Appendix

STATISTICAL TREND DETECTION AND ANALYSIS METHODS[5]

TRENDS IN ANNUAL SUMMARY STATISTICS—PARAMETRIC METHODS

This section discusses some proposed classical methods for evaluating trends in annual summary statistics. Many analyses require that either the consecutive concentrations or the corresponding annual summary statistics are approximately normally distributed and independent. Various tests of these assumptions are applied to the estimated *residuals*: the observed concentration or annual summary statistic minus the estimated mean (predicted by the statistical model). Since means are derived from data, residuals are not exactly independent even if the model exactly describes the underlying distributions, but the proposed tests should work reasonably well. Since these tests use general linear models and assume normality of the concentrations or annual summary statistics, it follows that residuals will be normally distributed if the statistical model is correct.

Testing Independence

We describe two statistical methods for testing independence of residuals. The first, a nonparametric test, requires no further major assumptions about the data (in particular, it is applicable even if residuals are not normally distributed). The second requires that residuals be normally distributed. In general, the normality assumption is less important than the independence assumption as most of the proposed procedures give reasonable estimates of the trend even if the normality assumption is violated provided (a) the true distribution is not excessively skewed and (b) the independence assumption holds. The trend estimates will be quite poor if the independence assumption is too severely violated.

Before applying these procedures, the first step is to examine a time series plot of the residuals; the presence of clear patterns in the plot indicates significant departures from independence. For example, the residuals are not independent if the series tends to alternate between clusters of large positive residuals and large negative residuals. Unfortunately, this can also occur if the regression model (the assumed relation between the mean concentration or annual summary statistic and the regression terms such as site and year) is incorrect. It is not always possible to distinguish which features defining the statistical model are representative of real world conditions and which are not.

[5] Some of the material in this appendix has been adapted from Cohen and Stoeckenius (1992).

Runs Test of Independence

The runs test of independence is a nonparametric method used to test the assumption of independence for data from a single site. It is not applicable when evaluating combined data from more than one site as the test statistic measures temporal rather than spatial dependence. Its advantage over the autocorrelation test described next is that specific parametric assumptions are not needed when applying the test.

Consecutive residuals are given plus or minus signs according to whether or not they exceed the median. The runs test is then based on the number of changes of sign in the series. A very small number of sign changes indicates a strong positive dependence between consecutive residuals; that is, both values tend to be high or low. A very large number of sign changes indicates a strong negative dependence; that is, high values tend to be followed by low values and vice versa. Thus if the number of sign changes is extreme in either direction, there is strong statistical evidence of dependence.

Standard texts give the theoretical mean M and standard deviation S of the number of runs calculated under the null hypothesis that consecutive observations are independent. It can also be shown that for large N the distribution of the number of runs is approximately normal. Thus the p-value of a given number of runs is approximately given by the tail area of a standard normal distribution that exceeds the standardized test statistic, |observed runs - M|/S. Equivalently, the null hypothesis of independence is rejected at the 5 percent level if this standardized test statistic exceeds $z_{0.025}$ (the value exceeded with a probability of 0.025 by a standard normal distribution). Such a case is "statistical evidence of dependence." (More exact probability calculations are available for small samples but are harder to implement.)

Autocorrelation Test of Independence

As an alternative to the runs test of independence we also propose the autocorrelation test to consecutive residuals from the same site. This test is based on values of the correlation coefficients calculated from all pairs of data points separated by k time periods, where k, the lag time, is between 1 and 5. Such autocorrelation coefficients between consecutive values of the same variable are called autocorrelation coefficients. Following standard practice, we shall use the following definition of $\hat{o}(k)$, the estimated autocorrelation coefficient between consecutive values k time periods apart:

$$\hat{o}(k) = \frac{\sum_{t=1}^{N-k} (X_t - \overline{X})(X_{t+k} - \overline{X})}{\sum_{t=1}^{N} (X_t - \overline{X})^2}$$

where $\overline{X}$ denotes the arithmetic mean of all N values (observations, predictions, or residuals). the value of $\hat{o}(k)$ will lead to zero as the sample size N tends to infinity, if the consecutive values are independent. Otherwise, the limiting value is between -1 and +1. Thus, large values of $|\hat{\sigma}(k)|$ are evidence of dependence.

The statistical test of independence uses the standardized autocorrelation coefficient, defined as the autocorrelation coefficient $\hat{\sigma}(k)$ divided by an estimate of its standard deviation. We propose the use of the estimate of standard deviation given by

$$Variance\ [\hat{\sigma}(k)] = \frac{n - k}{n(n + 2)}$$

This formula gives the exact variance of $\hat{\sigma}(k)$ in the case where the X observations are independent and normally distributed, and where the true mean replaces the sample mean, $\bar{X}$, in the formula for $\hat{\sigma}(k)$. The approximate statistical significance of the standardized correlation coefficient is found from standard normal tables, assuming normally distributed data.

Testing for Normality

If the above procedures show that the residuals are approximately independent then various procedures can be used to test for normality. It is extremely important to apply these tests to the residuals and not the concentrations or annual summary statistics, because any trend in the mean would invalidate these procedures. These methods assume that all the data (i.e. residuals) come from a common population and therefore have a constant mean and variance. If the overall statistical model assumes that a transformation of the raw data (e.g. a logarithmic transformation) is required to convert them to normality, then it will be necessary to calculate the residuals from the transformed data.

The simplest approach and an important first step is to prepare a histogram of the residuals that gives the frequencies of each possible value, grouped into class intervals of equal sizes. If the residuals are normally distributed then the histogram should approximately follow the classic bell-shaped normal curve. More quantitative procedures are described below: the Kolmogorov-Smirnov test of normality and the Shapiro-Wilks test. Note that if the normality test indicates a departure from normality then it may be that the statistical model is correct except that the error distribution is not normal, or that the statistical model and normality is correct but the data are not independent, or that the statistical model (regression terms) is wrong. If only the normality of the error distribution is wrong, then it may be desirable to use the proposed non-parametric procedures. If only the independence assumption is wrong (see previous section) then it will be appropriate to apply the time series methods discussed below.

Kolmogorov-Smirnov Test of Normality

For each residual v in the data set, the distribution function, F(v), is defined as the fraction of values less than or equal to v:

$$F(v) = (\text{Number of residuals} \leq v)/(\text{Number of residuals})$$

Correspondingly, using statistical tables, we can also calculate the distribution function G(v) for an infinite sample from the hypothetical normal distribution (with the same mean and variance as in the data).

The Kolmogorov-Smirnov test is based on the maximum difference between F(v) and G(v) as v varies across all possible values. The greater the observed maximum difference, the greater the evidence against the normality assumption. The critical value of the test statistic (exceeded with a 5 percent probability if normality holds) can be approximated using standard asymptotically valid formulae.

Shapiro-Wilks Test of Normality

Another standard test of normality is the Shapiro-Wilks test described in many standard statistical texts and in the SAS software manuals (SAS procedure UNIVARIATE). This test is based on the ratio of the optimal estimate of the sample variance based on squaring a linear combination of order statistics to the sample variance. For large samples the test is asymptotically equivalent to plotting the data on a normal probability plot and calculating the Pearson correlation coefficient.

Simple Linear Regression

One of the simplest methods of trend detection and estimation is simple linear regression (Chock et al., 1982; Kumar and Chock, 1984; Wackter and Bayly, 1987; SCAQMD, 1991; EPA, 1974). In this method the expected annual summary statistic is assumed to be linear in the year, so the estimated means plot as a straight line. This statistical model is usually fitted by least squares, which is defined as finding the straight line such that the squared errors about that line (squared differences between the observed annual summary statistic and the estimate from the straight line trend) are minimized. That method is most appropriate when the annual summary statistics are normally distributed with a constant variance. For the set of annual summary statistics from a single site rather than averages across multiple sites and/or multiple days, the assumption of normality may be less tenable than an assumption of log-normality, so one can apply simple linear regression to the logged concentrations and then transform from the straight line in log space back to an exponential curve for the mean concentrations (Chock et al., 1982, Kumar and Chock, 1984).

General Linear Regression

The simplest approach uses the set of annual summary statistics from a single site and fits a linear trend using simple linear regression, as described above. To estimate polynomial or other trend curves at a single site, other terms are used in the linear regression model. For example, instead of expressing the annual mean as a straight line function of the calendar year, higher powers of the calendar year can be added to the regression equation and hence a quadratic or higher order polynomial trend curve can be estimated.

Since the above method only uses the annual summary statistics it is not a very powerful method for detecting the trend at a single site (i.e., there is a high probability that the slope or trend curve will not be found to be significantly different from the case of no trend - the same mean every year).

Use of Raw Concentration Data

If the summary statistic is an annual mean of the consecutive concentrations, and certain assumptions are made, then much more powerful statistical techniques can be used to estimate the trend at a single site. The simplest analysis assumes that the consecutive measurements within an ozone season (or PAMS monitoring season) are approximately independent and have the same mean and variance; under those assumptions the trend in the annual mean can be evaluated by fitting the regression model to the raw data instead of the annual means. Analyses taking into account the possible dependencies between consecutive measurements at the same site are discussed in the section on time series analyses.

Multiple Sites

Various extensions of the simple and general linear regression approach can be used to evaluate overall trends at multiple sites. Since emissions control measures will generally have different effects on ambient measurements in different locations (depending on where the emissions reductions occur as well as meteorological factors), it would not be expected that trends in VOC, VOC species, air toxics, and ozone would be exactly the same at every PAMS site but they might be reasonably similar. If data from several sites can be combined into one calculation of the estimated trend (assumed consistent across those sites), then the estimate will be much more accurate than a trend estimate calculated from data at a single site. This type of analysis will be extremely useful in the early years of PAMS data collection since there will not be enough data at each site to get very reliable site-specific trend estimates.

The simplest method averages the annual summary statistics across all the sites and estimates the trends in these spatial averages using linear regression with one value for each year. Since the method is simple and requires few additional assumptions, we recommend using this approach to get a quick and useful overall estimate of the trend. The method does not require assumptions about the variation of the mean or trend from site to site. The method also has the advantage of being quite robust, in the sense that the mean across a large number of sites can be expected to follow the normal distribution fairly closely even if the site means do not; this follows from the central limit theorem.

A slightly more complicated method applies the same regression model to the data set consisting of the annual statistic for each site and year. Since a site effect is not assumed, this method effectively assumes that the underlying mean is the same at every site. This assumption is less tenable than the assumption of a consistent trend across sites and therefore this regression approach is not recommended except in certain situations. The approach is reasonable for the analysis of nearby sites known to have similar ambient concentration levels. The approach also may be useful in the early years of the PAMS program, where there may be insufficient data to accurately estimate separate means for each site. Usually a better approach is the two-way analysis of variance method described next.

Two-Way Analysis of Variance

A very reasonable approach to the analysis of data from several sites assumes that the annual summary statistic for a given site and year varies both with the site and year. For example, if annual summary statistics are available for every site, two-way analysis of variance can be used (Pollack and Stocking, 1989; Cohen and Pollack, 1991; Capel et al., 1983; Pollack et al., 1984; Pollack and Hunt, 1985; EPA, 1984-1991). The two-way analysis of variance method assumes that the annual summary statistic for a given site and year is the sum of a site effect, a year effect, and a normally distributed error term. In particular, this approach does not assume a specific trend curve since the form of the year effect is not specified. The errors are assumed to be independent, with mean zero and a constant variance. The main advantage of this method is that it accounts for the dependence between annual composite site averages caused by site effects. A significant trend determined from the two-way analysis of variance corresponds to a statistically significant year effect.

The two-way analysis of variance method can also be applied in cases where some annual summary statistics are missing for some sites (often because the available data are insufficient to satisfy the validity criteria for the annual summary statistic). In this case the approach is equivalent to using the general linear model on the available data to estimate the site and year effects, and then estimating any missing value as the sum of the estimated site and year effects corresponding to the particular missing site and year (Pollack and Stocking, 1989). One disadvantage of this method is that data from all sites are used to fill in the missing values, rather than data from local sites. Cohen and Pollack (1991) provide an extension of the two-way analysis of variance approach that deals with this problem by allowing the year effects to depend on the region. Regions are defined by combining nearby sites in such a way that the mean squared error in the fitted general linear model is minimized.

In the two-way analysis of variance approach, a trend is indicated by statistically significant differences between the mean annual summary statistics for a pair of years. An often useful plot is given by graphing the annual means with simultaneous confidence intervals defined such that two means are significantly different if the corresponding confidence intervals do not overlap (Pollack and Stocking, 1989; Pollack et al., 1984; Pollack and Hunt, 1985; EPA, 1984-1991). For example, Figure A-1 (from Pollack et al., 1984) illustrates simultaneous confidence intervals for four years of data. Since the plotted confidence intervals overlap for years 1 and 2 but not for years 1 and 3, years 1 and 2 are not significantly different, but years 1 and 3 are significantly different.

The Tukey studentized range technique used to derive the simultaneous confidence intervals in Figure A-1 was described by Pollack and others (1984; Pollack and Hunt, 1985). Since the number of possible simultaneous comparisons is k(k-1)/2, where k is the often large number of years of data analyzed, testing each pair at the usual 5 percent significance level would lead to a high probability that at least one difference would be declared significant when in fact there are no real trends (since each comparison would then have a 5 percent probability of being declared statistically significant). To treat this problem, the confidence intervals were computed in such a way that if there is no real trend, the probability that every pair of annual confidence intervals overlaps is 95 percent. Thus the probability of erroneously determining one or more significant differences is 5 percent. The same Tukey studentized range technique is applicable in other cases where a general linear model is used to estimate the year effects.

One important problem with the two-way analysis of variance method is that significant year effects may be due to almost entirely to differences in meteorological conditions for different years. Thus these year effects may not be due to long term emissions trends. One way of dealing with this problem is to adjust the annual summary statistics to account for possible annual meteorological differences before fitting the linear model. Such an adjustment can be based on a statistical regression analysis simply by adding additional terms representing meteorological factors to the site and year effects; for example a term giving the annual mean summer temperature might be very useful for an analysis of ozone data.

The two-way analysis of variance approach can be modified to allow for specific trend functions by assuming that the year effects are certain functions of the year, but the site effects are arbitrary. The case of a linear trend but arbitrary site effects is used by Capel and others (1983). One major advantage of this modified approach is that the year effects will then probably not be confounded with annual meteorological conditions, since any long term trends in meteorology are likely to be negligible compared to trends in emissions.

Other General Linear Model Approaches

The above set of analyses are important examples of the general linear model approach to estimating and testing for trends. These analyses can be extended in various ways that are far too numerous to be described here. In a similar manner to the time series approaches described below, the regression models fitted to the annual summary statistics can be enhanced by the addition of various explanatory variables, such as annual summary statistics for meteorology or emissions, or terms to represent abrupt step changes in the annual summary statistics. Other enhanced analyses use multivariate methods that combine the set of measurements at different sites, or for different pollutants, into a single vector and thus explicitly model the correlations between the concentrations at different sites or for different pollutants.

If the consecutive measurements at a given site within the same PAMS measurement season can be assumed to be approximately independent then more powerful statistical methods can be used to estimate the trends by fitting a general linear model to these raw concentrations rather than the annual means. The general linear model can then explicitly take into account the effects of the variation in the mean during the week (for CO, for example) or from month to month during the ozone season by adding appropriate terms to the fitted general linear model. Another possibility is to assume that the mean concentration on the *i*th day of year y is equal to a multiple of 365y + i; this formula recognizes the fact that if there is a linear trend in the annual mean, then it is plausible that the mean at the beginning of the PAMS season will be different to the mean at the end of the season and will slowly decrease during the measurement period (after adjusting for the effects of the month and the day of the week.)

Trend Detection Probabilities

An important issue for air quality managers is the probability that a trend in the annual summary statistics can be detected using an appropriate statistical analysis of the PAMS data, since this gives the managers some measure of the effectiveness of these data. The calculated probability of detecting a trend depends upon the assumptions that are made about the statistical distributions of the ambient concentration data and the amount of valid data that are collected each year. Such a calculation was made using speciation data collected in the Summer 1990 Atlanta Ozone Study (Cohen and Stoeckenius, 1982) where it was found that a three-percent annual trend in early morning mean total VOC could be detected with a 60 percent probability assuming 90 days of valid data per year at 4 sites. Similar results were found for specific VOC species (e.g. benzene and acetylene) and for daily means. Note that these calculations considered the question of detecting ANY non-zero linear trend in regions with similar ambient concentration distributions to Atlanta assuming that the true means decreased by three percent per year. The probability of detecting a specific non-zero trend (e.g. two percent or greater) will be even lower.

NONPARAMETRIC METHODS

The assumptions of independent normally distributed errors with the same constant variance used in several of the trend analyses described above may not be applicable. In order to test for and/or estimate the trend without making such distributional assumptions, various nonparametric methods can been used. Nonparametric methods proposed and applied in the literature for trend analysis include the use of Spearman's rho (Kolaz and Swinford, 1988, 1989; Sweitzer and Kolaz, 1984; Lettenmaier, 1976; EPA, 1974), and Kendall's tau (Hirsch et al., 1982; Hirsch and Slack, 1984; Freas and Sieurin, 1977) for trend detection, and the use of the Theil/Sen slope estimator for linear trend estimation (Hirsch et al., 1982; Freas and Sieurin, 1977). Since only very general distributional assumptions are made (concerning the dependence structure), the results are valid under very general conditions but the methods will have lower power (trend detection probability) compared to parametric tests in cases where the parametric model assumptions are reasonable approximations. The nonparametric tests can have much greater power than parametric tests when the distributional requirements of the parametric test are violated (Lettenmaier, 1976).

The main advantages of the non-parametric procedures over parametric alternatives are that the procedures can be used without making too many assumptions about the underlying concentration distributions and. in many cases, their relative simplicity. One disadvantage of these approaches is the relatively low power (i.e. a low probability of detecting a trend) in cases where the assumptions for a corresponding parametric test are reasonable. Another disadvantage for some of these procedures is that the non-parametric test may only be able to determine whether a statistically significant trend exists (trend detection) and cannot determine the size of the trend. Obviously an air quality manager will usually prefer to have an estimate of the trend. For parametric tests the results can usually be expressed as a confidence interval for the trend. If the confidence interval does not contain zero then a non-zero trend has been detected. Several nonparametric methods (such as the Spearman's rho and chi-square tests described below and in EPA, 1974) are used for trend detection but cannot be used for trend estimation.

In this description we include some of the more standard non-parametric approaches that have been used in the past for trend analysis and could easily be adapted to the PAMS data. These analyses use data from a single site and do not combine results from different sites. A wide variety of other more general non-parametric tests applicable to data from multiple sites could be derived. A general approach that often produces reasonable non-parametric analyses from a parametric analysis is to apply a standard test based on a fitted general linear model but apply it to the ranks instead of the raw data. Thus any of the general linear model approaches described above based on annual summary statistics could be converted into non-parametric procedures by replacing the highest annual summary statistic by 1, the next highest by 2, and so on. A similar approach could be used for the general linear models applied to the daily summary statistics.

Spearman's Rho Test of Trend

The Spearman's rho test of trend (Kolaz and Swinford, 1988, 1989; Sweitzer and Kolaz, 1984; Lettenmaier, 1976; EPA, 1974) is based on Spearman's rho statistic, which is the standard Pearson correlation coefficient between the rank of the annual summary statistics and the year. The rank is 1 for the highest summary statistic, 2 for the second highest, and so on. If there is no trend and all observations are independent, then all rank orderings are equally likely. This fact is used to calculate the statistical significance of the Spearman's rho statistic; a value significantly different from zero implies a significant trend. If ties in the annual summary statistics are present, then the significance level has to be adjusted to account for the number of ties. In paper 12 a comparison of the power (trend detection probability) of Spearman's rho with the power of simple linear regression shows that the nonparametric test can be almost as efficient as the simple linear regression t test even when the normality assumption holds. The linear regression power calculations in (Lettenmaier, 1976) are based on formulae that are incorrect for small samples but approximately correct for large samples (see formula 3b in Lettenmaier, 1976). Thus the reported results in that paper may be inaccurate for small samples and should be used with caution.

Kendall's Tau Test of Trend

Kendall's tau is an alternative nonparametric statistic that can be used to test for trend (Hirsch et al., 1982; Hirsch and Slack, 1984; Freas and Sieurin, 1977). This statistic can be calculated as the number of possible pairs of years for which the ordering of the years is the same as the ordering of the annual summary statistics (the lower annual statistic occurs in the earlier year) less the number of possible pairs of years with the reverse ordering. If there is no trend and all observations are independent, then all rank orderings of the annual statistics are equally likely; this result is used to compute the statistical significance of the tau statistic. Adjustments for tied annual summary statistics are described in the cited articles.

Adjustments of Kendall's tau for seasonality (Hirsch et al., 1982) and serial dependence (Hirsch and Slack, 1984) have been proposed and investigated in the context of water quality data analysis. A seasonally adjusted Kendall's tau (Hirsch et al., 1982) allows for different annual means and trends in different calendar months by adding up the 12 Kendall's tau statistics from each month. In paper 13 the null distribution of this statistic (when there are no trends) is calculated assuming values from different calendar months are independent. In paper 14 the null distribution is calculated assuming values in different months can be correlated. Both papers include calculations of the power of these tests for simulated data. The power of the seasonal and serial dependence adjusted Kendall's tau is greater than the power of the simpler seasonal dependence adjusted Kendall's tau if there is serial dependence, but is less in the independent case.

Kendall's tau test of trend is related to the Theil/Sen non-parametric slope estimator (Freas and Sieurin, 1977), which gives an estimate of the assumed linear trend. This estimator is the median of all possible ratios of the change in the annual summary statistic from one year to a later year divided by the number of years separating the two values. If the trends differ by calendar month, then the same calculation can be applied to the monthly summary statistics by only considering ratios for values in the same month, i.e., that differ by an exact multiple of 12 months (Hirsch et al., 1982).

TIME SERIES MODELS

Most of the above procedures require that consecutive concentrations or annual summary statistics are approximately independent. This assumption can be tested using the procedures described above under "Tests of Independence." Usually this issue is most important when the raw concentration data are analyzed, since measurements separated by shorter time intervals are generally more likely to be dependent. Note that we propose to analyze daily summary statistics calculated from the hourly or three-hourly PAMS data, rather than dealing with the raw hourly or three hourly data. This approach avoids technical difficulties associated with the fact that the PAMS monitoring scheme allows for multiple samples on a given day but sampling need not be every day (it can be every third or every sixth day).

One simple approach to treating the serial dependence between consecutive measurements is to decrease the sampling frequency by using only every second, third, fourth,.. measurement. This simple approach has the disadvantage of throwing away valuable information, but if the autocorrelation is very high then the amount of additional information in the dropped measurements will be small. A useful rule of thumb to determine how much data to drop is found using the autocorrelation coefficients calculated in the autocorrelation test of independence described above. If the lag k correlation is the first non-significant correlation then one might use every k sample values and assume those values are approximately independent.

A technically much better approach, but one that requires significantly more sophisticated analysis and computation, is to fit a time series model that explicitly takes the dependence into account. In this discussion we mainly present models based on normally distributed data although the statistical literature does include methods for dealing with more general error distributions.

A wide variety of different statistical models can be used for these analyses, allowing for different annual trend functions, different dependencies between consecutive measurements, the inclusion of day of the week, seasonal, and/or meteorological factors, the inclusion of spatial dependencies (for data at different monitoring sites), and the inclusion of intervention terms to account for relatively abrupt step changes in the ozone and ozone precursor concentrations. In fact most of the models described above in the subsection "Other general linear model approaches" could be analyzed using a time series approach by adding to the model the error auto-correlations (the general linear model approach assumes independent, and hence, uncorrelated errors). The biggest limitations are a) the availability of software for the proposed analysis, and b) the need to consider and compare a large number of possible time series models.

For most trend analyses the time domain approach to time series analysis is more appropriate than a frequency domain approach, which would analyze the series by examining the inherent periodicities. The auto-regressive integrated moving average (ARIMA) modeling approach is a general model that is likely to encompass the dependencies in the PAMS data. The simplest case, conceptually, is an autoregressive process with regression terms. Each daily summary statistic is assumed to be the sum of regression terms plus an autoregressive error term. The regression terms can represent the annual trend, the month to month variation within a year, variation within a week, site effects, meteorological measurements, and similar explanatory variables. The autoregressive error term is generated from a statistical model which assumes that each value is an error term plus a constant times the previous value, another constant times the second previous value, and so on up to a certain lag. The error terms are assumed to be independently drawn from a normal distribution, and are often referred to as white noise.

More sophisticated ARIMA models allow for moving average terms, differencing, and seasonal differencing. A moving average term can be added to an autoregressive model by replacing the independent error terms with sums of innovation terms; the ith error term is the sum of the ith innovation plus a multiple of the i-1th innovation, plus a multiple of the i-2th innovation, and so on; the innovations are assumed to be white noise. Differencing expands the set of possible models by assuming that the ARIMA model applies to the difference between one daily value and the following daily value. Seasonal differencing expands the process further by using differences between values a fixed number of periods apart (e.g. 12 month seasonal differences could be used to take into account monthly effects.)

Differencing will affect the trend estimation. For example, if there is an assumed linear trend in the annual means, then the mean of the differences between concentrations 365 days apart will be the trend rate (slope of the trend line in the annual means).

The number of possible time series models is immense and it can require considerable expertise to select the best model, or even to select a good fitting model. The methods suggested above for testing normality and independence can be applied to appropriately defined residuals. A useful definition is to calculate the residual as the difference between the observed daily summary statistic and the best model prediction based on all observed data up to the previous day. For application of the statistical tests for independence and normality it will first be necessary to divide these residuals by their estimated variances (since the variances of consecutive residuals are not equal). A commonly used test for time series analysis is the portmanteau test which is based on sum up to lag k of the squared autocorrelations for the residuals. This statistic is compared with a chi-square distribution to determine significance (large values imply a poor fit).

The fitting of these models to the raw data can be performed using modern time series software. A crucial requirement for the analysis of PAMS data is that the approach allows for model fitting even if there are substantial blocks of missing values; PAMS data will be collected during summer months only at many sites. One very useful software package that allows for almost any pattern of missing values (except for values at the beginning of the series) is the Splus package (a product of Statistical Sciences, Inc.). This software can fit any type of univariate ARIMA model including regression terms and seasonal differencing, taking into account missing values. The method is a version of the maximum likelihood method using a Kalman filter and a state space representation.
The extreme value model of Smith (1989) described in the next subsection is an example of a time series model that includes trends and serial dependence and is based on extreme value distributions rather than normal distributions. Software for such complex statistical analyses is not directly available in commercial software packages and so these analyses can require a substantial programming effort.

PROCEDURES BASED ON EXTREME VALUES AND EXCEEDANCES

As discussed above, the proposed methods in this section are designed for the analysis of trends in ozone rather than other PAMS species. We shall discusses methods based on fitting extreme value distributions to the daily maxima, and also discuss methods based on the analysis of annual exceedance rates, defined as the number of days per year that the daily maximum exceeds the NAAQS.

The theory of extreme values can be used to estimate the distribution of the annual maximum hourly concentration and/or the second up to the kth highest daily maximum hourly concentration, and to estimate the distribution of the number of days for which the daily maximum exceeds a high threshold (that may or may not be the ozone NAAQS). This section considers some trend analyses based on these approximations. We begin with a simple, but not very powerful, non-parametric technique, based on the chi-square distribution, to compare exceedance rates in different years. We then present the use of the Poisson process approximation for the daily exceedances. This approach is now being routinely applied for the EPA Trends reports (EPA, 1984-1991). Other applications of extreme value theory that have been used in the past are also presented. These other methods are potentially very useful in analyzing ozone trends but may be too complex for routine application.

Chi-Square Test of Trend

The Chi-square test of trend (EPA, 1974) is a simple test primarily applicable to ozone data to compare exceedance rates (exceedances per year) for two different years. A simple two-by-two table is created giving the number of NAAQS exceedance days and NAAQS non-exceedance days for each year. If there were no trend, then the proportions of exceedance days per year would be equal for both years. The differences between the observed numbers of exceedance days and the expected numbers in the case of no trend can be used to compute a chi-squared statistic. Because of the minimal amount of information used to compute this trend test statistic, the test has the disadvantage of having a very low trend detection probability which in most cases outweighs the advantage of simplicity.

Poisson Process Approximation

The simplest approach that uses extreme value theory is based on the result that exceedance days will follow a Poisson process in the limit, provided that the dependence between daily concentrations separated by a given number of days decreases sufficiently fast as the separation increases. Assuming that the numbers of exceedances for different sites are approximately independent, it follows that the total number of exceedance days for a given year (summed over the sites) will approximately have a Poisson distribution. The Poisson distribution has a variance equal to the mean, and the maximum likelihood estimate of this parameter will be the observed number of exceedances. If annual exceedances are averaged across a large number of sites, then the annual average number of exceedances per site will be approximately normally distributed, with a mean estimated by the annual average number of exceedance days per site and a variance estimated by the annual average number of exceedance days per site divided by the number of sites.

The Poisson distribution model for the exceedance rates can be used to calculate simultaneous confidence intervals for the annual mean number of exceedances per site using the Bonferroni method. (The Tukey studentized range method is not applicable in this case because the variance varies from year to year.) This approach is derived in Pollack et al. (1984) and has been applied in several of the annual EPA Trends Reports (EPA, 1984-1991). The Bonferroni approach is used in these situations to compute confidence intervals such that the probability of erroneously determining one or more significant differences is 5 percent or less. In general these Bonferroni intervals are wider than the unknown width needed to exactly attain an overall 5 percent error probability, i.e., the Bonferroni intervals are an upper bound approximation to the exact 95 percent simultaneous confidence intervals.

Advanced Methods

More powerful but much more complex procedures based on extreme value theory fit detailed statistical models to the process of exceedances (i.e. the sequence of records that scores each day as an exceedance or not an exceedance) or to the daily maxima themselves. We shall describe here some specific statistical papers by Smith and Shively describing the results of these approaches. These methods are potentially very useful but are probably too complex for routine use.

Shively (1991) used an approach similar to the Poisson exceedance count model to estimate the long-term trend in ozone exceedance rates for Houston daily maxima. The sequence of daily exceedances of the selected high ozone threshold was modeled as a non-homogeneous Poisson process. Thus the exceedances were assumed to follow a Poisson process with a rate that was not constant. The logarithm of the exceedance rate for a given day is the sum of multiples of certain meteorological measurements for that day and of a multiple of the calendar year. The calendar year multiple gives the estimated trend.

In another paper, Shively (1990) used the limiting joint extreme value distribution for the k highest daily maximum hourly ozone concentrations for each of several years. This limiting distribution assumes that all daily maxima are approximately independent, and that for each year the daily maxima have the same distribution. The location parameter (a parameter related to the mean of the limiting distribution) was assumed to change linearly with the year. The maximum likelihood method was used to estimate the parameters, but a bootstrap method was used to determine the statistical significance of the trend, since the amount of data used in the analysis was too small to apply asymptotic theory for the significance test.

The methods of Smith (1989) use the latest advances in extreme value theory to derive a very complete description of the sequence of daily maxima that incorporates the most general limiting extreme value distribution for the upper tail, the possible clustering of exceedances, seasonal trends (within year), and annual trends (across years). Since exceedance days often cluster together in cases of strong serial dependence, Smith fitted the trend model to all hourly ozone concentrations greater than a high threshold separated in time by more than a cluster interval; if more than one hourly exceedance of the threshold occurred within the cluster interval, only the highest of the cluster exceedances was used to fit the model. To fit ozone data from Houston, Texas (1973-1986) various thresholds (0.08, 0.10, 0.12, 0.16, 0.20, 0.26, 0.28, and 0.30 ppm) and two alternative cluster intervals (24 and 72 hours) were used with somewhat different results.

According to the limiting extreme value theory model, the cluster exceedances occur according to a Poisson process, and the distribution of the cluster maximum concentration is the tail generalized Pareto distribution (GPD). The tail GPD has a location, scale, and shape parameter. To treat seasonality, which is variation within the year, the scale and shape parameters differ by the calendar month (or pair of months), but are the same for every year. To treat trend, which is variation from year to year, the location parameter was assumed to be an intercept plus a slope parameter multiplied by the calendar year; both intercept and slope vary by calendar month (or pair of months). This complex model was fitted by the maximum likelihood method.

The extreme value theory model in Smith (1989) is the most realistic application of extreme value theory to ozone data since it incorporates serial dependence, seasonal dependence, annual trends, and the most general limiting extreme value distribution. The routine application of Smith's model by air quality managers is difficult because of the computational difficulties in fitting the model and the very difficult problem of selecting reasonable choices for the threshold and cluster interval; Smith's selections for Houston are likely to be inappropriate for many other cities. An even more complete analysis would include terms representing meteorological effects into the extreme value theory model and allow for nonlinear trend functions.

References

Bloomfield, P., G. Oehlert, M. L. Thompson, and S. Zeger. 1983. A frequency domain analysis of trends in Dobson total ozone records. J. Geophysical Res. , 88(C13):8512-8522.

Capel, J., T. R. Johnson, and T. McCurdy. 1983. "Analysis of Ozone Trends for Selected Indices of Daily Maximum Air Quality Data." Air Pollution Control Association Annual Meeting, Atlanta, Georgia (June 19-24, 1983).

Chock, D. P., S. Kumar, and R. W. Herrmann. 1982. An analysis of trends in oxidant air quality in the South Coast Air Basin of California. Atmos. Environ., 16(11):2615-2624.

Cohen, J. P., and A. K. Pollack. 1991. "General Linear Models Approach to Estimating National Air Quality Trends Assuming Different Regional Trends." Systems Applications International, San Rafael, California (SYSAPP-91/035).

EPA. 1973. The National Air Monitoring Program: Air Quality and Emissions Trends - Annual Report. U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Research Triangle Park, North Carolina (450/1-73-001a and b).

EPA. 1974. Guideline for the Evaluation of Air Quality Trends. Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency.

Freas, W. A., and E. Sieurin. 1977. "A Nonparametric Calibration Procedure for Multi-Source Urban Air Pollution Dispersion Models." Fifth Conference on Probability and Statistics in Atmospheric Sciences, American Meteorological Society, Las Vegas, Nevada.

Hirsch, R. M., and J. R. Slack. 1984. A nonparametric trend test for seasonal data with serial dependence. Water Resources Res., 20(6):727-732.

Hirsch, R. M., J. R. Slack, and R. A. Smith. 1982. Techniques of trend analysis for monthly water quality data. Water Resources Res. , 18(1):107-121.

Kolaz, D. J., and R. L. Swinford. 1988. "Ozone Air Quality: How Does Chicago Rate?" 81st Annual Meeting of the Air Pollution Control Association, Dallas, Texas (June 1988).

Kolaz, D. J., and R. L. Swinford. 1989. "Ozone Trends in the Greater Chicago Area." Ozone Conference on "Federal Controls for Ozone Around Lake Michigan," Lake Michigan States' Section and Wisconsin Chapter of the Air and Waste Management Association (October 12-13, 1989).

Kumar, S., and D. P. Chock. 1984. An update on oxidant trends in the South Coast Air Basin of California. Atmos. Environ., 18(10):2131-2134.

Lettenmaier, D. P. 1976. Detection of trends in water quality data from records with dependent observations. Water Resources Res., 12(5):1037-1046.

Pollack, A. K., and W. F. Hunt 1985. "Analysis of Trends and Variability in Extreme and Annual Average Sulfur Dioxide Concentrations." Air Pollution Control Association Specialty Conference on "Quality Assurance in Air Pollution Measurements," Boulder, Colorado.

Pollack, A. K., W. F. Hunt, Jr., and T. C. Curran. 1984. "Analysis of Variance Applied to National Ozone Air Quality Trends." 77th Annual Meeting of the Air Pollution Control Association, San Francisco, California (June 24-29, 1984).

Pollack, A. K., and T. S. Stocking. 1989. "General Linear Models Approach to Estimating National Air Quality Trends." Systems Applications, Inc., San Rafael, California (SYSAPP-89/098).

Reinsel, G., G. C. Tiao, M. N. Wang, R. Lewis, and D. Nychka. 1981. Statistical analysis of stratospheric ozone data for the detection of trends. Atmos. Environ., 15(9):1569-1577.

SCAQMD. 1991. "Final Air Quality Management Plan 1991 Revision. Final Appendix II-B: Air Quality Trends in California's South Coast and Southeast Desert Air Basins, 1976-1990." South Coast Air Quality Management District.

Shively, T. S. 1990. An analysis of the long-term trend in ozone data from two Houston, Texas monitoring sites. Atmos. Environ., 24B(2):293-301.

Shively, T. S. 1991. An analysis of the trend in ground-level ozone using nonhomogeneous Poisson processes. Atmos. Environ., 25B(3):387-395.

Smith, R. L. 1989. Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone. Statist. Sciences, 4:367-393.

Sweitzer, T. A., and D. J. Kolaz. 1984. "An Assessment of the Influence of Meteorology on the Trend of Ozone Concentrations in the Chicago Area." Air Pollution Control Association Specialty Conference on "Quality Assurance in Air Pollution Measurements," Boulder, Colorado (October 14-18, 1984).

Wackter, D. J., and P. V. Bayly. 1987. "The Effectiveness of Connecticut's SIP on Reducing Ozone Levels from 1976 through 1987." Air Pollution Control Association Specialty Conference on "The Scientific and Technical Issues Facing Post-1987 Ozone Control Strategies," Hartford, Connecticut (November 1987).