APPENDIX A

Statistical Approaches to Ozone Trend Detection

Panel Members

P. Bloomfield, Chair

D. R. Brillinger

D. W. Nychka

R. S. Stolarski

NC 999967

Appendix A

Statistical Approaches to Ozone Trend Detection

Contents

1.0 CURRENT STATUS OF STATISTICAL ANALYSES

The statistical analyses compiled by Reinsel, Tiao, and their coworkers provide the most comprehensive work on trend direction. This group's efforts represent a long-term involvement with ozone data. Their methods have been published in peer-reviewed statistical journals (Reinsel and Tiao, 1987), and previous discrepancies with other approaches have been reconciled (Hill et al., 1986). Stemming from this statistical work are some recent analyses concerning seasonal trends in total ozone for some of the stations in the Dobson network, reported in Chapter 4 of this report. There is a substantial amount of analysis of ozone measurements in the meteorological literature, such as the work by Angell and Korshover (1983b). However, because such studies do not adequately adjust for the short-term correlation in ozone over time, their results are of limited value for drawing conclusions about trends in ozone. For these reasons, this review will concentrate on the methods developed by Reinsel and Tiao.

1.1 Ground-Based Total Column Ozone

The most complete statistical analysis of ground-based ozone measurements is by Reinsel et al. (1987). Based on the measurements of a global network of 36 Dobson spectrophotometers, an average trend of -0.026 ± 0.092 percent per year[1] was estimated over the period 1970–1984[2]. Perhaps the most important aspect of interpreting these results is to distinguish between the average trend associated with these 36 locations and a global trend in total column ozone. They need not be the same. Although satellite data (see Section 1.3) suggest that trend analysis based on the Dobson network is representative of a global trend, more investigation in this area is needed. At present it is uncertain how to extrapolate the average trend among the Dobson network to a global trend for the entire atmosphere.

In order to discuss the assumptions that lead to the average trend cited above and to compare this analysis with other work, it is helpful to describe the statistical model used by Reinsel and Tiao. This model accounts for several factors that influence trend detection: the seasonal behavior of ozone, the relationship of ozone with the solar activity, and short-term autocorrelation within the ozone series. Let Y_t represent the monthly average total ozone recorded from a particular Dobson station. This observation is assumed to have the following decomposition:

$$Y_t = \mu + S_t + \omega X_t + \gamma Z_t + N_t. \tag{1}$$

In the above expression, μ represents a mean level, S_t is a seasonal component, X_t is a ramp function modeling a linear trend in ozone beginning in 1970, Z_t is the monthly average of 10.7 cm solar flux (or a smoothed version of it), and, finally, N_t is a random variable representing the short-term variation in ozone. The parameter ω represents the trend in ozone. Using this statistical model and some assumptions on the short-term variation, it is possible to estimate ω and also derive a measure of the uncertainty in this estimate.

It is well accepted that the monthly fluctuations in ozone are not independent of one another and tend to be positively correlated. One way of accounting for this behavior is to assume that

[1] In this appendix all ± limits correspond to 95 percent confidence intervals.

[2] This trend estimate has been updated to -0.05 ± 0.07 percent per year during the longer period of 1970–1986 for a network of 43 Dobson stations (Reinsel, 1988, personal communication).

$\{N_t\}$ is approximated by an autoregressive time series (see Sections 2.4, 2.5, and 3.3 for more detail). Another approach using spectral analysis is given by Bloomfield et al. (1983). This part of the statistical model is necessary in order to obtain reliable measures of uncertainty for the trend estimate. Unfortunately, the trend estimates reported by Angell and Korshover (1983b) and Heath (1986) based on ordinary least-squares are of unknown accuracy because the standard errors (SE's) do not allow for autocorrelation. This issue is discussed at length in Section 2.

The other components in this model are also related to the estimate of ω. Modeling seasonality of the ozone measurements reduces the variance in the data and thus improves the accuracy of the trend estimate. The inclusion of the solar flux reduces bias in the estimate of ω by distinguishing behavior in ozone due to solar cycle from other long-term trends.

The statistical model described above is used by Reinsel and Tiao to obtain trend estimates for each station. Using a random effects model to describe variation between stations, these estimates are combined to yield an average trend for the Dobson network. In order to be able to interpret this estimated average trend, this random effects model will be briefly described. The 36 stations can be divided into seven geographic regions (see Table 1, from Hill et al., 1986). Let $\omega_{i,j}$ denote the actual trend in total ozone for the ith station in the jth region. This trend is assumed to satisfy the equation

$$\omega_{i,j} = \omega + \alpha_j + \beta_{i,j}$$

where ω is the actual "global" trend in ozone, α_j is a zero mean random variable that reflects the variability in trends between different regions, and $\beta_{i,j}$ is a zero mean random variable that reflects the variability of trends within a region. It should be noted that the term $\beta_{i,j}$ not only represents variability due to meteorological effects but also accounts for spurious trends such as calibration problems for particular stations. With this decomposition, the estimated trend for a particular station can be expressed as

$$\hat{\omega}_{i,j} = \omega + \alpha_j + \beta_{i,j} + E_{i,j}$$

where $E_{i,j}$ is the error in the trend estimate due to the variability of the ozone within a station's record.

This "random effects" model can be used to derive an estimate for ω that is a weighted average of the individual trend estimates; a standard error for this estimate can also be calculated. This approach for combining individual trend estimates has an advantage over a simple average because it adjusts for correlation among stations within the same region. The spatial correlation implied by this model, however, has a simple structure that may not be a good approximation of the ozone field. It assumes that all stations within a region are equally correlated while stations in different regions are independent. A more serious problem is interpreting the parameter ω. Although Reinsel et al. (1987) refer to ω as a global trend in ozone, this is an assumption and is not implied by the random effects model. A more precise definition of ω is that it represents the component of trend in total ozone that is common to all the stations used in the analysis.

An alternative to the time domain approach of Reinsel and Tiao can be found in Bloomfield et al. (1983). Rather than using a random effects model to combine estimated trends, a similar model is used to construct an average ozone series. The parameter ω is then estimated from the

Fourier transform of this single series. This frequency domain analysis has the advantage that less need be assumed about the structure of the short- and long-term variation in ozone. One

Table 1. Dobson Network Used in Total Ozone Trend Analysis

Region	Station	Location	Data Period
North America	Resolute, Canada	75°N, 95°W	Jan. 1958–Dec. 1984
	Churchill, Canada	59°N, 94°W	Jan. 1965–Dec. 1984
	Edmonton, Canada*	54°N, 114°W	April 1958–Dec. 1984
	Goose, Canada	53°N, 60°W	Jan. 1962–Dec. 1984
	Caribou, USA	47°N, 68°W	Jan. 1963–Dec. 1984
	Bismarck, USA	47°N, 101°W	Jan. 1963–Dec. 1984
	Toronto, Canada	44°N, 79°W	Jan. 1960–Dec. 1984
	Boulder, USA	40°N, 105°W	Jan. 1964–Dec. 1984
	Nashville, USA	36°N, 87°W	March 1963–Dec. 1984
Europe	Aarhus, Denmark	56°N, 10°E	Jan. 1958–Dec. 1984
	Lerwick, United Kingdom	60°N, 1°W	Jan. 1969–Dec. 1984
	Bracknell, United Kingdom	51°N, 1°W	Jan. 1969–Dec. 1984
	Potsdam, GDR	52°N, 13°E	Jan. 1964–Dec. 1984
	Belsk, Poland	52°N, 21°E	April 1963–Dec. 1984
	Hradec Kralove, Czechoslovakia	50°N, 16°E	Aug. 1961–Dec. 1984
	Hohenpeissenberg, FRG	48°N, 11°E	Jan. 1968–Dec. 1984
	Arosa, Switzerland*	47°N, 10°E	Jan. 1958–Dec. 1984
	Mont Louis, France	42°N, 2°E	March 1962–Dec. 1979
	Vigna di Valle, Italy	42°N, 12°E	Jan. 1958–Dec. 1984
	Cagliari/Elmas, Italy	39°N, 9°E	Jan. 1958–June 1984
	Lisbon, Portugal	39°N, 9°W	Oct. 1967–Dec. 1984
India	Srinagar, India	34°N, 74°E	Feb. 1964–Nov. 1984
	Quetta, Pakistan	30°N, 67°E	Jan. 1964–Dec. 1984
	New Delhi, India	29°N, 77°E	Jan. 1960–Dec. 1984
	Varanasi, India	25°N, 83°E	Jan. 1964–Dec. 1984
	Mount Abu, India	25°N, 73°E	Nov. 1969–Feb. 1982
	Kodaikanal, India*	10°N, 77°E	Jan. 1961–Dec. 1983
Japan	Sapporo	43°N, 141°E	Feb. 1958–Dec. 1984
	Tateno*	36°N, 140°E	Jan. 1958–Dec. 1984
	Kagoshima	32°N, 131°E	Jan. 1960–Dec. 1984
Australia	Brisbane	27°S, 153°E	Jan. 1958–April 1983
	Perth	32°S, 116°E	March 1969–April 1983
	Aspendale*	38°S, 145°E	Jan. 1958–Dec. 1982
	Macquarie Island*	54°S, 159°E	April 1963–Aug. 1981
South America	Huancayo, Peru*	12°S, 75°W	March 1964–Dec. 1984
	Buenos Aires, Argentina*	35°S, 58°W	Oct. 1965–Dec. 1984
Pacific	Mauna Loa, Hawaii*	20°N, 156°W	Jan. 1964–Dec. 1984

*Nine-station network used by Hill et al. (1977).

disadvantage of this model, however, is that stations within a particular region are constrained to having the same spectrum. This assumption is also made among the regional effects. Although this analysis yielded a trend estimate that differed slightly from that of Reinsel and his coworkers, most of the discrepancy can be explained by different sensitivities to extreme ozone values and to the solar variable, by different lengths of data records, and by different methods for modeling the seasonal variance (Hill et al., 1986).

Recent results reported in Chapter 4 suggest that the trend in ozone may depend on the season. For Arosa and several other northern stations, the percent losses in total ozone tended to be greater in the winter. Thus, the estimated trend may be more sensitive to the statistical treatment of the seasonal component than would otherwise have been expected. Another implication of these results is that the sensitivity of the trend estimates may be improved by concentrating on the months in which a larger trend is expected. Because monthly ozone values are autocorrelated, the estimated trends for each month will also be correlated. This feature makes it difficult to interpret the 12 individual trend estimates. However, these estimated monthly trends can be averaged to yield an estimate of the annual trend; a standard error for this estimate can be computed using the dispersion matrix of the individual estimates. Some of the statistical issues of coping with the seasonality in ozone are discussed in Section 3.

1.2 Ground-Based Ozone Profiles (Umkehr)

Statistical analysis of the Umkehr data supplied by some stations in the Dobson network is motivated by the suggestion that the main depletion of ozone due to the release of chlorofluoromethanes (CFM's) will occur at altitudes between 35 to 40 km (Umkehr layers 7 and 8). Trend analysis that concentrates on this segment of the stratosphere may be more sensitive in detecting a depletion due to CFM's than an analysis based on total column ozone. In fact, Reinsel et al. (1987) report a statistically significant negative trend in layers 7 and 8. However, these results need to be qualified in two ways. First, the average trend estimate for a particular layer cannot be interpreted as a global trend, but rather refers to the average ozone in the layer above the Dobson stations taking Umkehr data. Also, Umkehr measurements are sensitive to stratospheric aerosols (see Chapters 3 and 5 of this report). Although Reinsel's method attempts to adjust for aerosols, all the bias caused by the presence of aerosols may not be removed.

The statistical methods applied to the Dobson profile data are similar to the analysis of the total column measurements; therefore, only some specific remarks will be made. To account for the dependence of the Umkehr measurements on aerosols, an additional variable was included in the model (1). This is the atmospheric transmission of solar radiation measured at Mauna Loa, Hawaii. Although these transmission data are specific to the integrated amount of aerosols in the stratosphere over Mauna Loa, it is assumed in Reinsel et al. (1987) that a smoothed and possibly lagged version of this series may be appropriate at other locations. One weakness in this analysis is that the choices for smoothing and lagging the transmission series were not based on a specific statistical model. Nevertheless, the estimated effects due to aerosols on the different ozone layers are in agreement with theoretical predictions. Available aerosol data are discussed in Chapter 10. Reinsel and his coworkers are currently exploring the use of lidar measurements to improve aerosol corrections.

An earlier analysis of Umkehr data by Bloomfield et al. (1982) came to different conclusions from those of Reinsel et al. (1987). Specifically, the earlier analysis detected no significant negative trends. However, the earlier analysis included no attempt to correct for aerosol effects, and was based on older data. Since the statistical methods used were similar to those of Reinsel et al. (1987) in other respects, it is to be expected that an updated analysis would give results largely in agreement with those of the later analysis.

1.3 Satellite Data

The Solar Backscatter Ultraviolet (SBUV) instrument aboard Nimbus–7 provides the most recent and the longest record of ozone measurements from a satellite (see Chapter 5). These data can be used in two different ways to improve trend detection. In contrast to the Dobson network, the Nimbus–7 ozone data provide nearly global coverage. Thus it is fairly simple to construct zonal or global series as aggregations of the raw data. Another use of the satellite data is to assess the global representativeness of the Dobson instrument network. Both of these topics are considered in Reinsel et al. (1988). This work gives an estimate of -0.35 ± 0.28 percent per year for the global trend in total ozone for November 1978 to September 1985, after adjusting for a drift in the SBUV instrument and accounting for solar activity[3]. Comparing the satellite measurements at the locations of the Dobson stations with the full record of measurements suggests that the average trend estimated for this network will be similar to a global trend.

This work uses the same model as that given in (1). The time scale is monthly means, and Y_t should be interpreted as some aggregation of the raw data either over blocks (10 degrees of latitude by 20 degrees of longitude), over zones (bands of 10 degrees latitude), or over the entire surface covered by the satellite (70°S latitude to 70°N latitude). In each of these cases, just as in the Dobson measurements, successive ozone values tend to be correlated. This autocorrelation must be taken into account to derive reliable uncertainty levels for estimated trends. One notable feature of this data set is its short length (7 years) relative to the Dobson data. Since this time series does not even span one solar cycle, the adjustment of the ozone series using a covariate for solar activity (such as 10.7 solar flux) is important. Reinsel et al. (1988) have included such a solar term in their analysis.

Trend analysis using the SBUV data alone is difficult because of an instrument drift (discussed in Chapter 2). Reinsel et al. (1988) estimate a linear drift to be -0.39 ± 0.11 percent per year using the Dobson network[4]. If the satellite data are adjusted by this estimated drift, then the standard error of the estimated trend must also reflect the uncertainty of the drift estimate. Reinsel et al. (1988) include the contribution of the drift uncertainty in the trend standard error, but their method is not entirely satisfactory.

The representativeness of the Dobson network was evaluated by taking a weighted average of all blocks containing the 36 stations used in Reinsel et al. (1988). When this series was subtracted from the global series, no significant trend was found in these differences ($-.06 \pm .12$ percent per year). Although these results suggest that a trend in the Dobson network may be similar to a global trend, there are some problems with this comparison. The blocks are not point measurements of total column ozone but are themselves averages over a substantial amount of surface area. This may cause closer agreement with the global series than might otherwise occur if one used the measurements taken at specific locations. Also, in the actual Dobson analysis,

[3] This trend estimate has been updated to -0.28 ± 0.22 percent per year for the period November 1978 through December 1986 (Reinsel, 1988, personal communication).

[4] This drift estimate has been updated to $-.40 \pm 0.11$ percent per year for the period November 1978 through December 1986 (Reinsel, 1988, personal communication).

individual trends are estimated for each station and then combined. This differs from first forming an average series from the blocks and then estimating the average trend from this single series.

2.0 TEMPORAL AND SPATIAL CORRELATION

An important contribution of statistical methods to problems such as estimation of trends in ozone data is attaching an appropriate measure of uncertainty to any trend estimate. In this section we shall review the calculation of such measures and discuss their validity.

2.1 Least-Squares Estimates

The calculation of the least-squares slope of a set of data is the most familiar example of trend estimation. Suppose that the data are $y_1, y_2, \ldots, y_n$. Then the least-squares estimate of slope is

$$\hat{\omega} = \frac{\sum_{t=1}^{n} (t - \bar{t})(y_t - \bar{y})}{\sum_{t=1}^{n} (t - \bar{t})^2} \qquad (2)$$

where

$$\bar{t} = (n + 1)/2 \text{ and } \bar{y} = (1/n) \sum_{t=1}^{n} y_t.$$

The full equation of the fitted line is

$$y = \bar{y} + \hat{\omega}(t - \bar{t}).$$

The *residuals* are the vertical distances from the data points (t, y_t) to the line, and are given by

$$r_t = y_t - \hat{y}_t = y_t - \{\bar{y} + \hat{\omega}(t - \bar{t})\}.$$

The *residual sum of squares* is just

$$\sum_{t=1}^{n} r_t^2,$$

and the *residual mean square* is this divided by the *degrees of freedom* $(n-2)$,

$$s^2 = \frac{1}{n-2} \sum_{t=1}^{n} r_t^2.$$

The *standard error* of the slope estimate $\hat{\omega}$ (as it is usually calculated) is

$$SE(\hat{\omega}) = \frac{s}{\sqrt{\sum_{t=1}^{n} (t - \bar{t})^2}}, \qquad (3)$$

and it is common for the estimated slope to be reported as "$\hat{\omega} \pm SE(\hat{\omega})$" or, for reasons described in the next section, as "$\hat{\omega} \pm 2\, SE(\hat{\omega})$". But what meaning can be attached to such a report?

2.2 Interpretation of Standard Error

We can give a firm interpretation to a standard error such as $SE(\hat{\omega})$ only by reference to some *model* for the randomness in the data. The simplest such model is

$$y_t = \mu + \omega(t - \bar{t}) + \epsilon_t, \tag{4}$$

where $\{\epsilon_1, \epsilon_2, \ldots, \epsilon_n\}$ are random errors, independently drawn from a Gaussian distribution with mean value 0 and variance σ^2. The constants μ, ω, and σ are the (statistical) *parameters* of the model, and $\bar{y}$, $\hat{\omega}$, and s are *estimates* of these parameters. The implication of such a model is that the observed set of data $\{y_1, y_2, \ldots, y_n\}$ is only one out of an infinitely large set of possible data sequences, each with different values of the ϵ's. The observed data are a *sample* (of size 1) from the *population* of possible sequences.

Since each possible sequence gives rise to a different value of $\hat{\omega}$, the model implies a distribution for $\hat{\omega}$, called its *sampling distribution*. The variance of this distribution is

$$\mathrm{var}(\hat{\omega}) = \frac{\sigma^2}{\sum_{t=1}^{n} (t - \bar{t})^2}, \tag{5}$$

and, consequently, the standard error calculated in (3) can be regarded as an estimate of the square root of this variance.

It follows that for an appropriate constant $t_{n-2}^{.95}$, the range of values

$$\hat{\omega} \pm t_{n-2}^{.95} SE(\hat{\omega})$$

has a 95 percent chance of containing the "true" slope value, ω, or, in a certain sense,

$$Pr\{\hat{\omega} - t_{n-2}^{.95} SE(\hat{\omega}) \leq \omega \leq \hat{\omega} + t_{n-2}^{.95} SE(\hat{\omega})\} = .95 \tag{6}$$

Since the tabulated values for $t_{n-2}^{.95}$ are all at least 1.96, and for $n > 60$ are at most 2, the interval is often approximated by

$$\hat{\omega} \pm 2SE(\hat{\omega}).$$

Thus, "2σ" limits can be interpreted as giving an approximate 95 percent confidence interval.

2.3 Limitations of the Model

Evidently the credibility of the standard error (3) depends on the credibility of the model (4), in light of the data. It is often clear by cursory inspection that the Gaussian distribution is not a good model for the distribution of the errors ϵ_t. The Gaussian distribution is symmetric about its center and has relatively short tails, behavioral aspects that are often not shared by real data. However, it is known that the confidence interval statement (6) is not drastically affected by such deviations from the model (the effects of such deviations in the simpler case of estimating a mean are discussed by Benjamini, 1983).

A more serious problem is that the model (4) states that the data are *independent* of each other, and hence *uncorrelated*. In many sets of real data, however, especially those collected sequentially in time (*time series* data), cursory inspection reveals *serial* (or *temporal*) *correlation* in the data. Unfortunately, the effect of correlations among the data, including serial correlation, is generally to invalidate any confidence interpretation of the conventional standard error formula (3).

How is correlation visible? Positive correlation among consecutive observations means that a value higher than expected (that is, one with a positive error ϵ) is likely to be followed by another observation with a positive ϵ. Such *persistence* of deviations above or below what is expected is characteristic of many real series, including ozone data and most of the related meteorological data. Figure 1 shows annual mean ozone levels at Arosa, Switzerland, 1933 to 1982, from Birrer (1975) and Dütsch (1984b). Also shown is the (least-squares) regression line for the data, with equation

$$y = 336.4 - 0.1648\,(\text{year} - 1957.5).$$

The occurrence of several values in a row above or below the regression line is evidence of positive serial correlation, though it is less strong in this series than in many. Figure 2 shows the same data with least-squares lines fitted through data for several successive 10-year periods. This is essentially the same as the analysis of Bishop and Hill (1982). The coefficients of the lines are given in Table 2. The average of the nine SE's is 0.819, whereas the standard deviation of the calculated trends is 1.199. Thus the calculated trends show 46 percent more variation than their standard errors suggest they should. For series with stronger correlation between consecutive values, such as monthly station data and regional or global average series, the actual variability can exceed the estimated standard error by much more than in this case.

Where does this leave us? The standard formula (3) is, in general, meaningful only in the context of the model (4), but this model is untenable for most of the data we need to analyze for trend. Evidently we need a more tenable model, and a new formula for standard error that has the desired confidence interpretation (6) in the context of that model.

We shall see that other models may suggest other formulas for slope estimates, as well as for their standard errors. In general, these other slope estimates differ little from the ordinary least-squares estimates (2), and, in fact, (2) is asymptotically efficient in the presence of stationary autocorrelated noise (Grenander, 1954). Thus, the problem lies mainly in obtaining valid standard errors, not in calculating the slope estimate itself.

Table 2. Coefficients of least-squares lines

Start year	End year	Centercept[a]	Slope	SE (slope)
1933	1942	340.633	1.526	1.276
1938	1947	338.458	−0.678	1.397
1943	1952	335.892	1.753	0.676
1948	1957	337.450	0.224	0.799
1953	1962	339.600	0.490	0.747
1958	1967	337.976	−1.915	0.754
1963	1972	335.576	0.643	0.727
1968	1977	334.392	−1.166	0.578
1973	1982	330.375	0.044	0.440

[a] The *centercept* of the line is its height at the midpoint of the interval to which it is fit.

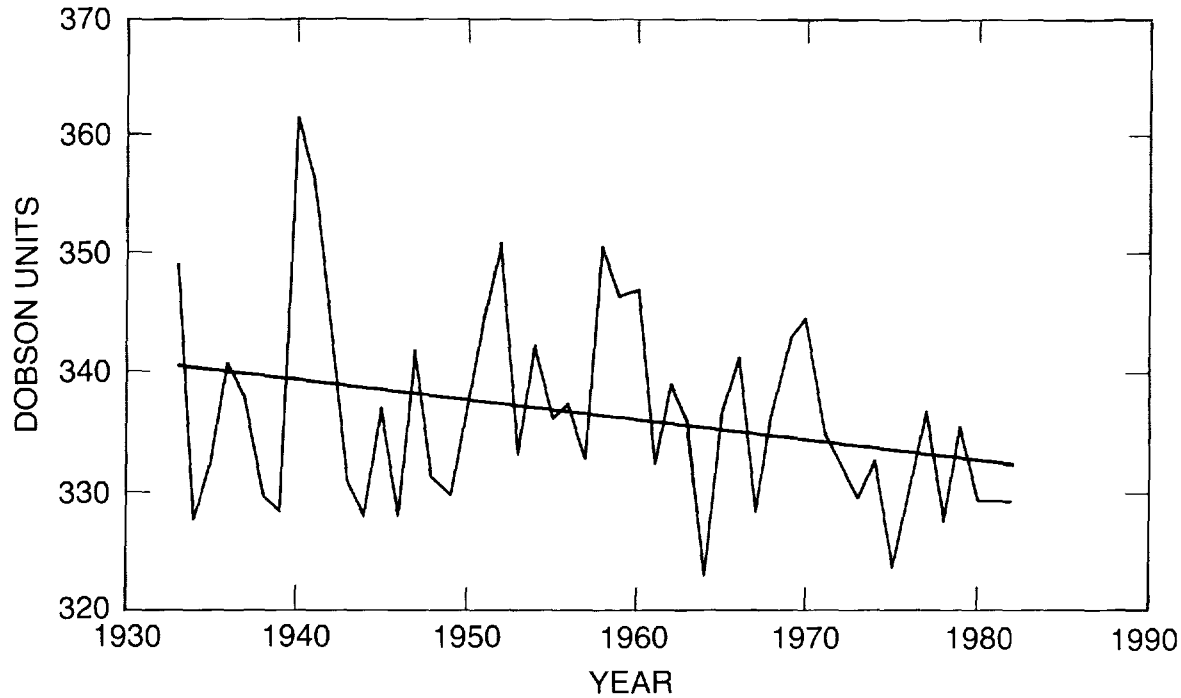

Figure 1: Annual mean ozone levels at Arosa, Switzerland, 1983–1982. (Unweighted averages of monthly means, with missing months replaced by least-squares estimates.)

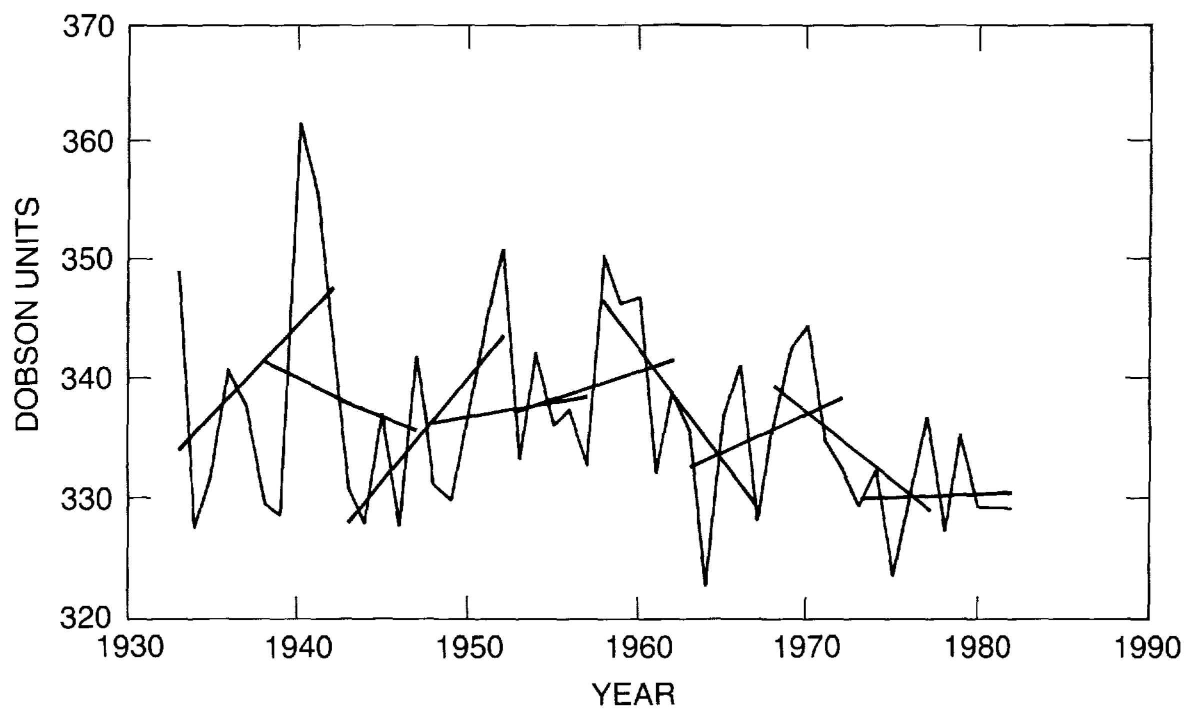

Figure 2: Arosa data with least-squares fitted lines.

2.4 An Elementary Model for Temporal Correlation

Models for serially correlated data have been discussed extensively in the statistical literature (see, for example, Box and Jenkins, 1976). One of the simplest is the (first-order) *autoregressive model*

$$\eta_t = \phi\eta_{t-1} + \epsilon_t, \ |\phi| < 1, \tag{7}$$

where the ϵ's are, as in (4), independent from mean 0 and variance σ^2. If we introduce the *backshift operator* B, defined by

$$B\eta_t = \eta_{t-1},$$

and more generally

$$B^h\eta_t = \eta_{t-h}, \ -\infty < h < \infty,$$

the model (7) may be written

$$\eta_t = \phi B\eta_t + \epsilon_t$$

whence

$$(1 - \phi B)\eta_t = \epsilon_t$$

and

$$\eta_t = (1 - \phi B)^{-1}\epsilon_t.$$

For this model, the serial covariances (or *autocovariances*) are

$$\gamma_h = \text{cov}\,(\eta_t, \eta_{t-h}) = \frac{\sigma^2\phi^{|h|}}{1 - \phi^2}, \ -\infty < h < \infty$$

and the serial correlations are

$$\rho_h = \text{corr}\,(\eta_t, \eta_{t-h}) = \phi^{|h|}, \ -\infty < h < \infty. \tag{8}$$

Since ϕ can be arbitrarily close to 1, this model can display correlations that are also close to 1, and that die away very slowly as the time separation h increases.

Now suppose that we wish to estimate a trend in the presence of correlated errors with the structure (7). That is, suppose that we wish to estimate ω in the equation

$$y_t = \mu + \omega(t - \bar{t}) + \eta_t,$$

where $\{\eta_t\}$ satisfies (7). If we know the value of ϕ, the simplest solution is to construct a new set of data

$$\begin{aligned}y_t^* &= y_t - \phi y_{t-1}\\ &= \{\mu + \omega(t - \overline{t}) + \eta_t\} - \phi\{\mu + \omega(t - 1 - \overline{t}) + \eta_{t-1}\}\\ &= (1 - \phi)(\mu + \omega/2) + \omega(1 - \phi)(t - \overline{t} - 1/2) + \eta_t - \phi\eta_{t-1}\\ &= (1 - \phi)(\mu + \omega/2) + \omega(1 - \phi)(t - \overline{t} - 1/2) + \epsilon_t,\end{aligned}$$

for $t = 2,\ldots,n$. If we write

$$x_t^* = (1 - \phi)(t - \overline{t} - 1/2)$$

and

$$\mu^* = (1 - \phi)\mu + (1 - \phi)\omega/2,$$

then the modified data satisfy

$$y_t^* = \mu^* + \omega x_t^* + \epsilon_t,$$

which is in the same form as the original equation (4), with errors ϵ_t that are now *uncorrelated*. It is therefore legitimate to estimate ω by least-squares, and to report its standard error as in equation (3). The resulting estimate is

$$\hat{\omega}_{ar1} = \frac{\sum_{t=2}^{n} x_t^*(y_t^* - \overline{y}^*)}{\sum_{t=2}^{n} x_t^{*2}}$$

with standard error

$$SE(\hat{\omega}_{\mathrm{ar1}}) = \frac{s_{ar1}}{\sqrt{\sum_{t=2}^{n} x_t^{*2}}}$$

where

$$s_{ar1}^2 = \frac{1}{n-3}\sum_{t=2}^{n}(y_t^* - \overline{y}^* - \hat{\omega}_{ar1}x_t^*)^2.$$

It may be shown that if n is large and $|\phi|$ is not close to 1, the new slope estimate $\hat{\omega}_{ar1}$ is very close to the estimate $\hat{\omega}$ of equation (2). However, the new standard error may be quite different from the earlier version, and is typically larger. In fact, we can show that if n is large and $|\phi|$ is not close to 1, then the (valid) standard error of $\hat{\omega}_{ar1}$ is larger than the (invalid) standard error of the least-squares estimate $\hat{\omega}$ by the factor

$$\sqrt{(1 + \phi)/(1 - \phi)}.$$

For the Arosa annual average data, the estimated value of ϕ is around 0.25, which would make this factor approximately 1.29. Since we observed 46 percent more variation among the slopes than was suggested by the least-squares standard error, rather than 29 percent, there is even more variability among the calculated trends than is predicted by the model (7).

Reinsel et al. (1987) found the month-to-month correlation for SBUV total ozone column data averaged from 70°S to 70°N to be around 0.84, which would make this factor 3.39. Thus, for ozone data averaged over a large region, in this case most of the globe, the effect of ignoring serial correlation can be a reported standard error of less than a third of its true value. Reinsel et al. (1988) give several illustrations of the difference between the valid and invalid standard errors.

In practice, of course, we will not know the value of ϕ, but it may be estimated from residuals and, provided the series length is adequate, used essentially as if it were known. An iterative procedure in which ϕ and ω are updated alternately was suggested by Cochrane and Orcutt (1949), and has been shown to yield maximum conditional likelihood estimates (Sargan, 1964). The procedure is discussed by Shumway (1988, Section 3.5).

2.5 More General Temporal Models

Equation (7) is often not the appropriate model for a given set of data; in practice, serial correlations may not show the simple exponential decay of equation (8). However, it is the simplest of a family of models (the *ARMA* family) that can be used in this way to model serial correlation and to provide valid standard errors for parameter estimates. The ARMA model of order (p,q) has the form

$$\eta_t = \sum_{r=1}^{p} \phi_r \eta_{t-r} + \sum_{s=0}^{q} \theta_s \epsilon_{t-s}. \tag{9}$$

In terms of the backshift operator, this may be written

$$\eta_t = \frac{\theta(B)}{\phi(B)} \epsilon_t$$

where the polynomials $\phi(z)$ and $\theta(z)$ are defined by

$$\phi(z) = 1 - \sum_{r=1}^{p} \phi_r z^r, \ \theta(z) = \sum_{s=0}^{q} \theta_s z^s.$$

The constraint $|\phi| < 1$ in the first-order model is generalized to the requirements

$$\phi(z) \neq 0 \text{ for } |z| \leq 1, \ \theta(z) \neq 0 \text{ for } |z| < 1.$$

For this model, the autocorrelations

$$\rho_h = \mathrm{corr}(\eta_t, \eta_{t-h})$$

cannot be written down explicitly, but satisfy the difference equation

$$\rho_h = \sum_{r=1}^{p} \phi_r \rho_{h-r}, \ h > q.$$

They therefore still decay to zero at an exponential *rate* as $h \to \infty$, but not as a simple exponential *sequence*.

As in the case of the AR(1) model (7), the slope of a trend line should be estimated not by ordinary least-squares, but by some method that takes the covariance structure of the data into account. Various methods are available, and some are implemented in the more complete statistical packages. It is again true that the slope estimate itself is generally close to the one obtained by ordinary least-squares, but its standard error may be very different when calculated appropriately. The approximate ratio of the correct standard error to the incorrect one can again be calculated, but it is more easily expressed in terms of the *power spectrum* of the model than its autocovariances.

The power spectrum of a stationary time series is the function whose Fourier coefficients are the autocovariances

$$s(f) = \sum_{h=-\infty}^{\infty} \gamma_h \epsilon^{-2\pi i f h}, \; 0 \leq f < 1,$$

from which it follows that

$$\gamma_h = \int_0^1 e^{2\pi i f h}\, s(f)\, df, \; -\infty < h < \infty.$$

The power spectrum of the ARMA(p, q) model (9) is

$$s_{ARMA}(f) = \sigma^2 \frac{|\theta(e^{2\pi i f})|^2}{|\phi(e^{2\pi i f})|^2}.$$

Grenander (1954) showed that the variance of the sampling distribution of a trend estimated from a stretch of data of length n is, for large n, approximately

$$\operatorname{var}(\hat{\omega}) = \frac{s(0)}{\sum_{t=1}^{n} (t - \bar{t})^2}, \tag{10}$$

Comparison with equation (5) shows that the factor σ^2 in the numerator has been replaced by $s(0)$, the power spectrum evaluated at zero frequency. Equation (10) can be used as the basis for computing the approximate standard error of the trend estimate, or merely to get an indication of how much the standard error differs from the result given by the formula (3), the ratio being just

$$\sqrt{s(0) \,/ \int_0^1 s(f)df}\,.$$

2.6 Spatial Correlation

In the previous sections we discussed the complications that arise when we analyze data collected over time, with correlation among observations that are close together in time (serial correlation, or *temporal* correlation). Measurements made close together in *space* also tend to be correlated, and when such measurements are analyzed jointly, their *spatial* correlation also needs to be taken into account.

One general approach to analysis of spatiotemporal data is based on models generalized from the strictly temporal models described above. However, this approach requires the data to be collected on a regular grid of locations, and to show some homogeneity in their correlation structure. One or both of these features is absent in most of the data that have been collected in studies of ozone and other atmospheric trends. However, more ad hoc solutions can be found for certain problems.

One such problem is the need to combine quantities calculated from data collected at different locations. Suppose that $\hat{\theta}_1$ and $\hat{\theta}_2$ are estimates of similar quantities at two locations. Each $\hat{\theta}$ might be a trend estimate, or an average ozone level over a decade, or a change in such an average from one decade to the next. The standard error of the average

$$\hat{\theta}_. = (\hat{\theta}_1 + \hat{\theta}_2)/2$$

is given by

$$SE\,(\hat{\theta}_.) = \sqrt{\mathrm{var}\,(\hat{\theta}_.)}$$

where

$$\mathrm{var}(\hat{\theta}_.) = (1/4)\,\{\mathrm{var}(\hat{\theta}_1) + \mathrm{var}(\hat{\theta}_2) + 2\mathrm{cov}(\hat{\theta}_1, \hat{\theta}_2)\}.$$

The two variance terms can usually be obtained from the analysis of the data for locations 1 and 2 separately, but the covariance term depends on the joint behavior of the two sets of data, and therefore can be calculated only from a joint analysis. While this is not a difficult analysis to carry out, it is not one that the commonly available statistical packages are set up to handle. Reinsel and Tiao (1987) describe a method that amounts to doing just this, in their analysis of an entire Dobson network, but the method is not easy to implement with standard software. As a result, it is more convenient to calculate summary statistics for a number of locations in a slightly different way.

Consider, for instance, the problem of calculating an average trend for all Dobson stations in a certain latitude band. Since it is straightforward to calculate a trend for each station, it might seem that the way to proceed is to simply combine the individual trends. However, to compute the standard error for this average we would need to know all the covariances among the individual station trends. We can, instead, form an average ozone time series for the band, then estimate the trend in this single series. The standard error of the resulting estimate may then be obtained in the usual way. The estimated trend in the average series is, in general, close to the average of the individual trends, and may actually be a more appropriate quantity to consider.

The weakness of this approach is that the analysis of the average series can account only for the statistical fluctuations that are visible in the average. Thus, phenomena that induce spurious trends in the individual station records, for example instrument drift, pass undetected, while in Reinsel and Tiao's approach they show up as variations among the individual trend estimates. As a result, the standard error obtained from analysis of an average series may need to be inflated to allow for such phenomena, while the standard error calculated from Reinsel and Tiao's method automatically allows for them. Neither approach can handle spurious trends common to all stations, so in either case some allowance must be made for these.

3.0 SEASONALITY IN OZONE DATA

In this section we review various ways in which ozone data show seasonal behavior, the problems such behavior raises, and some of the solutions used.

3.1 Seasonal Structure in the Mean

Seasonal structure is one aspect of many time series that needs to be considered when estimating a trend. It is usually easy to separate seasonal behavior from trend, so the issue is not one of confusion between the two so much as estimating the size of the random errors in the data.

There are two basic strategies for coping with seasonality in the mean. One is to use *seasonal adjustment* to remove the seasonal behavior, and then to estimate trend from the *seasonally adjusted* data. The second is to fit a model that includes both seasonal structure and a trend term, so that both parts are estimated simultaneously. The former is preferred in exploratory analysis as it allows more flexibility in estimating the trend. For instance, a graph of the adjusted data will often suggest what kind of trend is actually present, and will typically provide a preliminary estimate of its magnitude. By contrast, simultaneous estimation is preferable when a model has been chosen and its parameters are being estimated. One advantage of simultaneous estimation is that it usually provides standard errors for all estimated parameters, as well as their correlations.

In either case, it is necessary to specify the seasonal structure to be removed or incorporated into the model, respectively. The simplest approach is to allow an arbitrary mean for each month[5]. If we write the data as

$$y_t = S_t + z_t,$$

where S_t and z_t represent the seasonal and nonseasonal parts of the data, respectively, then this amounts to putting

$$S_t = \mu_i \text{ if data month } t \text{ falls in calendar month } i,\ i = 1,2,\ldots 12.$$

Here, μ_1 is the mean of January data, μ_2 is the mean of February data, and so on.

This approach has the merits of simplicity and ease of interpretation, but in short series it may be undesirable to use as many as 12 parameters to describe seasonal structure. An alternative approach is to use a sine-cosine expansion. In this case, we put

$$S_t = \mu + \sum_{j=1}^{J} \{A_j \cos(2\pi jt/12) + B_j \sin(2\pi jt/12)\}$$

where $J \leq 5$, using $2J + 1$ parameters. Often $J = 1$ or $J = 2$ is sufficient, corresponding to fitting an annual wave with three parameters, or annual and semiannual waves requiring five parameters, respectively. If we set $J = 6$ with the constraint $B_6 = 0$, there are again 12 parameters; this is effectively the same as allowing an arbitrary set of monthly means.

Figure 3 shows the results of these two approaches for the monthly Arosa data. The asterisks indicate the monthly means; the curve is constructed from a sine-cosine expansion with $J = 2$ (estimated by ordinary least-squares). This two-frequency model fits most of the monthly means

[5] For ease of exposition we deal only with the case of monthly data. The modifications necessary for other types of seasonal behavior are clear.

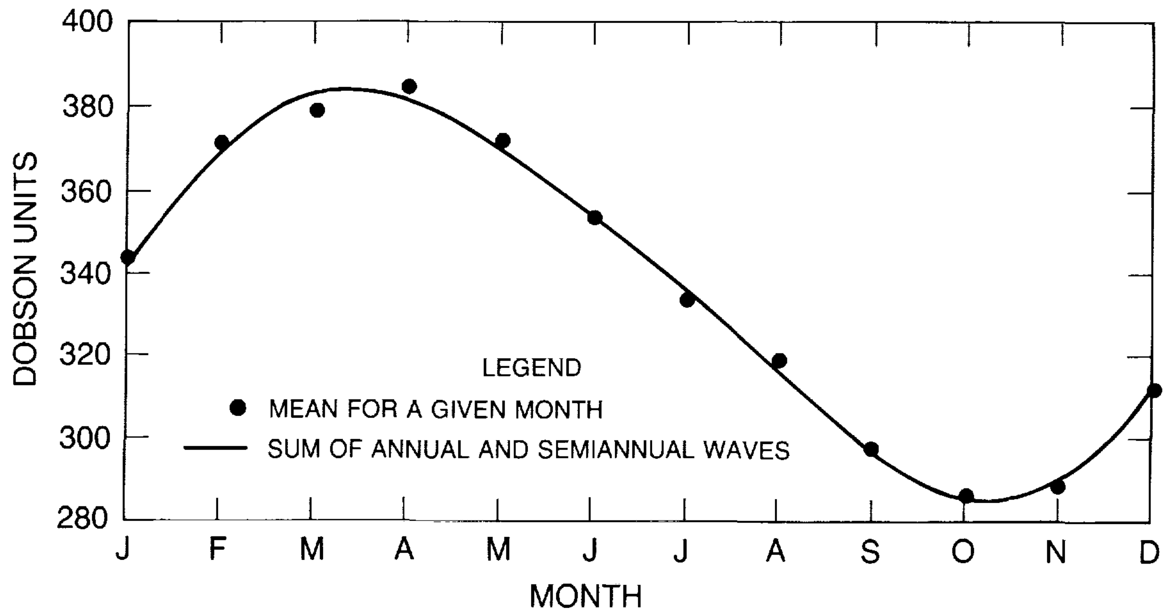

Figure 3: Seasonal structure of monthly Arosa data.

well in the graph, but the lack of fit in March and April could be critical in a trend study (see Section 3.2 below).

3.2 Seasonal Structure In the Trend

The issue of seasonal structure in the mean of a series discussed in the previous section may also extend to estimating a trend, if theory or observation suggests that the trend may vary from season to season.

Suppose, for instance, that trend is being measured by a ramp function such as

$$
\begin{aligned}
r_t &= \begin{cases} \lambda & \text{if } t < t_0 \\ \lambda + \omega\,(t - t_0) & \text{if } t \geq t_0 \end{cases} \\
&= \lambda + \omega\,(t - t_0)_+,
\end{aligned}
$$

where $(t - t_0)_+$ is defined by

$$
(t - t_0)_+ = \begin{cases} 0 & \text{if } t < t_0 \\ (t - t_0) & \text{if } t \geq t_0. \end{cases}
$$

The function r_t has the constant level λ up to month t_0, and increases with slope ω units per month after month t_0. As in the previous section, we have various ways of allowing the magnitude of the trend to vary from month to month. The simplest is to replace the single λ by 12 λ's and the single ω by 12 ω's:

$$
r_t = \lambda_i + \omega_i\,(t - t_0)_+ \text{ if data month } t \text{ falls in calendar month } i.
$$

Seasonal trend estimates of this form are reported in Chapter 4.

Again, however, we might not want to allow completely arbitrary variation in the trends from one month to the next. One alternative would be to use sine-cosine expansions as before:

$$r_t = \lambda_0 + \sum_{j=1}^{J}\{A_j\cos(2\pi jt/12) + B_j\sin(2\pi jt/12)\}$$
$$+ \Big[\ \omega_0 + \sum_{j=1}^{J}\{C_j\cos(2\pi jt/12) + D_j\sin(2\pi jt/12)\}\Big](t - t_0)_+.$$

Notice that we have used the same number of terms in each expansion. Statistical considerations such as lack of significance of some coefficients might suggest dropping terms from one sum but not the other, but care would have to be used. Omission of significant terms from one sum would introduce bias into estimates of coefficients in the other, and possibly into other coefficients in the same sum. In particular, bias could arise if a sine-cosine expansion were used for, say, the level term, and arbitrary constants were used for the trend, since using arbitrary constants is effectively the same as taking $J = 6$ (and omitting the final sine term).

3.3 Seasonal Structure in the Correlations

In Section 2 we have discussed the use of time-series models to represent the correlation among measurements such as ozone columns at different times at the same location (or geographical region). One way in which those models are not sufficiently general is that they have *stationary* covariance structure:

$$\text{cov}(y_t, y_{t-h}) = \text{a function only of } h.$$

However, most ozone time series have covariances that depend on the season as well as the time separation h. This is most easily seen in the monthly dependence of the standard deviation. As was mentioned in Section 1.1, failure to allow for such seasonality in covariance structure distorts any analysis by placing equal weight on the more variable winter data and the less variable summer data.

Two approaches have emerged for coping with this problem. Reinsel and his coworkers have developed and exploited a method that can be described as seasonally weighted least-squares, while the values reported in Chapter 4 were obtained essentially by standardization of the data before fitting the trend model. The two approaches may be compared in terms of a corresponding model for the noise term η_t. In each case, the analysis may be interpreted as using a modified ARMA model (cf. equation 9). Reinsel's approach corresponds to the model

$$\eta_t = \frac{\theta(B)}{\phi(B)}(\sigma_t\epsilon_t),$$

while the approach used in this report corresponds to

$$\eta_t = \sigma_t\frac{\theta(B)}{\phi(B)}\epsilon_t.$$

In each case, the scale factors σ_t are periodic, and reflect the seasonal dependence of the standard deviation of the noise η_t and of the innovations ϵ_t, respectively. Both general models appear plausible, but studies of which fits ozone data better have not been performed in depth.

4.0 Summary

Ready, widespread acceptance of the results of a scientific investigation depends critically on the credibility of the study, and, in the case of a data-oriented study, this in turn depends largely on two factors: the quality of the data on which the study is based and the quality of the subsequent analysis of those data. Data quality for each of the measuring systems is discussed in the relevant chapters; here we offer only broad comments. The primary focus of this appendix is on data analysis techniques, about which we offer more specific suggestions.

- *Trend estimates*—There are many ways to qualify the trend in a set of data, including fitting a straight line, fitting a ramp (or hockey stick) function, or comparing averages over different time windows. In each case, the resulting quantity should be accompanied by an appropriate standard error.

- *Standard errors*—A standard error has meaning only in the context of being a model for the way in which the data were sampled. If the data show evidence of spatial or temporal correlation, the sampling model must reflect this. Computer software for building the required sampling models and calculating the corresponding trend estimates and standard errors is widely available and should be used more extensively.

- *Current results*—Trend estimates based on fitting ramp functions reported in Chapter 4, and those of Reinsel, Tiao, and their coworkers, adequately account for serial correlation and its seasonal structure, and represent state-of-the-art estimates. Estimates of trends from Total Ozone Mapping Spectrometer (TOMS) data, also reported in Chapter 4, allow for serial correlation, but not for its seasonal structure. Although not state of the art, the results should be close. Standard errors of other trend estimates obtained by least-squares fitting without allowance for serial correlation may be incorrect by factors of more than three.

- *Ground-based total ozone column data*—It has been demonstrated that the quality of Dobson total ozone column data can be improved substantially by retroactive application of corrections based on calibration changes. This is best carried out by complete recalculation of each day's data on the basis of corrected algorithms, but useful improvements can be made by correcting monthly averages. Stations should be urged to give a higher priority to such adjustments to historical data, and to ensuring publication of the adjusted data by the World Ozone Center.

- *Ground-based ozone profile data*—The largest remaining question about the quality of the ozone profile data obtained by the Umkehr technique is the impact of aerosols. This question will have to be resolved before the Umkehr network can fulfill its dual roles of providing ground-based information about trends in ozone profiles and of providing ground truth for the validation of satelliteborne instruments.

- *Satellite ozone measurements*—The realization that the effect of diffuser plate degradation on Nimbus-7 cannot be uniquely separated from other instrument changes underscores the difficulty of maintaining measurement stability with satellite-based systems. This illustrates the continuing need for cross-checking all types of measurements, a need about which there has been a tendency to become complacent.