical value for testing the null hypothesis $H_0:\xi=0$ using the data on Y. A second limit is the "detection limit or true signal level (concentration) which will reliably produce observed signals which lead to detection." This may be interpreted as the concentration ξ which has a high power for the detection hypothesis. Finally there is the determination or method quantitation limit: the "true signal level (concentration) which will be expected to provide measurements of adequate precision for quantitative determination (as opposed to qualitative detection)." Currie's approach thus divided the measurement axis into regions of unreliable detection, detection but unreliable measurement, and reliable measurement.

The definition of detection limits may be quite important when decisions are based on proper detection of hazardous compounds. In compliance monitoring for example, measurements may be taken from several wells around a toxic waste site to assess if hazardous material is entering the groundwater. The sampling protocol typically monitors well water in a routine fashion unless a toxin or toxic indicator is detected; then, more intensive (and costly) monitoring is applied. Gibbons (1994; 1996) noted, however, that the simple decision theoretic approach to setting a detection limit may be flawed for this problem. His work focused on two important questions: (i) if k wells are sampled, what is the probability that at least one exceeds a regulated standard due to instrument error when in fact all are in compliance [cf. Lambert *et al.* (1991)]? And, (ii) how should a method detection limit (MDL) be estimated when the variance increases with concentration? These problems have quite different perspectives on what the MDL is, and estimates of the limit can differ widely. What seems important is that definitions of detection limits need to be stated carefully in experimental protocols and scientific reports.

Even if properly defined, use of MDLs for estimation and testing purposes is irregular and uneven in the literature, primarily due to minimal statistical input on how to quantify MDLs and non-detects. Common usage replaces non-detects with a single value, such as $0, \frac{1}{2}$MDL, or MDL. Clearly, this is too simplistic for most applications — although Davis (1994) indicates selected cases where the results can be reasonable — and only slightly greater effort is required to improve the estimation process. For example, Akritas *et al.* (1994) illustrate a robust parametric method for quantifying non-detects, using simple probability plot regression. Given N observations of which n_C are below a defined MDL, the $N - n_C$ detected (non-censored) observations are plotted on normal (or log-normal, etc.) probability paper and then a straight line is fit through them. Extrapolated back into the non-detect region, the line provides estimates for the non-detected values.

A hypothetical example given by Akritas *et al.* illustrates the approach. Table 1 contains data, Y_i (i=1,...,50), generated from a distribution such that $\log\{Y\} \sim N(0,1)$. Suppose detection can only occur if the observation exceeds 0.4. Thus $n_C = 10$ observations are non-detects in Table 1. Plotting the logs of the 40 detected values against quantiles from a standard normal and then fitting

a straight line to them produces estimates for the log-non-detects on a line with intercept –0.0689 and slope 1.064. Table 2 gives the corresponding estimates (on the original scale), along with the more naive 0, $\frac{1}{2}$MDL, and MDL values. Improvements in estimating the non-detects are evidenced, although there appears to be a slight upward bias in the estimates with these data. Akritas *et al.* (1994) discuss modifications and improvements to address variance and bias reduction for this and other approaches with detection limit data.

In these areas, many further problems remain unresolved. For example, the chemical monitoring problem typically involves multivariate data, bringing with it the potential for multiple censoring (El-Shaarawi and Naderi, 1991). How can the ideas of detection limits be extended to the multivariate setting, and how should detection limit studies be designed in this case? How should monitoring programs apportion resources between the expense of a detection limit study and the larger monitoring effort? Also, Lambert *et al.* (1991) and Gibbons (1994) emphasize that non-detects in the field are different from those in the laboratory, and issues important to one area may be only secondary in the other. Indeed, Davis (1994) notes the need for more emphasis on random effects and variance components in the analysis of data from field studies.

It is worth noting that Currie's (1968) seminal work focused largely on quality control and error rates. In actual applications, there may be errors associated with technicians, machines, and laboratories, as well as confounding effects due to the manner in which the soil or water is collected and treated following collection. The data collection effort may also introduce new complexities. For example, when data have been collected over time, there may be multiple limits due to ongoing improvements in measurement. Or, it is often assumed that an observation measured as a non-detect actually corresponds to a value below the detection limit. As pointed out by Lambert *et al.* (1991), however, this is not true; in fact, depending on how the limit is defined and what is studied, there may be cases where a known concentration is recorded as a non-detect.

10. COMBINING ENVIRONMENTAL INFORMATION

Another increasingly important issue in the environmental sciences is the need to combine information from diverse sources that relate to a common endpoint. Statistical techniques for data

combination continue to develop, however, and the issue of combining environmental information is a very active area of statistical and applied subject-matter research. A common rubric for combining independent results is *meta-analysis* (Hedges and Olkin, 1985), where the goal is to bring together results of different studies, re-analyze the disparate results within the context of their common endpoints, and provide a quantitative analysis of the phenomenon of interest based on the combined data. With many environmental endpoints, however, the effects of interest often are small and therefore hard to detect with limited sample sizes; or, data on many multiple endpoints may mask small or highly-localized effects.

In this section, we assume the different studies are considering similar endpoints, and that the data derived from them will provide essentially similar information when associated with similar study conditions. This is a form of *homogeneity* or *exchangeability* among the data sources. For this setting, we discuss how some attempts at solving these data combination problems have led to a number of interesting environmetric developments and modifications of standard statistical methods.

10.1 Combining *P*-values

Perhaps the best-known method of combining information is Fisher's inverse χ^2 method (Fisher, 1948), where individual P-values, P_k, from K independent studies are combined (k=1,...,K). The result is a combined P-value: $X_F^2 = -2\sum_{k=1}^{K}\log(P_k)$, which is compared to a χ^2 reference distribution with 2K *df*. For example, when characterizing or remediating environmental waste sites, levels of various toxic chemicals are recorded at a single site to identify if a particular clean-up technology is operating properly. The data are collected at K different locations within the site, requiring efficient combination as part of the reporting process. One questions whether the overall clean-up been successful, or is more effort required?

If there are M different chemicals' concentrations recorded at each location, this may be viewed as a multiparameter hypothesis testing problem, where we observe K independent P-vectors $\mathbf{Z}_k = [Z_{k_1},...,Z_{k_M}]'$, each with common mean $\boldsymbol{\mu}_k = [\mu_{k_1},...,\mu_{k_M}]'$ and possibly-unequal covariance matrices $\boldsymbol{\Sigma}_k$. Then, we test if $\boldsymbol{\mu}_k$ has exceeded some known threshold vector $\boldsymbol{\mu}^o = [\mu^o_{k_1},...,\mu^o_{k_M}]'$.

If the null hypothesis of no exceedance is rejected, the clean-up has not been successful, and requires continuation.

For the special case $\Sigma_k = \sigma_k^2 \mathbf{I}$ (k=1,...,K), and assuming normality, each separate location provides an independent F-statistic for testing $\mu_k = \mu^o$. Combination of the information to achieve an omnibus test of the clean-up's effectiveness across all M chemicals can be achieved via Fisher's method: take the individual *P*-values, P_k, based on each F_k, and compute X_F^2. If X_F^2 is larger than an upper-α $\chi^2(2K)$ critical point, conclude that the clean-up requires continuation.

Alternatively, in selected cases it is possible to derive a more powerful combined test, by taking advantage of possible correlations between the $\mathbf{Z}_k$ values. Consider the case K=2: let R be the observed correlation coefficient between $\mathbf{Z}_1 - \mu^o$ and $\mathbf{Z}_2 - \mu^o$. Under H_0, R possesses a distribution whose density is proportional to $(1 - r^2)^{(M-3)/2}$, and this allows for calculation of a one-sided *P*-value, say, P_R. Then, Mathew *et al.* (1993) show that Fisher-combination of the *P*-values P_1, P_2, and P_R, via

$$X_R^2 = -2\{\log(P_1) + \log(P_2) + \log(P_R)\},$$

yields a more powerful test statistic than X_F^2, referencing X_R^2 to a $\chi^2(6)$ distribution. In this particular environmental clean-up application, there are the concerns that statistics based on the simple correlation may also be sensitive to cases where the observed vector drops well below the threshold level, or whether the sample sizes are large enough to assure reasonable power to detect departures from H_0. Also, it may be more appropriate to test for threshold exceedance in population extremes, rather than population means. Useful modifications and extensions may be possible in these regards, however, and development of this environmetric application is an important area of further research.

10.2 Hierarchical Bayesian Methods for Combining Information from Multiple Studies

Some fascinating applications of combining environmental data involve settings where a hierarchical model is posited, and appeal is made to some form of Bayesian or empirical Bayesian analysis. For instance, following on analyses by Hasselblad (1994), DuMouchel (1994) considered nine separate studies from North America and western Europe on toxicity to the

airborne irritant nitrogen dioxide (NO_2). The studies reported adverse lower respiratory symptoms after NO_2-exposure in children aged 5 to 12 years, using odds ratios to quantify any increased risk of lower respiratory distress.

To adjust for possible sources of heterogeneity across differences in design and subject characteristics among the nine studies, DuMouchel employed a hierarchical regression model. A key feature was inclusion of covariates that represented the different sources of heterogeneity. For each study the outcome of interest was the odds ratio of exposure for responding subjects ("cases") to odds of exposure for non-responding, healthy subjects ("controls"). Denote the log-odds ratios as ψ_i (i=1,...,9), and assume $\psi_i \sim$ (indep.)$N(\theta_i, s_i^2)$. DuMouchel made s_i proportional to $\log\{\upsilon_i/\lambda_i\}$, where υ_i and λ_i are the upper and lower 95% confidence limits on the odds ratio reported in the *i*th study. The hierarchical feature assumed $\theta_i \sim$ (indep.)$N(\eta_i(\boldsymbol{\beta}),\tau^2)$, where $\eta_i(\boldsymbol{\beta})$ is a linear predictor encompassing the regression feature, $\eta_i(\boldsymbol{\beta}) = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_P x_{Pi}$, and τ^2 is a hierarchical variance parameter (DuMouchel and Harris, 1983). The x_{ki} terms represent covariates that quantify the known sources of heterogeneity. DuMouchel set P=3 and defined the covariates as indicators that identified if the *i*th study failed to correct for (1) background smoking, (2) NO_2 measurement heterogeneity, or (3) subject gender. In this way, the $\boldsymbol{\beta}$ parameters act to correct the log-odds ratios for any single study's failure to correct for these factors.

If in the parameter hierarchy the prior quantities $\boldsymbol{\beta}$ and τ^2 are unknown (as is common), further hierarchical hyper-prior distributions can be assigned to them. Often, the hyper-priors are taken as diffuse functions in order to represent a form of vague prior knowledge. The various levels of the hierarchy are then combined in standard fashion to yield posterior specifications for the parameters of interest, here, the expected log-odds ratios θ_i. Point estimates are taken as the posterior means of the θ_is, and standard errors are available as the square roots of the posterior variances of the θ_is. Applied to the nine NO_2 studies, DuMouchel's hierarchical regression model produced posterior interval estimates based on normal approximations for the posterior log-odds in which five of the nine studies exhibited significant posterior increases in odds of disease. [Unadjusted for the hierarchical model effects, only four of the nine separate studies were viewed as significant; see

Piegorsch and Cox (1996).] The hierarchical model was able to synthesize information across the ensemble of data, helping to more sensitively identify the significant effects.

Hierarchical Bayesian analyses such as this represent important advances for understanding the complex effects of environmental toxins, and further formulations and applications of such models represent important examples of advanced statistical research in the environmental sciences. We expect their development will continue, as this and other recent examples (Warren-Hicks and Wolpert, 1994; Consonni and Veronese, 1995; Dominici *et al.*, 1997) have begun to illustrate.

11. SUMMARY

The many environmetric problems described above represent only a sampling of the great diversity of challenging issues in quantitative environmental research, and of the great diversity of views on how to solve them. A goal of our presentation has been to mirror both forms of diversity, illustrating that a great many perspectives exist on the nature of "environmental statistics." Many of these areas remain open for further advancement, and as we have noted throughout, any such advances in both the science and the statistics cannot occur without greater multidisciplinary collaboration among subject-matter scientists, social/public policy makers, and statisticians. We encourage statistical *and* subject-matter readers to assume these challenges, and in doing so, to work towards better multidisciplinary interaction. The resulting quantitative methodology will best represent good statistics, good science, and good regulation/public policy.

Acknowledgments

Thanks are due to Professor George Casella for his encouragement during the development of the manuscript, and to the Editor, an anonymous reviewer, and Professor G.P. Patil for their helpful comments. Research by Dr. Piegorsch was supported in part by grant #CA76031 from the U.S. National Cancer Institute, and that by Dr. R. L. Smith was supported in part by grant #DMS-9705166 from the U.S. National Science Foundation and by the U.S. Environmental Protection Agency Cooperative Agreement #CR819638-01-0 at the National Institute of Statistical Sciences.

REFERENCES

Akritas, M. G., Ruscitti, T. F., and Patil, G. P. (1994). Statistical analysis for censored environmental data. In *Handbook of Statistics Volume 12: Environmental Statistics*, Patil, G. P. and Rao, C. R. (eds.), 221-242. Amsterdam: North-Holland/Elsevier.

Altman, N. S. (1990). Kernel smoothing of data with correlated errors. *Journal of the American Statistical Association* **85**, 749-759.

Anderson, C. W. (1994). The aggregate excess measure of severity of extreme events. *Journal of Research of the National Institute of Standards and Technology* **99**, 555-592.

Anderson, C. W., and Nadarajah, S. (1993). Environmental factors affecting reservoir safety. In *Statistics for the Environment*, Barnett, V. and Turkman, K. F. (eds.), 163-182. New York: John Wiley & Sons.

Anderson, D. R., White, G., and Burnham, K. P. (1995). Some specialized risk assessment methodologies for vertebrate populations. *Environmental and Ecological Statistics* **2**, 91-115.

Armitage, P., and Doll, R. (1954). The age distribution of cancer and a multi-stage theory of carcinogenesis. *British Journal of Cancer* **8**, 1-12.

ASC Committee on Environmental Improvement (1980). Principles of environmental analysis. *Analytical Chemistry* **55**, 2242-2249.

Bailar, J. C. (1988). Foreword: The Ninth Symposium on Statistics and the Environment. *Statistical Science* **3**, 274.

Bailer, A. J., and Portier, C. J. (1994). Modeling risks from water contaminants: The application of concentration-response models. In *Water Contamination and Health: Integration of Exposure Assessment, Toxicology, and Risk Assessment*, Rhoda, G. and Wang, M. (eds.), 447-466. New York: Marcel Dekker.

Barnett, V., and Turkman, K. F. (1993). *Statistics for the Environment.* New York: John Wiley & Sons.

Barnett, V., and Turkman, K. F. (1994). *Statistics for the Environment 2: Water Related Issues.* New York: John Wiley & Sons.

Barnett, V., and Turkman, K. F. (1997). *Statistics for the Environment 3: Pollution Assessment and Control.* New York: John Wiley & Sons.

Bloomfield, P. (1992). Trends in global temperatures. *Climatic Change* **21**, 1-16.

Bloomfield, P., and Nychka, D. (1992). Climate spectra and detecting climate change. *Climatic Change* **21**, 275-288.

Bloomfield, P., Oehlert, G., Thompson, M. L., and Zeger, S. (1983). A frequency domain analysis of trends in Dobson total ozone records. *Journal of Geophysics Research* **88**, 8512-8522.

Brown, P. J., Le, N. D., and Zidek, J. V. (1994). Multivariate spatial interpolation and exposure to air pollutants. *Canadian Journal of Statistics* **22**, 489-505.

Buishand, T. A. (1993). Rainfall depth-duration frequency curves: a problem of dependent extremes. In *Statistics for the Environment*, Barnett, V. and Turkman, K. F. (eds.), 183-197. New York: John Wiley & Sons.

Cicmanec, J. L., Dourson, M. L., and Hertzberg, R. C. (1996). Noncancer risk assessment: Present and emerging issues. In *Toxicology and Risk Assessment. Principles, Methods, and Applications*, Fan, A. M. and Chang, L. W. (eds.), 293-312. New York: Marcel Dekker.

Cliff, A. D., Haggett, P., Ord, J. K., Bassett, K. A., and Davies, R. B. (1975). *Elements of Spatial Structure: A Quantitative Approach.* New York: Cambridge University Press.

Coles, S. G. (1994). A temporal study of extreme rainfall. In *Statistics for the Environment 2. Water Related Issues*, Barnett, V. and Turkman, K. F. (eds.), 61-78. New York: John Wiley & Sons.

Coles, S. G., and Tawn, J. A. (1994). Statistical methods for multivariate extremes: an application to structural design (with discussion). *Applied Statistics* **43**, 1-48.

Consonni, G., and Veronese, P. (1995). A Bayesian method for combining results from several binomial experiments. *Journal of the American Statistical Association* **90**, 935-944.

Cothern, C. R., and Ross, N. P. (1994). *Environmental Statistics, Assessment, and Forecasting.* Boca Raton, FL: Lewis Publishers.

Cressie, N. A. C. (1993). *Statistics for Spatial Data.* New York: John Wiley & Sons.

Crump, K., and Howe, R. (1985). A review of methods for calculating confidence limits in low dose extrapolation. In *Toxicological Risk Assessment, Volume I: Biological and Statistical Criteria*, Clayson, D. B., Krewski, D. and Munro, I. (eds.), 187-203. Boca Raton, FL: CRC Press.

Currie, L. A. (1968). Limits for qualitative detection and quantitative determination: application to radiochemistry. *Analytical Chemistry* **40**, 586-593.

Davis, C. B. (1994). Environmental regulatory statistics. In *Handbook of Statistics Volume 12: Environmental Statistics*, Patil, G. P. and Rao, C. R. (eds.), 817-865. Amsterdam: North-Holland/Elsevier.

Davison, A. C., and Smith, R. L. (1990). Models for exceedances over high thresholds (with discussion). *Journal of the Royal Statistical Society, series B* **52**, 393-442.

Diggle, P. J. (1990). A point process modelling approach to raised incidence of a rare phenomenon in the vicinity of a prespecified point. *Journal of the Royal Statistical Society, series A* **153**, 349-362.

Diggle, P. J., and Rowlingson, B. S. (1994). A conditional approach to point process modelling of elevated risk. *Journal of the Royal Statistical Society, series A* **157**, 433-440.

Diggle, P. J., S, M., Elliot, P., and Shaddick, G. (1997). Regression modelling of disease risk in relation to point sources. *Journal of the Royal Statistical Society, series A* **160**, 491-505.

Dixon, M. J., and Tawn, J. A. (1994). Extreme sea levels: Modelling interaction between tide and surge. In *Statistics for the Environment 2: Water Related Issues*, Barnett, V. and Turkman, K. F. (eds.), 221-232. New York: John Wiley & Sons.

Dominici, F., Parmigiani, G., Reckhow, K. H., and Wolpert, R. L. (1997). Combining information from related regressions. *Journal of Agricultural, Biological, and Environmental Statistics* **2**, 313-332.

DuMouchel, W. (1994). Hierarchical Bayes linear models for meta-analysis. Technical Report number 27. National Institute of Statistical Sciences.

DuMouchel, W. M., and Harris, J. E. (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.

Duncan, G., and Kalton, G. (1987). Issues of design and analysis of surveys across time. *International Statistical Review* **55**, 97-117.

Egbert, G. D., and Lettenmeier, D. P. (1986). Stochastic modeling of the space-time structure of atmospheric chemical deposition. *Water Resources Research* **14**, 135-143.

El-Shaarawi, A., and Naderi, A. (1991). Statistical inference from multiply censored environmental data. *Environmental Monitoring and Assessment* **17**, 339-347.

Elliot, P., Martuzzi, M., and Shaddick, G. (1995). Spatial statistical methods in environmental epidemiology: a critique. *Statistical Methods in Medical Research* **4**, 137-159.

Esterby, S. R. (1996). Review of methods for the detection and estimation of trends with emphasis on water quality applications. *Hydrological Processes* **10**, 127-150.

Eynon, B. P. (1988). Statistical analysis of precipitation chemistry measurements over the eastern United States. Part II: Kriging analysis of regional patterns and trends. *Journal of Applied Meteorology* **27**, 1334-1343.

Eynon, B. P., and Switzer, P. (1983). The variability of rainfall acidity. *The Canadian Journal of Statistics* **11**, 11-24.

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

Freedman, D. A., and Zeisel, H. (1988). From mouse-to-man: The quantitative assessment of cancer risks (with discussion). *Statistical Science* **3**, 3-56.

Gibbons, R. D. (1994). *Statistical Methods for Groundwater Monitoring*. New York: John Wiley & Sons.

Gibbons, R. D. (1996). Some conceptual and statistical issues in analysis of groundwater monitoring data. *Environmetrics* **7**, 185-199.

Gomes, M. I. (1993). On the estimation of parameters of rare events in environmental time series. In *Statistics for the Environment*, Barnett, V. and Turkman, K. F. (eds.), 225-241. New York: John Wiley & Sons.

Gumbel, E. J. (1958). *Statistics of Extremes*. New York: Columbia University Press.

Günther, O. (1997). *Environmental Information Systems*. New York: Springer-Verlag.

Guttorp, P., Meiring, W., and Sampson, P. D. (1994). A space-time analysis of ground-level ozone. *Environmetrics* **5**, 241-254.

Guttorp, P., and Sampson, P. D. (1994). Methods for estimating heterogeneous spatial covariance functions with environmental applications. In *Handbook of Statistics Volume 12: Environmental Statistics*, Patil, G. P. and Rao, C. R. (eds.), 661-690. New York: North-Holland/Elsevier.

Guttorp, P., Sampson, P. D., and Newman, K. (1992). Nonparametric estimation of spatial covariance with application monitoring network evaluation. In *Statistics in the Environmental and Earth Sciences*, Walden, A. T. and Guttorp, P. (eds.), 39-57. London: Edward Arnold.

Haas, T. C. (1995). Local prediction of a spatio-temporal process with an application to wet sulfate deposition. *Journal of the American Statistical Association* **90**, 1189-1199.

Hallenbeck, W. H. (1993). *Quantitative Risk Assessment for Environmental and Occupational Health*. Boca Raton, FL: CRC Press.

Handcock, M. S., and Wallis, J. R. (1994). An approach to statistical spatial-temporal modeling of meteorological fields (with discussion). *Journal of the American Statistical Association* **89**, 368-390.

Haseman, J. K. (1984). Statistical issues in the design, analysis and interpretation of animal carcinogenicity studies. *Environmental Health Perspectives* **58**, 385-392.

Hasselblad, V. (1994). Meta-analysis in environmental statistics. In *Handbook of Statistics Volume 12: Environmental Statistics*, Patil, G. P. and Rao, C. R. (eds.), 691-716. New York: North-Holland/Elsevier.

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

Hegerl, G., von Storch, H., Hasselmann, K., Santer, B. D., Cubasch, U., and Jones, P. D. (1996). Detecting greenhouse-gas-induced climate change with an optimal fingerprint method. *Journal of Climate* **9**, 2281-2306.

Helsel, D. R., and Hirsch, R. M. (1992). *Statistical Methods in Water Resources*. Amsterdam: Elsevier.

Hewitt, C. N. (1992). *Methods of Environmental Data Analysis*. Amsterdam: Elsevier Applied Science.

Hoel, D. G. (1985). Mathematical dose-response models and their application to risk estimation. In *Methods for Estimating Risk of Chemical Injury: Human and Non-Human Biota and Ecosystems*, Vouk, V. B., Butler, G. C., Hoel, D. G. and Peakall, D. B. (eds.), 347-359. New York: John Wiley & Sons.

Høst, G., Omre, H., and Switzer, P. (1995). Spatial interpolation errors for monitoring data. *Journal of the American Statistical Association* **90**, 853-861.

Hsing, T., Hüsler, J., and Leadbetter, M. R. (1988). On the exceedance point process for a stationary sequence. *Probability Theory and Related Fields* **78**, 97-112.

Hunter, J. S. (1994). Environmetrics: An emerging science. In *Handbook of Statistics Volume 12: Environmental Statistics*, Patil, G. P. and Rao, C. R. (eds.), 1-7. New York: North-Holland/Elsevier.

Klemes, V. (1994). Statistics and probability: Wrong remedies for a confused hydrologic modeler. In *Statistics for the Environment 2. Water Related Issues*, Barnett, V. and Turkman, K. F. (eds.), 345-367. New York: John Wiley & Sons.

Kodell, R. L., and Park, C. N. (1995). Linear extrapolation in cancer risk assessment. In *Low-Dose Extrapolation of Cancer Risks: Issues and Perspectives*, Olin, S., Farland, W., Park, C., Rhomberg, L., Scheuplein, R., Starr, T. and Wilson, J. (eds.), 87-104. Washington, DC: ILSI Press.

Krewski, D., and Brown, C. (1981). Carcinogenic risk assessment: a guide to the literature. *Biometrics* **37**, 353-366.

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

Le, N. D., Sun, W., and Zidek, J. V. (1997). Bayesian multivariate spatial interpolation with data missing by design. *Journal of the Royal Statistical Society, series B* **59**, 501-510.

Le, N. D., and Zidek, J. V. (1992). Interpolation with uncertain spatial covariance: A Bayesian alternative to kriging. *Journal of Multivariate Analysis* **43**, 351-374.

Leadbetter, M. R. (1983). Extremes and local dependence in stationary sequences. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **65**, 291-306.

Leadbetter, M. R., Lindgren, G., and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Series*. New York: Springer-Verlag.

Leslie, P. H. (1945). On the use of matrices in certain population mathematics. *Biometrika* **33**, 183-212.

Lindgren, G., Zetterqvist, I., and Holmstedt, G. (1993). Estimation of quantiles in airborne pollution: Measuring pollution around gas furnaces. In *Statistics for the Environment*, Barnett, V. and Turkman, K. F. (eds.), 111-122. New York: John Wiley & Sons.

Little, M. P. (1995). Are two mutations sufficient to cause cancer? Some generalizations of the two-mutation model of carcinogenesis of Moolgavkar, Venzon, and Knudson, and of the multistage model of Armitage and Doll. *Biometrics* **51**, 1278-1291.

Loader, C., and Switzer, P. (1992). Spatial covariance estimation for monitoring data. In *Statistics in the Environmental and Earth Sciences*, Walden, A. T. and Guttorp, P. (eds.), 52-70. New York: Halsted Press.

Luebeck, E. G., and Moolgavkar, S. H. (1996). Biologically based cancer modeling. In *Toxicology and Risk Assessment. Principles, Methods, and Applications*, Fan, A. M. and Chang, L. W. (eds.), 533-555. New York: Marcel Dekker.

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

McCready, M., Patel, S., and Rennolls, K. (1997). An investigation of the effect of road traffic pollution on asthma, using geographical information systems. In *Statistics for the Environment 3: Pollution Assessment and Control*, Barnett, V. and Turkman, K. F. (eds.), 287-299. New York: John Wiley & Sons.

Meiring, W., Monestiez, P., Sampson, P. D., and Guttorp, P. (1997). Developments in the modelling of nonstationary spatial covariance structure from space-time monitoring data. In *Geostatistics Wollongong '96, Vol. 1*, Baafi, E. Y. and Schofield, N. (eds.), 162-173. Dordrecht, The Netherlands: Kluwer.

Meng, Z., Dabdub, D., and Seinfeld, J. H. (1997). Chemical coupling between atmospheric ozone and particulate matter. *Science* **277**, 116-119.

Moolgavkar, S. H., and Venzon, D. J. (1979). Two-event models for carcinogenesis: incidence curves for childhood and adult tumors. *Mathematical Biosciences* **47**, 55-77.

Niu, X., and Tiao, G. C. (1995). Modeling satellite ozone data. *Journal of the American Statistical Association* **90**, 969-983.

Oehlert, G. W. (1993). Regional trends in sulfate wet deposition. *Journal of the American Statistical Association* **88**, 390-399.

Olkin, I., Sacks, J., Blumstein, A., Eddy, A., Eddy, W., Jurs, P., Kruskal, W., Kurtz, T., McDonald, G. C., Peierls, R., Shaman, P., and Spurgeon, W. (1990). Cross-disciplinary research in the statistical sciences. *Statistical Science* **5**, 121-146.

Ott, W. R. (1995). *Environmental Statistics and Data Analysis*. Boca Raton, FL: CRC Press.

Patil, G. P., and Rao, C. R. (1994). *Handbook of Statistics Volume 12: Environmental Statistics*. Amsterdam: North-Holland/Elsevier.

Pearson, J. C. G., and Turton, A. (1993). *Statistical Methods in Environmental Health*. New York: Chapman & Hall.

Pickands, J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics* **3**, 119-131.

Piegorsch, W. W., and Cox, L. H. (1996). Combining environmental information II: Environmental epidemiology and toxicology. *Environmetrics* **7**, 309-324.

Portier, C. J. (1989). Quantitative risk assessment. In *Carcinogenicity and Pesticides. Principles, Issues, and Relationships*, Ragsdale, N. N. and Menzer, R. E. (eds.), 164-174. Washington, DC: American Chemical Society.

Portier, C. J., and Kopp-Schneider, A. (1991). A multistage model of carcinogenesis incorporating DNA damage and repair. *Risk Analysis* **11**, 535-543.

Redmond, C. K. (1991). Introduction: Methods for environmental risk assessment. *Environmental Health Perspectives* **90**, 157.

Reed, D. W., and Stewart, E. J. (1994). Inter-site and inter-duration dependence on rainfall extremes. In *Statistics for the Environment 2. Water Related Issues*, Barnett, V. and Turkman, K. F. (eds.), 125-146. New York: John Wiley & Sons.

Resnick, S. (1987). *Extreme Values, Point Processes and Regular Variation*. New York: Springer-Verlag.

Rivoirard, J. (1994). *Introduction to Disjunctive Kriging and Non-linear Geostatistics*. Oxford: Oxford University Press.

Samet, J. M., Zeger, S. L., and Berhane, K. (1995). The association of mortality and particulate air pollution. In *Particulate Air Pollution and Daily Mortality: Replication and Validation of Selected Studies. The Phase I Report of the Particle Epidemiology Evaluation Project* 1-104. Cambridge MA: Health Effects Institute.

Samet, J. M., Zeger, S. L., Kelsall, J. E., and Xu, J. (1997). Air Pollution, Weather and Mortality in Philadelphia, 1973-1988. In *Particulate Air Pollution and Daily Mortality: Analyses of the Effects of Weather and Multiple Air Pollutants. The Phase IB Report of the Particle Epidemiology Evaluation Project* 1-29. Cambridge MA: Health Effects Institute.

Sampson, P. D., and Guttorp, P. (1992). Nonparametric estimation of nonstationary spatial covariance structure. *Journal of the American Statistical Association* **87**, 108-119.

Santer, B. D., Taylor, K. E., Wigley, T. M. L., Johns, T. C., Jones, P. D., Karoly, D. J., Mitchell, J. F. B., Oort, A. H., Penner, J. E., Ramaswamy, V., Schwarzkopf, M. D., Stouffer, R. J., and Tett, S. (1996). A search for human influences on the thermal structure of the atmosphere. *Nature* **382**, 39-46.

Scheiner, S. M., and Gurevitch, J. (1993). *The Design and Analysis of Ecological Experiments*. New York: Chapman & Hall.

Schoenfeld, D. A. (1986). Confidence bounds for normal means under order restrictions, with application to dose-response curves, toxicology experiments, and low-dose extrapolation. *Journal of the American Statistical Association* **81**, 186-195.

Schwartz, J. (1993). Air pollution and daily mortality in Birmingham, Alabama. *American Journal of Epidemiology* **137**, 1136-1147.

Schwartz, J., and Marcus, A. (1990). Mortality and air pollution in London: A time series analysis. *American Journal of Epidemiology* **131**, 185-194.

Seber, G. A. F., and Thompson, S. K. (1994). Environmental adaptive sampling. In *Handbook of Statistics Volume 12: Environmental Statistics*, Patil, G. P. and Rao, C. R. (eds.), 201-220. New York: North-Holland/Elsevier.

Smith, R. L. (1987). Estimating the tails of probability distributions. *Annals of Statistics* **15**, 1174-1207.

Smith, R. L. (1989). Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone. *Statistical Science* **4**, 367-393.

Smith, R. L. (1993). Long-range dependence and global warming. In *Statistics for the Environment*, Barnett, V. and Turkman, K. F. (eds.), 141-161. New York: John Wiley & Sons.

Smith, R. L., and Chen, F.-L. (1996). Regression in long-memory time series. In *Athens Conference on Applied Probability and Time Series, Volume II: Time Series Analysis in Memory of E. J. Hannan*, Robinson, P. M. and Rosenblatt, M. (eds.), 378-391. New York: Springer-Verlag.

Smith, R. L., Davis, J. M., Sacks, J., Speckman, P., and Styer, P. E. (1997a). Air pollution and daily mortality in Birmingham, Alabama: A Reappraisal. Technical Report. National Institute of Statistical Sciences.

Smith, R. L., Davis, J. M., and Speckman, P. (1998). Assessing the human health risk of environmental particulates. In *Statistics for the Environment 4*, Barnett, V., Turkman, K. F. and Stein, A. (eds.), (to appear). Chichester: John Wiley & Sons.

Smith, R. L., and Huang, L.-S. (1994). Modeling high threshold exceedances of urban ozone. Technical Report number 6. National Institute of Statistical Sciences.

Smith, R. L., and Shively, T. S. (1995). A point process approach to modeling trends in tropospheric ozone. *Atmospheric Environment* **29**, 3489-3499.

Smith, R. L., Tawn, J. A., and Coles, S. G. (1997b). Markov chain models for threshold exceedances. *Biometrika* **84**, 249-268.

Smith, R. L., and Weissman, I. (1994). Estimating the extremal index. *Journal of the Royal Statistical Society, series B* **56**, 515-528.

Solow, A. R. (1994). Statistical methods in atmospheric science. In *Handbook of Statistics Volume 12: Environmental Statistics*, Patil, G. P. and Rao, C. R. (eds.), 717-734. New York: North-Holland/Elsevier.

Solow, A. R., and Polasky, S. (1994). Measuring biological diversity (with discussion). *Environmental and Ecological Statistics* **1**, 95-107.

Speed, T. (1993). Modelling and managing a salmon population. In *Statistics for the Environment*, Barnett, V. and Turkman, K. F. (eds.), 267-292. New York: John Wiley & Sons.

Stehman, S. V., and Overton, W. S. (1994). Environmental sampling and monitoring. In *Handbook of Statistics Volume 12: Environmental Statistics*, Patil, G. P. and Rao, C. R. (eds.), 263-306. New York: North-Holland/Elsevier.

Stone, R. A. (1988). Investigations of excess environmental risks around putative sources: statistical problems and a proposed test. *Statistics in Medicine* **7**, 649-660.

Styer, P. E., McMillan, N., Gao, F., Davis, J., and Sacks, J. (1995). Effect of outdoor airborne particulate matter on daily death counts. *Environmental Health Perspectives* **103**, 490-497.

Tawn, J. A. (1993). Extreme sea levels. In *Statistics for the Environment*, Barnett, V. and Turkman, K. F. (eds.), 243-263. New York: John Wiley & Sons.

Tawn, J. A., Dixon, M. J., and Woodworth, P. L. (1994). Trends in sea levels. In *Statistics for the Environment 2: Water Related Issues*, Barnett, V. and Turkman, K. F. (eds.), 147-181. New York: John Wiley & Sons.

Urquhardt, S., Overton, W. S., and Irkes, D. S. (1993). Comparing sampling designs for monitoring ecological status and trends: Impact of temporal patterns. In *Statistics for the Environment*, Barnett, V. and Turkman, K. F. (eds.), 71-85. New York: John Wiley & Sons.

van der Hoeven, N. (1997). How to measure no effect. Part III: Statistical aspects of NOEC, EC*x* and NEC estimates. *Environmetrics* **8**, 255-261.

van Leeuwen, D. M., Murray, L. W., and Urquhardt, N. S. (1996). A mixed model with both fixed and random trend components across time. *Journal of Agricultural, Biological, and Environmental Statistics* **1**, 435-453.

Vrijling, J. K. (1994). Sea-level rise: A potential threat. In *Statistics for the Environment 2. Water Related Issues*, Barnett, V. and Turkman, K. F. (eds.), 183-204. New York: John Wiley & Sons.

Wahba, G. (1990). *Spline Models for Observational Data*. Philadelphia: Society for Industrial and Applied Mathematics.

Walden, A. T., and Guttorp, P. (1992). *Statistics in the Environmental and Earth Sciences*. New York: Halsted Press.

Warren-Hicks, W. J., and Wolpert, R. L. (1994). Predictive models of fish response to acidification: Using Bayesian inference to combine laboratory and field measurements. In *Environmental Statistics, Assessment, and Forecasting*, Cothern, C. R. and Ross, N. P. (eds.), 99-112. Boca Raton, FL: Lewis Publishers.

Yajima, Y. (1991). On estimation of a regression model with long-memory stationary errors. *Annals of Statistics* **19**, 158-177.

Zidek, J. V. (1997). Interpolating air pollution for health impact assessment. In *Statistics for the Environment 3: Pollution Assessment and Control*, Barnett, V. and Turkman, K. F. (eds.), 251-268. New York: John Wiley & Sons.

TABLE 1

N = 50 observations (ordered) from a lognormal distribution with mean $e^{1/2} = 1.649$ and variance $e^2 - e = 4.671$, from Akritas et al. (1994); asterisks indicate MDL at 0.4

*0.1007	*0.1113	*0.1167	*0.1848	*0.2531	*0.2621	*0.2747	*0.2915
*0.3529	*0.3629	0.4014	0.4136	0.4154	0.4182	0.4443	0.4696
0.5282	0.5386	0.5648	0.6152	0.6779	0.7485	0.7508	0.8170
1.1029	1.1210	1.1938	1.3856	1.4381	1.4526	1.5144	1.5220
1.5470	1.5472	1.6761	1.7095	2.0705	2.2106	2.2118	2.2966
2.4249	2.4444	2.4615	2.7230	2.7461	3.3692	4.4418	4.8487
			4.8613	6.9258			

TABLE 2

$N - n_C = 10$ non-detected observations (ordered) from Table 1 assuming the MDL is 0.4, along with estimated values from Akritas et al. (1994)

actual values	probability plot regressed	0	$\frac{1}{2}$MDL	MDL
0.1007	0.1040	0.0	0.2	0.4
0.1113	0.1434	0.0	0.2	0.4
0.1167	0.1765	0.0	0.2	0.4
0.1848	0.2068	0.0	0.2	0.4
0.2531	0.2357	0.0	0.2	0.4
0.2621	0.2639	0.0	0.2	0.4
0.2747	0.2917	0.0	0.2	0.4
0.2915	0.3194	0.0	0.2	0.4
0.3529	0.3472	0.0	0.2	0.4
0.3629	0.3754	0.0	0.2	0.4

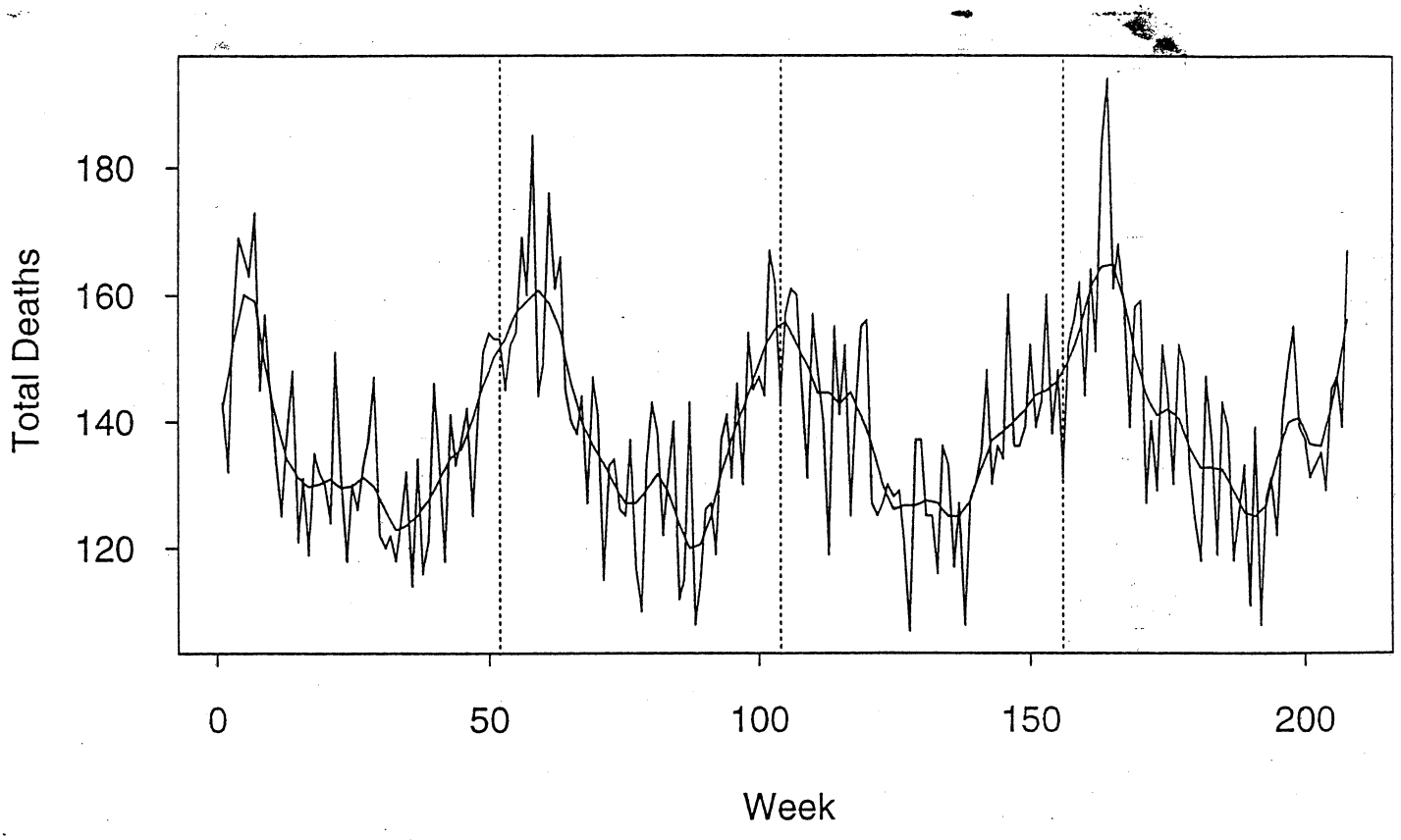

FIG. 1. *Birmingham weekly deaths for 1985-1988 together with smoothed curve obtained via LOESS fit. The dotted vertical lines denote the ends of each year (weeks 52, 104, 156).*

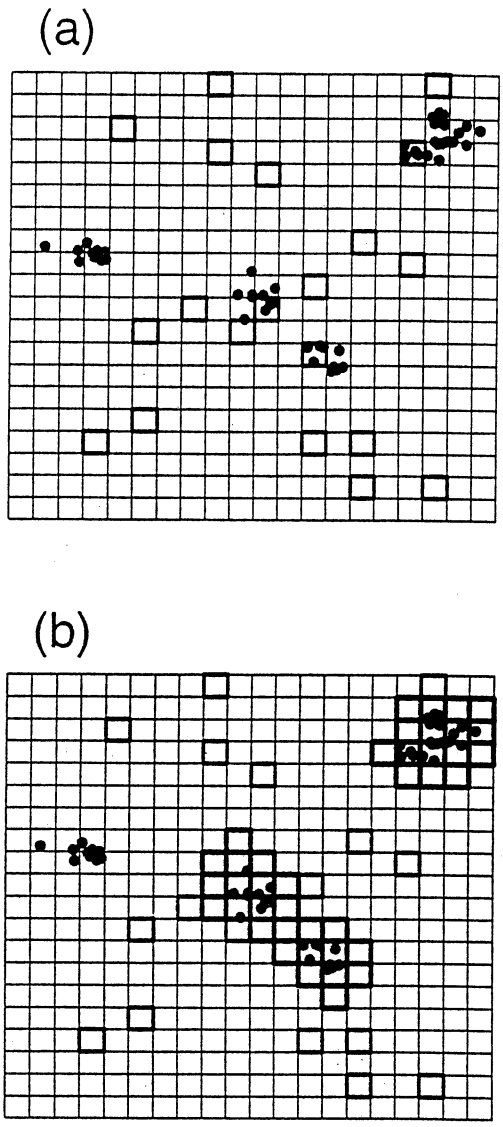

FIG. 2. *An illustration of adaptive sampling for hot spot identification with a 400-unit population and a simple random sample of size 20: After initial sampling (a), the 4 nearest neighbors of any contaminated sample unit are also inspected (b).*