The effect of air pollution & climate on hospital admissions for chronic obstructive airways disease: a non-parametric alternative

Short title	The effect of air pollution & climate on hospital admissions

Authors

Bircan Erbas

Department of General Practice & Public Health,

The University of Melbourne, VIC 3010, Australia.

Rob J Hyndman

Department of Econometrics and Business Statistics,

Monash University, VIC 3800, Australia.

Email: Rob.Hyndman@buseco.monash.edu.au.

Telephone: +613 9905 2358. Fax: +613 9905 5474.

Corresponding author: Rob Hyndman

The effect of air pollution & climate on hospital admissions for chronic obstructive airways disease: a non-parametric alternative

Bircan Erbas[1] and Rob J Hyndman[2]

11 May 2000

Abstract: Medical researchers have discovered an alarming increase in rates of respiratory disease in many countries. Exposure to environmental factors such as air pollution, ozone and pollen seem to be possible causes. In recent years several studies have shown a short term association between exposure to air borne particles and ozone and an increase in respiratory hospital admissions.

The aim of this paper is to examine methodological issues in epidemiological time series studies involving count data. As a vehicle of illustration, we consider the effect of air pollution and climate on daily counts of hospital admissions for chronic obstructive airways disease. We apply non-parametric (Generalized Additive) models and parametric (Generalized Linear) models, and examine the relationship between hospital admissions for chronic obstructive airways disease and the short term exposure to pollutants.

We show that the Generalized Additive Model is superior to the Generalized Linear Model when capturing the non-linear pattern of the response variable with a number of the covariates.

One problem in epidemiological time series studies is incorporating autocorrelation in the statistical models. We show that correct specification of the non-linear functional form, and allowing for seasonality, will in some cases remove the autocorrelation inherent in morbidity/mortality data.

We also present an alternative residual analysis when the response is not normally distributed.

[1]Department of General Practice & Public Health, The University of Melbourne, VIC 3010, Australia.
[2]Department of Econometrics and Business Statistics, Monash University, VIC 3800, Australia.

Introduction

In recent years several studies from a number of groups in America[1], Europe[2,3] and more recently Australia[4] have shown a short term association between exposure to low levels of pollutants and daily hospital admissions for Chronic Obstructive Pulmonary Disease (COPD).

COPD is a cardiorespiratory disease characterized by chronic and usually progressive impairment of air flow due to obstruction, damage and disorganization of the airways[2]. Individuals with advanced COPD may experience respiratory stress when exposed to toxic pollutants which may precipitate an admission to hospital[5]. Thus emergency room admissions for COPD has been accepted as a good indicator of the short term respiratory effects of air pollution[3,6].

Pollutants including sulfur dioxide, nitrogen dioxide and suspended particulates have been recognized as respiratory irritants. In fact, current biological knowledge suggests that sulfur dioxide can exacerbate COPD[3]. Ozone also produces a range of complex pulmonary responses when levels are between 0.08 and 0.20ppm[7].

In Melbourne, Australia, the air quality has steadily improved over the past fifteen years and is currently ranked as "good" by international standards[8]. The largest polluter of air in Melbourne is motor vehicles. Motor vehicles are responsible for about half of Melbourne's air pollution in Summer, contributing 65% of nitrogen oxide emissions[8] and 25% of airborne particles. Both levels of ozone and nitrogen dioxide have remained consistently below the acceptable air quality standards for Melbourne.

To quantify the relationship between hospital admissions for COPD and exposure to air pollutants, many authors have proposed a variety of statistical methodologies depending on the objectives, and understanding of the problem[9].

The most commonly used statistical methodology has been Autoregressive Poisson models developed for the APHEA (Air Pollution and Health, a European Approach) project[1].

This is a parametric method where a Poisson distribution is assumed for the response and an autoregressive process is assumed for the correlation structure in the residuals. However in some instances allowing nonlinearity can avoid the problem of autocorrelation and thereby simplify the model. Thus it is imperative that the functional form between the response and the covariates is specified correctly before any assumption is made on the correlation structure of the residuals.

In this paper we will present the application of two alternative statistical methodologies: Generalized Linear Models[10] and Generalized Additive Models[11] to examine the relationship between hospital admissions for COPD and short term exposure to pollutants. We shall demonstrate that the use of nonparametric (and therefore nonlinear) regression relationships avoids the need to allow for autocorrelation in the models. The resulting models are also more easily and directly interpretable than the Autoregressive Poisson models which are commonly used. We will also present alternative residual analysis in non-normal regression situations where customary definitions of residuals are less useful[12].

Thus our objective is to concentrate on methodological issues and to further develop the methodology for the detection of short term health effects in the analysis of epidemiological time series data.

Exploratory data analysis

COPD hospital admissions, air pollution and climatic data

Daily counts of hospital admissions for ICD 496 (chronic obstructive airways disease) were obtained from the Public Health & Development Division of the Department of Human Services, Victoria, Australia. Air pollution data was obtained from the Environment Protection Authority (EPA), which maintains a network of 12 monitoring stations around Melbourne. Daily maximum hourly levels of Nitrogen Dioxide (NO_2), Ozone

Table 1: *Descriptive measures of Hospital admissions for COPD pollutants and climatic variables in Melbourne, Australia for 1989 to 1992*

			Percentile					
Variable	**Mean**	**Min**	10th	25th	50th	75th	90th	**Max**
COPD	10^a	0	5	8	10	13	16	28
Nitrogen dioxide (NO_2)^b	2.45	0.45	1.23	1.74	2.38	3.03	3.73	9.33
Air Particle Index (API)^c	0.86	0.33	0.43	0.5	0.66	1.02	1.56	4.85
Sulfur dioxide (SO_2)^b	0.60	0	0.13	0.27	0.5	0.77	1.15	3.34
Ozone (O_3)^b	2.78	0.75	1.78	2.1	2.45	2.96	4.28	14.35
Humidity (hu)^d	69.76	24.66	55.82	63.42	70.4	76.86	82.86	94.15
Dew point temperature (wb)^d	11.13	3.45	6.97	8.46	10.75	13.49	15.85	21.21
Dry bulb temperature (db)^d	14.17	5.06	8.91	10.66	13.47	17.28	20.30	32.48

^{a}The median
^{b}pphm
^{c}bscat/10^4m
^{d}degrees

(O_3), Sulfur Dioxide (SO_2) and particulate (API) were obtained. Three hourly humidity (hu), dry bulb temperature (db) and dew point temperature (wb) were obtained from the Commonwealth Bureau of Meteorology which has four major stations in the Melbourne metropolitan area.

Seasonality and pairwise relationships

Figure 1 shows the result of a cubic smoothing spline applied to the daily number of hospital admissions for COPD as a function of time from July 1989 to December 1992. Inspection of Figure 1 clearly shows a strong seasonal pattern. This strong seasonal pattern may be due to inherent seasonality in COPD admissions or it may be induced by seasonality in some meteorological or pollutant covariates.

A scatter plot is a graphical tool that allows a visual examination of the possible functional form between the response (COPD admissions) and each of the pollutants. A more powerful graphical tool is the matrix of pairwise scatter plots where pairwise scatter plots of all the pollutants and climatic data are constructed. The pairwise scatter plot in Figure 2 shows some of the response variable and the key explanatory variables.

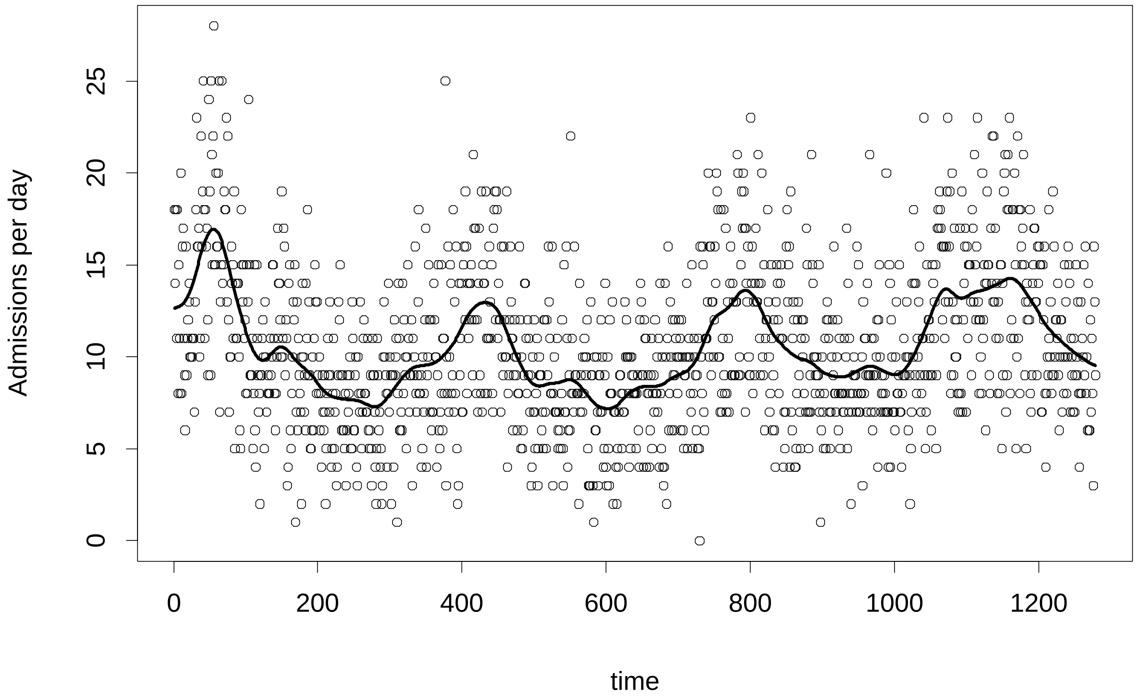

Figure 1: *Smoothing spline of daily counts of hospital admissions for Chronic Obstructive Pulmonary disease in Melbourne, Australia from 1 July 1989 to 31 December 1992*

It is evident that the functional forms between the response variable and the climatic variables—humidity (hu), dry bulbs temperature (db) and dew point temperature (wb)—are non-linear. This non-linear relationship is consistent with results reported previously[4,1]. The functional form between COPD and the pollutants is not as clear as the climatic variables.

The pairwise scatter plot in Figure 3 shows the same variables after each of them has been seasonally adjusted. The seasonal adjustment was carried out using the classical decomposition method[13] with seasonal period equal to 365. Now any relationship between COPD and the explanatory variables is not evident, suggesting that what was seen in Figure 2 was induced by the seasonal variation in COPD and the climatic variables, rather than any direct relationship between the variables. For this reason, it will be important to include regressors in our models which allow for seasonality.

In the models we apply in Sections 3 and 5 we shall separate out the effect of seasonality so that the relationship between response and explanatory variables can be studied with-

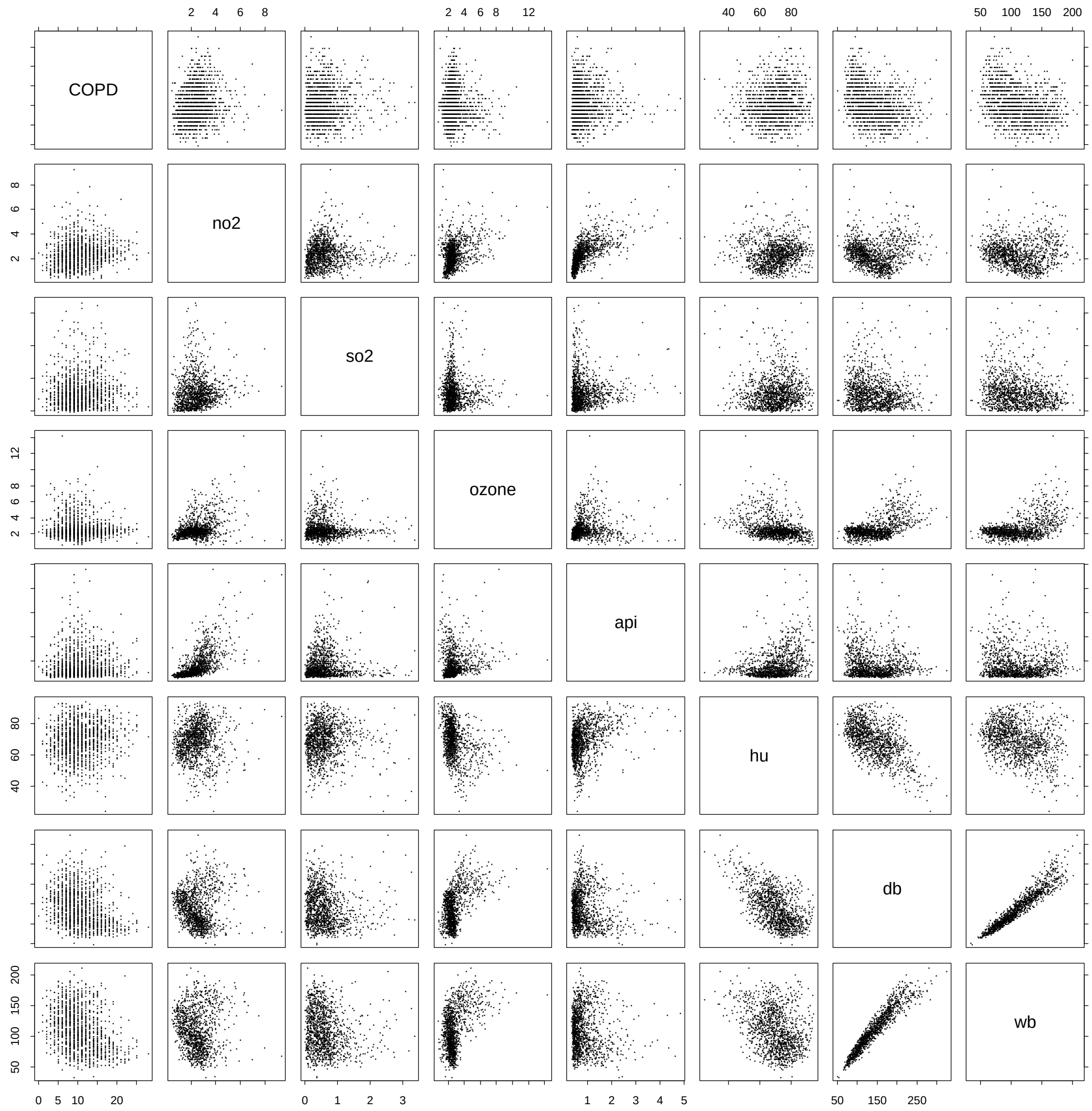

Figure 2: *Pairwise scatter plots for hospital admissions for COPD, pollutants and climatic data*

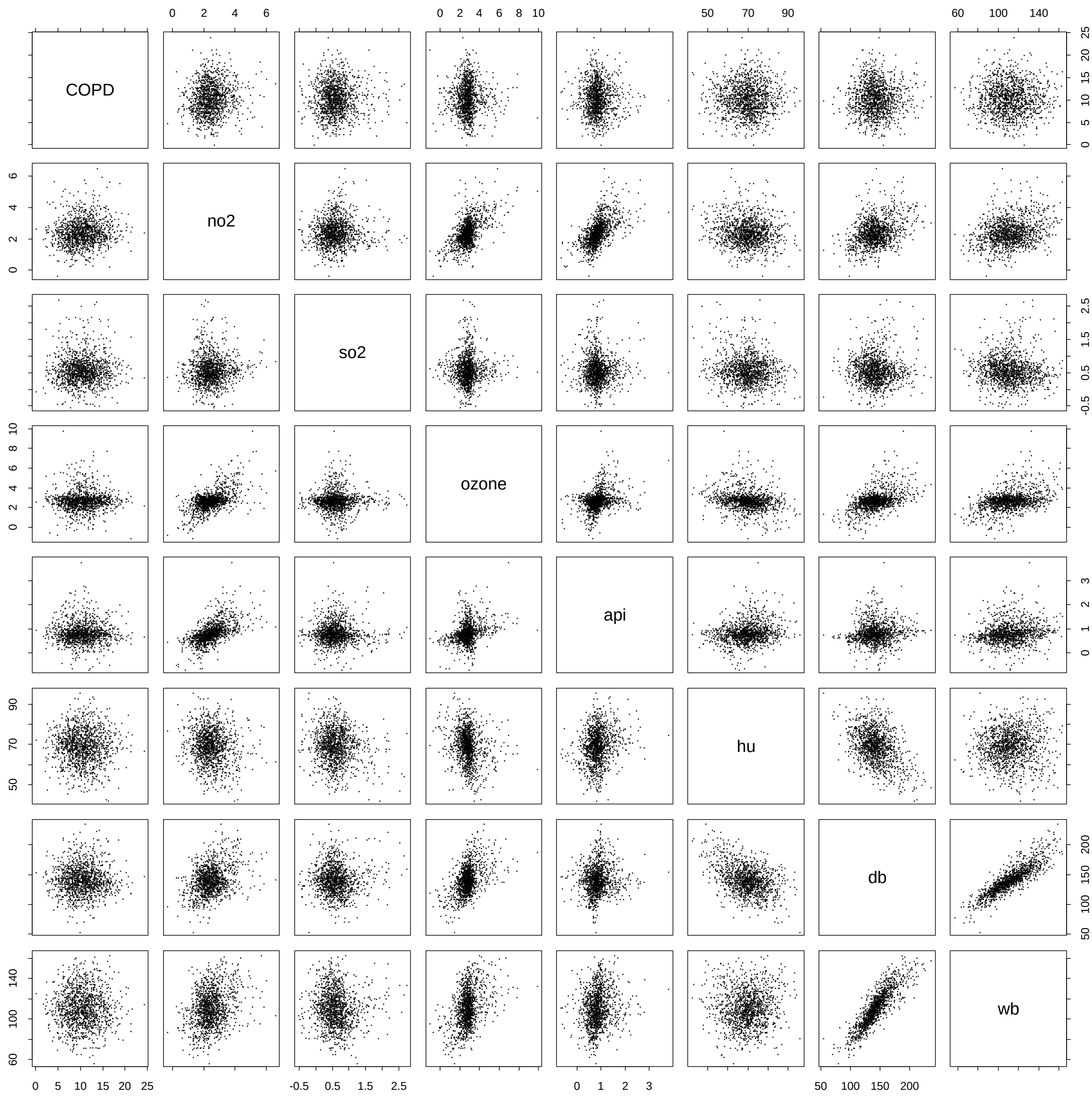

Figure 3: *Pairwise scatter plots for hospital admissions for COPD, pollutants and climatic data. All variables seasonally adjusted.*

out seasonal confounding. This will be done using Fourier series to model the seasonal variation.

Inspection of Figure 3 also shows colinearity between dew point and dry bulb temperature. If GLMs and GAMs are used in the presence of colinearity, the estimators will have large variances, and it will be difficult to interpret the effects[14]. To eliminate the effects of colinearity we will use only one of dew point or dry bulb temperature.

Generalized Linear Modelling approach

The Generalized Linear Model (GLM) is a natural extension of classical linear models. GLMs allow us to study the effects of covariates on a response variable Y_t that may have a non-Normal error distribution.

In classical linear models we can estimate the expected value of a response variable Y_t, conditional on a linear function of predictor variable, $X_{1,t}, X_{2,t}, \ldots, X_{r,t}$ as follows

$$\mathrm{E}(Y_t|X_t) = \beta_0 + \sum_{i=1}^{r} \beta_i X_{t,i} + Z_t \tag{1}$$

where $Z_t \stackrel{d}{=} NID(0, \sigma^2)$ and $t = 1, \ldots, n$.

For a GLM we estimate the expected value of the response variable, as a function of a linear combination of the predictors, i.e.,

$$\mathrm{E}(Y_t|X_t) = \eta \left(\beta_0 + \sum_{i=1}^{r} \beta_i X_{t,i} \right). \tag{2}$$

Here the function η^{-1} is called the link function.

In this paper the response variable Y_t denotes hospital admissions. Because these are count data, they have typically been modeled as a Poisson process[15] with a log link function so that

$$\eta(\mu) = \exp(\mu). \tag{3}$$

The variance of Y_t is a function of the mean response defined by

$$\mathrm{Var}(Y) = \phi\mu \tag{4}$$

where ϕ is the dispersion parameter.

For a Poisson distribution, $\phi = 1$. However, when modeling hospital admissions data we cannot assume that the dispersion parameter in (4) is equal to 1, since hospital admissions data are frequently overdispersed[1]. Rather, we use quasi-likelihood estimation[10].

Table 2 shows the results of a GLM analysis with COPD as the response variable and dummies for day of week, a quadratic time trend, and linear functional forms for the relationship between the response and the explanatory variables. Lags of up to 2 days were included in the analysis for each pollutant and climatic variable. We use Fourier series terms to model the seasonal pattern in the data. We included terms for $\cos(2\pi kt/365)$ and $\sin(2\pi kt/365)$ for $k = 1, 2, 3, 4$.

We used an efficient step-wise selection method in S-Plus to select regressors. Akaike's Information Criterion[16] was used to evaluate different models and the model with the smallest AIC was chosen as the final model.

Thus, with the GLM we have ignored the possible non-linear functional forms between the response and the explanatory variables (apart from time). This model results in a statistically significant effect ($P = 0.0785$) of same day nitrogen dioxide on COPD hospital admissions. An increase in daily maximum hourly level of nitrogen dioxide from the 10th to 90th percentile (Table 1) is statistically associated with an increase of 4.89% hospital admissions for COPD. A statistically significant lag 1 ($P = 0.0633$) and lag 2 ($P = 0.0295$) effect of particulates was also observed.

Table 2: *Regression coefficients, standard errors and P-values obtained by a GLM analysis*

Parameter	**Estimate**	**Standard Error**	**P-value**
Intercept	2.4638	0.0829	0.0000
$NO_{2,t}$	0.0191	0.0108	0.0783
api_{t-1}	0.0432	0.0232	0.0625
api_{t-2}	-0.0499	0.0227	0.0285
t	1.0699	0.3349	0.0014
t^2	2.5302	0.3617	0.0000
D_1	-0.0419	0.0171	0.0146
D_2	-0.0160	0.0101	0.1116
D_3	-0.0057	0.0071	0.4238
D_4	-0.0058	0.0056	0.2975
D_5	-0.0504	0.0052	0.0000
D_6	-0.0216	0.0042	0.0000
$\sin(2\pi t/365)$	0.1750	0.0152	0.0000
$\cos(2\pi t/365)$	0.1625	0.0175	0.0000
$\sin(4\pi kt/365)$	0.0455	0.0146	0.0018
$\cos(4\pi kt/365)$	0.0072	0.0145	0.6178
$\sin(6\pi kt/365)$	-0.0273	0.0145	0.0599
$\cos(6\pi kt/365)$	-0.0466	0.0142	0.0011
$\sin(8\pi kt/365)$	-0.0396	0.0141	0.0050
$\cos(8\pi kt/365)$	-0.0161	0.0139	0.2473
hu_t	-0.0028	0.0011	0.0110

D_i denotes day of week dummy variables (D_1 = Monday, D_2 = Tuesday, etc.), $NO_{2,t}$ nitrogen dioxide lags, api_{t-i} lagged particulates.

Residual Analysis

Graphical analyses of residuals are central to statistical model building. These plots indicate the effectiveness of the chosen statistical model. When modeling the relationship between hospital admissions for COPD with pollutants and climatic variables, a number of residual analysis techniques could be employed.

Plots of the residuals from the final model with each predictor included in the final model can identify incorrectly specified relationships between the predictors and the response variable; that is, whether non-linear relationships have been incorrectly modeled as linear.

Plots of the residuals from the final model with each of the excluded predictors from the final model will identify any remaining patterns that should be modeled.

Conditioning plots[17] of the residuals with any two predictors identify any remaining patterns and interactions between the two predictor variables.

An autocorrelation plot of the residuals will show any remaining autocorrelation that needs to be identified and modeled[18].

In classical linear regression, the residuals are normally distributed with equal variance. In Generalized Linear Modeling the two most commonly used types of residuals are the Pearson and Deviance residuals[11]. In non-normal situations where the response takes on a small number of distinct values (such as when the response is Poisson and has small mean), then the Deviance and Pearson residuals are far from normally distributed. Furthermore, residual plots against explanatory variables can be misleading.

For these reasons, Dunn & Smyth[12] propose randomized quantile residuals for non-normal response models. These residuals are computed by inverting the fitted distribution function at each value of the response variable and then finding the equivalent standard normal quantile. These residuals then become exactly standard normal, and provide more readily interpretable residual plots. Randomized quantile residuals will be used throughout this paper because of the discrete and overdispersed nature of the response.

Plots based on the quantile residuals are shown in Figure 4. The left plot shows a smoothing spline time trend fitted to the quantile residuals. The non-linear relationship demonstrates that the quadratic time trend used in the GLM has not adequately captured the non-linearity in the trend. The right plot shows the autocorrelation function for the quantile residuals. There is some significant autocorrelation in the residuals at lags 1, 3 and 7 which should be accounted for in the model. Unfortunately, allowing for autocorrelation in GLMs is not straightforward[35].

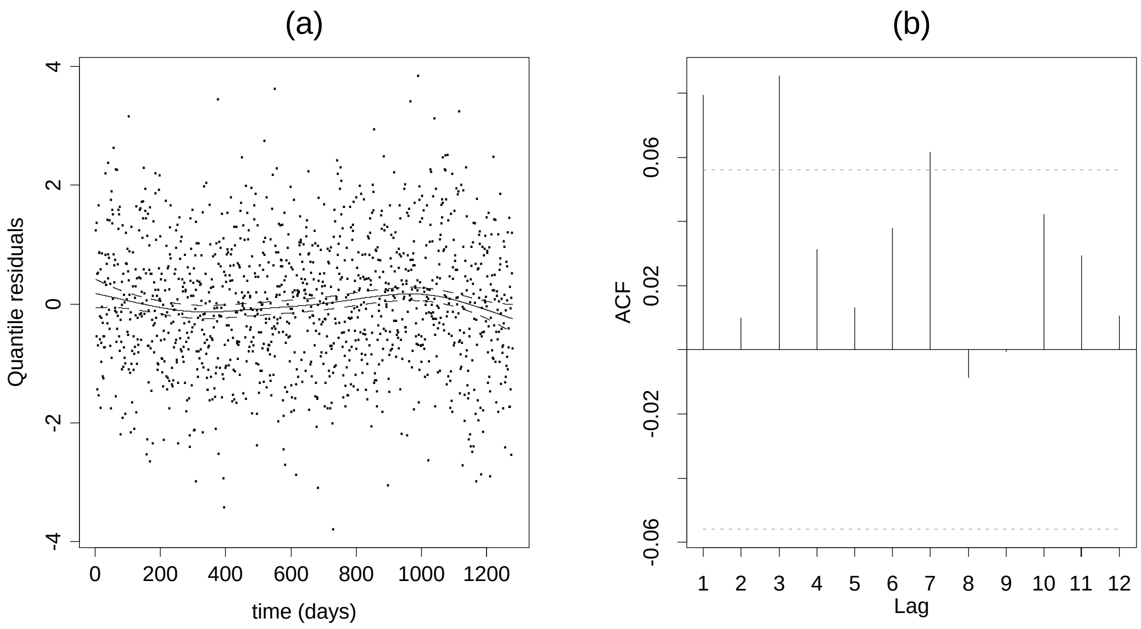

Figure 4: *(a) A smoothing spline fitted to the quantile residuals with time as a covariate. The dashed lines denote 95% pointwise confidence intervals around the line. There is a clear non-linear relationship demonstrating that a quadratic time trend has not captured the relationship adequately. (b) The autocorrelation function for the quantile residuals. There is some significant autocorrelation in the residuals which should be accounted for.*

Generalized Additive Modelling approach

Until recently many studies have hypothesized a linear function for the association between hospital admissions and pollutants or climatic variables, with no real empirical or biological knowledge. It is possible that the true association between the hospital admissions and pollution/climatic variables may be linear or non-linear or a combination of both as reported elsewhere[1,19]. Each explanatory variable may have a different functional form with hospital admissions.

A nonparametric alternative to the parametric GLM is the Generalized Additive Model (GAM). GAMs allow a nonparametric relationship between the response variable and each predictor using scatterplot smoothers[20].

As for a GLM, we use a link function of the expected value of the response variable, which is based on nonparametric functions of the predictors. That is,

$$\mathrm{E}(Y_t|X_t) = \eta\left(\beta_0 + \sum_{i=1}^{r} g_i(X_{t,i})\right). \tag{5}$$

The estimated nonparametric functions g_i are similar to coefficients in a linear model. A linear function for any g_i can be fitted to obtain semi-parametric models.

Cubic smoothing splines will be used to estimate the smooth relationship between the response variable and each of the predictors in this paper. Splines provide a much better solution than polynomials since polynomials fit the data globally while splines fit the data locally. For each variable, we set the degrees of freedom of the smoothing spline to be equal to four. This was felt to provide a reasonable balance between overfitting the data and allowing for sufficient curvature when it was required. (An alternative but more computationally intensive approach would be to use a bandwidth selection procedure such as Generalized Cross-Validation[21,22].)

As for the GLM, we used a step-wise selection method to choose the regressors, selecting the model with the smallest AIC.

To explore the effect of the pollutants we fitted two GAMs. The first model contains no explicit seasonal terms and includes dummies for day of week and a smooth term for the non-linear time trend. We obtained the following model using step-wise selection:

$$\begin{aligned}\mathrm{E}(Y_t|X_t) &= \exp[\beta_0 + g_1(\mathrm{db}_t) + g_2(\mathrm{ozone}_{t-2}) \\ &\quad + g_3(\mathrm{SO}_{2,t-2}) + g_4(\mathrm{db}_{t-2}) + \beta_5\mathrm{hu}_t + g_6(\mathrm{t}) \\ &\quad + \beta_7 D_1 + \beta_8 D_2 + \beta_9 D_3 + \beta_{10} D_4 + \beta_{11} D_5 + \beta_{12} D_6]\end{aligned} \tag{6}$$

where X_t contains all covariates and $Y_t|X_t$ is Pseudo-Poisson (i.e., Poisson with over-dispersion).

The second model uses the same covariates with the addition of Fourier series terms to

Table 3: *Regression coefficients of linear terms, standard errors and P-values obtained by the seasonal GAM analysis*

Parameter	**Estimate**	**Standard Error**	**P-value**
Intercept	2.0335	0.0912	0.0000
$NO_{2,t}$	0.0249	0.0102	0.0146
ozone_{t-2}	0.0208	0.0096	0.0296
api_{t-2}	-0.0363	0.0205	0.0777
hu_{t-2}	0.0020	0.0011	0.0791
D_1	-0.0426	0.0170	0.0124
D_2	-0.0147	0.0100	0.1402
D_3	-0.0051	0.0071	0.4735
D_4	-0.0067	0.0056	0.2251
D_5	-0.0505	0.0051	0.0000
D_6	-0.0221	0.0041	0.0000
$\sin(2\pi t/365)$	0.1797	0.0145	0.0000
$\cos(2\pi t/365)$	0.0370	0.0144	0.0105
$\sin(4\pi kt/365)$	-0.0206	0.0143	0.1506
$\cos(4\pi kt/365)$	-0.0389	0.0140	0.0057
$\sin(6\pi kt/365)$	0.1384	0.0177	0.0000
$\cos(6\pi kt/365)$	0.0044	0.0145	0.7599
$\sin(8\pi kt/365)$	-0.0511	0.0141	0.0003
$\cos(8\pi kt/365)$	-0.0152	0.0138	0.2703

model seasonality. The selected model was

$$\begin{aligned} \text{E}(Y_t|X_t) &= \exp[\beta_0 + \beta_1 \text{NO}_{2,t} + \beta_2 \text{ozone}_{t-2} + g_3(\text{SO}_{2,t-2}) + \beta_4 \text{api}_{t-2} + \\ &\quad \beta_5 \text{hu}_{t-2} + g_6(\text{t}) + \beta_7 D_1 + \beta_8 D_2 + \beta_9 D_3 + \beta_{10} D_4 + \beta_{11} D_5 + \\ &\quad \beta_{12} D_6 + \sum_{k=1}^{4} [\gamma_k \cos(2\pi kt/365) + \theta_k \sin(2\pi kt/365)]] \end{aligned} \tag{7}$$

where $Y_t|X_t$ is Pseudo-Poisson. Table 3 shows the coefficients for the linear terms in model (7) which can be interpreted in terms of increased daily hospital admissions. For example, an increase in daily maximum hourly concentration of nitrogen dioxide from the 10th to 90th percentile results in an increase in daily COPD hospital admissions of $e^{0.0249(3.73-1.23)} - 1 = 6.42\%$ (95%CI: 1.13 to 11.99, $P = 0.0145$).

If we compare the results from the GLM analysis in Table 2 and the results of the GAM analysis in equation (7) we notice some striking features.

Figure 5 shows a comparison of the functional form for the time trend, using both the GLM and the GAM. Clearly the assumption that the non-linear functional form between time and COPD hospital admissions is quadratic is incorrect. Making no assumptions for the functional form of the time trend in the right figure of Figure 5 has allowed us to capture the non-linear time trend appropriately.

A comparison of equation (6) and equation (7) also shows some interesting results. Irrespective of the seasonal adjustment, Sulfur dioxide lagged 2 days is a statistically significant non-linear pollutant, shown in Figure 6. This significant effect of Sulfur dioxide on COPD hospital admissions is indicative of an independent effect of seasonality. However, this is not the case with the rest of the pollutants. It seems that the effects of Nitrogen dioxide, Particulates and ozone are sufficiently small that they seem not to be significant when there is no seasonal adjustment, in equation (6). In equation (6) dew point temperature (wb) and dry bulbs temperature (db) are acting as seasonal proxies because they have no further effect after the allowance for seasonality in equation (7).

Ozone is a seasonal variable. Without seasonal adjustment in equation (6), the observed effect is a combination of a real effect and an effect due to its confounding with seasonal-

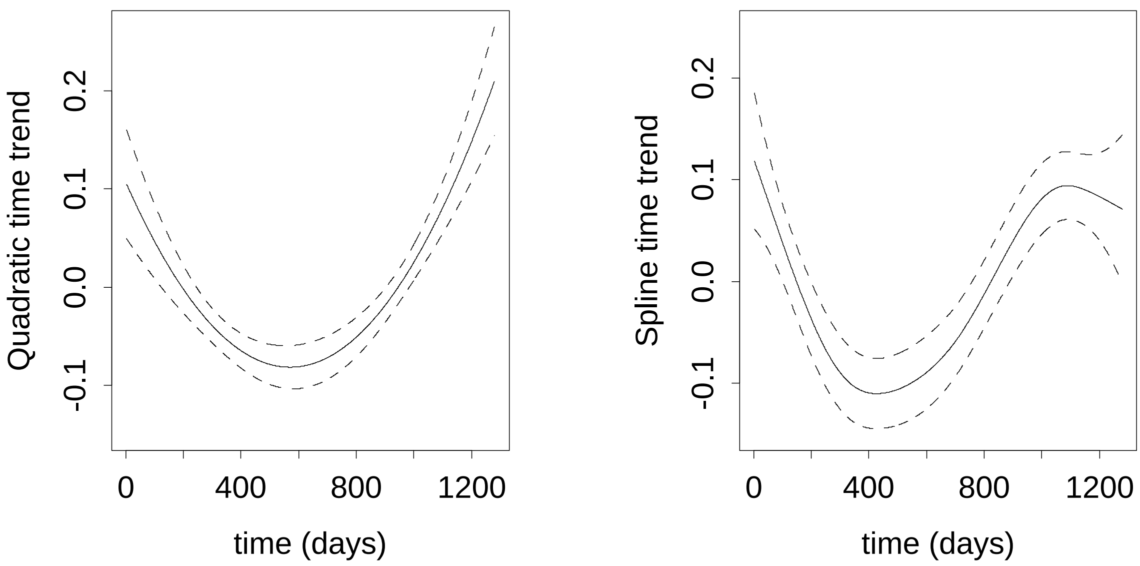

Figure 5: *The left figure shows the nonlinear fit for time using a generalized linear model. The right figure plots the spline fit for time using a generalized additive model.*

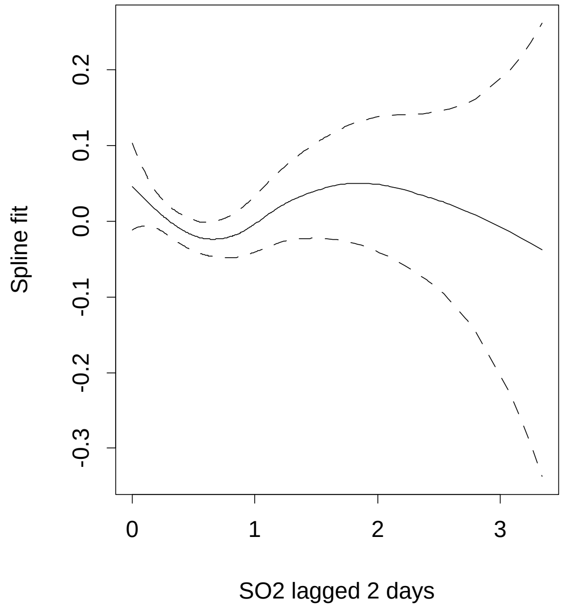

Figure 6: *A smooth plot of Sulfur dioxide lagged 2 days from the fit of a generalized additive model in equation (7).*

ity. In equation (7) we observe the true linear effect of ozone after seasonal adjustment.

From this analysis, it is clear that each of the pollutants has an important effect on hospital admissions for COPD. However, it is difficult to separate these effects clearly because they will have some confounding with each other.

Also, it seems that confounding with seasonality is not really an important factor here, except in understanding the effect of ozone. It seems that seasonality masks the effect of some of the covariates and this is an important finding in the analysis and modeling of morbidity/mortality data.

Serial correlation in the residuals is also an important methodological issue in morbidity/mortality research. Figure 7 shows the autocorrelation plots of the quantile residuals from the non seasonal GAM in equation (6) and the seasonal GAM in equation (7). After correctly specifying the non-linear functional form between the response and the covariates and after allowing for seasonality, it is clear that there is little significant pattern remaining in the autocorrelation plot of the quantile residuals.

Discussion

Pope & Schwartz[15] state that in time series studies of pulmonary health, the data cannot be fully utilized without adequately dealing with non-linearity in the covariates and time. The present study has attempted to adequately address these issues with appropriate statistical methodology.

European studies[2,23,24] and various American studies[25,26] have shown a consistent statistical effect of ozone on hospital admissions for COPD. We have shown a statistically significant lag 2 linear effect of ozone on COPD admissions which is consistent with results from Schouten et al.[24], who have also reported significant lag 2 results for ozone on COPD hospital admissions in Rotterdam (Netherlands). Others have shown an instantaneous and lag 1 effect of ozone on COPD admissions.

Some European studies have also shown a statistically significant effect of nitrogen dioxide on hospital admissions. The present study has clearly shown a same day statistically significant linear effect of nitrogen dioxide on hospital admissions for COPD. Although

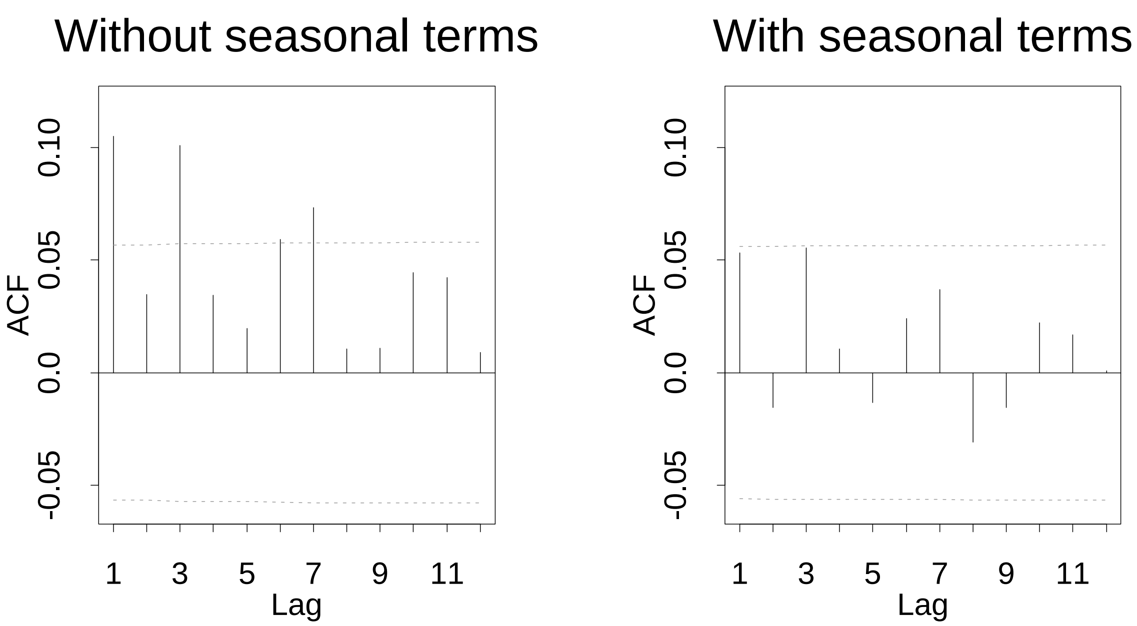

Figure 7: *The left figure depicts any remaining autocorrelation after fitting the GAM in equation (6). The right figure shows any remaining autocorrelation after fitting the GAM in equation (7).*

the level of nitrogen dioxide has remained consistently below the set guidelines for Melbourne it is clear from our study that current levels of nitrogen dioxide are likely to be harmful to people with COPD.

Lagged 2 days sulfur dioxide has shown a statistically significant non-linear effect on hospital admissions for COPD even after the control for the confounding effect of seasonality and climatic variables.

Our results may not be directly comparable with those obtained in North America and Europe due to the differences in the chemical nature and in the composition of the whole pollution mixture[2]. In addition, inadequate control for seasonal and confounding variables and the possibility of overfiltering may also cause discrepancies in the results of the effects of the pollutants[24] from those studies.

In Sydney, Australia, Morgan et al.[4] have shown that nitrogen dioxide and particulates are associated with an increase in hospital admissions for COPD in the elderly. The present study has not shown a statistically significant effect of same day or lag 1 day particulates. In our data, nephelometery light-scattering values were used as a proxy for fine particulate mass concentration during 1989-1992. This may have reduced the strength of the effect of particulates on hospital admissions for COPD.

Generalized Additive Models have only recently been used in the analysis of admissions/mortality data[27,28,29]. However, most, if not all, European, American and Australian studies have used parametric statistical methods to model the non-linear functional form of the response variable and covariates. Therefore, it is difficult to compare the nonparametric results from our study.

However, the advantage of the nonparametric GAM compared to the parametric GLM is clearly demonstrated in this paper. In particular, the non-linear time trend under a quadratic assumption in the GLM and the smooth function in the GAM has shown that the GAM clearly has captured this non-linear pattern. Also, the GAM has captured the nonlinear smooth function for Sulfur dioxide (lagged 2 days), which was incorrectly specified as a linear function in the GLM analysis. These findings demonstrate the superiority of GAMs compared to GLMs.

Residual analysis is an integral part in the diagnostic checking stage of statistical modeling. We have used graphical analysis of residuals to identify incorrectly specified models, deviations from normality, and serial correlation of the residuals.

In morbidity/mortality research, it is difficult to determine the type of residuals used. Many studies haven't reported the type of residuals used[30,31]. The standardized and deviance residuals have been reported by Wordley et al.[32]; raw residuals were reported by Simpson et al.[33] and Schwartz[34]. However, many studies do not include a residual analysis at all. In this paper we have proposed the random quantile residuals, these residuals should be used in over-dispersed and non-normal situations when the deviance and Pearson residuals can be non-normal.

Autocorrelation is an important methodological issues in morbidity and mortality research. Statistical methodology to deal with autocorrelation in GLM and GAMs is still under development[35]. In this paper we have shown that allowing non-linear relationships and seasonal terms can remove the serial correlation pattern inherent in morbidity/mortality data and thereby greatly simplify the analysis.

References

1 Schwartz, J., Spix, C., Touloumi, G., Bacarova, L., Barumamdzadeh, T., Tertre, A le., Piekarski, T., Ponce de Leon, A., Ponka, A., Rossi, G., Saez, M., & Schouten, J.P. (1996) Methodological issues in studies of air pollution and daily counts of deaths or hospital admissions, *Journal of Epidemiology & Community Health*, **50** (suppl 1), s3–s11.

2 Anderson, H.R., Spix, C., Medina, S., Schouten, J., Castellsague, J., Rossi, G., Zmirou, D., Touloumi, G., Wojtyniak, B., Ponka, A.,

BACHAROVA, L., SCHWARTZ, J., & KATSOUYANNI, K. (1997) Air pollution and daily admissions for chronic obstructive pulmonary disease in 6 European cities: results from the APHEA project, *European Respiratory Journal*, **10**, 1064–1071.

3 SUNYER, J., SAEZ, M., MURILLO, C., CASTELLSAGUE, J., MARTINEZ, F., & ANTO, M.J. (1993) Air pollution and emergency room admissions for Chronic Obstructive Pulmonary Disease: a 5-year study, *American Journal of Epidemiology*, **137**, 701–705.

4 MORGAN, G., CORBETT, S., & WLODARCYZK, J. (1998) Air pollution and hospital admissions in Sydney, Australia, 1990-1994, *American Journal of Public Health*, **88**, 1761–1766.

5 SPIX, C., ANDERSON, H.R., SCHWARTZ, J., VIGOTTI, M.A., LETERTRE, A., VONK, J., TOULOUMI, G., BALDUCCI, F., PIEKARSKI, T., BACHAROVA, L., TOBIAS, A., PONKA, A., & KATSOUYANNI, K. (1998) Short-term effects of air pollution on hospital admissions of respiratory diseases in Europe: a quantitative summary of APHEA study results, *Archives of Environmental Health*, **53**, 54–64.

6 SUNYER, J., ANTO, J., MURILLO, C., & SAEZ, M. (1991) Effects of urban air pollution on emergency room admissions for chronic obstructive pulmonary disease, *American Journal of Epidemiology*, **134**, 277–288.

7 LIPPMANN, M (1989) Health effects of ozone. a critical review, *Journal of the Air Pollution Control Association*, **39**, 672–695.

8 EPA (2000) "EPA Air Quality Information", Environment Protection Authority, Victoria, Australia. http://www.epa.vic.gov.au/aq/info/ Accessed 8 May 2000.

9 GOLDSMITH, J., FRIGER, M., & ABRAMSON, M. (1996) Associations between health and air pollution in time-series analyses, *Arch. Environ. Health*, **51**, 359–367.

10 MCCULLAGH, P. & NELDER J.A. (1989) *Generalized Linear Models*, London: Chapman and Hall.

11 HASTIE, T. & TIBSHIRANI, R.J. (1990) *Generalized Additive Models*, London: Chapman and Hall.

12 DUNN, P., & SMYTH, G. (1996) Randomized quantile residuals, *Journal of Computational and Graphical Statistics*, **5**, 236–244.

13 MAKRIDAKIS, S., WHEELWRIGHT, S.C., & HYNDMAN, R.J. (1998) *Forecasting: methods and applications*, John Wiley & Sons: New York.

14 PITARD, A & VIEL, J (1997) Some methods to address collinearity among pollutants in epidemiological time series, *Statistics in Medicine*, **16**, 527–544.

15 POPE, A & SCHWARTZ, J (1996) Time series for the analysis of pulmonary health data, *American Journal of Respiratory Care & Medicine*, **154**, s229–s233.

16 AKAIKE, H. (1973) Information theory and an extension of the maximum likelihood principle, *2nd International Symposium on Information Theory*, B.N. Petrov & F. Csaki (eds), Adademiai Kidao, Budapest, 267–281.

17 CLEVELAND, W.S. (1993) *Visualizing data*, Hobart Press.

18 BOX, G.E.P. & JENKINS, G.M. (1979) *Time series analysis: forecasting and control*, San Francisco: Holden-Day.

19 THURSTON, G & KINNEY, P. (1995) Air pollution epidemiology: considerations in time series modeling, *Inhalation Toxicology*, **7**, 71–83.

20 CHAMBERS, J & HASTIE, T. (1992) *Statistical models in S*, California: Wadsworth & Brooks/Cole Advanced Books & Software.

21 WAHBA, G. (1977) A survey of some smoothing problems and the method of generalized cross-validation for solving them, In *Applications of Statistics*, ed., P.R. Krishnaiah, 507–523, North Holland: Amsterdam.

22 CRAVEN, P., & WAHBA, G. (1979) Smoothing noisy data with spline functions: estimating the degree of smoothing by the method of generalized cross-validation, *Numerical mathematics*, **31**, 377–403.

23 BURNETTE, R., DALES, R., KREWSKI, D., VINCENT, R., DANN, T., & BROOK, J.R. (1995) Associations between ambient particulates sulfate and admissions to Ontario

hospitals for cardiac and respiratory disease, *American Journal of Epidemiology*, **142**, 15–22.

24 SCHOUTEN, J. P., VONK, J.M., & GRAAF, A DE. (1996) Short term effects of air pollution on emergency hospital admissions for respiratory disease: results of the APHEA project in two major cities in The Netherlands, 1977-89, *Journal of Epidemiology and Community Health*, **50** (Suppl 1, S22–S29).

25 SCHWARTZ, J (1994) Air pollution and hospital admissions for the elderly in Birmingham, Alabama, *American Journal of Epidemiology*, **139**, 589–598.

26 SCHWARTZ, J (1994) Air pollution and hospital admissions for the elderly in Detroit, Michigan, *American Journal of Respiratory Critical Care Medicine*, **150**, 648–655.

27 SCHWARTZ, J (1994) Nonparametric smoothing in the analysis of air pollution and respiratory illness, *The Canadian Journal of Statistics*, **22**, 471–487.

28 SCHWARTZ, J (1994) Air pollution and daily mortality: a review and meta analysis, *Environmental Research*, **64**, 36–52.

29 HOEK, G., SCHWARTZ, J., GROOT, B., & EILERS, P. (1997) Effects of ambient particulate matter and ozone on daily mortality in Rotterdam, The Netherlands, *Archives of Environmental Health*, **52**, 455–463.

30 ABURTO-BORJA, V., LOOMIS, D.P., BANGDIWALA, S.I., SHY, C.M., & RASCON-PACHECO, R.A. (1997) Ozone, suspended particulates, and daily mortality in Mexico City, *American Journal of Epidemiology*, **145**, 258–268.

31 SCHWARTZ, J & MORRIS, R (1995) Air pollution and hospital admissions for Cardiovascular disease in Detroit, Michigan, *American Journal of Epidemiology*, **142**, 23–35.

32 WORDLEY, J., WALTERS, S., & AYRES, J.G. (1997) Short term variations in hospital admissions and mortality and particulate air pollution, *Occupational and Environmental Medicine*, **54**, 108–116.

33 SIMPSON, R., WILLIAMS, G., PTEROESCHEVSKY, A., MORGAN, G., & RUTHER-

FORD, S. (1997) Associations between outdoor air pollution and daily mortality in Brisbane, Australia, *Archives of Environmental Health*, **52**, 442–454.

34 SCHWARTZ, J (1995) Short term fluctuations in air pollution and hospital admissions of the elderly for respiratory disease, *Thorax*, **50**, 531–538.

35 BRUMBACK, B.A., RYAN, L.M., SCHWARTZ, J.D., NEAS, L.M., STARK, P.C., & BURGE, H.A. (2000) Transitional regression models, with application to environmental time series, *J. American Statistical Asssociation*, **95**, 16–27.