Air Pollution and Human Health: A Study Based On Emergency Room Visits Data from Prince George, British Columbia

James Ahkong, B. M. Golam Kibria and A. John Petkau
Statistical Consulting and Research Laboratory
Department of Statistics
University of British Columbia
Vancouver, B.C., V6T 1Z2
Canada

February 4, 2000

Acknowledgement

The authors would like to thank Mr. Dennis Fudge, Air Pollution Meteorologist, Pollution Prevention, Ministry of Environment, Lands and Parks, Prince George, Mr. Gary Myers, Superintendent of Climate Services, Pacific and Yukon Regions, Environment Canada, Ms. Angeli Mitra, Research Associate, Northern Interior Regional Health Unit, Prince George and Lecturer, Department of Nursing, University of Northern British Columbia, Prince George, and Dr. Lorna Medd, Medical Health Officer, Northern Interior Regional Health Unit, Prince George, for providing the air pollution, meteorological and emergency room visits data and clarifying various aspects of the data at the early stages. We are also thankful to Mr. Rick White of the Statistical Consulting and Research Laboratory, Department of Statistics, The University of British Columbia, for his advice throughout the course of the project.

Abstract

The possibility of relationships between ambient levels of air pollution and mortality/morbidity is of concern to environmental agencies throughout the world. This report presents the results of such a study based on emergency room (ER) visits from the single hospital in Prince George, a city of about 75,000 population in British Columbia. The current study is a follow-up to earlier studies based on hospital admissions reported in Knight et al (1988) and on emergency room (ER) visits reported in Knight et al (1989) and in McNeney and Petkau (1991). In this study, overdispersed Poisson regression models are developed to estimate the short-term increased risk of daily ER visits associated with three air pollutants (particulate matter less than 10 microns (PM_{10}), sulfur dioxide (SO_2) and ozone (O_3)), while controlling for temporal and meteorological effects.

A model is first developed for the temporal and meteorological effects based on loess smooths. This serves as the base model in developing the final models incorporating the effects of the pollutants. The pollutants were first considered individually and then simultaneously. The reduction in deviance due to the addition of the pollutant effects is used to assess the strength of the relationships between the pollutants and the ER visit counts.

When the pollutants are considered individually, PM_{10} at lags of one and three days, same day ozone, and sulfur dioxide at a lag of four days have non-negligible effects on the ER visit counts. A decrease of PM_{10} from 20 $\mu g/m^3$ to 10 $\mu g/m^3$ both one and three days previously is associated with a predicted decrease in ER visits counts of about 0.2% (95% CI[1] from -1.9% to 2.3%). The fitted relationship with ozone is quadratic which complicates the interpretation. A decrease in ozone from 30 $\mu g/m^3$ to 15 $\mu g/m^3$ on the same day is associated with a predicted increase in ER visit counts of 1.4% (95% CI from -2.8% to 5.9%). On the other hand, a decrease from 45 $\mu g/m^3$ to 30 $\mu g/m^3$ is predicted to result in a 3.5% (95% CI from -0.4% to 7.2%) decrease in the ER visit counts. A decrease in sulfur dioxide from 6 $\mu g/m^3$ to 3 $\mu g/m^3$ four days previously is associated with a predicted 0.4% (95% CI from 0.1% to 0.8%) decrease in the ER visit counts. When the pollutants are considered simultaneously, ozone has most significant effect on the ER visit counts, while PM_{10} has the least significant effect.

[1] CI stands for confidence interval

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In this study, we examine possible statistical relationships between air pollutants and emergency room (ER) visits based on data from a single community: Prince George, BC. Prince George is an industrial city with a population of about 75,000 located at the geographical centre of the Canadian province of British Columbia. The available data consists of daily counts of emergency room visits for respiratory diseases and corresponding values of the air pollution parameters particulate matter less than 10 microns in diameter (PM_{10}), sulfur dioxide (SO_2), and ozone (O_3) for the period of July 1, 1995 to December 31, 1997.

Detailed background information on the possible effects on respiratory hospital admissions in Prince George has been described by Knight et al (1988). Their study, based on data from April 1, 1984 to March 31, 1986, considered total reduced sulphate (TRS) and total suspended particulates (TSP) as the air pollution parameters and found that TRS and TSP (considered separately) do not have significant effects on hospital admissions. The follow-up study of Knight et al (1989) considered emergency room visit counts data instead of hospital admissions and concluded that TRS and TSP (considered separately) each have an effect on emergency room visits. However, the magnitude of the effects was very small in both cases. It was recommended to collect data on additional air quality parameters, such as sulfur dioxide, ozone, NO_2 and other pollutants. Another follow-up study by McNeney and Petkau (1991) considered the separate effects of TRS, TSP and sulfur dioxide on emergency room visits from April 1, 1986 to March 31, 1988. In this study it was evident that TRS and TSP have significant positive effects on emergency room visits during fall, significant negative effects during winter and small effects during spring and summer. Sulfur dioxide had a negligible effect on emergency room visits.

Our study is motivated by these previous studies. We consider PM_{10}, sulfur dioxide and ozone as air pollution parameters. We use over-dispersed Poisson regression models for the analysis of the data and provide detailed results of the analysis. The modeling of the effects of these pollutants on the emergency room visit counts will also take into account temporal and meteorological variables (temperature, relative humidity and wind).

Studies of the association between air pollution and either morbidity or mortality have been considered by various researchers, such as Samet et al (1981), Bates and Sizto (1987), Kelsall et al (1997) and, very recently, Kelsall et al (1999). These authors have considered various methods to assess the associations with air pollution. Detailed background and a list of earlier references related to this topic is provided by Knight et al (1988).

The organization of this report is as follows: Chapter 2 provides a description of the data. Chapter 3 provides a brief overview of the method of the data analysis. Chapter 4 summarizes the modeling of the temporal and meteorological effects. Chapter 5 summarizes the modeling of the pollutants, both individually and simultaneously. A summary and some concluding remarks are presented in Chapter 6. In addition, the results of some sensitivity analyses are presented in an appendix.

Chapter 2

Data Descriptions

2.1 Emergency Room (ER) Visits Data

The emergency room (ER) visit counts are of the number of ER visits to the single hospital in Prince George. The visits of interest for this study are those that occurred from July 1 1995 to December 31 1997. Only visits corresponding to certain code numbers representing a diagnosis of a specific disease are studied.[1] The visits are recorded on a daily basis, with only ER visits from residents of Prince George included in the study. A count of 1 was attributed to the day of the entry to the emergency room.

Each of the 915 days of data registered at least one visit and up to 32 visits were recorded on a given day. The time series plot of Figure C.1[2] indicates the presence of cyclical patterns in the data. The histogram of Figure C.2 shows the distribution of the ER visits counts to be only moderately positively skewed and centered around a value of about 10 visits. The summary of the ER visits counts in Table 2.1 (see p.4) shows mean and median values of close to 10 visits which agree with the histogram. Boxplots of the ER visits counts by day of the week and month of the year are presented in Figures C.3 and C.4 respectively.

[1]Discharge codes used were 346, 401, 410, 413, 429, 460, 461, 462, 463, 465, 486, 487, 490, 492, 493, 785 and 786.

[2]See Appendix C for all plots

Table 2.1: Overall ER Counts of Visits

Mean	10.76
SD	4.88
Minimum	1.00
1st Quartile	7.00
Median	10.00
3rd Quartile	13.00
Maximum	32.00

Table 2.2: ER Counts by Day of the Week

Day	*Mean*	*SD*
Monday	10.99	4.80
Tuesday	9.98	4.00
Wednesday	10.88	4.60
Thursday	12.70	4.80
Friday	9.92	4.26
Saturday	9.80	5.32
Sunday	11.05	5.62

The summary of the ER visits by day of the week in Table 2.2 as well as the boxplots (Figure C.3) show a tendency for higher counts during midweek (Thursday) and late weekends (Sunday and Monday). The summary of the ER visits by month of the year in Table 2.3 (see p.5) and the corresponding boxplots (Figure C.4), show lower counts in the summer months and higher counts in the winter months. January is the only month that is an exception to the overall pattern.

Table 2.3: ER Counts by Month of the Year

Month	*Mean*	*SD*
January	9.68	4.18
February	13.21	4.76
March	13.82	5.40
April	11.27	4.19
May	10.02	4.17
June	9.65	4.46
July	8.48	4.35
August	8.19	4.02
September	10.50	4.32
October	10.25	4.33
November	11.76	4.49
December	13.44	5.66

2.2 Meteorological Data

Measurements for three meteorological variables, temperature, relative humidity and wind were provided. All measurements for the meteorological data are from the Prince George airport.

2.2.1 Temperature

Temperature was recorded hourly in degrees celsius from July 1, 1995 to July 31, 1998. There were 36 out of the 27,408 hours of missing values for temperature. The algorithm used to fill-in these and all other missing values in the study is described in Appendix A. The filled-in hourly temperature data was then converted to daily data by taking the average over the 24 hours in a day.

Both the time series plot in Figure C.5 and the boxplots by month in Figure C.6 indicate a clear and very strong seasonal pattern in daily average temperature in Prince George. The winter produces low (and occasionally very low) temperatures and the summer produces warmer daily average temperatures. Therefore temperature is quite cyclical throughout a given year.

The summary statistics of the daily average temperatures provided in Table 2.4 (see p.6) show the high variability of temperature throughout the period of this study. Indeed this confirms that Prince George has a large temperature difference between winter and summer days.

Table 2.4: Daily Average Temperature in $^{\circ}C$

Mean	5.00
SD	10.67
Minimum	-32.00
1st Quartile	-1.00
Median	6.00
3rd Quartile	14.00
Maximum	24.00

2.2.2 Relative Humidity (RH)

The daily maximum and minimum values of relative humidity (RH) were measured as percentages from July 1, 1995 to Dec 31, 1997. There were 62 days of RH data missing for both the maximum and minimum readings (both had the exact same days missing). These missing values were once again filled-in separately for maximum and minimum RH data.

The time series plots of Figure C.7 for maximum and minimum RH indicate that the majority of maximum RH values exceed 90% while minimum RH values are very cyclical and have some values that exceed 90% and that fall below 50%. The boxplots in Figure C.8 indicate that maximum RH has higher values for the fall and for February, while minimum RH has higher values for late fall and early winter (November and December) and lower values for early summer (May and June). The summary statistics for the RH data given in Table 2.5 show the high variability of the minimum RH values and the maximum RH values being distributed mainly above 90%. Both RH values also have a large range.

Table 2.5: Relative Humidity in Percentage

	Daily Minimum	*Daily Maximum*
Mean	53.35	89.97
SD	18.79	10.24
Minimum	13.00	39.00
1st Quartile	39.00	87.00
Median	50.00	93.00
3rd Quartile	66.50	97.00
Maximum	100.00	100.00

2.2.3 Wind

Wind speed and direction was measured hourly over the same time period (July 1, 1995 to July 31, 1998) as the temperature and RH data. Wind direction was measured in degrees ranging from 0° to 359° and wind speed was measured in meters per second (m/sec). Wind had more missing values than temperature with 650 and 47 out of 27,408 hours of missing for direction and speed respectively. The filling-in algorithm cannot be used to fill-in wind direction directly since 359° and 0° must be considered adjacent for such directional data. The filling-in algorithm[3] would consider 359° and 0° to be completely different because it considers these values in a linear fashion. To allow this wind data to be used in the models, polar co-ordinates were used to convert wind direction and wind speed into "horizontal (H)" and "vertical (V)" components as follows:

$$\begin{aligned} \text{Wind H} &= \text{Wind speed} \times \cos(\text{Wind direction}) \\ \text{Wind V} &= \text{Wind speed} \times \sin(\text{Wind direction}). \end{aligned} \tag{2.2.1}$$

The missing values were then filled-in separately for each component and daily averages were subsequently evaluated. Wind H is interpreted as wind in the east-west direction while Wind V is wind in the north-south direction.

The time series plots and the boxplots by month in Figures C.9 and C.10 respectively do not show any strong seasonal patterns nor any powerful winds in the north, south, east or west directions. The summary of the wind components provided in Table 2.6 (see p.8) show relatively low values for both wind components. The horizontal wind component tends to be negative which indicates more westernly wind. The vertical wind component also tends to be negative indicating more southerly wind. Judging by the means and standard deviations, the horizontal wind component is somewhat more prominent than the vertical wind component.

[3] See Appendix A

Table 2.6: Daily Average Wind Components (m/sec)

	Wind H	*Wind V*
Mean	−0.43	−0.13
SD	1.94	1.26
Minimum	−8.34	−5.08
1st Quartile	−1.34	−0.82
Median	−0.36	−0.05
3rd Quartile	0.38	0.63
Maximum	7.62	4.50

2.3 Pollutants

The three pollutants considered in this study are particulate matter of 10 microns or less in diameter, ozone and sulfur dioxide. These three pollutants were measured hourly at the Plaza station from July 1, 1995 to July 31, 1998. Only data from the same time period as that of the ER visit counts was used in the analysis, but the data for 1998 were used to help fill-in missing values.

Each of the three pollutants have systematically missing data due to the need for calibration. Every 25 hours or so the measuring equipment must be re-calibrated, so no value for the pollutant is recorded in that hour. We simply handle these as missing values and fill them in along with the other missing values that occur. Also on November 1, 1997 the equipment used to measure all the pollutants was updated. After that date, the pollutants are recorded to the nearest decimal place; details are provided in the following subsections.

2.3.1 Particulate matter less than 10 microns in diameter (PM_{10})

PM_{10}, measured in micrograms per cubic meter ($\mu g/m^3$), had 340 out of 27,408 hours of missing data. Prior to November 1997, PM_{10} was recorded to the nearest integer, but after November 1997 decimal places were also included in the data. After filling-in, the hourly PM_{10} data were converted into daily averages.

The time series plot and boxplots of Figures C.11 and C.13 respectively, reveal at least three potential extreme values for PM_{10} (indicated by the 3 dashes at the top of the boxplots for February and March). These three values are quite a bit larger than 100 while the other values are all well

Table 2.7: Daily Average PM_{10} in $\mu g/m^3$

Mean	20.58
SD	14.59
Minimum	2.83
1st Quartile	11.33
Median	17.04
3rd Quartile	25.31
Maximum	150.60

below 100. PM_{10} plotted by month seems to have two cycles with peaks in February and September and lower values in the middle of summer and around the winter holidays. The histogram (Figure C.12) shows the skewness of the distribution of PM_{10} towards the lower values. The summary statistics for PM_{10} are provided in Table 2.7.

2.3.2 Ozone (O_3)

Ozone (O_3), measured in $\mu g/m^3$, had 1,328 out of 27,408 hours of missing data. Prior to November 1997, O_3 was recorded only in even-numbered integers but after the equipment update O_3 was recorded for every integer to the nearest decimal place. The filled-in O_3 data was converted into daily averages as well.

Ozone is seasonal as seen in the time series plot and boxplots of Figures C.14 and C.16 respectively. The highest values of O_3 are in the spring months, following the very sharp increase in the daily average values in February and March. It gradually decreases every subsequent month until the winter months which have the lowest O_3 values. The histogram of Figure C.15 shows O_3 to follow a modestly skewed (positively) distribution with a majority of its values at "medium" levels. The summary statistics for O_3 are provided in Table 2.8 (see p.10).

2.3.3 Sulfur Dioxide (SO_2)

Sulfur dioxide (SO_2), measured in $\mu g/m^3$, had 1,495 out of 27,048 hours of missing data. Prior to November 1997, SO_2 was truncated to certain values (for example, readings below 3 became 0). After the equipment update, SO_2 was measured to the nearest decimal place. Subsequently, the

Table 2.8: Daily Average O_3 in $\mu g/m^3$

Mean	33.02
SD	18.00
Minimum	1.42
1st Quartile	19.80
Median	31.25
3rd Quartile	45.06
Maximum	87.81

filled-in hourly data was also converted into a daily average.

The majority of the daily average SO_2 values are close to zero as seen in the time series plot, histogram and the boxplots of Figures C.17, C.18 and C.19 respectively. The largest values of SO_2 tend to occur in the winter months (December to February) when SO_2 values are also quite variable. SO_2 values are smaller and less variable in the summer months (June to August). The summary statistics for the SO_2 data are shown in Table 2.9.

Table 2.9: Daily Average SO_2 in $\mu g/m^3$

Mean	15.04
SD	21.67
Minimum	0.00
1st Quartile	2.40
Median	6.08
3rd Quartile	18.58
Maximum	170.00

Chapter 3

Data Analysis

The goal of the analysis is to examine the relationships between the levels of pollutants and the ER visit counts from July 1, 1995 to December 31, 1997. The underlying question is whether a change in pollutant levels induces a change in the ER visit counts, be it a same day effect or a delayed effect several days later. This question is addressed using regression techniques where the ER visit counts are regressed on temporal and meteorological variables as well as on pollutants. The details are discussed in chapters to follow.

From the summary statistics of the ER visits data given earlier, it is clear that many of the visit counts are quite small. Therefore using ordinary regression techniques (such as linear regression) is not likely to be adequate. Instead we used Poisson regression[1], which seems to be a reasonable way to proceed with the analysis.

[1]See Appendix B for a brief overview of this method.

Chapter 4

Preliminary Modeling

Before examining the relationships between the ER visit counts and the air pollutants, temporal and meteorological patterns must be taken into account. This is because we want to capture the true effects of the pollutants on the ER visit counts, so we must take into account other effects on the ER visit counts. Although there may well be other factors influencing the ER visit counts, the temporal and meteorological patterns used as predictors in the following sections accommodate the factors most often considered as potential confounders of relationships between the ER visit counts and air pollution parameters.

4.1 Temporal Patterns

Based on the time series plot (Figure C.1) of the ER visit counts, it is evident that there are clear seasonal patterns in the data. The counts in winter tend to be larger than in summer (see Figure C.4). There are also potential differences depending upon the day of the week (Figure C.3 and Table 2.2) with higher counts on Thursday, Sunday and Monday. In addition, there could be a linear trend over time in the counts if, for example, the population of Prince George changed in a systematic fashion over the time period of the study. A possible model[1] adjusting only for such

[1] All models are stated in a generalized form rather than in mathematical notation

temporal patterns is

$$\log \lambda_i = \text{Linear Time Effect} + \text{Day of the Week Effects} + \text{Seasonal Effect},$$

where λ_i is the admission rate to the emergency room for day i and i=1,...,915.

The counts by month of the year presented in Table 2.3 and boxplots in Figure C.4, show some seasonality in that months clustered together tend to have similar counts. However January and July have rather different counts from the adjacent months, so grouping months together into arbitrary "seasons" would not give a good fit. To best capture the seasonal patterns, a generalized additive model (based on a loess smooth) was fit using the S-Plus software package. This software automatically determines the smooth function of time which provides the best fit to the ER visits data by fitting regression models to the data locally.

The determination of the best-fitting smooth function depends on a specified span (or spacing). For example, if the value of the span is specified as 0.1, to fit a value at any particular time, the one-tenth of the data points that are closest (in time) to that time are used in a regression analysis (the form of the regression function can be specified as constant, linear, quadratic, etc). The fitted regression function is used to determine the value of the smooth function at the particular time and the entire process is repeated for each time point. The larger the span used, the smoother the resulting fit becomes. Hence large spans result in fits that are close to a straight line (if a linear regression is carried out), while small spans result in fits that track closer to the observed values of the response (ER visit counts in our case). So very large and very small spans are not usually recommended. In the body of this report, the span of 0.1 is used to avoid "over-smoothing" the temporal and meteorological effects. However, as a sensitivity analysis, we repeated some of the analyses with span of 0.5; these results are presented in Appendix D.

Taking into account over-dispersion, various models for the patterns over time were fitted as indicated in Table 4.1 (see p.14). The p-values in Table 4.1 are based on comparing the fitted model to an appropriate simpler model to determine the significance of adding the new effect by examining whether the reduction in the deviance associated with the larger model is substantial relative to the additional number of degrees of freedom used in the model[2]. The addition of the

[2] A p-value of 0.05 or less is usually taken as indicating that the addition of the new effect is an important improvement to the fit

Table 4.1: Temporal Models

Model Terms	Deviance	df	Δdev	Δdf	p-value
Constant Rates Model	2015.6	914.0			
Linear Time Effect	2009.5	913.0	6.1	1.0	0.10
Day of the Week Effects	1942.8	908.0	72.8	6.0	<0.001
Day of the Week Effects + Linear Time Effect	1936.7	907.0	6.1	1.0	0.09
Day of the Week Effects + Smooth Time Effect	1642.5	890.0	300.3	18.0	<0.001

day of the week effects to the constant rates model results in a substantial reduction in deviance (Δdev=72.8, Δdf=6, p-value < 0.001). In particular, this has a much greater impact than adding the linear time effect to the constant rates model (Δdev=6.1, Δdf=2, p-value=0.10). Adding the linear time effect to the model with only day of the week effects has almost the same impact (Δdev=6.1, Δdf=1, p-value=0.09) as adding the linear time effect to the constant rates model. On the other hand, adding a smooth time effect to the model which includes only the day of the week effects produces a very large reduction in deviance (Δdev=300.6, Δdf=18, p-value< 0.001). Since these temporal effects are incorporated mainly to account for systematic patterns over time in the count data, we use the smooth time effect to provide a better fit of the temporal patterns. Thus, the model accounting only for temporal effects is

$$\log \lambda_i = \text{Day of the Week Effects} + \text{Smooth Time Effect}. \tag{4.1.1}$$

4.2 Meteorological Effects

The meteorological variables included in this study are daily average temperature, daily maximum and minimum relative humidity (RH), and the daily average wind components[3]. Since the maximum RH is usually of more interest than minimum RH, initially only it is considered for inclusion in the models. Smooth functions (with span 0.1) of these meteorological variables are fit to best capture the systematic patterns in the data. Temperature, RH, Wind H and Wind V are smoothed individually, but the two wind components are always considered together. Various fitted models are presented in Table 4.2 (see p.15).

[3]Wind H and Wind V

Table 4.2: Meteorological models

Model Terms	Deviance	df	Δdev	Δdf	p-value
Base = Temporal Effects	1642.5	890.0			
Base + Temp	1624.9	880.4	17.6	9.6	0.44
Base + Wind	1578.6	849.5	63.9	40.5	0.66
Base + RH	1590.2	867.3	52.3	22.7	0.17
Base + RH + Temp	1576.0	857.7	14.2	9.6	0.62
Base + RH + Wind	1523.0	826.8	67.2	40.5	0.65

Adding each meteorological effect individually to the base model (the model accounting for temporal effects as discussed in the previous section), results in negligible reductions in deviance for temperature and wind (p-values are 0.44 and 0.66, respectively) and a somewhat larger but still insignificant reduction for RH (p-value=0.17). Adding either temperature or wind in addition to the smooth relative humidity effect has little impact on the fit (p-values are 0.62 and 0.65, respectively). Although the RH effect does not significantly improve the fit, it is the only meteorological factor which provides a reasonable reduction of the deviance (relative to the degrees of freedom associated with the factor). As there are still plenty of degrees of freedom remaining, we choose to include the RH effect in our preliminary model accounting for both temporal and meteorological effects.

Additional models involving more complicated combinations of the meteorological factors were also fitted. For example, minimum RH was also fitted together with maximum RH, but this reduced the deviance only by a small amount. Therefore, the new base model accounting for both meteorological and temporal effects is:

$$\log \lambda_i = \text{Day of the Week Effects} + \text{Smooth Time Effect} + \text{Smooth RH Effect}. \tag{4.2.2}$$

This base model represents the best-fitting relationship for these factors with the ER visit counts. In the remainder of this report we focus on pollutant effects on the ER visit counts, while taking the meteorological and temporal effects into account in the form of this base model.

Chapter 5

Modelling the Pollutants

Using the form of the temporal and meteorological effects selected in the previous chapter as a base model, the effects of the three pollutants on ER visit counts can now be studied. These pollutants will first be considered individually by adding appropriate effects to the base model as additional predictors. The reduction of deviance due to the addition of the effects for the individual pollutant will be assessed to determine whether there is a relationship between the pollutant and the ER visit counts. The regression coefficient of the pollutant will determine the nature of the relationship (whether it is positive or negative and how large). The pollutants will be examined for same day effects and also for lagged effects. In particular, lags up to 4 days for each pollutant will be considered for any relationship with ER visit counts.

The pollutant effects will not be smoothed, in contrast to how the temporal and meteorological effects were incorporated. Only very simple functions of the pollutants will be considered so that the fitted relationship between the pollutant and ER visit counts is easy to interpret. We will however use plots of the best-fitting smooth functions of the pollutants corresponding to different spans to guide our choice of the form of function to use for the model for each pollutant.

5.1 Particulate matter less than 10 microns (PM_{10})

As an initial exploratory step, we must check to see if fitting PM_{10} linearly is a reasonable way to proceed. This is done by analyzing how PM_{10} would be fitted if it were fit as a smooth function. Looking at the plots of the smoothed PM_{10} for different spans (Figure C.20) reveals that fitting PM_{10} linearly is a reasonable way to proceed because of the prevailing linear relationship with the ER visit counts. Initially, the effect of PM_{10} on the same day was explored by the addition of a linear effect

$$\log \lambda_i = \text{Base Model} + \text{PM}_i,$$

where PM_i is the PM_{10} measure on the ith day and i=1,...,915.

The addition of a linear effect for same day PM_{10} to the base model resulted in some reduction in deviance (Δdev=4.5, Δdf=1, p-value=0.12). Interestingly, this fitted relationship between PM_{10} and the the ER visit counts is negative (PM_{10} coefficient=-0.0014, SE=0.0009)[1] which indicates that an increase in PM_{10} is associated with a slight decrease in the ER visit counts. The autocorrelation structure for the residuals of this model show large[2] autocorrelations for lags 1 and 8 while other diagnostic plots do not reveal any problems with the model.

However, this fitted form assumes that the ER visit counts are related to PM_{10} only on the same day. It might be possible that increased levels of PM_{10} also affect the number of people visiting the emergency room on later days. Therefore the possibility of lagged pollutant effects should be explored. The addition of linear effects for PM_{10} at lags of 0 (same day) to 4 days was considered via models of the form:

$$\log \lambda_i = \text{Base Model} + \text{PM}_i + \text{PM}_{i-1} + \text{PM}_{i-2} + \text{PM}_{i-3} + \text{PM}_{i-4}.$$

When examining the effects of PM_{10} at a lag of 4 days, 4 days of data are lost from the analysis. This is because the ER visit count on day i is related to the PM_{10} level on day $i-4$, so the ER visit counts for days 1, 2, 3, and 4 do not enter into the analysis. To make the fits of different lag models comparable, these 4 days of data are withheld from all the fits, so the data of 911 days are used

[1] SE stands for standard error

[2] Autocorrelations exceeding ± 2 standard errors are considered large

Table 5.1: PM_{10} Models

Model Terms	Deviance	df	Δdev	Δdf	p-value
Base = Temporal and RH effects	1575.3	863.4			
Base + Lag 0	1570.8	862.4	4.5	1.0	0.12
Base + Lag 1	1571.7	862.4	3.6	1.0	0.16
Base + Lag 2	1574.8	862.4	0.5	1.0	0.59
Base + Lag 3	1570.6	862.4	4.7	1.0	0.11
Base + Lag 4	1570.6	862.4	4.7	1.0	0.11
Base + Lag 3 + Lag 0	1565.2	861.4	5.4	1.0	0.08
Base + Lag 3 + Lag 1	1563.6	861.4	7.0	1.0	0.05
Base + Lag 3 + Lag 2	1570.1	861.4	0.5	1.0	0.60
Base + Lag 3 + Lag 4	1569.4	861.4	1.2	1.0	0.42
Base + Lag 3 + Lag 1 + Lag 0	1562.7	860.4	0.9	1.0	0.50
Base + Lag 3 + Lag 1 + Lag 2	1562.7	860.4	0.9	1.0	0.49
Base + Lag 3 + Lag 1 + Lag 4	1562.9	860.4	0.7	1.0	0.55

instead of 915. The fitting of simultaneous effects for several lags was also explored. All lagged effects were modeled linearly based on plots similar to that of Figure C.20 for each individual lag effect. A summary of these model fits is presented in Table 5.1.

Considering first only the individual effects, adding the lag 3 effect to the base model produced the greatest reduction in deviance, although this reduction was only marginally greater than that associated with the lag 0 and lag 4 effects and in any case was not substantial (p-value=0.11). Adding the lag 1 effect to the model including the lag 3 effect produces the largest reduction in deviance which is marginally significant (p-value=0.05). Adding an individual lagged effect to the model including effects at lags 1 and 3 provides little additional reduction in the deviance. The overall reduction of deviance associated with including the lag 1 and lag 3 effects is statistically significant (Δdev=11.73, Δdf=2, p-value=0.04). Therefore, the effect of PM_{10} is fitted as follows:

$$\log \lambda_i = \text{Base Model} + \text{PM}_{i-1} + \text{PM}_{i-3}. \tag{5.1.1}$$

The estimated coefficients of PM_{10} at lag 1 and 3 in this model are -0.0020 (SE=0.0010) and 0.0022 (SE=0.0009) respectively. Taken at face value, this indicates that the effects of PM_{10} on the ER visit counts arise primarily through the difference of the PM_{10} levels 3 days and 1 day previously. The fitted model indicates that a decrease in PM_{10} from 3 days ago to 1 day ago is associated with an increase in the ER visit counts. Figure C.21 reveals some autocorrelation of the residuals for lags 1 and 8, a reasonably symmetric distribution of the residuals around 0, and no other systematic patterns in the residuals versus the fitted values from the model. These plots indicate no major problems with the model.

The boxplots of the daily average PM_{10} (Figure C.13) show three days with large values of over 100. Could expressing PM_{10} on a different scale intended to make these three values appear less extreme result in a different fitted model? Taking the logarithm of PM_{10} and fitting it as a similar model in (5.1.1) provides some reduction in deviance when compared to the base model (Δdev=10.57, Δdf=2, p-value=0.06). There is still a negative coefficient for lag 1 (PM coefficient=-0.053, SE = 0.020) and a positive coefficient for lag 3 (PM coefficient=0.043, SE = 0.024). However the reduction in deviance is not large enough to be considered significant and since the relationship with the ER visit counts is qualitatively the same as that obtained with the model (5.1.1), this transformation of the predictor PM_{10} does not result in drastic changes to the results. Therefore this model will not be considered further.

Another approach to examining the influence of these three extreme values is to replace these outliers with a moving average of the five previous readings (for each outlier). This results in these three values is being replaced by the adjusted values shown in Table 5.2.

Table 5.2: Changing of the Outliers

Original	Adjusted
125.64	26.06
150.04	25.63
150.63	26.30

For this adjusted data set, lag 3 provides the greatest reduction in deviance from the base model (Δdev=6.43, Δdf=1, p-value=0.06) and no other individual lag added to lag 3 yields a sizable reduction in the deviance (other p-values > 0.10). Thus after this modification of the three outliers, the best fitting model would be:

$$\log \lambda_i = \text{Base Model} + \text{PM}_{i-3}.$$

For this model, the estimated coefficient for PM_{10} at lag 3 is -0.0022 (SE = 0.0011) which has the opposite sign of the lag 3 PM_{10} effect in the model based on the original data. The residual plots for this model (not produced here) are quite similar to those for model (5.1.1). The only noteworthy difference is in the residual versus fitted values plot where the two data points to the far right in Figure C.21 are now closer to the rest of the data points. These two data points still have low residual values. So modifying these three outlying points in this way produces a different interpretation of the relationship between PM_{10} and ER visit counts but no other drastic change in the results. The reduction of deviance for the adjusted data (p-value=0.06) suggests that the addition of the fitted effects of PM_{10} in this model are not really significant, so perhaps they should not be taken too seriously. In fact neither model has overly convincing PM_{10} effects but the original model had a significant reduction in deviance so deserves attention. We will use (5.1.1) as our fitted model for the PM_{10} effects; the fitted effects will be discussed in more detail later.

5.2 Ozone (O_3)

As an initial exploratory step, the best-fitting smooth function of O_3 was examined. As seen in Figure C.22, the prevailing parabolic shape of the plots at different spans indicate that O_3 should perhaps be fitted quadratically. Choosing this approach, the O_3 effect is fitted by the following model:

$$\log \lambda_i = \text{Base Model} + \mathrm{O}_{3i} + \mathrm{O}_{3i}^2. \tag{5.2.2}$$

The addition of this quadratic effect for the same day O_3 to the base model yields a substantial reduction in the deviance (Δdev=14.5, Δdf=2, p-value=0.02). The estimated coefficients of the linear and quadratic components of O_3 are -0.00594 (SE=0.00251) and 0.00011 (SE=0.00003) respectively. This suggests that there is indeed a significant quadratic effect and a negative linear effect of O_3 on ER visits. Checking the diagnostic plots of the residuals does not reveal any major problems with the model.

Lagged O_3 effects were also examined. Since the same day O_3 effect was not modeled linearly, we must also determine, for each lag effect, the appropriate order to use for the fit. This is done with plots similar to Figure C.21. This resulted in lags 1 and 2 being modelled with 4th degree polynomial effects (a quadratic effect is a 2nd degree polynomial effect), while lags 0, 3 and 4 are modelled quadratically. The model for the lag j of O_3 effect is

$$\log \lambda_i = \text{Base Model} + \mathrm{O}_{3,i-j} + \mathrm{O}_{3,i-j}^2 + \cdots + \mathrm{O}_{3,i-j}^k, \tag{5.2.3}$$

Table 5.3: O_3 Models

Model Terms	Deviance	df	Δdev	Δdf	p-value
Base = Temporal and RH effects	1575.3	863.4			
Base + Lag 0	1560.8	861.4	14.5	2.0	0.02
Base + Lag 1	1573.6	859.4	1.7	4.0	0.92
Base + Lag 2	1572.3	859.4	3.0	4.0	0.80
Base + Lag 3	1573.9	861.4	1.4	2.0	0.69
Base + Lag 4	1573.5	861.4	1.8	2.0	0.62
Base + Lag 0 + Lag 1	1560.7	857.4	0.1	4.0	0.99
Base + Lag 0 + Lag 2	1558.5	857.4	2.3	4.0	0.86
Base + Lag 0 + Lag 3	1558.5	859.4	2.3	2.0	0.53
Base + Lag 0 + Lag 4	1559.1	859.4	1.7	2.0	0.62

where k is the order of the fit.

Once again 911 days are used to accommodate the fitting of lagged effects up to lag 4. A stepwise regression technique is used to determine any large O_3 effects by adding each lag effect individually to the base model. The results are provided in Table 5.3.

Adding each specific lag individually to the base model produces no large reduction of deviance except for the same day effect (Δdev=14.5, Δdf=2, p-value=0.02). Similarly, adding any individual lagged effect to the lag 0 model provides little reduction in the deviance. As a result only the same day effect of O_3 on the ER visit counts is of interest. Figure C.23 indicates no major problems with the same day effect model as we seen through the low autocorrelation and the fairly symmetric distribution of the residuals around zero. So the O_3 effects are modelled as in (5.2.2). The fitted effects will be discussed in more detail later.

5.3 Sulfur Dioxide (SO_2)

Initially, the best fitting smooth function of SO_2 was examined. The plots of the smooth functions of SO_2 at different spans (Figure C.24) indicate a linear trend, so a linear fits of SO_2 were examined. The effect of SO_2 on the same day was explored by the addition of a linear effect

$$\log \lambda_i = \text{Base Model} + \mathrm{SO}_{2i},$$

where SO_{2i} is the SO_2 measure on the ith day and i=1,...,915.

Table 5.4: SO_2 Models

Model Terms	Deviance	df	Δdev	Δdf	p-value
Base = Temporal and RH effects	1575.3	863.4			
Base + Lag 0	1571.5	862.4	3.8	1.0	0.15
Base + Lag 1	1569.9	862.4	5.4	1.0	0.09
Base + Lag 2	1572.4	862.4	2.9	1.0	0.21
Base + Lag 3	1574.8	862.4	0.5	1.0	0.59
Base + Lag 4	1566.1	862.4	9.2	1.0	0.02
Base + Lag 4 + Lag 0	1563.2	861.4	2.9	1.0	0.20
Base + Lag 4 + Lag 1	1559.8	861.4	6.3	1.0	0.06
Base + Lag 4 + Lag 2	1561.5	861.4	4.6	1.0	0.11
Base + Lag 4 + Lag 3	1566.1	861.4	0.0	1.0	0.92

The addition of a same day linear SO_2 effect does not reduce the deviance substantially (Δdev=3.8, Δdf=1, p-value=0.15). The estimated regression coefficient -0.0008 (SE=0.0006) suggests that SO_2 has a negative relationship with ER visit counts. The autocorrelation structure for the residuals of this model show autocorrelations for lags 1 and 8 while other diagnostic plots reveal no problems with the model.

Other lagged day effects (lags 0, 1, 2, 3, 4) were fit to see if there is a large reduction in deviance associated with different lags of SO_2. The lag j effect of SO_2 is modelled linearly as

$$\log \lambda_i = \text{Base Model} + SO_{2,i-\text{j}}.$$

Out of the 915 available days, 911 are used and the fits are summarized in Table 5.4.

Considering the individual lag effects, lag 4 results in substantial reduction in deviance (Δdev=9.2, Δdf=1, p-value=0.02). The estimated coefficient of SO_2 at lag 4 is 0.0015 (SE=0.0006) which is quite large and indicates a positive relationship with the emergency room visit counts. Figure C.25 shows some autocorrelation in the residuals from this model at lags 1 and 8 while the residual versus fitted values plot indicates that the residuals are symmetric around zero. There appear to be no major problems with the model.

Adding the lag 1 effect to the model with the lag 4 effect produces some reduction in deviance but this is not substantial (Δdev=6.3, Δdf=1, p-value=0.06). Adding any other individual lag to the model containing lags 1 and 4 produces negligible additional reduction in the deviance. (Longer lags were also considered but no substantial effects were detected.) The estimated coefficients for this model are -0.0012 (SE=0.0006) and 0.0015 (SE=0.0006) for lags 1 and 4 respectively. This gives a similar lag 4 effect as in the model containing only the lag 4 effect. Since the reduction of deviance associated with the additional inclusion of the lag 1 effect is not statistically significant (p-value=0.06), the following linear model is considered as the fitted model for SO_2:

$$\log \lambda_i = \text{Base Model} + \text{SO}_{2,i-4}. \tag{5.3.4}$$

The fitted SO_2 effect will be discussed in more detail later.

5.4 Multiple Pollutants

All of the pollutant effects that have been identified as being of some importance are now examined simultaneously to explore their joint effects on the ER visit counts. Using the best-fitting forms for the individual pollutants, the initial full model for all pollutants is

$$\log \lambda_i = \text{Base Model} + \text{PM}_{i-1} + \text{PM}_{i-3} + \text{O}_{3i} + \text{O}_{3i}^2 + \text{SO}_{2,i-4}. \tag{5.4.5}$$

Backward elimination is used to determine which pollutants have the greater effects on the ER visit counts. The pollutant removed from the model at any stage is the one which results in the least significant change in deviance, namely the elimination corresponding to the greatest p-value. The results are presented in Table 5.5 (see p.24). Considering each pollutant individually for removal from the model with all the pollutants, Table 5.5 indicates that PM_{10} is least important (p-value=0.08). Ozone and sulfur dioxide now remain in the model and each of these pollutants is considered for elimination. Removing either results in similar p-values, both of which are significant (p-value=0.02 for removal of O_3, p-value=0.03 for removal of SO_2). Hence both of these pollutants seem to have an effect on the ER visit counts. If one pollutant is to be removed from the model, based on the p-values it would be SO_2. Judged in this fashion, O_3 is the pollutant with the most effect on the ER visit counts while PM_{10} has the least effect.

It should be noted from the results of Table 5.5 that each of the pollutants, but particularly O_3 and SO_2, appear to have their own effects on the ER visit counts. The estimated regression coefficients for each model are presented in Table 5.6. The estimated regression coefficients do not vary greatly across models, although the linear O_3 effect seems to decrease somewhat in magnitude

Table 5.5: Multiple Pollutant Models

Model Terms	Deviance	df	Δdev	Δdf	p-value
All Pollutants	1542.7	858.4			
Ozone + Sulfur Dioxide	1551.8	860.4	9.1	2.0	0.08
PM_{10}+ Sulfur Dioxide	1556.6	860.4	13.9	2.0	0.02
PM_{10}+ Ozone	1550.3	859.4	7.6	1.0	0.04
Sulfur Dioxide	1566.1	862.4	14.3	2.0	0.02
Ozone	1560.8	861.4	9.0	1.0	0.03

when models with fewer pollutants are fitted and the lag 3 PM_{10} effect increases slightly when another pollutant is dropped from the model.

Table 5.6: Estimated Regression Coefficients$\times 10^{-3}$(SE $\times$ 10^{-3})

Model Terms	PM_{i-1}	PM_{i-3}	O_{3i}	O_{3i}^2	$SO_{2,i-4}$
All Pollutants	−2.2 (1.0)	1.5 (0.9)	−7.2 (2.6)	0.11 (0.03)	1.5 (0.6)
Ozone + Sulfur Dioxide	—	—	−6.3 (3.3)	0.11 (0.03)	1.5 (0.6)
PM_{10}+ Sulfur Dioxide	−2.1 (1.0)	1.7 (0.9)	—	—	1.4 (0.6)
PM_{10}+ Ozone	−2.1 (1.0)	2.1 (0.9)	−6.7 (3.4)	0.11 (0.03)	—
Sulfur Dioxide	—	—	—	—	1.5 (0.6)
Ozone	—	—	−5.9 (2.5)	0.11 (0.03)	—

Chapter 6

Summary and Concluding Remarks

In this report Poisson regression models have been used to examine possible relationships between the ER visits and ambient levels of air pollutants in Prince George, B.C. The results of the data analysis considering each pollutant modelled individually indicate that there are non-negligible relationships. However each pollutant is modelled differently, with the strongest relationships occurring at different lags and with O_3 reflecting a non-linear relationship.

The fit of PM_{10} results in the use of a lag 1 and lag 3 day model with lag 1 having a negative effect (-0.0020, SE=0.0010) and lag 3 having a positive effect (0.0022, SE=0.0009). These fitted relationships can be interpreted as follows: Suppose the PM_{10} measurement was 20 $\mu g/m^3$ for both one day and three days previously. If the PM_{10} measure decreased by 10 $\mu g/m^3$ (or 50%) for both days while meteorological effects and any other effects remained constant, the ER visit counts would be predicted to decrease by about 0.2% (95% confidence interval from -1.9% to 2.3%). However, this is assuming that the PM_{10} measure remains constant both one and three days ago. Another scenario could be a decrease in PM_{10} for only one of the lags. For example, if the PM_{10} measure one day ago decreased from 20 $\mu g/m^3$ to 10 $\mu g/m^3$ and the PM_{10} measure three days ago remained at 20 $\mu g/m^3$ (all other factors remaining constant), the ER visit counts would be predicted to increase by about 2.0% (95% confidence interval from 0.0% to 4.1%). The form of the fitted relationship makes it quite difficult to describe the apparent effects of PM_{10} on the ER visit counts.

The fitted model for O_3 contains both linear (-0.0059, SE=0.0025) and quadratic same day effects (0.00011, SE=0.00003). Same day O_3 has a significant relationship with the ER visit counts which can be interpreted as follows: Suppose that today's O_3 measure was 30 $\mu g/m^3$. A decrease from 30 μg/m^3 to 15 μg/m^3 (or 50%) (all other factors remaining constant) would be predicted to result in an increase in the ER visit counts by about 1.4% (95% confidence interval from -2.8% to 5.9%). However, if today's O_3 measure was 45 μg/m^3 and a decrease to 30 μg/m^3 occurs, the predicted result is a decrease in the ER visit counts by about 3.5% (95% confidence interval from -0.4% to 7.2%). Because of the quadratic relationship, a fixed decrease in O_3 from two different values results in different predicted impacts on the ER visit counts.

The fitted model for sulfur dioxide has a positive linear effect at lag 4 (0.0015, SE=0.0006). The lag 4 effect is interpreted as follows: Suppose that the measure of sulfur dioxide 4 days ago was 6 μg/m^3. A decrease in the sulfur dioxide measure 4 days ago of 3 μg/m^3 (or 50%) (all other factors remaining constant) results in a predicted decrease in the ER visit counts by about 0.4% (95% confidence interval from 0.1% to 0.8%).

The purpose of this study was to examine the relationships between emergency room visits and ambient levels of air pollutants. Based on the available data (from July 1995 to December 1997), the fitted models indicate that individually PM_{10}, SO_2, and O_3 have statistically significant effects on emergency room visits. Considering the pollutants simultaneously indicates that ozone has the most significant effect and PM_{10} has least significant effect on emergency room visits. These findings support the recommendation of Knight et al (1989) that data be collected on sulfur dioxide and ozone among other pollutants as additional air quality parameters. Of course, all the findings reported here have been derived strictly on an empirical basis; the biological plausibility of these fitted models must also be considered.

Finally we note that this study is based only on the pollution data recorded at Plaza station in Prince George; a more comprehensive study would include readings from multiple stations throughout Prince George. Also, the study time frame of two and a half years is quite short given the modest size of the daily ER visit counts. A more comprehensive study might usefully be based on a longer time frame. Such a study would allow more detailed examination of the functional form of the relationships between the ER visit counts and the different pollutants and would also allow examination of the consistency of these empirical relationships across different subcategories of the ER visit counts.

References

Bates, D. V. and Sizto, R. (1987). Air pollution and hospital admissions in Southern Ontario: The acid summer haze effect. *Environmental Research*, **43**, 317-331.

Kelsall, J. E., Samet, J. M., Zeger, S. L., and Xu, J. (1997). Air pollution and mortality in Philadelphia, *American Journal of Epidemiology*, **146**, 750-762.

Kelsall, J. E., Zeger, S. L., and Samet, J. M. (1999). Frequency domain log-linear models; air pollution and mortality, *Applied Statistics*, **48**, 331-334.

Knight, K. Leroux, B., Millar, J. and Petkau, A. J. (1988). Air pollution and human health: A study based on hospital admission data from Prince George, British Columbia. Technical Report # 128, Department of Statistics, The University of British Columbia, Vancouver, Canada.

Knight, K. Leroux, B., Millar, J. and Petkau, A. J. (1989). Air pollution and human health: A study based on emergency room visits data from Prince George, British Columbia. Technical Report # 136, Department of Statistics, The University of British Columbia, Vancouver, Canada.

Lawless, J. F. (1987) Negative binomial regression models. *Canadian Journal of Statistics*, **15**, 209-225.

McNeney, B. and Petkau, A. J. (1991). Air pollution and human health: A follow-up study based on emergency room visits data from Prince George, British Columbia. Report prepared under contract with the Ministry of Health, Province of British Columbia.

Samet, J. M., Bishop, Y., Speizer, F. E., Spengler, J. D., and Ferris, B. F. (1981). The relationship between air pollution and emergency room visits in an industrial community. *Journal of the Air Pollution Control Association*, **31**, 236-240.

Zeger, S. L. (1988). A regression models for time series count. *Biometrika*, **75**, 621-629.

Appendix A

Method for Filling-In Missing Values

The filling-in algorithm uses time series models to forecast or backcast for missing values. For each variable with a missing value, a linear time series model was fitted taking into account hour of the day effects (if the missing values are hourly) and any significant cyclical effects usually attributed to the seasonality of the data. The t^{th} data value is expressed as:

$$y_t = F_t + R_t, \tag{A.0.1}$$

where $t = 1, ..., n$, and

$$F_t = \text{Hour of the Day Effects} + \sum_{\mathrm{j}=1}^{k} [\beta_{\mathrm{j}} \cos(2\pi \mathrm{jt}/8766) + \gamma_{\mathrm{j}} \sin(2\pi \mathrm{jt}/8766)], \tag{A.0.2}$$

where k = number of cycles to be included in the fit and R_t = residual value for the t^{th} data value.

This linear time series model was fitted by carrying out a linear regression of the response variable (y_t) on the hour of the day effects (if the data is hourly) and the cyclical effects. The resulting model yields fitted values ($\hat{y_t}$) and residual values ($R_t = y_t - \hat{y_t}$) corresponding to each observed value of y_t. The residuals can now be modeled using an autoregressive model:

$$R_t = \phi_1 R_{t-1} + \phi_2 R_{t-2} + ... + \phi_i R_{t-i} + \epsilon_t, \tag{A.0.3}$$

where i = autoregressive order of the time series model, ϕ_j = autocorrelation coefficient of j^{th} order ($j = 1, \ldots, i$), and ϵ_t is assumed to have a mean of zero and a constant variance.

Using an autoregressive model requires a specified order of the fit for the residuals. In this study, several models of different orders were examined. The model with the best autocorrelation structure in its residuals was chosen by examining the plots of the autocorrelations of the residuals. The model with the lowest autocorrelations was deemed the best model. The residuals corresponding to the missing entries are then predicted using the residuals of non-missing entries and the estimated autocorrelation coefficients obtained from the fit of (A.0.3). Once the residuals corresponding to the missing values are predicted, the actual missing values can be predicted using the fit of (A.0.1).

Appendix B

Poisson Regression Models

Poisson regression models are useful for modeling count data. Emergency room visits can be considered as rare events and countable, so a plausible initial model is that the daily numbers of emergency room (ER) visits follow a Poisson distribution with possibly different means. The mean number of visits on a given day will be described by a function of the available covariates (a set of predictors), which includes temperature, wind, relative humidity, and the pollutants: PM_{10}, sulfur dioxide (SO_2) and ozone (O_3).

The use of Poisson regression depends on the following two basic assumptions:

- the expected number of visits on a given day depends on time and pollution and not on the past number of visits.
- daily visits are conditionally independent given temporal, meteorological and pollution covariates.

The first assumption is reasonable in this case because the diagnoses under consideration in this study are not infectious. The following model can accommodate the case where the mean depends explicitly on the past visits as well. Suppose $\mathbf{t}_i$ is a vector of temporal and meteorological covariates and $\mathbf{u}_i$ is a vector of pollution covariates, both for day i; $\mathbf{u}_i$ can include values of pollution variables prior to day i as well as those values on day i. If y_i denotes the number of ER visits on day i, a log-linear model assumes that y_i has a Poisson distribution with mean $\lambda_i(\mathbf{t}_i, \mathbf{u}_i)$, where

$$\log \lambda_i(\mathbf{t}_i, \mathbf{u}_i) = \alpha + \mathbf{t}_i'\boldsymbol{\beta} + \mathbf{u}_i'\boldsymbol{\gamma},$$

with α, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$ as unknown parameters (regression coefficients to be fitted).

Given a set of observations $(y_1, \mathbf{t}_1, \mathbf{u}_1)$, $(y_2, \mathbf{t}_2, \mathbf{u}_2)$, ...,$(y_n, \mathbf{t}_n, \mathbf{u}_n)$, the maximum likelihood estimates of the parameters α, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$ can be determined by maximizing the log-likelihood function:

$$l(\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}) = \sum_{i=1}^{n} [\ y_i \log(\alpha + \mathbf{t}_i'\boldsymbol{\beta} + \mathbf{u}_i'\boldsymbol{\gamma}) - (\alpha + \mathbf{t}_i'\boldsymbol{\beta} + \mathbf{u}_i'\boldsymbol{\gamma}) - \log(y_i!)\ ]$$

over all possible values of the parameters α, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$; the estimated values are denoted by $\hat{\alpha}$, $\hat{\boldsymbol{\beta}}$, and $\hat{\boldsymbol{\gamma}}$ respectively. We have carried out the Poisson regression analyses (and all other analyses) using S-Plus software.

One practical aspect of fitting Poisson regression models concerns over-dispersion. If the variance of the daily visit counts given the temporal, meteorological and pollution variables is greater than the mean, then the visits are said to be "over-dispersed". One approach to this problem is to introduce a dispersion parameter $\phi > 1$ and assume that the variance of the visits on day i is $\phi \times \lambda(\mathbf{t}_i, \mathbf{u}_i)$. A discussion of over-dispersion is given by Knight et al (1989) and Kelsall et al (1999) among others. This represents the simplest implementation of a general approach to fitting models in complicated settings referred to as quasi-likelihood. We briefly describe the procedure for statistical inference in this more general model.

The regression parameters of the model are again estimated by maximizing the Poisson log-likelihood function (without a dispersion parameter). However, it is necessary to take into account the dispersion parameter for assessing the statistical significance of the estimated parameters and in assessing the plausibility of model reductions. To discuss this in the context of the model reduction, assume that we begin with one model (the full model) and wish to test the suitability of another model (the reduced model), which arises from the full model by setting certain unknown parameters (e.g. regression coefficients) to zero; statisticians refer to such a reduced model as being "nested" within the full model. The difference in the deviances of the models is then approximately distributed as $\phi \times \chi^2$ with degrees of freedom (df) given by the difference in the df associated with the two deviances. Therefore, the difference in deviances[1] must be divided by an estimate of the dispersion parameter before it can be compared to a chi-square distribution to assess whether the reduction from the full model to the reduced model is permissible. Here we will use the estimate of the overdispersion parameter based on the deviance of the full model, which leads to the following

[1]The larger the discrepancy between the observed values and the values predicted for a particular model, the larger the deviance associated with that model.

statistic for testing the reduction to the reduced model:

$$\Delta' \mathrm{dev} = \frac{\text{deviance (reduced)} - \text{deviance (full)}}{\text{deviance (full)}/\mathrm{df(full)}}$$

This statistic is compared to a chi-square with df given by $\Delta \mathrm{df} = \mathrm{df(reduced)} - \mathrm{df(full)}$. For more details about Poisson regression models, readers are referred to Lawless (1987), Zeger (1988), Knight et al (1989), and Kelsall et al (1999).

Appendix C

Descriptive Plots

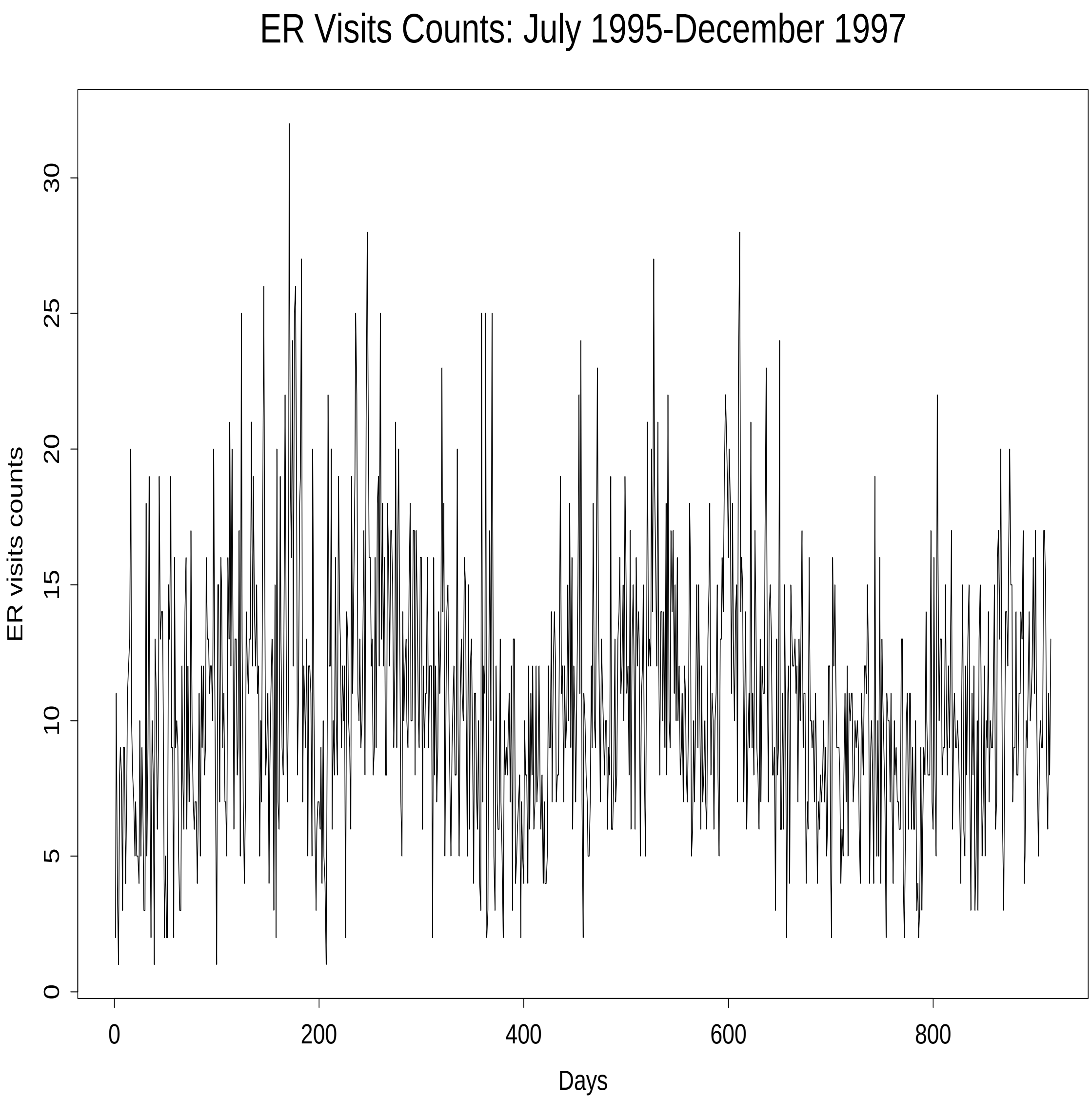


Figure C.1: Time series plot of ER counts of visits (Day 0 = June 30 1995)

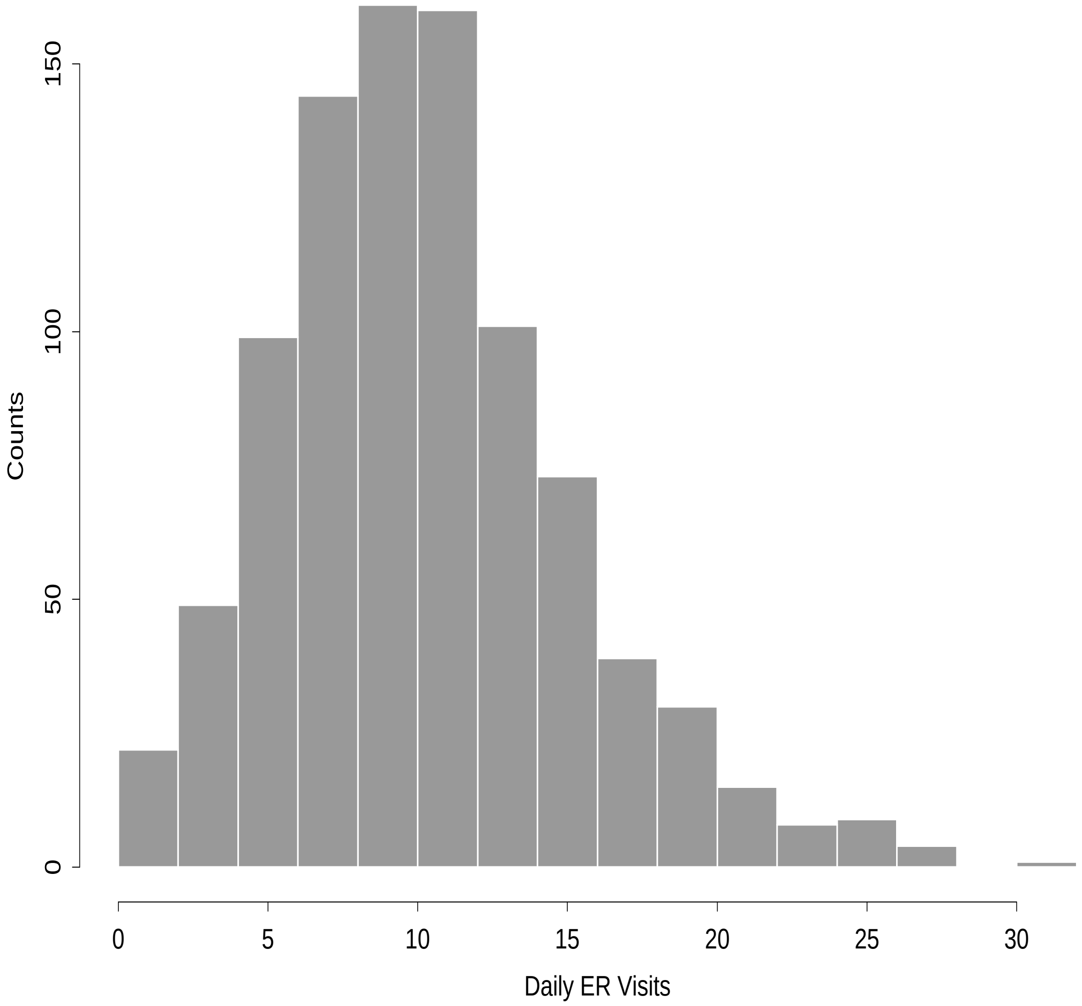


Figure C.2: Histogram of ER counts of visits

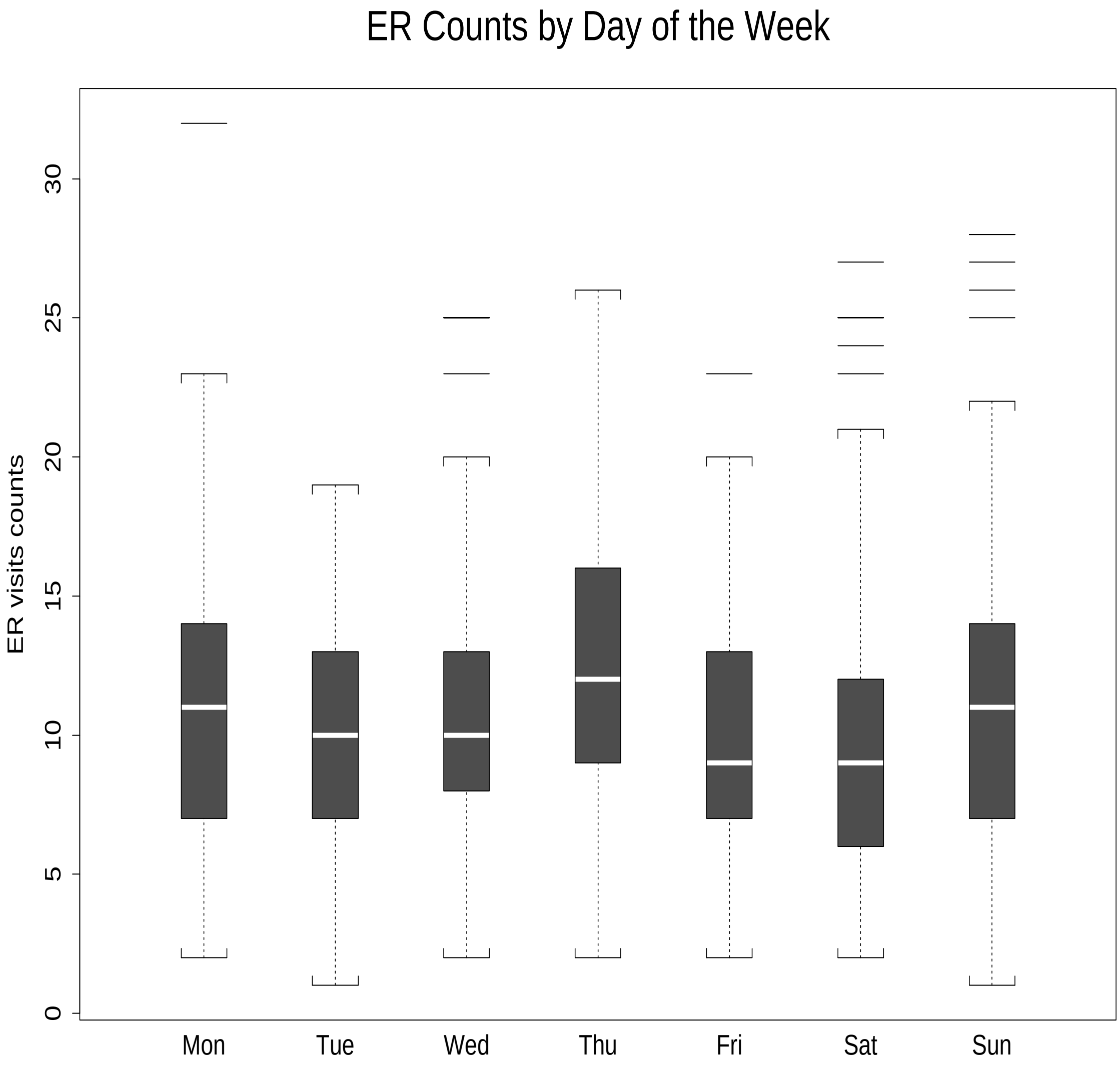


Figure C.3: Boxplots of ER counts by day of the week

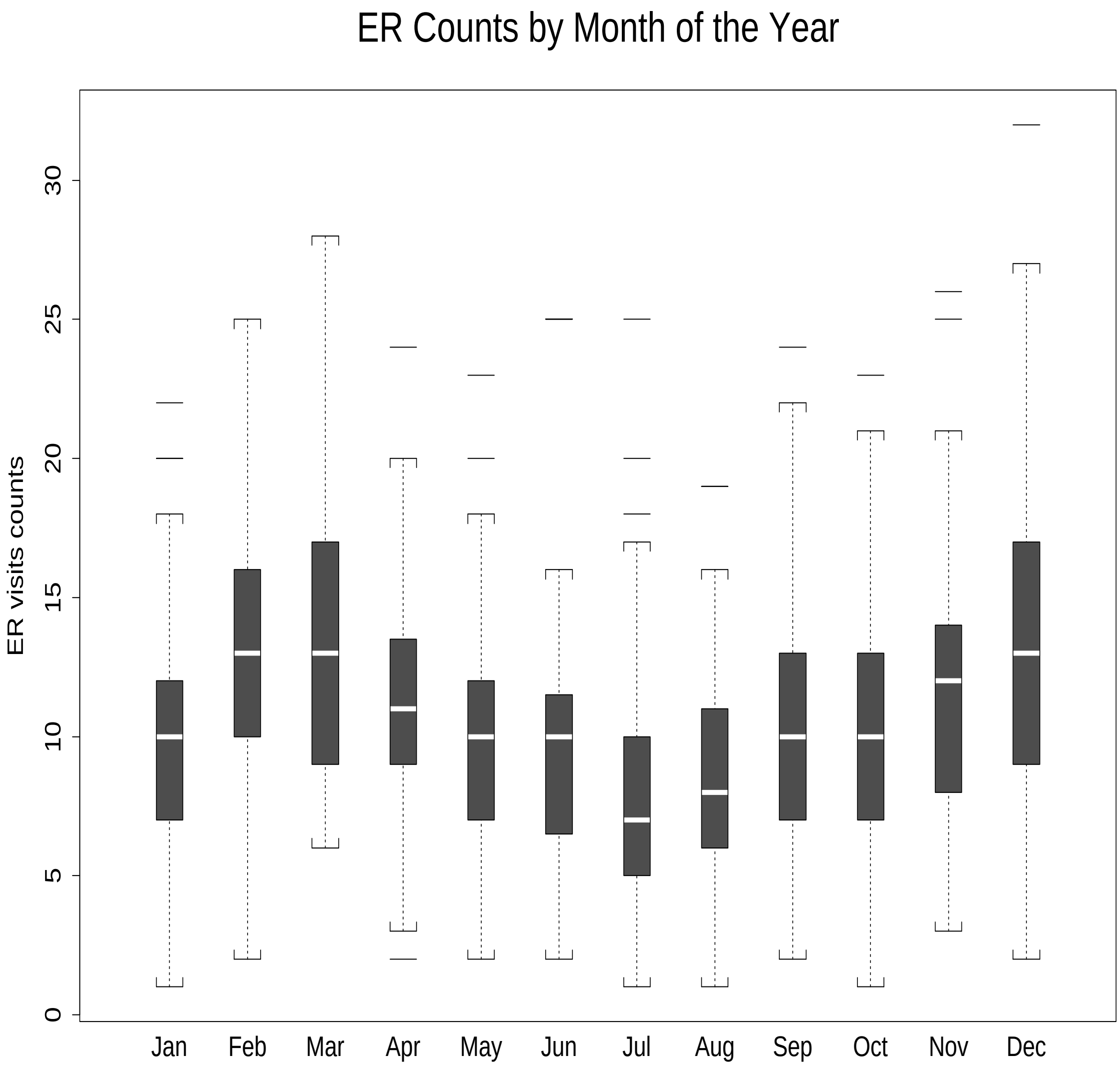


Figure C.4: Boxplots of ER counts by month

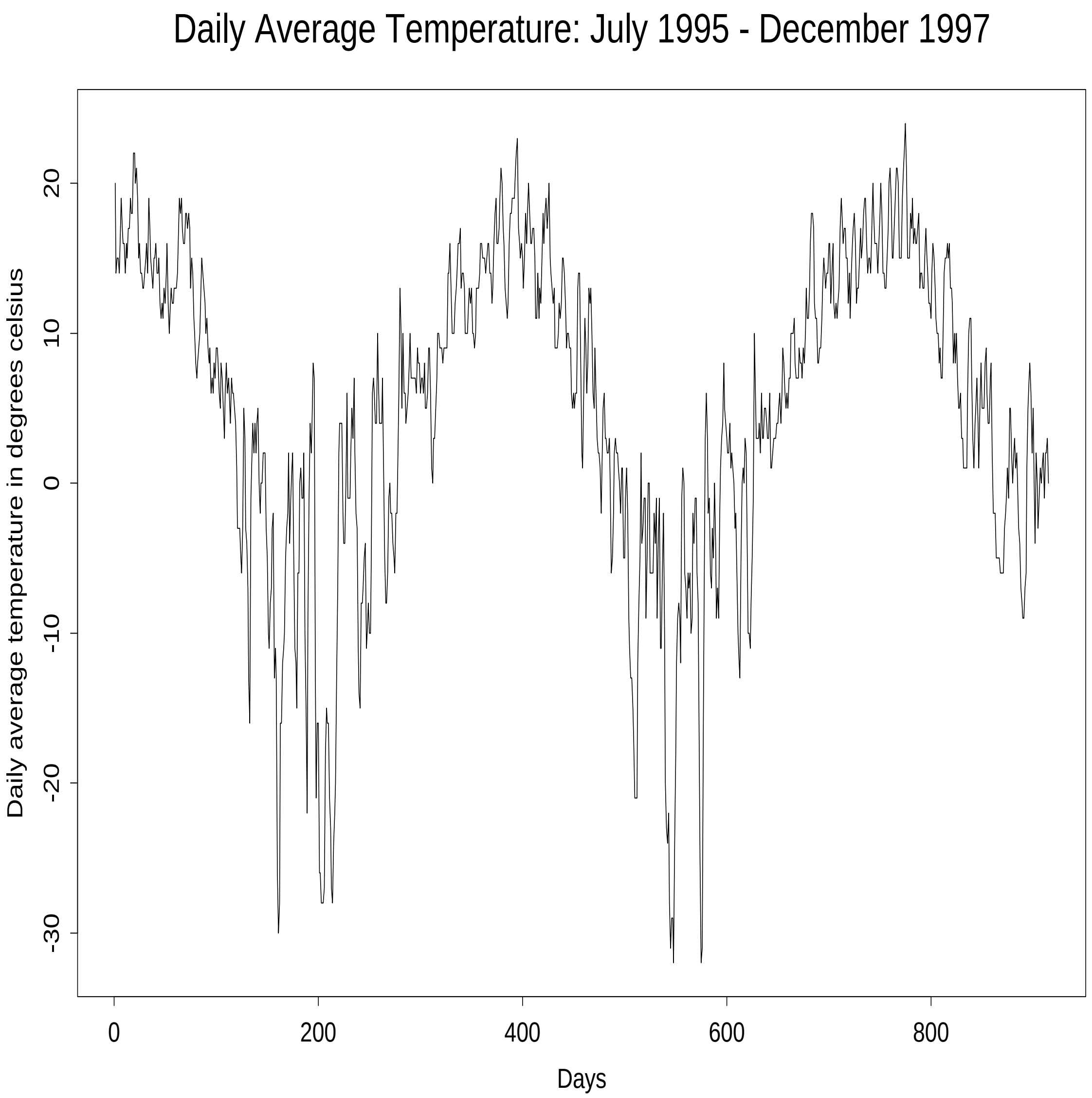


Figure C.5: Time series plot of daily average temperature($^{\circ}C$) (Day 0 = June 30 1995)

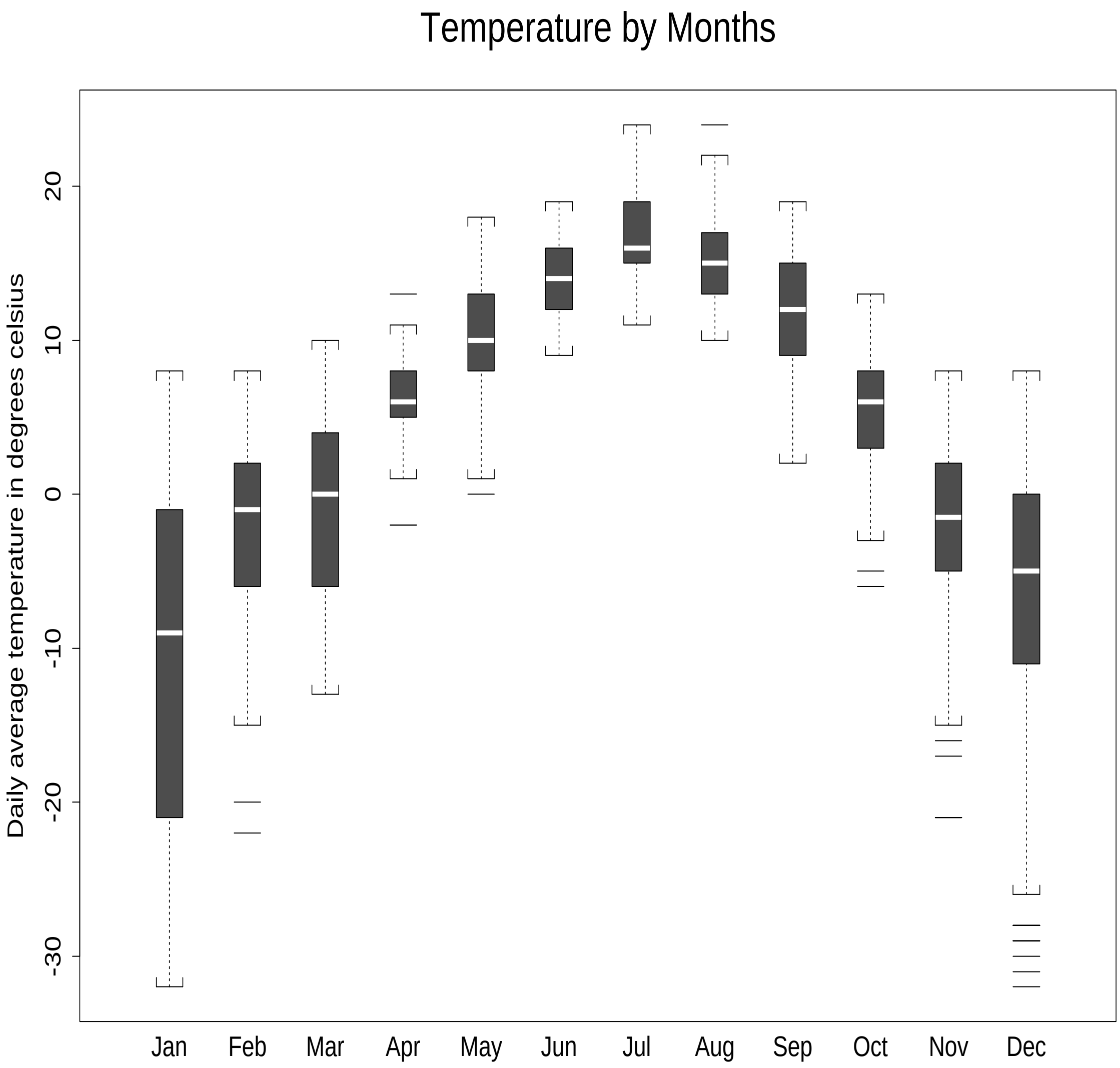


Figure C.6: Boxplots of daily average temperature($^{\circ}C$) by month

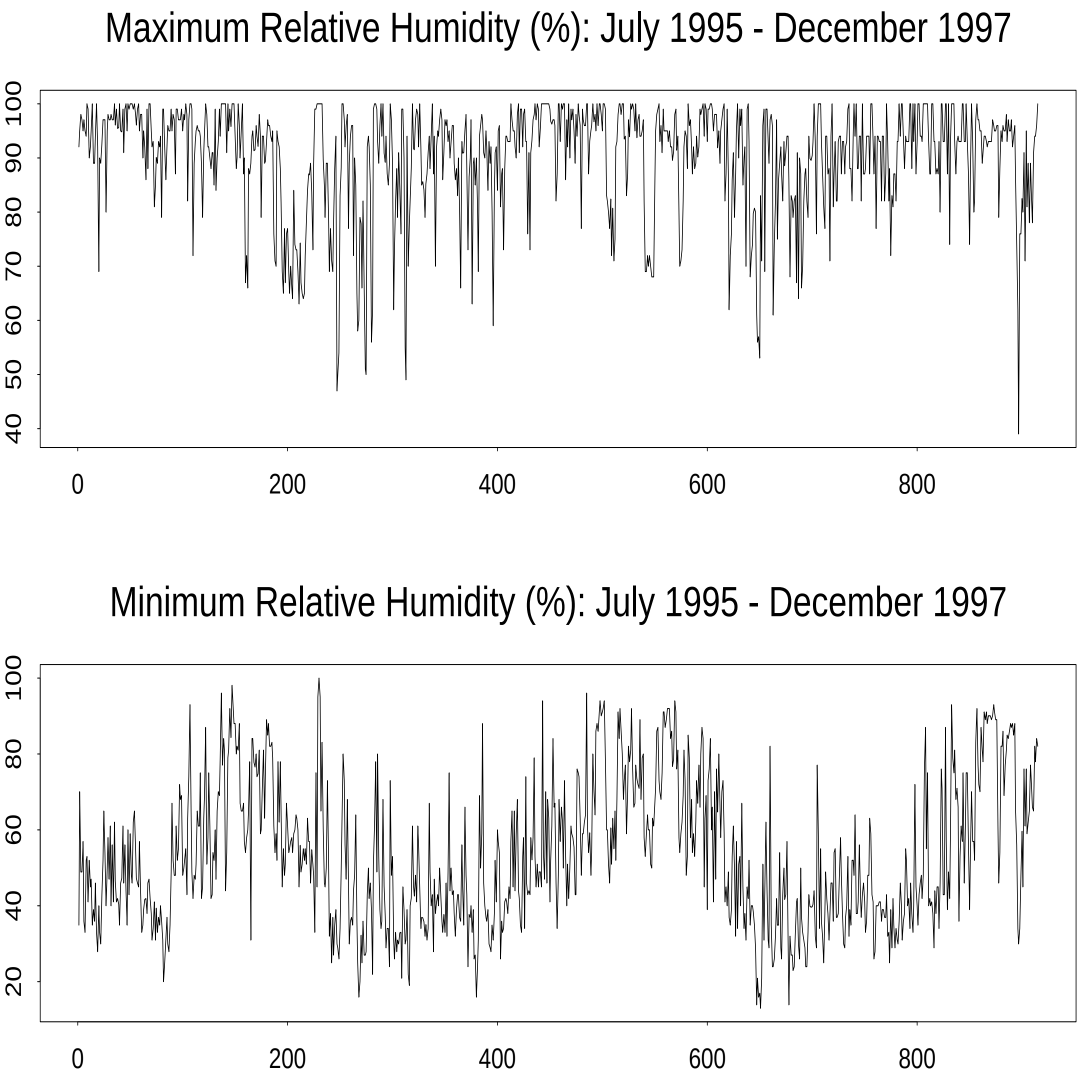


Figure C.7: Time series plot of maximum and minimum relative humidity (%) (Day 0 = June 30 1995)

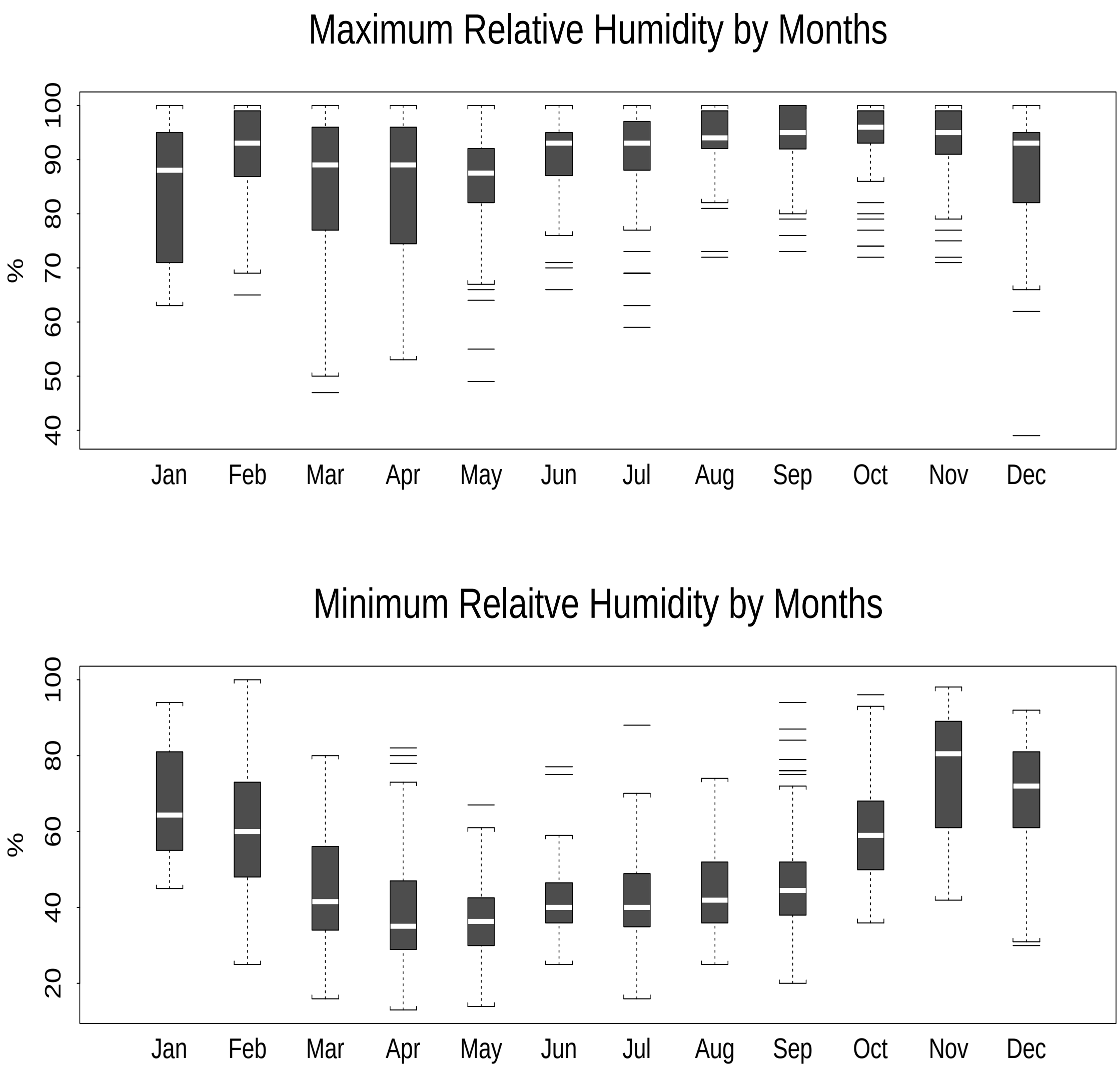
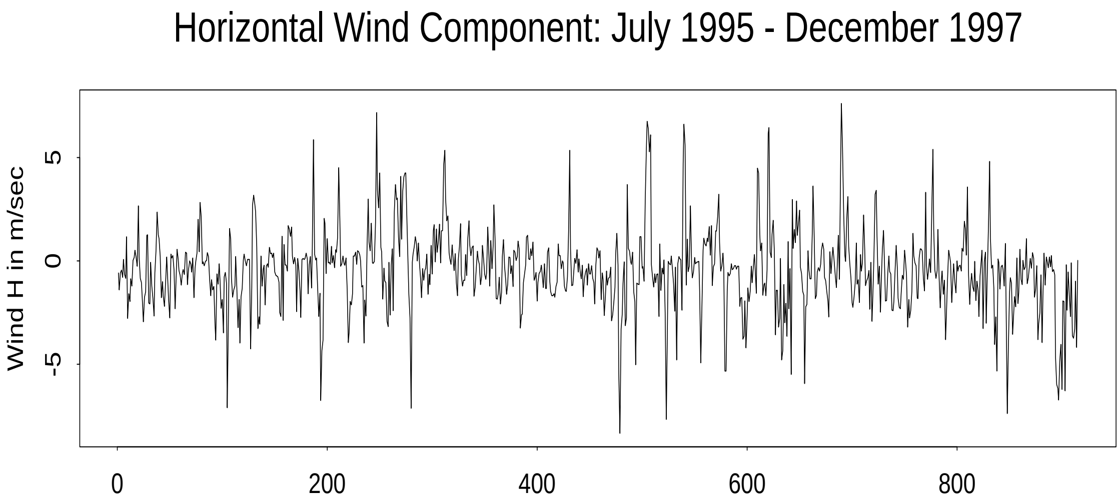


Figure C.8: Boxplots of maximum and minimum relative humidity (%) by month



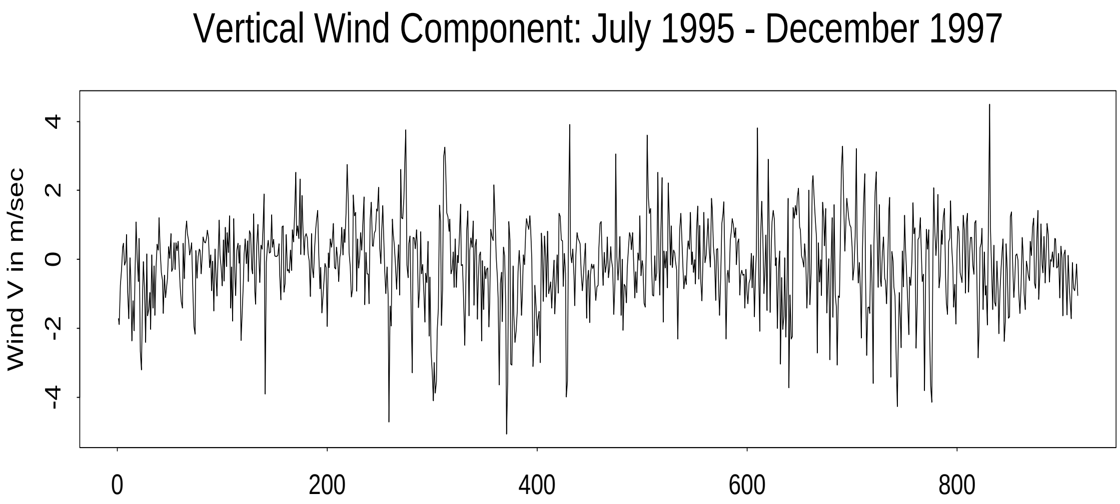


Figure C.9: Time series plot of horizontal and vertical wind components(m/sec) (Day 0 = June 30 1995)

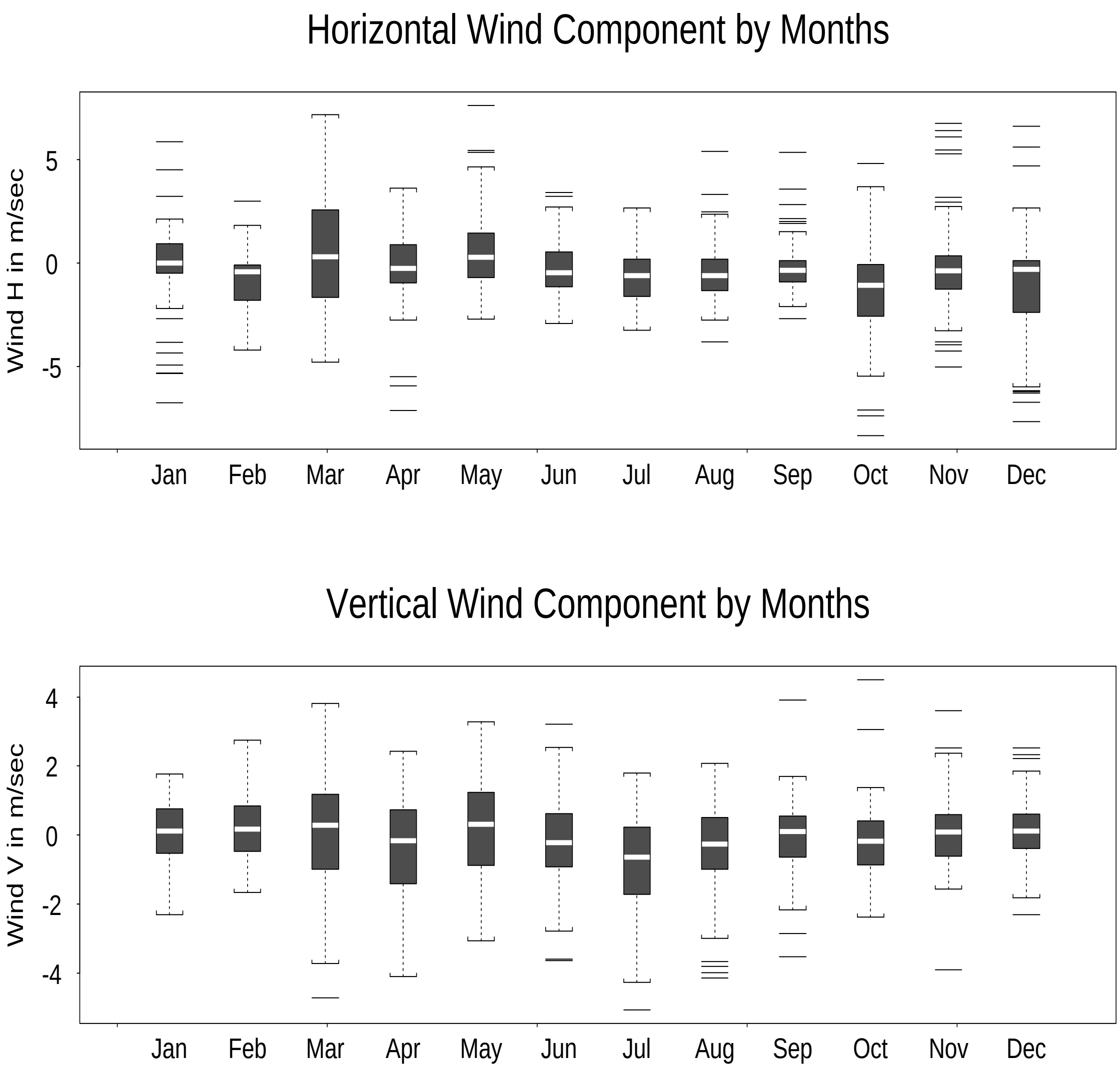


Figure C.10: Boxplots of horizontal and vertical wind(m/sec) by month

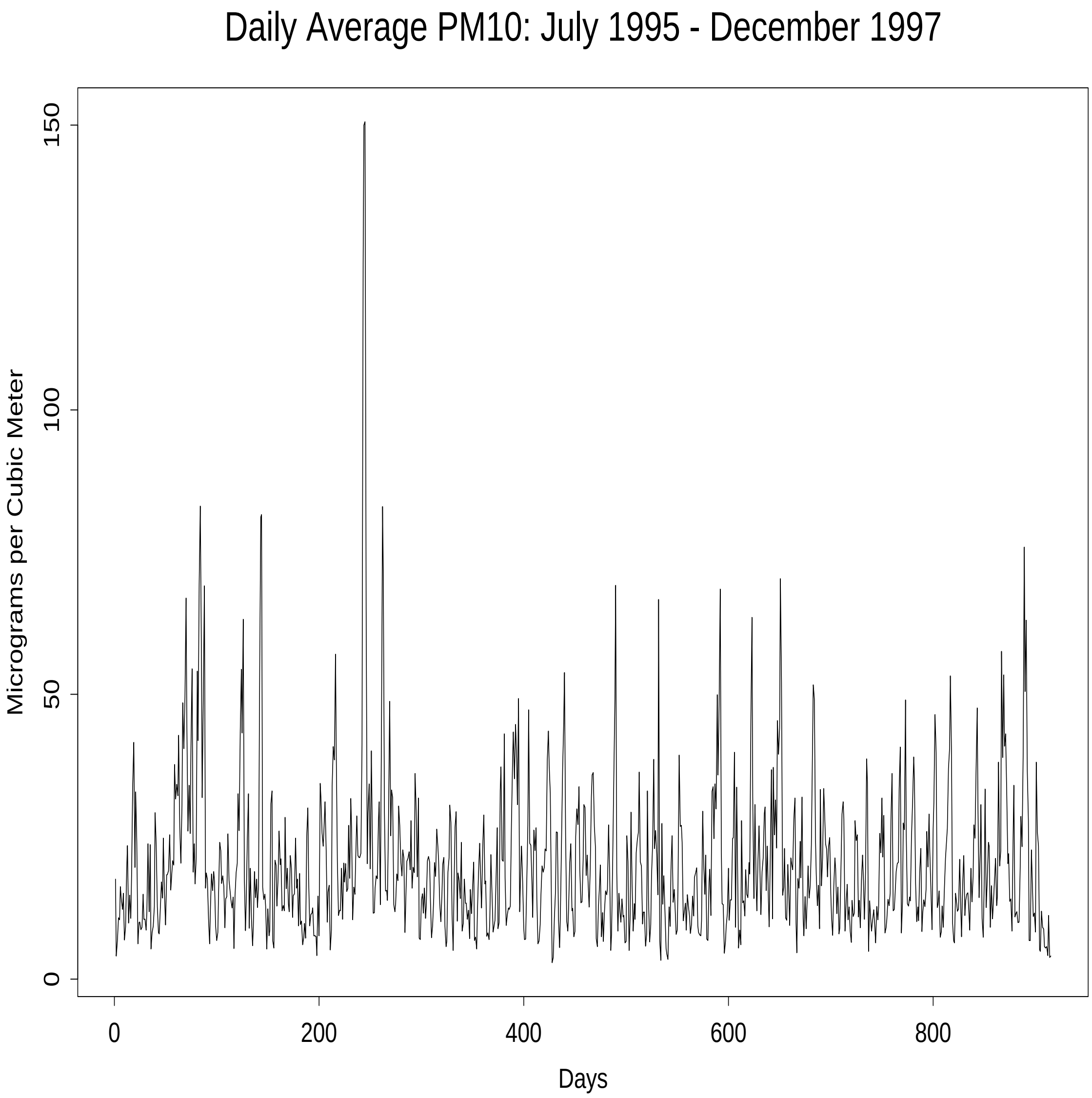


Figure C.11: Time series plot of PM_{10} ($\mu g/m^3$) (Day 0 = June 30 1995)

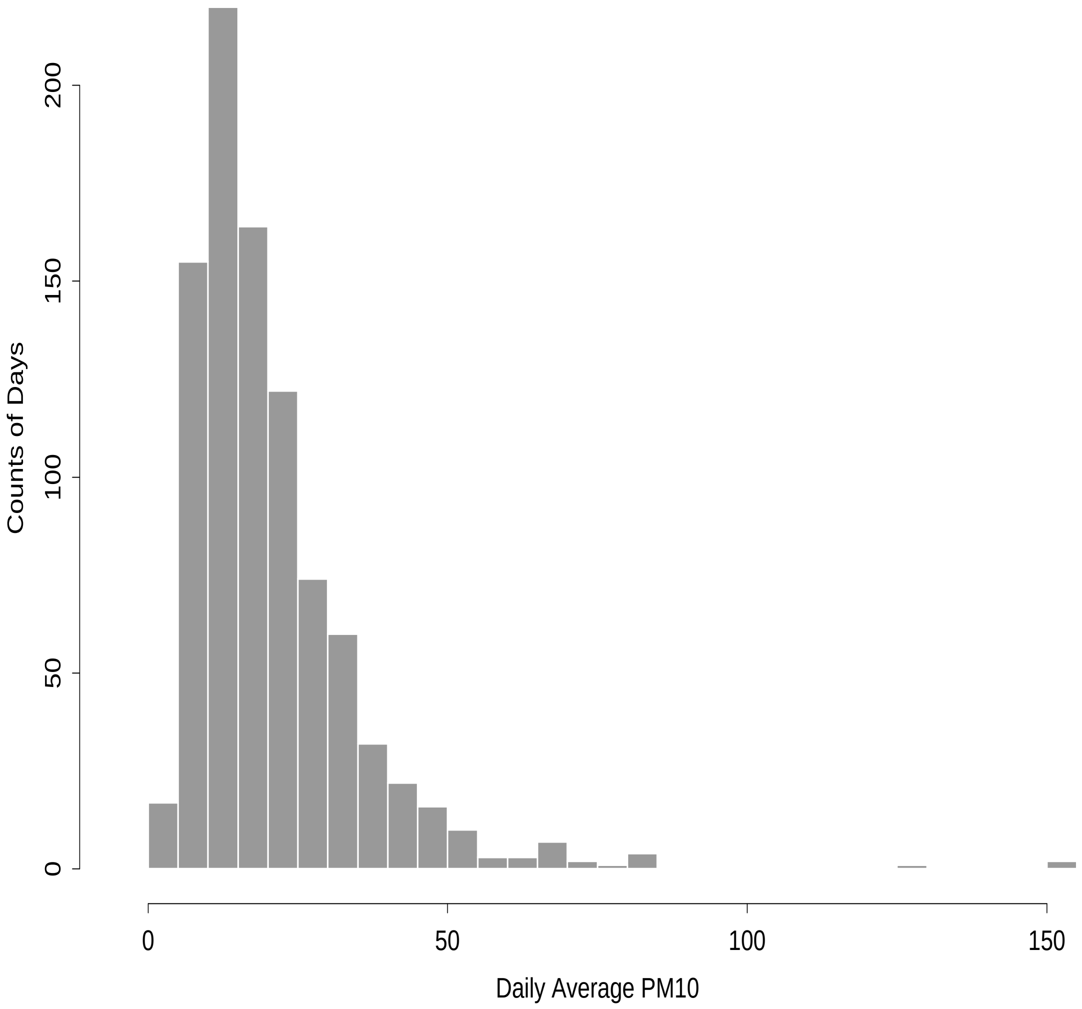


Figure C.12: Histogram of PM_{10} ($\mu g/m^3$)

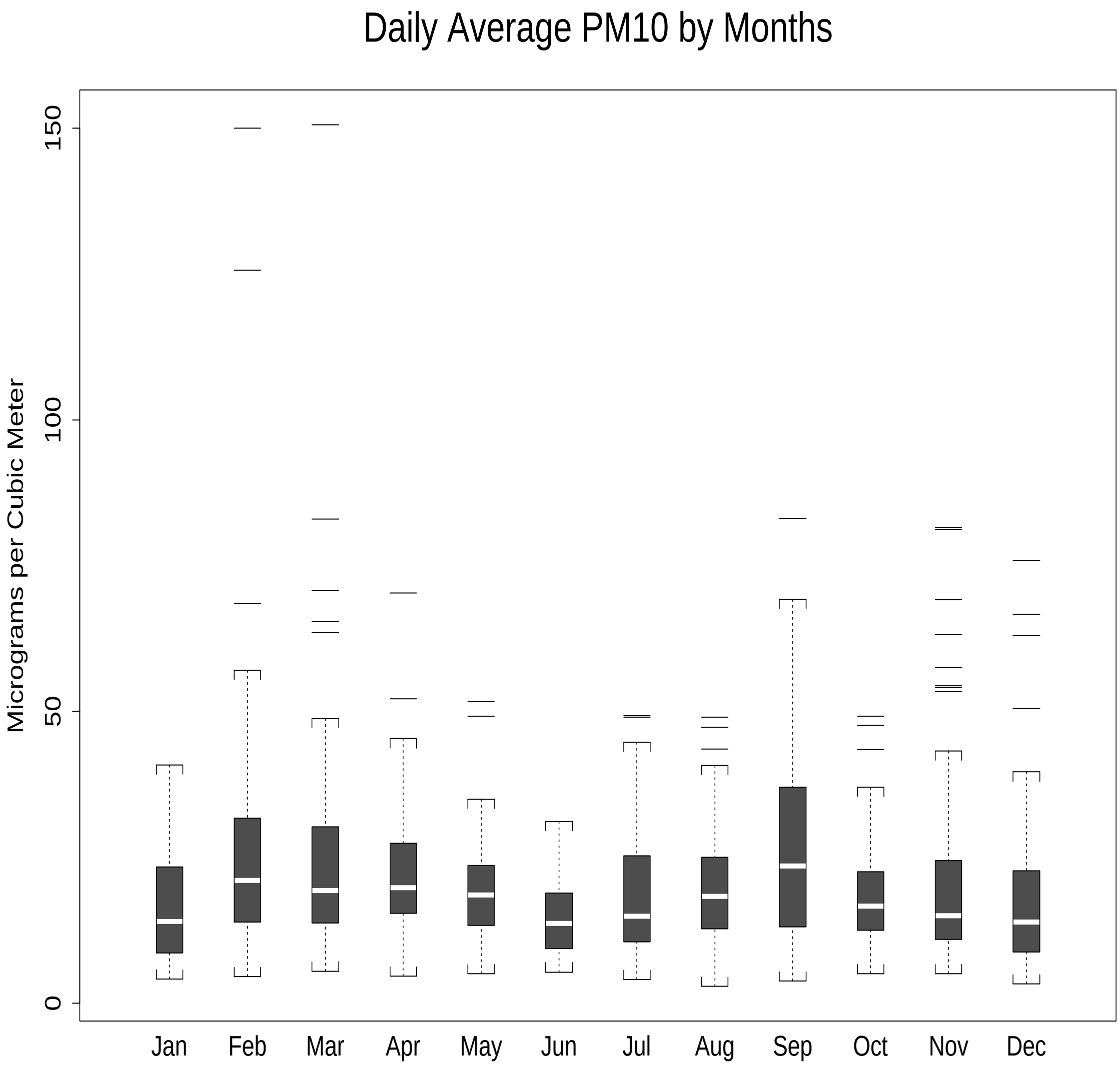


Figure C.13: Boxplots of PM_{10} ($\mu g/m^3$) by month

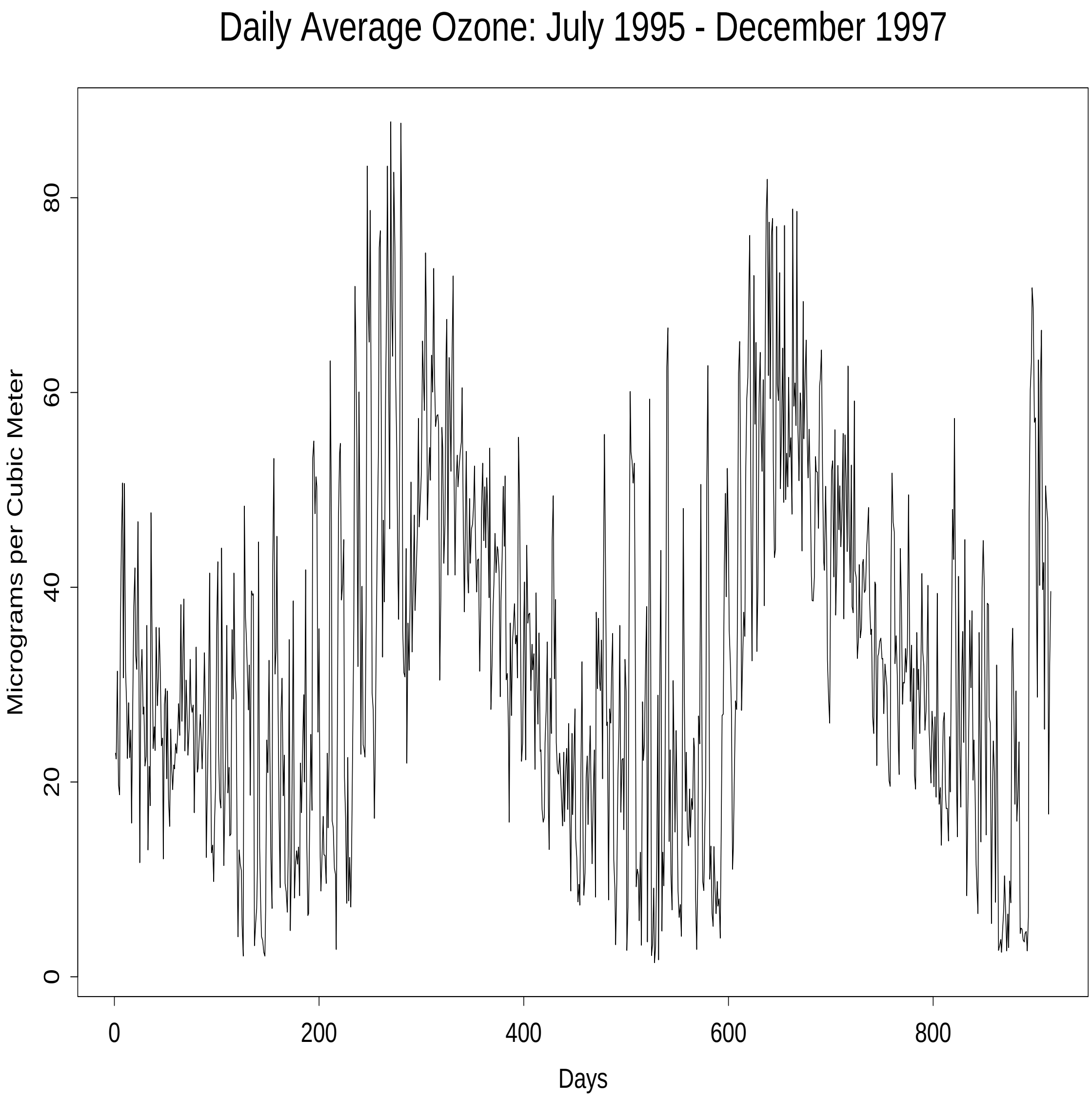


Figure C.14: Time series plot of ozone (μg/m^3) (Day 0 = June 30 1995)

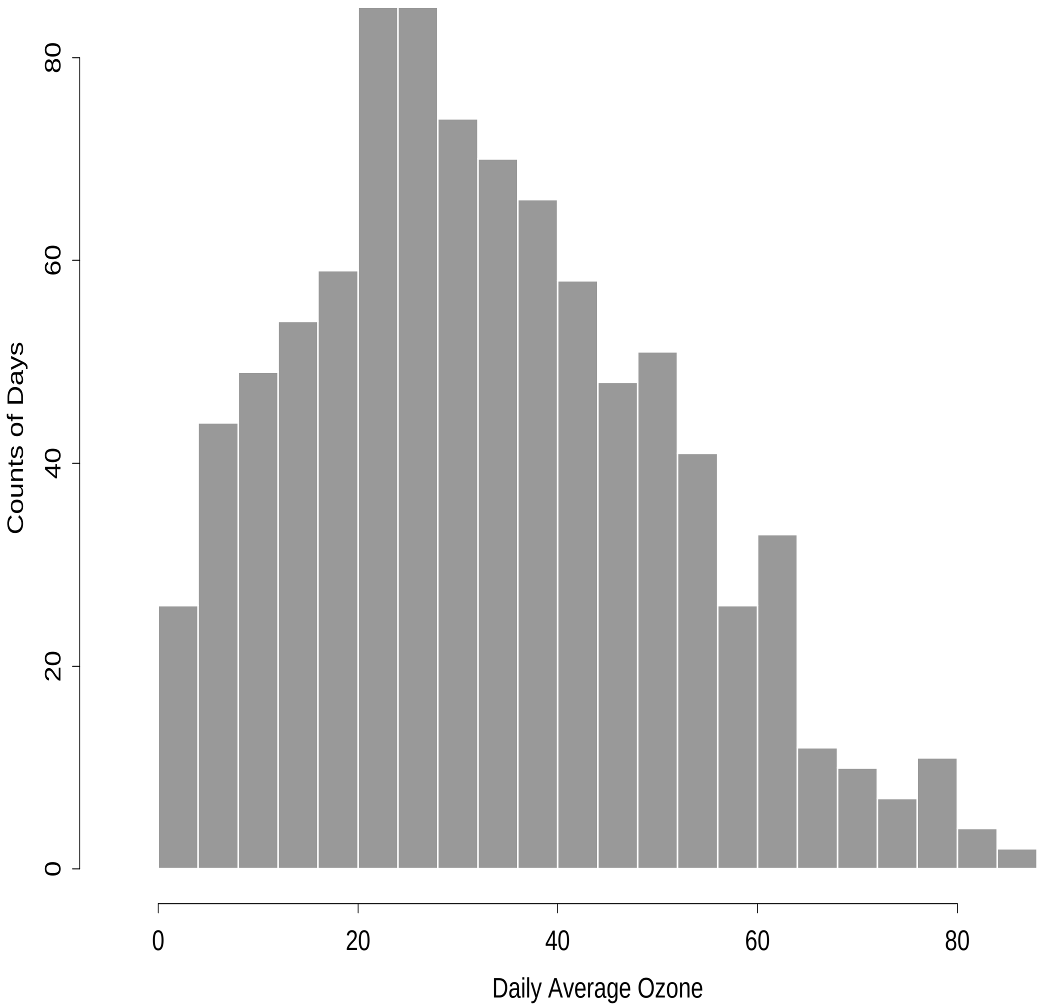


Figure C.15: Histogram of ozone (μg/m^3)

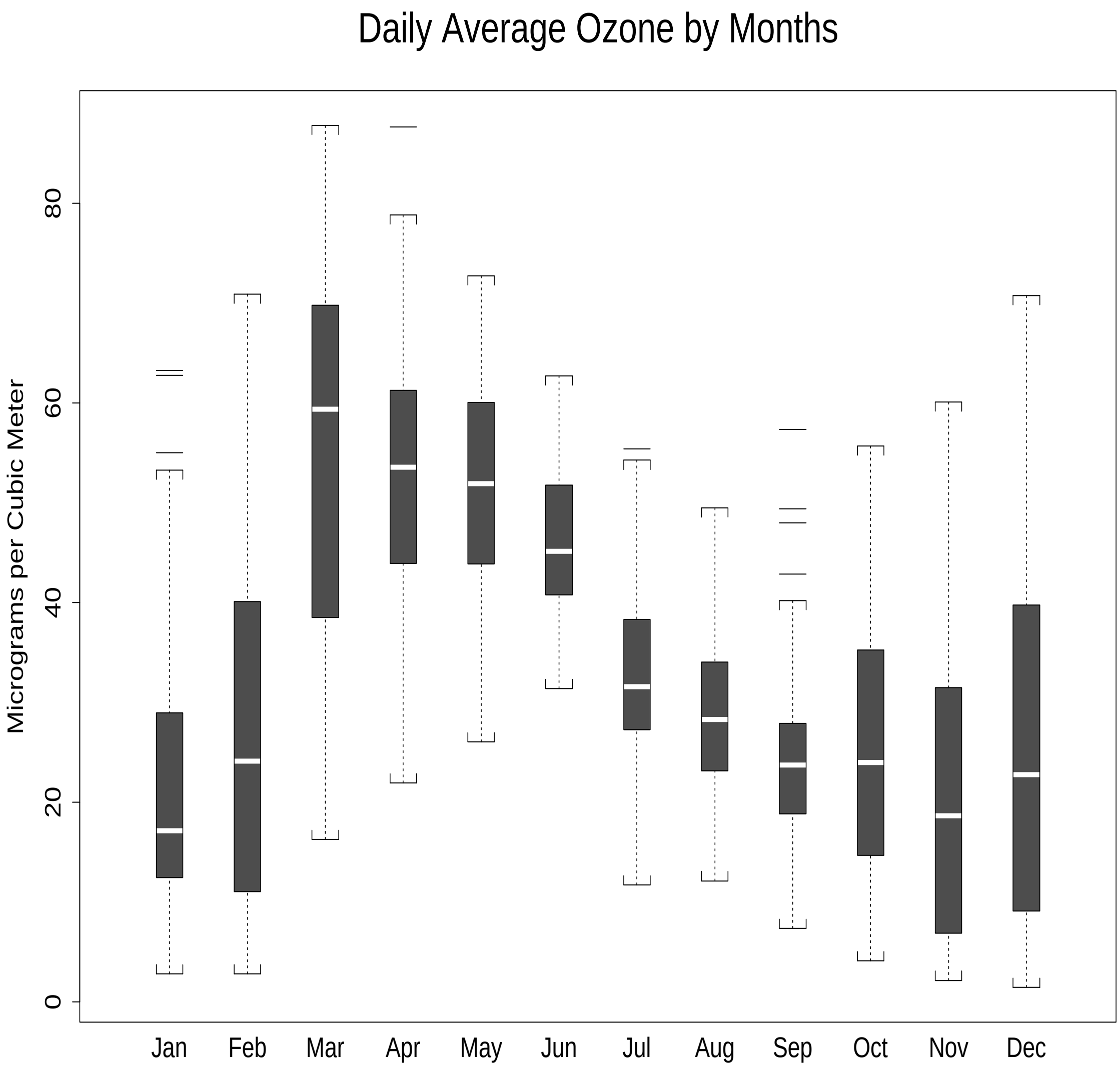


Figure C.16: Boxplots of ozone (μg/m^3) by month

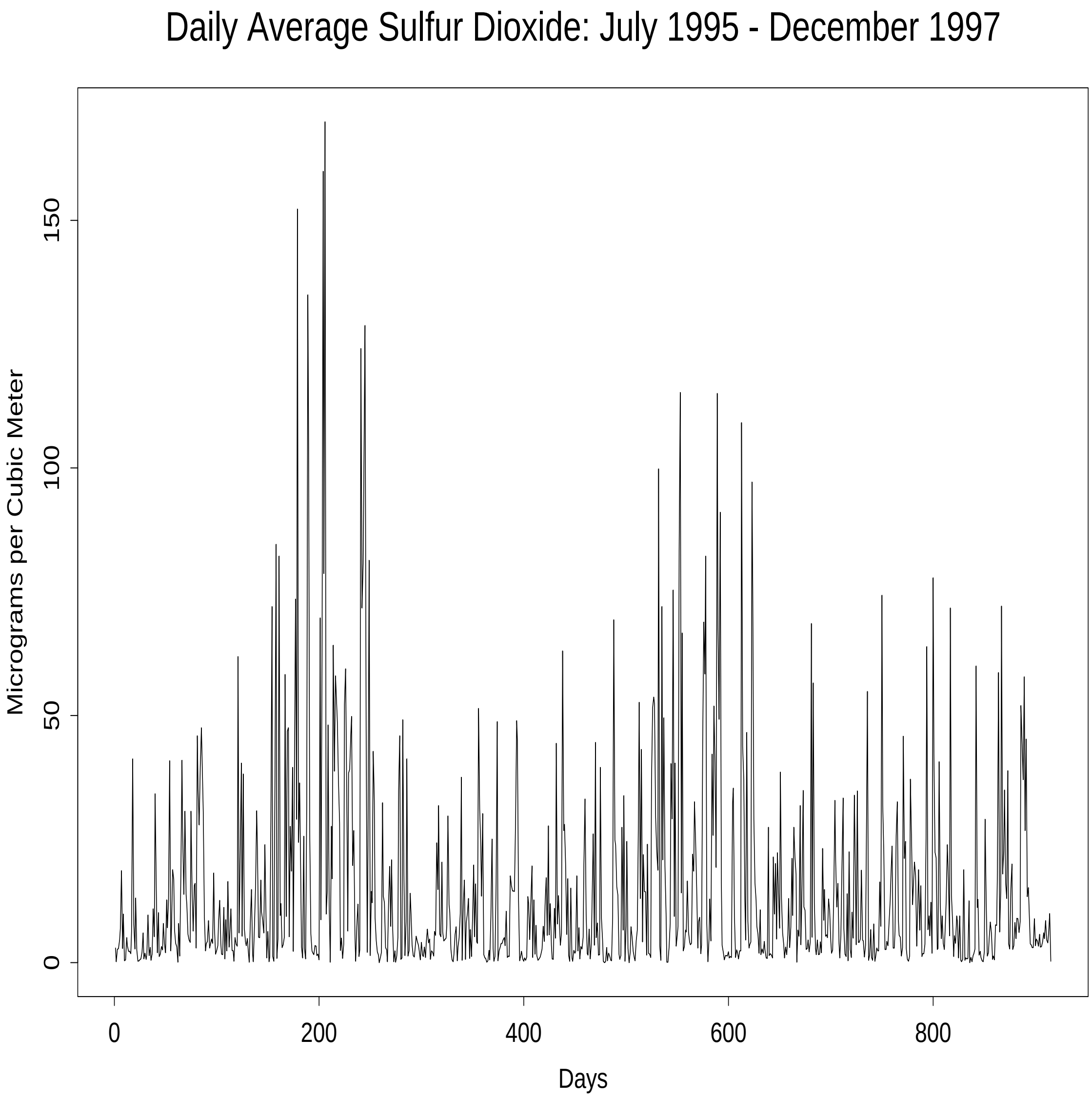


Figure C.17: Time series plot of sulfur dioxide (μg/m^3)

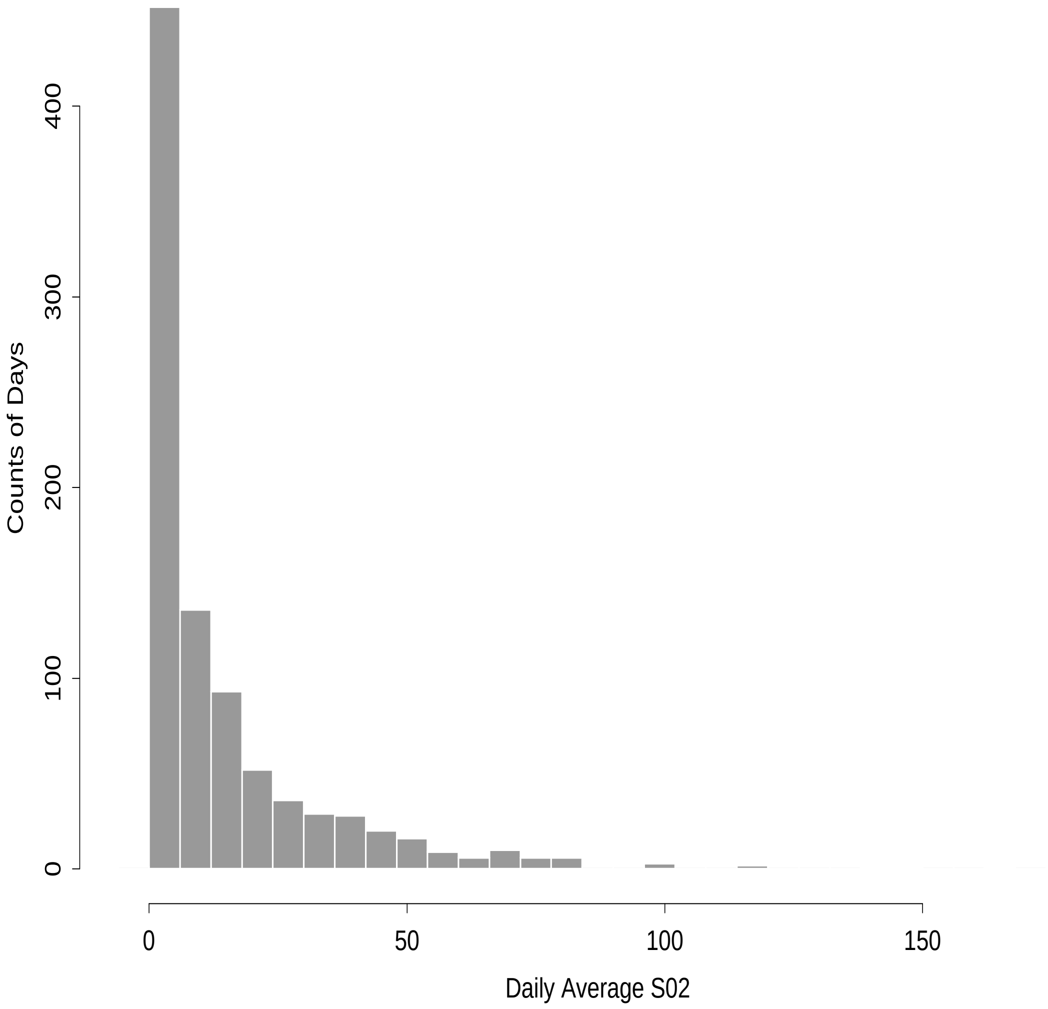


Figure C.18: Histogram of sulfur dioxide (μg/m^3)

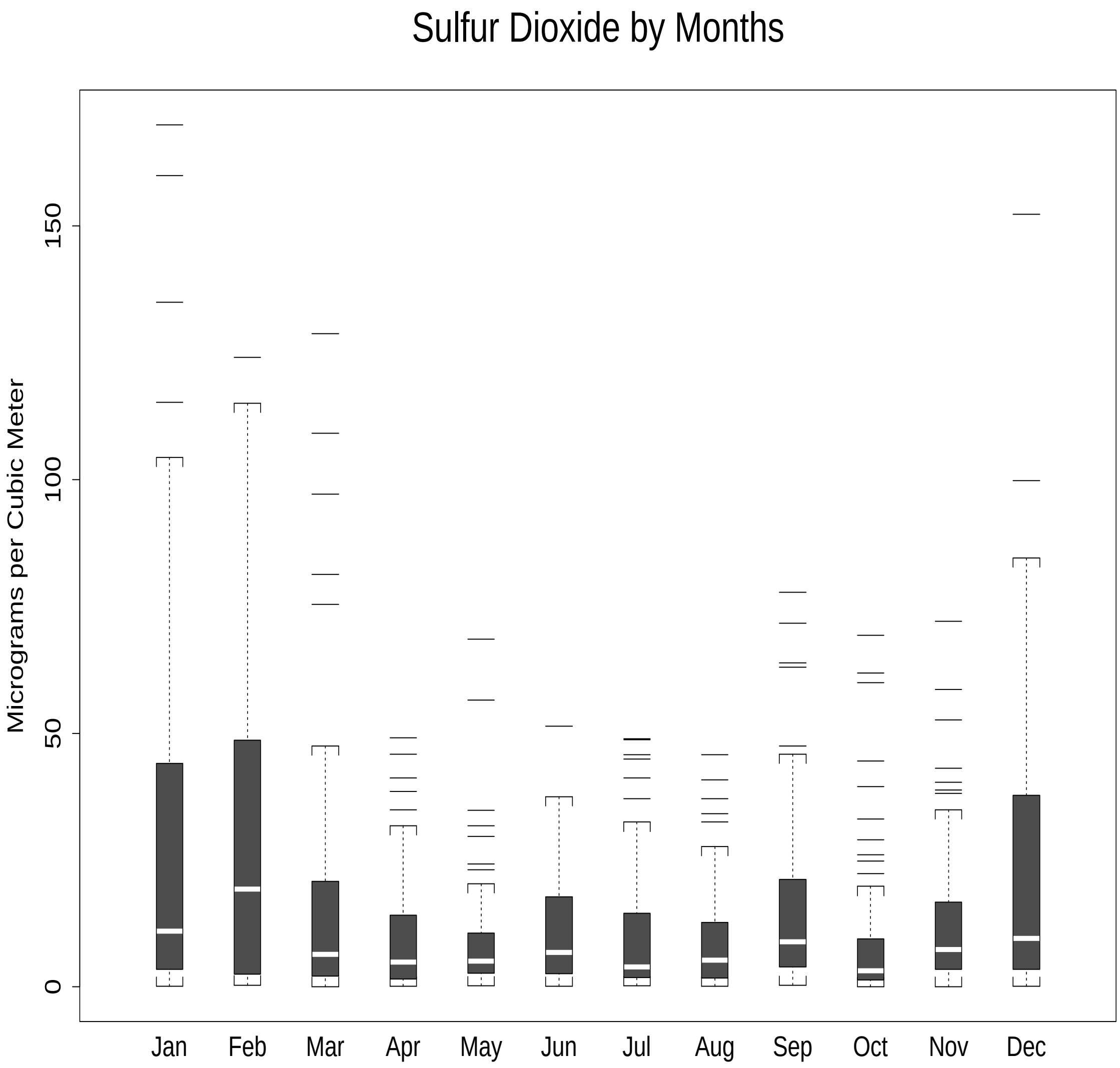


Figure C.19: Boxplots of sulfur dioxide ($\mu g/m^3$) by month

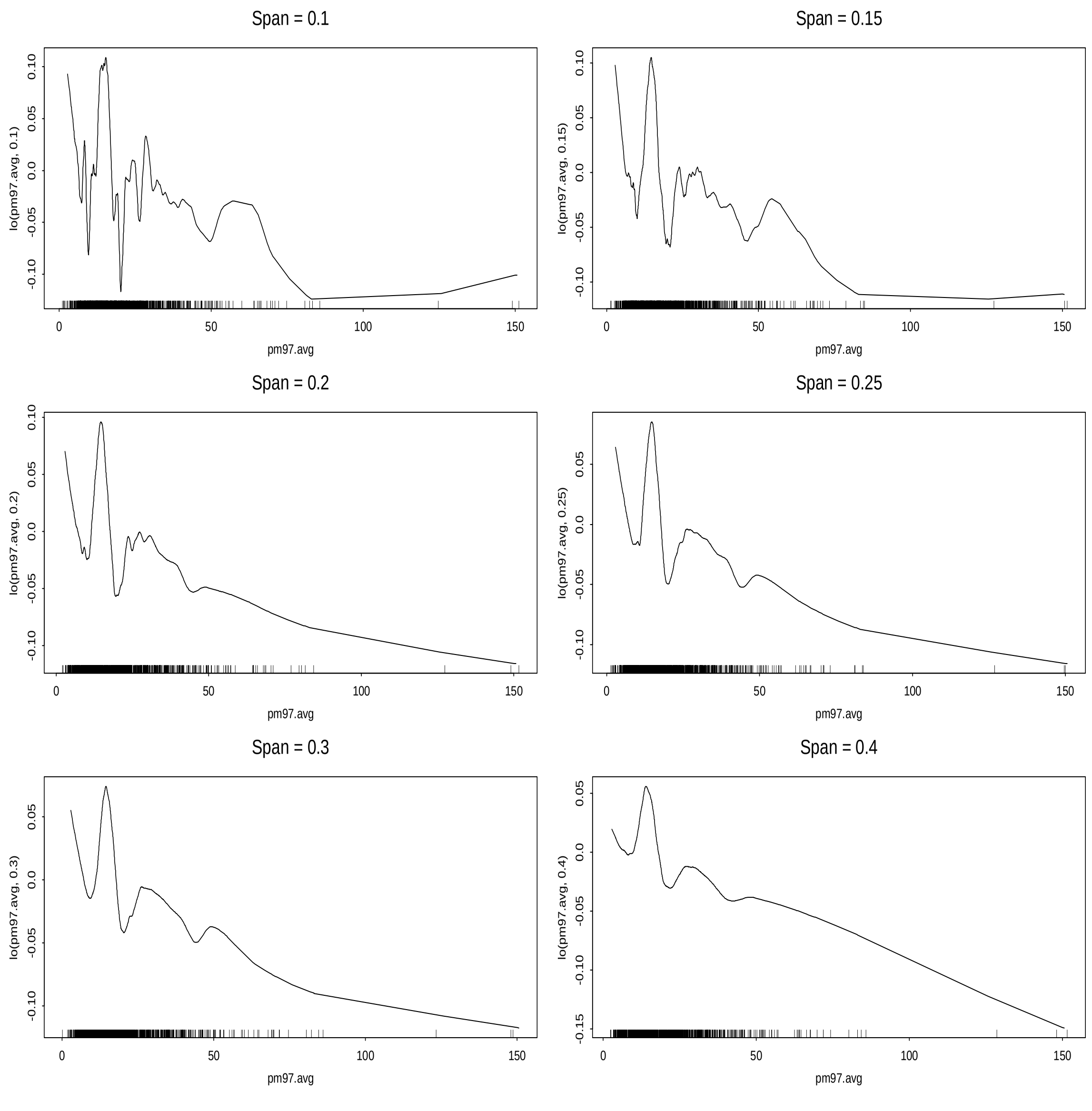


Figure C.20: Smoothed PM_{10} plot at different spans

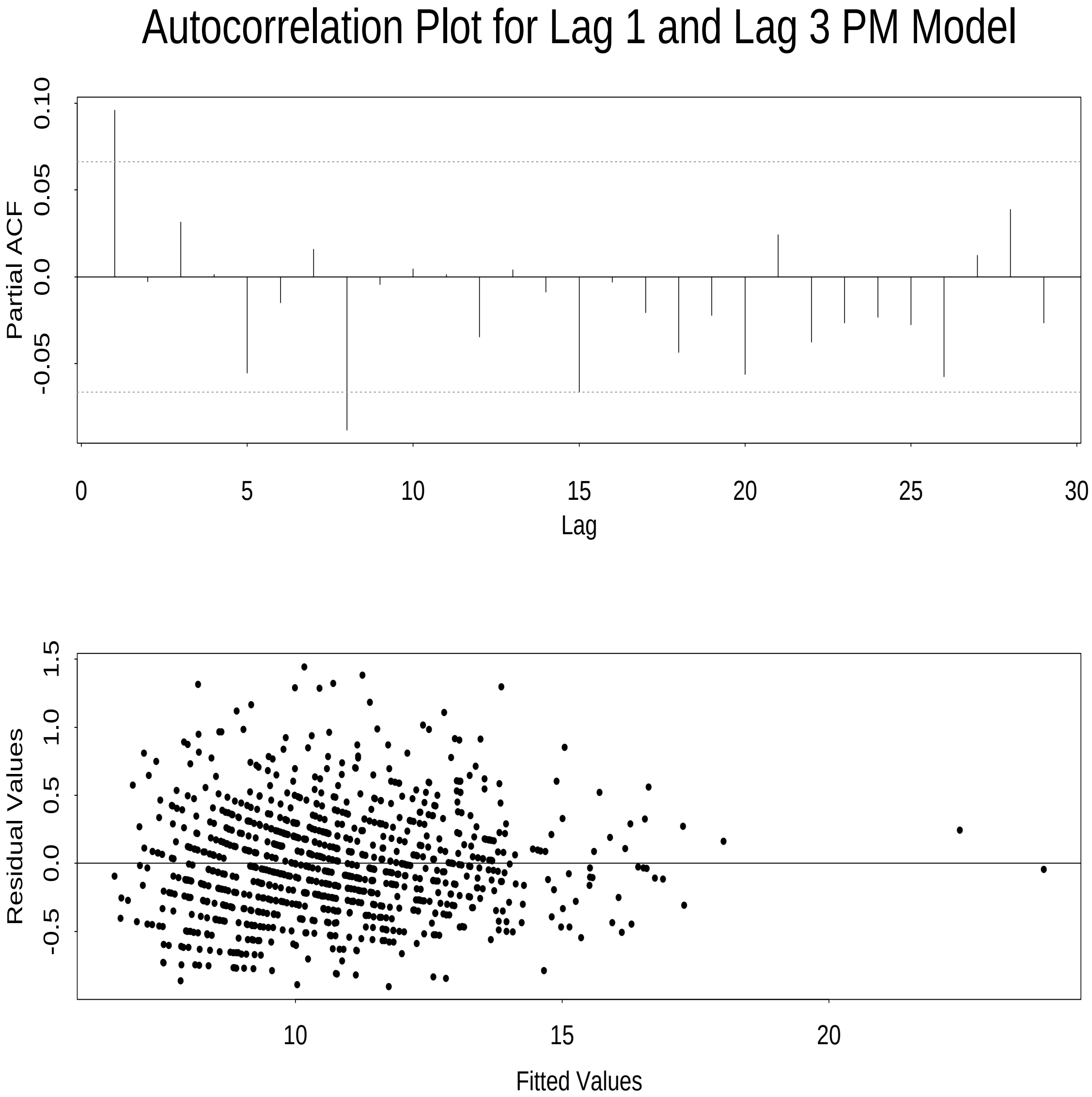


Figure C.21: Autocorrelation function of PM_{10} lags 1 and 3 model residuals ($\pm$ 2 standard errors indicated by dashed lines) and residual versus fitted values plot

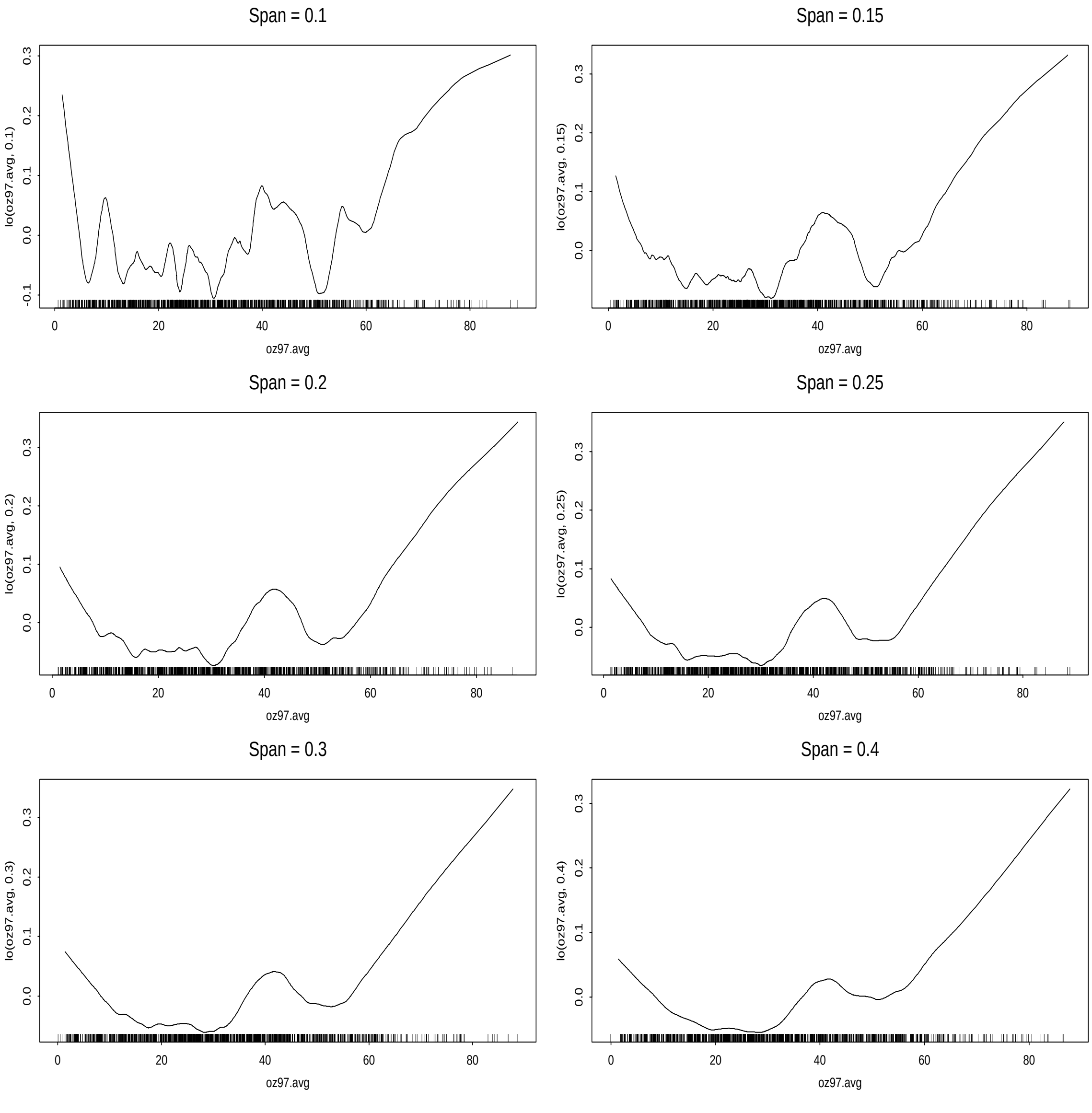


Figure C.22: Smoothed Ozone plot at different spans

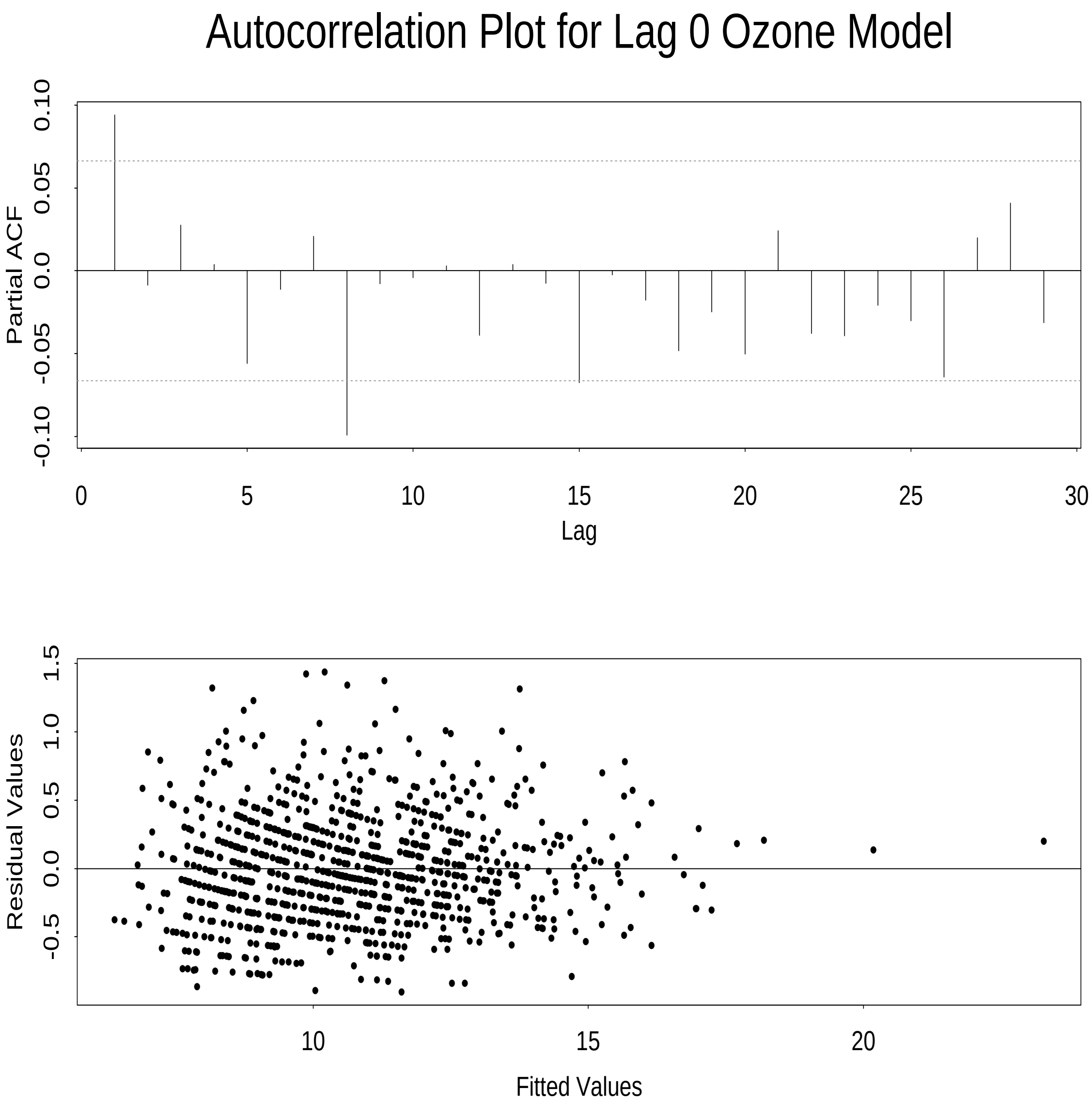


Figure C.23: Autocorrelation function of Ozone lag 0 model residuals (± 2 standard errors indicated by dashed lines) and residual versus fitted values plot

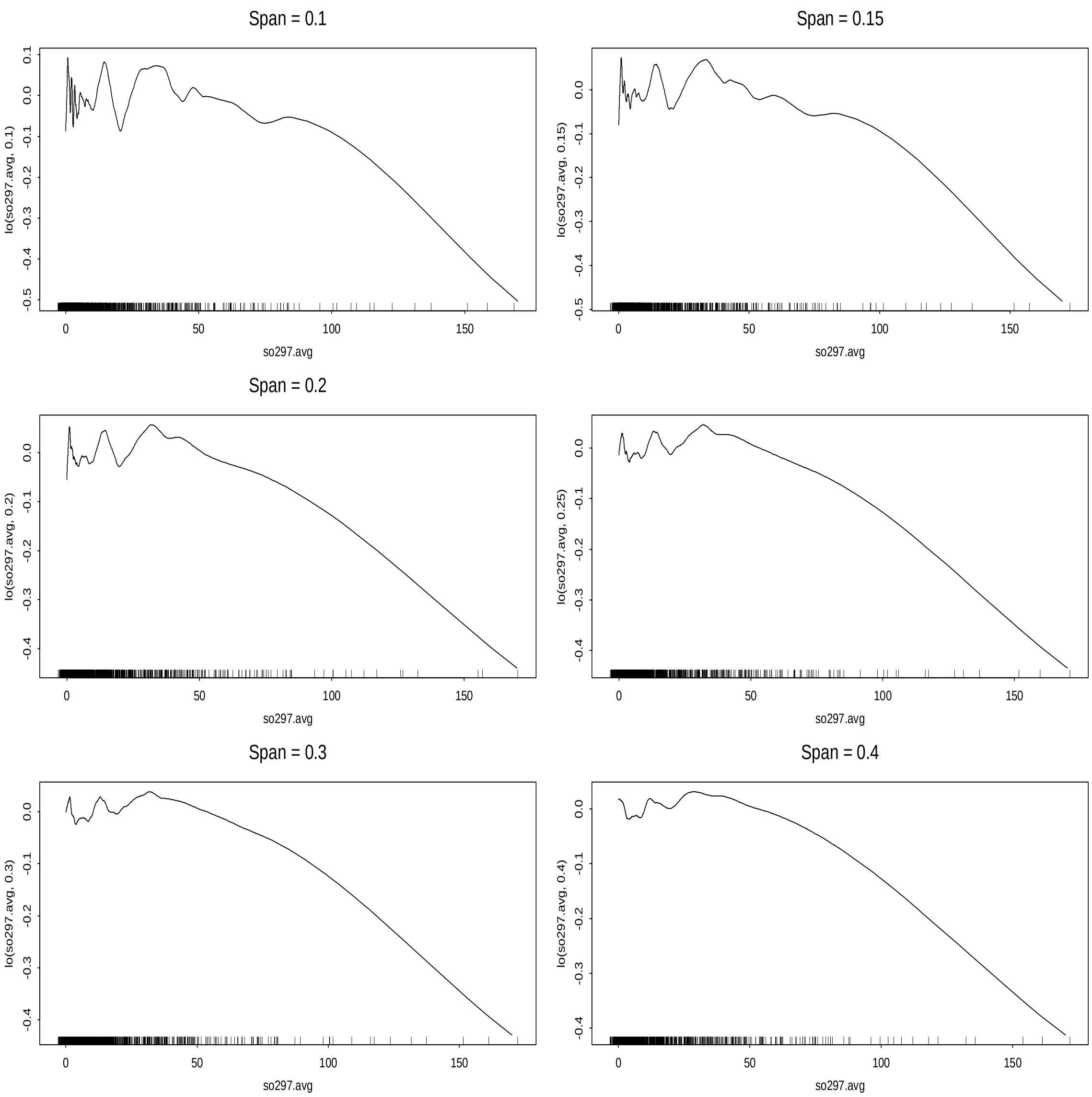


Figure C.24: Smoothed Sulfur Dioxide plot at different spans

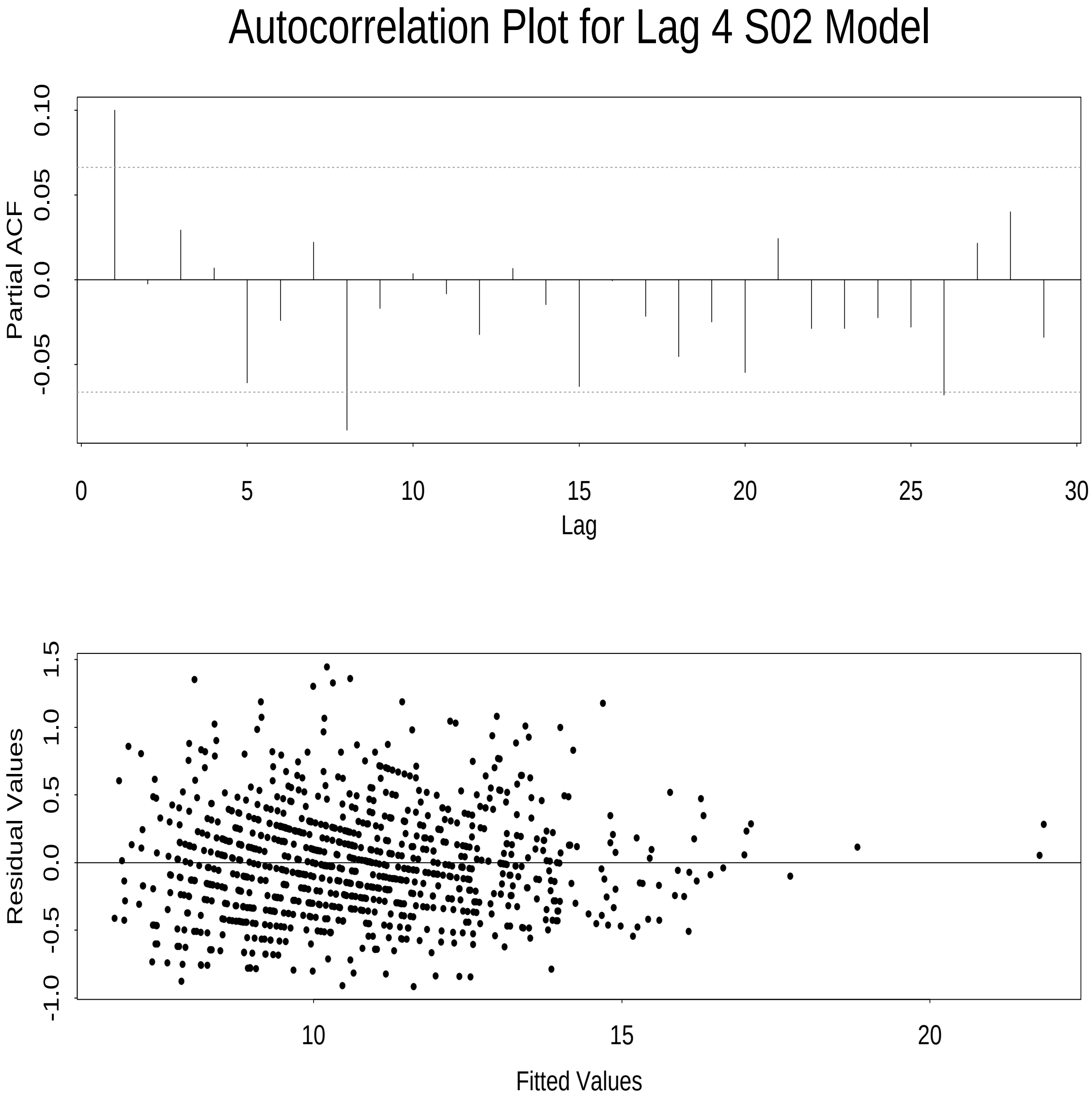


Figure C.25: Autocorrelation function of $S0_2$ lag 4 model residuals ($\pm$ 2 standard errors indicated by dashed lines) and residual versus fitted values plot

Appendix D

Sensitivity Analysis

As a sensitivity analysis for the loess smooth fitting of the temporal and meteorological effects, we repeated some of the analyses with a span of 0.5. Some of these results are presented in this appendix.

Temporal Effects

The results for various models for the patterns over times are presented in Table D.1. The only change to the results in Table 4.1 is for the model with a smooth time effect. Note that although the change in deviance due to the smooth time effect is considerably less here, the model accounting for the temporal effects is of the same form as before:

$$\log \lambda_i = \text{Day of the Week Effects} + \text{Smooth Time Effect}. \qquad \text{(D.0.1)}$$

Table D.1: Temporal Models

Model Terms	Deviance	df	Δdev	Δdf	p-value
Constant Rates Model	2015.6	914.0			
Linear Time Effect	2009.5	913.0	6.1	1.0	0.10
Day of the Week Effects	1942.8	908.0	72.8	6.0	<0.001
Day of the Week Effects + Linear Time Effect	1936.7	907.0	6.1	1.0	0.10
Day of the Week Effects + Smooth Time Effect	1834.3	904.7	108.5	3.0	<0.001

Table D.2: Meteorological Models

Model Terms	Deviance	df	Δdev	Δdf	p-value
Base = Temporal Effects	1834.3	904.7			
Base + Temp	1728.4	901.4	105.9	3.3	<0.001
Base + Wind	1807.4	894.2	26.9	10.5	0.24
Base + RH	1807.3	900.4	27.0	4.4	0.01
Base + Temp + Wind	1711.9	888.2	16.5	13.2	0.82
Base + Temp+ RH	1716.3	897.1	12.1	4.4	0.21

Meteorological Effects

The results for various models also incorporating smooth functions (with span of 0.5) of the meteorological effects are provided in Table D.2. Considering the meteorological effects individually, temperature is clearly the most important. Adding either wind or relative humidity in addition to the smooth temperature effect has little impact on the fit. Hence we obtain a base model accounting for both temporal and meteorological effects of the form:

$$\log \lambda_i = \text{Day of the Week Effects} + \text{Smooth Time Effect} + \text{Smooth Temp Effect}, \tag{D.0.2}$$

which differs from that identified in Chapter 4. In the remainder of this appendix, we examine the extent to which this change in the base model might affect the assessed relationships between the individual pollutants and the emergency room visit counts.

D.1 Particulate matter less than 10 microns (PM_{10})

The effect of PM_{10} on the same day was explored by the addition of a linear effect

$$\log \lambda_i = \text{Base Model} + \text{PM}_i.$$

The addition of this linear effect resulted in a small reduction in deviance (Δdev=1.8, Δdf=1, p-value=0.34). This fitted relationship between PM_{10} and the ER visit counts is negative (PM_{10} coefficient=−0.0009, SE=0.0010) which indicates that an increase in PM_{10} is associated with a slight decrease in the ER visit counts.

The addition of linear effects for PM_{10} at lags of 0 (same day) to 4 days and fitting of simultaneous linear effects for several lags was also explored. A summary of these model fits are presented in Table D.3 (see p.61).

Table D.3: PM_{10} Models

Model Terms	Deviance	df	Δdev	Δdf	p-value
Base = Temporal+Temp	1710.1	897.4			
Base + Lag 0	1708.3	896.4	1.8	1.0	0.34
Base + Lag 1	1709.2	896.4	0.9	1.0	0.50
Base + Lag 2	1706.5	896.4	3.6	1.0	0.17
Base + Lag 3	1698.7	896.4	11.4	1.0	0.01
Base + Lag 4	1698.8	896.4	11.3	1.0	0.01
Base + Lag 3 + Lag 0	1695.0	895.4	3.7	1.0	0.16
Base + Lag 3 + Lag 1	1693.6	895.4	5.1	1.0	0.11
Base + Lag 3 + Lag 2	1698.5	895.4	0.2	1.0	0.76
Base + Lag 3 + Lag 4	1696.4	895.4	2.3	1.0	0.27
Base + Lag 4 + Lag 0	1696.6	895.4	2.2	1.0	0.28
Base + Lag 4 + Lag 1	1696.4	895.4	2.4	1.0	0.26
Base + Lag 4 + Lag 2	1698.3	895.4	0.5	1.0	0.64

Considering first only the individual effects, adding the lag 3 effect to the base model produced the greatest reduction in deviance (Δdev=11.4, Δdf=1, p-value=0.01). Adding the lag 1 effect in addition to the lag 3 effect produces some reduction in deviance (p-value=0.11). Adding an individual lagged effect to the model with both lags 1 and 3 provides negligible additional reduction in the deviance (results not shown). The overall reduction of deviance associated with including the lag 1 and lag 3 effects is statistically significant (Δdev=16.5, Δdf=2, p-value=0.01). Therefore, the effect of PM_{10} is fitted as follows:

$$\log \lambda_i = \text{Base Model } + \mathrm{PM}_{i-1} + \mathrm{PM}_{i-3}. \tag{D.1.3}$$

The estimated coefficients of PM_{10} at lag 1 and 3 in this model are -0.0017 (SE=0.0010) and 0.0029 (SE=0.0009) respectively. The form of the model (D.1.3) with PM_{10} at lags 1 and 3 is selected partly on the basis of the statistical results and partly to allow a direct comparison to the model (5.1.1).

D.2 Ozone (O_3)

As an initial exploratory step, ozone is fitted by the following model:

$$\log \lambda_i = \text{Base Model} + \mathrm{O}_{3i} + \mathrm{O}_{3i}^2. \tag{D.2.4}$$

Table D.4: O_3 Models

Model Terms	Deviance	df	Δdev	Δdf	p-value
Base = Temporal+Temp	1710.1	897.4			
Base + Lag 0	1690.6	895.4	19.5	2.0	0.01
Base + Lag 1	1706.2	893.4	3.9	4.0	0.73
Base + Lag 2	1706.3	893.4	3.8	4.0	0.74
Base + Lag 3	1709.2	895.4	0.9	2.0	0.79
Base + Lag 4	1709.1	895.4	1.0	2.0	0.77
Base + Lag 0 + Lag 1	1690.2	891.4	0.4	4.0	0.99
Base + Lag 0 + Lag 2	1684.9	891.4	5.7	4.0	0.55
Base + Lag 0 + Lag 3	1685.6	893.4	5.0	2.0	0.27
Base + Lag 0 + Lag 4	1687.0	893.4	3.6	2.0	0.38

As indicated by Table D.4, the addition of this quadratic effect for same day O_3 to the base model yields a substantial reduction in the deviance (Δdev=19.5, Δdf=2, p-value=0.01). The estimated coefficients of the linear and quadratic components of O_3 are -0.0073 (SE=0.0028) and 0.00011 (SE=0.00004) respectively. This suggests that there is indeed a significant negative linear effect and a positive quadratic effect of O_3 on ER visits.

Adding a quadratic effect for lag 0, 3, or 4, or a 4^{th} degree polynomial effect for lag 1 or 2 of O_3 to the base model produces no large reduction of deviance except for the same day effect. Similarly, adding any of these individual lagged effects to the model with the quadratic lag 0 effect included provides little reduction in the deviance. As a result only the same day effect of O_3 on the ER visit counts is of interest.

D.3 Sulfur Dioxide (SO_2)

The effect of SO_2 on the same day was explored by the addition of a linear effect

$$\log \lambda_i = \text{Base Model} + \text{SO}_{2i}.$$

The addition of this effect does not reduce the deviance substantially (Δdev=1.4, Δdf=1, p-value=0.40). The estimated regression coefficient of SO_2 is -0.0006 (SE=0.0007). This suggests that SO_2 has a negative relationship with the ER visit counts.

Linear effects at other lags (lags 0, 1, 2, 3, 4) were fit to see if there is a large reduction in deviance associated with different lags of SO_2. The fits of these models are summarized in Table D.5 (see p.63).

Table D.5: SO_2 Models

Model Terms	Deviance	df	Δdev	Δdf	p-value
Base = Temporal+Temp	1710.1	897.4			
Base + Lag 0	1708.7	896.4	1.4	1.0	0.40
Base + Lag 1	1707.4	896.4	2.7	1.0	0.24
Base + Lag 2	1709.9	896.4	0.2	1.0	0.76
Base + Lag 3	1706.6	896.4	3.5	1.0	0.18
Base + Lag 4	1694.4	896.4	15.7	1.0	0.004
Base + Lag 4 + Lag 0	1693.1	895.4	1.3	1.0	0.41
Base + Lag 4 + Lag 1	1689.0	895.4	5.3	1.0	0.09
Base + Lag 4 + Lag 2	1692.4	895.4	2.0	1.0	0.31
Base + Lag 4 + Lag 3	1694.1	895.4	0.3	1.0	0.72

Among the individual lag effects, lag 4 results in substantial reduction in deviance (Δdev=15.7, Δdf=1, p-value=0.004). The estimated coefficient of SO_2 at lag 4 is 0.0018 (SE=0.0006) which is quite large and indicates a positive relationship between the emergency room visit counts and sulfur dioxide.

Adding the lag 1 effect to the model with the lag 4 effect produces some reduction in deviance but this is not too substantial (Δdev=5.3, Δdf=1, p-value=0.09). Adding any other individual lag to the model containing effects for lags 1 and 4 produces negligible additional reduction in deviance (results not shown). The estimated coefficients for this model are -0.0010 (SE=0.0007) and 0.0019 (SE=0.0006) for lags 1 and 4 respectively. This gives a similar lag 4 effect as in the model containing only the lag 4 effect. Since the reduction of deviance associated with including the lag 4 effect is statistically significant (p-value=0.004), the following model is considered as the fitted model for SO_2:

$$\log \lambda_i = \text{Base Model} + \text{SO}_{2,i-4}. \tag{D.3.5}$$

This model is of the same form as the fitted model for SO_2 in the main text, thereby facilitating direct comparisons.

D.4 Comparison of the results for span 0.1 and 0.5.

Table D.6 compares the predictions obtained for the impact of the individual pollutants on the emergency room visit counts when the base model is fit with span of 0.1 and 0.5. If we use a span

Table D.6: Predicted percent decreases in ER visit counts

		Span=0.1		Span=0.5	
Pollutant	Decrease in Pollutant	Estimate	95% CI	Estimate	95% CI
PM_{10}	Lag 3: 20 to 10 $\mu g/m^3$	−2.0%	(−4.1%, 0.0%)	−0.9%	(−2.9%, 0.9%)
PM_{10}	Lags 1, 3: 20 to 10 $\mu g/m^3$	0.2%	(−1.9%, 2.3%)	1.2%	(−0.9%, 3.4%)
Ozone	Lag 0: 30 to 15 $\mu g/m^3$	−1.4%	(−5.9%, 2.8%)	−3.5%	(−7.2%, −0.1%)
Ozone	Lag 0: 45 to 30 $\mu g/m^3$	3.5%	(−0.4%, 7.2%)	1.4%	(−0.9%, 3.7%)
SO_2	Lag 4: 6 to 3 $\mu g/m^3$	0.4%	(0.1%, 0.8%)	0.5%	(0.2%, 0.9%)

of 0.5 instead of 0.1 for fitting the base model, the signs of the coefficients of the pollutants do not change, although their estimated effects (and corresponding SEs) differ somewhat. A decrease in PM_{10} for both lag 1 and lag 3 from 20 $\mu g/m^3$ to 10 $\mu g/m^3$ (50%), while controlling the temporal and meteorological effects, is associated with a predicted 0.2% decrease in the ER visit counts for span 0.1, but a 1.2% decrease for span 0.5. A decrease in same day ozone from 30 $\mu g/m^3$ to 15 $\mu g/m^3$ (50%) is associated with a predicted increase of 1.4% in the ER visit counts for span 0.1 and an increase of 3.5% for span 0.5. If, on the other hand, same day ozone decreases from 45 $\mu g/m^3$ to 30 $\mu g/m^3$ (33.3%), then the ER visit counts are predicted to decrease by 3.5% for span 0.1 and decrease by 1.4% for span 0.5. A decrease in sulfur dioxide from 6 $\mu g/m^3$ to 3 $\mu g/m^3$ (50%) 4 days ago, is associated with a decrease of 0.4% in the ER visit counts for span 0.1 and 0.5% for span 0.5. Although the predicted decreases with the span of 0.5 differ somewhat from those with the span of 0.1, overall, both sets of decreases are quite small and the qualitative interpretation differs little.