NPL Report CMSC 15/02

Method for Evaluating Trends in Ozone Concentration Data and its Application to Data from the UK Rural Ozone Monitoring Network

M. G. Cox, P. M. Harris, M. J. T. Milton
and P. T. Woods

November 2002

Method for Evaluating Trends in Ozone Concentration Data and its Application to Data from the UK Rural Ozone Monitoring Network

M. G. Cox and P. M. Harris
Centre for Mathematics and Scientific Computing

M. J. T. Milton and P. T. Woods
Centre for Optical and Analytical Measurement

November 2002

ABSTRACT

A method is developed to evaluate long-term trends in the concentration of ozone at the sixteen sites in the UK Rural Ozone Monitoring Network. The method is based on a regression model incorporating "seasonal" factors that is fitted to the ozone concentration data formed into weekly, monthly, quarterly or yearly averages. The uncertainties associated with the results derived from the regression model are estimated by bootstrap resampling from the residual deviations. This approach has the advantage over traditional approaches that it does not require any assumption to be made about the distribution of the residual errors that are unexplained by the model.

The validity of the method is verified in a number of ways. It is shown that trends estimated from measured data can be sensitive to "missing" data points, and a recommendation is made for using the model in conjunction with data averaged over an appropriate time interval in order to reduce this sensitivity and to have other advantageous properties. Additionally, bootstrap resampling is used to estimate the uncertainty associated with measures of the trend aggregated over all sixteen sites in the UK Rural Ozone Monitoring Network

Key Words

air quality bootstrap resampling long-term trends ozone regression statistics

© Crown Copyright 2002
Reproduced by Permission of the Controller of HMSO

ISSN 1471–0005

National Physical Laboratory
Queens Road, Teddington, Middlesex, TW11 0LW

Extracts from this report may be reproduced provided that the source
is acknowledged and the extract is not taken out of context

Approved on behalf of the Managing Director, NPL
by Dr D Rayner, Head, Centre for Mathematics and Scientific Computing

CONTENTS

1. Introduction

The growth in the extent and scope of air-quality monitoring has resulted in the accumulation of substantial archives of data measured at automatic stations. One of the most important questions that this data is intended to address is whether ambient air quality is improving or deteriorating from year to year. Although this question appears to be straightforward, in practice long-term trends are difficult to quantify accurately because of the presence of large seasonal and other effects in the data (Eldred and Cahill, 1994; Oltmans *et al*, 1998).

In this paper we consider "raw data" consisting of hourly averages of ozone concentration measured at the sixteen sites in the UK Rural Ozone Monitoring Network (PORG, 1993). The sites include a variety of different rural locations such as remote hilltops, coastal areas and undeveloped areas between population centres. The locations of these sites within the UK are indicated on the map shown in Figure 1. The ambient ozone concentration is recorded automatically and averaged over fifteen-minute intervals. It is then ratified, calibrated and "quality assured" prior to being placed on a public access database (the Department of the Environment's National Air Quality Information Archive at http://www.aeat.co.uk/netcen/aqarchive/auto.html) as one-hour averages. Although data is available at some sites from 1973, this study used a common seven-year period, January 1989 to December 1995, to allow comparison of the results obtained at different sites and to permit aggregation of these results over the whole network.

The first question addressed in this paper is "What is the long-term trend over the seven-year period in ozone concentration at each site, and what are the uncertainties associated with these estimates?". This question is established as a formal statistical hypothesis and the results of our analysis are used to test the hypothesis for each site. The precise definition of "long-term trend in ozone concentration" is crucial to the hypothesis and its testing. In this study, the arithmetic mean of the hourly data, evaluated over four time periods (yearly, quarterly, monthly and weekly), is used to measure ozone concentration. Then, an underlying straight-line trend in these measures, expressed as a *gradient* with respect to time, is used to indicate the long-term trend.

One approach to estimating the long-term trend in this type of data is to fit a straight-line to the measured data. The gradient of the fitted line is used to indicate the long-term trend, and its associated confidence interval can be evaluated under the assumption that the residual errors associated with the fitted line are normally distributed. Such an approach is somewhat simplistic because it does not take explicit account of periodic trends in the data (such as seasonal variation) and it is statistically unsound because the residual deviations from the model depart appreciably from normality.

The approach we have developed involves the use of a class of regression models incorporating seasonal factors in addition to an underlying long-term trend. The seasonal factors are used to describe yearly periodic trends in the data. The approach adopted to determine the sampling distributions of the estimated underlying trend, and hence the confidence intervals associated with these estimates, is non-parametric because no assumption is made about the distribution of the residual errors.

An alternative approach to the analysis of time series of air quality data is to use a regression model together with information about parameters that may have a causal link to the data being modelled. Such an approach is well suited to the study of the causes of changes in air quality and to extrapolation to forecast future values (Shi and Harrison, 1997). However, the method developed here is better suited to studies of the trend of data sets that may be carried out prior to more detailed regression modelling of other inputs. The approach can be used for any statistical measure applied to the data sets, for example the mean, standard deviation or 95th percentile point of the cumulative distribution.

As with the application of statistics in many areas, tests must be carried out in order to demonstrate that the results obtained are valid. In order to justify that the results obtained from the method are statistically meaningful, it is necessary to demonstrate that the model chosen provides a satisfactory explanation of the data. In this case, it is important to show that the number of seasonal factors used is appropriate, and that the underlying trend is predominantly linear. The method used to evaluate the confidence intervals associated with the estimates relies on bootstrap resampling from the distribution of residual errors. This approach requires some further demonstration that there is no significant bias introduced and that the confidence intervals are reliable. Finally, it is important to confirm that the model is sufficiently robust to deal with possible gaps in the data.

We have also extended the scope and application of the approach to address the second question "How can the information from the individual sites be aggregated to estimate the long-term trend in ozone concentration over the whole network, and what is uncertainty associated with such an estimate?". In the same way as for the analysis of a single site, this question is established as a formal statistical hypothesis and the results of our analyses are used to test the hypothesis.

The method we have used is to combine the results of the analyses for the individual sites to obtain an aggregated measure. Bootstrap resampling is then used to estimate the sampling distribution of the aggregated measure from which confidence intervals associated with the long-term aggregated trend values are derived.

The paper is organised as follows. In Section 2 we describe the regression models used to fit the data. Initially, these are applied to yearly averages of the data and are subsequently generalised by the inclusion of seasonal factors to data representing averages over other time periods. In Section 3 the use of bootstrap resampling for the evaluation of uncertainties is discussed. In Section 4 the results for individual sites are presented. In Section 5 we validate the model used in terms of the number of seasonal factors and the nature of the underlying trend. In Section 6 we demonstrate the reliability of the results returned by bootstrap resampling, and in Section 7 we consider the important issue of the influence of missing data points on the results. In Section 8 we describe the aggregation of the site results to derive measures for the network as a whole. Section 9 contains a discussion of the methods developed in this study and presents recommendations concerning their application. Section 10 contains our conclusions.

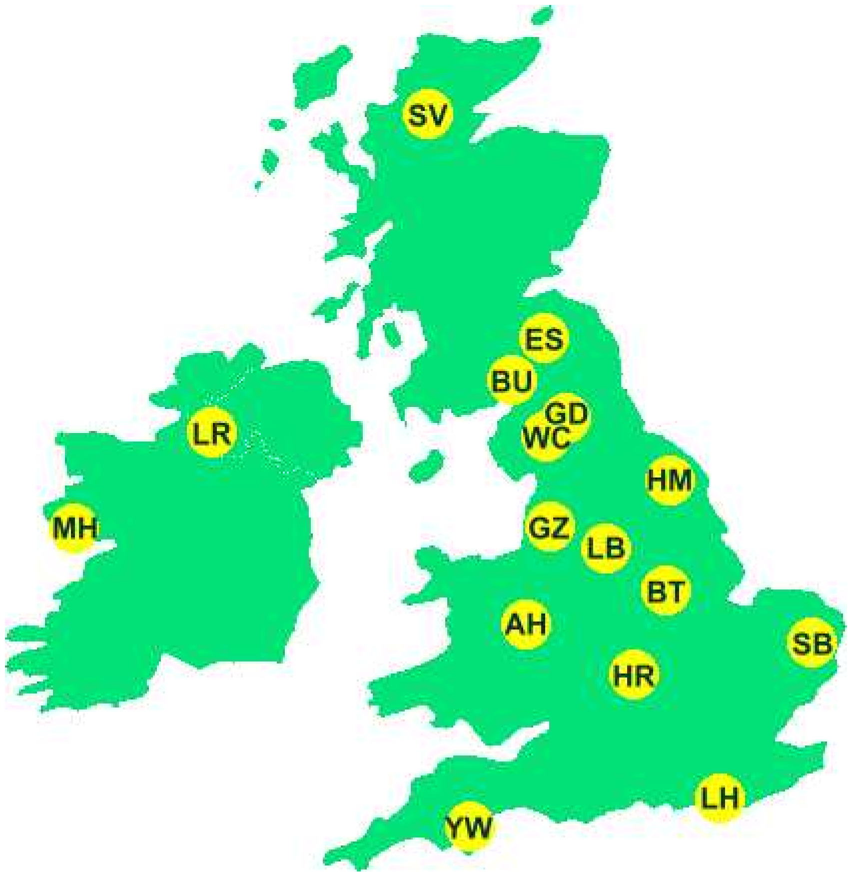

Figure 1: Locations of the monitoring stations forming the UK Rural Ozone Monitoring Network.

2. Data fitting with regression models

2.1 Straight-line regression model

A straightforward approach to estimating the long-term trend in a data series is to calculate yearly averages of the data and then to fit a straight-line regression model to these averages. Suppose y_i*, i = 1, ..., n_y, denotes the value in parts-per-billion (ppb) of ozone concentration obtained by taking the arithmetic mean of hourly data from year i. The regression model for these values takes the form:

$$y_i{*} = F_y(t_i, a, b_1) = at_i + b_1, \quad t_i = i,$$

where a represents the gradient of the underlying straight-line trend in ozone concentration (measured in ppb per year), and b_1 is the value of ozone concentration at the start of the first year.

The parameters in the regression model are determined by minimising the difference between the measured data values y_i and the model values y_i^* in a weighted least-squares sense, i.e., by minimising the weighted residual sum of squares

$$\sum w_i^2 e_i^2, \qquad e_i = y_i - y_i{*}.$$

In principle, the raw data referred to in Section 1 is available as a continuous time series of hourly averages. In practice, there are missing data as a result of occasional malfunction of the automatic measuring stations. As a consequence, there can be gaps in the data sequence spanning hours, days or weeks. The method of analysis indicated above accounts for such gaps by introducing "weights" w_i that reflect the quantity of data actually available within the time periods being analysed:

$$w_i = \frac{\text{number of hourly values recorded in year } i}{\text{total number of hours in year } i}. \tag{1}$$

The values w_i range from unity if all the data is present to zero if no data is present. (A discussion of the limitations of the use of weights to reflect data capture is given in Section 7.)

An example of some raw data and the corresponding regression model is illustrated in Figures 2 and 3. Figure 2 shows the distribution of hourly averages at a single site (Aston Hill) for the period January 1989 to December 1995. Figure 3 shows the straight-line fit to data obtained by averaging the hourly values for each year and the corresponding model residuals.

Although this approach is relatively straightforward to implement, it has several disadvantages, in particular:

1. It "smooths out" much of the fine structure (and hence information) in the raw data.

2. It reduces a relatively large number of data points to a number for which only "small-sample" statistical theory is applicable.

3. It does not provide realistic confidence intervals associated with the gradient (or any other modelled parameter) without making additional assumptions (such as that of normality). This is usually not a good assumption since the residuals defined by the fitted model are not purely

estimates of measurement error but are predominantly quantities unexplained by the model which behave in a random or near-random manner.

For these reasons, the approach is too limited to be universally applicable to the determination of long-term trends in data. Although it is useful as part of a more comprehensive study involving the analysis of data averaged over shorter time periods, it should be extended so that it (i) models the data in a way that takes better account of its inherent statistical variability, (ii) quantifies the trends and their associated confidence intervals, and (iii) can be used to aggregate the results obtained from many sites.

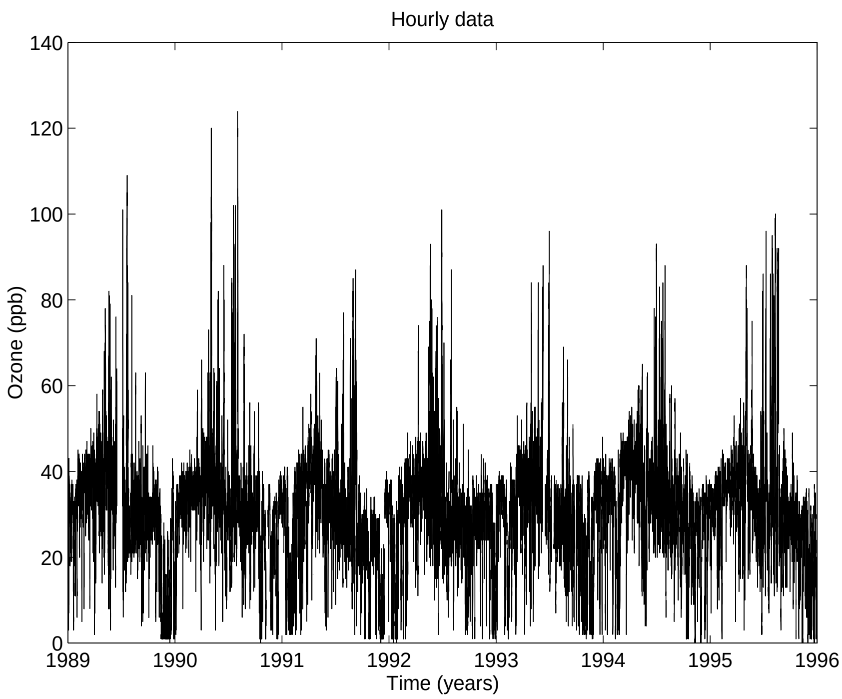

Figure 2: Time series of hourly averaged ozone concentration at Aston Hill (AH) for the period January 1989 to December 1995.

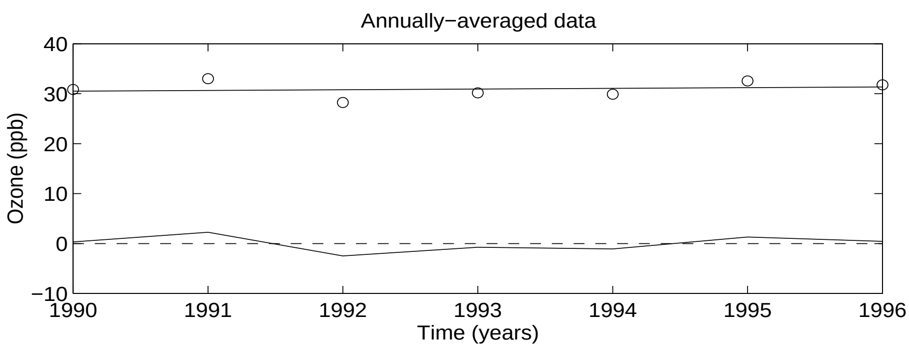

Figure 3: Fit of straight-line regression model to mean ozone concentration data obtained by taking yearly averages of the data shown in Figure 1. The upper trace shows the data (small circles) and the fitted model. The weighted residuals (straight line segments joining their values) and the “zero line” are shown below.

2.2 Regression models incorporating seasonal factors

Some of the drawbacks of the simple straight-line regression model can be overcome by the use of a model that includes an underlying straight-line trend together with a set of factors to account for periodic variations. We refer to these as “seasonal factors”, and use 1, 4, 12 or 52 factors according to whether yearly-, quarterly-, monthly- or weekly-averaged data is being analysed. They are not intended to explain the nature of the periodic variations in the data, but are used as a statistical tool to remove some periodic behaviour in order to identify the underlying trend more clearly.

As an example, suppose $q_{i,j}*$, $j = 1, ..., 4$, are the values of ozone concentration obtained by averaging over each of the four quarters of year i. Then, the values for the first two years are modelled using an underlying straight-line trend with gradient a and four seasonal factors $b_1, ..., b_4$ as follows:

$$q_{1,1}* = \frac{1}{4}a + b_1, \quad q_{1,2}* = \frac{2}{4}a + b_2, \quad q_{1,3}* = \frac{3}{4}a + b_3, \quad q_{1,4}* = 1a + b_4,$$
$$q_{2,1}* = \frac{5}{4}a + b_1, \quad q_{2,2}* = \frac{6}{4}a + b_2, \quad q_{2,3}* = \frac{7}{4}a + b_3, \quad q_{2,4}* = 2a + b_4.$$

In the general case, if $v_{i,j}*$, $j = 1, ..., n$, are the values obtained by averaging over n equal time periods in year i, the regression model takes the form:

$$v_{i,j}* = F_v\left(t_{i,j}, a, \boldsymbol{b}\right) = at_{i,j} + b_{j,} \quad t_{i,j} = (i-1) + \frac{j}{n}, \quad j = 1, \ldots, n,$$

where a is the gradient of the underlying straight-line trend and $\boldsymbol{b} = (b_1, ..., b_n)^{\mathrm{T}}$ are the seasonal factors. The parameters a and $\boldsymbol{b}$ in the regression model are determined by minimising the difference between the data and model values in a weighted least-squares sense as described in Section 2.1. The weights are defined analogously to equation (1) for the appropriate time period. A further extension of this model to allow the underlying trend to take a more general form is considered in Section 5.

3. Evaluation of uncertainties

The regression models described above can be used to evaluate standard uncertainties associated with the model parameters. A 95% confidence interval associated with a model parameter is given by $[x - ks, x + ks]$, where x is the estimated value of the parameter, s its associated standard uncertainty, and k a coverage factor (BIPM *et al*, 1995). The coverage factor k depends on the distribution of the weighted residuals. If we assume that they are distributed normally with a mean of zero, then k is approximately two and the 95% confidence interval would be approximately $[x - 2s, x + 2s]$.[1]

Data averaging constitutes the initial part of the process of analysing the data using the regression models described here. According to the well-known Central Limit theorem (Kendall and Stuart, Vol. 1, 1966, p193), the mean of a sample of values drawn independently from the same distribution tends to normality as the size of the sample tends to infinity. Hence, it might be anticipated that Gaussian statistics would be applicable to the models considered here. However, the hourly figures on which the averages are based are *serially correlated* (Kendall and Stuart, Vol. 3, 1966, p361). The

[1] We use the notation $[x - U, x + U]$ to denote a confidence interval associated with a measurand X, where x is a best estimate of X and U an expanded uncertainty. This form is preferable to the recommendation (BIPM *et al*, 1995) $x \pm U$, because the latter cannot be generalised to asymmetric confidence intervals of the form $[x - U_1, x + U_2]$ used here.

extent of this serial correlation is clearly evidenced by forming the set of serial correlations of lag 1, 2, ... (Kendall and Stuart, Vol. 3, 1966, p362). Since the assumption of independence does not hold, it is imprudent to assume that the conditions of the Central Limit theorem apply. Hence, we make no assumption about the nature of the errors in the data and accordingly use bootstrap resampling (Efron (1982), Efron and Tibshirani (1993)) to estimate the confidence limits associated with the model parameters rather than rely on Gaussian statistics. Ott (1995, p175) also notes that hourly readings of water quality or ambient air quality usually have strong serial dependencies, and states that this fact can influence the calculation of confidence intervals.

The technique of bootstrap resampling enables non-normally distributed data to be treated robustly (Gatz and Smith, Parts I and II, 1995). It involves the repeated resampling of the residual deviations to generate new data sets that can be fitted with a regression model to establish new estimates for the model parameters. In this application, if the parameters a defining the gradient of the underlying trend and $\boldsymbol{b}$ for the seasonal factors are computed as described above then we determine the model values $f_{i,j}$ and model weighted residuals $r_{i,j}$ from

$$f_{i,j} = F_v\left(t_{i,j}, a, \boldsymbol{b}\right) \qquad \text{and} \qquad r_{i,j} = w_{i,j}\left(v_{i,j} - f_{i,j}\right)$$

The values $r_{i,j}$ form a sample from the distribution of weighted residuals of the fitted model with respect to possible values $v_{i,j}$ for ozone concentration, and therefore provide a (discrete) approximation to this distribution.

A new set of simulated data $v_{i,j}^{+}$, generally different from the values $v_{i,j}$, is constructed by adding to each model value $f_{i,j}$ a random sample r taken from the set of values $r_{i,j}$ (assuming replacement) as follows:

$$v_{i,j}^{+} = f_{i,j} + r.$$

The set of values $v_{i,j}^{+} - f_{i,j}$ forms a *bootstrap* sample from the set of residuals $r_{i,j}$. Having constructed the simulated data $v_{i,j}^{+}$, we may then process it in any way we choose. In particular, we may compute estimates a^{+} and $\boldsymbol{b}^{+}$ of the parameters in the model that fits this data in the manner described in Section 2.2.

This procedure is repeated a large number of times (N) to generate the $1 \times N$ matrix A and the $n \times N$ matrix B. The structure of each of these matrices is:

$$\begin{array}{ccc} & N \text{ bootstrap samples} & \\ A = & [a^{(1)} \; \dots \; a^{(l)} \; \dots \; a^{(N)}] & \text{gradient of trend} \end{array}$$

and

$$\begin{array}{ccc} & N \text{ bootstrap samples} & \\ B = & \begin{bmatrix} b_1^{(1)} & \dots & b_1^{(l)} & \dots & b_1^{(N)} \\ \vdots & & \vdots & & \vdots \\ b_n^{(1)} & \dots & b_n^{(l)} & \dots & b_n^{(N)} \end{bmatrix} & n \text{ seasonal factors} \end{array}$$

Each row of these matrices contains a sample from the distribution for the corresponding parameter, and therefore provides a (discrete) approximation to this distribution. Since the elements in the $1 \times N$

matrix A form a sample from the distribution for the gradient in ozone concentration, the 2.5 and 97.5 percentiles of this empirical distribution specify a 95% confidence interval associated with the value of the trend.

In this way we may derive confidence intervals at any level associated with any of the model parameters without the need to make assumptions about the distribution of residuals associated with the model. In fact, we may obtain confidence intervals associated with any quantity *derived* from the model parameters. An example of this is described later in this report, where methods for aggregating the results obtained from many sites are discussed.

There are many commercial software packages and libraries available that can be used to solve the least-squares problems described above and in Section 2. In Lawson and Hanson (1974), for example, an approach is described that is numerically stable and, furthermore, by exploiting the fact that the least-squares observation matrix is unchanged for all bootstrap samples, is efficient when used as part of a bootstrap resampling scheme.

4. Results for individual site data

The null hypothesis to be tested is that *there is no underlying straight-line trend over the time span of the data, i.e., the gradient of the underlying long-term trend in the regression model is zero*. The regression model described in Section 2.2 is fitted by weighted least-squares and the sampling distribution of the gradient of the underlying straight-line trend term determined empirically using bootstrap resampling. If the 95% confidence interval associated with the gradient, computed from this empirical distribution, does not contain zero then, in a formal statistical sense, there is reason to doubt the hypothesis that ozone concentration is not changing.

In Table 1 we list estimates of the gradient (in ppb per year) of the underlying straight-line trend obtained from fitted regression models to data for each of the sixteen sites. The table includes estimates obtained for data representing the arithmetic mean over the four time periods. Figure 4 illustrates the fits to mean ozone concentration data at one of these sites (Aston Hill). The sampling distribution of the trend for each time period estimated using bootstrap resampling is shown in Figure 5. The distributions are presented as frequency distributions, and the 95% confidence intervals obtained from the distribution are also marked.

Finally, Figure 6 shows the estimates of the trend at each site together with their associated 95% confidence intervals obtained by bootstrap resampling. The results for the weekly-averaged data indicate that there is evidence to doubt the hypothesis for seven of the sixteen sites. Furthermore, for six of these, we conclude that at the 95% confidence level the trend in ozone concentration is increasing, whereas for the remaining site it is decreasing.

For comparative purposes we also computed the 95% confidence intervals associated with the estimates in Table 1 under the (unsupported) assumption that Gaussian statistics apply. It was observed that in all cases the non-parametric confidence interval lay wholly within that for Gaussian statistics. A valid bootstrap approach can be expected to provide reliable confidence intervals that may be smaller than or greater than those obtained under the assumption that Gaussian statistics apply. Whether the intervals are smaller or greater depends on the nature of the sampling distribution, which we set out to estimate in an unbiased way. We believe our approach to be valid because of the careful model validation carried out (Sections 5 and 6). Although we offer no explanation at this stage why the confidence intervals obtained from the bootstrap are consistently smaller in this particular application, we intend carrying out subsequent work to understand better the reasons for this effect. It is to be noted that when applied to simulated data containing *Gaussian* noise, the computed confidence intervals were essentially identical, a demonstration of the fact that the observed differences for real data are not an artefact introduced by the bootstrap technique.

		Gradient of the underlying straight-line trend			
Site		Yearly	Quarterly	Monthly	Weekly
Aston Hill	AH	0.139	0.120	0.126	0.118
Bottesford	BT	0.443	0.433	0.427	0.431
Bush	BU	0.075	0.076	0.103	0.100
Eskdalemuir	ES	-0.098	-0.052	-0.047	-0.047
Great Dun Fell	GD	0.265	0.289	0.288	0.297
Glazebury	GZ	-0.084	0.022	0.039	0.042
High Muffles	HM	0.412	0.362	0.377	0.369
Harwell	HR	0.447	0.414	0.457	0.463
Ladybower	LB	0.038	0.127	0.108	0.124
Lullington Heath	LH	-0.011	-0.029	-0.043	-0.036
Lough Navar	LR	-0.013	-0.033	-0.041	-0.031
Mace Head	MH	0.160	0.123	0.099	0.097
Sibton	SB	0.130	0.157	0.157	0.162
Strath Vaich	SV	0.394	0.335	0.313	0.340
Wharleycroft	WC	0.655	0.558	0.535	0.548
Yarner Wood	YW	-0.404	-0.390	-0.408	-0.400

Table 1: Estimates of the gradient (in ppb per year) of the underlying straight-line trend obtained from fitted regression models to ozone concentration data for all sixteen sites. Results for data representing the mean of ozone concentration over yearly, quarterly, monthly and weekly time periods are shown.

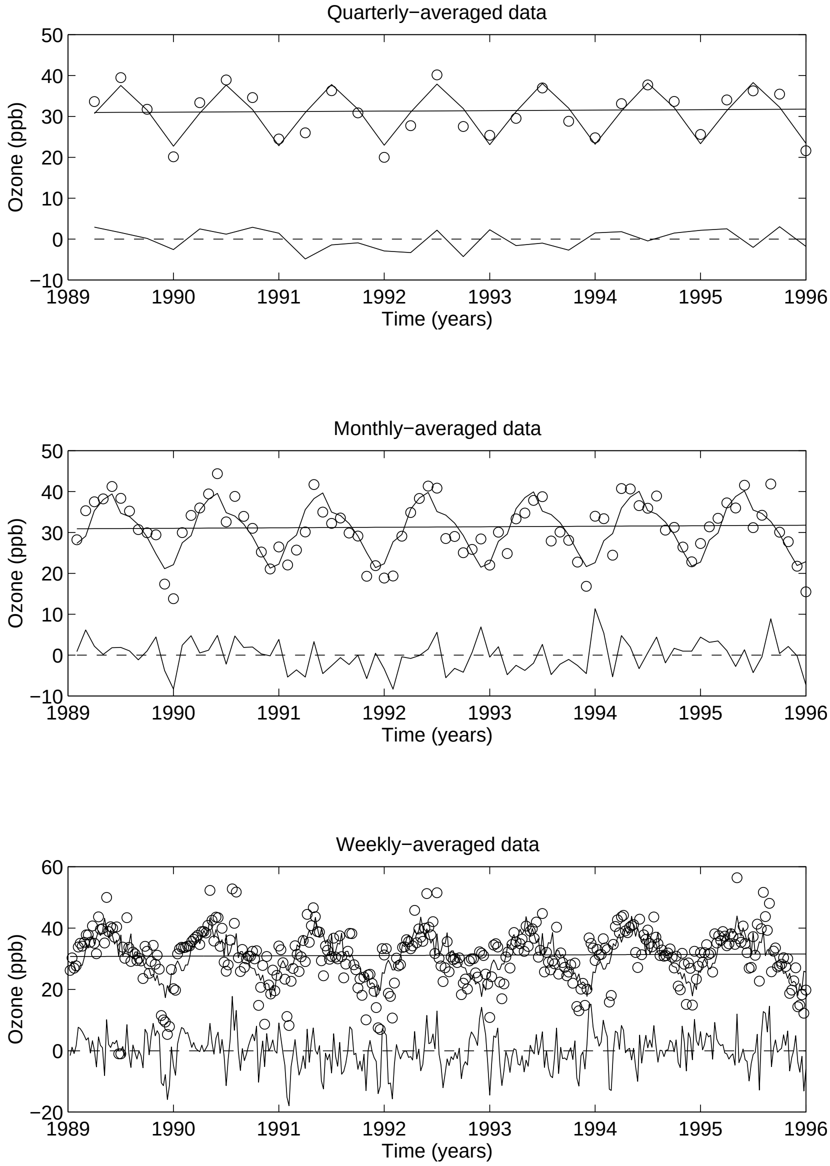

Figure 4: Fit of regression model to mean ozone concentration data obtained by averaging over quarterly, monthly and weekly time periods for a single site (Aston Hill). For each fit, the data (small circles), fitted model (straight line segments joining model values) and underlying trend are shown in the upper trace. The weighted residuals (straight line segments joining their values) and the “zero line” are shown below.

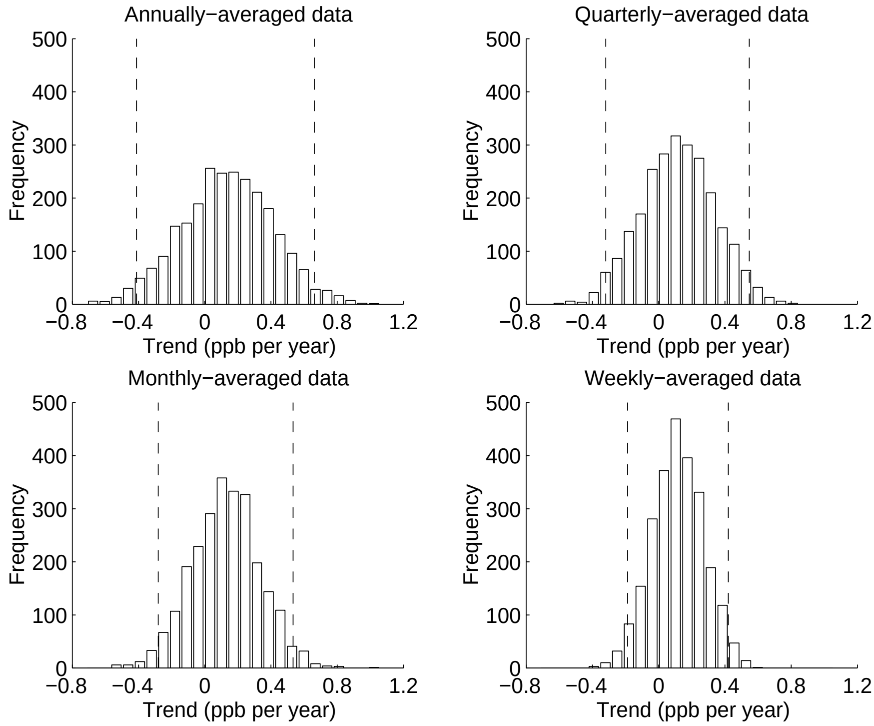

Figure 5: Sampling distributions for the long-term trend in ozone concentration at the single site (Aston Hill) for the four time periods estimated using bootstrap resampling. The 95% confidence intervals obtained from the distributions are also marked.

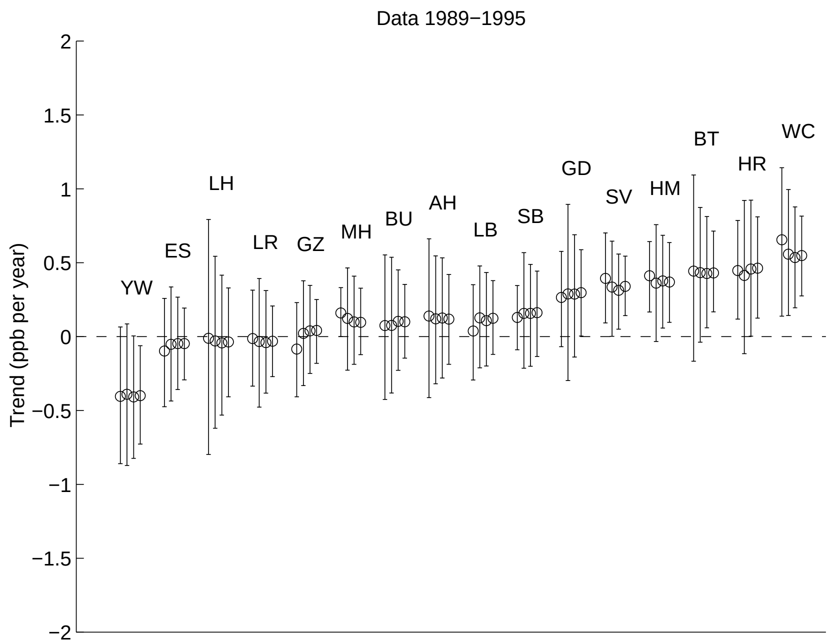

Figure 6: Best estimates (from regression) and associated 95% confidence intervals (from bootstrap resampling) for fits to mean ozone concentration data for all sixteen sites. Each group of intervals relates to, from left to right, yearly, quarterly, monthly and weekly estimates. The sites are ordered according to the values of the regression estimates for weekly averaged data.

5. Model Validation

The regression models used to fit the data consist of a number of seasonal factors together with a linear underlying trend. In the general case, if $v_{i,j}*$, $j = 1, ..., n$, are the values obtained by averaging over n equal time periods the hourly data in year i, the model takes the form:

$$v_{i,j}* = F_v\left(t_{i,j}, a, \boldsymbol{b}\right) = at_{i,j} + b_j, \quad t_{i,j} = (i-1) + \frac{j}{n}, \quad j = 1, \ldots, n,$$

where a is the gradient of the underlying straight-line trend and $\boldsymbol{b} = (b_1, ..., b_n)^T$ are the seasonal factors.

5.1 Number of seasonal factors

An important objective of validating this model is to establish some justification for the number of seasonal factors used. Models with 1, 4, 12 and 52 factors (according to whether the data was averaged on a yearly, quarterly, monthly or weekly basis) have been compared. This choice can be validated by examining the root-mean-square residual values associated with the regression fits for each number of seasonal factors.

In Table 2 the values of s_n, the root-mean-square residual value for a regression model having n seasonal factors fitted to the ozone concentration data averaged over n time periods in a year, are listed for each site. The quantity s_n is an estimate of σ_n, the standard deviation of the part of the data that is not explained by the model. It will be a valid estimate if the resulting model residuals can be regarded as independent and identically distributed. As a consequence of the data averaging process, the values of σ_n will be proportional to $\sqrt{n}$ (i.e., they increase with n). If the values of s_n are also so proportional, the models can be regarded as being equally effective in explaining the data. Referring to Table 2, as expected, s_n increases with n. Under the above assumptions the s_n would be in proportion to the square roots of 1, 4, 12 and 52, (viz, to 1.0, 2.0, 3.5 and 7.2). However, the values of s_n generally decrease at a slower rate than this; the only exceptions to this are sites identified as HM, MH and SB (see Table 2). This implies that the use of a greater number of seasonal factors generally gives greater modelling power and hence provides a better explanation of the inherent variability of the data.

5.2 Nonlinear underlying trends

In Section 5.1 we considered the effect of the number of seasonal factors used in the model on the quality of the fit to the data. Another consideration is the form of the underlying trend. In this section we consider whether there are any benefits in using a quadratic or cubic polynomial, rather than a linear function.

Table 3 lists the root-mean-square residuals s_n for fits to the data at a single site (Aston Hill) over the four time periods and with different degrees d for the polynomial used to describe the underlying trend. The 95% confidence limits associated with s_n for the straight-line trend model are obtained from the distribution of s_n determined by evaluating this statistic for 2500 bootstrap samples of the weighted residuals (Efron, 1982).

Site		Yearly	Quarterly	Monthly	Weekly
Aston Hill	AH	1.74	2.60	4.11	6.36
Bottesford	BT	2.05	2.75	3.74	5.54
Bush	BU	1.55	2.72	3.58	5.04
Eskdalemuir	ES	1.18	2.38	3.28	5.09
Great Dun Fell	GD	1.01	3.60	4.25	6.11
Glazebury	GZ	1.01	2.08	3.01	4.45
High Muffles	HM	0.77	2.37	3.08	5.52
Harwell	HR	1.09	2.94	4.49	6.79
Ladybower	LB	1.03	2.08	3.26	5.33
Lullington Heath	LH	2.45	3.48	4.94	7.74
Lough Navar	LR	1.03	2.52	3.53	5.09
Mace Head	MH	0.55	2.13	3.05	4.81
Sibton	SB	0.68	2.29	3.48	5.98
Strath Vaich	SV	0.96	1.90	2.67	4.04
Wharleycroft	WC	1.55	2.48	3.51	5.57
Yarner Wood	YW	1.52	2.90	4.36	6.93

Table 2: Root-mean-square residual values s_n (in ppb) for the fitted regression models to ozone concentration data for each of the sixteen sites. Results for data representing the arithmetic mean over yearly, quarterly, monthly and weekly time periods are shown.

For all the time periods, the value of s_n corresponding to the choice of a polynomial of degree $d = 2$ or 3 lies well within the 95% confidence interval associated with s_n for the straight-line trend model. Consequently, we conclude that there is no benefit in using a quadratic or cubic polynomial, rather than a linear function, to describe the underlying trend for this site. Analysis of the results for the other fifteen sites leads us to draw the same conclusion for all sites.

We also provide in Table 3, for the straight-line trend model and the Aston Hill data, values of a statistic defined to be the ratio of the length of the 95% confidence interval to the value of s_n, that compares with the "coefficient of variation" quoted in conventional statistical analyses (Kendall and Stuart, 1966). The value of this statistic is smallest for data averaged over weekly time periods. This suggests that the regression model with weekly factors yields, in a relative sense, the most tightly defined root-mean-square residual value s_n for this data. The same result applies for the other fifteen sites.

	d = 1			d = 2	d = 3
	s_n	$[u_L, u_U]$	$(u_U - u_L)/s_n$	s_n	s_n
Year	1.74	[0.91, 2.37]	0.84	1.69	1.95
Quarter	2.60	[2.15, 3.05]	0.35	2.50	2.47
Month	4.11	[3.48, 4.72]	0.30	4.04	4.01
Week	6.36	[5.86, 6.86]	0.16	6.31	6.28

Table 3: Root-mean-square residual s_n (in ppb) for the fits to the mean data for a single site (Aston Hill) for the four time periods using a polynomial of degree d to describe the underlying trend. The 95% confidence interval $[u_L, u_U]$ associated with s_n (obtained from bootstrap resampling) for the straight-line trend model is also presented, together with the ratio $(u_U - u_L)/s_n$.

6. Validity of the resampling process

6.1 Bias

It is possible for the bootstrap resampling process employed in our analysis to introduce biased confidence intervals. Efron and Tibshirani (1993) describe how the bias may be quantified and, if it is large, how bias-corrected intervals may be estimated. The bias correction z_0, in the notation of Efron and Tibshirani (1993), for a statistic θ is given by $z_0 = \Phi^{-1}(\rho)$, where Φ^{-1} indicates the inverse function of the standard normal cumulative distribution function and ρ is the proportion of bootstrap replicates less than the original estimate. A value of $\rho = ½$, giving $z_0 = 0$, implies that there is no bias.

When applied to the bootstrap replicates of the gradient estimated for all four time periods, we find that the values of ρ for the sixteen rural sites are all close to one half, ranging from 0.48 to 0.53, and scattered (apparently) randomly about one half. We conclude that the use of the bootstrap for this purpose is not introducing significant bias, and therefore a bias correction is not necessary.

6.2 Reliability of estimated confidence limits

Although bootstrap resampling has been used to estimate non-parametric 95% confidence intervals associated with the underlying trend, it is necessary to consider the confidence that can reasonably be placed on these intervals for the finite number N of resamplings (N = 2500 in our case) used. Berthouex and Brown (1994, p68) state that the precision of the quantiles so estimated decreases rapidly as the estimates move towards the extreme tails of the distribution. They provide formulae for quantifying this precision, as follows. Let p denote the (fractional) quantile of interest (0.025 and 0.975 in our case).

Then, 95% confidence intervals associated the pth quantile are (for a large sample, as in our case)

$$[\,p(N+1) - 1.96\{Np(1-p)\}^{1/2},\ \ p(N+1) + 1.96\{Np(1-p)\}^{1/2}\,].$$

(The value $p(N + 1)$ is that of the pth quantile used in our analysis.) The corresponding values of the statistic of interest (trend in our case) are then obtained immediately from the corresponding values in the empirical error distribution.

Figure 7 shows the cumulative distribution function for values of the slope obtained from the model fitted to weekly-averaged data for the site Aston Hill. The 95% confidence interval associated with the slope is [–0.187, 0.420]. The confidence limits in this interval correspond to, respectively, the averages of the 62nd and 63rd values, and the 2438th and 2439th values, in the (inverse) cumulative distribution function. These limits are indicated in Figure 6 using solid vertical lines. Application of the above formula with $p = 0.025$ and then $p = 0.075$ gives, approximately, [47, 78] and [2423, 2454] for the 95% confidence intervals associated with the two quantiles. This yields [–0.210, –0.169] and [0.407, 0.441] for the 95% confidence intervals associated with the endpoints of the 95% confidence interval associated with the slope. The positions of the confidence limits on the two quantiles are illustrated in Figure 6 using broken vertical lines. We find that even allowing for this "additional" uncertainty, the confidence interval is shorter than that predicted by Gaussian statistics. This result applies for all sixteen sites and all four time periods.

7. Input data quality

The raw data used in this study is nominally available as hourly averages (i.e., as a *time series*) without interruption. In practice, there will be missing data as a result of occasional malfunction of the automatic measurement stations. A consequence of malfunction is that there can be gaps in the data sequence spanning hours, days or weeks. The method of analysis presented earlier in this report accounts for such gaps by introducing "weights" w_i that reflect the quantity of data actually available within the time periods being analysed. For example, for the analysis of yearly averaged data, the weights take the form:

$$w_i = \frac{\text{number of hourly values recorded in year } i}{\text{total number of hours in year } i}.$$

The reliability of the results will depend on the manner in which data capture is handled. Data capture is an important issue and impacts on the balance between the economics of obtaining regular measurements and the quality of the results obtained from measurement analysis. Its detailed consideration is beyond the scope of this paper. However, our study has led to a number of recommendations.

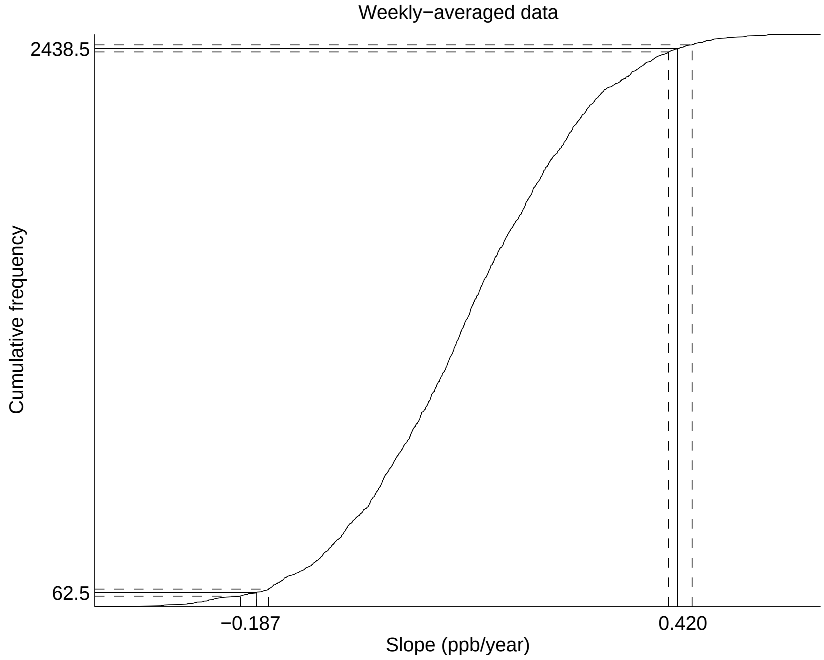

Figure 7: Cumulative frequency distribution resulting from resampling the underlying trend in ozone concentration at the single site (Aston Hill) for weekly-averaged data. The 95% confidence limits are shown (solid lines) together with the 95% confidence intervals associated with these limits (broken lines).

The use of weights to reflect percentage data capture can be criticised on a number of grounds. Firstly, the data value (time-averaged concentration) is associated with the same time value as it would be had there been 100% data capture. This association will be reasonable if the gaps in the data are approximately uniformly distributed over the time interval. However, if the gaps are significantly greater at one end of the interval, an averaged data value will be biased positionally in time. Secondly, if gaps in the data occur where extreme values of ozone concentration are expected, the resulting time-averaged concentration will itself be biased; for example, missing large values in the summer will give a time-averaged concentration that is too small.

Figure 8 illustrates the problems that can arise when there are periods of data missing. Two time series of hourly averaged data are shown. The first is the data for Aston Hill, and the second is derived from this time series by removing the following periods of data: the first six months of 1989, the months January to March and October to December of 1993 and 1994, and the last six months of 1995. Below each time series is shown the fit of the straight-line regression model to mean ozone concentration data obtained by taking yearly averages and weighting these values in the manner described above.

For the second of the time series, we see that because the data is not uniformly distributed over each year, the yearly values are biased positionally in time. The yearly averaged values are positioned at the end of the year from which they are derived and, consequently, the yearly average for 1995 is located in a period for which there is no data. (If we had chosen to locate the value at the start of the year, the result would apply to the 1989 value instead.) Furthermore, the removal of hourly data from the winter months of 1993 and 1994 gives yearly averages for these years that are high compared with the values for the first time series. The estimate of the underlying trend for the second time series is consequently higher than for the first time series. The inclusion of weights does not compensate for the absence of data in this, albeit extreme, example.

The use of models with larger numbers of seasonal factors can generally be expected to handle gaps in the data better. Consider the extreme case in which the data capture over months May-August of a particular year was zero, but was 100% for the remaining eight months. In such a case the use of yearly averages would give a value for that year that is likely to be appreciably smaller than its true (unknown) average. As a consequence, the trend line based on yearly averages would be influenced, particularly if the year in question were the first or last in the sequence being analysed. If, conversely, monthly averages were used, there would be data with 100% data capture for months 1–4 and 9–12 and zero data capture for months 5-8. All but the central set of months would contribute meaningful values to the analysis. The central months would make no contribution (through missing, but not biased, data), and the overall analysis would automatically take account of data present in the corresponding months of other years to define the parameters of the model.

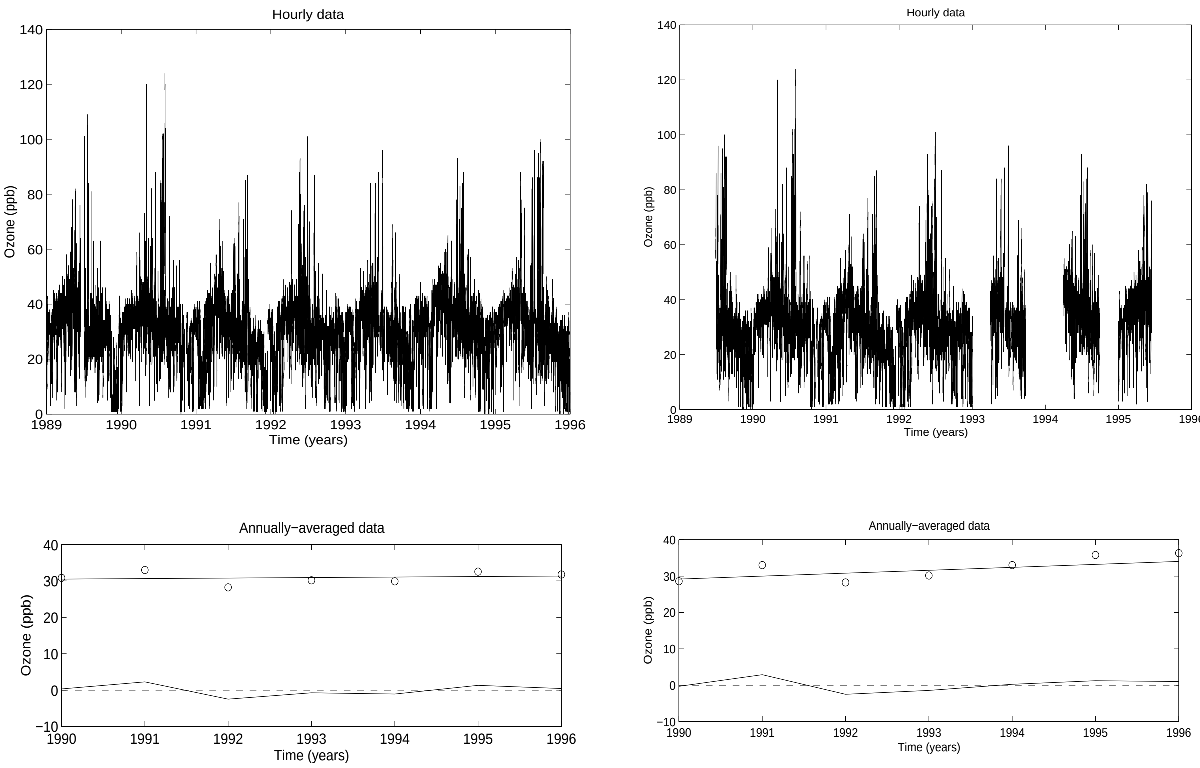

Figure 8: Time series of hourly averaged ozone concentration data at Aston Hill with and without periods of data missing. The fits of a straight-line regression model to mean ozone concentration data obtained by taking yearly averages of the hourly data are shown below.

8. Estimating trends across all sites

In Section 4 the long-term trend in ozone concentration at each site was reported. In this section we aim to consider how the results obtained for each site can be aggregated to evaluate the long-term trend over the whole network. This is equivalent to testing the statistical hypothesis that *there is no long-term change in ozone concentration over the network.*

There are many ways of combining the results from the sixteen sites to obtain representative values for the whole network. In this work two approaches have been considered:

1. The mean of the individual site values is taken.

2. The number of sites at which the long-term trend is increasing.

Computing either of these statistics is straightforward, but evaluating their associated uncertainties is more difficult because only sixteen data points are available. We show here how bootstrap resampling (Efron (1982), Efron and Tibshirani (1993)) can be used to overcome this problem.

As a result of the calculations described earlier, for each of the p sites, a row vector A_i, i =1, ..., p, of length N, that describes the sampling distribution of the gradient in ozone concentration was calculated. These vectors are arranged into a single $p \times N$ matrix G that contains all N = 2500 bootstrap estimates from all sixteen sites:

$$\begin{array}{ccc} & \text{2500 bootstrap samples} & \\ G = & \begin{bmatrix} \cdots & \cdots & \mathrm{A}_1 & \cdots & \cdots \\ & & \vdots & & \\ \cdots & \cdots & \mathrm{A}_i & \cdots & \cdots \\ & & \vdots & & \\ \cdots & \cdots & \mathrm{A}_{16} & \cdots & \cdots \end{bmatrix} & \begin{array}{c} \text{Site 1} \\ \\ \text{Site } i \\ \\ \text{Site 16} \end{array} \end{array}$$

This array of bootstrap estimates can form the basis for estimating the confidence interval associated with any estimator formed by taking a function (for example, the arithmetic mean, or the number of positive values) of the gradients. For a specific estimator, we form its value for each column of the matrix G, (i.e., for each bootstrap sample). The result is a $1 \times N$ matrix whose elements estimate the sampling distribution for that estimator. The 2.5 and 97.5 percentiles of this empirical distribution specify a 95% confidence interval associated with the value of the estimator.

8.1 Average over all sites

A straightforward approach to evaluating the trend over the whole network is to take the average of the trends at each of the sites. To estimate the confidence interval associated with this value, we average the elements in the columns of the matrix G to form 2500 bootstrap values of the network average. These values estimate the sampling distribution of the network average. In Table 4 we list the average of the gradient values obtained from the original regression fits to the data for each site. This provides an estimate of the overall trend value, and its associated 95% confidence interval obtained from the estimated sampling distribution. In Figure 9 we show the estimated sampling distributions as frequency distributions for each time period. The 95% confidence intervals are also shown.

For no time period does the 95% confidence interval contain zero. Consequently, the results for all time periods indicate that there is evidence to doubt the hypothesis that there is no long-term change in the mean concentration of ozone over the whole network.

	Mean trend	
	Estimate	95% conf. int.
Year	0.159	[0.055, 0.263]
Quarter	0.157	[0.045, 0.266]
Month	0.156	[0.065, 0.249]
Week	0.161	[0.091, 0.231]

Table 4: Estimates of, and 95% confidence intervals associated with, the mean long-term trend across all sixteen sites. Results for data representing the mean of ozone concentration over yearly, quarterly, monthly and weekly time periods are shown.

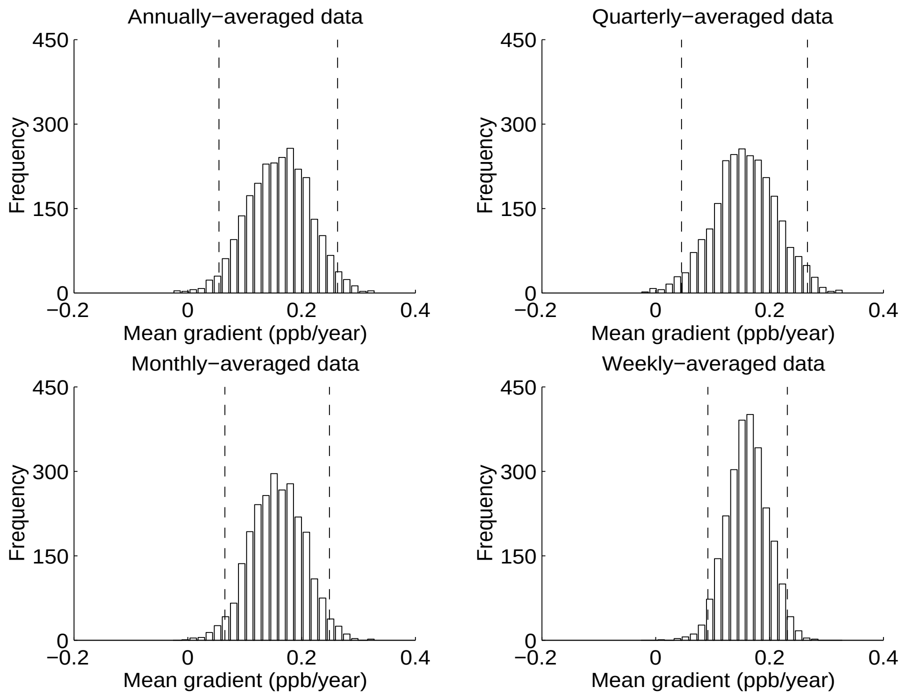

Figure 9: Sampling distribution for the average long-term trend in ozone concentration across all sixteen sites for the four time periods estimated using bootstrap resampling. The 95% confidence intervals obtained from the distributions are also marked.

8.2 Number of sites with a positive trend

An alternative approach to aggregating the results over all sites is to count the number of sites with a positive trend. This number provides a discrete measure of the underlying trend in ozone concentration for the network: if there are significantly more sites with a positive trend than with a negative trend, this is interpreted as indicating that ozone concentration is increasing over the network. To estimate the confidence interval associated with this aggregated measure, we count the number of positive gradient values in each column of the matrix G. The resulting bootstrap values estimate the sampling distribution of the number of positive gradient values. Table 5 lists the number of positive gradient values obtained from the original regression fits to the data for each site. This provides an estimate of the number of positive trends across the network, and its associated 95% confidence interval obtained from the estimated sampling distribution. In Figure 10 we show the estimated sampling distributions as frequency distributions for each time period. The 95% confidence intervals are also shown.

Our interpretation of the null hypothesis that *there is no long-term change in ozone concentration over the network* is that *the number of sites with a positive trend is equal to the number with a negative trend*. For a network consisting of sixteen sites, the null hypothesis to be tested is that *there are eight sites with a positive trend*. The results for the monthly- and weekly-averaged data indicate that at the 95% level there is evidence to doubt this hypothesis. The results for the other two time periods are finely balanced in this respect.

	No. of positive trends	
	Estimate	95% conf. int.
Year	11	[8, 14]
Quarter	12	[8, 14]
Month	12	[9, 14]
Week	12	[9, 14]

Table 5: Estimates of, and 95% confidence intervals associated with, the number of positive trends across all sixteen sites. Results for data representing the mean of ozone concentration over yearly, quarterly, monthly and weekly time periods are shown.

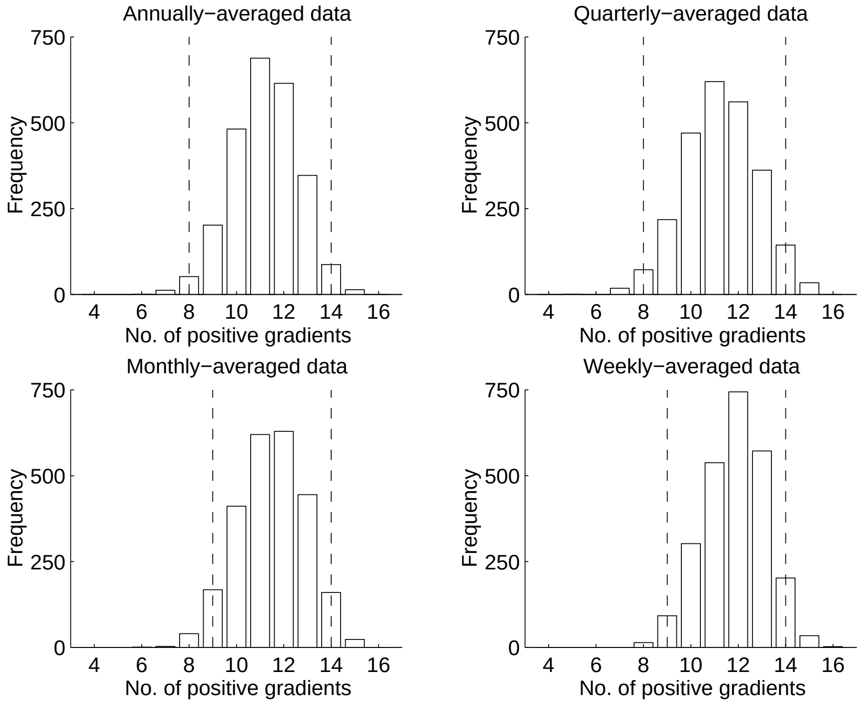

Figure 10: Sampling distribution for the number of sites with a positive trend across all sixteen sites for the four time periods estimated using bootstrap resampling. The 95% confidence intervals obtained from the distributions are also marked.

9. Discussion

As a result of this work, it is recommended that ozone concentration data is analysed with regression models of the type presented in Section 2. This involves using weighted models containing an underlying linear trend function and a number of seasonal factors. Moreover, it is recommended that 95% confidence intervals associated with the parameters of these models are estimated using bootstrap resampling. These recommendations are made for the following reasons:

1. The long-term trend is available directly as one of the parameters in these models. Furthermore, the seasonal factors provide information about the variation in ozone concentration over and above any underlying trend.

2. These models may be used to allow for the effects of percentage data capture by introducing weights associated with the time-averaged data.

3. The use of bootstrap resampling for evaluating confidence intervals has significant advantages over the use of conventional Gaussian statistics. For example, no assumption is made concerning the error distribution of the un-modelled part of the data and, in general, tighter confidence intervals are obtained by this means.

4. Since the regression models form a *family* of models (with respect to the number of seasonal factors and degree of the polynomial representing the underlying trend), within-family *model validation* can be carried out to determine the appropriateness of the selected model.

The models operate with data constituting time-averaged values derived from the hourly averages of ozone concentration measured at the sixteen sites in the UK Rural Ozone Monitoring Network. Of the possible time-averaged values that could be taken for this purpose, it is recommended that weekly averages should be used for the following reasons:

1. The use of weekly averages rather than averages over *longer* time periods is likely to take fuller account of the fine structure present in the data.

2. The use of averages over time periods *shorter* than one week is likely to introduce more detail into the analysis than can be justified statistically, particularly in regard to the large number of seasonal factors (365 if days are used) and the difficulty of "alignment" of corresponding days because of the "Leap Year" effect.

3. There is a *tendency* for the confidence interval for models with larger numbers of seasonal factors to provide shorter confidence intervals for the trend parameter (Figure 5).

4. The use of models with larger numbers of seasonal factors can generally be expected to handle gaps in the data better.

An examination of Figure 6 indicates that there is reasonable consistency between the estimates provided by the models for each time interval. In most cases the 95% confidence intervals overlap strongly. (It must be stressed, however, that this observation does not apply consistently: see, e.g., the sites Sibton and Great Dun Fell. An extreme example can easily be constructed for which a yearly-based model fits the data *exactly* and a quarterly-based model does not, giving rise to an infinitely better confidence interval (of zero length) for the first model and one of finite length for the second.)

10. Conclusions

The accurate evaluation of long-term trends in ozone concentration data presents a challenge for statistical analysis. In this type of work, the data is often aggregated so that the number of data points fitted by the model is limited. Additionally, the residual deviations may not be distributed normally. Consequently, conventional methods for evaluating confidence intervals may give misleading results. This work contends that regression models with factors representing seasonal changes can provide a robust basis for the evaluation of long-term trends. The model residuals themselves have been used to estimate the distribution function of the variability in the data that is unaccounted for by the model. By extensive bootstrap resampling of this distribution we have estimated the distribution function of the long-term trend and calculated its associated 95% confidence interval without any additional assumptions. We believe that the approach used here has advantages over other methods that are in widespread use. We have carried out a number of tests to demonstrate that it is robust and to validate its use in this application. We have also explored the difficulties of evaluating trends in data sets that have missing values, and the aggregation of the results for individual sites to give measures of the long-term trend in ozone concentration over a network of sites.

We have also considered the question "Is ozone increasing in the UK?". Our most robust answer, based on the analysis of weekly averages, is that the mean yearly increase between 1989 and 1995 across the sixteen sites in the UK Rural Ozone Monitoring Network is 0.161 ppb per year with an associated 95% confidence interval of 0.091 ppb to 0.231 ppb. The resampling technique used to obtain the estimated confidence interval does not assume that the measured data (or the mean trend at each site) are normally distributed. We believe the confidence interval calculated this way gives a good reflection of the basic variability of the data. The approach is straightforward to implement and has applications to other types of ambient air quality data.

11. Acknowledgements

We would like to thank Dr P Ciarlini (CNR-Istituto di Applicazioni Calcolo, Rome) and Mr G F Miller (ex-NPL) for helpful discussions during this work. We would also like to thank our colleague Godfrey Cowin for his assistance in preparing this report.

12. References

BERTHOUEX, P. MAC and BROWN, L.C. (1994). *Statistics for Environmental Engineers*. CRC Press, Inc., US.

BIPM, IEC, IFCC, ISO, IUPAC, IUPAP and OIML. (1995). Guide to the Expression of Uncertainty in Measurement. ISBN 92–67–10188–9, Second Edition.

EFRON, B. (1982). *The Jackknife, the Bootstrap, and Other Resampling Plans.* Society for Industrial and Applied Mathematics, Philadelphia.

EFRON, B. and TIBSHIRANI, R. (1993). An Introduction to the Bootstrap. Monographs on Statistics and Applied Probability 57. Chapman and Hall.

ELDRED, R.A. and CAHILL, T.A. (1994).Trends in elemental concentrations of fine particles at remote sites in the United States of America. *Atmospheric Environment*, **28** (5), pp 1009–1019.

GATZ, D.F. and SMITH, L. (1995). The standard error of a weighted mean concentration – I. Bootstrapping vs other methods. *Atmospheric Environment*, **29** (11), pp 1185–1193.

GATZ, D.F. and SMITH, L. (1995). The standard error of a weighted mean concentration – II. Estimating confidence intervals. *Atmospheric Environment*, **29** (11), pp 1195–1200.

KENDALL, M.G. and STUART, A. (1966). *The Advanced Theory of Statistics. Volume 1: Distribution Theory*. Charles Griffin and Company Limited, London.

KENDALL, M.G. and STUART, A. (1966). *The Advanced Theory of Statistics. Volume 3: Design and Analysis, and Time Series*. Charles Griffin and Company Limited, London.

LAWSON, C.L. and HANSON, R.J. (1974). *Solving Least Squares Problems*. Prentice Hall, Englewood Cliffs, New Jersey.

OLTMANS, S.J. *et al* (1998). Trends in ozone in the troposphere. *Geophysical Research Letters*, **25** (2), pp 139–142.

OTT, W.R. (1995). *Environmental Statistics and Data Analysis*. CRC Press, Inc., US.

PORG (1993). Ozone in the United Kingdom 1992. Third report of the United Kingdom Photochemical Oxidants Review Group, Harwell laboratory, Oxfordshire, UK.

SHI, J.P. and HARRISON, R.M. (1997). Regression modelling of hourly NOx and NO2 concentrations in urban air in London. *Atmospheric Environment*, **31** (24), pp 4081–4094.