

Statistical methodological aspects of modelling relationships
between air pollution, temperature and health



Gavin Shaddick

Imperial College of Medicine at St Mary's

Submitted for the degree of Doctor of Philosophy
and the Diploma of Membership of Imperial College

Abstract

Recently, the short-term, or *acute*, health effects of air pollution, have received considerable attention and positive associations between daily measures of air pollution and adverse health outcomes have been reported in a large number of studies. In many of these studies, temperature has been treated as little more than a nuisance variable, both through adjusting for day-to-day variability and also as a result of controlling for seasonality. However, the effect of temperature is potentially an important determinant of health in its own right. Meanwhile, less attention has been given to investigating the *chronic* effects of air pollution, i.e. the association between health outcomes and long-term exposures to air pollution, possibly over several years. This thesis considers a number of methodological issues encountered in modelling the acute effects on health of both air pollution and temperature, and the chronic effects of air pollution. Methods are proposed for the estimation of missing values and the spatial-temporal modelling of routinely measured pollution data, and for dealing with the effects of measurement error, sensitivity to unmeasured variables, the use of non-independent outcomes over time, and the problems of model selection and comparison. Statistical methodologies that acknowledge these issues are developed and applied to a series of substantive epidemiological enquiries, including the long term effects of air pollution, the acute effects of cold temperatures, and producing stable estimates of pollution levels for use in studies of the acute effects of air pollution. Although the modelling approaches presented here were developed to facilitate the analyses of specific enquires, the issues encountered are not exclusive to the applications under discussion, but are general to many epidemiological analyses.

Acknowledgements

I would like to thank my supervisors, Jon Wakefield and Paul Elliott. In particular, I would like to thank Jon for his guidance and support throughout our time in the Department of Epidemiology and Public Health and beyond, and not least for encouraging me to return and embark on this piece of work. I would like to thank Paul for finding a way for me to complete this thesis on a part-time basis. I would like to thank colleagues in the department for their assistance, both with this thesis and in other matters. I would also like to thank members of the department for help in obtaining the data required for this thesis. Finally, I would like to thank Tracey for her support, understanding and for making it possible for me to work on this thesis at the expense of many other things.

Declaration

The work in this thesis is my own. The health data used in Chapters 2 and 3, and data on socio-economic deprivation were extracted from the national database held within the Small Area Health Statistics Unit, which is located in the Department of Epidemiology and Public Health, Imperial College Faculty of Medicine at St. Mary's. The air pollution exposure data used in Chapter 3 were supplied by colleagues from the same department. The temperature data used in Chapter 2 were supplied by the Meteorological Office and the pollution data used in Chapter 5 were retrieved from the UK National Air Quality Information Archive.

Parts of the methodology and study design presented in Chapter 3 were presented by the author at the twelfth annual conference of the International Society of Environmental Epidemiology, Buffalo, USA. 2000. *Briggs, D, Elliott, P and Shaddick, G. Changing geographic relationships between air pollution and cardio-respiratory mortality in Great Britain, 1962-1996.*

The methodology presented in Chapter 4 was presented by the author at the tenth annual conference of the International Society of Environmental Epidemiology, Boston, USA. 1998. *Shaddick G, Wakefield J., Elliott P. Methodological aspects of modelling spatial and temporal relationships between air pollution and respiratory health.*

Parts of the work presented in Chapter 5 have been published in the Journal of the Royal Statistical Society, Series C (Applied Statistics, Vol 51: Issue 3. 2002) with the title *Modelling multiple pollutants at multiple sites*, jointly with Jon Wakefield. The methodology from this chapter was also presented at two international conferences. The first was the twelfth annual conference of the International Society of Environmental Epidemiology, Buffalo, USA. 2000. *Shaddick G and Wakefield, J. Modelling levels of pollution for use in time series studies examining the relationship between air pollution and health.* The second was at the conference of the International Environmentalmetrics Society, Portland, USA. 2001 *Shaddick G and Wakefield, J. Modelling multiple pollutants at multiple sites.*

Contents

1	Introduction	1
2	Statistical review	5
2.1	Likelihood approach	6
2.1.1	Quasi-likelihood	7
2.1.2	Likelihood ratio tests	7
2.2	Bayesian approach	8
2.2.1	Introduction to Bayesian inference	8
2.2.2	Bayesian point estimates	9
2.2.3	Computation	10
2.3	Regression modelling	11
2.3.1	Generalised linear models	12
2.3.2	Generalised additive models	14
2.3.3	Hierarchical models	15
2.3.4	Generalised estimating equations	19
2.4	Model selection and comparison	21
2.4.1	Effect of selection on properties of estimators	21
2.4.2	Selection procedures	24
2.4.3	Bayes factors	28
2.4.4	Bayesian model averaging	30
2.5	Temporal modelling	35
2.5.1	Stationary processes	36
2.5.2	Trend and seasonality	37

2.5.3	State space models	39
2.6	Spatial modelling	41
2.6.1	Stationary random fields	41
2.6.2	Prediction at unsampled locations	45
2.6.3	Non-stationary random fields	46
2.7	Spatial-temporal modelling	49
2.7.1	Separable models	49
2.7.2	Non-Separable models	51
3	Modelling the association between mortality and long term exposure to air pollution	54
3.1	Introduction	54
3.1.1	Epidemiology of chronic health effects	55
3.2	Data description	57
3.2.1	Health, socio-economic and demographic data	58
3.2.2	Air pollution data	60
3.3	Statistical Methodology	61
3.3.1	Modelling the cumulative effects of exposure	61
3.3.2	Model framework	66
3.4	Health period specific analyses	67
3.4.1	Models	68
3.4.2	Inference	70
3.4.3	Results	70
3.5	Combined health period analysis	77
3.5.1	Models	77
3.5.2	Inference	79
3.5.3	Results	79
4	Modelling the effect of cold winter temperatures on mortality	85
4.1	Introduction	85
4.1.1	Epidemiology of effects of cold temperatures on mortality	86

4.2	Data Description	88
4.3	Exploratory analysis	91
4.3.1	Models	91
4.3.2	Inference	93
4.3.3	Results	93
4.4	Full Analysis	98
4.4.1	Models	98
4.4.2	Bayesian model averaging	99
4.4.3	Inference	100
4.4.4	Results	101
5	Modelling Daily Multivariate Pollutant Data at Multiple Sites	109
5.1	Introduction	109
5.1.1	Epidemiology of acute health effects	111
5.2	Data description	117
5.3	Single pollutant, single monitoring site	123
5.3.1	Model	123
5.3.2	Inference	124
5.3.3	Results	125
5.4	Single pollutant, multiple monitoring sites	129
5.4.1	Model	129
5.4.2	Inference	130
5.4.3	Results	131
5.5	Multiple pollutants, single monitoring site	133
5.5.1	Model	133
5.5.2	Results	134
5.6	Multiple pollutants, multiple monitoring sites	138
5.6.1	The general model	138
5.6.2	Results	140

6	Discussion	144
6.1	Modelling the association between mortality and long term exposure to air pollution	145
6.2	Modelling the effect of cold winter temperatures on mortality	147
6.3	Modelling daily multivariate pollutant data at multiple sites	148
6.4	Summary of the epidemiological findings in the case studies	149
6.5	Key statistical issues: Over-dispersion	149
6.6	Key statistical issues: Model selection	150
6.7	Further statistical issues and developments	152
6.8	Conclusion	154

List of Tables

- 2.1 Possible design matrices and corresponding regression coefficients when using combinations of two covariates 23
- 2.2 Possible models and posterior model probabilities using combinations of two covariates, with $\epsilon_i(i = 1, 2, 3, 4) \sim \text{i.i.d } N(0, \sigma_i^2)$ 30
- 3.1 Summary of respiratory mortality at ward level in four health periods, (1981-84, 1985-88, 1989-92, 1993-96) at ward level. Means, standard deviations, medians, ranges and interquartile ranges are given for observed (Obs) and age-sex adjusted expected numbers of cases (Exp), together with measures of relative risk (O/E ratio) and the number of wards (N). 60
- 3.2 Summary of SO₂ measurements at ward level from the seven pollution periods (1966-69, 1970-73, 1974-77, 1978-81, 1982-85, 1986-89, 1990-93): means, standard deviations, medians, ranges, interquartile ranges and number of wards (N). 63
- 3.3 Correlations between SO₂ measurements in the seven exposure periods 65
- 3.4 Study design: Four health periods (with number of wards, n_t) with associated pollution periods (p) producing lagged exposures (l). Lagged exposures are 0-4, 5-8, 9-12 and 13-16 years for $l = 1, \dots, 4$ respectively. 67
- 3.5 Relative risks, with 95% confidence / credible intervals for increase of 10 ppb of SO₂ over the previous four years, with and without adjustment for deprivation. Results are for four models; likelihood based Poisson, Bayesian Poisson, Quasi-likelihood (with dispersion parameter) and Bayesian Poisson with log-normal (LN) random effects (with variance of random effects.) 71
- 3.6 Relative risks, with 95% confidence / credible intervals for increase of 10 ppb of SO₂ over the previous sixteen years, with and without adjustment for deprivation. Results are for four models; likelihood based Poisson, Bayesian Poisson, Quasi-likelihood (with dispersion parameter) and Bayesian Poisson with log-normal (LN) random effects (with variance of random effects.) 73

3.7	Relative risks, with 95% confidence/credible intervals for increase of 10 ppb of SO ₂ for cumulative effect of exposure over the previous 0-4, 5-8, 9-12 and 13-16 years, with adjustment for deprivation. Results are for four models; likelihood based Poisson, Bayesian Poisson, Quasi-likelihood (with dispersion parameter) and Bayesian Poisson with log-normal (LN) random effects (with variance of random effects).	74
3.8	Deviances and degrees of freedom for quasi-likelihood models relating various lagged exposures (previous 4 years, 16 years (total) and for the cumulative effects of 0-4, 5-8, 9-12 and 13-16 years (incremental)) of SO ₂ . Changes in deviance from the null model are given, with associated p-values. Results are given with and without adjustment for deprivation, after allowance for over-dispersion.	76
3.9	Relative risks, with 95% credible intervals for increase of 10 ppb of SO ₂ for effects of exposure over the previous four years, with and without adjustment for deprivation. Variances of random effects are given from Bayesian hierarchical models. HP denotes health period (HP1=1981-84, HP2=1985-88, HP3=1989-92, HP4=1993-96).	80
3.10	Relative risks, with 95% credible intervals for increase of 10 ppb of SO ₂ for effects of exposure over the previous sixteen years, with and without adjustment for deprivation. Variance of random effects are given from Bayesian hierarchical models. HP denotes health period (HP1=1981-84, HP2=1985-88, HP3=1989-92, HP4=1993-96).	81
3.11	Relative risks, with 95% credible intervals, for increase of 10 ppb of SO ₂ with variance of random effects, for simplified models of the effects of exposure over the previous 0-4, 5-8, 9-12 and 13-16 years, with adjustment for deprivation. The models are (1) without different effects of health period, (2) without different effects of lagged exposures and (3) no effects of different health periods or lags. HP denotes health period (HP1=1981-84, HP2=1985-88, HP3=1989-92, HP4=1993-96).	82
4.1	Summaries of all cause mortality and influenza deaths, by year, London, winters (November-February), 1981-97. Yearly totals, and means of daily counts are shown together with standard deviations.	90
4.2	Summary of all cause mortality and minimum temperatures, by month, season and over the entire study period (1981-97). Means of daily mean values are shown together with standard deviations.	91
4.3	Number of consecutive days at given temperature or less, with mean number of daily deaths during winters, London, 1981-97.	92

4.4	Relative risks with 95% confidence intervals for decrease of 1 ⁰ C in lagged temperature. Left hand results are for a non-threshold model, right hand side for a model with a change point at 10 ⁰ C. The top half of the table gives results for models with lags fitted separately, the bottom when fitted together in a single model.	98
4.5	Relative risks with 95% confidence intervals for decrease of 1 ⁰ C in lagged temperature. Results are given for Poisson and quasi-likelihood models with individual lagged effects fitted in separate models, together with deviances, degrees of freedom and changes in deviances (adjusted for over-dispersion). Dispersion parameters from the quasi-likelihood models are also given (disp). Changes in deviances are for cumulative addition of yearly term and daily influenza deaths to null model, and then for each lagged value fitted separately to model containing year and influenza terms.	101
4.6	Relative risks with 95% confidence intervals for decrease of 1 ⁰ C in lagged temperature. Results are given for Poisson and quasi-likelihood models with individual lagged effects fitted together in a single model. Deviances and degrees of freedom are given for series of models containing given lagged value and all previous lags. Changes in deviances (adjusted for over-dispersion) are given for cumulative addition of yearly term and daily influenza deaths to null model, and then between models differing sequentially by a single lagged value. Dispersion parameters from the quasi-likelihood models are also given (disp).	102
4.7	Results from BMA for models with non-zero posterior model probabilities (pmp) when modelling the effect of lagged values of temperature. Coefficients are given from each of the models, together with support for including each variable in the model. All models include yearly term and daily influenza deaths.	103
4.8	Relative risks corresponding to a unit change (drop of 1 ⁰ C) for same day and lagged temperatures (up to 7 days), calculated using Bayesian Model Averaging.	104
4.9	Results from using Bayesian model averaging for prediction using a range of temperature scenarios. Expected daily deaths and corresponding standard deviations are given for a range of low temperatures (5 ⁰ C to -2 ⁰ C) for periods of up to eight days.	105
5.1	Summary of pollutants measured, and periods of operation, at eight sites in London, 1994–97. The total number of days of operation are given for each pollutant at each site together with the percentage of missing observations. The units are μgm^{-3} for PM ₁₀ , parts per billion for SO ₂ and NO and parts per million for CO.	118

5.2	Correlation/covariance matrix for (logged) pollutants and temperature, Bloomsbury, 1994–97. The variances lie on the diagonal, with covariances above and correlations below.	120
5.3	Correlation/covariance matrix for (logged) values of PM_{10} , measured at eight sites, 1994–97. The variances lie on the diagonal, with covariances above and correlations below.	122
5.4	Observation and system variances, with coefficients for temperature, for two pollutants, PM_{10} and SO_2 , modelled individually at two sites, Bloomsbury and Bexley. The three sets of results show the effect of adding linear and quadratic effects of temperature to the model.	126
5.5	Observation and system variances, with coefficients for temperature, for PM_{10} modelled jointly at eight sites, but independently of measurements of other pollutants. Site effects, and variances are presented for both independent and multivariate spatial models. The labelling is: 1 = Bexley, 2 = Bloomsbury, 3 = Brent, 4 = Eltham, 5 = Harringey, 6 = Hillingdon, 7 = North Kensington, 8 = Sutton.	132
5.6	Observation and system variances, with coefficients for temperature, for four pollutants modelled using the multivariate pollution model. Results are from the Bloomsbury site, modelled independently of measurements at other sites.	135
5.7	Posterior medians of the Σ_p matrix between the underlying levels of four pollutants modelled using the multivariate pollution model at the Bloomsbury site, modelled independently of measurements at other sites. The variances lie on the diagonal, with covariances above and correlations below.	135
5.8	Posterior medians from the general model with dependencies across time and pollutants, with pollutant specific spatial effects. Pollutant specific observation and system variances, together with coefficients for temperature, are presented together with site effects and variances, for the multivariate spatial components. The labelling is: 1 = Bexley, 2 = Bloomsbury, 3 = Brent, 4 = Eltham, 5 = Harringey, 6 = Hillingdon, 7 = North Kensington, 8 = Sutton.	141
5.9	Posterior medians of the Σ_p matrix between the underlying levels of four pollutants modelled using the multivariate pollution model with multivariate spatial effects for all eight monitoring sites. The variances lie on the diagonal, with covariances above and correlations below.	142

List of Figures

3.1	Relative risk (adjusted for age and sex) of all cause, respiratory and cardiovascular mortality (for all ages) against measure of socio-economic deprivation (Carstairs' score), for four health periods.	59
3.2	Histograms of SO ₂ exposure in seven exposure periods (1966-69, 1970-73, 1974-77, 1978-81, 1982-85, 1986-89, 1990-93).	62
3.3	Correlations between SO ₂ measurements in the seven exposure periods (1966-69, 1970-73, 1974-77, 1978-81, 1982-85, 1986-89, 1990-93).	63
3.4	Relative risk of respiratory mortality (adjusted for age, sex and deprivation) , for all ages, against lagged exposure of SO ₂ . Mortality data is for first health period (1981-85), exposure data for lags of 0-4, 5-8, 9-12 and 13-16 years). The line joins median relative risks of 20ppb wide categories.	64
3.5	Associations between ward level (log-normal) random effects from model including SO ₂ exposure in previous four years and deprivation. The four panels refer to data from four health periods (1981-84, 1985-88, 1989-92, 1993-96).	75
3.6	Associations between ward level random effects using Bayesian hierarchical model with exposure over previous four years and deprivation. The six panels show associations between (a) deprivation and random effects unadjusted for exposure and deprivation, (b) exposure and unadjusted random effects, (c) deprivation and random effects adjusted for exposure, (d) exposure and random effects adjusted for deprivation, (e) deprivation and random effects adjusted for exposure and deprivation and (f) exposure and random effects adjusted for exposure and deprivation.	83
3.7	Relative risks, with 95% credible intervals for increase of 10 ppb of SO ₂ for effects of exposure over the previous 0-4, 5-8, 9-12 and 13-16 years, with adjustment for deprivation. The three panels refer to models with different priors on the lagged effects; (a) independent (b) auto-regressive, smoothing over lags within health periods, and (c) auto-regressive, smoothing each lagged effect over health periods.	84

4.1	Time series plots of (a) daily deaths, (b) daily temperatures, (c) daily influenza deaths and (d) relationship between daily deaths and temperatures, London, winters (November–February), 1981–97.	89
4.2	The effects of different spans on modelling the relationship between (i) mortality and temperature (ii) mortality and time (seasonality). The span used to smooth time increases for pairs of plots from top left to bottom right. In each pair of plots, the left shows temperature and the right time.	94
4.3	Modelling the relationships between daily deaths and different lags of temperature. Each panel is the result of fitting a lowess smoother (span=2/3) to each lag separately within a GAM. Dotted lines represent two (pointwise) standard errors.	95
4.4	Modelling the relationships between daily deaths and different lags of temperature together in a single model, fitting a lowess smoother (span=2/3) to each lag within a GAM.	96
4.5	Results of fitting different spline models to the relationship between daily deaths and same day temperature. The number of knots increases from one (left hand panels) to three (right hand panels), and the degree of the polynomial from one (top) to three (bottom). Dotted lines represent two (pointwise) standard errors.	97
4.6	Model setup for BMA analysis of lagged effects of temperature: variables that are included in each of the candidate models. Models to the right of the dotted line are those containing the same day effect of temperature.	106
4.7	Components of Bayes factors approximations used in BMA analysis: changes in deviance, penalty term and approximations of Bayes factors. Models to the right of the dotted line are those containing the same day effect of temperature	107
4.8	Results from using Bayesian model averaging for prediction using a range of temperature scenarios. The number of excess daily deaths expected from low temperatures for periods of up to eight days are given shown, in comparison with a single day at 5 ⁰ . Temperatures range from 5 ⁰ C to -2 ⁰ C for periods of between one and eight days.	108
5.1	Time series plots of logged values of PM ₁₀ for eight sites in London, 1994–97 (missing values are marked on the horizontal axis as triangles).	117
5.2	Scatter plots showing relationships between the four pollutants (PM ₁₀ , SO ₂ , NO and CO) with temperature, at the Bloomsbury site, 1994–97. The lines are loess smoothers.	120

5.3	The (log) correlation between logged values of pollutants and distance between sites. The reference line is $\log(\text{correlation}) = -\phi_p \text{ distance}$, using the posterior median for ϕ_p from the spatial model for PM_{10}	121
5.4	Time series plot for a subset of 250 days: (a) recorded measurements and estimated $\hat{\theta}_t$ for (log) PM_{10} , Bloomsbury, triangles indicate missing values; (b) differences between recorded and estimated values. In (a) the dotted lines indicate approximate 90% intervals, in (b) the dotted lines are $\pm 2\hat{\sigma}_v$	128
5.5	Contour plot of (a) site locations, (b) site effects based on a 20x20 grid of locations without a pollution monitor and (c) the corresponding standard deviations of the site effects. Results are for PM_{10}	137
5.6	The relationship between temperature and (log) pollutant level.	143

Chapter 1

Introduction

The potential effects of ambient air pollution on health is a major issue in public health, and the subject has received a great deal of attention in recent years. The majority of studies investigating this issue have concentrated on acute, or short-term, effects, and a large number have reported associations between both mortality and morbidity with daily changes in routinely monitored measures of air pollution, including particulate matter, sulphur dioxide, oxides of nitrogen, carbon monoxide and ozone.

Much of the evidence for existing guidelines on permissible levels of air pollution is drawn from studies of this type, although there are important methodological issues which need to be addressed, especially when using routinely available health and air pollution data. Both are generally only available at the aggregate (group) rather than at the individual level. In particular, the aggregate data may be aggregated over time and/or space. In the likely presence of unmeasured confounders, analyses may be affected by within- and between-group variability. The presence of such variability may violate some of the fundamental assumptions made in the modelling process, and lead to *over-dispersion* in the outcome. More specifically, the majority of analyses assume that the outcome, for example daily deaths, follow a Poisson distribution, which may be untenable. Over-dispersion may also arise because of data anomalies, i.e. errors in the numerators and/or denominators. Its presence will lead to confidence intervals for the estimates of risk being too narrow and may effect model selection, through over-estimation of the changes in deviances (i.e. how well different models fit the data) between possible models. The latter point is important, as conclusions and interpretation of results are often made based on a single, or a small number of ‘final’ models, the choice of which may be brought into question.

In choosing such a ‘final’ model, many analyses involve fitting a series of regression models, many of which will contain highly correlated variables, e.g. the different lagged values of pollutants or temperature. This can make the choice of a single model, based on global measures of fit and subject matter considerations, highly problematic. Choosing a single model, which is then considered

to be correct when drawing inference about the parameters of interest or performing prediction, ignores the inherent uncertainty associated with the actual modelling process, in particular the uncertainty which arises through selecting a subset of variables for inclusion, together with that due to distributional assumptions and/or transformations. These issues are important to the process of public health policy and regulation affecting admissible concentrations of air pollutants as they may result in different effects size estimates (and their precision) for the various pollutants of interest.

In many of the studies of acute effects, temperature (both through adjustment for day-to-day variability and also as a result of controlling for seasonality) has been treated as little more than a nuisance variable. However, temperature is potentially an important determinant of health in its own right. Of particular public health interest is the question of the possible effects of cold temperatures, including the effect of prolonged periods of cold, e.g., the number of additional deaths that might be expected after a given number of consecutive days below a certain temperature.

Compared to acute effects of air pollution on health, rather less attention has been given to the investigation of chronic effects of air pollution, i.e. the association between health outcomes and long-term exposures to air pollution, possibly over several years. This is primarily due to the lack of availability of suitable data (including potential confounders), and it is uncertain to what extent, if at all, findings from studies of short-term effects can be extrapolated to longer term (chronic) effects. Whilst some or all of the effect of acute exposures may be attributed to ‘mortality-displacement’ (i.e. bringing forward health events which were likely to occur within a short time), chronic exposures may be implicated in fundamental disease causation (e.g. by affecting lung development or function, by sensitising people in early life to respiratory allergens or by provoking cell mutation).

This thesis addresses a number of statistical methodological issues encountered in modelling the acute effects, both of air pollution and temperature, and the chronic effects of air pollution, and by applying appropriate methodologies to national data and data from London, provides estimates of these various effects. The issues addressed include sensitivity to unmeasured variables, the use of repeated, or non-independent, measurements over time, the effects of measurement error, the problems of model selection and comparison and the spatio-temporal modelling of routinely measured pollution levels. Chapter 2 gives a critical appraisal of these issues and assessment of the statistical methods used in studies of the effects of air pollution on health, and introduces the statistical methodology used in this thesis. In the following three chapters, methodology that acknowledges these issues is developed and applied to three substantial epidemiological enquiries; the long-term effects of air pollution on health (Chapter 3), the effects of cold (winter) temperatures (Chapter 4) and obtaining valid estimates of pollution levels for use in studies of acute effects (Chapter 5). Each chapter begins with a description of the epidemiological motivation, together with a review of the relevant literature. This is followed by a detailed description of

the data used in the analyses, with particular reference to how knowledge of the data generating mechanism might be incorporated into the subsequent modelling. The statistical methodology is then presented in stages, starting with a description of initial, or exploratory, analyses, then building up, with increasing complexity, to a final analysis. Results and explanation are given for each of these stages. Chapter 6 provides the discussion of the results and of the issues explored in this thesis.

The individual chapters are now described in greater detail. Chapter 2 includes a critical review of the theory behind regression modelling and its use in epidemiological studies, with particular reference to those approaches used in studies of acute effects, including the issues of model selection and comparison. For each of the aspects of modelling discussed, examples are given of their use from the epidemiological literature. Chapter 2 also introduces the statistical methodology used in this thesis, including the Bayesian approach to inference (upon which the majority of the methodology developed here is based) with comparisons to traditional likelihood approaches. In addition, the issues involved in modelling data with a temporal and/or spatial structure are assessed, with particular reference to modelling underlying levels of pollution using measurements from multiple monitoring sites.

Almost all of the studies of the chronic effects of air pollution have focused only on concurrent exposures: that is, on the associations between health outcome and pollution levels measured in the same, or very recent, years. As such, these studies take no account either of possible latency effects (due to exposures earlier in life), or of the effects of cumulative exposures over many years. The approach presented in Chapter 3 addresses both of these issues at a much higher level of geographical resolution that has been possible before, and additionally allows the effects to vary over calendar time. This is achieved by using health outcomes from distinct time periods, each of which is associated with previous exposure over a number of years. By considering sub-divisions of these exposure periods, the effects of different lags can be examined. This is implemented within a Bayesian framework and applied to ward-level mortality data (respiratory deaths) from Great Britain during the period 1981–96 with associated exposures from ambient concentrations of sulphur dioxide from 1966–81 measured in the same wards. The effects of unmeasured confounders are considered, both in terms of possible over-dispersion and, as a number of the wards may be common to more than one of the time periods, in the possibility that the outcomes may not be considered independent. The effects of socio-economic deprivation, using a census based index, are also taken into account.

In Chapter 4, an alternative approach to traditional model selection is presented, that of Bayesian model averaging (BMA). This acknowledges the inherent uncertainty in the model selection procedure, by calculating posterior probabilities for each of the candidate models which allows weighted averages of the parameter estimates and of predicted values to be obtained, together with estimates of variability that incorporate between-model differences. This approach is applied to

modelling the association between short-term changes in temperature and mortality, using data from winters in London during the period 1981–1997. The question of whether prolonged periods of cold are especially important is addressed, using ‘averaged’ predictions for various temperature scenarios from the different models .

In conducting time-series studies to investigate the relationship between air pollution and health, it is important to have a good measure of the level of pollution on each of the study days. Often daily measurements are available from a number of monitoring sites across the study area. Each of these monitors may measure different sets of pollutants, there may be periods of missing data, and each of the recorded measurements is subject to error. In Chapter 5, a (hierarchical) dynamic linear model is proposed for the analysis of such data, in which the dependencies across time, space and pollutants are exploited. A Bayesian approach is adopted and applied to data collected at eight sites within London, measuring particulate matter (PM_{10}), carbon monoxide (CO), nitrogen oxide (NO) and sulphur dioxide (SO_2) over the period 1994–97. The estimates of the underlying levels of pollution can then be used to ‘map’ the pollution field and for subsequent health analysis, with uncertainty in the exposures being incorporated into the precision of the resulting estimates of risk.

Although the models presented in this thesis were developed to facilitate the analyses of specific epidemiological questions, the issues that they address and the approaches used are general, in that they can be readily extended to a wide range of hypotheses, for data at different spatial and temporal scales. The wider application of these approaches is discussed in Chapter 6, together with chapter-specific discussions of the results from the epidemiological investigations. Future developments of the statistical methodology are also considered.

Chapter 2

Statistical review

In this chapter, the statistical methodology used in examining the relationships between air pollution, temperature and health are reviewed. This includes a description of the methods and underlying statistical theory used both in this thesis and in the existing epidemiological literature. Section 2.1 gives a brief description of the use of likelihood, which forms the basis of much of the ‘traditional’ or frequentist approach to statistical inference commonly used in epidemiological studies. Section 2.2.1 introduces the Bayesian approach, upon which the majority of the methodology introduced in this thesis is based, with comparisons to the likelihood approach. The choice of prior distributions, central to the idea of Bayesian statistics, and the computational aspects are also described.

Section 2.3 reviews the theory behind regression modelling and its use in epidemiological studies, starting with a review of the methods used for estimation when fitting a Generalised Linear Model, which includes the important case of Poisson or log-linear models that are widely used in epidemiological studies (Section 2.3.1). An alternative to this parametric approach is presented in the form of Generalised Additive Models, in which the relationship between the response variable is modelled as a function of explanatory variables which have been ‘smoothed’ using non-parametric methods (Section 2.3.2). The concept of hierarchical models is then explored and methods of estimation, both traditional and Bayesian, are described with particular emphasis on how auto-correlation can be incorporated within the model structure. An alternative approach that is often used in air pollution studies to deal with auto-correlation, Generalised Estimating Equations, is also discussed (Section 2.3.4).

Section 2.4.2 concentrates on the important subject of model selection and comparison and the effect this can have on the inferential process. Methods for the selection of which variables to include in the regression model are compared, Bayes factors are introduced as a method for model comparison, which are an integral component of Bayesian Model Averaging, in which the inherent uncertainty associated with the modelling process is considered.

Much of the data used in studies examining the relationship between short term changes in levels of air pollution and health outcomes takes the form of daily measurements of one, or more, pollutants which are related to daily counts of mortality or morbidity. Aspects of modelling time series data such as this are discussed in Section 2.5. This includes a discussion of stationary processes, which underlie many of the common classes of time series models that are used to describe and forecast time series data. State space models, which provide a powerful framework for modelling auto-correlated data are discussed, and in particular, their close relationship to the structure of Bayesian hierarchical models, are also reviewed.

The last two sections of the statistical review contain a brief introduction the methodology used in spatio-temporal modelling of environmental variables. The concept of a stationary random field is developed from the definition of stationarity encountered in time series section, and methods for interpolating such a field, allowing prediction at locations where data is not recorded, are discussed. A brief description of the difficulties of dealing with non-stationary fields and of ways of transforming them into stationary fields is given. In the last section, these spatial models are expanded to include a temporal component, leading to either separable models, in which the spatial and temporal components are considered independent, or non-separable models, in which they are modelled jointly.

Throughout this chapter, the methods are reviewed within the context of studies of the effects of air pollution on health, from which examples are given.

2.1 Likelihood approach

The likelihood, $L_n(\theta|y) = f(y|\theta)$ is a measure of the *support* provided for a particular set of the parameters, θ , of a probability model by the data, $y = (y_1, \dots, y_n)'$. It is a relative measure and only determined by the values of the observed data, therefore any constants independent of θ can be omitted. It is often more convenient to consider the log likelihood, $l_n(\theta|y) = \log(L_n(\theta|y))$. If the response vector, y , consists of n independent and identically distributed observations, then the likelihood of θ is $L_n(\theta|y) = \prod_{i=1}^n L(\theta|y_i)$ and $l_n(\theta|y) = \sum_{i=1}^n l_n(\theta|y_i)$.

The first derivative of the log likelihood is known as the *score function*, U_n , which for a sample of size n is given by

$$U_n(\theta|y) = l'_n(\theta|y) = \frac{dl_n(\theta|y)}{d\theta} \quad (2.1)$$

The maximum likelihood estimate (MLE) is found by solving the equation $U_n(\theta|y) = 0$.

For reasons of clarity, unless stated, the following descriptions consider a univariate unknown parameter, θ . In regular problems analogous results for more than one parameter follow naturally using standard multivariate techniques such as those described in DeGroot (1986) and Chatfield

and Collins (1980).

2.1.1 Quasi-likelihood

Quasi-likelihood only requires that the first two moments of the data generating distribution are specified, i.e. the mean and variance. The integral $Q(\mu|y) = \int_y^\mu \frac{y-t}{V(t)} dt$ acts like a *quasi* likelihood even if it doesn't constitute a proper likelihood function and estimates of the parameters of the model can be obtained by maximizing under certain conditions. Further details of the use quasi-likelihood can be in Chapter 3 (in relation to over-dispersion in Poisson models, where the variance is greater than the mean) and examples of its implementation can be found in Section 2.3.4 of this chapter (with relation regression models where there is autocorrelation between the outcomes). Further technical details can be found in McCullagh and Nelder (1989).

2.1.2 Likelihood ratio tests

The likelihood ratio test is a useful way of comparing the fit of two competing nested models (where the set of parameters in one model is a subset of those in another). The models can be constructed to correspond to the null and alternative hypothesis of a test. If Ω is the parameter space of the larger model and Ω_0 that of the smaller model, then the models are nested if, $\Omega_0 \in \Omega$. The null hypothesis corresponds to $H_0 : \theta \in \Omega_0$ and the alternative $H_0 : \theta \in \Omega|\Omega_0$ (the set difference). The likelihood ratio test for comparing the two models, and thus testing the hypotheses, is twice the log of the difference between the maximumised likelihoods under the two models

$$2 \left(\frac{\max_{\Omega|\Omega_0} l_n(\theta)}{\max_{\Omega} l_n(\theta)} \right)$$

In the case of regression models, if a particular model is compared to the 'full' or saturated model then the *scaled deviance* can be calculated

$$\begin{aligned} \text{Deviance, } D &= -2 \log(L_k/L_m) \\ &= -2(l_k - l_m) \end{aligned} \tag{2.2}$$

where l_k is the log of the likelihood of model M_k and l_m that for the saturated model, M_m , which contains all n parameters. In what follows, the scaled deviance will be referred to simply as the *deviance*, D .

Large values of the deviance ($L_k \ll L_m$) suggest that the current model does not fit well. If the model in question has r_k variables then the scaled deviance is asymptotically χ^2 distributed with

$n - r_k$ degrees of freedom if the model is appropriate, if the model fits poorly then the deviance will be large.

Rather than make comparisons with the full model, M_m , it is generally more useful to make comparisons with another model with only one (or a few) extra parameters, i.e. comparing models M_k and M_q , with $r_k + 1$ and $r_q + 1$ parameters respectively. This is because, for all but the Normal linear case, the χ^2 test for change in deviances is a much better asymptotic approximation than the one using comparisons with the full model (McCullagh and Nelder (1989)). If $r_k < r_q$ then M_k is said to be nested within M_q and if $D_k = -2(l_k - l_m)$ and $D_q = -2(l_q - l_m)$ then

$$D_k - D_q = -2(l_k - l_q) \sim \chi^2_{r_q - r_k} \quad (2.3)$$

These results assume that the number of parameters is fixed as $n \rightarrow \infty$. Further discussion of the selection of variables for use in regression models can be found in Section 2.4.

2.2 Bayesian approach

The Bayesian philosophy considers that all uncertainty can be described by means of probability distributions, and is thus highly parametric. Whilst non-Bayesian inference can involve many technical difficulties, such as the choice of point estimators and test statistics, Bayesian inference ‘simply’ involves conditional probability calculations (although in practice, this can be anything but simple). What follows is a brief introduction to some of the concepts involved in performing Bayesian analysis, with particular reference to the methods used in this thesis, a more comprehensive account can be found in Gelman et al. (1995).

2.2.1 Introduction to Bayesian inference

When considering the effects of a set of covariates on a response variable, interest often lies in drawing conclusions about the unknown parameters, θ . In the case of a regression model these might include the set of coefficients, β and the variance, σ^2 . The interest might lie in using a regression equation to predict values, Z , of the response variable for particular values of the explanatory variables. When data, y , is observed, such inference will be expressed as $p(\theta|y)$ and $p(Z|y)$ for the cases of parameter estimation and prediction respectively. As in the previous section, results are given for a univariate unknown parameter, θ .

The joint probability distribution relating the parameters, and the observed values, $p(\theta, y)$, can be expressed as a product of two densities; the *prior distribution*, $p(\theta)$ and the density function of y given the value of the parameters, $p(y|\theta)$

$$p(\theta, y) = p(y|\theta)p(\theta) \quad (2.4)$$

Using Bayes' theorem, the *posterior density*, $p(\theta|y)$, representing the updated beliefs about θ after the data has been observed, can be found

$$p(\theta|y) = \frac{p(\theta, y)}{p(y)} = \frac{p(y|\theta)p(\theta)}{p(y)} \quad (2.5)$$

where $p(y) = \int_{\theta} p(y|\theta)p(\theta)d\theta$, over all the possible values of θ (for the case of a continuous parameter, θ).

The denominator, $p(y)$, does not depend on any unknown quantities (y has now been observed and is considered fixed) and is known as the *normalising constant*. Hence, equation(2.5) can be simplified to

$$p(\theta|y) \propto p(y|\theta)p(\theta) \quad (2.6)$$

The prior distribution allows knowledge from previous studies or experiments to be incorporated, but can introduce an increased level of complexity into the calculations. Although there may be cases where there are actual data available with which to construct priors, often decisions have to be based on knowledge of the literature and other sources. In this case, there will be doubts as to the accuracy of the prior distributions and it is important to assess the sensitivity of the posterior probabilities to the choice of priors. Although Bayes theorem implies that the prior distribution, $p(\theta)$ should be a probability distribution, this does not have to be the case. An *improper* prior can be unnormalised and doesn't even have to be integrable. If $p(\theta)$ is a non-negative function such that $\int p(\theta)d\theta = \infty$, but the normalising constant $p(y) = \int_{\theta} p(y|\theta)p(\theta)$ is still finite, then (2.5) still defines a probability density. However, great care must be taken with improper priors, particularly for model selection.

If the interest is in performing prediction, there are two cases to consider; before and after data has been observed. Before it has been observed, the *marginal distribution* of y , $p(y)$, which is also known as the *prior predictive distribution* gives a distribution function for y that does not depend on any previous observations. After the data, y , have been observed, a predicted value, Z , is obtained, as follows

$$\begin{aligned} p(Z|y) &= \int p(Z, \theta|y)d\theta \\ &= \int p(Z|y, \theta)p(\theta|y)d\theta \\ &= \int p(Z|\theta)p(\theta|y)d\theta \end{aligned} \quad (2.7)$$

This assumes that the predicted value, Z , is conditionally independent of the observed data y given θ .

2.2.2 Bayesian point estimates

Strictly speaking there is no need for point estimates within a Bayesian framework, as a parameter is considered a random variable and a lot more information is available from examining the entire

(posterior) distribution. However, point estimates can be useful in comparison and discussion and if a point estimate from the posterior distribution of the unknown parameters is required, then the posterior mean, $\bar{\theta} = E(\theta|y) = \int \theta p(\theta|y) d\theta$, for example, can be found, although this is not necessarily an easy task. Alternatively, if samples can be taken from the posterior distribution, using, for example Gibbs sampling (see Section 2.2.3), then the mean and median together with credible intervals can be obtained. A $100(1 - \alpha)\%$ *Credible Interval*, I , for this point estimate can be found by choosing a and b , with the set $I = (a, b)$ such that for a given α , $p(\theta \in I|y) = \int_b^a p(\theta|y) d\theta = 1 - \alpha$.

2.2.3 Computation

In stylised problems, so-called conjugate prior distributions are available that allow for analytical solutions to finding posterior distributions (see, for example, Gelman et al. (1995)). However, in most practical situations such conjugacy is either not available or overly restrictive and in order to evaluate the posterior model probabilities there are several computational matters which have to be considered. There is therefore a need to compute, or estimate, integrals of the form $p(y) = \int p(y|\theta)p(\theta)d\theta$.

Analytical approximations

One approach is to use an analytic approximation such as Laplace's method, which is based on the Taylor series expansion of a real valued function $f(u)$

$$\int e^{f(u)} du \approx (2\pi)^{r/2} |H|^{1/2} \exp\{f(u^*)\} \quad (2.8)$$

where r is the length of the vector u , u^* is the value of u at which f attains its maximum and H is minus the inverse Hessian information of f evaluated at u^* .

Examples of the use of this form of approximation can be found in Section 2.4.3.

MCMC methods

An alternative approach to finding the posterior probabilities, which avoids the problems of integration, is to use simulation techniques, such as Markov Chain Monte Carlo (MCMC), where a random sample from the posterior distribution is used to estimate $\int p(y|\theta)p(\theta)d\theta$. For further information on MCMC see Gilks et al. (1995) and Gamerman (1997). The Markov chain is constructed so that the stationary distribution is the posterior distribution. Two such Markov chains are the Metropolis-Hastings algorithm and the Gibbs sampler, which are now briefly described.

Starting with an initial value θ^0 , a sequence, $\theta^{(s)}, s = 1, \dots, S$ is generated. At each stage, a 'candidate' value, φ , is drawn from a *proposal distribution*, and the ratio of the (unnormalised) posterior probabilities under each of the values calculated. The candidate value is then accepted with probability r , given by

$$r = \min \left\{ \frac{p(y|\varphi)p(\varphi)}{p(y|\theta^0)p(\theta^0)}, 1 \right\} \quad (2.9)$$

If the candidate value is accepted, then $\theta_{s+1} = \varphi$, otherwise the chain does not move. For illustration, consider an initial value θ^0 , with 'candidate' value, φ , drawn from a normal (proposal) distribution with mean θ^0 and variance b^2 . The ratio of the (unnormalised) posterior probabilities under each of the values is calculated, and the minimum of that ratio and one taken. If a random sample is then taken from a Uniform distribution, $U \sim (0,1)$ and $U < r$, $\theta^1 = \varphi$, the value sampled from the Normal distribution, otherwise let $\theta^1 = \theta^0$, the original value. When this step is repeated N times, by drawing each of the candidate variables from $N(\theta^1, b^2), N(\theta^2, b^2), \dots etc$ then the resulting sample $\theta^0, \theta^1, \dots, \theta^N$ will approximate a random sample from the posterior distribution. In the Gibbs sampler, the proposal distribution is the full conditional distribution of the parameter in question, giving an acceptance probability of one.

2.3 Regression modelling

The use of regression models has become increasingly popular in epidemiological research, especially when dealing with large numbers of covariate strata, where the small numbers within each stratum makes the use of tabular techniques, such as Mantel-Haenszel, untenable. Fitting a regression model generally involves explicitly stating the form of the relationship between the explanatory and response variables being examined, and hence makes clear the assumptions that underpin the analysis. However, this may be at the expense of the level of understanding, especially when advanced statistical methods are used, and it is important to express the results in a form that can be easily interpreted and understood, for example in Poisson regression, rather than just presenting the regression coefficients in their raw form, they can be expressed as relative risks. There are many aspects to model fitting, including the choice of distribution (e.g. Normal, Poisson or Binomial), the choice of the regression (e.g. linear, log-linear or logistic), and how the actual computation is carried out. There are many reasons for performing a regression analysis and the intended use of the regression equation will influence the manner in which the model is constructed and interpreted. In the applications considered here, the interest is in explanation and prediction, but not prediction in the sense of the stock market in which prediction of future events alone is often important.

The most common reason for performing a regression analysis in epidemiology is to obtain estimates of the coefficients associated with the variables of interest, for example, the effect of an

increase in PM_{10} on the risk of respiratory death. In developing a model intended for estimating parameters the choice of which variables are included or eliminated is of great importance. If several of the variables are highly correlated, which is likely to be the case when using several different pollutants and particularly so when dealing with different lagged values of variables, including a variable which is highly correlated with the one of interest may dramatically alter the estimate of its effect or possibly even lead to the variable of interest being excluded from the model. In terms of interpretation and drawing conclusions, there is an obvious desire to have a relatively simple model that nevertheless includes all the important variables, but there are also important theoretical reasons for eliminating irrelevant variables from the model, details of which are given in Section 2.4.1.

As well as using the model to obtain parameter estimates, if there are measurements for various covariates the expected increase in the number of deaths can be predicted, for example given certain increases in the level of PM_{10} and levels of other pollutants and temperature. In comparison with the estimation of parameters, prediction is more tolerant towards the choice of variables, indeed, it can almost be considered fairly unimportant whether the variables might be causally related to the response variable, or that the model is realistic, although this would obviously be preferable. If the prediction is being performed within the same range of data as used in its construction, any variable that has predictive information on the response variable can be considered useful. Of course, more faith would be placed in predictions based on established causal relationships, if they exist, as these are likely to be less sensitive to unusual facets of the data and changes in the correlation structure of the explanatory variables.

Often, in epidemiology, the underlying aim behind using regression models, whether it is explicitly stated or not, is to develop a realistic model of the system, to identify the variables of interest and the strength of their effect. Though an epidemiological regression model cannot itself prove causality, that can only really be provided by randomised experiments, it can indicate the change in the response variable that might be associated with changes in exposure, either by prediction or estimation, which is a very useful tool in developing insight and understanding into possible causal relationships.

2.3.1 Generalised linear models

Generalized linear models (GLMs) are widely used in statistics and a thorough description is given in McCullagh and Nelder (1989). They are essentially an extension of linear models which take the form

1. $Y_i \sim N(\mu_i, \sigma^2)$, i.i.d. Normal observations
2. $g(\mu_i) = \eta_i$, an identity link function

3. $\eta_i = \sum_{j=1}^p x_{ij}\beta_j$, a linear predictor

The parameters, β can be estimated by solving the estimating equations,

$$U_n(\beta) = \sum_{i=1}^p \frac{d\mu_i}{d\beta} v_i^{-1} (y_i - \mu_i(\beta)) = 0 \quad (2.10)$$

The likelihood for a GLM has a general form for distributions in the exponential family of distributions, i.e. those of the form $f_y(y|\theta, \phi) = \exp\{(y\theta - b(\theta))/a(\phi)\} + c(y, \phi)\}$, where the expectation and variance of Y take the form $E(Y) = b'(\theta)$ and $VAR(Y) = v_i = b''(\theta)a(\phi)$ respectively. The function $a(\phi)$ is often of the form $a(\phi) = \phi/w$, where ϕ is the dispersion parameter and w the prior weights.

In the majority of studies examining the affects of air pollution on health, the outcome, e.g. number of daily deaths, is assumed to follow a Poisson distribution. In this case, $f_y(y|\theta = (\mu, \phi)) = \exp\{y \log \mu - \mu - \log y!\}$, therefore $\log \mu = \theta$, $b(\theta) = \exp(\theta)$, $c(y, \phi) = -\log y!$ and $a(\phi) = 1$.

In general, the link and error structure have to be explicitly defined. Poisson data, for example, take the form of counts and thus the outcome is in the range $[0, \infty)$. The natural link function is therefore $\log(\mu_i) = \gamma_i$, which restricts the outcome to the required range and assumes an underlying multiplicative relationship.

Testing between competing nested models is achieved using the difference in deviance, as described in Section 2.1.2. The deviance will be of the form $\sum_{i=1}^n 2\{y_i(\tilde{\theta}_i - \hat{\theta}_i) - b(\theta_i) + b(\hat{\theta}_i)\}/\phi = D(y|\hat{\theta})$, where $\hat{\theta}$ is the MLE under the model in question and $\tilde{\theta}$ is the MLE under the saturated model. The deviance for Poisson data is therefore $D = 2 \sum_{i=1}^n \{y \log(\tilde{\theta}/\hat{\theta}) - (y - \hat{\theta})\}$.

The MLEs in all but the Normal case cannot be found by simply maximizing the log likelihoods numerically. Instead, they can be found using Iteratively Re-weighted Least Squares (IRLS) (see McCullagh and Nelder (1989)).

The choice of a log link and variance that is proportional to the mean is the canonical one for the Poisson distribution and under regular Poisson assumptions, the variance is assumed to be equal to the mean, $VAR(Y_i) = E(Y_i)$, i.e. the dispersion parameter $\phi = 1$. When this is relaxed to proportionality, $VAR(Y_i) = \phi E(Y_i)$, the *dispersion*, or *scale*, parameter is assumed to be constant over all of the data. The estimating equations for this will generally be different from those obtained by weighted least squares, but solutions can be found using quasi-likelihood techniques (see Sections 2.1.1 and Chapters 3 and 4). This *over-dispersion* may arise in a number of different ways, causing the variance to be greater than the mean. One example is when there is inter-subject variability where the Poisson mean parameter is not constant, but is a random variable. If the mean is assumed to have a Gamma distribution, with mean μ and variance μ/ϕ then the distribution of the outcome, Y will follow a negative binomial distribution,

$p(y|\mu, \phi) = \frac{\Gamma(y+\phi\mu)\phi^\phi}{y!\Gamma(\phi\mu)(1+\phi)^{y+\phi\mu}}$, with $E(Y) = \mu$ and $VAR(Y) = \mu(1 + \phi)/\phi$. Further discussion of the effects of over-dispersion, the ways in which it can be represented, and the effect it can have on model comparison and the precision of risk estimates can be found in Chapters 3 and 4, where particular reference is made to the relationship between the presence of unmeasured confounders and over-dispersion.

2.3.2 Generalised additive models

Generalised additive models (GAMs) can be thought of non-parametric extensions of GLMs. In Section 2.3.1 the form of the GLM was given as $g(\mu_i) = g(E(y)) = X\beta$. In a GAM, instead of assuming dependence on the sum of linear predictors, the outcome is assumed to be dependent on a sum of smooth functions of the predictors

$$g(E(Y)) = \sum_j S_j(X_i) \quad (2.11)$$

where S_j are smoothing functions. GAMs therefore provide a flexible framework for controlling for non-linear dependence on potential covariates. They are fitted using an iterative procedure known as the backfitting algorithm, which simultaneously fits all functions, before smoothing the partial residuals of each variable in turn and iterating until the individual functions don't change (Hastie and Tibshirani (1990)).

Smoothers

The class of potential smoothers is very large and only a selection of the most commonly used are described here. For a more complete review see Hastie and Tibshirani (1990). Possibly the most intuitive smoother is the moving average or running mean, in which the value of x_i is replaced by the average of that value together with those within a defined period around t

$$S(X_t) = \sum_{k=-p}^p w_p X_{t+k} : i = p+1, \dots, n-p \quad (2.12)$$

where w_k are the weights and $p > 0$. Here, $2p+1$ is the order of the moving average, i.e. the number of points which are included in calculating the average.

Unless the number of points included in the calculation of the average is large, in which case information will be lost, the resulting line is unlikely to be very smooth due to the sudden drop in weights as points are not included in the calculation. However, if the weights were specified to follow a probability distribution, as in *density estimation* (Silverman (1986)), then a smooth set of weights can be defined. An alternative approach is to calculate a least squares regression line within each set of neighbouring points of the form

$$S(x_i) = \hat{\alpha}_{x_i} + \hat{\beta}_{x_i} x_i \quad (2.13)$$

The fitted value at x_i is then using as the smoothed value. The regression can be fitted using weights, w_i , which can be constructed to be decreasing with distance from x_i , giving a *loess* smoother (Hastie and Tibshirani (1990)).

Another approach is to fit a polynomial to the data, for example

$$S(x_i) = \sum_{k=0}^p \beta_k x_i^k \quad (2.14)$$

where the coefficients, β_k , are estimated from the data. As with the moving average approach, a choice of the extent of the smoothing (the order of the polynomial) has to be made. Simple polynomials are often useful as part of a model for the trend with additional terms describing the seasonality, but polynomials of higher order, whilst theoretically allowing very good fits to part of the data are likely to be too restrictive, especially if the model is to be used on another dataset or for forecasting out of the range of the current set.

Regression *splines* fit piecewise polynomials separated by *knots*, and usually are forced to join smoothly at the knots. When the first and second derivatives are continuous, the piecewise polynomial is known as a *cubic spline*. A drawback of this approach is the need to choose the number and position of the knots, although attempts have been made to automate the procedure (Denison et al. (1998)).

The number of time series studies that have used non-parametric smoothing functions to assess the relationship between air pollution and health has markedly increased as the software required to fit GAMs has become widely available (Chambers and Hastie (1993)), and it is now almost the standard approach for epidemiological studies of this type. For example, in analysing the relationship between PM_{10} and SO_2 and hospital admission in New Haven and Tacoma, U.S., Schwartz (1995) used a GAM to replicate the 19 day moving average used in Kinney and Ozkaynak (1991), modelling the outcome as Poisson outcome rather than being restricted to the normal distribution as a consequence of using moving averages of the actual outcome variable. Further examples of the use of GAMs are given in Section 2.5, where they are used to allow for long-term patterns, such as seasonality, in time-series studies, enabling the associations between short-term changes in air pollution and health to be examined.

2.3.3 Hierarchical models

Many epidemiological studies involve data that inherently have a hierarchical structure, for example when assessing the differences in mortality rates across hospitals for a particular procedure, data may be measured both at the hospital ‘level’ (e.g size, speciality) and on individual patients within the hospitals (e.g. age, gender). An introduction to the need for hierarchical models and the implementation of the linear case are given in Goldstein (1987), Sullivan et al. (1999) and

Burton et al. (1998). In this section, a simple regression example is presented leading to a discussion of *Generalised Linear Mixed Models* (GLMM) (Breslow and Clayton (1993)), which are an extension of GLMs that allows the inclusion of random effects. Hierarchical models are also known as multi-level models and may be fitted using likelihood methods, but are naturally viewed from a Bayesian perspective.

Hierarchical models offer a convenient way of handling the different levels of correlation that may be present between outcomes in longitudinal studies where repeated measurements, j , may be made on a set of subjects, i . They provide a framework for incorporating the correlation that may be present between measurements made at different times but on the same individual. A comprehensive introduction to the subject of analysing longitudinal data of this type can be found in Diggle et al. (1994). For example, the first, or top level of the model might relate the outcome on individual i to a set of level 1 covariates

$$Y_{ij} = \alpha + \beta X_{ij} + \epsilon_{ij} \quad (2.15)$$

In a naive pooled analysis, the ϵ_{ij} 's would be considered *i.i.d* $N(0, \sigma^2)$. However, as each of the i sets of j readings made on each individual are likely to be correlated outcomes, this assumption is unlikely to be tenable. One approach would be to fit a different intercept term for each individual (a fixed effect) but this requires the ratio of J to I to be suitably large.

Alternatively, a random effect can be fitted, referring to the fact that the regression coefficients can vary from individual to individual. This alternative to fitting a separate intercept for each individual, i , assumes that there exists an overall intercept (α_0) for the whole population and that a separate intercept α_{0i} for each individual comes from a distribution with expectation α_0 and a suitable variance. For example $\alpha_{0i} \sim N(\alpha_0, \sigma_\alpha^2)$ or $u_i = \alpha_{0i} - \alpha_0 \sim N(0, \sigma_\alpha^2)$. Model (2.15) can therefore be expressed as

$$Y_{ij} = \alpha_0 + u_i + \beta X_{ij} + \epsilon_{ij} \quad (2.16)$$

with $E(\epsilon_{ij}) = 0, V(\epsilon) = \sigma_\epsilon^2, COV(\epsilon_{ij}, \epsilon_{ik}) = 0$ and $E(u_i) = 0, V(u_i) = \sigma_u^2, COV(u_i, u_j) = 0$. The variance of Y_{ij} is thus $\sigma_\epsilon^2 + \sigma_u^2$, where the variance between Y_{ij} within a single individual is σ_u^2 . The intra-subject correlation coefficient is therefore $\frac{\sigma_u^2}{\sigma_u^2 + \sigma_\epsilon^2}$. Given estimates of σ_u^2 and σ_ϵ^2 , the *exchangeable* correlation structure means that every observation within an individual is equally correlated with every other observation in that individual, i.e. $\rho_{jj} = 1, \rho_{jk} = \rho \ j \neq k$, the intra class correlation coefficient.

In non-hierarchical models, tests of statistical significance are associated with the addition (or deletion) of one or more parameters and the likelihood ratio test (section 2.1.2). However when drawing inference about random parameters, the usual null hypothesis, that a single variance is zero, will violate one of the standard regularity conditions needed to ensure that the asymptotic

distribution of the test statistic is χ^2 . In such cases, a standard χ^2 test will underestimate the significance of an observed departure from the null hypothesis (Laird and Ware (1982)).

Generalised Linear Mixed Models are essentially GLMs (section 2.3.1) that include one or more random effects. Conditional on random effects, b_i , the y_i are assumed independent, with expectation, $E(y_i|b) = \mu_i = g^{-1}(\gamma_i) = g^{-1}(X_i\beta + z_i b_i)$ and variance, $VAR(y_i|b) = \sigma^2 = a(\phi)v(\mu_i)$. The likelihood for N subjects, each with N_j measurements will thus be

$$L(\beta, b, \sigma^2|Y) \propto \prod_{i=1}^N \prod_{j=1}^{n_j} p(Y_{ij}|\beta, b_i, \sigma^2) \quad (2.17)$$

If the random effects are assumed to be normally distributed, $b \sim N(0, D)$, then within subject correlation can be taken into account whilst allowing each subject to have a unique covariance structure. The correlation structure can be defined using the covariance, D (rather than simply using $I\sigma$). In a Bayesian setting, where the covariance matrix will be assigned a prior distribution, D^{-1} is often assumed to have a Wishart prior, this being conjugate to the multivariate normal distribution.

Frequentist approach

The parameters of the distribution of the random effect are known as *hyperparameters*, θ and must also be estimated from the data. In the case of linear regression, the parameters, β, ϕ and θ can be estimated by maximum likelihood using iterative generalised least squares (Goldstein (1987)). In the more general case, the usual strategy is to attempt to estimate the parameters by using the integrated likelihood, details of which can be found in Laird and Ware (1982) and Diggle et al. (1994).

$$p(Y|\beta, \phi, \theta) \propto \int p(Y|\beta, b, \phi)p(b|\theta)db \quad (2.18)$$

In general, $\hat{b} = E(b|\hat{\theta}, \hat{\beta}, \hat{\phi})$, using the MLEs obtained from (2.18). Whilst this approach is feasible for Normal models, the computations are intractable in other cases, although they can be approximated. Breslow and Clayton (1993) review approaches to approximate these calculations, including penalised quasi-likelihood which is estimated using a variation of the iteratively re-weighted least squares calculations used for GLMs (see Section 2.3.1). The problem with this approach however, is that the underlying uncertainty in $\hat{\beta}, \hat{\phi}$ and $\hat{\theta}$ is not acknowledged.

Bayesian approach

In assigning the prior distributions of the hyperparameters, θ , the Bayesian approach provides a natural framework for dealing with hierarchical models, incorporating the uncertainty in the estimates of the parameters. For illustration, the set of level one (non-hyper) parameters are

denoted simply by β (omitting σ^2/ϕ for clarity), in which case the appropriate Bayesian posterior distribution is the vector (θ, β) , with joint prior distribution, $p(\theta, \beta) = p(\theta)p(\beta|\theta)$. The joint posterior distribution is then $p(\theta, \beta|y) \propto p(y|\beta)p(\theta, \beta)$. The hyperparameters can be estimated by obtaining the marginal posterior distribution, $p(\theta|y)$, which involves integrating the joint posterior distribution, $p(\theta|y) = \int p(\beta, \theta|y)d\beta$.

Hierarchical models have been used within the context of studies of air pollution and health in an attempt to obtain more stable estimates of the regression coefficients. With a suitably large period of study, it may be possible to perform separate analyses on independent subsets of the data and to collate the results, performing a within study *meta-analysis*. This was done by Schwartz and Marcus (1990) who fit separate autoregressive models to each of the 14 winters within the period 1958–72 and fit a normal random effects model using the average of the resulting estimates of the effect of particulate matter (and SO_2), and an estimate of the variance that allowed for the presence of more year-to-year variability than might be expected simply from the within year variances, due to omitted explanatory variables (which might include those with temporal structure).

In the APHEA study (Katsouyanni et al. (1995)), a similar method was used to pool the results from Poisson regressions from several cities in Europe. Touloumi et al. (1997) examined the effect of oxidant exposure in 6 cities and Zmirou et al. (1998), the effect of PM_{10} and SO_2 in 10 cities, with relation to daily mortality. In each, models were fitted separately for each city and the resulting estimates averaged, weighted by the inverse variance, implying a fixed effects model. However, the models fitted in each of the cities did not necessarily include the same set of variables apart from the pollutant in question, so the interpretation of the estimates will differ in terms of what they are adjusted for, and additional variation may have been introduced. Calculating such an average assumes that the estimates all came from the same underlying distribution from which a point estimate is a reasonable summary, an assumption that may not be tenable if the make up of the pollutants was different in the cities, i.e. if they were being generated by different mechanisms, such as industry, heating, cars, etc... In an attempt to acknowledge this, where there was evidence of significant heterogeneity in the estimates, a random effects model was used assuming that the individual regression coefficients are a sample from an underlying normal distribution with variance equal to the between-cities variance, which was estimated from the data.

These studies have a number of advantages over traditional meta-analyses in that there is no selection bias, beyond any present in the original choice of cities included in the study, no cities or results were excluded and to a large extent, the methods used to produce the estimates were the same. Although different explanatory variables may have been used, these differences are likely to be much less than collecting results from different studies with completely different approaches to the data collection and model formation.

Domininci et al. (2000) used a Bayesian hierarchical model to combine estimates of the effect of PM_{10} and SO_2 on daily mortality from the twenty largest cities in the U.S. A GAM was fitted separately for each city, modelling deaths as a Poisson regression with smoothed functions of time and lagged temperature. The relative risks from each city, β^c , (representing short-term changes in air pollution) were then assumed to come from an underlying normal distribution $\beta^c|\alpha, \Sigma \sim N(z^c\alpha, \Sigma)$, where z^c are city specific covariates. The trend and weather terms are treated as nuisance parameters, η^c and are assumed to be independent across cities. A fully Bayesian analysis of this would involve assigning probability distributions (and associated hyperparameters) to all the parameters at the second level (the first level being the regression equation itself), whether of interest or nuisance. Integrating the joint posterior distribution of all the parameters with respect to the nuisance parameters, would then give the marginal posterior distributions of the β^c 's, i.e. $p(\beta^c|\text{data, hyperparameters}) = \int p(\beta^c, \eta^c|\text{data, hyperparameters})d\eta^c$. This is difficult in practice as the regression was performed using GAMs, and although the pollutant was used as a linear term, the other explanatory variables were modelled non-parametrically. Instead, after the first stage, the regression parameters are taken as given and essentially transferred to a separate analysis to describe the variability between cities. In this analysis, the likelihood estimates of β^c and η^c were assumed to be multivariate normal with means and covariance matrices equal to the corresponding MLEs. The marginal likelihood of β^c then has a normal distribution. An alternative approach would be to use MCMC to sample from the posterior. This was done for a sample of the cities, but it was said to be computationally prohibitive for the whole analysis. The covariance matrix, Σ , was assumed to have a Wishart prior, for which the off-diagonals (the correlation between cities) took into account the distance between cities by using an inverse exponential distance function, although it seems unlikely that there would be any meaningful spatial structure at this resolution, as the cities were spread across the United States, with very large distances between them, and so the idea that correlations between the cities might be based on their geographical proximity seems unlikely.

2.3.4 Generalised estimating equations

An alternative approach to hierarchical models for allowing for the possibility of serial correlation (and over-dispersion) that has been used widely in air pollution epidemiology are Generalised Estimating Equations (GEEs). GEEs extend generalised linear models to the extent that the correlation between individuals is taken into consideration in the formulation of the model, the estimation of the regression coefficients and their standard errors. Here, the concepts behind GEEs are briefly introduced together with examples of how they have been used in epidemiological studies. Further details of GEE models can be found in Liang and Zeger (1986), Zeger et al. (1988), Diggle et al. (1994) and Zeger and Liang (1992).

Unlike a hierarchical model, where the structure of the covariance model is explicitly defined (2.3.3), a standard GEE is a marginal model. In a marginal model, the outcome, Y_i , is related to a linear predictor which contains only fixed effects and the correlation or variance parameters are estimated from the residuals. Comparisons between the uses of GEEs and hierarchical models for analysing repeated measures data are given in Omar et al. (1999) and Burton et al. (1998).

In a GEE model, only the first two moments of the joint distribution of Y_i are specified. For this reason, a multivariate analogue of the quasi-likelihood (Section 2.1.1) is used. The regression coefficients, β , can be estimated by simultaneously solving variants of the estimating equation used in GLMs (equation 2.10).

These models do not explicitly model the correlation structure, and as such they are of little use when it is the correlation structure that is of primary interest. Consequently, the basic correlation structure should be chosen before starting the analysis with the aim being to estimate the parameters that characterise the chosen structure. Examples of the correlation structure include independence; where each observation in an individual is completely uncorrelated with every other observation in that individual; exchangeable or compound symmetry, where each observation within an individual is equally correlated with every other observation within that individual; and autoregressive, where two observations taken close (in time or space) within an individual, are likely to be more closely correlated than observations taken far apart.

GEEs have been used widely in time series analyses of air pollution and health, to allow for the serial correlation that is likely to be present in the residuals from a regression model that assumes independence. Examples can be found in Schwartz (1994a), Schwartz and Dockery (1992a) and Schwartz (1991), who analysed the effect of particulate matter on daily mortality in Birmingham, Alabama, Philadelphia, and Detroit, U.S. respectively. Log-linear regression model were used with an autoregressive structure to allow for the autocorrelation between daily mortality counts. An exchangeable covariance structure was used by Schwartz and Dockery (1992b) and comparison made with an autoregressive structure, however there was no noticeable difference in the results. It is notable that the model building and selection in these examples was performed on regular GLMs assuming independence, with the final chosen model fit using GEEs, and it is not clear whether this might have an effect on the suitability of the model that is used for estimation.

GEEs were used in a different way by Moolgavkar et al. (1995), in an analysis of data from Philadelphia, U.S. for 1973–88. They performed log-linear regression, but instead of using day as the unit of analysis in the repeated measures of the GEE, they performed separate analyses for each season, for example 16 summers, and used the season as the unit of observation in the GEE and using the independence structure. Using the seasons as replications, more stable estimates could be obtained than if a single observation was used. In this sense, it is the same as the rationale as for the hierarchical model of Schwartz and Marcus (1990) described in Section 2.3.3,

but achieved through using a marginal model rather than a hierarchical one.

2.4 Model selection and comparison

In constructing a regression model, decisions will have to be made on which variables should be included, and what form they should take, i.e. should they be transformed or categorised. The choice of variables and the methods used may, in part, be determined by the reasoning behind the regression analysis, whether for estimation or prediction.

2.4.1 Effect of selection on properties of estimators

This section concentrates on the important subject of subset selection, the choice of which variables to include in the regression equation, and the effect this has on the inferential process.

In choosing which models should be considered, it is important that they have some plausibility outside the realms of statistical significance. Careful consideration of prior information is advocated by Rothman and Greenland (1998), although it is acknowledged that in most cases the prior knowledge available is likely to be too limited to provide much guidance in model selection (Section 2.2.1 describes how, in a Bayesian setting, if such information were available, it could be directly incorporated into the analysis). They suggests the selection procedure should start with a *credible model*, which would be considered ‘compatible with available information’ and gives the example of modelling cancer rates, where such a model would include age and sex. Variables are then added to the starting model, their importance, and thus whether they are retained, being assessed, based on their contribution to the fit of the model to the data. There are many techniques for automatically selecting variables to be included in a model, and a selection of these are described in Section 2.4.1, but the danger in using these so called ‘black box’ routines is that a known confounder may be excluded from the model, possibly in favour of a variable with no biological plausibility. In addition to a credible starting model, the final model should also not conflict with prior information. The choice of which other variables are included can have a significant effect on the interpretation of the effect of the variables of interest, although if the model is to be used only for prediction there is more tolerance to the choice of variables, a point emphasised by Clayton and Hills (1993).

In this section, the effect of variable selection on the results of the regression equation are explored. It should be noted that the theoretical effects described hold for cases when the selection is not based on information from the current data, which is often not the case. In order to completely eliminate the effect of bias caused by the selection of variables, completely independent data sets would be required for (i) the selection of the model and for (ii) the estimation of the regression coefficients. However, this unlikely to be an efficient approach in practice as less information will

be available for the actual estimation of the parameters.

For illustration, the effects of variable selection are demonstrated using linear regression, although the concepts generalise to other forms which may be better suited to an epidemiological investigation, such as Poisson regression.

Given a response variable, Y , with a single possible explanatory variable, X , to which a linear regression of the form $E(Y|X) = \alpha + \beta X$ is fitted, then an unbiased estimate, $\hat{\beta}$ of β can be found and then tested to see whether it is significantly different from zero. If it is then β is included in the equation, if not then it is dropped. Using this procedure means the estimate $\hat{\beta}$ will no longer be unbiased for β , as it depends on decisions made based on the data. If β is included in the model, rather than just fitting the intercept term, then the expectation of $\hat{\beta}$ should be written

$$E(\hat{\beta} | \hat{\beta} \text{ is significantly different from } 0)$$

which will not be equal to β . Chatfield (1995) describes a *pre-test estimator* which also considers *not* fitting the line, when $\beta = 0$, in order to give the unconditional estimate of $\hat{\beta}_{PT}$

$$\hat{\beta}_{PT} = \begin{cases} \hat{\beta}, & \text{if } \hat{\beta} \text{ is significant,} \\ 0, & \text{otherwise.} \end{cases}$$

From which it is obvious that $E(\hat{\beta}_{PT})$ is not generally equal to the unconditional estimator, $E(\hat{\beta})$. The effect of variable selection is investigated more fully later in this section, but briefly, Chatfield advises that (i) least squares theory does not apply when the same data are used to formulate and fit a model, and (ii) the analyst must always be clear exactly what any inference is conditioned on.

In order to investigate further the effects of variable selection on the estimates of the parameters and prediction, consider a linear model with n observations, each with a response variable y and a set of p possible covariates $x_1, \dots, x_p$. The full model, but not necessarily the ‘true’ underlying model, is obtained by fitting all the available variables,

$$y_i = \beta_0 + \sum_{j=1}^p \beta_j x_{ij} + \epsilon_i \quad (2.19)$$

Here, the residual terms, ϵ_i , are assumed independently and identically distributed, and in the case of Normal linear regression are such that $\epsilon_i \sim N(0, \sigma^2)$.

The full model is expressed in matrix notation as

$$Y = X\beta + \epsilon \quad (2.20)$$

where Y is the vector (of length n) of observed responses, X the full design matrix of dimension $n \times (p + 1)$, containing all of the p variables for which information is available together with a column representing the intercept term (which is assumed to be included in all of the models), β

Table 2.1: Possible design matrices and corresponding regression coefficients when using combinations of two covariates

Model, k	X_k	r_k	β^k	$X_{k'}$	$p + 1 - r_k$	$\beta^{k'}$
1	$[1_n]$	1	β_0^1	$[X_1, X_2]$	2	$(\beta_1^1, \beta_2^1)'$
2	$[1_n, X_1]$	2	$(\beta_0^2, \beta_1^2)'$	X_2	1	β_2^2
3	$[1_n, X_2]$	2	$(\beta_0^3, \beta_2^3)'$	X_1	1	β_1^3
4	$[1_n, X_1, X_2]$	3	$(\beta_0^4, \beta_1^4, \beta_2^4)'$	-	0	-

is the $p + 1$ vector of unknown regression coefficients and ϵ the n vector representing the random component.

With p potential covariates, by considering all the possible models that comprise combinations of each of them being included/excluded there will be $m = 2^p$ possible models, M_k ($k = 1, \dots, m$). Assuming that model, M_k , contains r_k regressors, so that $p + 1 - r_k$ have been omitted. For model, M_k , let X_k and β^k represent the design matrix and parameter vector respectively and $X_{k'}$ and $\beta^{k'}$ the complement, so that $X = (X_k, X_{k'})$ and $\beta^T = \{(\beta^k)', (\beta^{k'})'\}$.

A natural ordering is used, so that $k = 1$ represents the null model containing just the intercept term with design matrix $X_1 = 1_n$, an n vector of one's. The full model is represented by $X_m = X$, the complete $(n \times (p + 1))$ design matrix containing all the variables.

To clarify this notation, an example with two possible covariates, X_1 and X_2 is presented. Here, $p = 2$ and therefore the total number of possible models is $m = 2^2 = 4$. The full design matrix, X can be written as $[1_n, X_1, X_2]$, which is equivalent to X_4 . The four possible models are shown in Table 2.1.

As discussed above, from maximum likelihood theory, if the full model is correct then the expected value of $\hat{\beta}$ is equal to the true value of β , however, if the model includes variables that are extraneous then this is no longer true. If the true model is $Y = X_k \beta^k + \epsilon$ but the full model is fitted then the expected value and variance of $\hat{\beta}$ will be as follows

$$\begin{aligned}
 E(\hat{\beta}_k) &= \hat{\beta}_k + (X_k' X_k)^{-1} (X_k' X_{k'}) \hat{\beta}_{k'} \\
 &= \beta^k + A \beta^{k'} \\
 &= \text{true value} + \text{bias term}
 \end{aligned} \tag{2.21}$$

$$VAR(\hat{\beta}^k) = (X_k' X_{k'})^{-1} \sigma^2 \tag{2.22}$$

where $A = (X_k' X_k)^{-1} (X_k' X_{k'})$

This shows the bias in the regression coefficients that can be expected if the extraneous variables contained in $X_{k'}$ are included in the model. This bias will exist unless (i) $\beta_k = 0$, the true coefficients of the additional variables are actually zero and thus the variable has no effect on the response, or (ii) $X_k' X_{k'} = 0$, there is no correlation between the two sets of variables meaning that the inclusion of non-important variables will not have an effect on the explanatory power (and the

estimates) of the variables of interest. This is often a product of design in a controlled experiment, as in the case of randomised trials where randomisation is used to allow the investigator to separate the effects of different covariates, but is unlikely to be the case in observational studies.

It can be shown that the components of $\hat{\beta}^k$ will generally have less variation than those from the full model, $\hat{\beta}$. This means that the precision of the estimates can be reduced by using a subset model, even if the terms excluded are not completely extraneous, i.e. do not have a true coefficient of zero. It can also be shown that the estimate of the variance, $\hat{\sigma}_k$, will generally be biased upwards.

The reason that this result is not taken to the extreme case, where just the intercept would be fit and thus predict the same value irrespective of the values of the covariates, is because of the bias that is introduced when important variables are ignored. As variables are added to the model, the usual trade-off between reducing the bias and increasing the variance has to be considered. This bias term can also be expressed in terms of the mean square error (MSE) of the estimates $\hat{\beta}$, where

$$\begin{aligned} MSE(\hat{\beta}) &= E(\hat{\beta}^k - \beta^k)(\hat{\beta}^k - \beta^k)' \\ &= (X_k' X_k)^{-1} \sigma^2 + A \beta_{k'} \beta_{k'}' A' \\ &= \text{variance} + (\text{bias})^2 \end{aligned} \tag{2.23}$$

If a variable is included that has no effect on the response then the variance will be increased without a decrease in the bias. If a variable has a small effect on the response, then the reduction in bias has to be weighed against the increase in variance.

The bias in the regression estimates and the change in the variability associated with variable selection will have an associated effect on any predicted values generated from the regression equation. If a set of explanatory variables is considered $x' = (x_k' x_{k'}^*)$ (comprising values for both the important and extraneous variables) then the full model will produce response, $\hat{Y} = x' \beta$, with mean $x' \hat{\beta}$ and prediction variance,

$$\text{VAR}(\hat{Y}) = \sigma^2 (x' (X' X)^{-1} x) \tag{2.24}$$

As the point predictions are essentially based on the individual estimates, the variances decrease as terms are excluded from the model.

2.4.2 Selection procedures

Section 2.1.2 described how changes in deviances between nested models can be tested to assess differences in model fit. Procedures have been developed based on such tests which aim to identify good (though not necessarily the best) subset models, generally they aim to use less

computation that would be necessary if all the possible regressions were performed and examined. These methods are often referred to as *stepwise regression procedures*. Dependent on the method, subset models are identified sequentially by adding or deleting variables. Three, commonly used, stepwise methods are described here. These methods are in no way guaranteed to find the ‘best’ or even the same subset of variables.

Forward selection of variables adds one variable at a time to the previously chosen subset. At the beginning of the process, the first variable chosen is that which accounts for the largest amount of variation in the dependent variable, i.e. the variable having the highest correlation with Y . At each successive step, each of the remaining variables are added one at a time to the model and that which results in the largest reduction in the residual sum of squares is retained. This will be the variable with the highest correlation with the *residuals* from the current model. A stopping rule is needed to terminate the process, or it would continue until all the variables have been included.

If RSS_k denotes the residual sum of squares with r_k variables in the model and if the smallest RSS which can be obtained by adding another variable is RSS_q , for new model, M_q , then

$$R = \frac{RSS_k - RSS_q}{RSS_q / (n - r_k - 2)} \quad (2.25)$$

is calculated and compared to a ‘F-to-enter’ value, F_e , i.e. if R is greater than F_e , then the variable is added to the selected set.

Backward elimination is similar but starts with the full model and then chooses a variable to delete at each step according to which would result in the smallest increase in the residual sum of squares (the variable with the smallest partial sum of squares). Again, a stopping rule is needed or the process will continue until all the variables have been deleted.

If RSS_k denotes the residual sum of squares with $r_k + 1$ variables in the model and if the smallest RSS which can be obtained by deleting any variable, thus making the smaller model M_t , is RSS_t then

$$R = \frac{RSS_t - RSS_k}{RSS_k / (n - r_k - 1)} \quad (2.26)$$

is calculated and compared to the ‘F-to-stay’ value, F_s , if R is less than F_s , then the variable is deleted from the selected set.

It should be noted that although the terms ‘F-to-enter’ and ‘F-to-stay’ suggest that the ratios, R , follow an F -distribution, this is not the case. If, after r_k variables have been entered and the next variable is chosen at random then the assumptions required to use an F -test (that the

residuals are independently, identically and Normally distributed) are met. However, if the next variable is chosen so as to maximise R then the distribution will not be an F -distribution. Using approximations to the distribution of R to obtain nominal 5% points for the 'F-to-enter' taken from the F -distribution have been shown to give true significance levels in excess of 50% (Draper et al. (1979)).

Stepwise selection is a combination of the previous two approaches, neither of which takes in to account the effect that adding or deleting a variable can have on the contributions of the other variables previously included in the model. A variable that has been added to the model early on may become unimportant after another, possibly highly correlated, variable has been added. This is certainly going to be the case when considering the effects of pollutants and temperature. For this reason, stepwise selection is an extension of forward selection, that re-checks at each step the importance of the variables currently included in the model. If the partial sum of squares for any previously included variable(s) do not meet some predetermined criteria then those variables are excluded from the model before the next addition step is attempted.

The stopping rules, referred to as 'F-to-stay' and 'F-to-go' are both required in stepwise selection. However, there is no reason why they should be set at the same level, indeed it might be preferable to have a more relaxed criterion for inclusion as this would force the procedure to consider larger numbers of subset models.

One criticism of these selection methods is that they imply an order of importance to the variables. This can be misleading since, for example, in forward selection, the first variable that was included is often deemed unimportant when other variables have been included. Indeed, it could be the first variable to be excluded when backwards elimination is applied. It should also be remembered that the choice of variables in the final model can be determined to varying extent by the particular data under consideration, the selection method used and the stopping rule. Unless an independent dataset is used for the selection process, the bias in the individual parameter estimates will also be affected by these factors. The more extensive the search for the model, the more chance there is that the final model might be an extremely good fit to the data in question, but possibly with little other general use. By basing the inference on the results of a single model, the possibility of extreme or non-typical values in the data affecting the parameter estimates will be increased.

A danger in using these methods is that they might appear to replace the need for the investigator to obtain a genuine feel for the data, with respect to the problem under consideration, and may mask the possible effects that model assumptions and variable selection might have on the inferential process.

There are alternatives to the forward and backward selection criteria, including the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). The AIC chooses the

model which maximises

$$l(\hat{\theta}) - r_k \quad (2.27)$$

where $l(\hat{\theta})$ is the log-likelihood evaluated at the MLE and r_k is the number of parameters in the model. The penalty term, $-r_k$, aims to guard against models being chosen simply because they contain more variables and thus will have greater explanatory power, however limited the increase.

The BIC chooses the model which maximises

$$l(\hat{\theta}) - (r_k/2) \log n \quad (2.28)$$

This has an additional penalty term which considers the sample size, guarding against accepting variables into the model which might be considered ‘significant’ just because they are based on large numbers. The BIC will tend to favour simpler models than the AIC and also has the benefit of being closely related to Bayes Factors (see Section 2.4.3).

In studies of air pollution and health it is common for regression models to contain a number of highly correlated pollutants in the same regression and to attempt to draw inferences about which variables were causal, or at least had the strongest relationship, from that model. Many studies have used stepwise procedures, including many highly correlated pollutants and meteorological variables in the candidate list. For example, Kinney and Ozkaynak (1991) performed multiple regression using backward selection on 5 pollutants and 3 temperature variables in Los Angeles County. After pre-whitening both the outcome (daily deaths) and the explanatory variables using moving averages, they found significant correlations only between deaths and same day levels of the pollutants and temperature, except for ozone, for which a 1 day lag was significant. These variables were then included in the multiple regression on which backwards selection was performed, with the resulting model including same day NO_2 and temperature, and 1 day lag O_3 . The difficulties in using an automatic procedure to select between highly correlated variables have been discussed, and here there was the additional complication of the use of filtering on the data. Although this was an attempt to eliminate the problems of auto-correlation and cross correlation between the variables, the choice of which lagged variables to include in the initial model could be determined to a large extent by the exact filtering mechanism. For example, if a particular lag did not show significant correlation with the outcome after the filtering, it was not a candidate for inclusion in the initial model, even before the selection procedure was performed. The authors performed some sensitivity analysis, performing the analysis with several alternative filters, and also without any filtering, and roughly consistent results were found for filters removing cycles of > 7 , > 15 , > 30 and > 90 days, but not for > 365 days or without filtering. However, this sensitivity analysis was only performed for the effect of filtering on the final selection of (lagged) variables included in the chosen model. There was no investigation of

the effect of using different filtering mechanisms on the original list of all the possible variables, which due to the high correlations present could well have led to a different choice of final model using the automated procedure.

Given the problems discussed with automatic procedures of this type, the high levels of correlation between the explanatory variables and the generally low explanatory power of air pollution for mortality or morbidity, such an approach is unlikely to produce stable conclusions. The APHEA study (Katsouyanni et al. (1995)) took a different approach. Initially, single pollutant models were fitted and if one or more pollutant was associated with the outcome then attempts were made to separate the effects by examining associations with one pollutant stratified by the other. There was also the advantage of having a multi-city study, in which the relative levels of the pollutants were unlikely to be the same in each city, which could help in identifying interactions and masking of the effects of some of the pollutants. However, the model building was performed using a log transformed outcome with the final model using Poisson regression (Katsouyanni et al. (1995)). A similar approach was used by Gouveia and Fletcher (2000) who used linear regression with a log transformed outcome during the model building process, with different models being compared using F-tests. After a single final model was chosen, Poisson regression was used to examine the association between lagged values of five pollutants (PM_{10} , SO_2 , NO_2 , CO and O_3) and respiratory diseases in Sao Paulo, Brazil.

It is not clear what effect this strategy of basing the model building and the estimation on different distributional assumptions will have, but changing the modelling assumptions within an analysis in this way is surprisingly common in air pollution studies. This is often done to benefit from the simplicity associated with using the Normal distribution, even if the Poisson has been stated to be more appropriate. Given the relative ease with which more appropriate models can often be used, there seems little reason to introduce this additional complication to the analysis. In another example, Rahlenbeck and Kahl (1996) analysed the effect of particulate matter and SO_2 during winters in East Berlin, Germany (1981-99). They acknowledge the data as Poisson, and despite a relatively low number of average daily deaths (38.6, $sd=7$), they used normal regression to allow moving averages of the outcome variable (differences with 15 and 29 day means) in order to handle temporal correlation in a simple manner.

2.4.3 Bayes factors

In this section, an alternative to the likelihood ratio (LR) based methods of model comparison described in Section (2.1.2) is introduced. In common with those methods, Bayes factors give an indication of how much more likely the data are to have arisen under one model than another, but unlike the LR tests, the models do not have to be nested. Further details on Bayes factors, their interpretation and calculation can be found in Kass and Raftery (1995) and O'Hagan (1995).

For illustration, the simplest case is considered, where there are just two possible models, M_0 and M_1 . If the two models are assigned prior probabilities $p(M_0)$ and $p(M_1)$ then these are compared with the probabilities of the data under each of the models, $P(y|M_i), i = 0, 1$, the *prior predictives*, to obtain *posterior model probabilities*, $p(M_0|y)$ and $p(M_1|y)$. Contained in each model is the uncertainty associated with its parameters. For each of the models, M_k ,

$$p(y|M_k) = \int p(y|\theta_k, M_k)p(\theta_k|M_k)d\theta_k \quad (2.29)$$

From Bayes Theorem the probability that the data was generated under the first model can be obtained,

$$p(M_0|y) = \frac{p(y|M_0)p(M_0)}{p(y)} = \frac{p(y|M_0)p(M_0)}{p(y|M_0)p(M_0) + p(y|M_1)p(M_1)} \quad (2.30)$$

As there are only two models under question the posterior odds of model 1 in favour of model 2 can be expressed as

$$\frac{p(M_0|y)}{p(M_1|y)} = \frac{p(y|M_0)}{p(y|M_1)} \cdot \frac{p(M_0)}{p(M_1)} \quad (2.31)$$

The term $p(y|M_0)/p(y|M_1)$ is known as the *Bayes factor*, B_{01} in favour of model 0 over model 1. Therefore,

$$\text{Posterior Odds} = \text{Bayes factor} \times \text{Prior Odds} \quad (2.32)$$

In the simplest case where the prior probabilities of each model are equal, $p(M_0) = p(M_1) = 1/2$, then the posterior odds will equal the Bayes factor

$$\frac{p(M_0|y)}{p(M_1|y)} = \frac{p(y|M_0)}{p(y|M_1)} = B_{01} \quad (2.33)$$

As previously mentioned (in Section 2.2.3), there are often difficulties associated with the, often complex, integrals involved in obtaining the posteriors. One approach, suggested by Kass and Wasserman (1995), to approximating the posterior model probabilities is to restrict the choice of priors in such a way that approximations to the posterior model probabilities can be made that have no dependence on the prior distributions and don't involve performing integration. They suggest approximating the posterior model probabilities using the Bayesian Information Criteria, BIC, which was introduced in Section 2.4.2 under its more common use as a measure of model fit. When using the BIC approximation, the posterior probabilities are given by $\hat{P}(M_k|Y) = L_n(\hat{\theta})_k - \frac{r_k}{2} \log n$, where $L_n(\hat{\theta})_k$ is the log-likelihood evaluated at the MLE under model k and r_k is the number of parameters in the model. This is an approximation of order $O_P(1)$. For information, the notation, $O_P()$, known as 'big O_P ' indicates that the error term series is *bounded in probability*, i.e. for sufficiently large enough n , the sequence of error terms is bounded by some constant, K . The smaller the order of the error term sequence (for example,

$1/n, 1/n^2, \dots$) the smaller the error terms become for large n , and thus the better the approximation. Further details on the subject of *stochastic order* are available in Bishop et al. (1975).

To calculate Bayes factors for generalised linear models, Raftery and Richardson (1996) suggest a (Laplace) approximation of the form seen in Section 2.2.3. When this is used to approximate the Bayes factor (relative to the null model) the result is

$$\hat{m}_k = L(\hat{\theta}|Y)p(\hat{\theta})2p^{r_k/2}|H|^{1/2} \quad (2.34)$$

where $\hat{\theta}$ is the mode of the posterior $P(\theta|Y)$ and H is the matrix of second derivatives of the log-posterior evaluated at θ . This approximation has an error of order $O_P(n^{-1})$. In order to use this, the posterior mode $\hat{\theta}$ and the inverse Hessian of the log-posterior evaluated at θ are required, which are not routinely available when using standard statistical software packages.

Raftery and Richardson (1996) suggest a way of using values of the MLE, Deviance and the Fisher information matrix, I_θ which are widely available when fitting generalised linear models (see Section 2.3.1 and McCullagh and Nelder (1989)) to estimate Bayes Factors for comparing two models, again using the Bayesian Information Criteria. This approximation is of the order, $O_P(n^{-1})$ when the expected Fisher information is used for F_k , but this can be improved to $O_P(n^{-1/2})$ if the observed information is used. Further details of this method, and its use in model averaging can be found in Chapter 4.

2.4.4 Bayesian model averaging

In Section 2.4.2, the traditional approach to model ‘selection’ was considered, that of choosing a single model which is then considered to be correct when drawing inference about the parameters of interest or performing prediction. This ignores the inherent uncertainty associated with the actual modelling process, and in particular the uncertainty which arises through selecting a subset of variables for inclusion, together with that due to distributional assumptions and/or transformations.

Table 2.2: Possible models and posterior model probabilities using combinations of two covariates, with $\epsilon_i (i = 1, 2, 3, 4) \sim \text{i.i.d } N(0, \sigma_i^2)$

Model	Formula	Posterior Probability
M_1	$Y = \beta_0^1 + \epsilon_1$	$p(M_1 y)$
M_2	$Y = \beta_0^2 + \beta_1^2 X_1 + \epsilon_2$	$p(M_2 y)$
M_3	$Y = \beta_0^3 + \beta_2^3 X_2 + \epsilon_3$	$p(M_3 y)$
M_4	$Y = \beta_0^4 + \beta_1^4 X_1 + \beta_2^4 X_2 + \epsilon_4$	$p(M_4 y)$

This section describes an alternative approach where, rather than choosing a single model, information from several models is combined to obtain averaged estimates and/or predicted values. To illustrate this concept, consider the notation for the regression example introduced in Section 2.4.1 with two potential covariates, X_1 and X_2 , and $m = 4$ possible models, which are shown in Table 2.2.

The choice is then to

1. Choose the single model with the highest posterior probability, $p(M_i|Y)$, and use this to draw inference about the effects of the covariates (or to predict), ignoring the model uncertainty in the same way as would be done by picking a single model as the result of some likelihood comparison or stepwise procedure.
2. Examine the estimates from all four of the model separately, which introduces the problem of multiple comparisons and might not very useful if a single estimate/predicted value is required.
3. Combine the information from the four models.

The last approach is the essence of the Bayesian modelling averaging approach, inference on the parameters of interest and/or prediction is performed using weighted averages of the values from the individual models. The process is carried out within a Bayesian framework; each of the candidate models is assigned a prior probability which is updated using information from the data to obtain posterior model probabilities.

Consideration should be given to the fact that inference about an unknown parameter will be dependent on the structure of the model under examination, M_k , and the set of other variables included in that model. In order to calculate posterior probabilities for the parameters, one should condition not only on the observed data but also on the model in question. This is now incorporated into the model uncertainty framework. Given m candidate models, $M_1, \dots, M_m$ ($m = 2^p$), each with prior probability, $p(M_i)$ (the simplest example being where all the models are considered equally likely before any data is observed, where $p(M_k) = 1/m$), then the posterior probability for model, M_k can be found using Bayes' theorem in the following way

$$p(M_k|Y) = \frac{p(Y|M_k)p(M_k)}{\sum_{l=1}^m p(Y|M_l)p(M_l)} \quad (2.35)$$

Contained in each model is the uncertainty associated with its parameters. For ease of illustration, the following description is given in terms of the regression parameters, β . For each of the models, M_k ,

$$p(Y|M_k) = \int p(Y|\beta_k, M_k)p(\beta_k|M_k)d\beta_k \quad (2.36)$$

Therefore, given prior probability distributions for the unknown parameters, $p(\beta|M_k)$, under a particular model, M_k , then, conditioning on the model, gives

$$p(\beta|Y, M_k) = \frac{p(Y|\beta, M_k)p(\beta|M_k)}{p(Y|M_k)}$$

with $p(Y|M_k) = \int p(Y|\beta, M_k)p(\beta|M_k)d\beta$.

In order to predict a future observation, $Z = \hat{Y}$, using the model in question, the predictive distribution is given by

$$p(Z|y, M_k) = \int p(Z|\theta, M_k)p(\theta|y, M_k)d\theta \quad (2.37)$$

The Bayes Factors comparing two models, M_k and M_l , is then defined as

$$B_{kl} = \frac{p(M_k|y)}{p(M_l|y)} = \frac{\int p(y|\beta_k)p_k(\beta_k)d\beta_k}{\int p(y|\beta_l)p_l(\beta_l)d\beta_l} \quad (2.38)$$

When the posterior model probabilities have been obtained, the weighted posterior distributions of the individual parameters can be calculated, allowing the effect of individual covariates to be assessed. The weighted posterior density, mean and variance of the parameters of interest, β_j can be calculated as follows

$$\begin{aligned} p(\beta_j|y) &= \sum_{l=1}^m p(\beta_j^l|y, M_l)p(M_l|y) \\ E(\beta_j|y) &= \sum_{l=1}^m \bar{\beta}_j^l p(M_l|y) \end{aligned} \quad (2.39)$$

$$\begin{aligned} \text{VAR}(\beta_j|y) &= E_{M|y}[\text{VAR}(\beta_j|y, M_l)] + \text{VAR}_{M|y}[E(\beta_j|y, M_l)] \\ &= \sum_{l=1}^m (\sigma_j^l)^2 p(M_l|y) + \sum_{l=1}^m (\bar{\beta}_j^l - \bar{\beta}_j)^2 p(M_l|y) \\ &= \text{within model uncertainty} + \text{between model uncertainty} \end{aligned} \quad (2.40)$$

where $\bar{\beta}_j^l = E(\beta_j|y, M_l)$, is the posterior mean of value of the regression coefficient, β_j , under model M_l , and $\bar{\beta}_j = E(\beta_j|y)$.

When performing prediction of a new observation, $\hat{Y} = E(Z|y, X^*)$, using a particular model, the predictive distribution is given by

$$p(Z|y, M_k) = \int p(Z|\beta^k, X^*, M_k)p(\beta^k|y, M_k)d\beta^k \quad (2.41)$$

When all models are considered, the resulting prediction will be a weighted average of the predicted values from all the possible models

$$p(Z|X^*) = \sum_{l=1}^m P_l(Z|M_l, y, X^*)p(M_l|y) \quad (2.42)$$

With expected value

$$\begin{aligned} E[Z|y, X^*] &= E_{M|y}[E[Z|y, X^*, M]] \\ &= \sum_{l=1}^m E[Z|y, X^*, M_l]p(M_l|y) \end{aligned} \quad (2.43)$$

and variance

$$\begin{aligned}
 V[Z|y, X^*] &= E_{M|y}[VAR[Z|y, X^*, M]] + VAR_{M|y}[E[Z|y, X^*, M]] \\
 &= \sum_{l=1}^m VAR[Z|y, X^*, M_l]p(M_l|Y) + \sum_{l=1}^m (Z_l - \bar{Z})^2 p(M_l|y)
 \end{aligned} \tag{2.44}$$

where Z_l is the predicted value using model M_l , and $\bar{Z} = E[Z|y, X^*]$

Interpretation

It is natural to ask what is the interpretation of a coefficient arising from a Bayesian model averaging procedure. For illustration, consider again the case of Normal linear regression with just two covariates, X_1 and X_2 as described in Table 2.2. Considering Z , the predicted value given values of X_1 and X_2 ,

$$E(Z|\beta_0, \beta_1, \dots, \beta_p, X_1, X_2, y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \tag{2.45}$$

Expanding this expression over all the possible values of β gives

$$\begin{aligned}
 E(Z|X, y) &= E_{\beta|y}(E(Z|X, \beta_0, \beta_1, \beta_2, y)) \\
 &= E_{\beta|y}(\beta_0 + \beta_1 X_1 + \beta_2 X_2, y) \\
 &= E(\beta_0|y) + E(\beta_1|y)X_1 + E(\beta_2|y)X_2
 \end{aligned} \tag{2.46}$$

Taking expectations over the set of possible models gives

$$\begin{aligned}
 E(Z|X) &= E_{M|y}(E(Z|X, M_k, y)) \\
 &= E_{M|y}(E_{\beta|y}(E(Z|X, M_k, \beta, y))) \\
 &= \sum_{l=1}^4 E_{\beta|y}(E(Z|X, M_l, \beta, X, y))p(M_l|y)
 \end{aligned} \tag{2.47}$$

where $X = (X_1, X_2)'$.

Hence

$$\begin{aligned}
 E(Z|X) &= E(\beta_0|y, X, M_1)p(M_1|y) \\
 &+ \{E(\beta_0|y, X, M_2) + E(\beta_1|y, X, M_2)\}p(M_2|y) \\
 &+ \{E(\beta_0|y, X, M_3) + E(\beta_2|y, X, M_3)\}p(M_3|y) \\
 &+ \{E(\beta_0|y, X, M_4) + E(\beta_1|y, X, M_4) + E(\beta_2|y, X, M_4)\}p(M_4|y)
 \end{aligned} \tag{2.48}$$

In the above, in model M_2 for example, $E(Z|X, \beta_0, \beta_1) = \beta_0 + \beta_1 X_1$, so that even if the value of X_2 is known, it isn't used in that model. If $\beta^l = (\beta_0^l, \beta_1^l, \beta_2^l)$ is the set of parameter estimates

found under model l then

$$\begin{aligned}
 E(Z|X_1, X_2, \beta^1, \beta^2, \beta^3, \beta^4, y) &= \beta_0^1 p(M_1|y) \\
 &+ [\beta_0^2 + \beta_1^2 X_1] p(M_2|y) \\
 &+ [\beta_0^3 + \beta_2^3 X_2] p(M_3|y) \\
 &+ [\beta_0^4 + \beta_1^4 X_1 + \beta_2^4 X_2] p(M_4|y)
 \end{aligned} \tag{2.49}$$

A unit increase in X_1 will therefore result in a predicted value of Z'

$$\begin{aligned}
 E(Z'|X_1 + 1, X_2, \beta^1, \beta^2, \beta^3, \beta^4, y) &= \beta_0^1 p(M_1|y) \\
 &+ [\beta_0^2 + \beta_1^2 (X_1 + 1)] p(M_2|y) \\
 &+ [\beta_0^3 + \beta_2^3 X_2] p(M_3|y) \\
 &+ [\beta_0^4 + \beta_1^4 (X_1 + 1) + \beta_2^4 X_2] p(M_4|y)
 \end{aligned} \tag{2.50}$$

The difference being

$$\begin{aligned}
 E(Z'|X_1 + 1, X_2, \beta^1, \beta^2, \beta^3, \beta^4, y) - E(Z|X_1, X_2, \beta^1, \beta^2, \beta^3, \beta^4, y) \\
 = \beta_1^1 p(M_2|y) - \beta_1^4 p(M_4|y)
 \end{aligned} \tag{2.51}$$

This weighted average (by the posterior model probabilities) of the effects from each of the models in which X_1 is included, gives a natural interpretation for the a regression coefficient obtained using Bayesian model averaging.

Unlike the traditional approach of selecting just one model in which the estimate of a parameter of interest might be highly effected by the choice of which other variables which were in/excluded in the model, this estimate allows for the fact that other variables may also be in the model. An estimate of the weight of support for including X_1 in the model can be found by summing the posterior probabilities for the models that include the variable

$$p(X_1 \text{ should be included}) = \sum_{l=1}^m P(M_l|Y, \text{ model } l \text{ contains } X_1) \tag{2.52}$$

Only models that contain the particular variable, X_1 , and that have non-zero posterior probabilities will contribute to this sum, as is the case for the weighted averages in equation (2.47).

In studies of the effects of air pollution on mortality, it is often more appropriate to use Poisson regression, rather than simple linear regression. In this case, the expectation of the predicted value, Z , in the two covariate example is given by

$$E(Z|\beta_0, \beta_1, \beta_2, X_1, X_2, y) = \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2, y) \tag{2.53}$$

When this expectation is expanded to condition on the possible models, as in equation (2.47), it

gives

$$\begin{aligned}
 E(Z|X) &= \sum_{l=1}^4 E_{\beta}(E(Z|M_l, \beta, X, y))p(M_l|y) \\
 &= \exp(\beta_0^1|y)p(M_1|y) \\
 &+ \exp(\beta_0^2 + \beta_1^2 X_1|y)p(M_2|y) \\
 &+ \exp(\beta_0^3 + \beta_2^3 X_2|y)p(M_3|y) \\
 &+ \exp(\beta_0^4 + \beta_1^4 X_1 + \beta_2^4 X_2|y)p(M_4|y)
 \end{aligned} \tag{2.54}$$

A unit change in X_1 (with X_2 remaining constant) again only results in changes in the predictions from the two models that contain X_1 (2 and 4), difference between the two predicted values is now represented by

$$\begin{aligned}
 &p(M_2|y)[\exp(\beta_0^2 + \beta_1^2(X_1 + 1))/\exp(\beta_0^2 + \beta_1^2 X_1)] \\
 &+ p(M_4|y)[\exp(\beta_0^4 + \beta_1^4(X_1 + 1) + \beta_2^4 X_2)/\exp(\beta_0^4 + \beta_1^4 X_1 + \beta_2^4 X_2)] \\
 &= p(M_2|y)[\exp(\beta_1^2(X_1 + 1))/\exp(\beta_1^2 X_1)] \\
 &+ p(M_4|y)[\exp(\beta_1^4(X_1 + 1) + \beta_2^4 X_2)/\exp(\beta_1^4 X_1 + \beta_2^4 X_2)]
 \end{aligned} \tag{2.55}$$

This non-linear expression of the regression coefficient for X_1 is not the same as would be obtained by simply calculating a weighted average of the coefficients from each models and taking the exponent, i.e. $\exp(p(M_2|y)\beta_1^1 + p(M_4|y)\beta_1^4)$. Despite this problem, the models can still be used for prediction, as the individual predictions are made under the conditions of the specific models and the resultant predicted values averaged to produce a single final predictive value. For prediction, the fact that the effects of the individual parameters have been lost is not important, as the interest is in the combined effect of all the covariates. In the case of pollution, for example, the ability to interpret the effect of a $25\mu\text{g}/\text{m}^{-3}$ increase of PM_{10} on the risk of respiratory deaths for all values of PM_{10} would be lost, but it could be found for categories of values, such as 0–25, 25–50, etc... and the ability to predict the additional number of respiratory deaths associated with an increase in PM_{10} , given values for other covariates, such as SO_2 , CO , O_3 and temperature is retained. This approach is explored in Chapter 4, where BMA is used in examining the associations between cold winter temperatures and mortality.

2.5 Temporal modelling

Much of the data used in studies examining the relationship between short term changes in levels of air pollution and health outcomes takes the form of daily measurements of one, or more, pollutants which are related to daily counts of mortality or morbidity. The outcomes are likely to exhibit temporal correlation, i.e. there is likely to be a high degree of correlation between values of the outcome, Y_t , especially within close periods of time. This is not necessarily because

the outcome, e.g. number of daily deaths, is causally related to the number the day before, as might be the case with a contagious epidemic, but because the underlying risk factors in which interest lies, such as temperature and pollution, themselves tend to be highly correlated day to day. The values of Y_t cannot then be considered mutually independent and this must be considered in any analysis. Classical time series composition and analysis is primarily interested in modelling the behaviour of the response variable, rather than its relationship with a set of explanatory variables, although the methods used can be very useful in understanding the nature of any serial dependence in residuals and thus in constructing suitable models. There are many comprehensive texts on the subject of time series and forecasting and only a brief review is presented here. For a more complete treatment of the subject see Harvey (1981), Diggle (1991), Hamilton (1994), Chatfield (1996) and Chatfield (2000). When the continuous observations are considered, they are usually discretised by taking samples at equal intervals of time (e.g. day) to give discrete time series. The main aims of traditional time series analysis are description, modelling, forecasting and control. The main reasons for time series in epidemiology are modelling where the response can also depend on past or present values of other explanatory variables. Classical time series modelling aims to decompose the variation in the series into

<i>Trend</i>	T_t	long term movements in the mean
<i>Seasonality</i>	S_t	annual cyclical fluctuations
<i>Cycles</i>	C_t	other cyclical variations, at frequencies less than or greater than a year
<i>Residuals</i>	E_t	other random or systematic fluctuations

After modelling the trend, seasonality and cyclic components, there may still be autocorrelation in the residual term, due to short term dependencies in the data, and it is this fourth component that is the one of interest, as any systematic changes in the outcome, for example daily mortality, may be associated with short term changes in air pollution and temperature.

2.5.1 Stationary processes

A time series is said to be stationary if the distributional structure of Y_t is unaffected by a shift in time. *Strict stationarity* means that for any choice of s , and times $t_i, i = 1, \dots, m$, the joint probability distribution of $Y_{t_i+s}, i = 1, \dots, m$ is the same for all s . i.e. the joint distribution does not depend on the location in time. So, if a process is stationary, observations from any time period can be used to make inference about the overall underlying structure.

A weaker assumption is one of *weak* or *second order* stationarity where the mean, $E(Y_t) = \mu_t$ is constant for all t , and the auto-covariance function (ACF), $\gamma_{ts} = COV(Y_s, Y_t)$ or $\gamma_{ts} = COV(Y_s -$

$\mu_s, Y_t - \mu_t$) depends only on the distance (in time) between s and t and not their actual location, e.g. $\gamma_{1,5}$ would be equal to $\gamma_{11,15}$. This, on its own, would not imply strict stationarity.

2.5.2 Trend and seasonality

The type of models that are commonly considered in air pollution epidemiology are the form

$$Y_t = \mu_t + \epsilon_t \quad (2.56)$$

where ϵ_t is a (weakly) stationary random function, which is considered as ‘noise’, and μ_t is the *trend*.

In order to use a model of this type, assumptions have to be made about the form of either the trend or the noise function. The trend is the long term change in the underlying process. A simple linear trend, $\mu_t = \alpha + \beta t$ is often referred to as a *deterministic* or *global* trend. Alternatively, a *local* trend can be modelled that evolves through time, allowing the parameters to be dependent on time, $\mu_t = \alpha_t + \beta_t t$. A local trend can also take the form of a recursive equation, such as in state space models (see Section 2.5.3), e.g. $\mu_t = \mu_{t-1} + \beta_t t$.

Formally, the trend, μ_t , of a time series is defined as the expectation of the set of random variables Y_t , i.e. $\mu_t = E(Y_t)$. Often it is treated as a linear term within a model, indicating the general pattern of rise or fall of the outcome variable, e.g. the number of daily deaths, over the period of study. If the trend can be modelled successfully, then the noise term, ϵ_t , may be considered to be stationary.

Smoothing

In deciding what form the model for the trend should take, it is useful to try and decompose the time series into its ‘rough’ (noise) and ‘smooth’ components. Possibly one of the most intuitive methods is that of calculating a moving average, as described in Section 2.3.2. Modelling the trend in this way can be done using non-parametric functions within the framework of General Additive Models (see section 2.3.2), in which a variety of smoothing techniques can be used in combination to model the underlying trend and the seasonal patterns commonly present in time series data. More sophisticated approaches might involve fitting different functions, such as a loess curve with suitably chosen levels of smoothing, or spans of time (see Section 2.3.2).

A moving average approach was used by Mazumdar et al. (1982a), who used a 15 day moving average when analysing daily counts of death in winters in London, UK in the 1960s. This took into account long wavelength patterns in the data, but rather than using density estimation, the weights dropped from 1 to 0 between the seventh and eighth day which created distortions in the

smoothed data. Moving averages were also used by Kinney and Ozkaynak (1991) who investigated the relationships between daily death counts (all cause, respiratory and cardiovascular) and levels of five pollutants (O_x , SO_2 , NO_2 , CO and KM , a measure of optical reflectance) together with temperature variables in Los Angeles County during the period 1970–79. They explored the temporal structure of any associations between mortality and the pollutants (and temperature) by pre-whitening the data and then examining the cross-correlations. The pre-whitening consisted of transforming both the input and output series using a filter designed to remove all the auto-correlation in the individual variables, in this case an $AR(2)$ function and 365 day differencing was applied to each variable, leaving residuals that had little auto-correlation. They then used multiple regression to assess the effect of the short term changes in the pollutants on the different outcomes, after subtracting weighted 19-day moving averages from each variable, including the outcome. In both these examples, the data were assumed to be Gaussian, which may have been plausible in large cities such as London (mean daily deaths almost 300 (Schwartz and Marcus (1990))) and Los Angeles (mean daily deaths=152), but most studies of air pollution and health involve far fewer daily counts and whilst filtered Gaussian data is still Gaussian, this is not generally the case for other distributions, such as the Poisson.

Examples of the use of GAMs to model underlying trends and long-term patterns in time-series studies can be found in Schwartz (1994c), where they were used to analyse the relationship between particulate matter and daily mortality in Birmingham (Alabama, U.S.), between particulate air pollution, ozone and SO_2 and daily hospital admissions for respiratory illness (Philadelphia, U.S.) and between ozone and particulate matter for children in six cities in the eastern U.S. In each of these analyses, Poisson models were fitted using loess smoothers for time, temperature and a linear function of pollution, e.g. $\log(E(Y)) = \alpha + S_1(time) + S_2(temperature) + S_3(dewpoint) + \beta PM_{10} + \text{day of the week dummy variables}$. A similar model was used in Schwartz (1997), but in this case the effect of weather was modelled using regression splines. One of the major advantages of using non-linear functions could then be exploited, that of being able to include the possibility of thresholds, without forcing the epidemiologist to specify a functional form. In particular there appeared to be a non-linear relationship between temperature and mortality. However, unlike GLMs, there are no fixed degrees of freedom and so although deviances can be calculated, and ways of approximating the degrees have been proposed, there are no formal tests of differences between models (Hastie and Tibshirani (1990)). Despite the possibility of using a non-linear function for the variables of interest, namely air pollution, the final results were always presented for a linear term, in order to produce a single summary measure (relative risk per unit increase) and for comparability with other studies

Given the somewhat arbitrary choice of smoothing functions, such as loess, to remove long-term time trends or the number of degrees of freedom when fitting spline functions, it is important that sensitivity analysis is performed to assess the possible effects. This is particularly true

where decisions have to be made as to which data should be included in the analysis. Schwartz (1994e), for example, presents results from several models using different smoothers and where different sets of outliers are excluded, for example very hot or very humid days, and for different temperature variables, e.g. mean daily or nightly minimum. Although it is good practice to present the results from a number of different possible models, only a small selection can be presented and if they are not nested, or have any common structure, then it is difficult to make sensible comparisons between the different models, in addition to the theoretical issues of testing between non-parametric models of this type.

An example of the possible effects of different choices of smoothing is presented in Chapter 4, where the effects of cold temperatures are examined in relation to daily mortality. A series of loess smoothers with different spans are used to control for seasonality, resulting in markedly different results for the measure of interest, i.e. the association between short-term changes in temperature and daily mortality.

Although great effort has been given to dealing with temporal patterns in the studies described here, it is still likely that the residuals will exhibit auto-correlation. One approach is to include lagged values in the GAM, as in Schwartz (1995), where successive lagged terms were included in the model until there was no auto-correlation in the residuals. In Schwartz (1994e), the residuals of the chosen models were examined for auto-correlation, and if it was present, GEE models were fit, however this meant reverting to parametric models and essentially performing a different analysis. In a re-analysis of the data from Birmingham U.S., Kelsall et al. (1997) use splines within a GAM to control for temporal effects, assuming constant over-dispersion over time. Estimates of over-dispersion, beyond that which would be expected for Poisson data, would indicate that less of the variation was being accounted for by the model, and would include the effects of residual auto-correlation.

2.5.3 State space models

State space models arise from the notion that a measured signal will be contaminated by noise. In the simplest case, a state space model would consist of two equations, or levels; the observation equation and the system equation. In the observation equation, the ‘signal’ at time t is taken to be a combination of the variables in an underlying *state vector* also at time t , i.e. $Y_t = F_t\theta_t + u_t$, where $u_t \sim (0, \sigma_u^2)$. The state vector is the minimum set of information from the past data which contains all the information necessary for predicting future behaviour. So, the state vector has the Markov property that the latest value is all that is required. Whilst it may not be possible to observe all (or even any) of the elements of the state vector, θ_t directly, assumptions can be made about how it changes in time, leading to the system equation, $\theta_t = G_t\theta_{t-1} + \epsilon_w$, where $\epsilon_w \sim N(0, \sigma_w^2)$. Once data have been observed, inference can be made about this underlying state

vector, which can often be thought of as the true underlying level of the signal. Due to the set up of the state space model, inference can be performed sequentially, updating the estimates of the parameters as more data become available over time.

Normal Dynamic Linear Models (DLMs)

Analogous to the idea of state space models is that of dynamic linear models (DLMs) (West et al. (1985)). Here the joint distribution of $(Y_1, \dots, Y_T, \theta_1, \dots, \theta_T)$ is considered multivariate normal, with the measured values, y_t , normally distributed with an expected value equal to some function of the underlying true value (the state vector), $F_t\theta_t$, with associated error (or variance term) V_t . The current value of the underlying state vector is again normally distributed, with expected value equal to that of a function of the previous value ($G_t\theta_{t-1}$), with covariance matrix, W_t . G_t is known as the transition matrix of the system, if F_t, G_t, V_t, W_t are not functions of t , then the combination of observation and system equations gives a constant DLM

$$\begin{aligned} \text{Observation equation, } Y_t|\theta_t &\sim N(F_t\theta_t, V_t) \\ \text{System equation, } \theta_t|\theta_{t-1} &\sim N(G_t\theta_{t-1}, W_t) \\ \theta_1 &\sim N(m_1^*, c_1^*) \end{aligned} \tag{2.57}$$

This can also be presented in additive form, giving the traditional form of a state space model

$$\begin{aligned} \text{Observation equation, } Y_t &= F_t\theta_t + v_t \\ \text{System equation, } \theta_t &= G_t\theta_{t-1} + w_t \end{aligned} \tag{2.58}$$

Where $v_t \sim N(0, V_t), w_t \sim N(0, W_t)$, with $V_t \perp\!\!\!\perp \theta_t, W_t \perp\!\!\!\perp \theta_{t-1}, (V_t, W_t) \perp\!\!\!\perp \theta_{t-1}, y^{t-1}$.

Estimating the parameters

Having constructed a Normal DLM, the main parameter of interest, θ_t , can be estimated for each time, t . In the simplest case, $F_t = G_t = 1, V_t = V, W_t = W$, which is known as a *steady model*. In the case where the two variances are known, the resulting estimate can be found easily using Bayes Theorem. By considering the likelihood of the observed data and treating the underlying level as a ‘prior’ for the mean parameter, θ , then the posterior distribution can also be seen to be Normal with the mean, $E(\theta_t|y^t)$, being a weighted average of the observed data, y_t and the previous day’s underlying level, θ_{t-1} .

In the case of univariate θ ,

$$E(\theta_t|y^t) = E(\theta_{t-1}|y^{t-1}) + A_t(y_t - E(\theta_{t-1}|y^{t-1})) \tag{2.59}$$

where A_t is a weighting factor based on the ratio of two variances, $A_t = V(\theta|y^t)/V(y_t|\theta)$.

This is recognisable as one of the updating formula for the Kalman Filter (Meinhold and Singpurwalla (1983), Harvey (1981)).

It is unlikely in practice that the variances will be known, in which case they too will have to be estimated from the data. One approach, used in Chapter 5, is to use a fully Bayesian approach, assigning probability distributions to all the unknown parameters, and then sample from their posterior distributions using MCMC techniques. Details of this approach can be found in Gamerman and Smith (1996) and Carter and Kohn (1994) with a description of the general framework of dynamic hierarchical models in Gamerman and Migon (1993).

2.6 Spatial modelling

The modelling of environmental variables in time and space has a considerable literature and in this section only a brief introduction to the methodology used is presented. Many of these methods are based on the idea of a stationary random field, and much of the initial theory is similar to that used in time series modelling applied to higher dimensions. Firstly, in the purely spatial context, methods for estimating the spatial correlation, or covariance, are discussed, including methods for interpolation, such as kriging. When there is also a temporal aspect to the data, the idea of separable models is considered, which greatly simplifies the modelling process, as well as models which aim to specifically model time-space interactions.

2.6.1 Stationary random fields

A spatial random field is a stochastic process over a region, $\{Z(s) : s \in D\}$ where s is a location in Euclidean space $\mathbb{R}^d$. For example, if x-y co-ordinates, elevation and time were considered, $d = 4$. This underlying process is not directly measurable, but realisations of it can be obtained by taking measurements, possibly with error, at a set of known locations (gauged sites), $s_1, \dots, s_S$. One way of expressing the random field is as a combination of the overall trend together with a spatial structure, for example

$$\begin{aligned} X_s &= Z_s + \epsilon_s \\ Z_s &= \mu_s + E_s \\ \mu_s &= \sum_{j=1}^p \beta_j f_j(d_s) \end{aligned} \tag{2.60}$$

where ϵ_s is measurement error, μ_s is the mean of the underlying process, modelled as a function of the location, information about which is contained in d_s , with associated coefficients, β , $E_s = \sigma_z^2 E'$ represents the spatial structure, where σ_E^2 is the spatial variance.

The observed data, $X_s, s \neq s', s \in D$ at the first level of the model are considered conditionally independent given the value of the underlying process, Z_s . The second level describes the true underlying process as a combination of an trend (mean), μ_s , and a random process, E_s , which has with spatial structure in its covariance, $COV(E_s, E_{s'})$. In a fully Bayesian analysis, the third level of the model assigns prior distributions to the hyperparameters from level 2. An alternative approach is to estimate them from the data (Diggle et al. (1998)).

In a purely spatial analysis, repeated observations at a specific location over time are treated as independent realisations of the underlying process. The mean and covariance function are then estimated from the data and used to describe the spatial trend and association and for prediction at unsampled locations.

As with the analysis of time series, or temporal, data the concept of stationarity, as described in Section 2.5.1, is fundamental to much of the methods applied in spatial analysis. In this context, a stationary process has no spatial trend, i.e. a constant mean, with the covariance between two points dependent only on their distance between them and not their actual locations. If a process is stationary, observations from any area within the region can be used to make inference about the overall underlying structure. The most common stationary assumptions are (1) second order or weak stationarity, and (2) intrinsic stationarity.

Second order stationarity

$$\begin{aligned} E(Z_s) &= E(Z_{s'}) && \text{for all locations } s, s' \in D \\ COV(Z_s, Z_{s'}) &= C(|s - s'|) = C(d) \end{aligned} \quad (2.61)$$

This implies that the variance is constant over the entire region and that the covariance between two locations depends only on the distance, d , between them, known as *isotropic*.

A special case is the *Gaussian random field* in which the realisations come from the multivariate normal distribution with mean zero, $E(Z_s) = 0$ and isotropic covariance, $COV(Z_s, Z_{s'}) = \sigma_s^2 \rho(|s - s'|)$. It has both strong stationarity (where the joint distribution between locations is constant over the entire region), which implies second order stationarity and using Gaussian assumptions makes the theory for estimation and prediction considerably more straightforward than might otherwise be the case.

A relaxed form of stationarity is *intrinsic stationarity*, which is based on the difference between two locations, the difference in means is zero and the difference in variances is defined through the *semi - variogram*.

$$\frac{1}{2}VAR(Z_s, Z_{s'}) = \gamma(|s - s'|) = \gamma(d) \quad (2.62)$$

This implies that the variance of the difference must be the same everywhere in the region, but does not require that the variance of X_s is constant over the entire region. If two locations

are close together, their difference would typically be small, and hence so would the variance of their differences. As the locations get farther apart, their differences get larger and usually the variance of the difference will increase. If the relationship between distance and covariance (or the variogram) differs according to the direction of the difference between the locations then the spatial structure is *anisotropic*.

The covariance function and the semi-variogram are both functions that summarize the strength of association between responses as a function of distance, and possibly direction. In both time series, and spatial analysis, it is necessary to make some assumptions about stationarity in order to estimate the correlation pattern and variance as there are no independent realizations. Instead, different parts of the region are used as if they were replications of the underlying process and then a common covariance function for all parts of the regions estimated.

At zero (distance) lag, the variogram is zero, as the value at a single spot is constant and has no variance. The curve may then rise and reach a plateau, indicating that past a certain distance, the correlation between two units is zero. Such a plateau will occur when the semi-variogram reaches the variance of X . Assuming second order stationarity, the relationship between the covariance and the semi-variogram is as follows,

$$\begin{aligned}\gamma(d) &= \frac{1}{2} \text{VAR}(Z_s - Z_{s'}) \\ &= \frac{1}{2} \{ \text{VAR}(Z_s) + \text{VAR}(Z_{s'}) - 2\text{COV}(Z_s, Z_{s'}) \}\end{aligned}\tag{2.63}$$

If $\text{VAR}(Z_s) = \text{VAR}(Z_{s'}) = \sigma_s^2$ and $C(Z_s, Z_{s'}) = C(d)$ for all s, s' , then

$$\begin{aligned}\gamma(d) &= \frac{1}{2} (2\sigma_s^2 - 2C(d)) \\ &= \sigma_s^2 - C(d)\end{aligned}$$

Therefore, at $d = 0$, the semi-variogram is $\gamma(0) = \sigma_s^2 - C(0) = \sigma_s^2 - \sigma_s^2 = 0$, where $C(0) = \text{COV}(Z_s, Z_s) = \text{VAR}(Z_s)$ and when d is large, i.e. when $C(d) = 0$, $\gamma(d) = \sigma_s^2 - C(d) = \sigma_s^2$. Valid variograms must be *conditional negative definite* (Cressie (1993)), namely $\sum_{i=1}^m \sum_{j=1}^m a_i a_j 2\gamma(Z_s, Z_{s'}) \leq 0$, for any $\sum_{i=1}^m a_i = 0$. Given an intrinsically stationary process, they must satisfy the condition,

$$2\gamma(d)/d^2 \rightarrow 0, \quad \text{as } d \rightarrow \infty\tag{2.64}$$

In practice, the semi-variogram is often preferred to the covariance function because of the relaxed rules of stationarity and also because it only uses pairs of locations d units apart, and doesn't involve the overall mean, so if there is a shift in the mean (trend) which is not explicitly modelled, the semi-variogram is likely to be less affected than the covariance function.

After data has been de-trended, and stationarity achieved, semi-variogram models can be applied. Although it is possible to create empirical estimates of the semi-variogram, these are often not

ideal for the purpose of modelling spatial data as there may only be a few estimates at particular lag distances and so estimates of the covariance may not be stable. In addition, the matrix of sample covariances, C , may not be positive definite and thus not a valid covariance matrix ($\text{VAR}(q'Cq) \geq 0$). Often, it will be better to pool the data at different lags and fit a smooth curve, using an appropriate model. Such semi-variogram models do not always produce well defined covariance matrixes for the underlying processes as often the exact variance function is not specified.

A common class of models used for variogram models is the Matern Class.

Matern class of models

$$\gamma(d|\theta) = \begin{cases} 0 & d = 0 \\ \theta_1 \frac{1}{2^{\theta_2-1}\Gamma(\theta_2)} \left(\frac{2d\sqrt{\theta_2}}{\theta_1}\right)^{\theta_2} K_{\theta_2} \left(\frac{2d\sqrt{\theta_2}}{\theta_1}\right) & d > 0 \end{cases}$$

Where $\theta_1 > 0$ is a scalable parameter controlling the range of the spatial correlation, and $\theta_2 > 0$ is the smoothness parameter. K_{θ_2} is a the modified Bessel function of order θ_2 . The exponential model is a special case of this, with $\theta_2 = 1/2$. This is the form used in Chapter 5 to describe the correlations between measurements of pollution at different monitoring sites, i.e. based on the inverse of the distance between them.

Exponential model

$$\gamma(d|\theta) = \begin{cases} 0 & d = 0 \\ \theta_1 \exp(-\theta_2 d) & d > 0 \end{cases}$$

The limiting case of the Matern class of models, when $\theta_2 \rightarrow \infty$, is the Gaussian model.

The semi-variogram of two observations taken at the same place is zero by definition since $\text{VAR}(Z_s - Z_{s'}) = 0$, however there may be ‘microscale variations’ and measurement error, so that two measurements taken at the same location are not the exactly same. This *nugget effect*, or nugget variance, is a discontinuity in the semi-variogram at the origin, the difference between zero and at a lag distance just greater than zero. This effect can be modelled explicitly, but this may not be necessary as the Matern class of models is very flexible and can take many different shapes near $d = 0$ using the smoothness parameter, θ_2 .

The simplest method of fitting a variogram model is that of least squares, although this assumes that the observations are assumed to be uncorrelated with each other and to have the same variance. This is unlikely to be the case when fitting a variogram as the estimates for different lags may be based on different numbers of pairs and will be correlated with each other. Simple weighted least squares can be used under certain stationarity and sampling conditions (Cressie (1993)), or the observations can be transformed to ensure that they are not correlated, i.e. $(X^* = (V')^{-1/2} X$,

where $V = VAR(Z)$, which results in performing generalised least squares regression.

2.6.2 Prediction at unsampled locations

If the goal of the spatial analysis is to predict unobserved values at locations where data has not been sampled then a simple approach would be to calculate a weighted average of all points within a certain neighbourhood of a chosen point, s_0 . The weights may be chosen to reflect the distance between an observed location X_s and the point in question, using the inverse of the squared distance i.e. $\sum_{i=1}^S d_i^{-2} X_s$. There are many possible ways of weighting the samples, all leading to different estimators. An estimator used extensively in the geostatistical literature is *kriging* (Cressie (1993)). In ordinary kriging, the estimator incorporates correlations among the Z_s into the weights for predicting Z_0 . The weights are based on the covariances among points in the sample and the covariances between sample points and the point to be predicted. To compute kriging estimates the covariances among all points, $\tilde{C}_{ss'} = COV(X_s, X_{s'})$, and between each of the observed points and the point to be predicted, $C_{s0} = COV(X_s, X_0)$, need to be estimated using a covariance function or semi-variogram. The kriging estimator is the optimal estimator in that it minimizes the prediction variance or mean squared prediction error, $E(X_s - \hat{X}_s)^2$. Ordinary kriging gives optimal predictions under the assumption that the mean is constant (but unknown), with the weights obtained by solving the system of *kriging equations*.

$$w = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \\ \lambda \end{bmatrix} = \begin{bmatrix} C_{11} & \dots & C_{1n} & 1 \\ \vdots & \dots & & \vdots \\ C_{n1} & \dots & C_{nn} & 1 \\ 1 & \dots & 0 & 0 \end{bmatrix} \quad (2.65)$$

where λ is a Lagrange multiplier that constrains $\sum w_i = 1$.

Only intrinsic stationarity is required to obtain the ordinary kriging estimator, and so the results can thus be derived using semivariograms instead of covariances.

If one or more covariates are measured along with the response variable, and if the cross correlation functions are known or can be estimated, then the covariates can be used to improve prediction, even when they are not measured at the same locations as the response variable. This is especially useful when there are few observations of the variable to be predicted, but many observations of some other variable which is highly correlated with the first. The co-kriging prediction is then a weighted average of nearby values of all the variables.

Using the hierarchical model (2.60), when prior probability distributions are assigned to the unknown parameters in $\mu_s = \sum_{j=1}^p \beta_j f_j(d_s)$, estimation from the resulting Bayesian hierarchical

model is known as *Bayesian kriging* (Cressie (1993)). The posterior distributions for the underlying level at point not included in the sampled locations is

$$\begin{aligned} p(Z_0|X) \propto p(Z_0, X) &= \int \dots \int p(X, Z, Z_0) dZ \\ &= \int \dots \int \left\{ \prod_{i=1}^S p(Z_i|X_i) \right\} p(Z_0|Z) p(Z) dZ \end{aligned} \quad (2.66)$$

This form can be further expanded to incorporate the conditioning on the parameters within the model, i.e. $p(Z|\beta, \theta)$, where β are the coefficients in the mean term and θ those in the variogram/covariance function. Samples from the posterior can be obtained by MCMC simulation, or in the case where $p(X|Z)$ and mean $p(Z)$ are Gaussian, the posterior will be Gaussian. Within a Bayesian framework, Berger et al. (2001) provide an interesting discussion of the choice of priors in spatial models, and in particular distinguish between noninformative priors which lead to proper posteriors and those that do not.

If the random field is assumed Gaussian, then *linear kriging* can be performed. In this case, conditional on Z , the data, X_s are mutually independent Gaussian random variables with expectations $\mu_s + Z_s$ and common variance, σ_x^2 . The kriging predictor, T , is then a linear function of Z (Diggle et al. (1998)). If the predictor $\hat{Z}_0$ is not linear in Z , as would be used if the outcome variable was in the form of counts, or proportions, then $E(X_s|Z_s) = \mu_s = g(\sum_{j=1}^p \beta_j f_j(d_s))$, for some known function, f and explanatory variables, d_s at s , with associated coefficients, β . In such cases, a *generalized linear predictor* can be calculated, Diggle et al. (1998).

Holland et al. (2000) use kriging as part of a hierarchical model to model the spatial effects of SO_2 in the Eastern U.S. At the first level, the logs of weekly measurements of SO_2 are smoothed as functions of the period and the temperature using a GAM. The model was used to estimate the true (unobserved and unknown) value of a Gaussian random field. In a non-Bayesian analysis, the analytical estimates of the variances are likely to be underestimated as the covariance parameters are assumed to be known, when in fact they are estimated. This problem is not easily remedied within a MLE framework, but by adopting a Bayesian approach, posterior distributions can be computed with allowance for the extra uncertainty. Holland et al. used MCMC methods to draw samples from the posterior distribution, although it was not a fully Bayesian analysis, as the estimates obtained from the GAM in the first level are treated as given, albeit with measurement error.

2.6.3 Non-stationary random fields

Perhaps the simplest approach to dealing with non-stationary data is to divide the region into a number of subregions, each of which has a sufficient number of points to estimate a (stationary) variogram. The overall spatial process and its covariance structure can then be represented by a weighted average of the locally stationary processes (Cressie (1986)). Ordinary kriging assumes

that the mean is constant over the entire region of study, but in *universal kriging* a non-constant mean is represented as a function of the location, a polynomial function of finite order, k . The assumption that $E(Z_s - Z'_s) = 0$ no longer holds and the method assumes that the variogram and the order k of the polynomial are known, unfortunately k is never known and has to be guessed from the data as is γ which has to be estimated from the residuals. Other approaches include *median polish kriging* and *intrinsic random functions of order k* (Cressie (1986)). The former involves modelling the mean term using the medians of rows and columns of a spatial grid in such a way that the resulting residual terms are stationary and can so can be used with traditional geostatistical techniques. The choice of the spatial grid is critical to this approach, and attempts are often made to define it in meaningful directions determined by the particular application, for example where geology might dictate exploration drilling. The latter has more general model assumptions than the kriging methods, and involves estimating k and the *generalized covariance function* from k^{th} order differences, however its use in practice is limited as it is not generally possible to estimate the covariance function and the order k of the differencing again has to be guessed, and not estimated from the data.

Spatial deformation

In the methodology described so far, two major assumptions have been required, stationarity and isotropy, which are unlikely to hold in many environmental problems. One approach to dealing with a non-stationary and anisotropic process was suggested by Sampson and Guttorp (1992), who describe a method which transforms the space where the monitoring sites lie into a latent space where stationarity and isotropy hold. The correlation structure of the spatial process is expressed as a function of Euclidean distances between site locations after a bijective transformation of the geographical co-ordinates system.

$$D(Z_s, Z_{s'}) = \gamma_\theta(|f(Z_s) - f(Z_{s'})|) \quad (2.67)$$

where $f(.)$ is the 1:1 transformation that expresses the spatial non-stationarity and anisotropy and γ_θ belongs to the parametric family of valid isotropic variograms with unknown parameters, θ . This maps the geographical co-ordinates in the G plane to the D plane, in which the distance between locations actually represents the difference in the variances, and thus the variogram. The transformation essentially stretches the G -space in regions of low spatial correlation and contracts it in regions of higher spatial correlation, so that an isotropic (stationary) variogram can be used.

Sampson and Guttorp use multi-dimensional scaling to produce the transformation, searching for configurations that fulfill a constraint akin to weighted least squares of distances in D space on spatial dispersions in the G plane, with weights proportional to the inverse of the squared distance in G space. There are important choices to be made in the initial configuration as

different configurations will often result in convergence to different final configurations. However, prior information is available from the geographic configuration in G space.

Recently, Damian et al. (2000) and Schmidt and O'Hagan (2000) have independently proposed Bayesian approaches which incorporate the uncertainty in the choice of transformation function and the effect this has on the subsequent estimation. Schmidt and O'Hagan use a Gaussian process prior for the deformation to D space whilst Damian et al. use a prior based on thin plate splines. Both approaches use MCMC simulation to sample from the posterior.

When the given locations have been mapped in D space, the transformation has to be extrapolated to the entire area for prediction at unmeasured sites, which will then be transformed back into G space. Sampson and Guttorp and Damian et al. achieve the extrapolation using thin plate splines, Schmidt and O'Hagan extend the covariance matrix of the multivariate normal to include the unmeasured locations (ungauged sites). For clarity, the notation indicating that the predictions of the underlying level, $Z|X$, are conditional on the observed data is dropped giving,

$$\begin{pmatrix} Z_0 \\ Z \end{pmatrix} \sim N_{S+1} \left(0, \begin{bmatrix} \Sigma_{uu} & \Sigma_{ug} \\ \Sigma_{gu} & \Sigma_{gg} \end{bmatrix} \right) \quad (2.68)$$

where Σ_{gg} is the covariance between gauged, Σ_{gg} between ungauged sites and Σ_{gu} between gauged and ungauged sites. The covariance matrix, Σ_{gu} was modelled using the sum of S exponential variogram models (2.65) of the form $\sum_{i=1}^S \theta_1 \exp\{-\theta_2 |Z_0 - Z_i|^2\}$. The mean and variance of the predictive distribution are therefore of the form given $(Z_0|Z) \sim N(\Sigma_{ug}\Sigma_{gg}^{-1}(Z - \mu), \Sigma_{uu} - \Sigma'_{ug}\Sigma_{gg}^{-1}\Sigma_{ug})$.

Le et al. (1997) use the spatial-deformation approach to provide a method for dealing with non-stationary multivariate random fields when data are missing by design. Within a Bayesian framework, they assume a multivariate Gaussian process with mean $Z\beta$ and covariance matrix, Σ , with conjugate priors being used for both (normal and inverse Wishart respectively). They partition the covariance matrix into gauged and ungauged sites as in (2.68) and interpolate using the predictive distribution, multivariate t , conditional on the observations at the measured locations. The prior covariance matrix is assumed to be of the form $\Phi = \Lambda \otimes \Omega$, splitting into that between sites and that between responses. A two stage procedure is used; first the parameters of the covariance between gauged sites, Λ_g are estimated using a modified EM algorithm, allowing for data missing by design, secondly Λ_g is extended to all measured and unmeasured sites, Λ , using the semi-parametric approach suggested in Sampson and Guttorp (1992). The use of the inverse Wishart may be a limitation, with a single parameter reflecting the uncertainty associated with the hyper-covariance, especially when data were missing by design and sites may measure different items and some sites may have very few observations compared with others.

2.7 Spatial-temporal modelling

The methods described in the previous section assume that there are temporally independent replications, however the likely presence of temporal correlation will lead to a space-time covariance structure. If the temporal aspect is incorporated into the structure of (2.60), then Z_{st} can be represented as a space-time random field, which can again be expressed in terms of a hierarchical model, for example

$$\begin{aligned} X_{st} &= Z_{st} + \epsilon_{st} \\ Z_{st} &= \mu_{st} + E_s \\ \mu_{st} &= \int w_{\Delta}(\mu_{u,t-\Delta}) du + \eta_{st} \end{aligned} \tag{2.69}$$

where $\epsilon_{s,t}$ is an independent random, or measurement, error term, $\mu_{s,t}$ is a space-time mean field (trend) and E_s is a spatial process, independent of time. The third level indicates that the mean at μ_{st} depends dynamically on the values $\mu_{u,t-\Delta}$, for a spatial area, u and locations in time within Δ of t . So, if s is close in space to s' , then the dynamic aspect will depend on overlapping neighbourhoods and thus μ_{st} and $\mu_{s't}$ will be highly correlated. For the model to be consistent, the covariance structure should be independent of the time interval, Δ . In most applications, modelling the entire spatial-temporal structure will be impractical because of the high dimensionality. A number of approaches have been suggested to deal with this directly, and these are discussed in Section 2.7.2.

Models where the space-time structure is modelled jointly are known as *non-separable* models. A simpler, though not always appropriate approach, is to consider *separable* models, in which the temporal and spatial components are modelled separately, with no interaction. These models impose a particular type of independence between space and time components. If $\rho_{ss'}$ denote the correlation between observations Z_{st} and $Z_{s't}$, $\rho_{tt'}$ the correlation between Z_{st} and $Z_{st'}$, and $\rho_{ss',tt'}$ the correlation between $Z_{s,t}$ and $Z_{ss',tt'}$. Then the correlation of a separable model is

$$\rho_{ss',tt'} = \rho_{ss'} \times \rho_{tt'} \tag{2.70}$$

2.7.1 Separable models

In terms of model (2.69), the dynamic term in the third level would separate into $\mu_{st} = \sum_{i=1}^p \phi_i \mu_{s,t-i}$, leaving the spatial component in E_s , which is independent of time. Due to the reduction in computation that comes with this approach, earlier work on space-time modelling tended to be based on analysing the temporal and spatial aspects separately, and then to combine the chosen models in a single separable model.

Sun et al. (2000) use the approach suggested by Le et al. (1997) to develop a spatial predictive distribution for the space-time field of daily ambient PM_{10} in Vancouver, Canada. However, for simplicity, they analysed each of the monitoring sites separately and chose a single, $\text{AR}(1)$, model to represent the temporal structure. They identify the possibility that spatial correlation between sites might ‘leak’ into the lagged values of the series, due to modelling each site univariately. This would not have happened if a multivariate auto-regression was used, but such an approach may be infeasible with a large number of monitored and unmonitored sites. They then used the de-trended residuals to obtain posterior distributions of the covariance matrix, which was then extended to include unmonitored sites using the semi-parametric approach suggested by Sampson and Guttorp (1992). Guttorp et al. (1994) modelled the spatial covariances of hourly ozone levels using the Sampson and Guttorp spatial covariance approach, and allowed the parameters of the model to vary as a function of time of day.

Haslett and Raftery (1989) modelled wind velocity measurements in Ireland using a combination of kriging and ARMA time series models (see Chatfield (1996)) to predict values at gauged sites that have short runs of data. They use kriging estimates of the variance of the random field, using an exponential model for the covariance-distance relationship. They built space-time models by linking models for the observed data, X , directly at each of the sites through a spatially correlated set of innovations whose covariance structure was derived from the underlying continuous spatial process, which was assumed to be realised independently at each t . They then combine the spatial and temporal aspects in a model of the form

$$\begin{aligned} Z_{it} &= \mu_i + \nabla^{-d} \phi(B)^{-1} \theta(B) \epsilon_{it} \\ \epsilon_t &= (\epsilon_{1t}, \dots, \epsilon_{mt})' \sim \text{MVN}(0, \sigma_\epsilon^2 E_s) \end{aligned} \quad (2.71)$$

where the matrix E_s represents the spatial correlation, and does not depend on time. ∇^{-d} is the Binomial expansion of $(1-B)^{-d}$, where B is the backshift operator and d can be non-integer value, leading to fractional differencing and long memory processes. They assume a common ARIMA model for each site and estimation is performed using MLE techniques (Chatfield (1996)). In terms of model (2.69), the observed data is modelled directly, i.e. there is measurement error term in level one. A limitation of this method is that the estimator cannot be used for prediction at sites where there are no data at all, and is thus useful only for completing datasets from sites where data has actually been collected.

In an early Bayesian application, Handcock and Wallis (1994) consider the spatio-temporal modelling of winter temperature data but their approach was to carry out separate spatial analyses in each year using a Gaussian random field model. The mean and covariance parameters of this model were then examined temporally and found to be stable, which they concluded was justification for the use of a simple model that essentially treated spatial and temporal aspects separately. They developed a stationary spatial-temporal random field for mean temperature using 88 weather

stations in northern U.S. They first develop a Gaussian random field using a Bayesian kriging approach as a function of latitude, longitude and elevation, examining variograms in different directions to look for anisotropy. Having produced a spatial structure, they consider the yearly temperatures by examining each site separately, fitting AR(1) models and finding that only 3 out of the 88 were significant, leading to the conclusion that there was little evidence of short term temporal dependency. They then examined long term dependency using ARIMA models with non-integral or fractional differencing, finding little evidence of long memory dependence and also, that the spatial structure was changing over time.

2.7.2 Non-Separable models

The problems of high-dimensionality in modelling the entire space-time structure were addressed by Mardia et al. (1998) and Wilke and Cressie (1999) who reduce the dimensionality of the mean term in similar fashion to principal components analysis (see, for example, Chatfield and Collins (1980)), to an orthonormal sequence of deterministic spatial functions. Mardia et al. (1998) describe these functions as principal and trend fields, with the trend fields being functions of the co-ordinates of the sites and the principal fields selected from kriging estimates for a set of ‘normative’ sites. The hierarchical model (2.69) is now expressed as,

$$\begin{aligned} X_{st} &= Z_{st} + \epsilon_{st} \\ Z_{st} &= \sum_{j=1}^p A_{jt} \phi_j(s) + E_s \\ A_t &= H A_{t-1} + J \eta_t \end{aligned} \tag{2.72}$$

where either H and/or J must be non-diagonal in order to incorporate spatial structure into the temporal component. Mardia et al. (1998) term their method the ‘kriged Kalman filter’ (KKF) and outline a likelihood-based estimation strategy. Wilke and Cressie (1999) give an (approximate) Bayesian estimation approach. Besides the estimation techniques, the difference between the two approaches, is that there is no measurement error term (ϵ_{st}) in the first level of the KKF. This results in the error term consisting of two components, $\epsilon_{st} + E_s$, and as there is no attempt to incorporate the inherent variability in predicting the underlying process from the data, introducing the possibility of oversmoothed estimates (Cressie and Wilke (1998)). The KKF computes $E(\sum_{j=1}^p a_{jt} \phi_j(s) | Z)$ whereas Wilke and Cressie compute $E(X_{st} | Z)$.

Wilke et al. (1998) present a fully Bayesian hierarchical model in which they approach the problems of high dimensionality by explicitly modelling parameters lower in the hierarchy. They expand the second level of (2.69) to include site specific means, μ_s , and long term temporal patterns, $\beta_{1s} \cos(\omega t) + \beta_{2s} \sin(\omega t)$. They use a vector autoregressive process (VAR) (Chatfield (1996)) to model the dynamic aspect of the space-time interaction, A_{st} , i.e. $A_t = H A_{t-1} + \eta_{st}$. The updating parameter, H_t , was simplified by using a ‘nearest neighbour’ VAR (Cressie (1993))

in which $A_{st} = a_{st}A_{s,t-1} + bA_{s^1,t-1} + cA_{s^2,t-1} + dA_{s^3,t-1} + eA_{s^4,t-1} + \eta_{st}$, where $|s^i - s| = 1$, the four locations within distance one of s . In this form, the autoregressive parameter for the location in question, a_{st} , varies spatially, but b, c, d and e do not. Consequentially, the model has no spatial interaction between A_{st} and $A_{s't} : s \neq s'$. Estimation was achieved using MCMC which was simplified to some extent by using Gaussian error terms at each level, except for the variance hyperparameters, which were inverse Gamma.

Brown et al. (2001) considered the spatio-temporal modelling of radar-rainfall data in Lancashire, UK. They consider both separable and non-separable models. In the former, they describe a state space model with observation equation $X_t = A + B_t X_t + \epsilon_t$ and second level being a first order vector autoregressive process, VAR(1) (Chatfield (1996), Cressie (1993)), $A_{st} - \mu_{st} = H(A_{t-1} - \mu_{st}) + J\eta_{st}$. The covariance structure can be expressed as $\Gamma_k = H^k(H\Gamma_0 H' + JJ')$, with J influencing the purely spatial covariance, and H describing how the spatial covariance decays at larger time lags.

Brown et al. (2001) use a non-separable model in which the integral representing the space-time field in the third level of (2.69) is replaced by fitting a Gaussian density to the autoregressive parameters, A_{ts} , which ‘blurs’ the field, as described in Brown et al. (2000), allowing interaction between different sites at different time points. If two locations, s and s' are close in space, then the values of A_{st} and $A_{s't}$ will be calculated using values from overlapping neighbourhoods, leading to the outcomes being highly correlated. Fitting the model, which they achieve using MLE, requires the covariance matrix of all the grid points not just the gauged points, making the Kalman filter difficult. Instead they construct a complete multivariate model with the likelihood evaluated by inverting the covariance matrix. As previously mentioned, in most applications the size of the covariance matrix will prohibit using it directly, initiating the use of data reduction methods. In this particular application, the problems of dimensionality were lessened by being able to delete data points where there was no rainfall, as they added no information. The resulting covariance matrix, whilst still large (2000×2000) was then invertible.

Recently, Tonellato (2001) used a univariate auto-regressive process measured with error, within a Bayesian framework, to model hourly measurements of a CO from a small number of sites. The approach used is similar, albeit for a single pollutant, to that proposed in Chapter 5, where further details can be found.

A spatial-temporal model for hourly ozone measurements was developed by Carroll et al. (1997). The model, $X_{s,t} = \mu_t + \epsilon_{s,t}$ combined a trend term incorporating temperature and hourly/monthly effects, $\mu_t = \alpha_{hour} + \beta_{month} + \beta_1 temp_t + \beta_2 temp_t^2$, which was constant over space, and an error model in which the correlation in the residuals was a nonlinear function of time and space. In particular the spatial structure was a function of the lag between observations, $COV(\epsilon_{s,t}, \epsilon_{s',t'}) = \sigma^2 \rho(d, v)$,

where d is the distance between sites and $v = |t' - t|$ is the time difference, with the correlation being given by

$$\rho(d, v) = \begin{cases} 1 & d = v = 0 \\ \phi_v^d \psi_v & d \text{ otherwise} \end{cases}$$

where $\log(\psi_v) = a_0 + a_1 v + a_2 v^2$ and $\log(\phi_v) = b_0 + b_1 v + b_2 v^2$. The correlation of the random field is thus a product of two factors, the first, ψ_v^d depends on both the time and space, the second only on the time difference. Unfortunately, as Cressie (1997) pointed out, this correlation function is not positive definite. Using results from the model, there were occasions when $COV(Z_{st}, Z_{s't'}) > COV(Z_{st}, Z_{st})$. This highlights a genuine lack of a rich set of functions that can be used as space-time correlation functions.

Chapter 3

Modelling the association between mortality and long term exposure to air pollution

3.1 Introduction

Recently, there has been an increase in the evidence of an association between exposures to air pollution and human morbidity and mortality. Much of this evidence, however, comes from studies of acute health effects, that is, from associations between short-term exposures to air pollution and mortality on either the same day or preceding days. A wide range of pollutants have been implicated in these effects, but particular attention has tended to focus on particulate matter, measured in various ways (e.g. as PM_{10} , $PM_{2.5}$, TSP, black smoke) and oxides of sulphur (SO_x). Further details of the acute effects of air pollution can be found in Chapter 5.

Rather less attention has been given to investigating chronic effects of air pollution, i.e. the association between health outcomes and long-term exposures to air pollution, possibly over several years, due to the lack of availability of suitable data and problems of confounding, but it is uncertain to what extent, if at all, findings from studies of short term effects can be extrapolated to longer term (chronic) effects. Whilst some or all of the effect of acute exposures may be attributed to 'mortality displacement' (i.e. bringing forward health events which were likely to occur within a short time), chronic exposures may be implicated in fundamental disease causation (e.g. by sensitising people in early life to respiratory allergens or by provoking cell mutation).

Almost all the chronic studies reported to date have also focused only on concurrent exposures: that is, on the associations between health outcome and pollution levels measured in the same, or very recent, years. As such, these studies take no account either of possible latency effects (e.g. due to exposures earlier in life), or of the effects of cumulative exposures over many years. The

approach presented here is novel in that it addresses both of these issues at a much higher level of geographical resolution that has been possible before, and additionally allows the effects to vary over calendar time. This is achieved by using health outcomes from distinct time periods, each of which is associated with previous exposure over a number of years. By considering sub-divisions of these exposure periods, the effects of different lags can be examined. The model that is used here has four separate periods of health outcome at the ward level, with SO_2 exposure data over the previous sixteen years, within which, the separate effects of exposure periods of four years can be assessed. However, the model is a general one and can be used with any time scale.

The structure of this chapter is as follows. In Section 3.1.1, previous studies of the effects of long term exposure to air pollution on health are reviewed. Section 3.2 describes the health, exposure, socio-economic and demographic data used in the analyses. Section 3.3.1 provides a description of how cumulative exposures are modelled within the general model framework. The analyses consist of two approaches; the first considers four independent time periods, the second uses data from all of the time periods in a combined analysis. The models and results for these two sets of analyses are presented in Sections 3.4 and 3.5 respectively.

3.1.1 Epidemiology of chronic health effects

In comparison to that of short term effects, evidence of chronic effects of air pollution on health is far weaker. Typically disease occurrence is compared cross-sectionally across two or more contrasting environments (e.g. polluted/unpolluted cities), though interpretation is greatly complicated by the potential for unmeasured confounding, including the effects of smoking.

Key studies have concentrated on mortality, which is the focus here, though various studies of morbidity and lung function among children and adults in more and less polluted environments have been carried out, tending to find higher symptoms reported and worse lung function in the more polluted areas. Examples include the sixcities study (Dockery et al. (1989)) and the National Health and Nutrition Examination Studies (NHANES I and II) (Chestnut et al. (1991) and Schwartz (1989)) which analysed childrens lung function at the individual level. Both found small negative associations between lung function and particulate air pollution (less than 2% decline per $10\mu\text{gm}^{-3}$ increase in PM_{10}), after controlling for age, sex, race, height and weight. In a study of data from 24 U.S. cities, risks of relatively large deficits in lung function (15% or more) were observed in the most polluted cities (Raizenne et al. (1996)). There have also been several studies of the associations between air pollution and chronic respiratory diseases. These include the six cities study (Dockery et al. (1989)), NHANE I (Schwartz (1989)), the 24 cities study (Dockery et al. (1993)) and a study of Seventh Day Adventists (who had never smoked) in California (Euler et al. (1997)). All of these found significant associations (typically a 10-25% increase per $10\mu\text{gm}^{-3}$) between changes in particulate air pollution and respiratory symptoms.

For further details of these studies and a comprehensive review of the long-term effects of air pollution, see Wilson and Spengler (1996) and COMEAP (1995).

Possibly the earliest study of the effect of air pollution in Britain was that by Daly (1959) in which data on coal deliveries was used to provide an index of exposure to air pollution for 83 county boroughs, covering the major urban areas in England and Wales. Significant and strong correlations were found between this index and mortality rates for a range of causes, including bronchitis, pneumonia, respiratory tuberculosis, lung cancer and all non-respiratory diseases for the period from 1948-54 (different periods were used for some health endpoints because of data limitations). Attempts were made to adjust for socio-economic confounding, by considering the partial correlation coefficient after controlling for social class, overcrowding, population density and education (which also correlated strongly with both the index of pollution and mortality rates). This suggested that about one-third of the variation in mortality might be attributed to the independent effect of air pollution, equivalent to about 7,500 deaths per year for bronchitis. When relationships were also examined for lung cancer, incorporating data from a national survey to allow for smoking habits, no significant effects were found. A decade later, Gardiner and Crawford (1969) conducted a similar study in 61 county and London boroughs in England and Wales, covering the seven-year periods 1948-54 and 1958-64. Ranking of the boroughs in terms of their death rates was similar for the two study periods. Multiple regression techniques were used to analyse the combined effects of social factors (based on a composite measure of nine socio-economic indices), air pollution (measured on the basis of domestic coal usage), water hardness, latitude and rainfall on mortality rates for all deaths, lung cancer, stomach cancer, bronchitis and cardiovascular diseases. For males, air pollution was significant in models for all health outcomes; for females, the effects were generally weaker and often non-significant. Chinn et al. (1981) carried out a further study of 116 county and London boroughs in England and Wales, covering the years 1969-73. Again, multiple regression methods were used to examine relationships between mortality and a range of social and environmental factors, including monitored smoke and SO₂ concentrations. For a subset of the boroughs, associations with domestic coal usage were also examined. No consistent associations were found between either smoke or sulphur dioxide and mortality, in the combined models. When analysed independently, however, the index of coal consumption did show a weak relationship with all cause mortality in males and a somewhat stronger relationship with stomach cancer in females. The conclusion was that the results suggest 'a declining effect on mortality as pollution levels have fallen since the mid-1950s'.

These were all ecological studies carried out at relatively broad geographical scale: the old county boroughs, for example, typically contained between 50,000 and 1 million people. They all faced severe methodological problems, not least in acquiring valid estimates of air pollution and in allowing for socio-economic confounding. In an attempt to resolve these problems, designs including a longitudinal component have been used in two major studies in the U.S. The Harvard

Six Cities Study (Dockery et al. (1993)), followed a cohort of over 8000 adults in six cities in the U.S., from 1974 to the late 1980s. Exposure data were obtained for each city using a centrally-located pollution monitor, and health and individual level-confounder data were acquired through questionnaires and interviews. Mortality data were also obtained from the National Death Index for the years 1979–89. After controlling for confounding at city level, strong associations were found between mortality rates in the six cities and concentrations of fine ($\text{PM}_{2.5}$, PM_{10} or PM_{15}) and sulphate particles: across the cities, 26% higher all-cause mortality was found for PM_{10} concentration that varied from $18.2 \mu\text{gm}^{-3}$ (for the least polluted city) to $46.5 \mu\text{gm}^{-3}$ (for the most polluted). A recent reappraisal of the study results, with stronger control for socio-economic confounding, has confirmed these associations (Krewski et al. (2000)). Increases in risk of a similar magnitude were obtained by the American Cancer Society (ACS) study of 151 cities (Pope et al. (1995)). This used data for over 550,000 adults over the period 1982–89. Air pollution concentrations for sulphates were obtained for all cities from the national monitoring network; for fine particles data were available for 50 cities. When adjusted for smoking, education and other social factors, all-cause mortality rates were found to be 15% higher in the most polluted city, compared to the least, for sulphate particles, and 17% for fine particles. However, care is needed in extrapolating from findings in the USA to Europe or other areas, where the composition of air pollution, as well as climatic, socio-economic and lifestyle factors are very different (COMEAP (2001)).

To date, almost all of the studies of the chronic effects of pollution have considered only concurrent exposures. Over recent decades, air pollution concentrations have generally fallen, in response to industrial and technological changes and more rigorous regulation. At the same time, the character of air pollution has changed markedly, as domestic and industrial coal-burning has declined and emissions from road traffic have increased. Health risks determined on the basis only of current or recent exposures may therefore be misleading, especially for older age groups who may in the past have been subject to very different exposure regimes. One recent study which has considered historic exposures and possible latency effects is that reported by Nyberg et al. (2000). A case control design was used to examine relationships between air pollution (NO_x/NO_2 and SO_2) during the previous 30 years and lung cancer in Stockholm, Sweden. Results showed that the strongest associations with modelled NO_2 concentrations occurred for a lag of 20 years; little association was observed with SO_2 .

3.2 Data description

Since 1981, registration of all deaths in Great Britain have included the postcode of residence. Using geographical information system (GIS) techniques, this mortality data can be linked to

electoral wards and hence to the underlying populations at risk, and to socio-economic data from census. There is a long-established network of monitoring stations providing standardised data across the UK on major pollutants, especially black smoke and sulphur dioxide (SO₂). These monitoring stations can also be linked to wards through the National Grid reference of the site. This allows analyses of the association between health and air pollution to be performed at a much higher geographical resolution, i.e. at the level of electoral wards, than previous ecological studies of chronic effects of air pollution.

3.2.1 Health, socio-economic and demographic data

The health data consist of mortality counts for respiratory diseases (ICD9, 460-519). These data were extracted for all ages by ward from national postcoded mortality data, by age and sex, for four time periods: 1981–84, 1985–88, 1989–92 and 1993–96, henceforth referred to as *health periods*. Smoking is known to be a major risk factor for cardio-respiratory illness and it is known that smoking habits vary with social class (Kleinschmidt et al. (1995)) and may therefore correlate with pollution levels, and act as a potential confounder. Although routine data on smoking levels at small area level are not available in Great Britain, strong correlations have, however, been demonstrated on several occasions between smoking rates and the Carstairs index of deprivation, which has also been shown to be a strong predictor of disease risk (Carstairs and Morris (1989)). The index is derived from a weighted aggregation of data on four census variables: unemployment, overcrowding, car ownership and social class. It was computed for each ward using data from the 1991 census. Expected numbers, standardised by age and sex, were calculated for each ward using national mortality counts and population data from the 1991 census. The relationship between all cause mortality, cardiovascular and respiratory deaths and Carstairs score for the four health periods can be seen in Figure 3.1. Strong positive relationships are observed for all three, and in particular for respiratory mortality (i.e. higher mortality rates in the more deprived areas).

Summaries of the observed and age-sex standardised expected numbers together with a measure of the relative risk in each health period are shown in Table 3.1. The relative risks in each period are greater than one, due to the socio-economic profile of the wards included in the study, which is more deprived than if a random sample was taken of all wards in Great Britain. In many small-area epidemiological studies, for example Elliott et al. (1992), Elliott et al. (1996), Dolk et al. (1997), quintiles of Carstairs score have been used to control for deprivation, with the quintiles being defined on a national level. Using this approach here would mean that over 70% of the wards in this study lie in the two most deprived quintiles, which would greatly reduce its discriminatory power, in that there would be little to differentiate between a large number of the wards as they would be assigned the same score (quintile). For this reason, adjustment for deprivation is based

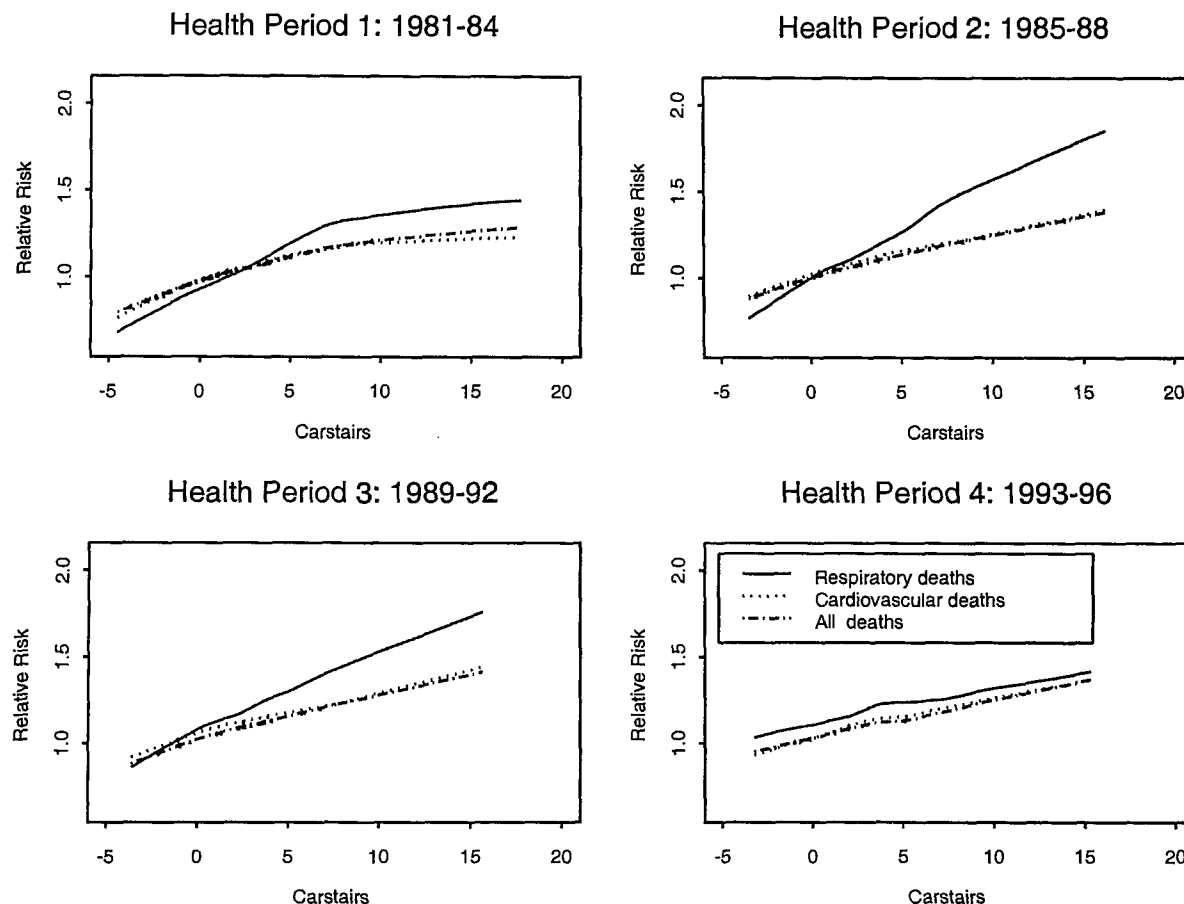


Figure 3.1: Relative risk (adjusted for age and sex) of all cause, respiratory and cardiovascular mortality (for all ages) against measure of socio-economic deprivation (Carstairs' score), for four health periods.

on deciles of Carstairs score, with the deciles being calculated using the wards under study.

The apparent decrease in the observed counts in the second and third periods is likely to be an artefact due to changes in the way ICD-9 rules for selecting the underlying cause of death were interpreted in England and Wales (known as 'Rule 3') (Rooney and Devis (1996)). From 1984, this rule was interpreted differently than before in England and Wales, resulting in an abrupt fall in mortality rates for pneumonia and other less common conditions regarded as terminal, and a corresponding apparent rise in mortality from many chronic conditions, including diseases of the nervous system and mental disorders. The anomaly was reversed by the introduction of an automated system for coding the cause of death in 1993. However, the expected numbers in the two affected periods are based on the observed counts for the entire country (via standardisation), and as a consequence, they too are reduced, leaving the estimates of risk unaffected.

Table 3.1: Summary of respiratory mortality at ward level in four health periods, (1981-84, 1985-88, 1989-92, 1993-96) at ward level. Means, standard deviations, medians, ranges and interquartile ranges are given for observed (Obs) and age-sex adjusted expected numbers of cases (Exp), together with measures of relative risk (O/E ratio) and the number of wards (N).

	N	Min.	25%	50%	75%	Max.	Mean	SD
1981-84								
Obs	391	1.0	36.0	61.0	91.0	247.0	67.6	43.5
Exp	-	0.5	37.7	59.3	79.4	189.4	60.7	31.0
O/E	-	0.0	0.8	1.0	1.3	4.7	1.1	0.5
1985-88								
Obs	222	2.0	32.0	54.0	78.0	187.0	57.6	33.8
Exp	-	0.9	27.3	45.8	61.2	153.6	47.1	24.8
O/E	-	0.4	1.0	1.2	1.4	4.8	1.2	0.5
1989-92								
Obs	191	1.0	30.2	57.5	79.8	177.0	60.1	36.0
Exp	-	1.4	29.3	47.0	61.6	160.9	47.5	24.9
O/E	-	0.2	1.0	1.2	1.5	2.9	1.3	0.4
1993-96								
Obs	123	3.0	41.5	68.5	107.5	217.0	77.9	46.6
Exp	-	1.9	40.3	63.5	83.3	220.2	64.0	33.9
O/E	-	0.5	1.0	1.1	1.5	2.1	1.2	0.3

3.2.2 Air pollution data

The data on annual concentrations of SO₂ were obtained from the UK National Air Quality Information Archive (data on black smoke was also obtained, but the analysis of those data were not include in this thesis). This is a routinely maintained data set prepared by the National Environmental Technology Centre (NETCEN) on behalf of the UK Department of the Environment, Food and Rural Affairs (DEFRA), which collates information from all the national air quality monitoring networks. The monitoring network, which measures both SO₂ and black smoke, was established in the early 1960s, and by 1971 included over 2000 sites. As levels of black smoke and SO₂ pollution have declined, the network has been progressively rationalised and reduced and currently comprises approximately 220 sites. Over this period, many sites have also been moved or replaced in order to reflect changing patterns and levels of pollution, and to reduce redundancy in the network. There is the possibility of selection bias, if the monitoring sites are kept in polluted areas; however this would only really be a problem in the unlikely event that the areas containing monitoring sites were also chosen because of high mortality, which would lead to informative sampling. The network is non-automated, and concentrations are measured (usually on a daily basis) by acid titration for SO₂ and by measurement of reflectance from a filter for black smoke. Air pollution statistics for each site are reported in 'pollution years' (April-March) and include mean concentration, standard deviation and number of valid reporting days. Data were obtained for all sites operating within the study period, a total of 3016 sites, and the data

filtered to remove all sites which were located in non-residential areas, which fell outside the study area (defined as Great Britain, meaning that Northern Ireland sites were excluded) or for which air pollution data were lacking. This resulted in a valid dataset of 2328 sites measuring SO₂, which were then matched to electoral wards, by dropping the site co-ordinates onto a digital map of ward boundaries, within a GIS. Ward-level pollution values for each year were generated as follows: where a single site operated within a ward, annual values for this site were used to give a ward-level annual concentration; where one site was replaced by another, at a different location in the same ward, data for the two sites were joined 'end-on' to provide a single, continuous record for the ward; where two or more sites operated concurrently within a ward, the site with the shorter run of data was discarded and data only for the longer-running site were used to provide a ward-level concentration.

The ward data were then grouped into 4-year *exposure periods* (1966-69, 1970-73,...,1990-93), and averaged. Averages were only computed where at least two years in the exposure period had at least 75% of days (i.e. 274 days) of valid data; otherwise the ward was dropped from that exposure period. Measurements of SO₂ are given in parts per billion (ppb).

Table 3.2 shows summary statistics for SO₂ concentrations, by exposure period, for the wards used in the analysis. Average concentrations for SO₂ declined rapidly during the first five exposure periods, from 43.7 (18.9) ppb for 1966-69 to 17.7(5.5) ppb for 1982-85. They increased to 25.5 (7.7) ppb for 1986-89, then fell again to 12.2 (4.8) for the last exposure period, 1990–1993. Histograms showing these changes in levels of SO₂ can be seen in Figure 3.2. Comparisons of ward level pollutant concentrations between different exposure periods can be seen in Figure 3.3 and Table 3.3 and show strong correlations (0.79-0.93) for consecutive blocks (i.e. a 4 year difference); decreasing with increasing time interval (0.55-0.67 for a 12 year difference).

Figure 3.4 gives an initial look at differences in the relative risks calculated for each ward (adjusted for age, sex and deprivation) associated with different (lagged) exposures of SO₂. There appears to be a positive association between the mortality data for first health period (1981-85) in relationship to exposure data for lags of 0-4, 5-8, 9-12 and 13-16 years, with similar associations being observed in the other health periods (not shown).

3.3 Statistical Methodology

3.3.1 Modelling the cumulative effects of exposure

In order to assess the effect of air pollution on health, models are required that relate risk to the exposure, both in terms of the degree of exposure and the time over which exposure occurred. In cohort studies of individuals, such models need to account for the duration of exposure, time since

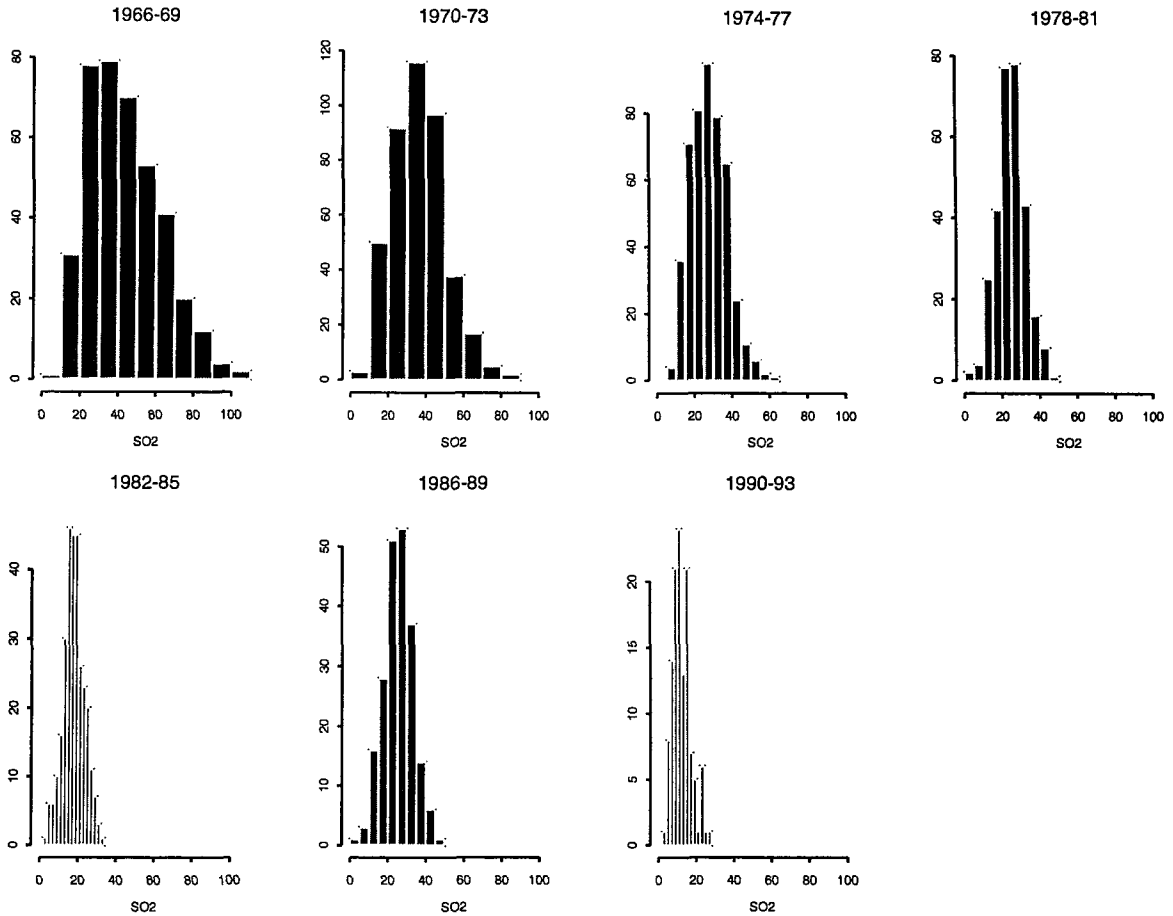


Figure 3.2: Histograms of SO_2 exposure in seven exposure periods (1966-69, 1970-73, 1974-77, 1978-81, 1982-85, 1986-89, 1990-93).

first exposure, time since exposure ceased and the age at first exposure (Breslow and Day (1980), Waternaux et al. (1989)). For the development of carcinogenesis, complex multistage models have been developed that use well defined dose-response relationships (Dewanji et al. (1999)). However, when using aggregated daily mortality counts for a specific day, e.g. day or health period, and specified area, detailed exposure histories and other information are generally not available.

Considering for ease of illustration, a generic area, if Y_t is the health outcome at time t , e.g. the number of respiratory deaths or on a single day or other period of time, and the true exposure history is $X(u), 0 \leq u \leq t$, then the outcome is modelled as a function of the exposure history.

$$E(Y_t) = f(X(u); \quad 0 \leq u \leq t) \quad (3.1)$$

As true personal exposure to air pollutants is unmeasurable over lifetime, as they depend on ambient levels and integrated time-activity, the term 'exposure' here relates to cumulative ambient outdoor concentrations of air pollutants, measured at the aggregate area level. The summaries of

Table 3.2: Summary of SO₂ measurements at ward level from the seven pollution periods (1966-69, 1970-73, 1974-77, 1978-81, 1982-85, 1986-89, 1990-93): means, standard deviations, medians, ranges, interquartile ranges and number of wards (N).

	N	Min.	25%	50%	75%	Max.	Mean	SD
1966-69	391	9.8	28.4	41.0	55.6	105.3	43.7	18.9
1970-73	420	5.9	26.0	35.8	44.9	88.7	36.3	13.9
1974-77	475	5.3	20.7	27.6	34.2	64.6	27.9	9.5
1978-81	296	4.3	20.3	25.0	29.4	50.0	24.9	7.5
1982-85	210	3.0	14.2	17.4	20.8	32.6	17.7	5.5
1986-89	123	4.8	21.0	25.6	30.4	50.0	25.5	7.7
1990-93	123	3.3	9.0	11.5	15.1	26.5	12.2	4.8

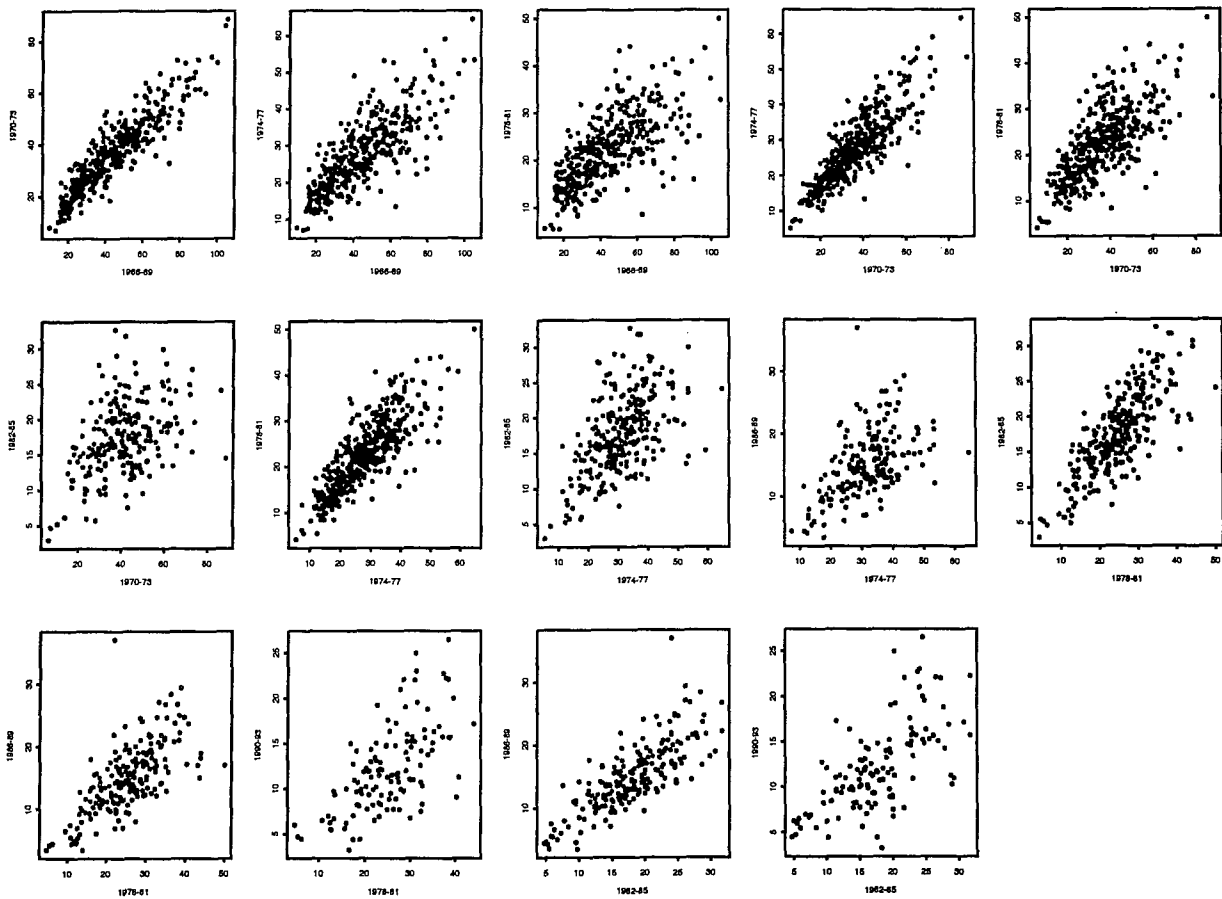


Figure 3.3: Correlations between SO₂ measurements in the seven exposure periods (1966-69, 1970-73, 1974-77, 1978-81, 1982-85, 1986-89, 1990-93).

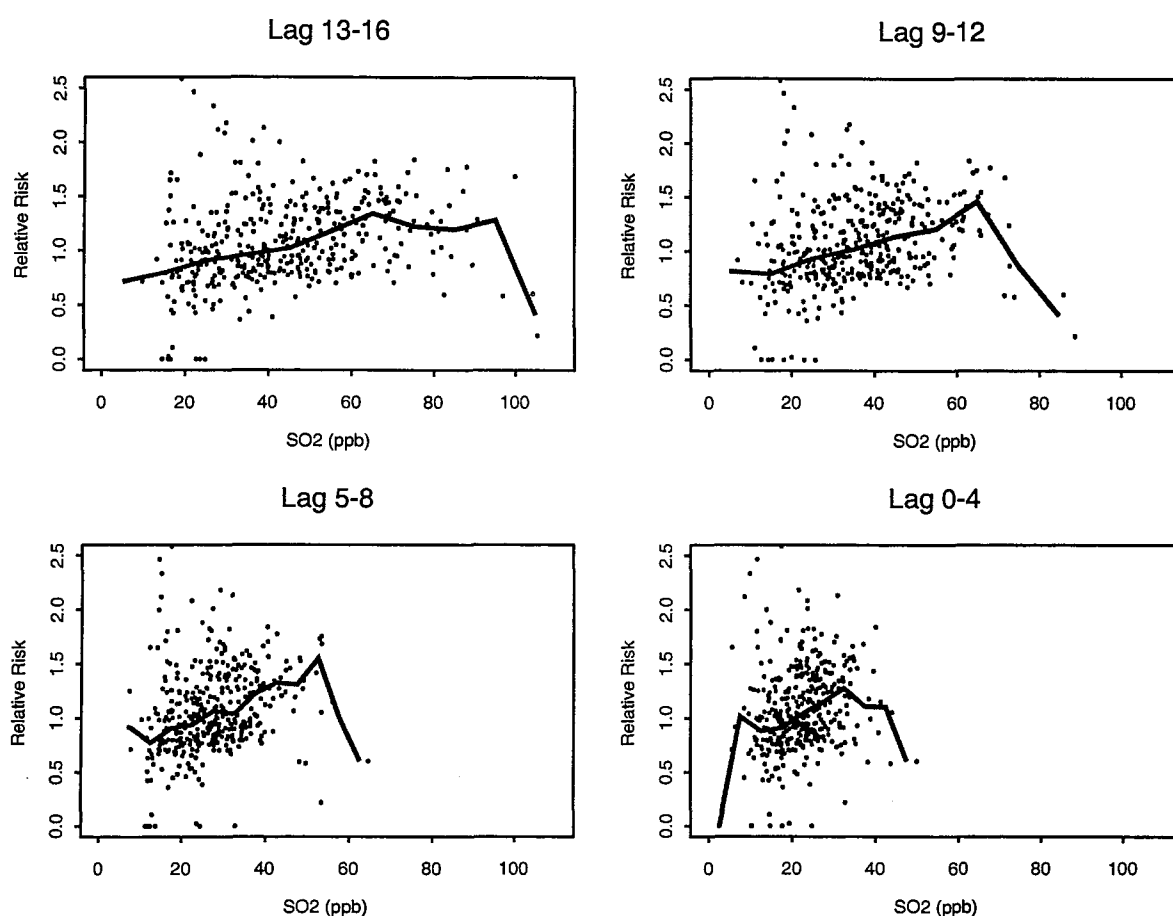


Figure 3.4: Relative risk of respiratory mortality (adjusted for age, sex and deprivation) , for all ages, against lagged exposure of SO₂. Mortality data is for first health period (1981-85), exposure data for lags of 0-4, 5-8, 9-12 and 13-16 years). The line joins median relative risks of 20ppb wide categories.

Table 3.3: Correlations between SO₂ measurements in the seven exposure periods

	1966-69	1970-73	1974-77	1978-81	1982-85	1986-89	1990-93
1966-69	1.00	0.93	0.80	0.67	-	-	-
1970-73	0.93	1.00	0.86	0.71	0.55	-	-
1974-77	0.80	0.86	1.00	0.87	0.71	0.64	-
1978-81	0.67	0.71	0.87	1.00	0.86	0.79	-
1982-85	-	0.55	0.71	0.86	1.00	0.87	-
1986-89	-	-	0.64	0.79	0.87	1.00	0.79
1990-93	-	-	-	-	-	0.79	1.00

the exposure history are therefore constructed based on available data, X_t .

If it is assumed that $X(u)$ is piecewise continuous, then the cumulative exposure up to and including time t is

$$\int_0^t X(u)du \quad (3.2)$$

Rather than just considering the effect of the total exposure over a period of time, the contributions from intervals within the period may be of interest, in which case (3.2) can be expressed in the form of weighted integrals (Breslow et al. (1983), Bandeen-Roche et al. (1999)).

$$C_t = \int_0^{u=t} W(t-u)X(u)du \quad (3.3)$$

where the weights, $W(t-u)$, determine the aspect of the exposure being summarized, for example if the weights are of the form $W(u) = \min(1, u/b)$, then the exposures are phased in linearly over a period of length b until reaching their maximum. This can allow for delayed as well as cumulative effects depending on the form of the weights. In individual studies, the form of the cumulative exposure can be explicitly modelled, for example in the case of exposure to asbestos fibers, where the rate of elimination of the fibers from the lungs, λ , may be incorporated and the model takes the form $W(u) = \{1 - \exp(-\lambda u)\}/\lambda$ (Berry et al. (1979)).

Since exposure to air pollution starts in infancy, the lower limit of the integral will not be zero, instead the sum is likely to be over a specified period of time. If the weights are of the form

$$W(u) = \begin{cases} 1/(b-a) & \text{for } a \leq u < b \\ 0 & \text{otherwise} \end{cases} \quad (3.4)$$

then the summary will represent the average for the period $(t-b, t-a]$, $0 \leq a < b \leq t$. For example, when studying the short term affects of air pollution, if $a = 0$ and $b = 2$, then $W(t-u)$ would represent a three day mean.

When dealing with mortality counts, and exposure measurements, made at discrete times, the integral in (3.3) can be approximated by a summation over a suitable discretisation.

$$C_t = \sum_{k=0}^t W_{t-k} X_k \quad (3.5)$$

If the probability of disease given cumulative exposure is assumed to be proportional to $\exp(\gamma C_t)$, i.e. a log-linear model in cumulative exposure, then a Poisson model can be used to estimate the weights, W_{t-k} in (3.5). Assuming that $Y_t \sim P(E_t \mu_t)$ where E_t represents the expected number of cases then

$$\log \mu_t = \alpha + \gamma \sum_{k=0}^t W_{t-k} X_k = \alpha + \sum_{k=0}^t \beta_{t-k} X_{t-k} \quad (3.6)$$

Hence the parameters, β_{t-k} represent the effect of exposure k time periods ago. Comparison with (3.3) and (3.5) shows that $\beta_{t-k} = \gamma W_{t-k}$.

Recently, developments have been made in specifying the shape of the distributions of the weights, W_{t-k} , within aggregate level studies examining the short term effects air pollution on health. Schwartz (2000) describes the use of a distributed lag model, where the weights fit a polynomial function (Harvey (1981)). This requires assumptions to be made on the maximum lags that are likely to have an effect, in terms of the polynomial used, but has the advantage of increasing the stability of the individual estimates where there is high collinearity between the explanatory variables (Zanobetti et al. (2000)).

3.3.2 Model framework

In this chapter, the effects of different lagged exposures are examined on the health outcome, with the possibility that their effects may change over time. The time aspect is incorporated by considering health data from four distinct periods, each of four years, covering 1981–84, 1985–88, 1989–92 and 1993–96, indexed by $t = 1, \dots, 4$. In addition, the exposure data are available for seven periods of four years, 1966–69, 1970–73, ..., 1990–93, indexed by $p = 1, \dots, 7$. In order to produce a set of consistent lagged exposures over time, four of the seven exposure periods, $p = t + j : j = 0, \dots, 3$, are selected for each of the four health periods to give exposure measures over the previous sixteen years. For simplicity of notation, the lagged effects of exposure are indexed by $l = 4 - (p - t)$, referring to the exposure l periods (four year blocks) previously, with $l = 1$ the shortest, and $l = 4$ the longest lags. For example, in the third health period (1989–92, $t = 3$), the second lag, $l = 2$ refers to exposure 5–8 years previously (1982–85).

The combinations of health and exposure periods are shown in Table 3.4, and resulted in $n_1 = 391$ wards contributing exposure data (1966–81) for first health period (1981–84), $n_2 = 222$ for the second (health, 1985–88; exposure, 1970–85), $n_t = 191$ for the third (health, 1989–92; exposure, 1974–89) and $n_t = 123$ for the fourth (health, 1993–96; exposure 1978–93). The term ‘lag’ here,

Health Period	n_t	t	Pollution period	p	lag (l)
1981-84	391	1	1966-69	1	4
1981-84		1	1970-73	2	3
1981-84		1	1974-77	3	2
1981-84		1	1978-81	4	1
1985-88	222	2	1970-73	2	4
1985-88		2	1974-77	3	3
1985-88		2	1978-81	4	2
1985-88		2	1982-85	5	1
1989-92	191	3	1974-77	3	4
1989-92		3	1978-81	4	3
1989-92		3	1982-85	5	2
1989-92		3	1986-89	6	1
1993-96	123	4	1978-81	4	4
1993-96		4	1982-85	5	3
1993-96		4	1986-89	6	2
1993-96		4	1990-93	7	1

Table 3.4: Study design: Four health periods (with number of wards, n_t) with associated pollution periods (p) producing lagged exposures (l). Lagged exposures are 0-4, 5-8, 9-12 and 13-16 years for $l = 1, \dots, 4$ respectively.

therefore refers to the lag before the first year of each health period. The result is an $n_t \times 4$ matrix of exposure data, consisting of elements, X_{tl} for each health period and lag, $t = 1, \dots, 4$, $l = 1, \dots, 4$.

It should be noted that the number of areas, n_t , varies over health period, and that due to the nature of the data collection, some areas are common to more than one health period. Also, when the measurements from a particular exposure period are associated with different health time periods, the sets of areas included in the two health period analyses *may be different*. For example, the set of areas in health period one and pollution period two (1970–73), $X_{1,3}$, used to obtain the second longest lagged effect are not necessarily the same as those used in association with health period 2 (1985–88), $X_{2,4}$, to obtain the longest lag.

3.4 Health period specific analyses

The first part of the analysis considers each of the four health periods separately in order to gain an understanding of the data, to assess the modelling assumptions and to aid the interpretation of the full analysis, using data from all the health periods. It should be noted that, as a number of areas are common to more than one of the health period analyses, in the likely presence of unmeasured confounders, the results from the separate analyses cannot be regarded as independent. This point is discussed in greater detail in Section 3.5.1, which contains details of the models for combined analyses.

3.4.1 Models

Considering a single health period, if Y_i denotes the number of deaths in area i , it is assumed that $Y_i \sim P(E_i\mu_i)$, where E_i represents the pre-calculated age-sex standardised expected number of cases. The rate in each area, μ_i is modelled as a function of the lagged exposures, together with an area-level measure of socio-economic deprivation, Z_i ,

$$\log \mu_i = \beta_0 + \sum_{l=1}^4 \beta_l X_{li} + \beta_d Z_i \quad (3.7)$$

where β_l represents the effect of exposure over the previous l periods (blocks of four years) and β_d is the effect of deprivation.

In addition to investigating the cumulative effects of the individual four year periods of pollution, the effects of just the shortest lagged exposure (0-4 years, $l = 1$)

$$\log(\mu_i) = \beta_0 + \beta_1 X_{1i} + \beta_d X_{di} \quad (3.8)$$

and of the summed exposure of the previous 16 years ($\sum_{j=1}^4 X_{ij}$) are examined separately.

$$\log(\mu_i) = \beta_0 + \beta_c \sum_{j=1}^4 X_{ji} + \beta_d X_{di} \quad (3.9)$$

where β_1 and β_c are the regression coefficients associated with the effects of exposure in the previous four years and over sixteen years respectively.

There is a strong possibility of over-dispersion in the Poisson models (i.e., where the variance is greater than the mean) arising from the presence of unmeasured confounders. These may be operating at the individual level, e.g. smoking, or at the area level, e.g. residual socio-economic confounding. Over dispersion may also arise because of data anomalies, i.e. errors in the numerators and/or denominators, e.g. due to migration.

For each of the three forms of exposure models, four inferential approaches were carried out, with increasing complexity, in an attempt to explore the effects of unmeasured confounders.

1. Poisson, using a likelihood based generalised linear model
2. Poisson, within a Bayesian framework
3. Quasi-likelihood
4. Poisson, with ward level (log-normal) random effects in a Bayesian hierarchical model

The first two models would be expected to yield near identical results, the difference between them being the methods used for inference, namely likelihood and Bayesian (see Section 3.4.2).

Over-dispersion

The likelihood-based approach theoretically allows the different models, and therefore the effect of different exposure scenarios, to be compared, but makes no allowance for the extra-Poisson variability that may be present. This will lead to confidence intervals for the estimates of risk being too narrow and changes in deviances, used to compare models, being too small. An attempt to correct these effects, and to assess the degree of over-dispersion, can be made using a quasi-likelihood model (see Section 2.1.1), in which the usual Poisson assumption, that the variance is equal to the mean, is relaxed, allowing it to be a function of the mean and an unknown constant, i.e. $Var(Y_i) = cE(Y_i)$. The quasi-likelihood approach requires specification of only of the mean-variance relationship rather than the full likelihood. The estimate of the ‘dispersion parameter’, c , can be obtained and used to adjust the variances and the changes in deviances between models. The fourth model attempts to incorporate the possibility of unmeasured confounders by allowing each ward to have its own intercept term, β_{0i} , enabling the underlying risk in each area, before the effects of pollution and deprivation are considered, to vary.

With a frequentist approach, it is not possible to fit a different fixed effect intercept term for each individual ward, due to the excess number of parameters and so a random effects model is used. This assumes that there is an overall intercept (α) with an additional effect, β_{0i} , for each ward which comes from a distribution with expectation zero and suitable variance, for example $\beta_{0i} \sim N(0, \sigma_w^2)$. Model (3.7) thus becomes,

$$\log \mu_i = \alpha + \beta_{0i} + \sum_{l=1}^4 \beta_l X_{li} + \beta_d Z_i \quad (3.10)$$

If a Gamma distribution, $Ga(a, b)$ were used for the distribution of the random effects, the marginal distribution of Y_{ti} is analytically tractable and takes the form of a negative binomial. Wakefield et al. (2000) discusses the effect of different formulations of this Gamma distribution. Two cases are of particular interest, relating to the methods for allowing over-dispersion used here. In the first, the random effects follow a distribution of the form $\beta_{0i} \sim Ga(E_i a_i b, E_i b)$ (where E_i is the expected number of cases in area i), in which case the marginal variance of $Y_i | a_i, b$ is $E(Y_i | a_i, b)(1 + 1/b)$. This linear function of the mean is close to the quasi-likelihood approach described above, with $c = (1 + 1/b)$. If $\beta_{0i} \sim Ga(b, b/a_i)$ then $V(Y_i | a_i, b) = E(Y_i | a_i, b)(1 + E(Y_i | a_i, b))/b$. Here, the variance is a quadratic function of the mean, which is also the case when a log normal model is used for the random effects, i.e. $\log(\beta_{0i}) = \alpha + \beta_{0i} + \sum_{l=1}^4 \beta_l X_{li} + \beta_d Z_i$ with $\beta_{0i} \sim N(0, \sigma_w^2)$. Although the marginal distribution of Y_i is intractable in this case, the mean and variance can be found and are

$$V(Y_i | \mu_i, \sigma_w^2) = E(Y_i | \mu_i, \sigma_w^2) \{1 + E(Y_i | \mu_i, \sigma_w^2)(\exp(\sigma_w^2) - 1)\} \quad (3.11)$$

$$\text{where } E(Y_i | \mu_i, \sigma_w^2) = E_i \exp(\mu_i + \sigma_w^2/2) \quad (3.12)$$

$$\text{and } \mu_i = \beta_0 + \sum_{l=1}^4 \beta_l X_{li} + \beta_d Z_i$$

As with the second example above, the variance is a quadratic function of the mean and so the results from the log-normal random effects model would not be expected to be the same as those from the quasi-likelihood model, where the variance is a linear function of the mean, although both will result in an increase in the variability of the estimates, and thus wider confidence/credible intervals.

Hyperpriors

In a Bayesian analysis, prior distributions must be assigned to all the parameters, including the intercept terms and the effects of pollution and deprivation. Within each separate health period analysis, normal priors are assumed for the intercept terms, $\alpha \sim N(c_0, C_0)$, the effects of pollution, $\beta_{tk} \sim N(c_e, C_e)$, and deprivation, $\beta_d \sim N(c_d, C_d)$. Here, c_0 will be a scalar, β_d a 10×1 matrix and c_e a 4×1 vector for the model with cumulative effects of all lags, or a scalar for the two models with a single exposure variable (0-4 or 0-16 years). The corresponding variance-covariance matrices are thus of size $C_0; 1 \times 1$, $C_d; 10 \times 10$ and $C_e; 4 \times 4$ (or 1×1 for the single exposure models). In the random effects model, where $\beta_{0i} \sim N(0, \sigma_w^2)$, a hyperprior distribution is also assigned to the variance term, when a Gamma distribution is specified for the precision of the random effects, specifically $\sigma_{\beta_0}^{-2} \sim Ga(a_{\beta_0}, b_{\beta_0})$, parameterised so that $E(\sigma_{\beta_0}^{-2}) = a_{\beta_0}/b_{\beta_0}$ and $V(\sigma_{\beta_0}^{-2}) = a_{\beta_0}/b_{\beta_0}^2$.

3.4.2 Inference

In the separate health period analyses, the first Poisson model is a standard GLM and estimates of the parameters are found using maximum likelihood and iteratively re-weighted least squares (IRLS). The Quasi-likelihood uses an adaption of the IRLS algorithm (McCullagh and Nelder (1989)).

For the Bayesian analyses, the posterior distributions will often involve high dimensional integration and may be analytically intractable. However, samples from this distribution may be generated in a straightforward fashion using Markov chain Monte Carlo (e.g., Smith and Roberts (1993)). This was performed using the WinBUGS software (Spiegelhalter et al. (1998)).

3.4.3 Results

The models described in Section 3.4.1 are now applied to the data from the four health periods. The analyses were performed with and without adjustment for deprivation. Prior sensitivity on the effect of the random effects is addressed; initially the prior distributions for the Bayesian models are specified as $\beta_0, \beta_d, \beta_k \sim N(0, 1000)$, corresponding to vague prior beliefs and $\sigma_w^{-2} \sim$

$Ga(1, 0.01)$ for the variance of the random effects. It should be noted that these priors assume that $\beta_k, k = 1, \dots, 4$ are independent.

Table 3.5: Relative risks, with 95% confidence / credible intervals for increase of 10 ppb of SO_2 over the previous four years, with and without adjustment for deprivation. Results are for four models; likelihood based Poisson, Bayesian Poisson, Quasi-likelihood (with dispersion parameter) and Bayesian Poisson with log-normal (LN) random effects (with variance of random effects.)

	Without depr.			With Depr		
	Rel. Risk	lower	upper	Rel. Risk	lower	upper
Health Period 1						
1. Poisson (lhd)	1.125	1.106	1.144	1.086	1.066	1.105
2. Poisson (Bayes)	1.125	1.106	1.144	1.087	1.069	1.106
3. Quasi	1.125	1.068	1.184	1.086	1.029	1.145
<i>Dispersion</i>	<i>9.190</i>	-	-	<i>8.952</i>	-	-
4. LN random effects	1.112	1.061	1.167	1.064	1.014	1.118
σ_w	<i>0.3175</i>	-	-	<i>0.2992</i>	-	-
Health Period 2						
1. Poisson (lhd)	1.094	1.055	1.135	1.073	1.033	1.115
2. Poisson (Bayes)	1.094	1.055	1.135	1.074	1.035	1.115
3. Quasi	1.094	0.999	1.199	1.073	0.984	1.171
<i>Dispersion</i>	<i>6.176</i>	-	-	<i>5.251</i>	-	-
4. LN random effects	1.094	1.008	1.192	1.072	0.9902	1.1580
σ_w	<i>0.2838</i>	-	-	<i>0.2380</i>	-	-
Health Period 3						
1. Poisson (lhd)	1.089	1.050	1.131	1.096	1.054	1.139
2. Poisson (Bayes)	1.089	1.049	1.130	1.094	1.055	1.137
3. Quasi	1.089	1.008	1.178	1.096	1.017	1.181
<i>Dispersion</i>	<i>4.424</i>	-	-	<i>3.835</i>	-	-
4. LN random effects	1.048	0.9670	1.135	1.058	0.9823	1.139
σ_w	<i>0.2464</i>	-	-	<i>0.2132</i>	-	-
Health Period 4						
1. Poisson (lhd)	1.125	1.076	1.176	1.133	1.082	1.186
2. Poisson (Bayes)	1.125	1.076	1.175	1.132	1.082	1.184
3. Quasi	1.125	1.014	1.248	1.133	1.024	1.254
<i>Dispersion</i>	<i>5.440</i>	-	-	<i>4.896</i>	-	-
4. LN random effects	1.140	1.026	1.272	1.157	1.043	1.286
σ_w	<i>0.2502</i>	-	-	<i>0.2330</i>	-	-

Table 3.5 gives results from the four different models in each health period for models using the most recent exposure periods, 0-4 years lag. For the likelihood based Poisson and the quasi-likelihood model, the MLEs are given together with 95% confidence intervals, with an estimate of the over-dispersion parameter for the latter. For the Bayesian models, posterior summaries, in the form of the median and 95% credible intervals are given. Results from the four modelling approaches were similar, with the two basic Poisson models (likelihood and Bayes) producing near identical results as expected, due to the relatively flat priors used. The effect of allowing for over-dispersion using quasi-likelihood and random effects produced similar increases in the

variability of the estimates. Significant risks (at the 5% level) were observed before adjustment for deprivation in all the health periods except with the quasi-likelihood model in period 2 and the random effects model in period 4. Increases of 11-12.5% per 10 ppb SO₂ were seen in the first health period, 9.5% in the second, 4.5-9% in the third and 12.5-14% in the fourth. In the first and third health periods, the random effects model gives somewhat lower estimates than the other models, but in the fourth it produced the highest estimates. High levels of over-dispersion were observed, with the estimates of the dispersion parameters using the quasi-likelihood model of 9.2, 6.2, 4.4 and 5.4 for the four health periods, strongly suggesting the presence of unmeasured confounders. The extent of the unmeasured covariates can be also seen in the variance of the random effects, σ_w , which indicates the extent to which the individual wards (as shown by the intercept terms) differ from the overall risk. These reduce when more information about the areas, i.e. deprivation, is included in the model, for example, the variance of the random effects drops by approximately 30 % in the second health period. These variances also follow the same pattern over time as the dispersion parameters, with the highest values observed in the first health period, dropping in the 2nd and 3rd, with a slight rise in the fourth.

Smaller excess risks are seen after adjustment for deprivation, but with no change in the significance of the results previously noted, except for the random effects model in health period 2. Increases of 6.5-8.5% were seen in the first health period, 7% in the second, 6-9.5% in the third and 13-15.5% in the fourth. The effect of adjusting for deprivation changes slightly over time, reducing the risk in the first two periods, but resulting in a (small) increase in the last two. There was no obvious association between the random effects with either exposure or deprivation after they were included in the model, as can be seen in Figure 3.5.

The results using the longest exposure period, i.e. over the previous 16 years, are presented in Table 3.6. Estimates of risk with 95% CIs are given for the four models, with indicators of over-dispersion, as presented in Table 3.5. It should be noted that the risks are smaller than those seen in Table 3.5 because in the first case $\exp(\beta)$ was the change in risk for a one unit increase in X , the four year exposure, while in the second $\exp(\beta)$ is the change for a one unit increase in the 16 year exposure. For example, increases of 1%, 1.5-2%, 1% and 3.5% were seen for health periods 1, 2, 3 and 4 respectively, after adjustment for deprivation. Similar patterns of over-dispersion (as seen in the quasi-likelihood dispersion parameter and σ_w) are seen, decreasing when deprivation is included and reducing over periods one to three, with a small increase in the fourth. Before adjusting for deprivation, significant results were seen in all health periods for all models except for the random effects model in period 3. Including deprivation caused a reduction in the risk for all four health periods, but with significant results for all but the quasi-likelihood and random effects models in period 3.

Table 3.7 gives summaries of the cumulative effects of the separate periods of lagged exposure (0-4, 5-8, 9-12 and 13-16 years), adjusted for deprivation, in the same form as Tables 3.5 and 3.6. The

Table 3.6: Relative risks, with 95% confidence / credible intervals for increase of 10 ppb of SO₂ over the previous sixteen years, with and without adjustment for deprivation. Results are for four models; likelihood based Poisson, Bayesian Poisson, Quasi-likelihood (with dispersion parameter) and Bayesian Poisson with log-normal (LN) random effects (with variance of random effects.)

	Without depr.			With Depr		
	Rel. Risk	lower	upper	Rel. Risk	lower	upper
Health Period 1						
1. Poisson (lhd)	1.020	1.017	1.022	1.012	1.009	1.015
2. Poisson (Bayes)	1.020	1.017	1.023	1.012	1.009	1.015
3. Quasi	1.020	1.012	1.028	1.012	1.003	1.021
<i>Dispersion</i>	9.358	-	-	9.202	-	-
4. LN random effects	1.021	1.013	1.029	1.011	1.003	1.019
σ_w	0.3134	-	-	0.2989	-	-
Health Period 2						
1. Poisson (lhd)	1.033	1.026	1.039	1.019	1.012	1.026
2. Poisson (Bayes)	1.033	1.026	1.039	1.018	1.012	1.025
3. Quasi	1.033	1.018	1.048	1.019	1.004	1.034
<i>Dispersion</i>	5.677	-	-	5.051	-	-
4. LN random effects	1.031	1.016	1.045	1.015	1.001	1.028
σ_w	0.2730	-	-	0.2363	-	-
Health Period 3						
1. Poisson (lhd)	1.018	1.009	1.027	1.012	1.004	1.021
2. Poisson (Bayes)	1.018	1.009	1.027	1.012	1.003	1.021
3. Quasi	1.018	1.000	1.036	1.012	0.995	1.030
<i>Dispersion</i>	4.436	-	-	3.835	-	-
4. LN random effects	1.013	0.9952	1.031	1.008	0.9915	1.024
σ_w	0.2464	-	-	0.2151	-	-
Health Period 4						
1. Poisson (lhd)	1.037	1.027	1.047	1.034	1.024	1.044
2. Poisson (Bayes)	1.037	1.027	1.047	1.034	1.024	1.044
3. Quasi	1.037	1.014	1.060	1.034	1.012	1.056
<i>Dispersion</i>	5.235	-	-	4.771	-	-
4. LN random effects	1.038	1.015	1.062	1.037	1.015	1.06
σ_w	0.2429	-	-	0.2284	-	-

Table 3.7: Relative risks, with 95% confidence/credible intervals for increase of 10 ppb of SO₂ for cumulative effect of exposure over the previous 0-4, 5-8, 9-12 and 13-16 years, with adjustment for deprivation. Results are for four models; likelihood based Poisson, Bayesian Poisson, Quasi-likelihood (with dispersion parameter) and Bayesian Poisson with log-normal (LN) random effects (with variance of random effects).

	lag	Health Period 1			Health Period 2		
		Rel. Risk	lower	upper	Rel. Risk	lower	upper
1. Poisson (lhd)	0-4	1.042	1.012	1.072	1.014	0.962	1.070
	5-8	1.061	1.031	1.092	1.029	0.981	1.079
	9-12	0.987	0.963	1.011	1.034	0.995	1.074
	13-16	0.999	0.984	1.014	1.004	0.982	1.027
2. Poisson (Bayes)	0-4	1.042	1.012	1.071	1.015	0.9626	1.070
	5-8	1.060	1.030	1.091	1.029	0.9826	1.079
	9-12	0.9869	0.9632	1.011	1.033	0.9950	1.073
	13-16	0.9984	0.9838	1.013	1.003	0.9814	1.026
3. Quasi	0-4	1.042	0.956	1.134	1.014	0.899	1.144
	5-8	1.061	0.974	1.155	1.029	0.925	1.145
	9-12	0.987	0.917	1.062	1.034	0.949	1.126
	13-16	0.999	0.955	1.044	1.004	0.954	1.057
<i>Dispersion</i>		9.003	-	-	5.089	-	-
4. LN random effects	0-4	1.012	0.9341	1.096	1.035	0.9227	1.156
	5-8	1.055	0.9778	1.148	1.013	0.9110	1.122
	9-12	0.9987	0.9168	1.055	1.031	0.9482	1.123
	13-16	1.006	0.9626	1.053	0.9981	0.9505	1.048
σ_w		0.2996	-	-	0.2381	-	-
	lag	Health Period 3			Health Period 4		
		Rel. Risk	lower	upper	Rel. Risk	lower	upper
1. Poisson (lhd)	0-4	1.128	1.065	1.196	1.012	0.947	1.081
	5-8	0.964	0.905	1.026	1.159	1.078	1.246
	9-12	1.027	0.975	1.082	0.890	0.833	0.951
	13-16	0.972	0.940	1.005	1.094	1.040	1.150
2. Poisson (Bayes)	0-4	1.122	1.060	1.178	1.014	0.9515	1.082
	5-8	0.9567	0.9076	1.027	1.150	1.072	1.231
	9-12	1.0242	0.9750	1.0780	0.8985	0.8431	0.9569
	13-16	0.9761	0.9453	1.007	1.088	1.038	1.142
3. Quasi	0-4	1.128	1.008	1.264	1.012	0.877	1.167
	5-8	0.964	0.853	1.089	1.159	0.992	1.355
	9-12	1.027	0.928	1.136	0.890	0.770	1.028
	13-16	0.972	0.911	1.038	1.094	0.982	1.218
<i>Dispersion</i>		3.798	-	-	4.678	-	-
4. LN random effects	0-4	1.067	0.9467	1.193	1.044	0.8938	1.214
	5-8	0.9822	0.8673	1.119	1.091	0.9150	1.318
	9-12	1.046	0.9382	1.164	0.9242	0.7796	1.083
	13-16	0.9626	0.8982	1.037	1.095	0.9728	1.237
σ_w		0.2146	-	-	0.2292	-	-

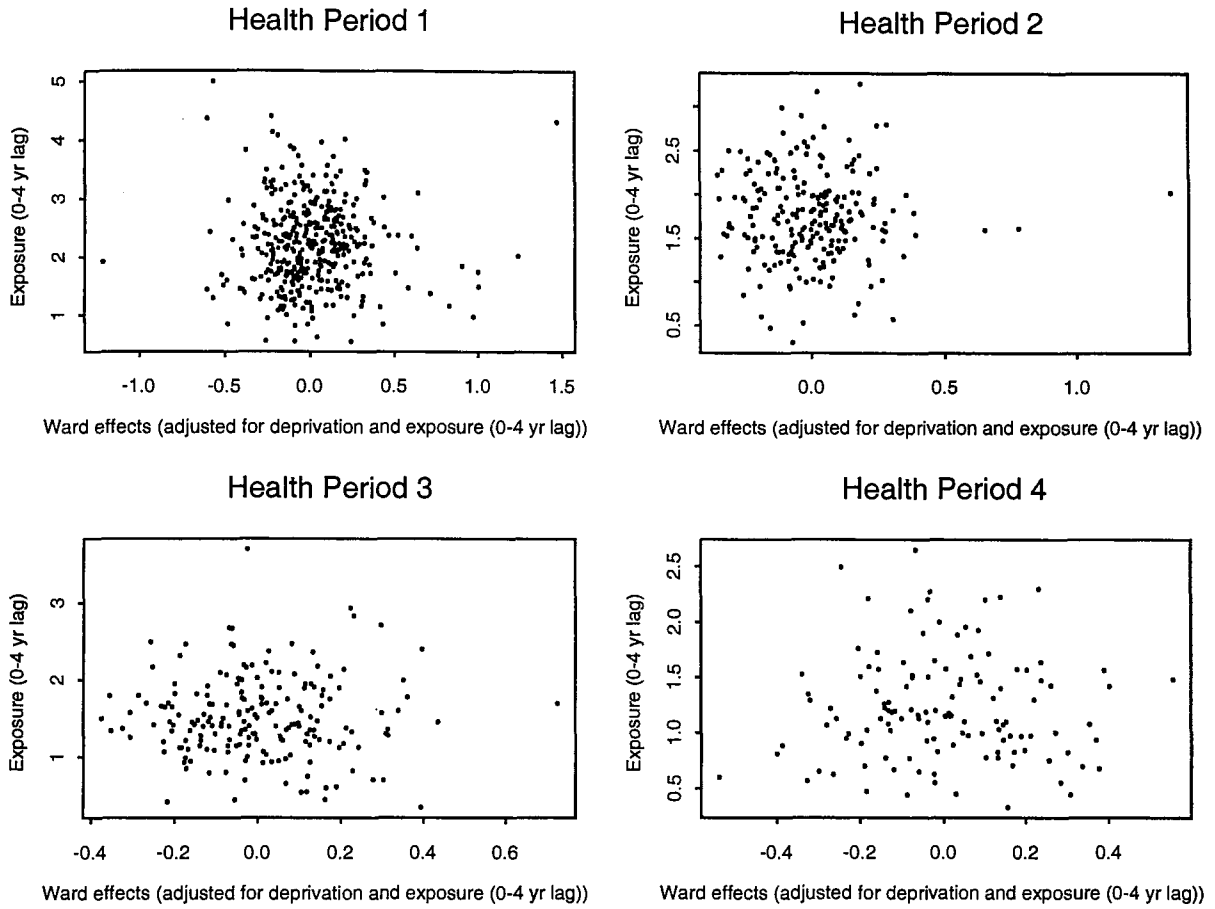


Figure 3.5: Associations between ward level (log-normal) random effects from model including SO_2 exposure in previous four years and deprivation. The four panels refer to data from four health periods (1981-84, 1985-88, 1989-92, 1993-96).

estimates of the coefficients for the separate lagged periods are based on highly correlated data, meaning that observing consistent patterns in the lags over health periods is difficult to interpret, and most of the individual estimates are non-significant. However, it appears that the greatest risk is associated with the most recent one or two exposure periods. Again, including deprivation in the model reduced the degree of over-dispersion, with reductions in both the dispersion parameter and σ_w (not shown). Similar patterns were also seen in the measures of over-dispersion over time, with the highest values again seen in the earlier periods.

The posterior medians presented in Tables 3.5, 3.6 and 3.7 appear to be robust to the choice of prior distribution for the variance of the random effects. The effect of using number of different prior combinations for the precision, including $Ga(1, 0.01)$, $Ga(0.5, 0.0005)$ and $Ga(0.001, 0.001)$, was assessed, with very little difference in the resulting estimates of risk or variances (not shown).

The deviances associated with the quasi-likelihood models of the different lagged exposures (0-4

Table 3.8: Deviances and degrees of freedom for quasi-likelihood models relating various lagged exposures (previous 4 years, 16 years (total) and for the cumulative effects of 0-4, 5-8, 9-12 and 13-16 years (incremental)) of SO₂. Changes in deviance from the null model are given, with associated p-values. Results are given with and without adjustment for deprivation, after allowance for over-dispersion.

1981-84	Without depr.				With Depr			
	Deviance	df	Adj. change	p	Deviance	df	Adj. change	p
Null	3332.2	390			2922.1	381		
Prev. 4 yrs	3151.1	389	-19.7	0.000	2842.4	380	-8.9	0.003
Prev. 16 yrs (total)	3122.9	389	-22.7	0.000	2857.7	380	-7.2	0.007
Prev. 16 yrs (incremental)	3092.0	386	-26.1	0.000	2824.3	377	-10.9	0.028
1985-88	Without depr.				With Depr			
	Deviance	df	Adj. change	p	Deviance	df	Adj. change	p
Null	1223.9	221			917.0	212		
Prev. 4 yrs	1200.7	220	-4.0	0.045	903.8	211	-2.6	0.107
Prev. 16 yrs (total)	1120.6	220	-17.9	0.000	886.4	211	-6.0	0.014
Prev. 16 yrs (incremental)	1112.5	217	-19.3	0.001	884.6	208	-6.4	0.173
1989-92	Without depr.				With Depr			
	Deviance	df	Adj. change	p	Deviance	df	Adj. change	p
Null	845.9	190			673.5	181		
Prev. 4 yrs	825.7	189	-4.6	0.033	651.9	180	-5.7	0.017
Prev. 16 yrs (total)	829.1	189	-3.8	0.052	666.0	180	-2.0	0.159
Prev. 16 yrs (incremental)	817.1	186	-6.5	0.165	647.6	177	-6.8	0.146
1993-96	Without depr.				With Depr			
	Deviance	df	Adj. change	p	Deviance	df	Adj. change	p
Null	686.3	122			579.8	113		
Prev. 4 yrs	659.7	121	-5.2	0.022	551.3	112	-6.1	0.014
Prev. 16 yrs (total)	632.0	121	-10.6	0.001	533.8	112	-9.8	0.002
Prev. 16 yrs (incremental)	596.9	118	-17.5	0.002	512.1	109	-14.5	0.006

years, 0-16 years and the cumulative effect of 0-4, 5-8, 9-12 and 13-16 years) of SO_2 are given in Table 3.8. The changes in deviances have been adjusted to allow for the over-dispersion, but significant differences are still observed. Each of the three exposure models resulted in a significant improvement over the null model, before and after including the effect of deprivation, except for the previous 16 years (total and incremental) in health period 3 and the previous 4 and 16 years (incremental) in health period 2, after adjustment for deprivation. Due to the high levels of collinearity, the effects of the different exposure measures are very similar and neither of them was significantly better than the others.

In the absence of deviances, which rely on likelihoods, model fit can be assessed informatively by examining changes in the variances of the random effects, a reduction in which would indicate that more of the variation is being explained by the variables in the model, and less being ‘mopped up’ by the random effects. Reductions are observed when extra information, in the form of exposure and deprivation, is incorporated into the model, but the reductions for all three models were virtually identical, again meaning that a single ‘best’ model cannot be chosen.

3.5 Combined health period analysis

In the next part of the analyses, data from all four health periods are modelled together. The model shown in (3.10) must therefore incorporate the repeated measures taken from some of the areas. If such information were ignored, then unmeasured confounders which are constant, or near constant over time, may cause dependence between the health outcomes, causing the assumption of independent outcomes to become untenable.

3.5.1 Models

First Stage

Model (3.10) is thus extended to include all the health periods, with an overall term, β_0 , with health period adjustments, β_t and ward level intercepts, β_{0i} (which are constant over time). The effect of deprivation, β_{dt} is also allowed to vary over time.

$$\log(\mu_{it}) = \alpha + \beta_{0i} + \sum_{l=1}^4 \beta_{tl} X_{tli} + \beta_{td} Z_{ti} \quad (3.13)$$

where X_{tli} is the level of pollution in area i for health period, $t = 1, \dots, 4$ and associated lag $l = 1, \dots, 4$

Second Stage

Random effects

The ward level intercepts (random effects) are again assumed to follow a normal distribution, i.e. $\log(\beta_{0i}) \sim N(0, \sigma_{\beta_0}^2)$, as described in Section 3.4.1.

Effects of lags and time

As this model now contains the effects of different lagged exposures over time (health periods), the coefficients, $\beta_{t,l}$ (the effect of exposure l periods previous to health period t) can vary over time. If this is not the case, and the effect of exposure l periods previously is constant over the four health periods then

$$\beta_{1,l} = \beta_{2,l} = \beta_{3,l} = \beta_{4,l} = \beta_l \quad (3.14)$$

Here, for example, the entries in Table 3.4 corresponding to $l = 3$, a lag of 9-12 years (health outcome 1981-84; exposure 1970-73, health outcome 1985-88; exposure 1974-77, health outcome 1989-92; exposure 1978-81, and health outcome 1993-96; exposure 1982-85) would be represented by a single coefficient β_2 .

Less plausibly, if the different lagged exposures all had the same effect, then

$$\beta_{t,t} = \beta_{t,1} = \beta_{t,2} = \beta_{t,3} = \beta_t \quad (3.15)$$

which still allows a health period effect. The simplest model would therefore have $\beta_{t,l} = \beta$, in which case each exposure period makes an equal contribution which is constant across health periods.

The estimates of the different effects of exposure over time obtained using model (3.10) will all come from data which are highly correlated, and it is likely that they may be individually unstable. In circumstances such as these, it may be possible to produce more stable estimates for a particular health period-lag effect by ‘borrowing’ information from similar effects. The approach is similar to that used by Breslow and Clayton (1993), who smoothed risks in an age-cohort model analysing breast cancer rates. They fitted a second order auto-regressive model to cohort effects, i.e. $\beta_k | \beta_1, \dots, \beta_{k-1} \sim N(2\beta_{k-1} - \beta_{k-2}, \tau)$, where β_k is the effect of the k^{th} cohort. Here, smoothed risk estimates are produced for the lags within each health period using a Gaussian auto-regression prior (Besag and Kooperberg (1995)). Instead of modelling the multivariate distribution of β to allow for dependencies between β_i and $\beta_j, i \neq j$, the univariate conditional distributions $\beta_i | \beta_j, i \neq j$ are modelled for values of j that are contained in a defined set, δ_i , of neighbours of i , i.e. those with similar health period/lag combinations. The conditional distributions are therefore, $N(\bar{\beta}_i, \frac{\sigma_{\beta}^2}{m_i})$, where $\bar{\beta}_i = 1/m_i \sum_{j \in \delta_i} \beta_j$, where m_i denotes the number contained in δ_i .

Two different approaches to smoothing the risks are considered; firstly, the risk estimates for the different lags are smoothed within each health period (3.16), and secondly, the estimates for each

lag are smoothed over health periods (3.17).

$$E(\beta_{t,k}) = \sum_{m=-1}^1 \beta_{t,k+m} \quad (3.16)$$

$$E(\beta_{t,k}) = \sum_{m=-1}^1 \beta_{t+m,k} \quad (3.17)$$

Third Stage - hyperpriors

Normal priors are assumed for the intercept terms, $\beta_0 \sim N(c_0, C_0)$, where c_0 is a 4×1 vector and C a 4×4 variance-covariance matrix, and the effect of deprivation, $\beta_d \sim N(c_d, C_d)$, where c_d is of size 40×1 (allowing the effect of deprivation to change over health periods) and C_d a 40×40 matrix. Normal priors are also assigned to the effects of exposure, $\beta_{tk} \sim N(c_e, C_e)$, with c_e a 16×1 vector and C_e a 16×16 matrix ($c_e; 4 \times 1, C_e; 4 \times 4$ for the single exposure models). A Gamma distribution is again specified for the precision of the random effects, $\sigma_{\beta_0}^{-2} \sim Ga(a_{\beta_0}, b_{\beta_0})$.

3.5.2 Inference

As in the health period specific analyses, the Bayesian analyses use Markov chain Monte Carlo to generate samples from the posterior distribution. This was performed using the WinBUGS software (Spiegelhalter et al. (1998)), noting that using the auto-regressive priors leads to cyclical graphs, in which case some of the conditional distributions have to be explicitly specified (Spiegelhalter et al. (1996)).

3.5.3 Results

Table 3.9 gives analogous results to those seen in Table 3.5, using exposure from the previous four years, but using data from all four health periods simultaneously. Results are given for the null model, with no effect of deprivation or exposure, for two models with independent and auto-regressive priors on the lagged effects, β_t , and for an overall model with no health period effect. Excesses in risk of 6-6.5%, 8%, 4.5% and 12.5% were seen before adjustment for deprivation in health periods 1, 2, 3 and 4 respectively. In comparison to the results from the health period specific analyses, using a random effects model, (Table 3.5), smaller excesses are seen in the first, second and fourth health periods. After adjustment for deprivation, excesses of 4.6%, 6.0-6.2%, 4.2-4.3% and 12.1-12.3% are seen for the four health periods, each of which is smaller than the corresponding risk from the health period specific analyses. Reductions in the underlying variance of the random effects are seen with adjustment for deprivation (which causes a reduction in risk in all periods), and also when exposure from the previous 4 years is added to the model. The effect of adding exposure and deprivation on the random effects can be seen in Figure 3.6. Before they are included in the model, they are correlated with the random effects, indicating that the

Table 3.9: Relative risks, with 95% credible intervals for increase of 10 ppb of SO₂ for effects of exposure over the previous four years, with and without adjustment for deprivation. Variances of random effects are given from Bayesian hierarchical models. HP denotes health period (HP1=1981-84, HP2=1985-88, HP3=1989-92, HP4=1993-96).

		Without Depr.			With Depr.		
		Rel. Risk	lower	upper	Rel. Risk	lower	upper
1. No exposure	σ_w	<i>0.3105</i>	-	-	<i>0.2819</i>	-	-
2. Independent β	HP1	1.065	1.032	1.098	1.046	1.015	1.079
	HP2	1.083	1.029	1.138	1.062	1.009	1.116
	HP3	1.045	0.9825	1.102	1.043	0.9873	1.102
	HP4	1.125	1.051	1.201	1.123	1.05	1.202
	σ_w	<i>0.3041</i>	-	-	<i>0.2790</i>	-	-
3. Auto-regressive β	HP1	1.060	1.034	1.098	1.046	1.014	1.079
	HP2	1.084	1.030	1.139	1.060	1.008	1.115
	HP3	1.047	0.9922	1.105	1.042	0.9873	1.100
	HP4	1.125	1.054	1.199	1.121	1.048	1.200
	σ_w	<i>0.3040</i>	-	-	<i>0.2787</i>	-	-
	σ_β	<i>0.7612</i>	-	-	<i>0.7629</i>	-	-
4. Overall		1.067	1.039	1.097	1.052	1.023	1.087
	σ_w	<i>0.3042</i>	-	-	<i>0.2782</i>	-	-

random effects contain information about both exposure and deprivation, as they are at this point unaccounted for. The correlation between the random effects and deprivation when exposure is included in the model, is greater than that between the random effects and exposure when deprivation is included in the model. When both are included in the model, there is no obvious correlation, indicating that the random effects then contain information only on unmeasured covariates. Without the effect of different health periods, the overall risk is 1.052 (1.023-1.087) after adjustment for deprivation, with very little difference in σ_w when compared to that seen when the model allowed for separate health period effects.

Table 3.10 presents summaries of the posterior distributions as in Table 3.9, but for exposure over the previous 16 years. They can be compared to the health period specific results seen in Table 3.6 but incorporate data from all the health periods simultaneously. Again smaller excess risks are seen in all but the third period, using both independent and auto-regressive priors, for which the results were similar. The risks seen for each health period show a similar pattern to the separate analyses, with the highest risks being observed in the fourth period. Excesses in risk of 1.3%, 2.6%, 0.9% and 2.2% were seen before adjustment for deprivation in health periods 1, 2, 3 and 4 respectively. The corresponding excesses after adjustment for deprivation were 0.6%, 1.2%, -0.2% and 1.4%. After adjustment for deprivation, significant increases in risk were seen in the first two health periods. Again, these are smaller than those observed in the health period specific analyses (Table 3.6). Significant risks are observed in all but the third period, with very little difference between the two priors. Similar decreases in the variance of the random effects,

Table 3.10: Relative risks, with 95% credible intervals for increase of 10 ppb of SO₂ for effects of exposure over the previous sixteen years, with and without adjustment for deprivation. Variance of random effects are given from Bayesian hierarchical models. HP denotes health period (HP1=1981-84, HP2=1985-88, HP3=1989-92, HP4=1993-96).

		Without Depr.			With Depr.		
		Rel. Risk	lower	upper	Rel. Risk	lower	upper
1. no exposure	σ_w	<i>0.3105</i>	-	-	<i>0.2819</i>	-	-
2. Independent β	HP1	1.013	1.007	1.019	1.006	1.000	1.012
	HP2	1.026	1.016	1.036	1.012	1.001	1.023
	HP3	1.009	0.9958	1.023	0.9981	0.9851	1.012
	HP4	1.022	1.005	1.039	1.014	0.9971	1.030
3. Auto-regressive β	σ_w	<i>0.2995</i>	-	-	<i>0.2791</i>	-	-
	HP1	1.013	1.007	1.019	1.006	1.000	1.012
	HP2	1.026	1.016	1.036	1.012	1.002	1.022
	HP3	1.009	0.9960	1.023	0.9981	0.9848	1.011
	HP4	1.022	1.005	1.039	1.014	0.9973	1.030
	σ_w	<i>0.2997</i>	-	-	<i>0.2788</i>	-	-
	σ_β	<i>0.7683</i>	-	-	<i>0.2791</i>	-	-
4. Overall		1.014	1.008	1.019	1.006	1.000	1.012
	σ_w	<i>0.2999</i>	-	-	<i>0.2793</i>	-	-

σ_w , are also seen with both deprivation and exposure. The overall result is again significant, 1.006 (1.000-1.012) with deprivation in the model, albeit with a smaller relative risk, again due to the increase in the range of the exposure data.

The results of modelling the cumulative effects of the 0-4, 5-8, 9-12 and 13-16 year lagged exposures can be seen in Figure 3.7. The medians and 95% credible intervals from the posterior distribution are shown for models using four different priors for β_{tk} ; (a) independent (b) auto-regressive, smoothing over lags within health periods, and (c) auto-regressive, smoothing each lagged effect over health periods. Little difference is seen between the results from the different models, with similar effects being seen to those from the individual health period analyses seen in Table 3.7, again most of the individual risk estimates are non-significant, except those associated with the most recent exposures, which were again significant in the first and fourth health periods, a result which seems robust to the choice of prior for β_{tk} .

The results of fitting simplified models, i.e. dropping the effect of health periods and/or the different lags can be seen in Table 3.11. Posterior medians with 95% credible intervals are given for the estimates of risk under each of the models, together with those of the variances of the random effects. In the first, the t subscript is dropped from β_{tk} , forcing the effect of the different lags to be constant over the four health periods. This resulted in elevated, but non-significant, risks for all but the shortest lag, which had a RR of 1.045 (1.013-1.078). This is however, smaller than the 12.4% excess seen in Table 3.9 when the 0-4 year lag was modelled independently, because

of the high collinearity with the other lagged values now in the model. When the k subscript was dropped, forcing the effect of the different lagged exposures to be the same, significantly elevated risks were seen in the first two health periods. The highest risk was again seen in the fourth period, but was non-significant (power was reduced because it included fewer areas). Fixing the effect of all health period-lag combinations to be the same, resulted in an overall increased risk of 1.007 (1.001-1.013).

Again, there was very little difference in the estimates of risks and variances seen in Tables 3.9, 3.10, 3.11 and Figure 3.7 when different prior distributions were used. For example, using $Ga(1, 0.01)$, $Ga(0.5, 0.0005)$ or $Ga(0.001, 0.001)$ priors for the random effect variance, had little change on the posterior medians of risk estimates or variances.

Table 3.11: Relative risks, with 95% credible intervals, for increase of 10 ppb of SO_2 with variance of random effects, for simplified models of the effects of exposure over the previous 0-4, 5-8, 9-12 and 13-16 years, with adjustment for deprivation. The models are (1) without different effects of health period, (2) without different effects of lagged exposures and (3) no effects of different health periods or lags. HP denotes health period (HP1=1981-84, HP2=1985-88, HP3=1989-92, HP4=1993-96).

		Rel. Risk	lower	upper
1. No effect of HP	0-4	1.045	1.013	1.078
	5-8	1.020	0.9919	1.048
	9-12	1.012	0.9896	1.034
	13-16	0.9856	0.9694	1.003
	σ_w	0.2787	-	-
2. No effect of lag	HP1	1.006	1.000	1.012
	HP2	1.012	1.002	1.022
	HP3	0.9983	0.9851	1.012
	HP4	1.014	0.9982	1.031
	σ_w	0.2789	-	-
3. No effect of HP-lag		1.007	1.001	1.013
	σ_w	0.2793	-	-

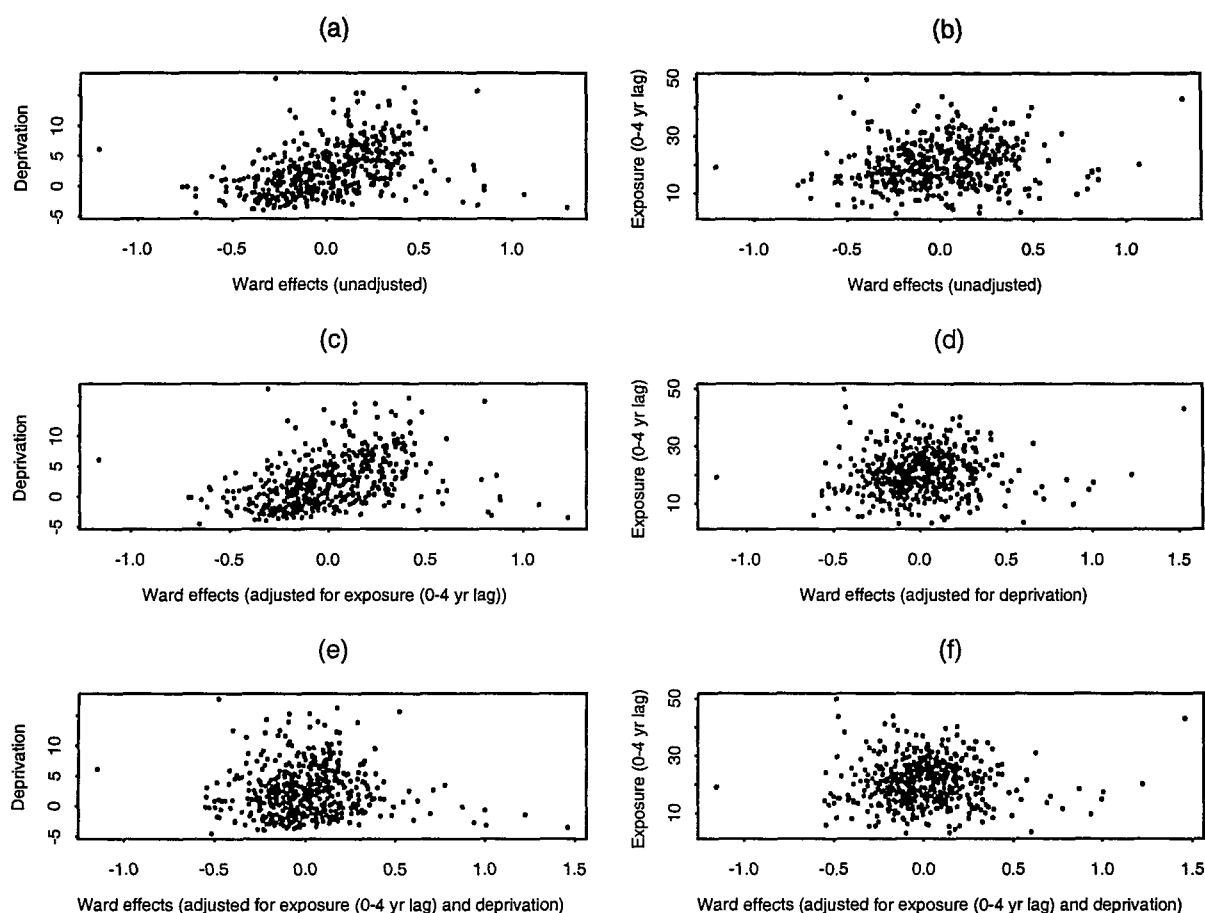


Figure 3.6: Associations between years ward level random effects using Bayesian hierarchical model with exposure over previous four years and deprivation. The six panels show associations between (a) deprivation and random effects unadjusted for exposure and deprivation, (b) exposure and unadjusted random effects, (c) deprivation and random effects adjusted for exposure, (d) exposure and random effects adjusted for deprivation, (e) deprivation and random effects adjusted for exposure and deprivation and (f) exposure and random effects adjusted for exposure and deprivation.

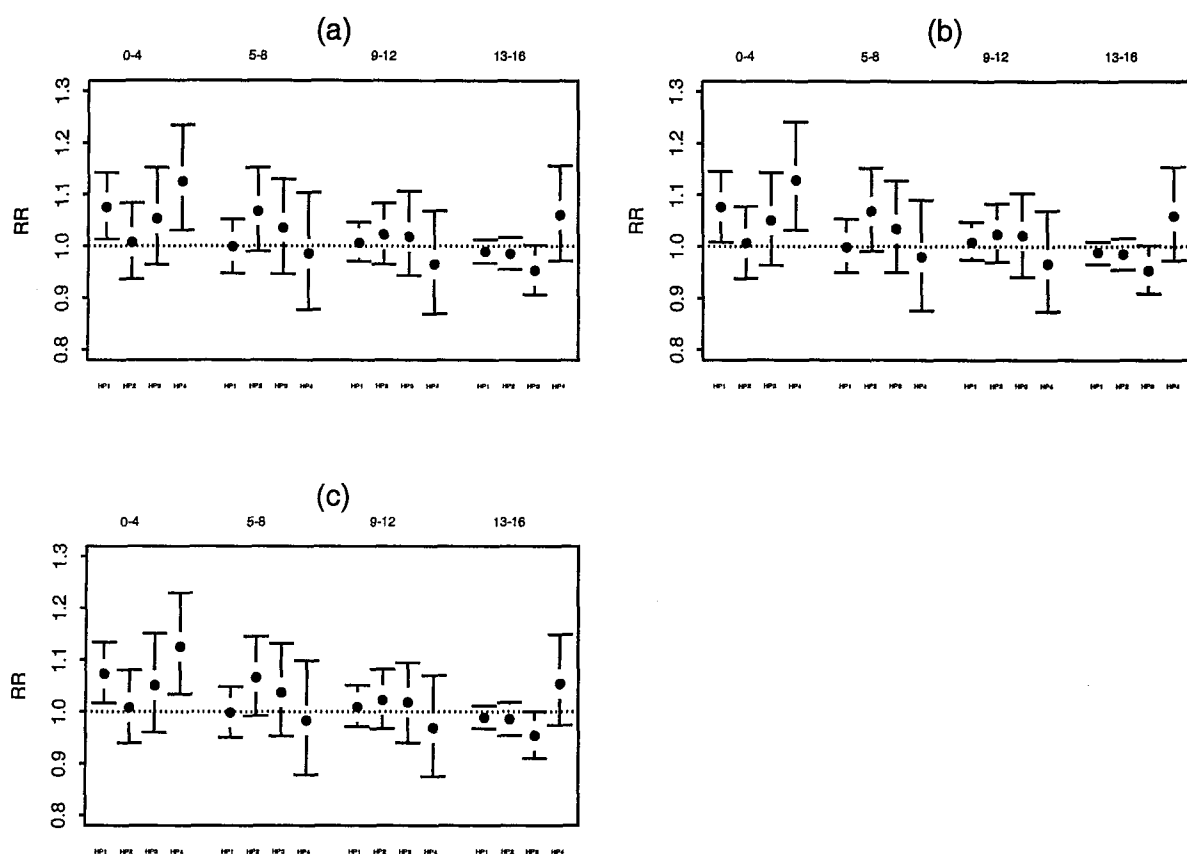


Figure 3.7: Relative risks, with 95% credible intervals for increase of 10 ppb of SO_2 for effects of exposure over the previous 0-4, 5-8, 9-12 and 13-16 years, with adjustment for deprivation. The three panels refer to models with different priors on the lagged effects; (a) independent (b) auto-regressive, smoothing over lags within health periods, and (c) auto-regressive, smoothing each lagged effect over health periods.

Chapter 4

Modelling the effect of cold winter temperatures on mortality

4.1 Introduction

In recent years, much attention has been given to the short-term, or acute, effects of air pollution and many studies have reported positive associations with both mortality and morbidity. For a review of a selection of these studies, see Chapter 5. In many of these studies, temperature, both through adjusting for day-to-day changes and also as a result of controlling for seasonality has been treated as little more than a nuisance variable which has to be controlled for. However, the effect of temperature is potentially an important determinant of health in its own right. There are many suggestions as to the possible mechanisms of the effect of cold temperature, of which Enquesselassie et al. (1993) describe three; (i) the failure to adapt physiologically to temperature changes or extremes causes stress on the heart, (ii) behavioral differences in summer and winter, including changes in diet, exercise and outdoor activity, and (iii) increased susceptibility to other causes which exhibit strong seasonal patterns, e.g. people already compromised by coronary disease may be more likely to suffer respiratory infections.

In this chapter, the potential short-term effects of temperature on health are considered, with particular focus on possible lagged effects and the effect of prolonged periods of cold. Such analyses have traditionally involved fitting a series of models, many of which contain highly correlated variables, i.e. the different lagged values of temperature, which can make the choice of a single, ‘final’ model highly problematic. An alternative approach is applied here, that of Bayesian model averaging (BMA) (which is described in Section 2.4.4), in which the inherent uncertainty in the model selection procedure is acknowledged, and posterior probabilities are calculated for each of the candidate models, allowing weighted averages of the parameter estimates and of predicted values to be calculated. The models presented are applied to data from London for the winters

within the period 1981-1997. The structure of the chapter is as follows. In Section 4.1.1, previous studies of the effects of cold temperatures on health are critically assessed. Section 4.2 describes the health and temperature data used in the analyses. Section 4.3 contains a description of an exploratory analysis, including details of the statistical models used (Section 4.3.1) and the results obtained (4.3.3). Section 4.4 contains the full analysis, with a description of the statistical methods in Section 4.4.1, together with details of the Bayesian Model Average approach (Section 4.4.2), which include a brief discussion of the choice of prior distributions and the computational methods and approximations that may have to be used to calculate some of the complex integrals that may be involved. Results from the full analyses are presented in Section 4.4.4.

4.1.1 Epidemiology of effects of cold temperatures on mortality

The associations between cold weather and increases in mortality have been observed over many years, first being reported in England and Wales by William Farr (Farr (1840) and Farr (1841)). An early study of daily mortality within winters was performed by the London County Council for 1954-57 (Gore and Shaddick (1958)), which reported an increase in cardiovascular mortality in the elderly during a period of cold temperature. Many of the studies considering the effects of cold temperatures on health have done so on a broad temporal scale, for example, comparing health outcomes in the winter months with those in the rest of the year. For example, Curwin and Devis (1988) used a linear regression model to compare deaths in winters (December-March) to those in summer (April-July) and Autumn (August-November) within England and Wales for 1949-85. They found the excess number of winter deaths was related both to cold temperatures and outbreaks of influenza. They concluded that over the entire study area of England and Wales, an average of 3.6 additional winter deaths were observed whenever an influenza death was recorded, and that a drop of 1°C below the average was associated with about 8000 excess winter deaths. They also found that the number of excess winter deaths was decreasing, by approximately 500 a year over the period of the study. In a further analysis of a recent subset of the data (1976-85), they found a similar increase in deaths per 1°C drop in temperature and noted that the number of influenza deaths observed in the early period (more than 15,000 a year) was greatly reduced, falling to less than 500 per year. A recent study by Aylin et al. (2001) compared the odds of death occurring in the winter period (December to March) compared to the rest of the year for those aged 65 and older, in Great Britain during 1986-96. They reported a significant association between mortality and temperature with a 1.5% increase in the odds of dying in winter for every 1°C lower 24 hour mean winter temperature, and an average of 30,000 excess winter deaths per year.

A number of studies have considered the effects of cold weather using daily time-series data, rather than a comparison over larger temporal scales. The majority of these have considered

associations only within the winter months, but others have attempted to model seasonal patterns in mortality and temperature over the entire year, before examining the short-term effects of changes in temperature.

Macfarlane (1977) analysed data for winters in 1958-65 within Greater London using data on mean daily temperatures from the meteorological station at Heathrow airport in addition to daily average levels of black smoke and SO₂. Estimating the effect of temperature was complicated by the fact that most short spells of cold temperature (3-5 days) were associated with temperature inversions, when the levels of pollution were also high. In times when pollution levels were relatively constant, an increase in cardiac deaths, mainly due to myocardial infarction, were observed during short periods of cold. Mackenbach et al. (1992) investigated the winter excess of all cause mortality in the Netherlands for 1979-87, using a Poisson regression model to relate nine-day moving averages of the observed number of daily deaths to that expected, adjusted for age and sex. Data from the entire year were used, with seasonality being modelled using a series of co-sine terms. A bimodal peak in all cause mortality was observed in the first months of the year, with the main contributions to an excess of 2964 deaths per winter period (comprising 2.5% of yearly deaths in the Netherlands), being cardiovascular (66%) and respiratory diseases (13%). They estimated that 12 excess winter deaths were associated with each influenza death, which is much higher than the 3.6 estimated for Britain by Curwin and Davis. By fitting models with and without seasonality terms, they attempted to estimate the period of excess winter mortality, finding the effect was greatest from November 11-April 14. In a later study, Mackenbach et al. (1993) assessed the effects of several meteorological variables including daily average, difference between daily minimum and maximum, precipitation, humidity and wind speed, together with air pollution (daily levels of SO₂). In order to incorporate an apparent non-linear relationship between average temperature and deaths, two categories were used, cold (< 16.5°C) and warm (> 16.5°C). Averages of these values for lagged periods of 1-5, 6-10, 11-15 and 16-20 days were used within a Poisson regression model together with lags of up to 20 days for pollution. When fitted in isolation, a significant effect of SO₂ was observed, but the effect was near zero when the meteorological variables were included in the model. They found that both high and low temperatures were associated with higher mortality, but that due to multi-collinearity, no obvious patterns could be observed in the lagged effects.

The Eurowinter Group (1997) assessed differences in the risk per 1°C fall in temperature (below 18°C) over several regions of Europe (north Finland, south Finland, Baden-Württemberg, the Netherlands, London, north Italy, Athens and Palermo). All regions showed more annual cold-related mortality than heat-related mortality, and over the seven regions, they estimated that the number of cold-related deaths was approximately 2003 per million (compared to 217 per million for heat related deaths). They found that the increases in risks were greater in warmer than in colder regions (Athens, 2.15% 95% CI (1.20-3.10), south Finland, 0.27% 95% (0.15-0.4)).

However, differences were also seen between regions with very similar mean winter temperatures (London, 7.6°C and northern Italy, 7.7°C with increases in risks of 1.37 and 0.51% per 1°C drop). Further analysis of data from southeast England, for 1979-94, (Donaldson (1997) and Keatinge et al. (2000)), suggested that three-day lagged values of temperature gave the steepest gradients for cold-related mortality, but that the effect reduced over the period of study (the risk per 1°C drop declining by 32%) and daily influenza deaths had little effect, in stark comparison to its importance in studies of the 1950-60s.

Excess deaths have also been associated with decreasing temperatures in countries which generally experience milder conditions than those encountered in Great Britain and northern Europe. In Barcelona, Spain, for example, where the average winter temperature is 10°C (yearly average, 15.7°C), Saez et al. (2000) estimated an increase in risk for IHD of 2.4% per 1°C drop below 4.7°C (increasing to approximately 4.4% when humidity was less than 70%). They analysed daily time-series of deaths (1986-1991), temperature, humidity and pollution (black smoke, SO_2 , NO_2 and O_3) using splines to investigate the possibility of non-linear relationship, and estimated a temperature threshold of 21°C , above which there was no effect of temperature on deaths, which increased to 23°C during periods of high humidity (above 85%). Similar results were seen in New South Wales, Australia (average daily temperature, 23°C), where coronary deaths were 20-40% more common in winter (June-August) and spring (September-November) than at other times of the year (Enquselassie et al. (1993)).

In places experiencing extreme cold, for example in Yekaterinburg, Russia where the average winter temperature is -6°C , the increase in deaths is somewhat smaller (Donaldson et al. (1998)). In Yatutsk, eastern Siberia, where the mean temperature for October-March is as low as -26°C , mortality remained unchanged as temperature decreased, even when temperatures fell to as low as -48°C (Donaldson et al. (1998)). This may have been due to differences in behaviour in colder compared to warmer climates, with people staying indoors in warm housing and wearing large amounts of warm clothing when outside (Eurowinter Group (1997)).

4.2 Data Description

The health data consist of daily counts of all cause mortality (ICD9, 0-999), extracted for all ages from national postcoded mortality data, for the period 1981-97. The data were obtained for all Enumeration Districts (EDs) within the London area, defined by ED codes beginning 01 (inner) and 02 (outer), containing an area roughly defined by the M25 orbital motorway.

Meteorological data were obtained from the Meteorological Office for the station at Kew Gardens, which recorded daily (24 hourly) minimum, maximum and average temperatures. As might be expected, there is very strong correlation between these three variables meaning that it would be

very difficult to isolate the effects of an individual measure of temperature from the others. For this reason, and as the particular interest here is in the effect of cold temperatures, the analyses is based on the minimum temperatures, but these could essentially be considered as a proxy for a whole range of climatic conditions. The study investigates the effect of cold temperatures in the winter seasons, defined as the four months November (1st) to February (28th, or 29th) .

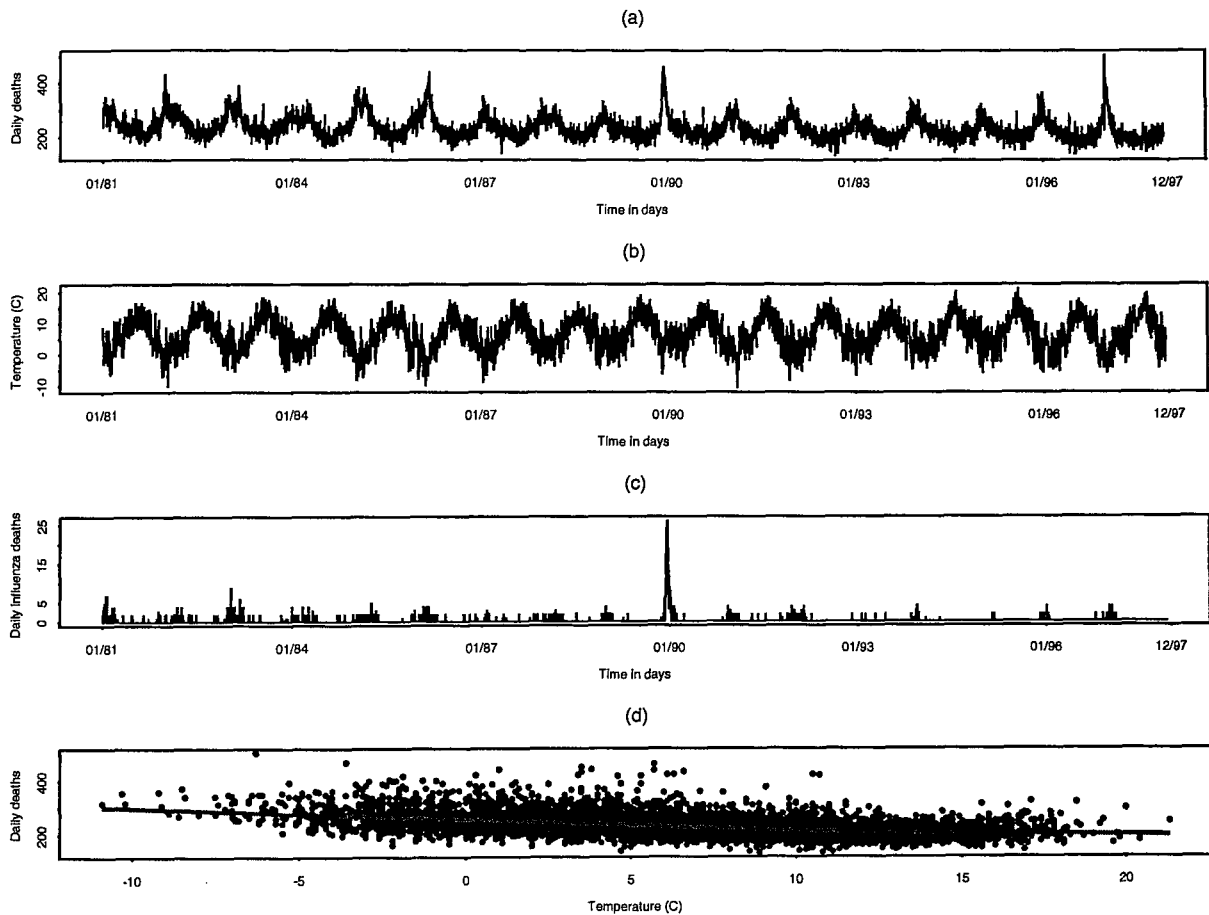


Figure 4.1: Time series plots of (a) daily deaths, (b) daily temperatures, (c) daily influenza deaths and (d) relationship between daily deaths and temperatures, London, winters (November–February), 1981-97.

Time series plots of daily deaths and temperatures can be seen in Figure 4.1. The seasonal patterns in temperature appear relatively constant. Table 4.1 shows the total number of deaths, together with the daily average and standard deviation, by year. There appears to be a gradual decline in the numbers of deaths over the study period. This suggests that any analyses should include the long term effects of time, i.e. year (it should be noted that the apparent drop in the total numbers of deaths during the last year of study is because data were not available for the last two weeks of the year; the average number of daily deaths is however consistent with the

previous values). The association between temperature and deaths is shown in Figure 4.1, which shows an increase in deaths with lower temperatures. Excess winter deaths have previously been shown to be strongly related to both cold temperatures and influenza deaths (e.g. Curwin and Devis (1988), Mackenbach et al. (1992)). Table 4.1 gives the total number of influenza deaths in each year of study, together with the daily averages (and standard deviations). Figure 4.1 also shows the number of daily influenza deaths over time. Although the daily counts are small (the daily average is approximately 0.1, with the winter average 0.5), they exhibit a number of peaks which indicate influenza epidemics, noticeably at the end of 1989, which coincided with a peak in all cause mortality.

Table 4.1: Summaries of all cause mortality and influenza deaths, by year, London, winters (November–February), 1981-97. Yearly totals, and means of daily counts are shown together with standard deviations.

	All deaths			Influenza deaths		
	total	mean	sd	total	mean	sd
1981	34907	288.5	36.0	133	0.4	1.0
1982	35003	289.3	33.8	77	0.2	0.6
1983	34912	288.5	33.7	116	0.3	0.9
1984	32248	264.3	26.2	49	0.1	0.5
1985	36212	299.3	38.8	77	0.2	0.6
1986	35829	296.1	48.9	66	0.2	0.6
1987	32341	267.3	30.8	36	0.1	0.4
1988	32917	269.8	20.9	52	0.1	0.5
1989	33836	279.6	68.2	347	1.0	3.5
1990	30245	250.0	30.1	84	0.2	0.9
1991	32400	267.8	35.1	38	0.1	0.4
1992	30686	251.5	31.2	33	0.1	0.4
1993	31047	256.6	31.3	43	0.1	0.5
1994	29393	242.9	30.2	2	0.0	0.0
1995	31454	260.0	33.6	22	0.1	0.3
1996	30568	250.6	39.9	17	0.0	0.3
1997	25348	258.7	71.9	28	0.1	0.4

Table 4.2 gives the average number of daily deaths by month and season, together with average (minimum) temperatures. Clear patterns are seen throughout the year, with the lowest temperatures being associated with higher numbers of deaths. The average number of deaths per day over the entire study period was 233, with a mean temperature of 6.6°C . In the winter months, the average minimum temperature was lower at 1.8°C , with the average number of daily deaths increasing to 276.

Table 4.3 gives the number of days at certain temperatures or colder, within the range 5°C to -2°C for periods of up to eight days. The average number of daily deaths is also given, and ranges from 272 at 5°C or less on a single day to 313 during periods of -2°C for eight consecutive days.

Table 4.2: Summary of all cause mortality and minimum temperatures, by month, season and over the entire study period (1981-97). Means of daily mean values are shown together with standard deviations.

	Deaths		Temp.	
	mean	sd	mean	sd
Jan	285.9	39.4	1.7	3.9
Feb	266.5	38.1	1.4	3.8
Mar	251.1	39.3	3.3	3.4
Apr	233.5	27.8	4.5	3.3
May	215.9	23.9	7.5	3.1
Jun	208.4	22.9	10.6	2.7
Jul	204.0	26.3	13.1	2.6
Aug	199.3	22.9	12.7	2.9
Sep	204.5	22.8	10.0	3.1
Oct	220.4	24.2	7.4	3.8
Nov	235.3	26.6	4.0	4.5
Dec	275.5	45.8	2.3	4.3
autumn	219.9	27.6	7.2	4.6
spring	233.5	34.3	5.1	3.7
summer	203.9	24.4	12.1	2.9
winter	276.3	42.0	1.8	4.0
overall	233.2	42.3	6.6	5.4

Increases in the number of deaths are observed both with decreasing temperature for a given number of days and also with increasing numbers of days at cold temperatures.

4.3 Exploratory analysis

4.3.1 Models

The number of daily deaths, Y_t , is examined with respect to daily (and lagged) values of temperature, X_t . One advantage in considering changes on a small temporal scale is that age-sex, or socio-economic structure of the population and many important potential confounding variables, such as smoking, are unlikely to confound the observed associations with temperature because the factors themselves, although unmeasured, do not vary daily with levels of temperature. However, there are other potential confounding variables that may also vary day-to-day, such as air pollution, which has been the focus of many studies of short-term changes in mortality (and morbidity). Studies investigating the short-term effects of air pollution, in which temperature is generally controlled for within the analysis rather than being the focus of attention, are reviewed in Chapter 5. As with all observational studies, there is, of course, still the possibility that the observed values may result from an association with an unmeasured factor, or one that is not included within the model being used. Over a prolonged study period, there may be long term

Table 4.3: Number of consecutive days at given temperature or less, with mean number of daily deaths during winters, London, 1981-97.

	Number of consecutive days							
	1 day	2 days	3 days	4 days	5 days	6 days	7 days	8 days
$\leq 5^{\circ}\text{C}$	1549	1362	1189	995	808	629	469	334
$\leq 4^{\circ}\text{C}$	1318	1086	894	718	543	406	275	175
$\leq 3^{\circ}\text{C}$	1144	888	699	538	382	273	181	99
$\leq 2^{\circ}\text{C}$	1002	740	567	418	278	194	122	63
$\leq 1^{\circ}\text{C}$	890	633	462	329	210	144	86	41
$\leq 0^{\circ}\text{C}$	795	545	383	266	164	114	62	29
$\leq -1^{\circ}\text{C}$	707	478	317	216	130	88	43	21
$\leq -2^{\circ}\text{C}$	629	419	262	178	105	68	31	15
	Mean daily deaths							
	1 day	2 days	3 days	4 days	5 days	6 days	7 days	8 days
$\leq 5^{\circ}\text{C}$	271.8	273.7	275.2	276.4	277.4	278.6	280.2	282.2
$\leq 4^{\circ}\text{C}$	273.4	276.1	278.3	280.7	282.7	285.7	287.9	290.6
$\leq 3^{\circ}\text{C}$	274.0	276.5	278.0	279.8	282.1	285.3	287.5	289.8
$\leq 2^{\circ}\text{C}$	275.7	279.3	281.4	284.0	286.3	290.3	292.8	295.1
$\leq 1^{\circ}\text{C}$	277.4	279.8	282.6	285.4	288.7	292.7	295.2	296.3
$\leq 0^{\circ}\text{C}$	277.5	280.8	284.6	288.2	288.5	289.3	294.5	298.2
$\leq -1^{\circ}\text{C}$	282.0	288.9	291.9	294.6	295.5	302.6	307.0	310.5
$\leq -2^{\circ}\text{C}$	284.2	293.7	298.5	300.6	302.0	310.0	313.3	313.2

changes in the patterns of mortality which should be incorporated into the model, achieved here by using a variable representing the year.

The number of daily deaths, Y_t , is assumed to be Poisson distributed, $Y_t \sim P(\mu_t)$ with the daily mean, μ_t being modelled as a function of lagged temperatures, X_{t-l} , and a yearly term, $year_t$. A number of previous studies have found significant relationships between daily counts of all cause mortality and influenza deaths (e.g. Curwin and Devis (1988), Mackenbach et al. (1992)), and their effect is assessed here by allowing the possibility of the number of daily influenza deaths, Z_t , to be included in the model.

$$\log(\mu_t) = \beta_0 + \beta_{year} year_t + \beta_{flu} Z_t + \sum_{l=0}^L \beta_l X_{t-l} \quad (4.1)$$

where β_l is the effect of the t^{th} lagged value of temperature ($t = 0$ indicates same day temperature), β_{year} is a linear term representing year and β_{flu} represents the effect of the number of daily influenza deaths.

Before embarking on fitting models within a formal framework, the relationships between mortality and temperature were explored by fitting General Additive Models (GAMs) (Hastie and Tibshirani (1990)). In a GAM, instead of assuming dependence on the sum of linear predictors, the outcome is assumed to be dependent on a sum of smooth functions of the predictors,

$X_j, j = 1, \dots, L.$

$$g\{E(Y)\} = \sum_j S_j(x_j) \quad (4.2)$$

where S_j are smoothing functions. GAMs therefore provide a flexible framework for controlling for non-linear dependence on potential covariates. Firstly, non-parametric smoothers (lowess) were fit to the different lags to examine the possibility of non-linear relationships, and possible thresholds.

$$\log(\mu_t) = S_1(X_t) + S_2(X_{t-1}) + \dots + S_7(X_{t-7}) \quad (4.3)$$

where S is the loess smoothing function.

A useful tool in smoothing data, is that of *splines* which allows the possibility of fitting almost any function representing μ_t (see Hastie and Tibshirani (1990)). To further investigate the possibility of thresholds, a series of spline models were fit, with different numbers of knot points and degrees of polynomial, again within the framework of GAMs (with S_j in (4.3) now being spline functions).

4.3.2 Inference

In the first, exploratory stage of examining the continuous effects of temperature, non-parametric relationships were fit within the framework of GAMs, which are fit using an iterative procedure known as the backfitting algorithm, which simultaneously fits all functions, before smoothing the partial residuals of each variable in turn and iterating until the individual functions no longer change (Hastie and Tibshirani (1990)). In the parametric models, for both continuous and discrete effects of temperature, the Poisson models are standard GLM and estimates of the parameters are found using maximum likelihood and iteratively re-weighted least squares (IRLS) (McCullagh and Nelder (1989)).

4.3.3 Results

The models described in Section 4.3 are now applied to the data from London. As discussed in Section 4.1.1, a number of studies have considered the short-term effects of cold temperatures whilst modelling temperatures throughout the entire year. This involves modelling the underlying seasonal patterns, using, for example, non-parametric smoothers or series of cosine/sine terms. The choice of these, and in particular their temporal resolution, can have a serious effect on the nature of the relationships between short-term changes of temperature and health outcome.

This can be seen in Figure 4.2 where the relationship between both temperature and mortality and time and mortality are shown in pairs of plots for various spans of time (using the lowess smoother), which decrease from the top to bottom. In the top pair, seasonality is inadequately modelled

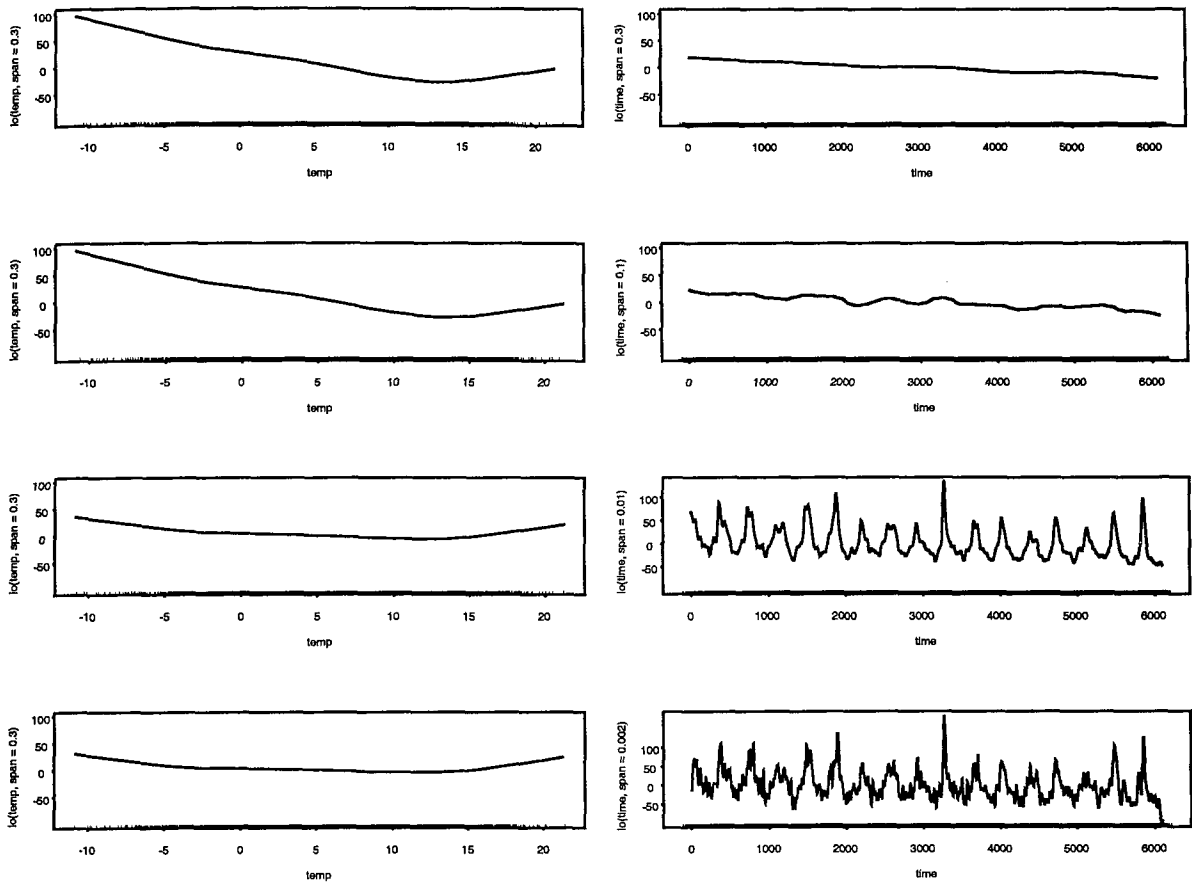


Figure 4.2: The effects of different spans on modelling the relationship between (i) mortality and temperature (ii) mortality and time (seasonality). The span used to smooth time increases for pairs of plots from top left to bottom right. In each pair of plots, the left shows temperature and the right time.

using a wide span (right hand plot). As seasonality is modelled more accurately, the relationship between temperature and mortality in cold weather (left hand plot) is less pronounced, as some of the explanatory power has been incorporated in to the seasonality term (the change of scale should be noted). When assessing the effect of short-term changes in the weather, care must be taken not to over-control for seasonality. An extreme case is seen in the bottom right hand pair of plots, where a 2 week span has been used to model time, giving a very good fit to the observed data, but causing the observed effect of cold weather on mortality to be greatly reduced. In the analyses that follows, only data from the winter months are considered, in an attempt to isolate the effects of cold temperature.

Figure 4.3 shows the result of fitting eight separate models, each containing a single lagged value of temperature, ranging from the same day value to a lag of seven days. The effect of temperature

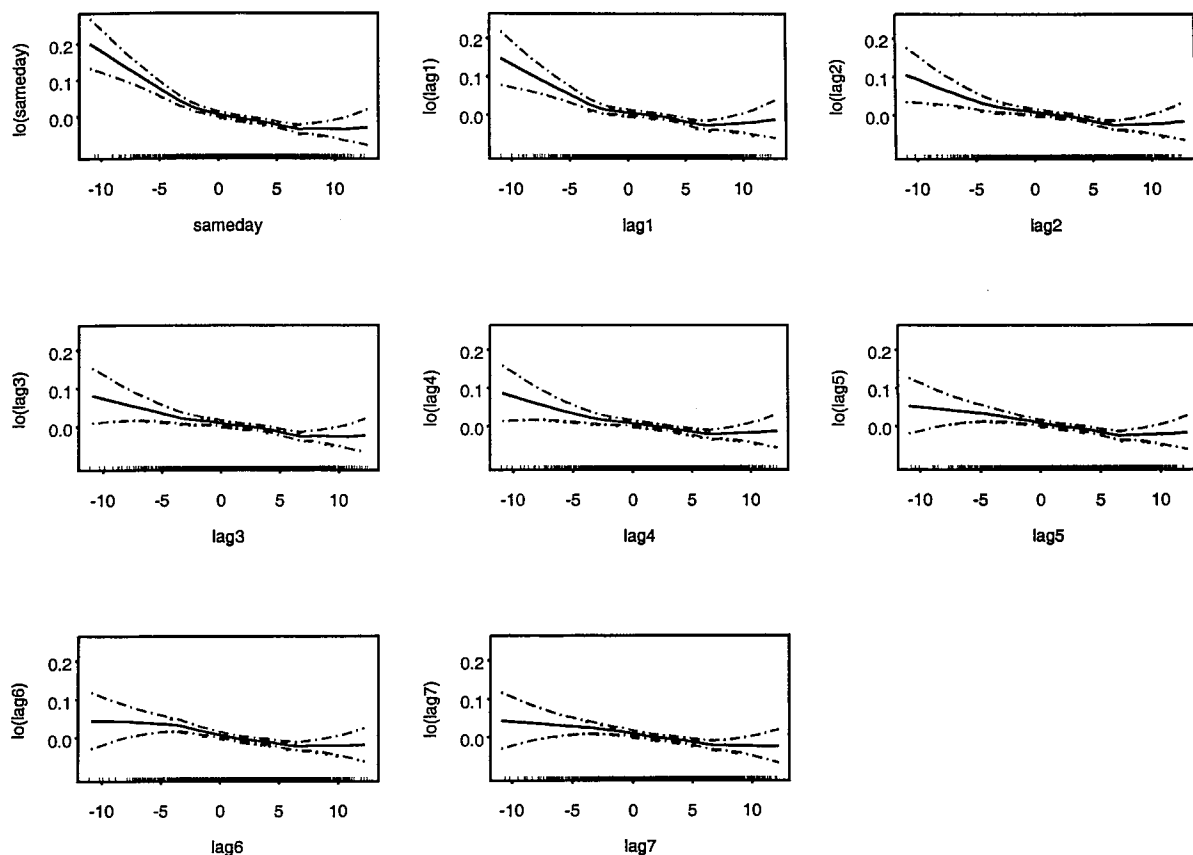


Figure 4.3: Modelling the relationships between daily deaths and different lags of temperature. Each panel is the result of fitting a lowess smoother (span=2/3) to each lag separately within a GAM. Dotted lines represent two (pointwise) standard errors.

in each is modelled as a smooth function (loess, span=2/3) within a GAM. Treated individually, each lag shows a strong negative association with mortality, with suggestion that the magnitude of the slope is reduced for the longer lags.

When the different lags are fitted together within a single model (Figure 4.4), the relationships with same day temperature is similar, however, the slopes associated with several of the lags are markedly non-linear, and in some cases the increase with warmer temperatures becomes more pronounced, e.g. one and four days. This is a result of the high collinearity between the variables, and will be a potential problem when interpreting the parameter estimates (and resulting relative risks) obtained from parametric models (GLMs). One approach to incorporating non-linear relationships within a parametric model framework is to use splines.

In Figure 4.5, a selection of the results of fitting a series of spline models to same day temperature are presented. The simplest case (top left panel) shows the result of fitting a single knot at 10^0C

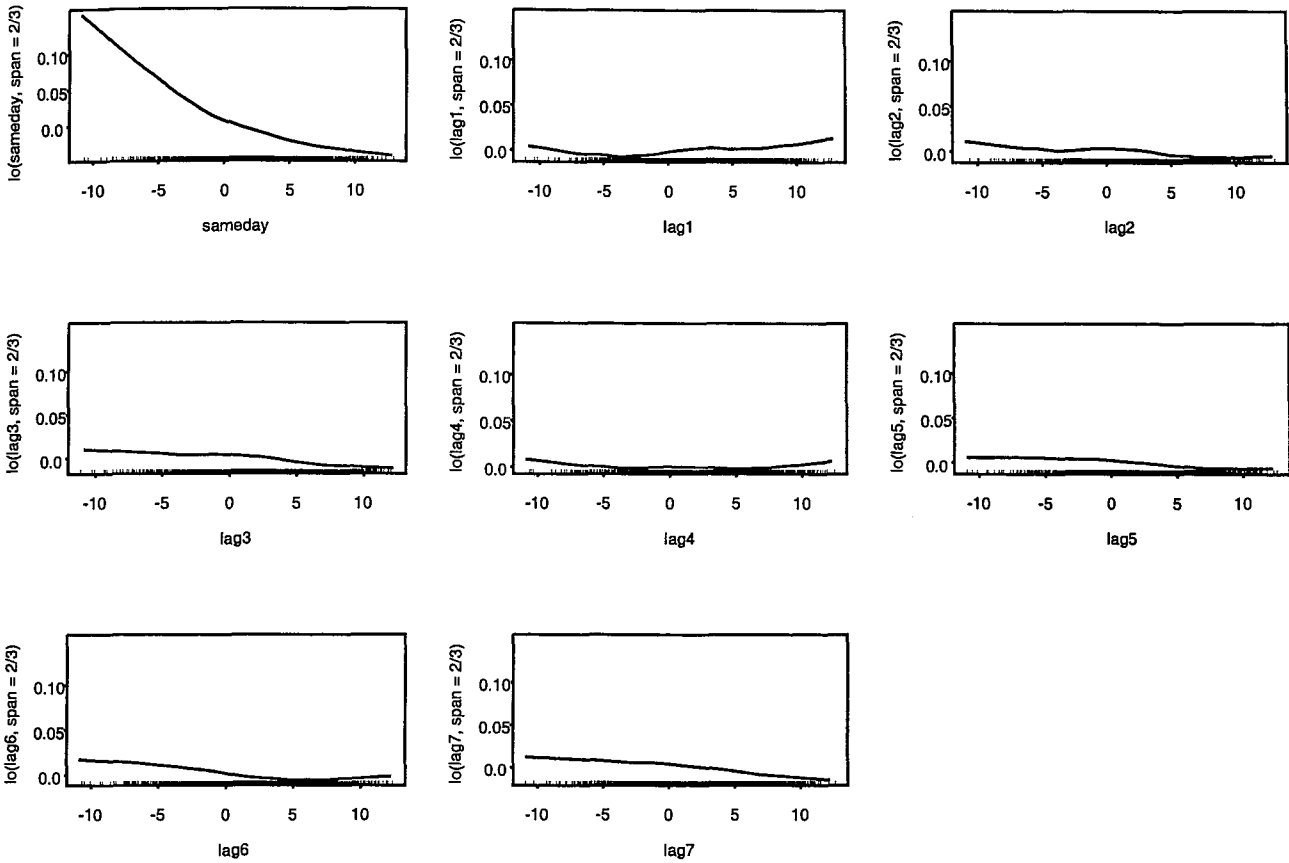


Figure 4.4: Modelling the relationships between daily deaths and different lags of temperature together in a single model, fitting a lowess smoother (span=2/3) to each lag within a GAM.

and allowing a polynomial of degree one. From left to right, the number of knots is increased, adding one firstly at 0°C and then at 10°C, and from top to bottom the degree of the polynomial is increase from one to three (a cubic spline). The appearance of the graphs is obviously highly determined by the arbitrary choice of knots, and to a lesser extent (in this example) by the degree of the polynomial. From inspection, there appears to be 2 distinct thresholds in the relationship, separating cold, mild and warmer temperatures.

The effect that fitting a threshold model can have on the resulting relative risks associated with the variable of interest is illustrated in Table 4.4. This shows the results of fitting the lagged effects of temperature as a single linear relationship (left hand side) and with a change point at 10°C (as seen in the top right panel of Figure 4.5). The top half of Table 4.4 gives the relative risks (and 95% CIs) from these models per 1°C decrease in temperature, with a single value presented from the non-threshold models and two for the threshold models, above and below the change point. When the lagged values are fitted separately in the non-threshold model, each results in a

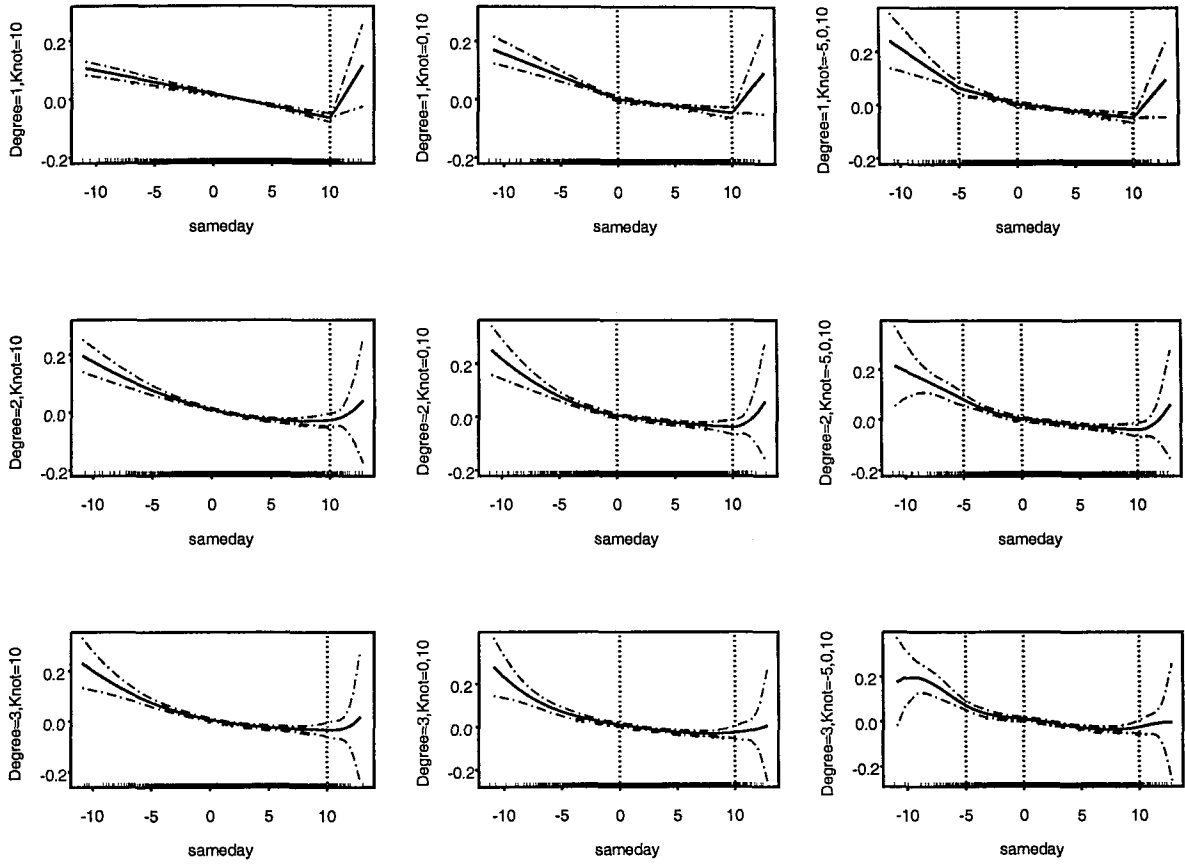


Figure 4.5: Results of fitting different spline models to the relationship between daily deaths and same day temperature. The number of knots increases from one (left hand panels) to three (right hand panels), and the degree of the polynomial from one (top) to three (bottom). Dotted lines represent two (pointwise) standard errors.

relative risk that is significantly greater than one. The same is seen for the colder ($< 10^0\text{C}$) part of the threshold model, with some non-significant values seen for the slope in the warmer section ($\geq 10^0\text{C}$). The largest excess in risk, 0.84% (0.76-0.91), is seen for sameday temperature at less than 10^0C in the threshold model. The negative association with warmer temperatures is also significant for same day temperature. The bottom half of Table 4.4 gives the relative risks (and 95% CIs) when the lagged effects are all fitted together in a single model. Again, the strongest effect (significant) is observed for same day temperature, but many of the other lags now show non-significant effects. As observed in Figure 4.4, there is no clear pattern in the risks of the lagged effects, although there is some suggestion of a decrease with increasing lag.

Table 4.4: Relative risks with 95% confidence intervals for decrease of 1°C in lagged temperature. Left hand results are for a non-threshold model, right hand side for a model with a change point at 10°C. The top half of the table gives results for models with lags fitted separately, the bottom when fitted together in a single model.

	No threshold			<10°C			≥ 10°C		
	RR	lower	upper	RR	lower	upper	RR	lower	upper
Lags fitted separately									
Same day	1.0079	1.0063	1.0094	1.0084	1.0076	1.0091	0.9458	0.9269	0.9650
Lag 1	1.0059	1.0043	1.0074	1.0059	1.0043	1.0074	0.9997	0.9886	1.0109
Lag 2	1.0050	1.0035	1.0066	1.0050	1.0035	1.0066	1.0003	0.9940	1.0066
Lag 3	1.0050	1.0035	1.0066	1.0050	1.0035	1.0066	1.0002	0.9940	1.0064
Lag 4	1.0046	1.0030	1.0061	1.0046	1.0030	1.0061	0.9995	0.9933	1.0058
Lag 5	1.0047	1.0031	1.0063	1.0047	1.0031	1.0063	1.0004	0.9946	1.0062
Lag 6	1.0048	1.0032	1.0063	1.0047	1.0032	1.0063	1.0006	0.9953	1.0060
Lag 7	1.0044	1.0028	1.0060	1.0044	1.0028	1.0060	0.9995	0.9937	1.0054
Lags fitted together									
Same day	1.0065	1.0045	1.0084	1.0071	1.0051	1.0091	0.9919	0.9145	1.0757
Lag 1	1.0000	0.9977	1.0023	1.0002	0.9979	1.0025	0.9881	0.9699	1.0067
Lag 2	1.0007	0.9984	1.0030	1.0004	0.9981	1.0027	1.0074	0.9917	1.0233
Lag 3	1.0015	0.9992	1.0038	1.0015	0.9991	1.0038	1.0052	0.9873	1.0234
Lag 4	1.0003	0.9980	1.0026	1.0004	0.9981	1.0028	0.9929	0.9751	1.0111
Lag 5	1.0010	0.9987	1.0033	1.0009	0.9986	1.0032	1.0009	0.9825	1.0196
Lag 6	1.0011	0.9988	1.0034	1.0008	0.9985	1.0031	1.0126	0.9945	1.0310
Lag 7	1.0018	0.9999	1.0038	1.0020	1.0000	1.0039	0.9935	0.9782	1.0090

4.4 Full Analysis

4.4.1 Models

The number of daily deaths, Y_t , is again assumed to be Poisson distributed, $Y_t \sim P(\mu_t)$ with

$$\log(\mu_t) = \beta_0 + \beta_{year} year_t + \beta_{flu} Z_t + \sum_{l=0}^L \beta_l X_{t-l} \quad (4.4)$$

where β_l is the effect of the t^{th} lagged value of temperature ($t = 0$ indicates same day temperature).

There is a strong possibility of over dispersion in the Poisson models (i.e., where the variance is greater than the mean) arising from the presence of unmeasured confounders or data anomalies. This will lead to confidence intervals for the estimates of risk being too narrow and changes in deviances, used to compare models, being too small. Further discussion of the effects of over-dispersion, and ways in which it can be modelled can be found in Chapter 3. An attempt to correct these effects, and to assess the degree of over-dispersion, can be made using a quasi-likelihood model (McCullagh and Nelder (1989)), in which the usual Poisson assumption, that the variance is equal to the mean, is relaxed, allowing it to be a function of the mean and an unknown constant, i.e. $Var(Y_i) = cE(Y_i)$. The quasi-likelihood approach requires specification of only of the mean-variance relationship rather than the full likelihood. The estimate of the

‘dispersion parameter’, c , can be obtained and used to adjust the variances and the changes in deviances between models.

4.4.2 Bayesian model averaging

The traditional approach to model ‘selection’, generally involves choosing a single model which is then considered to be correct when drawing inference about the parameters of interest or performing prediction. This ignores the inherent uncertainty associated with the actual modelling process, in particular the uncertainty which arises through selecting a subset of variables for inclusion, together with that due to distributional assumptions and/or transformations. A discussion of the theoretical and practical aspects of model selection can be found in Chapter 2, and reviews can be found in Draper (1995), Hocking (1976), Miller (1990) and Chatfield (1995).

Consideration should be given to the fact that inference about an unknown parameter will be dependent on the model under examination and on the set of other variables included in that model. In order to calculate posterior probabilities for the parameters, one should condition not only on the observed data but also on the model in question. In Bayesian model averaging, rather than choosing a single model, information from several models is combined to obtain averaged estimates and/or predicted values.

Given m candidate models, $M_1, \dots, M_m$, each with prior probability, $p(M_i)$ (the simplest example being where all the models are considered equally likely before any data are observed, i.e. $p(M_k) = 1/m$), then the posterior probability for model, M_k can be found using Bayes’ theorem

$$p(M_k|y) = \frac{p(y|M_k)p(M_k)}{\sum_{l=1}^m p(y|M_l)p(M_l)} \quad (4.5)$$

Obtaining the posterior model probabilities facilitates the calculation of Bayes Factors. In common with traditional likelihood methods (such as the comparison of deviances in GLMs), Bayes factors give an indication of how much more likely the data are to have arisen under one model than another, but unlike the LR tests, the models do not have to be nested. Further details on Bayes factors, their interpretation and implementation can be found in Kass and Raftery (1995), O’Hagan (1995) and in Chapter 2.

When the posterior model probabilities have been obtained, the weighted posterior distributions of the individual parameters can be calculated, allowing the effect of individual covariates to be assessed. The weighted posterior density, mean and variance of the parameters of interest can be calculated. Unlike the traditional approach of selecting just one model in which the estimate of a parameter of interest might be highly affected by the choice of which other variables were in/excluded in the model, this estimate allows for the fact that other variables may also be in the model. Although, in the linear case, the averaged coefficients have the same interpretation as in normal regression, i.e. they represent the change in the response associated with a unit change

in the explanatory variable, this may not be the case in non-linear examples, such as the log-linear models used in a Poisson regression used here. However, this is of little concern when performing prediction, where it is the combined explanatory power of the variables on the response that is important, rather than their individual effects. Further details of the interpretation of the parameters, and of prediction, can be found in Section 2.4.4.

Priors

As in any Bayesian analysis, prior distributions have to be assigned to all the parameters, and, in addition, to the models themselves. It is assumed that the priors for the regression coefficients, $\beta_j|M_k$, and the intercept terms, $\beta_0|M_k$ are both normally distributed, $N(\mu_{\beta_j}, \sigma_{\beta_j}^2)$ and $N(\mu_{\beta_0}, \sigma_{\beta_0}^2)$ respectively. *A priori*, the simplest case assumes that all models are considered equally likely, i.e. $p(M_k) = 1/m$, where m is the number of candidate models.

4.4.3 Inference

The Poisson models are standard GLM and estimates of the parameters are found using maximum likelihood and iteratively re-weighted least squares (IRLS) McCullagh and Nelder (1989). The Quasi-likelihood uses an adaption of the IRLS algorithm McCullagh and Nelder (1989).

The Bayes factors for comparing two models within the BMA are calculated using an approximation suggested by Raftery and Richardson (1996), which uses values of the MLE, Deviance and the Fisher information matrix, I_β which are widely available when fitting generalised linear models (McCullagh and Nelder (1989)). If the prior distribution of β_k for a particular model, M_k is such that $E(\beta_k|M) = \omega$ and its variance is W , then the following approximation can be used

$$2 \log B_{kl} \approx \chi^2 + (E_k - E_l) \quad (4.6)$$

where $\chi^2 = 2\{l_n(\hat{\beta}|y)_k - l(\hat{\beta}|y)_l\}$, the standard Deviance statistic between two nested models (M_k nested within M_l), and E_k is a penalty term which guards against models that provide a good fit simply because they contain a large number of variables, and thus have greater explanatory power, however limited the increase.

$$E_k = 2\lambda_k(\theta_k) + \lambda'_k(\theta_k)'(F_k + G_k)^{-1}\{2 - F_k(F_k + G_k)^{-1}\}\lambda'_k(\theta_k) - \log |F_k + G_k| + r_k \log(2p) \quad (4.7)$$

where $G_k = W_k^{-1}$, $\lambda_k(\theta_k) = \log P(\beta_k|M_k)$, the log of the prior of β_k and $\lambda'_k(\beta_k)$ is the vector (of length r_k) of first derivatives of $\lambda_k(\beta_k)$ with respect to the elements of θ_k . This approximation is of the order, $O_P(n^{-1})$ when the expected Fisher information is used for F_k , but this can be improved to $O_P(n^{-1/2})$ if the observed information is used.

4.4.4 Results

The results of fitting the different lagged values, from same day to 7 day lag, separately are shown in Table 4.5. Relative risks (and 95% confidence intervals) are given, firstly for standard Poisson models, and secondly using a quasi-likelihood model. The deviances and degrees of freedom are also presented, together with the dispersion parameters from the quasi-likelihood models. Changes in deviances are given for comparison between the null model and one containing a yearly term (second row), and then between the model containing year and one additionally containing daily influenza deaths. Both the yearly term and daily influenza deaths were found to be highly significant, even after adjusting the change in deviance for over-dispersion. There is a small, but significant, decline in the number of deaths by year over the period of study, and an increase of 2.8% (2.6-2.9) in the number of daily deaths associated with each influenza death.

The change in deviances for models containing a lagged value of temperature, adjusted for over-dispersion, are also given, with reference to the model containing the number of influenza deaths and a linear term representing year. Each of the individual lags produced a significant change in model fit when added separately, with the largest change in deviance being found for the same day value of temperature.

Table 4.5: Relative risks with 95% confidence intervals for decrease of 1°C in lagged temperature. Results are given for Poisson and quasi-likelihood models with individual lagged effects fitted in separate models, together with deviances, degrees of freedom and changes in deviances (adjusted for over-dispersion). Dispersion parameters from the quasi-likelihood models are also given (disp). Changes in deviances are for cumulative addition of yearly term and daily influenza deaths to null model, and then for each lagged value fitted separately to model containing year and influenza terms.

	Poisson			Quasi						
	RR	lower	upper	RR	lower	upper	disp	df	Dev.	Chg. Dev.
null	-	-	-	-	-	-	6.8	2037	13526.1	-
year	0.9898	0.9893	0.9903	0.9898	0.9884	0.9911	6.2	2036	12166.4	-219.1
flu	1.0276	1.0263	1.0289	1.0276	1.0247	1.0306	5.3	2035	10560.8	-301.9
Same day	1.0079	1.0072	1.0086	1.0079	1.0063	1.0094	5.0	2034	10045.2	-102.4
Lag 1	1.0059	1.0052	1.0066	1.0059	1.0043	1.0074	5.2	2034	10274.7	-55.5
Lag 2	1.0050	1.0044	1.0057	1.0050	1.0035	1.0066	5.2	2034	10349.4	-40.7
Lag 3	1.0050	1.0043	1.0057	1.0050	1.0035	1.0066	5.2	2034	10351.8	-40.2
Lag 4	1.0046	1.0039	1.0053	1.0046	1.0030	1.0061	5.2	2034	10387.7	-33.1
Lag 5	1.0047	1.0040	1.0054	1.0047	1.0031	1.0063	5.2	2034	10378.7	-34.9
Lag 6	1.0048	1.0041	1.0054	1.0048	1.0032	1.0063	5.2	2034	10373.6	-35.9
Lag 7	1.0044	1.0037	1.0051	1.0044	1.0028	1.0060	5.2	2034	10400.1	-30.7

The results of fitting the different lagged values together in a single model is shown in Table 4.6. A sequence of models are presented, starting with one containing a yearly term, influenza deaths and

same day temperature, with each of the subsequent models including an additional lagged value of temperature. The first three models, without lagged effects of temperature, are as in Table 4.5 and are included for comparison. Here, the changes in deviances provide a comparison between nested models, i.e. they assess the additional effect of adding the previous lagged value. After allowing for over-dispersion, only same day temperature has a significant increase in risk associated with a 1°C decrease in temperature (0.65%, (0.45-0.84)). The largest change in deviance occurred when same day temperature was added to a model containing influenza deaths and a yearly term. The addition of the one day lagged value did not result in a significant change in model fit, as was the case when the remaining lagged values were fitted sequentially.

Table 4.6: Relative risks with 95% confidence intervals for decrease of 1°C in lagged temperature. Results are given for Poisson and quasi-likelihood models with individual lagged effects fitted together in a single model. Deviances and degrees of freedom are given for series of models containing given lagged value and all previous lags. Changes in deviances (adjusted for over-dispersion) are given for cumulative addition of yearly term and daily influenza deaths to null model, and then between models differing sequentially by a single lagged value. Dispersion parameters from the quasi-likelihood models are also given (disp).

	Poisson			Quasi						
	RR	lower	upper	RR	lower	upper	disp	df	Dev.	Chg. Dev.
null	-	-	-	-	-	-	6.8	2037	13526.1	-
year	0.9898	0.9893	0.9903	0.9898	0.9884	0.9911	6.2	2036	12166.4	-219.1
flu	1.0276	1.0263	1.0289	1.0276	1.0247	1.0306	5.3	2035	10560.8	-301.9
Same day	1.0065	1.0056	1.0073	1.0065	1.0045	1.0084	5.0	2034	10045.2	-102.4
Lag 1	1.0000	0.9989	1.0010	1.0000	0.9977	1.0023	5.0	2033	10032.8	-105.0
Lag 2	1.0007	0.9996	1.0017	1.0007	0.9984	1.0030	5.0	2032	10002.8	-111.3
Lag 3	1.0015	1.0005	1.0025	1.0015	0.9992	1.0038	5.0	2031	9970.4	-118.2
Lag 4	1.0003	0.9993	1.0013	1.0003	0.9980	1.0026	5.0	2030	9956.7	-121.0
Lag 5	1.0010	0.9999	1.0020	1.0010	0.9987	1.0033	5.0	2029	9931.8	-126.3
Lag 6	1.0011	1.0001	1.0021	1.0011	0.9988	1.0034	5.0	2028	9906.5	-131.6
Lag 7	1.0018	1.0009	1.0027	1.0018	0.9999	1.0038	5.0	2027	9889.9	-135.2

Models for the lagged effects of temperature are now examined via Bayesian model averaging. The prior distributions for each of the parameters in the different models were assumed to be Normal with zero mean and unit variance, $\beta_j|M_k \sim N(0,1)$, and the intercept term assumed to be normally distributed with mean 0 and unit variance, $\beta_0 \sim N(0,1)$. Each of the models was considered equally likely, with $p(M_k) = 1/1024 = 2^p$, where $p = 10$ is the number of parameters under consideration (values from same day to 7 day lag, the yearly term and influenza deaths). The different models under consideration can be seen in Figure 4.6, which shows the parameters included in each of the 1024 candidate models, with the dotted line separating those models that include same day temperature from those that don't. In Figure 4.7, the results of fitting the models can be seen. The changes in deviance for each of the models are shown (in comparison

with a model including just the yearly term, which was considered fixed). As might be expected, the more complex models have higher values, but these are reduced by the penalty term ($E_k - E_0$), used in calculating the approximations of the Bayes factors. There is a clear difference between models including the same day temperature and those that exclude it, a result that is consistent with the initial analyses seen in Section 4.3.3. The results from the ten models which had non-negligible posterior model probabilities can be seen in Table 4.7. The parameter estimates for each of the models are given, together with their posterior model probabilities and resulting weighted averages of the coefficients and standard deviations, and a estimate of the support for including each variable in the model, the sum of the posterior model probabilities for the models that include that variable. The posterior distribution of the model probabilities is dominated by two models, having probabilities of 69% and 30%, indicating that there is a great deal of information within the data with which to update the priors. Both of these models contain the same day temperature, 3 day lagged value and a later lag (7 and 6 day respectively). The probabilities for the other models are much smaller, but all with non-negligible posterior model probabilities contain same day lagged temperature together with one or more later lags. The probability that the same day temperature should be included in the model is 1, as it is included in each model that has a non-zero probability. The probability that the 3 day lagged values is slightly lower at 99% and the probabilities that the 6 and 7 day lagged values should be included are of 30% and 70% respectively, with much lower values for the 4 and 5 day lagged values (approximately 0.3% and 0.7% respectively).

Table 4.7: Results from BMA for models with non-zero posterior model probabilities (pmp) when modelling the effect of lagged values of temperature. Coefficients are given from each of the models, together with support for including each variable in the model. All models include yearly term and daily influenza deaths.

Model	pmp	same day	lag1	lag2	lag3	lag4	lag5	lag6	lag7
914	0.6876	-0.0068	-	-	-0.0024	-	-	-	-0.0028
915	0.2969	-0.0067	-	-	-0.0023	-	-	-0.0028	-
902	0.0056	-0.0070	-	-	-	-	-0.0022	-	-0.0026
899	0.0033	-0.0072	-	-	-	-	-	-0.0034	-
906	0.0030	-0.0069	-	-	-	-0.0021	-	-	-0.0028
898	0.0013	-0.0074	-	-	-	-	-	-	-0.0033
918	0.0010	-0.0066	-	-	-0.0020	-	-0.0016	-	-0.0023
916	0.0007	-0.0066	-	-	-0.0022	-	-	-0.0017	-0.0018
907	0.0002	-0.0069	-	-	-	-0.0019	-	-0.0027	-
931	0.0001	-0.0066	-	-0.0018	-	-	-	-0.0030	-
$p(X_j \text{ included})$		1.0000	0.0000	0.0001	0.9862	0.0032	0.0066	0.3012	0.6992
bma mean		-0.00666	-	0.00000	-0.00230	-0.00001	-0.00001	-0.00083	-0.00192
bma sd		0.00037	-	0.00002	0.00046	0.00012	0.00017	0.00128	0.00130

For reasons discussed in Section 4.4.2, the averaged coefficients will not have the same interpretation as in normal regression, i.e. they don't represent the change in the response associated with

a unit change in the explanatory variable. However, it is possible to calculate such a measure, although it will be non-linear in X . Table 4.8 gives such values for a range of temperatures (5°C to -2°C) for each of the lagged variables (in each case the other variables are kept constant). Despite the non-linear relationship, there is little change over the small range of temperatures considered here, as would be expected when dealing with small increases in risk, as $\exp(\beta X) \approx 1 + \beta X$.

Table 4.8: Relative risks corresponding to a unit change (drop of 1°C) for same day and lagged temperatures (up to 7 days), calculated using Bayesian Model Averaging.

	same day	lag1	lag2	lag3	lag4	lag5	lag6	lag7
5°C								
4°C	1.00675	1.00000	1.00000	1.00235	1.00001	1.00001	1.00085	1.00197
3°C	1.00675	1.00000	1.00000	1.00235	1.00001	1.00001	1.00085	1.00197
2°C	1.00675	1.00000	1.00000	1.00235	1.00001	1.00001	1.00086	1.00198
1°C	1.00675	1.00000	1.00000	1.00235	1.00001	1.00001	1.00086	1.00198
0°C	1.00675	1.00000	1.00000	1.00235	1.00001	1.00001	1.00086	1.00198
-1°C	1.00675	1.00000	1.00000	1.00235	1.00001	1.00001	1.00086	1.00198
-2°C	1.00675	1.00000	1.00000	1.00235	1.00001	1.00001	1.00086	1.00198

The result of using the model averaging approach to produce predictions for a number of different temperature scenarios can be seen in Table 4.9 and Figure 4.8. The expected number of deaths was calculated for a range of temperatures (5°C to -2°C) over periods of time (1 to 8 days) for each of the candidate models and a weighted average (using the posterior model probabilities) calculated. The predicted number of deaths, together with standard deviations, can be seen in Table 4.9. This shows an increase of approximately 13 deaths per day when the (same day) temperature falls to -2°C (in comparison to 5°C). There is no change in the predicted number of excess deaths when temperature is cold for 2 or 3 days, as the models containing the shorter lags had zero, or near zero, posterior probabilities. There is an increase at four consecutive days of cold (3 day lag), little change in the estimate for 5 and 6 days, and then increases observed for 7 and 8 consecutive days, reflecting the effects of the two models containing 6 and 7 day lags with large posterior model probabilities. The effects of the posterior model probabilities, via the inclusion of different lagged values, on the predicted values can be seen in Figure 4.8. The expected number of excess deaths is given for a range of temperatures (5°C to -2°C) over periods of time (1 to 8 days) in comparison to that for a single day at 5°C . The increases at 3, 6 and 7 day lags (representing 4, 7 and 8 consecutive days of cold) observed in Table 4.9 can clearly be seen.

The standard deviations of the averaged predicted values are generally of similar magnitude, with the variability increasing for the more extreme temperature scenarios, i.e. long periods at very cold temperatures, which are not well represented in the data. The variability for the predicted values for 7 consecutive days of cold are noticeably greater than the others. This is due to greater

variability in the predicted values from the different models when the 6 day lagged value are used. The two ‘main’ models with large posterior model probabilities both include same day and 3 day lagged temperature, the difference between them being whether they also include the 6th or 7th lagged value. When using any of the lagged values up to and including the 5th day lagged value, there will be little between model variability in the predicted values, i.e. the two ‘main’ models produce very similar values. However, when using the 6 day lagged value, there are clear differences between the predictions from the model that includes the 6 day lagged value and the one that does not, resulting in larger between model variability (the second term in equation 2.44). This variability decreases again when considering 8 consecutive days of cold, as the values of the 6th and 7th lag will be the same (for example, when predicting the excess deaths when it is -2^o (for eight consecutive days) and, as their coefficients in the two models are similar, the predicted values are again similar.

Table 4.9: Results from using Bayesian model averaging for prediction using a range of temperature scenarios. Expected daily deaths and corresponding standard deviations are given for a range of low temperatures (5^oC to -2^oC) for periods of up to eight days.

Temperature (°C)	Predicted values							
	same day	lag1	lag2	lag3	lag4	lag5	lag6	lag7
5	276.1	-	-	-	-	-	-	-
4	277.9	277.9	277.9	278.6	278.6	278.6	278.8	279.3
3	279.8	279.8	279.8	281.1	281.1	281.1	281.5	282.6
2	281.6	281.6	281.6	283.6	283.6	283.6	284.3	286.0
1	283.5	283.5	283.5	286.1	286.2	286.2	287.1	289.3
0	285.4	285.4	285.4	288.7	288.7	288.8	290.0	292.8
-1	287.3	287.3	287.3	291.3	291.3	291.4	292.8	296.2
-2	289.3	287.3	289.3	293.9	294.0	294.0	295.7	299.1
Temperature (°C)	Standard deviations							
	same day	lag1	lag2	lag3	lag4	lag5	lag6	lag7
5	0.9	-	-	-	-	-	-	-
4	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.8
3	0.9	0.9	0.9	0.8	0.8	0.8	1.1	0.8
2	0.9	0.9	0.9	0.9	0.8	0.8	1.4	0.8
1	0.9	0.9	0.9	0.9	0.9	0.9	1.7	0.8
0	0.9	0.9	0.9	0.9	0.9	0.9	2.0	0.8
-1	1.0	1.0	1.0	1.0	1.0	1.0	2.3	0.9
-2	1.0	1.0	1.0	1.1	1.1	1.0	2.7	1.0

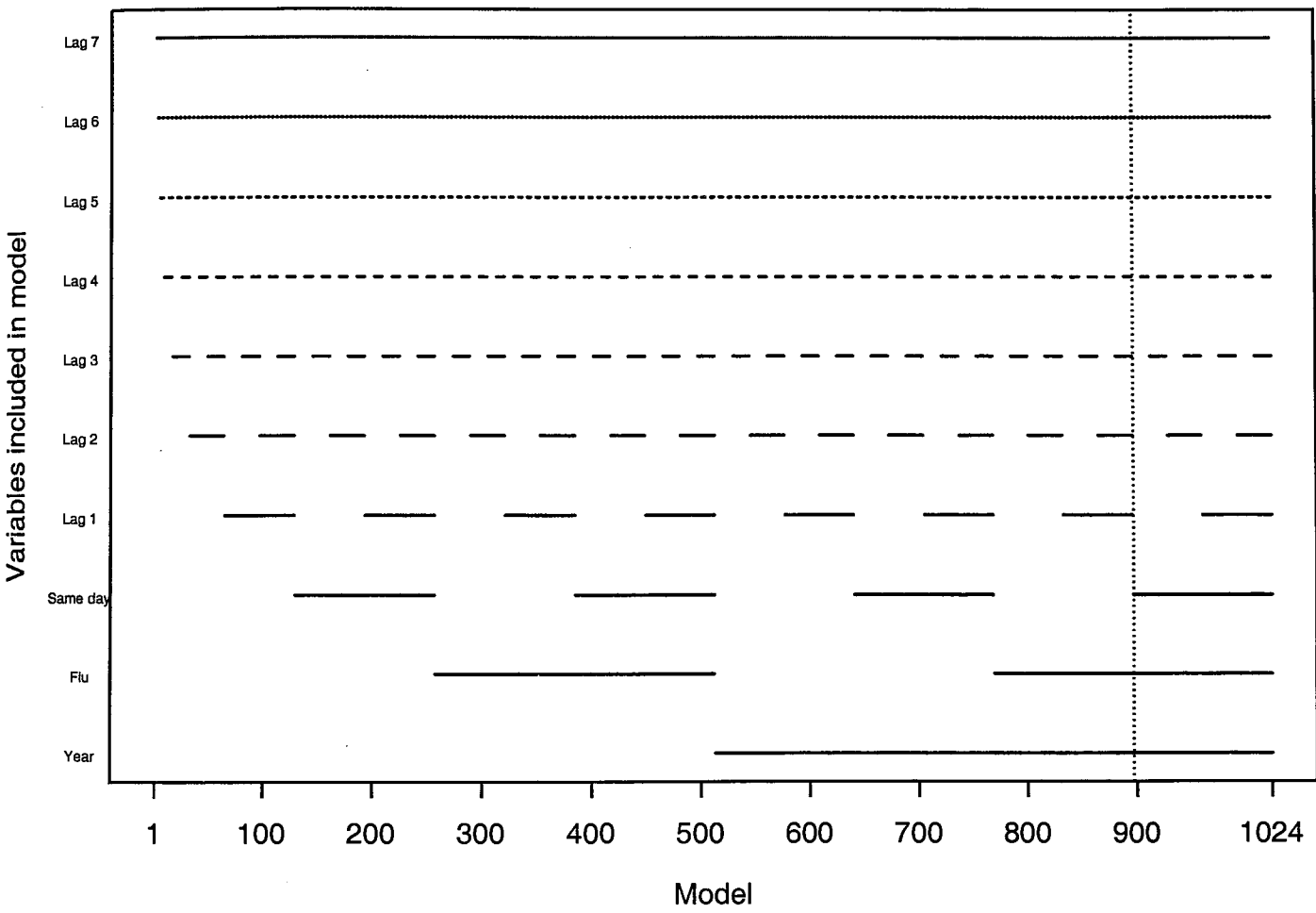


Figure 4.6: Model setup for BMA analysis of lagged effects of temperature: variables that are included in each of the candidate models. Models to the right of the dotted line are those containing the same day effect of temperature.

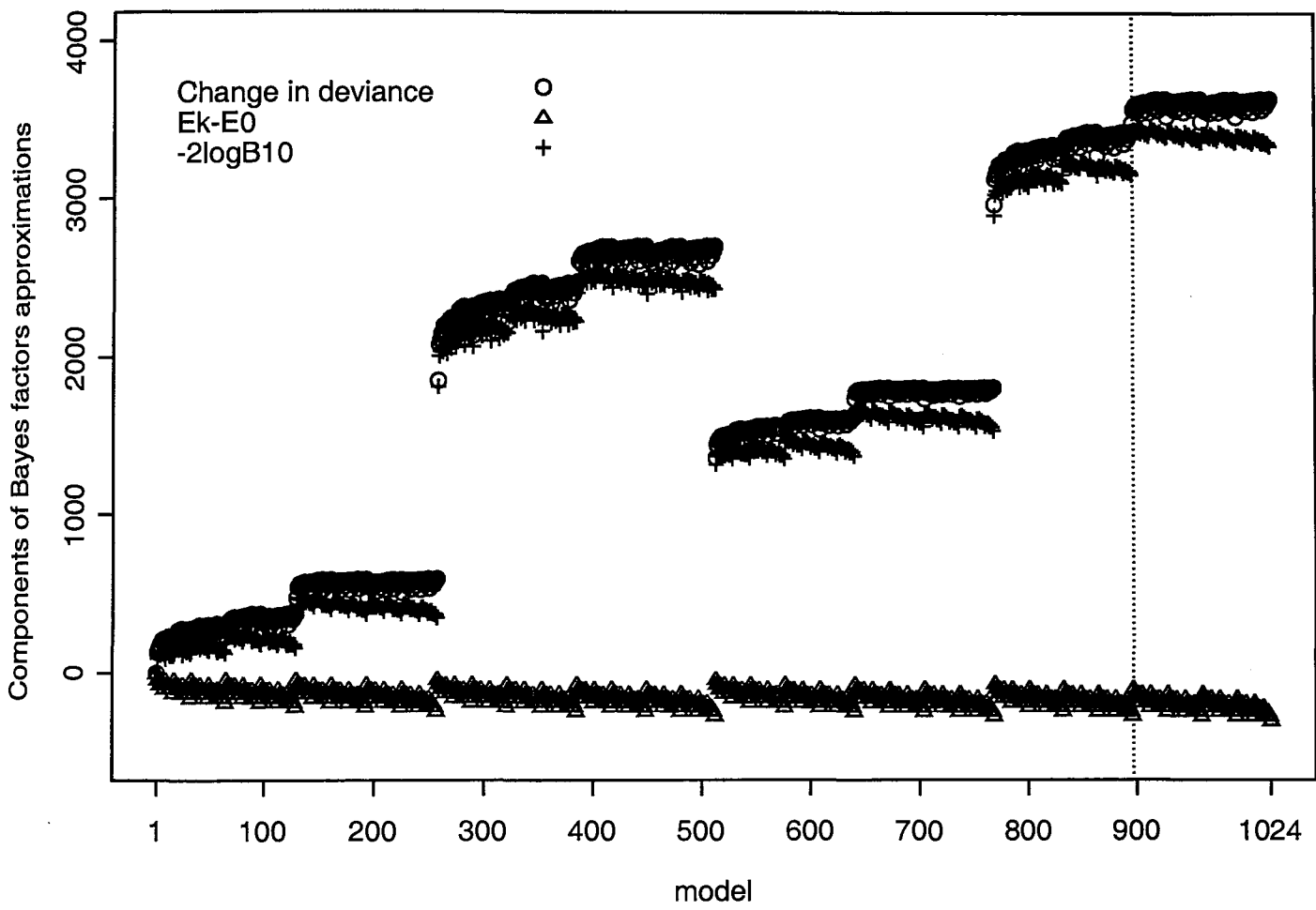


Figure 4.7: Components of Bayes factors approximations used in BMA analysis: changes in deviance, penalty term and approximations of Bayes factors. Models to the right of the dotted line are those containing the same day effect of temperature

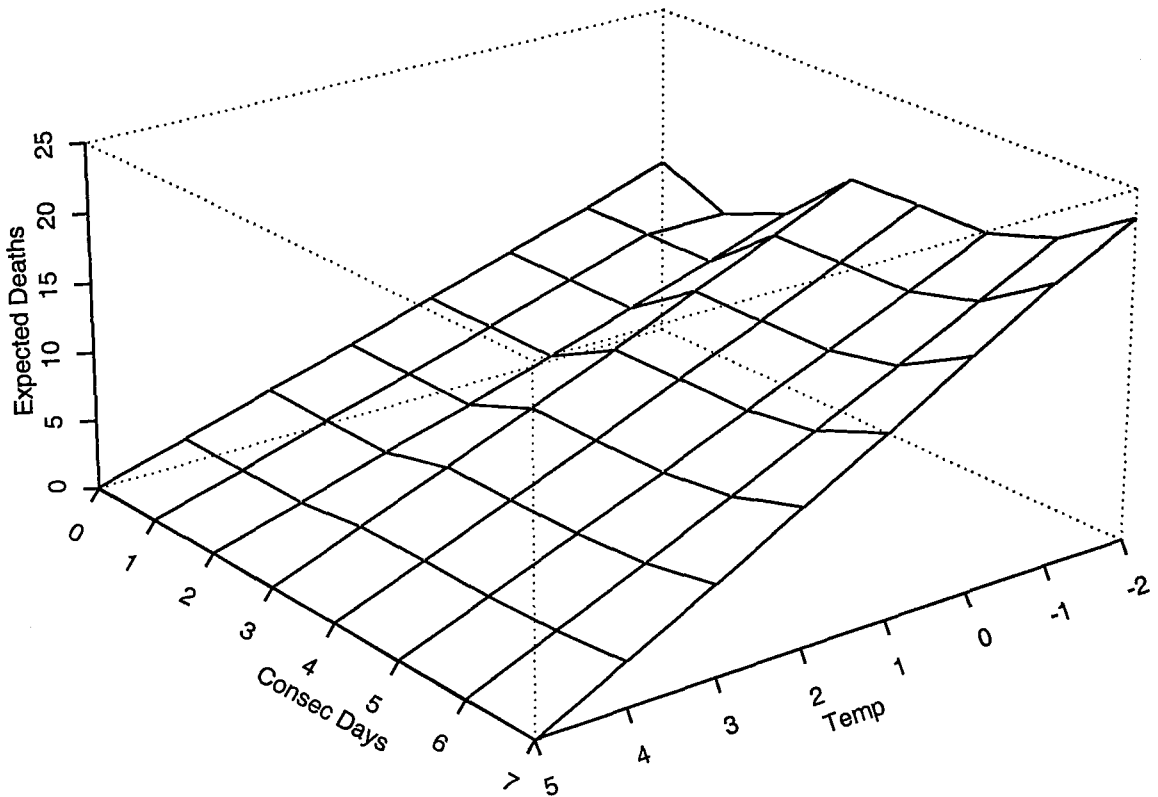


Figure 4.8: Results from using Bayesian model averaging for prediction using a range of temperature scenarios. The number of excess daily deaths expected from low temperatures for periods of up to eight days are given shown, in comparison with a single day at 5° . Temperatures range from 5°C to -2°C for periods of between one and eight days.

Chapter 5

Modelling Daily Multivariate Pollutant Data at Multiple Sites

5.1 Introduction

This chapter considers the spatio-temporal modelling of pollutants measured daily at a number of monitoring sites within an urban environment. The approach presented here was developed for use in studies investigating acute health effects of air pollution. Alternative objectives include the design problem of the positioning of a new monitoring site, or for regulatory purposes in order to determine whether environmental standards are being met.

Concern about the potential health effects of airborne pollution has increased over recent years, and numerous studies have shown associations between daily measures of pollution and adverse health effects, for both mortality and morbidity. While the long-term (or chronic) effects of air pollution are also of interest (and are discussed in further detail in Chapter 3), the great majority of studies, due to data availability and interpretational issues (such as the pattern of confounding), consider short-term, or acute, effects. While increases in adverse health events are associated with air pollution in most studies, it is unclear which are the actual pollutant(s) responsible, and the size of the effect. In conducting such time-series studies to investigate the relationship between air pollution and a health outcome, for example, respiratory mortality, it is important to have a good measure of the level of pollution on each of the study days. Often daily measurements are available from a number of monitoring sites across the study area. Each of these monitors may measure different sets of pollutants, there may be periods of missing data, due, for example, to a particular pollutant not being measured at a site, a monitor being inactive by design (for example, a six-day monitoring schedule), or because of an unreliable or faulty monitor, and each of the recorded measurements may be subject to error.

Ecological epidemiological studies examine exposure-risk associations at the level of the group

rather than the individual. Examination of ecological data can lead to a number of inferential difficulties, an umbrella term for these difficulties being ecological bias. Such bias may arise, for example, where there is within-group variability in exposures/confounders, or between-group confounding (see, for example, Greenland (1992)). In studies investigating the health effects of daily changes in air pollution the exposures (concentrations of outdoor air pollutants) are essentially treated as constant across the study area. Many studies are carried out in urban areas where there may be more than one monitoring site, in which case a daily average of readings from all sites is often used (for example, Schwartz and Dockery (1992a); Schwartz (1994a)). The use of such an average, or the readings from just one site, may be criticised on the grounds that pollution levels vary within an urban environment. There may be spatial differences due, for example, to local traffic conditions and point sources of pollution, and the ability of the pollutants to disperse because of the surrounding buildings.

In this chapter, a (hierarchical) dynamic linear model (DLM) is suggested for the analysis. Such DLMs are described in, for example, Pole et al. (1994) and Gamerman and Migon (1993) (who specifically consider hierarchical DLMs). At stage one of this hierarchy, each of the individual pollutants on any day are modelled as a function of the true underlying level, corrupted by measurement error. These true underlying levels are assumed to have structure in both space and time, and this structure is modelled at stage two of the hierarchy along with the relationship between pollutants. A Bayesian approach is adopted throughout with prior distributions being assigned to the unknown parameters and missing observations at stages one and two. Essentially, the multiple pollutants at any time point are modelled as arising from a multivariate Gaussian random field. This model addresses each of the inadequacies described in the observed data and, specifically, allows information from multiple sites on different pollutants to be combined, in order to provide an accurate level of pollution at each of the observed sites, or at locations previously unmeasured, though the latter is more dependent on modelling assumptions. At any location, monitored or otherwise, a measure of the uncertainty associated with the level can also be obtained. The latter is particularly useful for accounting for the variability in the pollution level, formally via errors-in-variables modelling (e.g., Carroll et al. (1995)), or informally when regression coefficients describing the relationship between health risk and pollution are interpreted.

The structure of this chapter is as follows. In Section 5.1.1 studies of the acute effects of exposure to air pollution on health are critically evaluated. Section 5.2 contains a description of the data used in the analysis, comprising four pollutants measured daily at eight monitoring sites in London over a four-year period. Details of the dependencies which might be expected in such data, and which are exploited within the modelling framework are discussed. The models are introduced in increasing levels of complexity, with the first, given in Section 5.3 considering the simple case of a single pollutant and a single monitoring site. It is natural to first consider the temporal model since the data across time are far more abundant than across space and across pollutants.

Section 5.4 stays with a single pollutant but extends the model to allow for multiple monitoring sites. Section 5.5 considers multiple pollutants at a single site and in Section 5.6, a final model that examines multiple pollutants at multiple sites is described, which combines features of the previous models in a general form. In each of Sections 5.3, 5.4, 5.5 and 5.6, the models are applied to the data from London.

5.1.1 Epidemiology of acute health effects

There is a considerable literature associated with the acute effects of air pollutants on health, for both mortality and morbidity, and a large number of epidemiological studies have reported associations at concentrations below existing guidelines and standards. A number of reviews of this literature are available, for example Pope et al. (1995), COMEAP (1995), Brunekreef et al. (1995) and Levy et al. (2000).

The earliest studies focused on severe air pollution episodes, with particular reference to a dramatic episode in London in the early 1950s. Later studies concentrated on the effect of much lower levels of pollution, using routinely measured pollution and health data, primarily in urban areas in the U.S. Recently, a small number of multi-city studies have been performed, in both the U.S. and Europe, using a common protocol for data management and analysis, in a set of cities, producing both city-specific, and overall combined estimates of risk.

The earliest epidemiological studies found that exposures to very high concentrations of particulate matter had adverse effects on human health. In an early study, Firket (1936) attributed 60 excess deaths to a severe smoke and fog episode in the Meuse Valley in Belgium in 1930. In London, 4000 excess deaths were associated with an episode of fog 1952 in which levels of black smoke exceeded $4,500 \mu\text{gm}^{-3}$ (Ministry of Health (1954)). London was the subject of a number of the earlier studies, due to its large population and the early availability of data on health outcomes and pollution levels (daily measurements of pollutants became available in 1965, although winter values were available from 1958). One of the first time-series analyses of daily mortality in relation to air pollution used black smoke and SO_2 data from London for the winter of 1958-59. Martin and Bradley (1960) estimated that both pollutants were associated with excess daily mortality when levels were higher than $750 \mu\text{gm}^{-3}$. Later reanalysis of the data by Ware et al. (1981) again found an association between pollutant levels and deaths, but no evidence for the threshold. A number of other studies have also used data from London from this period, including Mazumdar et al. (1982b), Shumway et al. (1983) Schwartz and Marcus (1990). All found associations between daily mortality and pollution, but with different conclusions about the relative effects of black smoke and SO_2 . For example Schwartz and Marcus (1990) found a small effect of SO_2 , Mazumdar et al. (1982b) no effect and Shumway et al. (1983) that the effects of black smoke and SO_2 were equally predictive. These differing conclusions may be due to the

different methods used in the analyses, and in particular in the way that temperature and seasonal patterns were modelled (Thurston and Kinney (1995)). Lippmann and Ito (1995) analysed the London data from 1965-72, paying particular attention to separating the effects of pollution from those of temperature, and found SO₂ to have the strongest effect. A recent study of daily mortality in London, 1987-92, found an increase of 2.5% (95%CI 0.9-4.1) associated with changes in black smoke (7-19 μgm^{-3}) on the previous day within the warm season, and 1.7% (0.8-2.6) for changes (8-23 $\mu\text{g}/\text{m}^{-3}$) over the entire year (for cardiovascular mortality the corresponding findings were 0.6% and 0.7% respectively, both non-significant, and no significant associations were found for corresponding differences in SO₂) (Anderson et al. (1996)). Analysis of data from London, 1992-94, produced significant associations between emergency hospital admissions for respiratory disease (all ages) and PM₁₀ and SO₂, with increases of 3% (1.1-4.9) and 2% (0.3-3.8) for differences between the 10th and 90th percentiles of the pollutant (15.8-46.5 μgm^{-3} for PM₁₀ and 13-31 μgm^{-3} for SO₂); significant increases were also seen in the 0-14 year age group (Atkinson et al. (1999)). Elsewhere in Great Britain, associations have been found between both smoke and SO₂ and asthma and respiratory admissions in Birmingham (Walters et al. (1994)), where, for example, 15.5 (95% CI, 5-25) additional respiratory admissions were associated with a 100 mgm^{-3} rise in SO₂. Associations were also found between hospital admissions for asthma and respiratory conditions and NO₂ over the wider area of the West Midlands (Walters et al. (1995)).

A large number of the more recent studies of the acute effects of air pollution have taken place in the U.S., using routinely collected health and pollution data, with broadly consistent effect estimates being found in a variety of study settings. Dockery and Pope (1994) reviewed eight studies in climatically diverse U.S. cities, finding a range of 0.7% to 1.6% for excess daily non-injury mortality associated with a 10 μgm^{-3} increase in PM₁₀, with a weighted mean of 1%. In four studies that included cause-specific mortality, increases of between 0.8% and 1.8% were found (weighted mean of 1.4%) for cardiovascular deaths and between 1.5% and 3.7% (weighted mean 3.5%) for respiratory deaths. Schwartz (1994a) extended the first analysis, but added two cities (London and Minneapolis) to the eight considered by Dockery and Pope, resulting in an estimate of a 0.6% (0.5-0.7) increase per 10 μm^{-3} .

A review of nine studies by Ostro (1993) gave a similar result: a 0.96% (weighted mean) increase in mortality per 10 μgm^{-3} increase in PM₁₀, with a one-standard-error range of 0.65% to 1.30%. A review of PM₁₀ mortality studies by Thurston (1996) found an overall mortality risk of 0.5 to 1% per 10 μgm^{-3} increase in PM₁₀, with higher relative risks indicated for the elderly and those with existing respiratory conditions. In the reviews by Dockery, Pope and Ostro, studies which used an exposure metric for exposure to particulate matter other than direct measurement of PM₁₀, e.g. total suspended particulates (TSP), black smoke (BS) or coefficient of haze (CoH), were converted into PM₁₀ units by applying data-derived correction factors. Thurston restricted the choice of studies to those which directly used PM₁₀ measurements or measurements with

on-site calibrations to PM_{10} . A meta-analysis performed by Schwartz (1994a) found an increased risk for non-accident mortality of 0.6% (0.5- 0.7) per $10 \mu g m^{-3}$ increase in TSP (an estimated 50-60% of TSP is $< 10 \mu g m^{-3}$ in diameter).

Recent studies in Latin American cities have also reported increased risks of mortality from particle pollution: a 1% increase in daily mortality for each $10 \mu g m^{-3}$ increase in PM_{10} in Santiago, Chile, (Ostro et al. (1996)) and a 0.6% increase in daily mortality for each $10 \mu g m^{-3}$ increase in TSP in Mexico City, Mexico (Barja-Aburto et al. (1997)). Gouveia and Fletcher (2000) examined the association between lagged values of five pollutants (PM_{10} , SO_2 , NO_2 , CO and O_3) and respiratory diseases in Sao Paulo, Brazil. Significant increases of daily admissions of children to hospital for total respiratory disease and pneumonia were associated with PM_{10} (9%), ozone (5-8%) and NO_2 (9%) (results are for an increase from the 10th to the 90th percentile of pollution measurements).

The relationships are often difficult to quantify because of the high correlations that may exist between the daily levels of particulate matter and other pollutants such as sulphur dioxide (SO_2), aerosol activity, ozone (O_3) or nitrogen dioxide (NO_2). Nevertheless, associations have also been shown in cases where there was no association reported with other pollutants, notably for SO_2 (Pope et al. (1992)). Information concerning the effect of a particular pollutant, such as PM_{10} , can be gained by assessing the effect of the pollutant across different locations that have differing levels of other pollutants, such as O_3 , or different correlations with other pollutants. For example, Schwartz and Dockery (1992a) found associations between particulate matter and mortality in Philadelphia when SO_2 , NO_2 , CO and O_3 were present and similar associations were found by Pope et al. (1992) in the Utah Valley, when CO was present and in Santa Clara where there were low levels of other pollutants (Fairley (1990)). The consistency of such findings across different pollution scenarios is the basis for much of the evidence of an adverse effect of ambient air pollution on health. Data from Philadelphia have been the subject of a number of analyses and much discussion considering the question of potential confounding of the particle-mortality association by other pollutants and by seasonal effects (Samet (1995), Moolgavkar et al. (1995), Dockery and Schwartz (1995)). Recently, a detailed comparison of different methods for controlling for seasonal and temperature effects using data from Philadelphia concluded that the associations were not spurious, but that TSP and SO_2 were strongly correlated, making it difficult to identify either as an independent predictor of excess mortality (Kelsall et al. (1997)).

If daily mortality is indeed associated with daily particle pollution, then it might be expected that associations with morbidity should also be observed. Several studies have found that increases in hospital admissions for a number of conditions are associated with increased levels of various pollutants (see reviews by Dockery and Pope (1994) and Pope et al. (1995)). Respiratory conditions were found to be associated with increased levels of sulphates and ozone, for example in southern Ontario, Canada, where 1% (0.4-1.5) and 0.8% (0.4-1.1) higher admissions were found for asthma

and all respiratory admission respectively, with 1.5% (0.8-2.2) higher COPD admissions (Burnett et al. (1994)). Consistent, and significant increases in COPD admissions with relation to PM_{10} have also been found in a number of other studies, for example an increased risk of 2.5% (0.3-4.7) was found in Birmingham, Alabama (Schwartz (1993)), 2.0% (0.9-3.2) in Detroit (Schwartz (1994b) and 4.6% (1.8-7.5) in Minneapolis/St. Paul (Schwartz (1994d)). Several studies have also found associations between short-term changes in air pollution on the number of emergency room visits, for example, Schwartz et al. (1993) found relationships with particulate matter, and Atkinson et al. (1999) with both PM_{10} and SO_2 .

There are inherent difficulties in interpreting, comparing and combining the results from these types of studies due to the differences in the pollutants measured, the statistical methodologies, the potential covariates that are adjusted for, and the lagged values that are used. A critical assessment and comparison of the different statistical methodologies, and in particular methods used for 'model-building', used in time-series studies such as these is given in Chapter 2. Even if the estimates of risk obtained from different studies all relate to levels of particulate pollution, the original measurements may have been different; the coefficient of haze (CoH), total suspended particulates (TSP) or black smoke (BS), as well as direct measurements of PM_{10} , and more recently, $PM_{2.5}$, each of which measures slightly different aspects of particulate matter. Although attempts have been made to standardise the measures of pollution by converting the measurements into 'equivalent' amounts of PM_{10} , for example $PM_{10} \approx 0.55$ TSP, $PM_{10} \approx CoH/0.55$, $PM_{10} \approx BS$ and $PM_{10} \approx PM_{2.5}/0.6$ (Dockery and Pope (1994)), great care is needed when comparing or combining the resulting estimates of risk, because the complex nature of the composition of particulate matter means that the different measures may measure different aspects of air pollution.

In Europe, the APHEA (Air Pollution and Health: a European Approach) project included data on mortality and hospital emergency admissions from 15 European cities in 10 countries, for which data were analysed by time-series methods according to a standard protocol, with the city-specific estimates being combined using meta-analysis. Results from the city specific analyses can be found in Katsouyanni (1996), whilst summaries and meta-analyses can be found in Touloumi et al. (1997), Katsouyanni et al. (1997), Anderson et al. (1997), Sunyer et al. (1997), Zmirou et al. (1998), Spix et al. (1998).

Broadly consistent associations were found in the six western European cities, which had wide differences in climate and environmental conditions. Using a random effects model to combine data from the individual cities, an overall excess of 2% (95% CI 1-3) was found in mortality per $10 \mu g m^{-3} PM_{10}$, with significant excesses also observed with changes in SO_2 (excess of 2.9% (2.3-3.5), per $50 \mu g m^{-3}$) and black smoke (3% (2-4), per $50 \mu g m^{-3}$) (Katsouyanni et al. (1997)). In the central European cities, the rises were smaller, and in the case of SO_2 and black smoke, non-significant; 4.3% (0.3-8.5) per $10 \mu g m^{-3} PM_{10}$, 0.8% (-0.7-2.4) per for a $50 \mu g m^{-3} SO_2$ and

0.4% (-0.3-1.1) per $50 \mu\text{g m}^{-3}$ black smoke. (Zmirou et al. (1998)) found significant associations between cause-specific (daily) deaths and black smoke and SO_2 , with pooled estimates from western European cities of a 2% (1-4) excess in cardiovascular mortality per $50 \mu\text{gm}^{-3}$ and 4% (1-6) per $50 \mu\text{gm}^{-3}$ SO_2 . For respiratory mortality, the corresponding excesses were 4% (2-7) for black smoke and 5% (3-7) for SO_2 . Associations were also found between changes in ozone and daily deaths in four of the cities (London, Athens, Barcelona and Paris) (Touloumi et al. (1997)). The combined result (using a random effects model) was an excess of 2.9% (1-4.9) for a $50 \mu\text{gm}^{-3}$ rise in ozone (or a 1.1% increase per 10 ppb). Significant associations were also found in the western European cities, specifically for cardiovascular (2%(1-3)) and respiratory (6%(2-10)) deaths (Zmirou et al. (1998)).

The NMMAPS study followed many of the same principles of APHEA, analysing and then combining data on daily mortality and the effect of PM_{10} and SO_2 from the 90 largest cities in the U.S. according to a standardised protocol for data management and analysis (Samet et al. (2000)). The first stage considered the largest 20 cities and is reported in Domininci et al. (2000), where a Bayesian hierarchical model was used to combine estimates from the individual cities. Overall, an increase of $10 \mu\text{gm}^{-3}$ in PM_{10} (one day lag) from the 20 cities analysis resulted in an excess of 0.48% (0.05-0.92) for all cause mortality (which increased with adjustment for ozone). This is slightly smaller than that reported from the APHEA study (Katsouyanni et al. (1997)), and the value of 0.7% found in a recent meta-analysis of 29 studies in 23 locations in Europe and North and South America (Levy et al. (2000)). It is somewhat lower than that found by Dockery and Pope (1994) and Schwartz (1994a), and although differences might be expected due to the different set of cities used and difference in the methods used, the results are not strictly comparable in that, although a single day's exposure was used in all the individual analysis used to produce the NMMAPS result, the meta-analyses included studies which used various lagged values and/or averages over different intervals. Also, whilst the overall estimate calculated in the NMMAPS study used data from all of the cities in the study, the meta-analyses of Dockery and Pope, and of Schwartz, used data only from published studies, meaning the result may have been distorted by publication bias.

In the second stage of the NMMAPS study the analysis was extended to incorporate data from the next 70 largest cities in the US. Whilst the first stage of the analysis (20 cities) focused primarily on estimating the overall effects of PM_{10} , the second (90 cities) specifically explored spatial heterogeneity and regional differences. A positive association between one day lagged PM_{10} and total mortality was found in most locations, with the size of the effect ranging from over 1% per $10 \mu\text{gm}^{-3}$ in New York City, San Diego and Oakland, to no (or an inverse) association in Dallas, Cleveland and Atlanta, where the daily average ranged from 15 to $53 \mu\text{gm}^{-3}$. Over all 90 cities, the NMMAPS study estimated 0.5% higher total non-accidental mortality associated with a $10 \mu\text{gm}^{-3}$ increase in PM_{10} . The effect of ozone (in the summer months) was estimated to

be 0.25% higher non-accidental mortality per 10 ppb, which is similar to that found by Thurston and Ito (1999) (0.36% per 10 ppb) from a meta-analysis of 15 studies, but smaller than that found in the APHEA study (Touloumi et al. (1997)). However, an inverse association was found in the winter months, possibly due to an unmeasured confounder.

Taken together, the studies are indicative of positive associations between a range of pollutants (PM_{10} , SO_2 , ozone) and various health outcomes, especially respiratory and cardiovascular, but the size of effect and its significance appear to have been overestimated, especially in earlier studies. With the notable exception of the more recent multi-city studies, there is no consistent approach to model selection, both in terms of which lagged values of pollutants are used (and reported) and in other confounders that are allowed for.

In many of these time-series studies examining the associations between short-term changes in air pollution and health, the exposure data consist of daily measurements of pollutants. These are often available from a number of monitoring sites across the study area, each of which monitors may measure different sets of pollutants, and in addition, they may be subject to error and may contain periods of missing data. In this chapter, a spatio-temporal model is proposed which addresses each of these points, and produces an underlying measure of the pollutants for each day together with a measure of uncertainty.

The modelling of environmental variables in time and space has a considerable literature, of which details can be found in the statistical review in Chapter 2. The model proposed here has relatively straightforward elements for each of the temporal, spatial, and multivariate pollutant components; the complexity arises from carrying out a fully Bayesian analysis of the complete structure. There are three components to the model; the first two of which, spatial and temporal, are comparable in some sense in that each of them has a natural order and may, for example, be measured on different scales. The third component, the multiple pollutants is not of this type as it is an ‘unordered’ variable. Consequently, the models specified to link the pollutants have a different form. In essence, there is an ‘interaction’ between the temporal and pollutant components, with the spatial component (which can be pollutant specific) being a ‘main effect’, independent of the temporal component. The temporal model used is very closely related to Kalman filtering in which the model structure allows a straightforward updating of the parameters to provide an approximate Bayesian solution (e.g. Meinhold and Singpurwalla (1983)). However, once the spatial model is incorporated this simple updating is no longer possible. Recently, independent research by Tonellato (2001) presented a similar modelling approach to that presented here, but only for a single pollutant, using a univariate auto-regressive process measured with error, for hourly measurements of CO from a small number of sites.

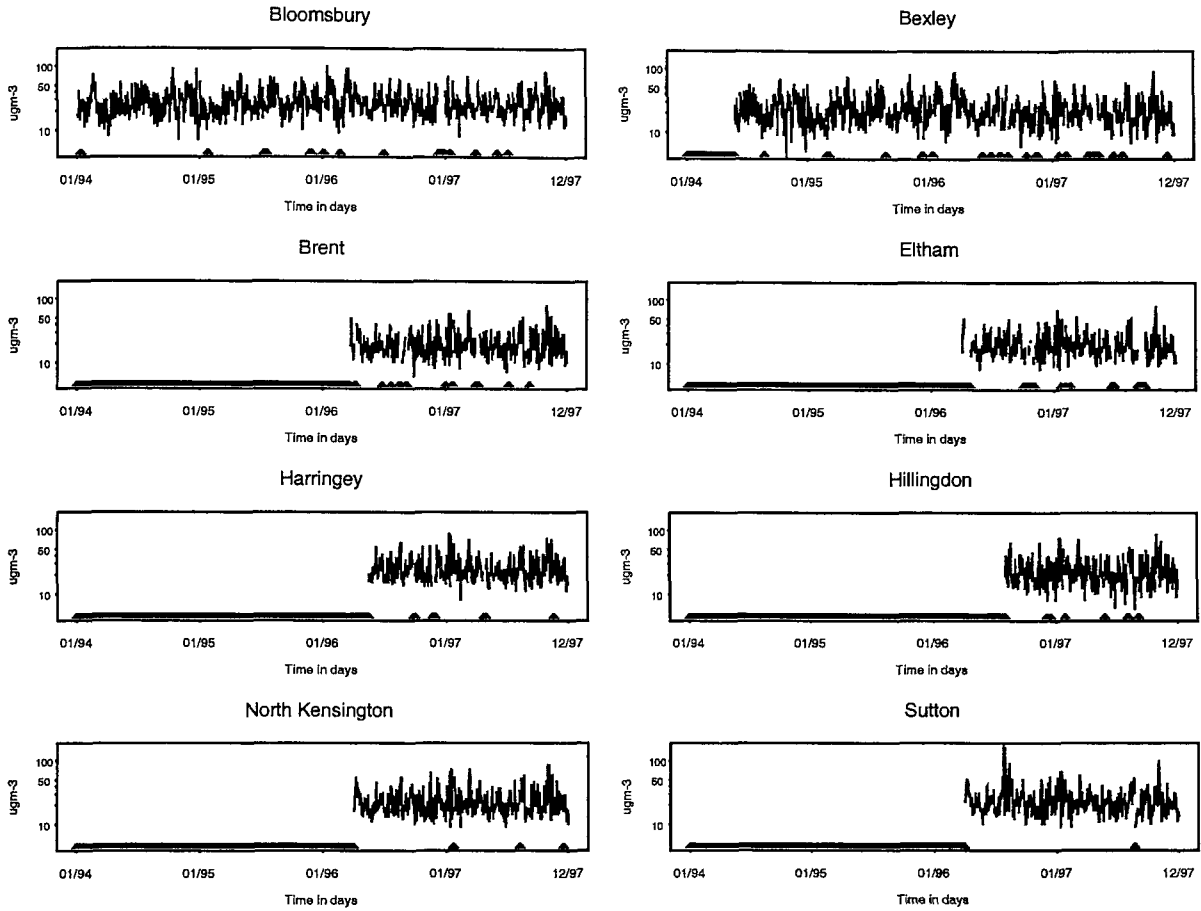


Figure 5.1: Time series plots of logged values of PM_{10} for eight sites in London, 1994–97 (missing values are marked on the horizontal axis as triangles).

5.2 Data description

The methods presented are illustrated using pollution data collected at eight monitoring sites within London measuring the pollutants particulate matter (PM_{10}), carbon monoxide (CO), nitrogen oxide (NO) and sulphur dioxide (SO_2) over the period 1994–1997. Table 5.1 summarises the periods of operation, and pollutants measured at each of the sites. All four pollutants were measured at five sites only; the periods of operation vary between one and four years, and the percentage of missing values within these periods of operation can be large. For example, 37% of the daily measurements for CO were missing at Hillingdon. Readings of zero for any of the pollutants were considered to be missing values, rather than implausibly low readings. Figure 5.1 shows time series plots for PM_{10} at all eight sites; which was unavailable at six sites for the first two years. There is no apparent trend in level over the time scale of the study.

Pollutant Dependence

The approach developed here exploits the dependencies that exist between pollutants, temporally and spatially, due to the processes by which they are formed. Increasingly, attention has been

Table 5.1: Summary of pollutants measured, and periods of operation, at eight sites in London, 1994-97. The total number of days of operation are given for each pollutant at each site together with the percentage of missing observations. The units are μgm^{-3} for PM_{10} , parts per billion for SO_2 and NO and parts per million for CO .

	Period	Total	Missing	%	Mean	Min.	25%	Med.	75%	Max.
Bexley										
PM_{10}	1994-97	1461	211	14.4	24.0	4.0	15.0	20.0	29.0	92.0
SO_2	1994, 1996-97	1095	178	16.3	6.9	1.0	3.0	4.0	8.0	76.0
NO	-	-	-	-	-	-	-	-	-	-
CO	1994-97	1461	192	13.1	0.5	0.1	0.3	0.4	0.5	4.4
Bloomsbury										
PM_{10}	1994-97	1461	61	4.2	28.0	7.0	19.0	24.0	34.0	103.0
SO_2	1994-97	1461	115	7.9	8.3	1.0	4.0	6.0	11.0	48.0
NO	1994-97	1461	44	3.0	42.4	4.0	19.0	30.0	50.0	467.0
CO	1994-97	1461	68	4.7	0.7	0.1	0.4	0.6	0.8	4.3
Brent										
PM_{10}	1996-97	731	120	16.4	20.8	6.0	14.0	18.0	25.0	82.0
SO_2	1996-97	731	33	4.5	4.4	1.0	2.0	3.0	5.2	20.0
NO	1996-97	731	57	7.8	23.8	1.0	5.0	8.0	22.5	414.0
CO	1996-97	366	15	4.1	0.5	0.1	0.2	0.3	0.7	5.0
Eltham										
PM_{10}	1996-97	731	166	22.7	21.2	8.0	15.0	18.0	25.0	81.0
SO_2	1996-97	731	91	12.4	4.6	1.0	2.0	3.0	5.0	40.0
NO	1996-97	731	95	13.0	21.7	1.0	5.0	9.0	20.0	339.0
CO	-	-	-	-	-	-	-	-	-	-
Harringey										
PM_{10}	1996-97	731	161	22.0	26.2	8.0	18.0	22.0	32.0	89.0
SO_2	-	-	-	-	-	-	-	-	-	-
NO	1996-97	731	139	19.0	63.3	5.0	28.0	43.0	68.6	562.0
CO	-	-	-	-	-	-	-	-	-	-
Hillingdon										
PM_{10}	1996-97	731	225	30.8	24.5	6.0	16.0	21.0	31.0	88.0
SO_2	1996-97	731	230	31.5	5.1	1.0	3.0	4.0	6.0	28.0
NO	1996-97	731	252	34.5	81.9	2.0	31.0	67.0	105.0	506.0
CO	1996-97	731	268	36.7	0.8	0.2	0.5	0.6	0.9	4.3
N. Kensington										
PM_{10}	1996-97	731	99	13.5	23.6	9.0	16.0	20.0	27.2	89.0
SO_2	1996-97	731	91	12.4	4.6	1.0	2.0	3.0	6.0	32.0
NO	1996-97	731	106	14.5	27.6	1.0	6.0	11.0	25.0	442.0
CO	1996-97	731	93	12.7	1.2	0.1	0.4	0.7	1.3	16.6
Sutton										
PM_{10}	1996-97	731	92	12.6	25.1	9.0	17.0	22.0	29.0	250.0
SO_2	1996-97	731	96	13.1	4.9	1.0	2.7	4.0	6.0	28.4
NO	1996-97	731	106	14.5	51.1	3.0	26.3	39.0	57.0	404.0
CO	1996-97	731	104	14.2	1.1	0.2	0.8	1.0	1.3	6.7

paid to the role of the small particulate matter known as PM_{10} (diameter $< 10\mu\text{gm}^{-3}$), and more recently even smaller particles ($\text{PM}_{2.5}$), with relation to adverse health outcomes. Such particles are often classed as either primary (emitted directly into the atmosphere) or secondary (formed or modified in the atmosphere) (Onursal and Gautan (1997), Richter and Williams (1998)). A major source of these fine particles, and of the other pollutants considered here, are combustion processes, in particular diesel combustion. Secondary particles are typically found when low volatility products are generated in the atmosphere, for example in the oxidation of sulphur dioxide (SO_2) to sulphuric acid. Sulphur dioxide, a corrosive acid gas, is primarily caused by power stations burning fossil fuels which contain sulphur. Since the decline in domestic coal burning in cities, emissions have decreased, but there is still concern about the combination of SO_2 and particulate matter. Carbon monoxide is a toxic gas emitted as a result of combustion processes, in urban areas it is almost entirely produced from road traffic emissions. The principle source of oxides of nitrogen, NO_x , which are formed during high temperature combustion processes, is also road traffic which is responsible for about 50% of all emissions. When NO is released into the atmosphere it reacts with ozone to create NO_2 . Associations between the levels of pollutants are also observed because of their relationship with meteorological conditions, such as wind direction and speed and the increased reactivity of primary pollutants in forming secondary pollutants at higher temperatures. Increased levels of pollution are also observed during temperature ‘inversions’, where the usual decrease in temperature with altitude, which causes hot pollutant gases to rise, is absent causing primary pollutants to be trapped near the ground, thus enhancing the production of secondary pollutants (Onursal and Gautan (1997)). Dependencies with temperature and certain pollutants will also be due, in part, to changing use of heating materials (ozone, which increases with temperature, is not considered here). Table 5.2 shows the sample correlations between the pollutants (logged values) and temperature, at the Bloomsbury site. Temperature is positively correlated with PM_{10} and inversely with SO_2 , NO and CO, with the strongest correlation being with NO and the weakest with PM_{10} . The apparent weak correlation with PM_{10} is likely to be due to its non-linear relationship with temperature. This can be seen in Figure 5.2, which shows the associations between the four pollutants and temperature at the Bloomsbury site. This is due to the different composition of PM_{10} over the year. High levels of particulate matter are observed in the summer months due to high levels of ozone, whereas in the winter high levels will be seen due to the increase in use of heating materials. It is noted that the associations between the pollutants appear relatively constant across sites (not shown), as do the associations with temperature.

Temporal Dependence

The pollutants considered here have varying atmospheric lifetimes, but all can linger in the environment. For example, nitrogen oxides have a lifetime of approximately one day and are then converted to nitric acid, which is removed from the atmosphere by deposition to the ground (for

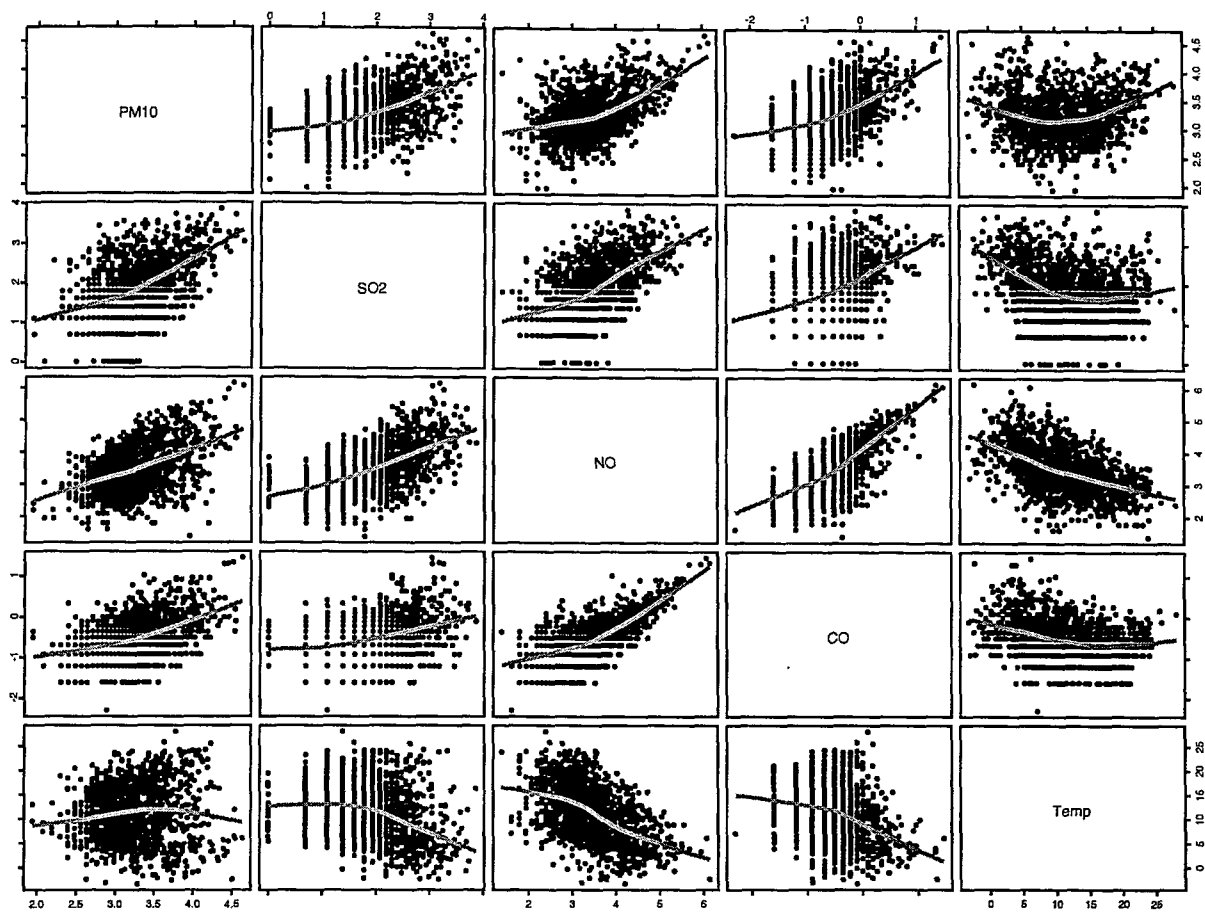


Figure 5.2: Scatter plots showing relationships between the four pollutants (PM₁₀, SO₂, NO and CO) with temperature, at the Bloomsbury site, 1994–97. The lines are loess smoothers.

Table 5.2: Correlation/covariance matrix for (logged) pollutants and temperature, Bloomsbury, 1994–97. The variances lie on the diagonal, with covariances above and correlations below.

	PM ₁₀	SO ₂	NO	CO	Temp
PM ₁₀	0.41 ²	0.15	0.14	0.091	0.23
SO ₂	0.49	0.72 ²	0.30	0.14	-1.21
NO	0.45	0.57	0.73 ²	0.26	-2.09
CO	0.45	0.38	0.71	0.49 ²	-0.82
Temp	0.10	-0.29	-0.50	-0.29	5.71 ²

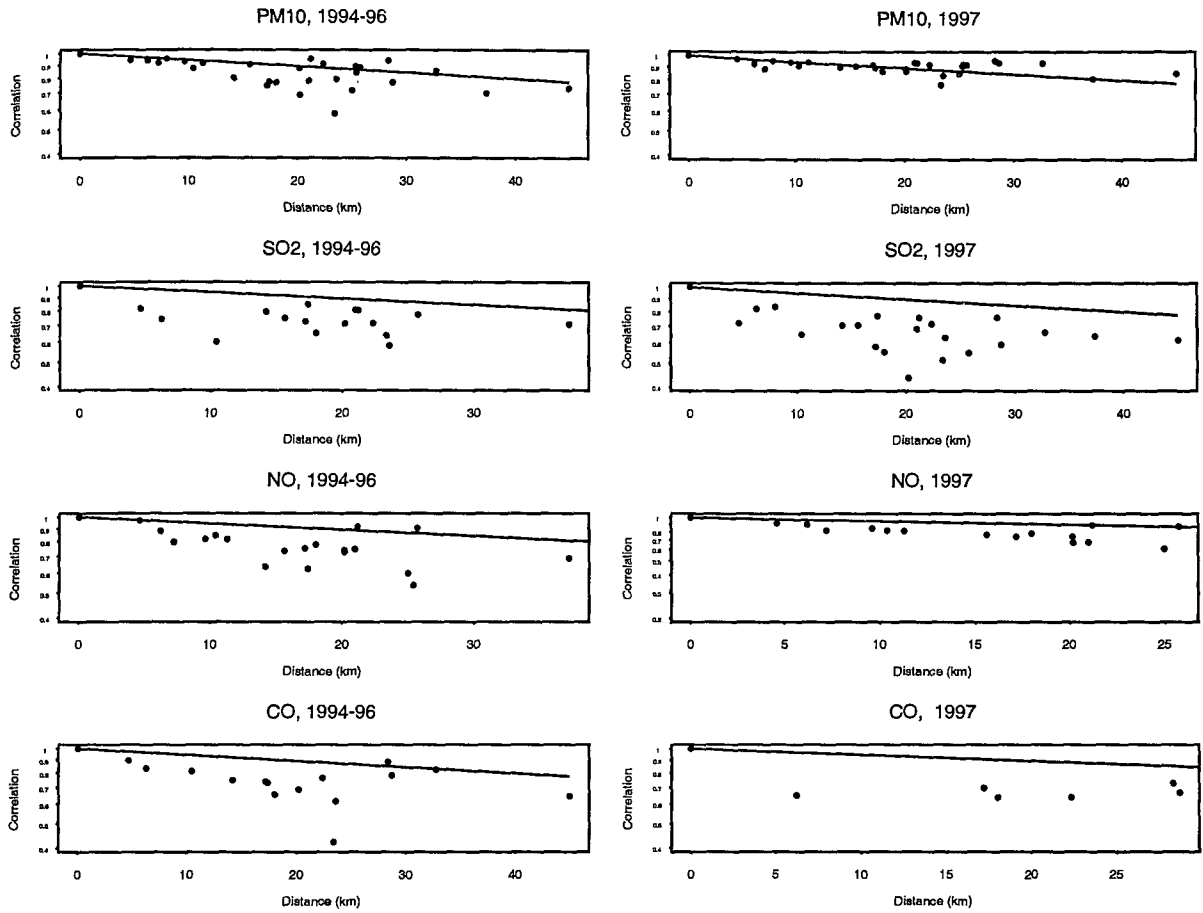


Figure 5.3: The (log) correlation between logged values of pollutants and distance between sites. The reference line is $\log(\text{correlation}) = -\phi_p \text{ distance}$, using the posterior median for ϕ_p from the spatial model for PM_{10} .

example, by rainwater), whereas CO can survive in the atmosphere for up to a month, before eventually oxidising to carbon dioxide. The atmospheric lifetime of particulate matter is strongly related to particle size, but may be as long as 10 days for particles of about 1mm in diameter (Onursal and Gautan (1997)). As previously mentioned, the dispersion of pollutants is also related to the meteorological conditions. From these considerations, strong daily dependence would be expected between pollutant measures on consecutive days.

Spatial Dependence

There is very high correlation between the readings for each of the pollutants from different monitoring sites. Figure 5.5(a) shows the position of the eight sites included in the study, and the correlations between the logged measurements of PM_{10} measured at the eight sites are presented in Table 5.3. High correlations between all pairs of measurements are observed at all sites. In Figure 5.3 the correlation between daily measurements of each (log) pollutant and the distance between the site at which they were measured, is plotted, for two different time periods. As expected, measurements from sites in close proximity to each other are, in general, more highly

Table 5.3: Correlation/covariance matrix for (logged) values of PM_{10} , measured at eight sites, 1994–97. The variances lie on the diagonal, with covariances above and correlations below.

	Bexley	Bloomsbury	Brent	Eltham	Harringey R.	Hillingdon	N. Kensington	Sutton
Bexley	0.46 ²	0.16	0.19	0.18	0.16	0.17	0.18	0.17
Bloomsbury	0.92	0.39 ²	0.15	0.15	0.14	0.15	0.15	0.14
Brent	0.92	0.90	0.43 ²	0.16	0.16	0.17	0.17	0.15
Eltham	0.95	0.91	0.91	0.42 ²	0.14	0.15	0.16	0.16
Harringey	0.90	0.89	0.94	0.87	0.39 ²	0.16	0.15	0.13
Hillingdon	0.83	0.83	0.88	0.79	0.91	0.45 ²	0.17	0.14
N.Kensington	0.95	0.93	0.96	0.94	0.94	0.87	0.42 ²	0.15
Sutton	0.89	0.84	0.84	0.90	0.82	0.72	0.88	0.42 ²

correlated than those further away, and this relationship appears relatively constant across time periods and pollutants.

Missing values

If there are a number of sites within the study area, the problem of daily missing values from one (or more) monitoring sites may be naively addressed by simply imputing the average readings from the other sites. Alternatively, standard multivariate techniques for imputing missing values may be used; in the APHEA study (Katsouyanni et al. (1995)), the mean level of all stations with complete data were regressed against the levels from a subset of the monitors, excluding a missing site. The authors note, however, that this approach is not practical if more than a small number of sites (for example, three) have missing values. Sheppard et al. (1999) estimated the missing values conditional on all non-missing measurements from the same day, assuming that the data arose from a multivariate distribution, with a constant covariance. Other studies have simply ignored days where there are missing values, which can lead to great data loss, particularly when lagged values are required for any subsequent health analysis.

Measurement error

Epidemiological studies investigating the relationship between air pollution and health rarely have personal exposures, but instead use measurements of ambient concentrations from one or a small number of monitors, and aggregate rates of mortality (or morbidity). The misclassification that may occur through using ecological rather than personal exposures is often referred to as measurement error. Here, the term ‘measurement error’ refers to the error that may arise when using recorded measurements of the pollutants; the possible effects of misclassification are discussed in detail elsewhere (e.g. Zeger et al. (2000)).

Data from continuous monitoring can generally be assumed to have a relatively small measurement error if, as is often the case, the sites are subject to strict quality assurance and validation procedures (as are the London data considered here). The measurements of PM_{10} from different sample based monitoring techniques can be affected by several factors, such as operating

temperature and filter media and history. Determination of the accuracy and precision of any given concentration is, therefore, liable to encompass a wide margin of error. The target figure in the data quality standards is a precision of $< 5\mu\text{gm}^{-3}$ for concentrations of less than $100\mu\text{gm}^{-3}$ (DETR (1998)). For SO_2 , the accuracy depends on the accuracy of the calibration standards and analyser stability, but based on long-term comparisons is estimated at about $\pm 1\%$ (DETR (1998)). Where non-continuous monitoring is used, then errors may be much greater. Passive samplers for NO_2 , for example, are generally required to have an accuracy of approximately $\pm 10 - 20\%$. Few studies have considered the effect of error in the actual recorded measurements of pollution. One pragmatic approach that has been used to assess the sensitivity of inference (e.g. Schwartz et al. (1996)) is to add normally distributed error components of varying sizes to the readings.

5.3 Single pollutant, single monitoring site

5.3.1 Model

Let Y_t denote the log of the observed pollution level on day t , θ_t the true pollution level, and $X'_t, t = 1, \dots, T$, a $q \times 1$ vector of daily explanatory variables, for example, meteorological variables such as temperature and humidity, which may depend on time. A hierarchical dynamic linear model is then described in three stages:

Stage One, Observed Data Model:

$$Y_t = X'_t\beta + \theta_t + v_t, \quad (5.1)$$

$t = 1, \dots, T$. This model contains a $q \times 1$ vector of regression parameters, β , and two error terms, one with temporal structure, θ_t , and one without structure, v_t , referred to as *measurement error*. It is also assumed that these errors are independent and identically distributed (i.i.d.) as $N(0, \sigma_v^2)$. It is also assumed that the variance does not depend on time, however the model is easily extendable to situations in which the measurement error may change as a function of t , for example, when a monitor is replaced. In terms of DLMS, this stage of the model is known as the *observation equation*. At this first stage, the daily measurements are therefore modelled as a function of the true underlying level of the pollutant with error.

Stage Two, Temporal Model:

$$\theta_t = \theta_{t-1} + w_t \quad (5.2)$$

with w_t i.i.d. as $N(0, \sigma_w^2)$. Hence this stage represents a first order smoothing model (e.g. West et al. (1985), Fahrmeir and Knorr-Held (2000)) with the true level on day t being modelled as a

function of the level on the previous day. The model is a limiting form of the autoregressive first order model and provides a non-stationary temporal model. Such an approach has been widely used (see for example, (Pole et al. (1994))). In terms of DLMs, (5.2) is the *system equation*, and θ_t , the *state*.

Stage Three, Hyperprior:

Prior distributions for β , σ_v^2 and σ_w^2 are specified at this stage. A normal prior $N(c, C)$ is assumed for β , where c is a $q \times 1$ vector and C a $q \times q$ variance-covariance matrix. Gamma priors are specified for the precisions, specifically $\sigma_v^{-2} \sim Ga(a_v, b_v)$ and $\sigma_w^{-2} \sim Ga(a_w, b_w)$. These are parameterised so that, for example, $E[\sigma_v^{-2}] = a_v/b_v$ and $\text{var}(\sigma_v^{-2}) = a_v/b_v^2$.

From a Bayesian perspective, the second (temporal) stage may be viewed as a prior distribution for $\theta' = (\theta_1, \dots, \theta_T)$, with

$$\begin{aligned} p(\theta|\sigma_w^2) &\propto \prod_{t=2}^T p(\theta_t|\theta_{t-1}, \sigma_w^2) \\ &\propto \exp \left\{ -\frac{1}{2\sigma_w^2} \sum_{t=2}^T (\theta_t - \theta_{t-1})^2 \right\} \\ &\propto \exp \left\{ -\frac{1}{2\sigma_w^2} \sum_{t=1}^T n_t \theta_t (\theta_t - \bar{\theta}_t) \right\} \end{aligned}$$

where n_t indicates the number of, and $\bar{\theta}$ the mean of, the neighbours of θ_t , i.e. θ_{t-1} and θ_{t+1} . The (autoregressive) prior distribution for θ used in (5.3), $p(\theta|\sigma_w^2)$, can therefore be expressed as

$$p(\theta_t|\theta_{-t}, \sigma_w^2) \sim \begin{cases} N(\theta_{t+1}, \sigma_w^2) & \text{for } t = 1, \\ N\left(\frac{\theta_{t-1} + \theta_{t+1}}{2}, \frac{\sigma_w^2}{2}\right) & \text{for } t = 2, \dots, T-1, \\ N(\theta_{t-1}, \sigma_w^2) & \text{for } t = T. \end{cases} \quad (5.3)$$

where θ_{-t} represents the vector of θ 's with θ_t removed. It is noted that σ_w^2 is a *conditional* variance and so it is not comparable to σ_v^2 . The joint distribution is improper and only expresses prior beliefs about differences in levels on neighbouring days, and no moments exist for the state process. Due to this impropriety, an intercept is not specified in (5.9). An intercept may be incorporated if an additional constraint is imposed, for example, $\sum \theta_t = 0$. Letting $y = (y_1, \dots, y_T)'$, the distribution $p(\theta|y)$ exists however, and may be evaluated using the moments of the collection of $\theta_1, \dots, \theta_T$ empirically.

5.3.2 Inference

The posterior distribution is given by

$$p(\theta, \beta_1, \sigma_v^2, \sigma_w^2|y) = p(y)^{-1} \left\{ \prod_{t=1}^T p(y_t|\theta_t, \beta_1, \sigma_v^2) \right\} \left\{ \prod_{t=2}^T p(\theta_t|\theta_{t-1}, \sigma_w^2) \right\} p(\theta_1)p(\beta_1)p(\sigma_v^2)p(\sigma_w^2) \quad (5.4)$$

which is analytically intractable but samples from this distribution may be generated in a straightforward fashion using Markov chain Monte Carlo (e.g., Smith and Roberts (1993)). This was performed using the WinBUGS software (Spiegelhalter et al. (1998)), noting that dealing with the cyclical graph that arises at stage two requires some of the conditional distributions to be explicitly specified (Spiegelhalter et al. (1996)). If the variances σ_v^2 and σ_w^2 were known then the Kalman filter could be applied (Meinhold and Singpurwalla (1983); Fahrmeir and Tutz (1994)) for efficient estimation.

In this context, the values of the pollutants on unmonitored (i.e. missing) days is of great interest; it should also be noted that analytical computation is hampered by missing values. Such values may be treated as parameters and the posterior is then obtained over these values and the model parameters. This approach to dealing with missing values may be easily implemented within the WinBUGS software. Inference on the parameters of interest is then performed via averaging over the distribution of the missing values. If, for storage reasons for example, it is impractical to save the samples on-line, samples may be generated from the posterior over the missing values retrospectively. If y_o denotes the observed values, y_m the missing values, and $\lambda = (\theta, \beta_1, \sigma_v^2, \sigma_w^2)'$, then samples from the distribution of missing values may be generated via

$$\begin{aligned} p(y_m|y_o) &= \int p(y_m|\lambda, y_o)p(\lambda|y_o)d\lambda \\ &\approx \frac{1}{K} \sum_{k=1}^K p(y_m|\lambda^{(k)}, y_o), \end{aligned}$$

where $\theta^{(k)} \sim p(\lambda|y_o)$, $k = 1, \dots, K$ denote K realizations of the Markov chain. This formulation can also be used for predicting pollution levels on future days.

5.3.3 Results

In general it is recommended that a number of initial analyses are performed to examine modelling assumptions. In particular, each pollutant may be modelled separately at each site, which, if all four pollutants were available at each of the sites, would give, in this example, 32 analyses. The fits to individual sites/pollutants may be examined and the plausibility of further modelling assumptions may also be assessed. Later models that incorporate data from multiple sites and more than one pollutant then basically smooth across these analyses (5.6). The spatial and temporal aspects can then be combined to carry out four analyses, one for each pollutant (Section 5.4). Similarly, for each site all four pollutants may be modelled simultaneously (5.5). Examination of changes in model parameters across these analyses allows an assessment of the impact of modelling assumptions.

The use of the single pollutant-single site model is now demonstrated using the London data. At this stage, separate analyses are performed for each pollutant at each site. The analyses

Table 5.4: Observation and system variances, with coefficients for temperature, for two pollutants, PM₁₀ and SO₂, modelled individually at two sites, Bloomsbury and Bexley. The three sets of results show the effect of adding linear and quadratic effects of temperature to the model.

	Without Temp.			Linear Temp.			Quad. Temp		
	Median	2.5%	97.5%	Median	2.5%	97.5%	Median	2.5%	97.5%
PM₁₀, Bloomsbury									
σ_v	0.2057	0.1822	0.2278	0.2839	0.2694	0.2989	0.2774	0.2639	0.2916
σ_w	0.2393	0.2143	0.2661	0.1710	0.1480	0.1977	0.1528	0.1313	0.1753
sd(θy)	0.3639	0.3459	0.3842	0.3263	0.3038	0.3511	0.3068	0.2856	0.3291
Temp.	-	-	-	-0.0006	-0.0092	0.0076	-0.0899	-0.1071	-0.0721
Temp. ²	-	-	-	-	-	-	0.0042	0.0035	0.0049
PM₁₀, Bexley									
σ_v	0.3235	0.3057	0.3421	0.3229	0.3049	0.3415	0.3144	0.2977	0.3319
σ_w	0.2037	0.1743	0.2357	0.2045	0.1760	0.2387	0.1848	0.1577	0.2138
sd(θy)	0.5435	0.4007	1.0250	0.4916	0.3796	0.9152	0.4281	0.3649	0.5880
Temp.	-	-	-	-0.0038	-0.0142	0.0071	-0.1055	-0.1260	-0.0850
Temp. ²	-	-	-	-	-	-	0.0047	0.0039	0.0056
SO₂, Bloomsbury									
σ_v	0.5709	0.5437	0.5989	0.5688	0.5426	0.5962	0.5632	0.5383	0.5897
σ_w	0.1690	0.1335	0.2095	0.1567	0.1246	0.1926	0.1349	0.1060	0.1679
sd(θy)	0.4680	0.4257	0.5119	0.4170	0.3743	0.4621	0.3886	0.3469	0.4353
Temp.	-	-	-	-0.0411	-0.0552	-0.0270	-0.1590	-0.1893	-0.1297
Temp. ²	-	-	-	-	-	-	0.0055	0.0043	0.0067
SO₂, Bexley									
σ_v	0.6433	0.6041	0.6862	0.6334	0.5949	0.6745	0.6192	0.5811	0.6605
σ_w	0.2655	0.2081	0.3313	0.2745	0.2216	0.3377	0.2488	0.1894	0.3128
sd(θy)	1.2880	0.9008	2.0620	1.1000	0.7747	2.0190	1.0860	0.6726	1.7760
Temp.	-	-	-	-0.0395	-0.0609	-0.0181	-0.2083	-0.2546	-0.1620
Temp. ²	-	-	-	-	-	-	0.0077	0.0058	0.0095

are performed with and without the use of daily temperature. In later sections the inference is refined as the data are linked across sites and pollutants via the hierarchy. Prior sensitivity is also addressed, and residuals examined to assess the assumptions of the model.

Initially the prior distributions are specified as $\sigma_v^{-2} \sim Ga(1, 0.01)$, $\sigma_w^{-2} \sim Ga(1, 0.01)$ and β_1, β_2 (the linear and quadratic effects of temperature) $\sim N(0, 1000)$, the latter corresponding to vague prior beliefs. The choices for the precisions give approximate values of the standard deviations of 0.1, i.e. on the original scale, a coefficient of variation of approximately 10% with a large spread. Two separate chains starting from different initial values were run for each model. Convergence was assessed by visual examination of the ‘time series’ plots of the samples for each chain, and computing the Gelman and Rubin statistic (Gelman and Rubin (1992)), which calculates the ratio of the between to within chain variability. The two Markov chains were run for 15,000 iterations, discarding the first 5,000 of each as ‘burn in’.

Table 5.4 contains posterior summaries for the results of two pollutants, PM₁₀ and SO₂, analysed at the two sites with the longest runs of data, Bloomsbury and Bexley. The estimates of the system variances for PM₁₀ from all eight sites were of similar magnitude, as were the measurement

variances. The day-to-day variability, as measured by $sd(\theta_t|y)$, showed greater differences over the sites, but was, in general, a greater component of variability than the measurement error component. For example, for PM_{10} at the two sites presented in Table 5.4, the variance of the measurement error accounts for around 25% of the total variability, i.e. $\sigma_v^2 + \text{var}(\theta|y)$. For prediction or inference for missing values, higher precision is obtained if the total variability is primarily from sources that are modelled via components that are smoothed, across time in this case. For SO_2 , there was greater overall variability than seen with PM_{10} , with σ_v , σ_w and $sd(\theta_t|y)$ generally being higher, with lowest variability (as with PM_{10}) being observed at the Bloomsbury site. In general, measurement error comprised a higher proportion of total variability than for PM_{10} . The variability observed in both NO and CO was greater with more heterogeneity over the eight sites than both PM_{10} and SO_2 , with NO showing the largest differences.

From the relationships observed in Table 5.2 and due to the mechanisms affecting the pollutant levels discussed in Section 5.1, a relationship with temperature would be expected. Examination of the scatterplots of PM_{10} versus temperature indicated that a simple linear relationship would not be an adequate description, due to the seasonal changes in the composition of PM_{10} , for example the larger contribution of ozone in the warmer months and from heating materials in the winter. Such a plot also showed large residual variability, indicating that the predictive effect of temperature would not be strong. The observation variances are relatively unchanged when temperature is added (apart from PM_{10} for Bloomsbury where the median increases), but the underlying variation, represented by $sd(\theta|y)$, are reduced, with the reductions being comparatively larger for PM_{10} than for the other pollutants.

Figure 5.4 shows the time series of a subset of 250 days from the observed data along with the posterior medians of θ_t , denoted $\hat{\theta}_t$ and 90% posterior intervals for the true θ_t . These intervals are for θ_t and not y_t , and so do not include measurement error. The increased width of the intervals at the missing values should be noted. Also shown are the residuals $Y_t - \hat{\theta}_t$ with the dotted lines representing $\pm 2\hat{\sigma}_v$ where $\hat{\sigma}_v$ denotes the posterior median. There do not appear to be systematic deviations.

The posterior medians presented in Table 5.4 appear to be relatively robust to the choice of prior distributions for σ_v^{-2} and σ_w^{-2} . A number of prior combinations were tried, including the choices $Ga(1, 0.01)$, $Ga(0.5, 0.0005)$ and $Ga(0.001, 0.001)$, with little difference in the resulting posterior medians.

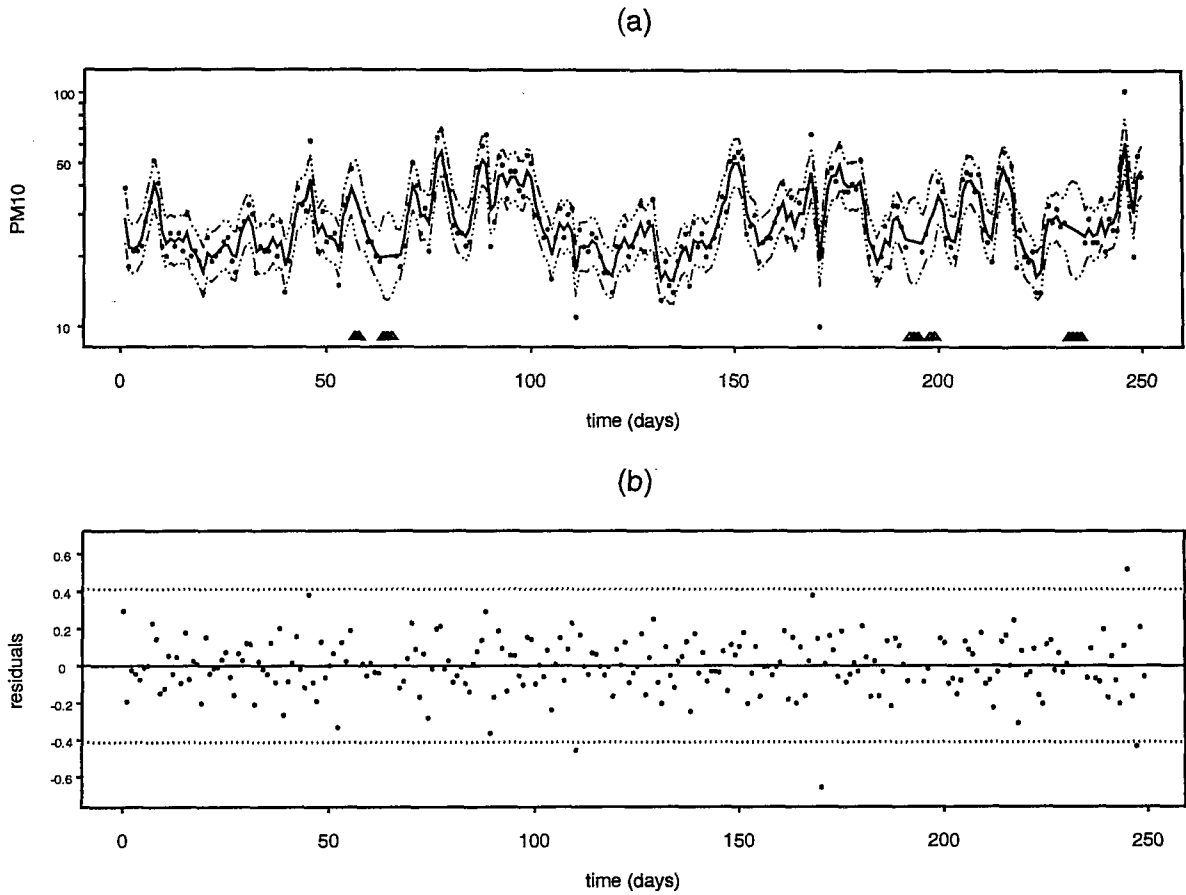


Figure 5.4: Time series plot for a subset of 250 days: (a) recorded measurements and estimated $\hat{\theta}_t$ for (log) PM_{10} , Bloomsbury, triangles indicate missing values; (b) differences between recorded and estimated values. In (a) the dotted lines indicate approximate 90% intervals, in (b) the dotted lines are $\pm 2\hat{\sigma}_v$.

5.4 Single pollutant, multiple monitoring sites

5.4.1 Model

The situation in which we have $S = 8$ monitoring sites measuring a single pollutant is now considered. The model assumes that the underlying autoregressive structure remains constant across sites but there is a constant adjustment in the mean level for site s by an amount m_s , $s = 1, \dots, S$. Let Y_{ts} denote the pollution level on day t at site s , X'_{st} , $t = 1, \dots, T$, $s = 1, \dots, S$, a $q \times 1$ vector of covariates that may depend on site as well as time. Site and time varying variables might include temperature variables, whilst temporally constant variables might reflect spatial characteristics of the site that are constant across time such as latitude and longitude (which could be used to remove any trend), or characteristics of the monitor, for example roadside or elevation. The model now consists of:

Stage One, Observed Data Model:

$$Y_{st} = X'_{st}\beta_1 + X'_s\beta_2 + m_s + \theta_t + v_{st}$$

with v_{st} i.i.d. as $N(0, \sigma_{vs}^2)$ and β_1, β_2 , $q_1 \times 1$ and $q_2 \times 1$ vectors of site/day and site only regression coefficients.

Stage Two (a), Temporal Model:

$$\theta_t = \theta_{t-1} + w_t \tag{5.5}$$

with w_t i.i.d. as $N(0, \sigma_w^2)$.

Stage Two (b), Spatial Model:

The collection of random effects, $m = (m_1, \dots, m_S)'$, are assumed to arise from the multivariate normal distribution

$$m \sim MVN(0_S, \sigma_m^2 \Sigma_m), \tag{5.6}$$

where 0_S is an $S \times 1$ vector of zeros, σ_m^2 the between-site variance and Σ_m is the $S \times S$ correlation matrix, in which element (s, s') represents the correlation between sites s and s' , $s, s' = 1, \dots, S$. This model is stationary (a point discussed in more detail later) and an isotropic covariance model assumed, in which the correlation between sites s and s' is assumed to be a function of the distance between them, i.e. $f(d_{ss'}, \phi)$, where $d_{ss'}$ represents the distance between sites s and s' (in km). Such models are frequently used in geostatistics (e.g. Cressie (1993)), and the specific form assumed here is,

$$f(d_{ss'}, \phi) = \exp(-\phi d_{ss'}) \tag{5.7}$$

where $\phi > 0$ describes the strength of the correlation. Note that the log correlation is linear in distance for this model and so the appropriateness of this form may be assessed using plots such as Figure 5.3. The use of a stationary and isotropic model with a single parameter is restrictive, but with only eight sites it is difficult to specify a more general model. In London, this assumption is likely to be more realistic than in other locations, since the meteorology and topography are relatively spatially stable. For more remarks concerning the specification of a spatial model see the discussion of Diggle et al. (1998). With a larger number of sites, a two parameter isotropic model could be used. For example, the Matérn class (e.g. Handcock and Wallis (1994)) has desirable qualities.

A simpler model that is clearly unrealistic but that we use for comparison is given by:

$$m_s \sim \text{i.i.d } N(0, \sigma_m^2), \quad (5.8)$$

$s = 1, \dots, S$. This model assumes that the site-specific levels are (conditionally) independent.

Stage Three, Hyperprior:

Prior distributions for $\beta_1, \beta_2, \sigma_{vs}^2, s = 1, \dots, S, \sigma_w^2, \phi$ and σ_m^2 are assigned at this stage. Priors for β and σ_w^2 are as in Section 5.3.1. Unless there is specific information to the contrary, i.e. that a monitor with different characteristics is used at a particular site, it is assumed $\sigma_{vs}^{-2} \sim Ga(a_v, b_v)$, $s = 1, \dots, S$. Additionally $\sigma_m^{-2} \sim Ga(a_m, b_m)$. A uniform prior is used for ϕ_p , with the limits being based on beliefs about the relationship between correlation and distance. For example, the distance, d , at which the correlation, ρ , between two sites might be expected to fall to a particular level would be $d = -\log(\rho)/\phi$.

Berger et al. (2001) provide an interesting discussion of the choice of priors in spatial models, and in particular show that the improper uniform prior on the positive real line leads to an improper posterior distribution. There is a lack of identifiability with this model since a constant pollution surface is consistent with both a non-zero mean process with zero correlation, and a zero-mean process with correlations of one, and the data alone cannot distinguish between these possibilities.

5.4.2 Inference

Using this model it is possible to estimate the site effects, and thus pollution levels, at locations where there is no monitoring site. Again, considering a generic pollutant; based on the posterior estimates of the site effects, m_s and the variance-covariance matrix $\sigma_m^2 \Sigma_m$, then for a site at a new location, m_{S+1} , the vector of levels at $S + 1$ locations $(m_1, \dots, m_S, m_{S+1})$ follows a multivariate normal distribution with zero mean and $(S + 1) \times (S + 1)$ variance-covariance matrix with upper $S \times S$ component given by $\sigma_m^2 \Sigma_m$, element $(S + 1) \times (S + 1)$ given by σ_m^2 , and final row, without this element, given by the correlations $\sigma_m^2 f(d_{s,S+1}, \phi)$, $s = 1, \dots, S$. The latter is denoted by the $S \times 1$ vector $\sigma_m^2 \Omega$. Letting $m = (m_1, \dots, m_S)'$, the conditional distribution of $m_{S+1}|m$ is, by

properties of the multivariate normal distribution, also normal with mean and variance given by

$$E[m_{S+1}|m] = \sigma_m^{-2} \Omega' \Sigma_m^{-1} m,$$

and

$$\text{var}(m_{S+1}|m) = \sigma_m^2 (1 - \Omega' \Sigma_m^{-1} \Omega)^{-1},$$

respectively. For exploratory purposes, the posterior medians (for example) of the parameters may be substituted into these expressions. A more accurate, though computationally expensive, strategy is to average the normal distributions that result from individual draws from the posterior distribution.

5.4.3 Results

Vague normal priors $N(0, 1000)$ were assigned to β_1 and β_2 , while for the precisions σ_{vs}^{-2} , $s = 1, \dots, S$, σ_w^{-2} and σ_m^2 , the distributions $Ga(1, 0.01)$ were assumed. With a small number of sites, there will be little information on ϕ , the parameter relating correlation with the distance between sites, and it will therefore be highly affected by the choice of prior. In this example, the limits of the uniform distribution were chosen to represent the cases where the correlation falls to 0.95 at a distance of 20km, reflecting very high correlation between all the sites, and 0.1 indicating far less spatial dependence.

The results of jointly modelling PM₁₀ at the eight sites, using exchangeable site effects (5.8) and those modelled multivariately according to (5.7) are shown in Table 5.5 and may be compared with those described in Section 5.3 and shown in Table 5.4. It is noted that the medians are virtually unchanged under the two spatial models, since there are large amounts of data at each site. The intervals are generally slightly wider under the spatial model reflecting the loss of information with dependent observations. The posterior for ϕ is diffuse due to the small number of sites, and it shows high correlation between sites measuring PM₁₀ within a urban area; the median corresponds to the correlation at 30km falling to 0.85 ($-\exp(\phi d)$) with a 95% credible interval of (0.75, 0.94). The effect of distance on the correlations between site measurements of the other pollutants was greater, with SO₂, NO and CO falling to approximately 0.75 (not shown).

In comparison with the results for pollutant site-specific analyses, there is a noticeable decrease in the observation variance at all of the sites. Although the system variances increase slightly for the Bloomsbury and Bexley sites, they decrease for the others, for which higher system variances were observed in the separate analyses (not shown). This reflects the fact that the underlying level, θ_t , is now 'responding' to eight series of data rather than just 'smoothing' one. This could bring into question the assumption of independence between space and time. Notwithstanding the advantages of borrowing information to produce estimates of the underlying levels for sites with unstable/or sparse data, if the objective of the analysis is to produce a set of estimates for

Table 5.5: Observation and system variances, with coefficients for temperature, for PM_{10} modelled jointly at eight sites, but independently of measurements of other pollutants. Site effects, and variances are presented for both independent and multivariate spatial models. The labelling is: 1 = Bexley, 2 = Bloomsbury, 3 = Brent, 4 = Eltham, 5 = Harringey, 6 = Hillingdon, 7 = North Kensington, 8 = Sutton.

	Independent			Multivariate		
	Median	2.5%	97.5%	Median	2.5%	97.5%
σ_{v1}	0.1361	0.1290	0.1437	0.1361	0.1291	0.1438
σ_{v2}	0.1279	0.1211	0.1349	0.1278	0.1211	0.1348
σ_{v3}	0.1042	0.0976	0.1115	0.1042	0.0975	0.1115
σ_{v4}	0.1248	0.1169	0.1334	0.1248	0.1170	0.1335
σ_{v5}	0.1222	0.1145	0.1306	0.1222	0.1145	0.1307
σ_{v6}	0.2116	0.1987	0.2266	0.2114	0.1984	0.2259
σ_{v7}	0.0632	0.0573	0.0695	0.0632	0.0573	0.0693
σ_{v8}	0.2290	0.2166	0.2425	0.2290	0.2166	0.2429
σ_w	0.3462	0.3326	0.3605	0.3465	0.3328	0.3609
$sd(\theta y)$	0.4062	0.4007	0.4143	0.4067	0.4005	0.4164
m_1	-0.0696	-0.0785	-0.0605	-0.0696	-0.0785	-0.0607
m_2	0.1341	0.1256	0.1429	0.1341	0.1257	0.1426
m_3	-0.1209	-0.1292	-0.1120	-0.1210	-0.1294	-0.1125
m_4	-0.1103	-0.1203	-0.0998	-0.1105	-0.1205	-0.1005
m_5	0.1098	0.1000	0.1200	0.1098	0.0999	0.1195
m_6	0.0131	-0.0043	0.0299	0.0132	-0.0032	0.0300
m_7	0.0031	-0.0031	0.0094	0.0030	-0.0031	0.0090
m_8	0.0410	0.0250	0.0575	0.0410	0.0250	0.0572
σ_m	0.0955	0.0627	0.2202	0.1019	0.0668	0.1794
ϕ	-	-	-	0.005675	0.002158	0.009778
Temp.	-0.0905	-0.1033	-0.0741	-0.0872	-0.1091	-0.0609
Temp. ²	0.0037	0.0031	0.0043	0.0035	0.0026	0.0044

a single site for which there is already a relatively stable series of data over the study period, i.e. Bloomsbury in this example, there may be little to be gained from incorporating data from other sites.

The coefficients of temperature, linear and quadratic, are similar to those seen in the separate analyses, but their addition to the model now accounts for a smaller reduction in the system variance (3.5%). Examining the average of the $S = 8$ measurement errors, the temporal, spatial and measurement error components account for, respectively, 80%, 10% and 10% of the total variability. Hence the variability due to the temporal components dominates that from the measurement error and spatial variability, as measured by σ_m .

Estimates of the differences in levels recorded at the eight sites are also given, the posterior medians of the site effects, m_s , range from -0.1209 to +0.1341. The site effects for the Bloomsbury and Bexley sites are +0.1341 and -0.0696 respectively, indicating the higher levels observed at the Bloomsbury site, which is in the centre of London (defined as an urban background site), compared to those at Bexley, which is on the outskirts of the city and defined as a suburban site. Similar results were observed for the other pollutants, with the site effects being more pronounced for CO and NO, reflecting the greater differences observed in the individual site analyses.

5.5 Multiple pollutants, single monitoring site

In this section, the case of a single monitoring site is again considered, but here all four pollutants are modelled simultaneously. As with the single pollutant case, the observed daily measurements, are assumed to arise from the true underlying level of the pollutant in question, measured with some error. At the second stage of the model, the fact that daily changes in each of the pollutants will be correlated is incorporated, by modelling the autoregressive structure of the underlying levels as multivariate normal with mean vector equal to the previous values of the estimates for each pollutants and a variance-covariance matrix reflecting both the underlying or system variance, and the correlation between the difference pollutants. Hence, this model allows the amount of temporal smoothing to be pollutant dependent.

5.5.1 Model

Let Y_{pt} denote the observed value of pollutant p on day t , $p = 1, \dots, P = 4$, $t = 1, \dots, T$. The model presented in Section 5.3.1 is extended to give:

Stage One, Observed Data Model:

$$Y_{pt} = X_t' \beta_p + \theta_{pt} + v_{pt}$$

with v_{pt} i.i.d. as $N(0, \sigma_{vp}^2)$ and β_p a $q \times 1$ vector of regression coefficients. The latter allows for the possibility that the relationship between pollution level and explanatory variables (e.g. temperature) may be pollutant dependent.

Stage Two, Temporal and Pollutant Model:

$$\theta_{pt} = \theta_{p,t-1} + w_{pt},$$

for $p = 1, \dots, P$. Here $w_t = (w_{1t}, \dots, w_{Pt})'$ are i.i.d. multivariate normal random variables with zero mean and variance-covariance matrix Σ_P . This matrix contains variances σ_{wp}^2 thus allowing different pollutants to have different amounts of temporal dependence, and $P(P-1)/2$ covariance terms reflecting the dependence (more precisely the covariance) between each of the pollutants, conditional on the previous day's values.

Stage Three, Hyperpriors:

Prior distributions for β_p , σ_{vp}^2 , Σ_P and σ_w^2 are assigned at this stage. The priors β_p and σ_w^2 are as in Section 5.3. Unless there is specific information to the contrary it is assumed that $\sigma_{vp}^{-2} \sim Ga(a_v, b_v)$, $p = 1, \dots, P$. For the covariance matrix, it is assumed that $\Sigma_P^{-1} \sim W_P(D, d)$ where $W_P(D, d)$ denotes a P -dimensional Wishart distribution with mean D and precision parameter d . The precision parameter is chosen to be $d = P$, the dimension of the covariance matrix, which corresponds to the flattest 'proper' distribution. Given that the modelling is performed on the logarithmic scale, the variances σ_{wp}^2 are approximately equal to the coefficient of variation of the underlying states on the original scale (which is a more natural scale on which to specify priors). For the mean, values are chosen such that the diagonals of the expected value (D/d) relate to the coefficient of variation that might be expected. The off-diagonal elements of D/d are chosen to reflect the expected correlations between the pollution-specific states.

5.5.2 Results

This model was applied to data on four pollutants (PM₁₀, SO₂, NO and CO) for each site individually. Again the priors $\sigma_{vp}^{-2} \sim Ga(1, 0.01)$, $p = 1, \dots, P$, and $\beta_1, \beta_2 \sim N(0, 1000)$, were assigned. For the parameters of the Wishart distribution, d was chosen to be equal to four, the dimension of Σ_P ; D was then chosen so that the diagonals of the expected value (D/d) represent a 10% coefficient of variation. The off-diagonals were taken to be zero.

The results of the model applied to data from the Bloomsbury site can be seen in Tables 5.6 and 5.7. The results for PM₁₀ shown in Table 5.6 are similar to those seen in the univariate example (Table 5.4), with a increase in the system variance σ_w^2 , an effect also seen at the other sites.

Table 5.6: Observation and system variances, with coefficients for temperature, for four pollutants modelled using the multivariate pollution model. Results are from the Bloomsbury site, modelled independently of measurements at other sites.

	Median	2.5%	97.5%
PM₁₀			
σ_{v1}	0.2258	0.2148	0.2375
σ_{w1}	0.3424	0.3124	0.3737
$sd(\theta_1 y)$	0.4099	0.3899	0.4299
Temp ₁	-0.0972	-0.1144	-0.0782
Temp ₁ ²	0.0043	0.0035	0.0050
SO₂			
σ_{v2}	0.4823	0.4607	0.5049
σ_{w2}	0.4715	0.4188	0.5260
$sd(\theta_2 y)$	0.5948	0.5503	0.6385
Temp ₂	-0.1663	-0.1981	-0.1357
Temp ₂ ²	0.0052	0.0038	0.0067
NO			
σ_{v3}	0.2502	0.2378	0.2633
σ_{w3}	0.3703	0.3357	0.4069
$sd(\theta_3 y)$	0.4795	0.4578	0.5015
Temp ₂	-0.0904	-0.1089	-0.0716
Temp ₃ ²	0.0031	0.0024	0.0037
CO			
σ_{v4}	0.3383	0.3220	0.3555
σ_{w4}	0.5511	0.5025	0.6020
$sd(\theta_4 y)$	0.6661	0.6350	0.6970
Temp ₄	-0.1385	-0.1652	-0.1122
Temp ₄ ²	0.0027	0.0016	0.0039

Table 5.7: Posterior medians of the Σ_p matrix between the underlying levels of four pollutants modelled using the multivariate pollution model at the Bloomsbury site, modelled independently of measurements at other sites. The variances lie on the diagonal, with covariances above and correlations below.

	PM ₁₀	SO ₂	NO	CO
PM ₁₀	0.3424 ²	0.1442	0.1039	0.1535
SO ₂	0.8806	0.4715 ²	0.1472	0.2315
NO	0.8192	0.8472	0.3703 ²	0.1866
CO	0.8134	0.9202	0.9146	0.5511 ²

The posterior medians of the (conditional) correlations between the underlying levels of the pollutants are given in Table 5.7. These are not strictly comparable to those seen in Table 5.2 as they are conditional on the previous day's level and do not include the non-negligible effect of independent measurement errors. Strong correlations are observed between all the parameters, meaning that in the presence of missing values of one pollutant, inference can be made by borrowing information from the non-missing values of the others.

Decomposing the total variability into that due to measurement error and that due to day-to-day variability it is apparent that the latter again dominates with 77%, 60%, 79% and 79% of the total variability being explained by the temporal variance for PM_{10} , SO_2 , NO and CO, respectively.

The effect of using different prior beliefs about the variation and correlation was examined by using Wishart distributions representing combinations of coefficients of variations ranging from 0.1 to 0.75 with correlations from 0.001 to 0.9. Very little effect on the resulting posterior medians, particularly for the correlation between the system parameters, was observed. This may be expected since there are abundant daily measurements.

Figures 5.5(b) and (c) show contour plots of the site effects for PM_{10} with standard deviations, using the posterior median values from the model, on a 20×20 grid covering the study area. As expected the variability of the effects increases with distance from the actual monitoring sites, and despite the small number of sites used, spatial patterns can be observed in the site effects, with higher values in the centre of the city. The predictions should be viewed with caution as the positions of the original monitors have not been accounted for, e.g. roadside, and this may strongly influence the level of recorded pollution (especially for NO and CO).

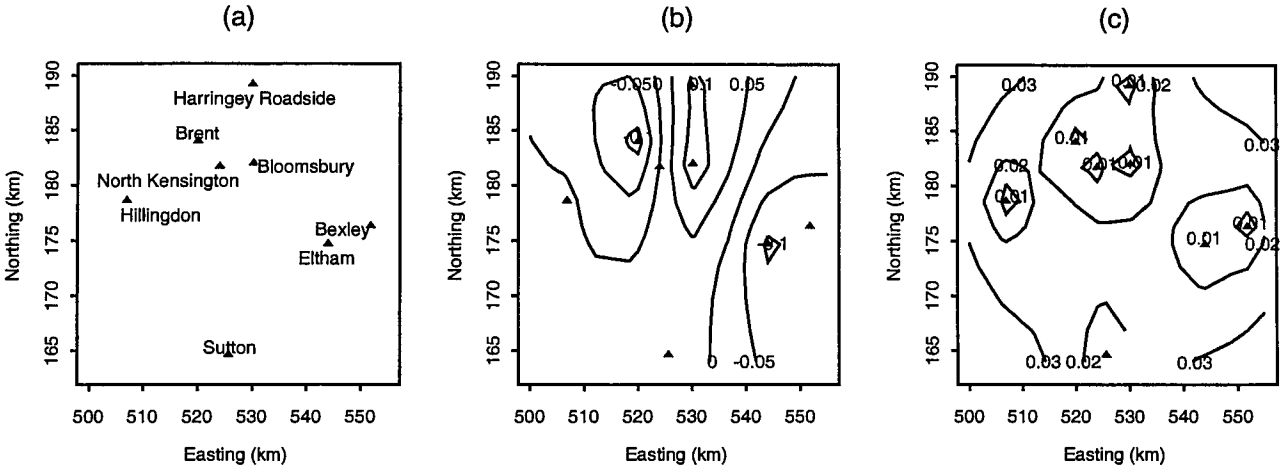


Figure 5.5: Contour plot of (a) site locations, (b) site effects based on a 20x20 grid of locations without a pollution monitor and (c) the corresponding standard deviations of the site effects. Results are for PM_{10} .

5.6 Multiple pollutants, multiple monitoring sites

In this section, the models discussed in the previous sections are combined for use where there are multiple pollutants being measured at a number of monitoring sites. The correlation between sites (Section 5.4) and that between daily values of the different pollutants (Section 5.5) are incorporated within a final model. This model is general in that it contains the simpler models previously described, allowing for a temporal-pollutant interaction and a spatial-pollutant interaction, with the spatial model being constant across time, isotropic and stationary. The general model and its assumptions are now described, together with summaries of those components previously described.

5.6.1 The general model

Stage One, Observed Data Model:

Let Y_{spt} denote the observed level of pollutant p at spatial location s on day t and assume

$$Y_{spt} = X_{spt}\beta + \theta_{pt} + m_s + v_{spt}, \quad (5.9)$$

In this model:

- v_{spt} represents the *measurement errors* which are assumed i.i.d. $N(0, \sigma_{sp}^2)$.
- m_s represents the spatial effect of being at site s .
- θ_{pt} is the error term that will induce temporal and pollutant dependence at Stage 2.
- The $q \times 1$ parameter vector β is a vector of regression coefficients.
- X_{spt} represents a $1 \times q$ vector of regressors that may change temporally (for example, temperature), and spatially. The latter may represent, for example, spatial characteristics of the site that may be constant across time such as latitude and longitude (which could be used to remove any trend), or characteristics of the monitor, for example roadside or elevation. The subscript p allows these effects to be pollutant specific.

At the first stage, the daily ($t = 1, \dots, T$) measurements of each pollutant ($p = 1, \dots, P$) at each monitoring site ($s = 1, \dots, S$) are modelled as a function of the true underlying level of the pollutant with a site adjustment and a pollutant-site specific error term.

Stage Two (a), Spatial/Pollutant Model:

As described in Section 5.4, the collection of random effects $m_p = (m_{p1}, \dots, m_{pS})'$, $p = 1, \dots, P$, is assumed to arise from the multivariate normal distribution

$$m_p \sim MVN(0_S, \sigma_{pm}^2 \Sigma_{pm}), \quad (5.10)$$

where 0_S is an $S \times 1$ vector of zeros, σ_{pm}^2 the between-site variance for pollutant p and Σ_{pm} is the $S \times S$ correlation matrix, in which element (s, s') represents the correlation between sites s and s' , $s, s' = 1, \dots, S$, for pollutant p .

Stage Two (b), Temporal/Pollutant Model:

$$\theta_{pt} = \theta_{p,t-1} + w_{pt}, \quad (5.11)$$

for $p = 1, \dots, P$. Here $w_t = (w_{1t}, \dots, w_{Pt})'$ are i.i.d. multivariate normal random variables with zero mean and variance-covariance matrix Σ_P . This matrix contains variances σ_{wp}^2 thus allowing different pollutants to have different amounts of temporal dependence, and $P(P-1)/2$ covariance terms reflecting the dependence (more precisely the covariance) between each of the pollutants, conditional on the previous day's values.

Stage Three, Hyperprior:

A normal prior $N(c, C)$ is assumed for β , where c is a $q \times 1$ vector and C a $q \times q$ variance-covariance matrix. Gamma priors are specified for the precisions, specifically $\sigma_{sp}^{-2} \sim Ga(a_v, b_v)$. The variance-covariance matrix, $\Sigma_P^{-1} \sim W_P(D, d)$ where $W_P(D, d)$ denotes a P -dimensional Wishart distribution with mean D and precision parameter d . Unless there is specific information to the contrary, i.e. that a monitor with different characteristics is used at a particular site, it is assumed $\sigma_{vs}^{-2} \sim Ga(a_v, b_v)$, $s = 1, \dots, S$. A uniform prior is used for ϕ_p , with the limits being based on beliefs about the relationship between correlation and distance.

The assumptions of the model include the following:

- The measurement error variance σ_{sp}^2 does not depend on time, although the model is easily extendable to situations in which the measurement error may change as a function of t .
- The relationship between the pollutants is constant over time.
- The relationship between the pollutants is spatially constant.
- The temporal and spatial components are independent.

Discussing this last point in more detail, for notational convenience, the dependence on p is suppressed and a generic pollutant considered. If data are missing at a site s' and at a time t' , these will be 'filled in' using $m_{s'}$ and $\theta_{t'}$, where the former component does not change with time. In the spatio-temporal literature, *separable* models are often considered, these impose a particular type of independence between space and time components. Let $\rho_{s'}$ denote the correlation between observations Y_s and $Y_{s+s'}$, $\rho_{tt'}$ the correlation between Y_t and $Y_{t+t'}$, and $\rho_{s't'}$ the correlation between $Y_{s+s', t+t'}$ and $Y_{s, t}$. Then the correlation of a stationary model is separable if

$$\rho_{s't'} = \rho_{s'} \times \rho_{t'}.$$

The model presented here exhibits a different kind of independence, though since it is non-stationary, the marginal distribution of Y cannot be evaluated. Consider the model

$$\theta_t = \alpha\theta_{t-1} + w_t,$$

with $|\alpha| < 1$. Combining this stationary time series model with the spatial model gives

$$\rho_{s't'} = q \times \rho_{s'} + (1 - q) \times \rho_{t'},$$

where $q = \sigma_m^2 / (\sigma_m^2 + \sigma_w^2)$, $\rho_{t'} = \alpha^{|t'|} / (1 - \alpha^2)$, illustrating that the ‘joint’ correlation is a weighted sum of the spatial and temporal components, with the weights being independent of t and s . The model presented here is obtained with $\alpha = 1$ as the limiting form of this model.

5.6.2 Results

The results from the final model are shown in Tables 5.8 and 5.9. Again there is very strong correlation between the system, or underlying, parameters (Table 5.9). The posterior medians of the system standard deviations, σ_{wp} , can be seen in Table 5.8 together with those for the site-pollutant random (measurement) errors and the coefficients for temperature. The results for PM_{10} shown in section 5.3 (Table 5.5) and the other pollutants (not shown) when they were modelled independently of the other pollutants are similar to the results from this multi-pollutant model. The site effects are also given, with slightly increased standard deviations, σ_{mp} , and show a pattern similar to those from the single pollutant models, again with those for NO and CO being more pronounced. The temporal component, represented by $\sigma(\theta|y)$, is also similar to that observed when PM_{10} was jointly modelled with the other pollutants at a single site (Section 5.5, Table 5.6). The measurement error (and temperature effects) are also consistent. The proportions of variability associated with the temporal, spatial and measurement error components are also similar to those previously observed, at 75%, 15% and 10% respectively for PM_{10} .

The results for SO_2 are also comparable to those seen from both the individually (Section 5.3, Table 5.4) and jointly (Section 5.5, Table 5.6) modelled pollutant effects. Again, the temporal effect dominates (65%), but with a larger effect of measurement error (29%) and a reduction in the spatial component (6%). In comparison to the results from the single site (multiple pollutant) models, the temporal variation associated with NO and CO is reduced, with a corresponding increase in the measurement error component.

Table 5.8: Posterior medians from the general model with dependencies across time and pollutants, with pollutant specific spatial effects. Pollutant specific observation and system variances, together with coefficients for temperature, are presented together with site effects and variances, for the multivariate spatial components. The labelling is: 1 = Bexley, 2 = Bloomsbury, 3 = Brent, 4 = Eltham, 5 = Harringey, 6 = Hillingdon, 7 = North Kensington, 8 = Sutton.

PM ₁₀							
	Median	2.5%	97.5%		Median	2.5%	97.5%
σ_{v11}	0.1380	0.1310	0.1456	m_{11}	-0.0696	-0.0785	-0.0607
σ_{v21}	0.1273	0.1209	0.1343	m_{21}	0.1344	0.1260	0.1427
σ_{v31}	0.1049	0.0983	0.1122	m_{31}	-0.1209	-0.1293	-0.1124
σ_{v41}	0.1251	0.1172	0.1339	m_{41}	-0.1105	-0.1204	-0.1005
σ_{v51}	0.1227	0.1150	0.1312	m_{51}	0.1098	0.1000	0.1195
σ_{v61}	0.2112	0.1983	0.2255	m_{61}	0.0128	-0.0039	0.0291
σ_{v71}	0.0623	0.0565	0.0684	m_{71}	0.0030	-0.0031	0.0090
σ_{v81}	0.2291	0.2166	0.2429	m_{81}	0.0411	0.0250	0.0570
σ_{w1}	0.3438	0.3302	0.3581	σ_{m1}	0.1297	0.0806	0.2597
$sd(\theta_1 y)$	0.4060	0.4005	0.4138	ϕ_1	0.00752	0.00181	0.01130
Temp. ₁	-0.0951	-0.1155	-0.0811	-	-	-	-
Temp. ₁ ²	0.0040	0.0034	0.0047	-	-	-	-
SO ₂							
	Median	2.5%	97.5%		Median	2.5%	97.5%
σ_{v12}	0.4811	0.4545	0.5099	m_{12}	0.0064	-0.0607	0.0783
σ_{v22}	0.4211	0.4012	0.4422	m_{22}	0.3084	0.2451	0.3756
σ_{v32}	0.4244	0.3992	0.4517	m_{32}	-0.2797	-0.3470	-0.2131
σ_{v42}	0.4743	0.4459	0.5052	m_{42}	-0.1915	-0.2601	-0.1223
σ_{v52}	0.1199	0.0519	0.6343	m_{52}	0.0965	-0.3228	0.5017
σ_{v62}	0.3551	0.3295	0.3830	m_{62}	0.1505	0.0842	0.2173
σ_{v72}	0.3822	0.3580	0.4085	m_{72}	-0.1495	-0.2149	-0.0848
σ_{v82}	0.4135	0.3885	0.4410	m_{82}	0.0594	-0.0057	0.1266
σ_{w2}	0.4710	0.4458	0.4973	σ_{m2}	0.2701	0.1651	0.5590
$sd(\theta_2 y)$	0.5719	0.5559	0.5880	ϕ_2	0.00871	0.00262	0.01138
Temp. ₂	-0.1429	-0.1663	-0.1174	-	-	-	-
Temp. ₂ ²	0.0048	0.0038	0.0058	-	-	-	-
NO							
	Median	2.5%	97.5%		Median	2.5%	97.5%
σ_{v13}	0.1196	0.0523	0.6354	m_{13}	0.1626	-1.2820	2.2640
σ_{v23}	0.2668	0.2509	0.2837	m_{23}	0.3293	0.0293	0.5361
σ_{v33}	0.6721	0.6349	0.7124	m_{33}	-0.8548	-1.1590	-0.6441
σ_{v43}	0.6660	0.6278	0.7067	m_{43}	-0.8725	-1.1730	-0.6616
σ_{v53}	0.3804	0.3557	0.4064	m_{53}	0.6277	0.3269	0.8358
σ_{v63}	0.8761	0.8211	0.9370	m_{63}	0.7097	0.4105	0.9264
σ_{v73}	0.5013	0.4712	0.5334	m_{73}	-0.6188	-0.9207	-0.4092
σ_{v83}	0.4407	0.4136	0.4699	m_{83}	0.5235	0.2223	0.7323
σ_{w3}	0.4994	0.4757	0.5241	σ_{m3}	0.8929	0.5204	2.0340
$sd(\theta_3 y)$	0.6250	0.6105	0.6432	ϕ_3	0.00842	0.00236	0.01138
Temp. ₃	-0.1366	-0.1615	-0.1124	-	-	-	-
Temp. ₃ ²	0.0026	0.0016	0.0036	-	-	-	-
CO							
	Median	2.5%	97.5%		Median	2.5%	97.5%
σ_{v14}	0.3580	0.3419	0.3748	m_{14}	-1.1940	-1.7390	-0.5778
σ_{v24}	0.2725	0.2592	0.2866	m_{24}	-0.6926	-1.2390	-0.0771
σ_{v34}	0.5987	0.5537	0.6491	m_{34}	-1.2810	-1.8270	-0.6542
σ_{v44}	0.1209	0.0526	0.6176	m_{44}	1.5530	-0.9085	5.4500
σ_{v54}	0.1202	0.0525	0.6330	m_{54}	2.7910	0.3848	7.9660
σ_{v64}	0.3465	0.3223	0.3729	m_{64}	-0.6946	-1.2390	-0.0825
σ_{v74}	0.6167	0.5825	0.6539	m_{74}	-0.5123	-1.0560	0.1012
σ_{v84}	0.2753	0.2572	0.2951	m_{84}	-0.1723	-0.7199	0.4416
σ_{w4}	0.2962	0.2806	0.3126	σ_{m4}	1.9380	0.6739	4.7520
$sd(\theta_4 y)$	0.4073	0.3961	0.4185	ϕ_4	0.00894	0.00310	0.01139
Temp. ₄	-0.0975	-0.1140	-0.0800	-	-	-	-
Temp. ₄ ²	0.0032	0.0025	0.0038	-	-	-	-

Whilst the spatial relationships for the different pollutants seen in Figure 5.3 appear somewhat similar, fitting a model with a single spatial effect for all four pollutants resulted in an oversimplification of the individual relationships (apparent from the different values of σ_{mp} and ϕ_p) and ignored the disparate scales of the pollution variables. The observed results from such a model appeared to be dominated by PM_{10} , which has the strongest correlation and for which there are most data available, at the expense of the others, notably NO and CO, which have more marked decreases in correlation with distance, but where the data are more sparse.

Figure 5.6 shows the relationship between temperature and (log) pollutant level with the fitted curve being obtained by substituting the posterior medians into the quadratic model. Clear curvature can be seen for all four pollutants implying that a linear form is not sufficient.

Table 5.9: Posterior medians of the Σ_p matrix between the underlying levels of four pollutants modelled using the multivariate pollution model with multivariate spatial effects for all eight monitoring sites. The variances lie on the diagonal, with covariances above and correlations below.

	PM_{10}	SO_2	NO	CO
PM_{10}	0.3438 ²	0.1142	0.1143	0.0705
SO_2	0.7055	0.4710 ²	0.1889	0.1022
NO	0.6655	0.8027	0.4994 ²	0.1417
CO	0.6924	0.7327	0.9578	0.2962 ²

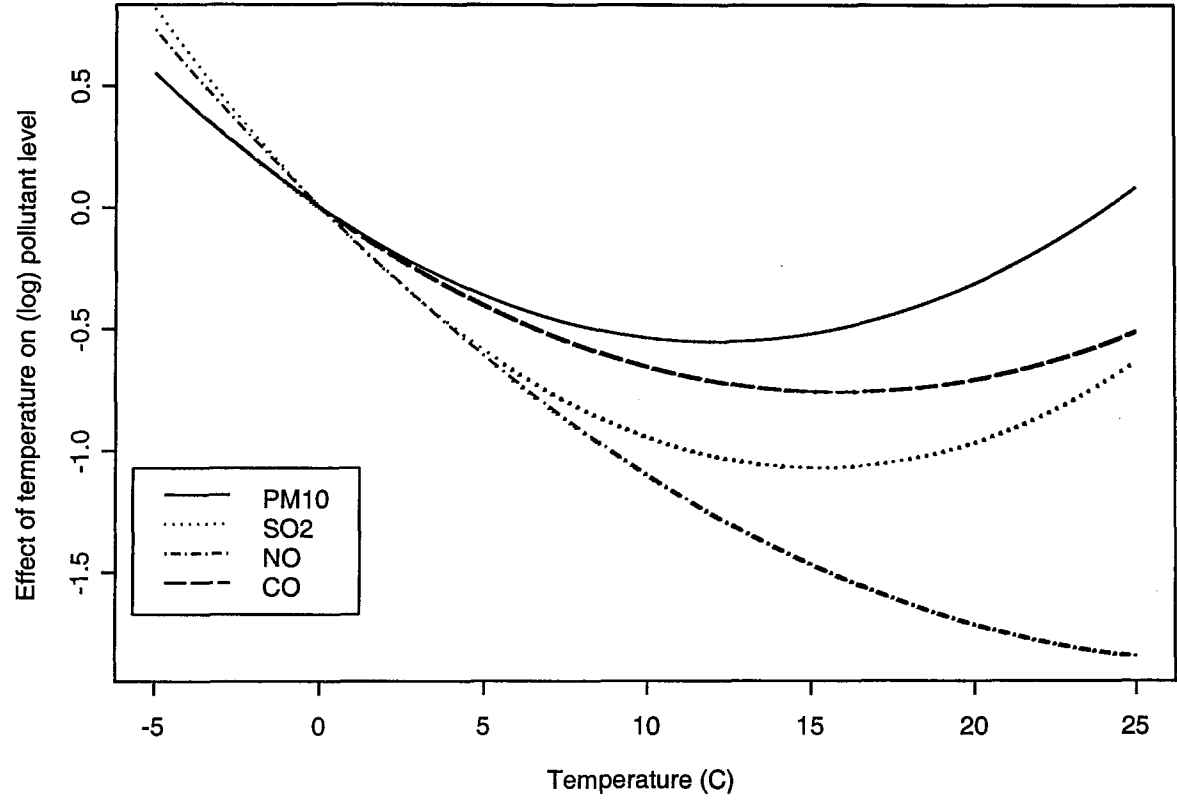


Figure 5.6: The relationship between temperature and (log) pollutant level.

Chapter 6

Discussion

The work presented in this thesis consists of two substantial epidemiological inquiries and a spatio-temporal analysis of routinely measured pollution levels. These are presented together with the development of statistical methods needed to address issues arising in ecological studies of the effects of air pollution and temperature on health. This chapter provides an overall discussion of the issues raised in this thesis, with specific reference to two particularly important areas, those of over-dispersion and model selection. The results from the individual studies are discussed, and possible future developments in the statistical methods are described.

In the first of the epidemiological studies, the effects of long-term exposure to SO_2 are examined in relation to respiratory deaths. Two major studies in the U.S. (Dockery et al. (1993), Pope et al. (1995)) have shown that those living in less polluted cities live longer than those in more polluted cities, however as yet there have been no similarly large scale studies within the U.K. Whereas the two U.S. studies focused on concurrent (or near concurrent) exposures and health outcome, the analyses presented here took into account different durations of exposure, whilst allowing for differences in historical levels of pollution. This analysis was also carried out at a much higher geographical resolution than the U.S. studies, with pollution levels being compared between a large number of small areas, rather than between a small number of cities (particularly in the six cities study (Dockery et al. (1993))). Significant increases in respiratory deaths were associated with areas experiencing higher pollution, with exposures up to the previous 16 years.

The second study examined the short-term effects of cold temperatures, with particular interest in the effects of prolonged periods of cold. Although the effects of temperature have previously been examined, it has often been within studies investigating the short-term effects of air pollution, where it is treated as a nuisance variable to be controlled for, rather than as a potentially important public health issue in its own right. Here, clear associations were observed between mortality and lagged values of temperature, with the largest excess in risk being associated with cold temperatures on the same day, with longer periods of cold being associated with additional excess

deaths.

In conducting any spatial epidemiological study, it is important to have a good measure of exposures within each area and/or time period. In time-series studies investigating the relationship between air pollution and health, difficulties often arise due to periods of missing data and because the exposure variables are often measured with error. In the third analysis in this thesis, a model is proposed that addresses each of these points, and is used to analyse multiple pollutant data from a number of monitoring sites. This produces estimates of pollution (together with measures of uncertainty) that can be used in health analyses.

Throughout the thesis, the issues arising in epidemiological studies performed using routine data at the aggregate level, including those encountered in the examples presented here, are discussed and statistical methodology developed to help address a number of them. The problems of unmeasured confounders, over-dispersion, correlated outcomes, model selection and uncertainty, missing data and measurement error are common to many epidemiological studies, including those in this thesis. Although each of these issues may be addressed using different approaches within different contexts, the ultimate aim in each case is to obtain an unbiased estimate of the 'true' effects, allowing for data inadequacies, confounding and imprecision.

The individual chapters are now discussed in detail.

6.1 Modelling the association between mortality and long term exposure to air pollution

Previous studies have focused on links between concurrent (or near-concurrent) exposures and health outcome, without consideration of possible latency effects over many years. In the study design used here, the nearest equivalent to the two large U.S. studies (the six-cities study (Dockery et al. (1993)) and the ACS study (Pope et al. (1995))), is to consider the association between mortality exposure in the immediately preceding 4 years. Overall, this gave a relative risk of 1.05 (1.02-1.09) per 10 ppb of SO_2 . The corresponding results from the individual health periods were 1.05 (1.02-1.08) for 1981-84, 1.06 (1.01-1.12) for 1985-88, 1.04 (0.99-1.10) for 1989-92 and 1.23 (1.05-1.20) for 1993-96. The effect of SO_2 was also assessed within the six-cities study, where a RR of 1.26 (1.08-1.47) was associated with a difference of approximately 22.4 ppb (comparing the least polluted city, Topeka, Kansas, with an average of 1.6 ppb with the most polluted city, Steubenville, Ohio, with an average of 24.0 ppb). The corresponding RR for such a difference from this analysis is approximately 1.12 (1.08-1.15). Both of the U.S. studies were prospective cohort studies, and thus had the advantage of allowing for adjustment for individual level covariates, such as smoking status, sex, race and possible occupational exposure. The study presented here was at the aggregate level, which did not allow for individual level adjustment. The geographic patterns

of SO₂ pollution shown by the data used here are, in many ways, also an industrial geography. In general, areas of higher SO₂ concentrations are those with higher levels of employment in manufacturing and heavy industry. Occupational exposures in these areas may be important contributors to respiratory mortality, especially amongst men. Some degree of control for this effect has been applied through adjustment for deprivation, by the Carstairs score, but residual confounding is likely to exist. The most important unmeasured confounder is likely to be smoking, and whilst the Carstairs score has been shown to relate to levels of smoking (Kleinschmidt et al. (1995)), it remains an ecological proxy for individual level smoking data. Moreover, the Carstairs index was developed in the 1980s, on the basis of data collected during the 1981 census. Since then, socio-economic structures and their expression through lifestyles has changed markedly; whether the Carstairs score has acted as a consistent proxy for smoking and other behaviours likely to affect respiratory mortality throughout the period of this study, and whether it adequately captures the health related effects of deprivation is open to some doubt. Thus, the Carstairs score may not have provided full adjustment for socio-economic confounding, especially in more recent health periods.

Despite the potential problems of confounding at the individual level, the study does have a number of advantages over the two U.S. studies; longer exposures were considered, all exposures preceded the health outcomes, the comparisons of pollution levels were made at a much higher geographical resolution (up to 400 small areas, rather than across cities), and the estimation of association were able to change over time, allowing for changes in the levels and composition of pollution.

In order to detect any differences in the effects of different lagged exposures on respiratory mortality, associations with SO₂ concentrations were analysed during each health period, for different exposure lags. The relative risks shown in Figure 3.7 in comparison with those shown in Tables 3.9 and 3.10 indicate the presence of latency effects, although there is some suggestion that these effects weaken for longer lags and times of exposure. With increasing lag between exposure and health outcome, there will also have been increased opportunity for migration into and out of the wards, resulting in increased exposure misclassification, and diluting the association between exposure and mortality.

Another question that can be addressed within this study design, is whether evidence exists for cumulative effects of exposure, over different lag periods: that is, whether variations in respiratory mortality can be attributed to a longer history of exposure. The longest exposure period considered here was over the previous sixteen years, for which relative risks of 1.006 (1.000-1.012), 1.012 (1.001-1.023), 0.9981 (0.9851-1.012) and 1.014 (0.9971-1.030) were found per 10 ppb of SO₂ (for 1981-84, 1985-88, 1989-92 and 1993-96 respectively), with an overall RR of 1.006 (1.000-1.012). The non-significant result seen in the fourth period, with the highest risk, may be due to the small number of wards with air pollution monitors present in the latter periods of the study.

For public health purposes, it may be useful to calculate individual ward estimates of risk, together with the associated random errors. Large values of individual ward random errors might indicate high levels of residual deprivation, or the presence of other strong confounding variables in that area, and might indicate the need for further, within-area, investigation.

Overall, the study has shown positive associations between long-term SO₂ exposure and respiratory mortality. Although the study design was intrinsically different, the excess risks were of similar magnitude to those previously observed in studies in the U.S., suggesting that the long term health risks of exposure to air pollution merit continued attention.

6.2 Modelling the effect of cold winter temperatures on mortality

A clear association between cold temperatures and mortality was observed, with a strong suggestion that same day temperature has the greatest explanatory power, with 0.8% (0.6%-0.9%) higher daily deaths associated with a 1⁰C drop in temperature, and excess risks generally decreasing with increasing lags. From a public health perspective, this is potentially a very important finding and could, for example, be used to allocate extra health care provision in periods of cold temperature. Taken in the context of the findings from the Eurowinter group (Eurowinter Group (1997) and other studies of the effects of temperature in extremely cold areas (Donaldson et al. (1998), Donaldson et al. (1998)), which found that the effect of decreases in (cold) temperatures were greater in warmer climates than in colder ones, it may be possible that the risk could be lessened by people being better prepared. For example, potentially susceptible groups such as the elderly could be advised on how to better prepare for prolonged periods of cold, e.g. by wearing more appropriate clothing.

In the Bayesian model averaging analysis, the probability that the same day temperature should be included in the model was one, as it was included in all models with non-zero posterior model probabilities. Lagged values from 3, 6 and 7 days had values of 99%, 30% and 70% respectively for the same measure, with other lagged values having negligible probabilities. The strong presence of same day temperature was also evident when predicting excess deaths for various temperature scenarios, where a decrease from 5⁰C to -2⁰C for same day temperature was associated with approximately 13 extra deaths in London (per day), a number which was relatively constant for periods of up to 3 days, after which the number of excess deaths increased to 18. This compares to corresponding changes in the observed data of 12.5 and 29 respectively (as shown in Table 4.3). Increases in the predicted excess deaths were then observed with longer periods of cold (at -2⁰C), the expected number of extra deaths increasing to 20 for periods of 7 days, and to 23 for 8 days. The corresponding changes in the observed data were both approximately 41.5. The smaller numbers predicted for these extreme scenarios reflect the fact that the model is essentially smoothing the relationship between temperature and mortality, over several different

possible models. There is therefore a question of whether the models are actually smoothing away exactly what, from a public health point of view, is of most interest, i.e. what happens in the most extreme cases (of cold temperature). This might be of greater concern if the conclusions were based on a single model, but is lessened to some extent when using BMA, as the conclusions are less likely to be affected by the choice of model, which might be affected by the particular dataset being used.

Of course, this analysis has not shown that temperature is the causal agent in the changes in daily deaths, and it may be that temperature is acting as a surrogate for one or more pollutants, or some other time-varying meteorological variable. Further analysis would, therefore, also include the effects of short-term changes in air pollution. Due to the high levels of collinearity between pollutants and temperature, it may not be possible to isolate the individual effects of any one variable, that would require studies with different temperature/pollutant ranges and associations. However, by using the Bayesian model averaging approach used here, the effect on deaths of various temperature scenarios performed here could be extended to incorporate changing levels of pollutants.

6.3 Modelling daily multivariate pollutant data at multiple sites

In this chapter, daily multiple pollution levels measured at a small set of spatial locations were modelled with the aim of producing stable estimates of pollutant levels, with measures of uncertainty, for use in subsequent health analysis. Initially, the aim of this analysis was to produce estimates of pollution for periods of missing data for use in time-series studies of the effects of short-term changes in air pollution. The model produces estimates, together with measures of uncertainty, for each day, whether missing or not, incorporating data from a number of different monitoring sites. It also allows predictions of pollutant levels to be made for locations without monitoring sites, in a similar way to kriging (for further details, see Chapter 2). This enables pollution ‘maps’ to be produced (for example, Figure 5.5 (b)), with the possibility of identifying areas of high pollution. Maps of the associated levels of uncertainty (Figure 5.5 (c)) may also help indicate where new monitoring sites should be located in order to produce more accurate estimates.

In this analysis, the spatial component (for PM_{10}) was found to be much smaller than the temporal one. This suggests that the approach of simply taking the means of the daily readings from a number of sites, as has been done in many time-series studies of air pollution (for example, Schwartz and Dockery (1992a)), may not be inappropriate. However, this ignores the uncertainty in the pollution levels, and in the subsequent mean, which may lead to the confidence intervals for the RRs being too narrow. It may also be problematic if data from every site is missing, which generally leads to that day (and the affected lags) being excluded from the analysis. It

should also be noted that this finding may also be specific to this dataset, and in particular the topology of the study area and the small number of sites considered. The assumption of low spatial heterogeneity of PM₁₀ (or other pollutants) should therefore be carefully examined in any new study, before using a simple measure such as the daily mean.

6.4 Summary of the epidemiological findings in the case studies

The study of the long-term effects of SO₂ on respiratory mortality provides more evidence of the chronic effects of air pollution. Despite potential confounding, it lends support to current efforts to reduce levels of air pollution. It is important to have estimates of the chronic effects, in addition to those based on acute studies, as the latter may, to some extent, merely represent displacement by a short period of time, rather than being ‘new’ cases/deaths. It should be noted that the exposure data from the early parts of the study period relate to historically high levels of pollution and so the results may not apply to current levels.

The study of the effects of cold temperatures on mortality confirmed and qualified the importance of temperature on mortality, which may have been inadequately allowed for in previous studies concentrating on pollutant effects. The quantification of the effects of temperature has important policy implications, although there is still the possibility that temperature is acting as a proxy for some other time varying factor(s).

The study of pollution levels in London gives a better characterisation of pollutant exposure for estimation of potential health effects. In addition, maps of the resulting estimates of pollution can help in identifying areas of high pollution, and maps of the variability could be used to indicate where future monitoring sites should be located, in order to obtain more accurate estimates of pollutant levels.

6.5 Key statistical issues: Over-dispersion

Important statistical issues in the analysis of the effects of air pollution, temperature and health have been examined and methods have been proposed that acknowledge many of the difficulties in using routinely collected data within studies performed at the aggregate level. In common with the majority of ecological epidemiological studies which examine exposure-risk associations at the level of the group rather than the individual, there can be a number of inferential difficulties, an umbrella term for these difficulties being ecological bias. Such bias may arise, for example, where there is within-group variability in exposures/confounders, or between, and within, group confounding (see, for example, Greenland (1992)). In the absence of knowledge of the within-group variability, as will be the case in the presence of unmeasured confounders, the common

assumption that the outcome (e.g. daily deaths) follows a Poisson distribution becomes untenable in the presence of over-dispersion. Failure to allow for over-dispersion in any analysis can seriously affect the results, both in terms of the precision of risk estimates and in model comparison. Consequently, the conclusions from an epidemiological study may be affected. For example, very large levels of over-dispersion were observed in two of the applications presented in this thesis; dispersion parameters as high as nine were found in the study of the long-term effects of air pollution (Chapter 3) and in the region of five in the study of the effect of cold temperatures (Chapter 4). In these examples, the usual Poisson assumption that the variance is equal to the mean is clearly incorrect (in the first case, it is greater by a factor of nine). The resulting widening of the confidence intervals can determine whether a relative risk is deemed significant or not, and there are examples in both Chapter 3 and Chapter 4 where any conclusions based on a naive understanding of the RRs estimates and their associated CIs would change entirely.

The other important aspect of over-dispersion is in the comparison of models, which is generally achieved by examining differences in deviances. If no allowance is made for over-dispersion, then differences between models will be over-estimated, and the adjustment can affect whether or not the inclusion of a particular parameter is deemed to have a significant effect on the fit of the model, and thus warrant inclusion. These problems will become particularly acute when using automatic selection procedures, such as stepwise selection, which are based entirely on such tests (without allowance for over-dispersion), without any user intervention. If large over-dispersion is present, then the selection of variables in the model could be greatly affected, and so interpretation of the observed estimates should be made with caution.

The problems associated with over-dispersion have largely been ignored in Poisson based time-series studies of air pollution (for example, Schwartz et al. (1993), Schwartz and Dockery (1992a), Katsouyanni et al. (1995), Gouveia and Fletcher (2000)). This may lead to problems of bias in the estimation of effect sizes, due to errors in the choice of which lagged value of the pollutant is used, or which other variables are included in the model. It will also effect the precision of the estimates, and thus the width of the confidence intervals, possibly leading to incorrect conclusions being drawn. This problem is likely to be exacerbated when combining results from a number of studies in meta-analyses (e.g. Schwartz (1994a)). This is an important point, as is it from studies of this type that much of the policy of the effects of air pollution on health, and regulatory guidelines, are based.

6.6 Key statistical issues: Model selection

The traditional approach to model ‘selection’ (as reviewed in detail in Chapter 2) involves choosing a single model which is then considered to be correct when drawing inference about the parameters of interest or performing prediction. This ignores the inherent uncertainty associated with

the actual modelling process, in particular the uncertainty which arises through selecting a subset of variables for inclusion, together with that due to distributional assumptions and/or transformations. Bayesian model averaging provides a framework in which this uncertainty is acknowledged, and incorporated within the precision of the resulting parameter estimates. Despite the importance of allowing for model uncertainty and the existence of a general strategy for dealing with it, there are still obstacles to its implementation. In the examples shown here, the numbers of possible models have been limited for illustrative and computational reasons. Theoretically it would be possible, and may be very useful, to be able to average over models containing not just different variables, but also different structures. For example, it should be possible to consider a number of different threshold values or different transformations, although this would quickly result in a very large numbers of candidate models. Suggestions have been made for dealing with the large number of models, such as ‘Occam’s Window’ (Raftery et al. (1997)), which selects a subset of the models initially considered, according to how the posterior probabilities compare with that of the model with the highest posterior probability. The averaging is then done over a much smaller set of models.

There is a fundamental question as to whether it actually makes sense to perform BMA if it is considered that there is a ‘true model’. If any of the candidate models is actually the correct one, then (unless it has a posterior model probability of one) the resulting ‘averaged’ model cannot be true, and will not provide the investigator with a true picture of how the process is operating, in terms of the estimation of the parameters. In general, and certainly in observational studies of this type, it is unlikely that the true functional form of the model would ever actually be known, and even if it were, it would be near impossible to measure all of the possible variables that may have an effect and therefore merit inclusion in the ‘true’ model. BMA, in taking information from a wide set of possible explanations of the process, drawing information from all the data that are available, could be perceived as more reliable than drawing inference from a single model, despite the knowledge that inference will be drawn from a ‘fictional’ model which is unlikely to exist in reality.

This may be a problem when explaining the model in terms of estimation of the parameters, particularly in non-linear cases, such as in the Poisson case generally used in epidemiological studies, but this is of little concern when performing prediction, where it is the combined explanatory power of the variables on the response that is important rather than their individual effects. It may not be possible to use an epidemiological study to isolate the individual lagged effect of a particular pollutant or temperature, however it has successfully been shown that prediction can be used to estimate the number of deaths for different exposure scenarios (which could be based on levels of several pollutants in addition to temperature) and thus calculate the number of extra deaths that would be expected under given scenarios. However, investigators should not be lulled into a false sense of security simply by using BMA, as the initial class of models is important and

priors are not always easy to specify.

6.7 Further statistical issues and developments

As discussed, large levels of over-dispersion were observed in the regression models in both of the epidemiological studies undertaken in this thesis. In the study of the long-term effects of air pollution, the effect of over-dispersion on the precision of the risk estimates has been assessed using quasi-likelihood and by fitting log-normal random effects within a hierarchical model. Further investigation might consider the sensitivity to these choices, by considering alternative structures for the random effects, e.g. using a negative binomial distribution, which could be fit using both likelihood and Bayesian approaches. In addition, alternative error distributions could be used, for example a Student's t distribution.

Also in this analysis, two approaches were given to produce 'smoothed' relative risks for each of the health period-lag combinations. It may be possible to use stronger smoothing models, e.g. a second order AR process, to produce more stable estimates, although in this case, there are only four different values for time (health period) and lag. However, the structure of the model is general, and could easily be extended to incorporate data at a higher temporal resolution, in which case it may be feasible to fit such models. A possible third approach might be perform the smoothing over both time and lags. However, allowance must be made for the fact that the two scales (time and lag) are not the same. Methods for dealing with this problem have been suggested by Besag and Higdon (1999), in the context of agricultural field experiments, who also propose ways of dealing with edge effects, by extending the edges of the design, i.e. adding additional health period-lag combinations, which are then treated as missing values with a Bayesian framework.

A useful extension to the Bayesian model averaging approach would allow non-linear models, the simplest approach might be to allow series of models with different thresholds, such as those seen in Figure 4.5, when modelling the effects of the entire range of temperatures in Chapter 4. This would greatly increase the number of models under consideration, and significantly increase the computational burden. In addition, at present there is no allowance for models with over-dispersion in BMA. One approach would be to try and incorporate quasi-likelihood models in the method used here, i.e. using the approximation based on results from generalised linear models. Another would be to incorporate random effects, modelled as log-normal or Gamma, into a hierarchical model structure, as described in Rothman and Greenland (1998).

In addressing the question of how prolonged periods of cold affect mortality, an alternative approach to using continuous variables would be to create discrete variables that represent the

different lagged values of temperature, enabling cold temperatures on consecutive days to be explicitly represented. For example, daily indicator variables, $U_{l,k}$, $l = 0, \dots, L$, $k = 1, \dots, K$, could be created to represent the lagged daily temperatures, where L is the maximum lag being considered and K the number of temperature categories. The representation of consecutive days at certain temperatures or less, can then be represented by the product of the individual daily variables. If, for example, the ranges of temperatures and lags considered are the same as that presented in Table 4.3, i.e. temperatures from 5 ($k=1$) to -2 ($k=8$) and lags from zero ($l=0$, same day) to 7 days ($l=8$). In this case, three consecutive days at 2C would be represented by $U_{1,4} \times U_{2,4} \times U_{3,4}$. These are interaction terms within a regression model, which would be fitted together with the main effects (individual daily variables) and lower interaction terms. In the comparison of such models, it would be possible to assess the *incremental* effects of consecutive days as well as the total effect of a number of consecutive days compared to that of a single day. Such an approach has certain limitations in that discretising data often leads to a loss of information, although if small enough categories were used (as in this example, with one category for each $^{\circ}\text{C}$), the loss should be minimal and it could provide a relatively simple method for dealing with non-linear relationships. In terms of implementation however, the inclusion of all the possible interactions representing consecutive days at a certain temperature will lead to the number of possible models being very large, and the subsequent computational burden may prove prohibitive, if performed within the context of BMA.

In modelling daily multivariate pollutant data at multiple sites, the model used was relatively simple, particularly the isotropic, stationary spatial component; this simplicity was necessary because of the paucity of spatial information and the computational expense. In London such a structure may not be too poor an approximation due to the topology, but in other locations with a more irregular topology the model may well be inadequate. It may be possible to ease the computational burden using extensions of the algorithms used for state space models (Carter and Kohn (1994)). Alternatively, approximations in the spirit of Wilke and Cressie (1999) may be used. Flexibility may also be added using the approach of Schmidt and O'Hagan (2000), though the computational burden will be high. However, the merits of refinements such as these will be largely determined by the quality of data that are available.

A number of other forms of data that commonly arise in studies such as these, may be incorporated within this framework. For example, 6-day pollution monitors, hourly data, and local traffic information could be incorporated if available.

As illustrated, estimates (and measures of uncertainty) can be made for locations at unmeasured sites though the accuracy of these estimates will crucially depend on the number of covariates (such as type of monitor) that have been incorporated. The width and symmetry of the intervals will also depend on the assumption of a Gaussian random field model. It has been seen here that spatial variability is much smaller than the temporal variability. The simple time series

model used here may be extended to use more general, and stationary, AR processes, which have the advantage that the variance parameters would be directly comparable to the other variances of the other components in the model, but may compromise the model's ability to model strong dependencies (for a discussion see Besag and Kooperberg (1995)). The parameters of such models could also be made location-specific. In a similar approach to other authors, independent priors are assumed for ϕ and σ_m within the spatial model, however this is not realistic and further research is required to find a more appropriate form. One possibility is to take priors such that $\sigma_m^2 |\Sigma_m|$, a measure of overall variability, is constant across models.

When the health study that motivated the exposure modelling in Chapter 5 is performed, the health and exposure data could be modelled jointly. Strictly, from a Bayesian perspective, this is the correct approach, for example, it allows 'feedback' from the health data to inform the exposure modelling, but it could lead to the health-exposure relationship being compromised. If there were a problem with the exposure modelling, which, for example, may be due to the difficulties in estimating the spatial effects, it could distort the health-exposure relationship. Another drawback of this approach is that it is computationally intensive. An alternative approach is to plug in the exposure levels, using for example, the median of the predictive distributions of observed pollutant levels and the predictive distributions themselves for unobserved levels. However, such an approach ignores the variability and so will in general produce interval estimates that are too narrow. A refinement is to use an errors-in-variables approach in which the pollution estimates and standard errors are used to inform a measurement error model, such a model does not allow feedback between the exposure and exposure-health components, but does allow the uncertainty in the pollutants to be acknowledged.

Given confidence in the spatial model, it would be possible to combine health data at the individual or small-area level, with exposure interpolations from this model. In the case of small-area data, a summary measure of all the pollutants may be used at, for example, the population weighted centroid. Strictly, as described in Wakefield and Salway (2001), the appropriate ecological model would integrate over the entire area, but this is not possible in closed form.

6.8 Conclusion

In this thesis, methods have been successfully developed to accommodate a number of the issues discussed in analysing data of the type commonly encountered in epidemiological studies of the effects of air pollution on health. These include allowance for unmeasured (non-confounding) variables, and the resulting presence of over-dispersion, the problems of model selection and comparison and the calculation of reliable estimates of underlying levels of pollution. The work presented here will serve as the basis for further epidemiological and statistical research. It should

be possible to analyse the relative effects of both temperature and several pollutants within the Bayesian model averaging framework, and to incorporate the modelled estimates of the underlying levels of pollutants, together with their uncertainty. Additionally, the models should make allowance for the possibility of over-dispersion. The resulting risk estimates, and interval estimates, would therefore take into account several difference sources of variation, including the effect of measurement error, the possibility of unmeasured variables and between model uncertainty. Such models will ultimately lead to more realistic estimates of the effects of air pollution and temperature on health to inform public health policy in this area.

Bibliography

- Anderson, H., A. Ponce de Leon, M. Bland, J. Bower, and D. Strachan (1996). Air pollution and daily mortality in London: 1987-92. *British Medical Journal* 12, 554-669.
- Anderson, H., C. Spix, and S. Medina (1997). Air pollution and daily admissions for chronic obstructive pulmonary disease in 6 European cities: Results from the APHEA project. *European Respiratory Journal* 10, 1064-71.
- Atkinson, R., S. Bremer, H. Anderson, D. Strachan, J. Bland, and A. Ponce de Leon (1999). Short-term associations between emergency hospital admissions for respiratory and cardiovascular disease and outdoor air pollution in London. *Archives of Environmental Health* 54, 398-411.
- Aylin, P., S. Morris, J. Wakefield, A. Grossinho, L. Jarup, and P. Elliott (2001). Temperature, housing, deprivation and their relationship to excess winter mortality in Great Britain, 1986-1996. *International Journal of Epidemiology* 30(5), 1100-1108.
- Bandein-Roche, K., C. Hall, W. Stewart, and S. Zeger (1999). Modelling disease progression in terms of exposure history. *Statistics in Medicine* 18, 2899-2916.
- Barja-Aburto, V., D. Loomis, S. Bangdiwala, C. Shy, and R. Rascon-Pacheco (1997). Ozone, suspended particulates, and daily mortality in Mexico City. *American Journal of Epidemiology* 145, 258-68.
- Berger, J., V. De Oliveira, and B. Sanso (2001). Objective Bayesian analysis of spatially correlated data. *Journal of the American Statistical Association* 96, 1361-1374.
- Berry, G., J. Gilson, S. Holmes, H. Lewinshon, and S. Roach (1979). Asbestosis: A study of dose-response relationship in an asbestos textile factory. *British Journal of Industrial Medicine* 36, 98-112.
- Besag, J. and D. Higdon (1999). Bayesian analysis of agricultural field experiments. *Journal of the Royal Statistical Society, Series B* 61, 691-746.
- Besag, J. and C. Kooperberg (1995). On conditional and intrinsic auto-regressions. *Biometrics* 82, 733-746.

- Bishop, Y., S. Feinburg, and P. Holland (1975). *Discrete Multivariate Analysis: Theory and Practice*. MIT Press.
- Breslow, N. and D. Clayton (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* 88(421), 9–25.
- Breslow, N. and N. Day (1980). *Statistical Methods in Cancer Research, Volume 2- The Analysis of Cohort Studies*. Scientific Publications No. 82. Lyon: International Agency for Research on Cancer.
- Breslow, N., J. Lubin, P. Marek, and B. Langholz (1983). Multiplicative models and cohort analysis. *Journal of the American Statistical Association* 78(381), 1–12.
- Brown, P., P. Diggle, M. Lord, and P. Young (2001). Space-time calibration of radar-rainfall data. *Journal of the Royal Statistical Society, Series C* 50(2), 221 – 241.
- Brown, P., K. Karsen, G. Roberts, and S. Tonellato (2000). Blur-generated non-separable space-time models. *Journal Royal Statistical Society B* 62, 847–860.
- Brunekreef, B., D. Dockery, and M. Krzyzanowski (1995). Epidemiologic studies on short-term effects of low levels of major ambient air pollution components. *Environmental Health Perspectives* 103(Suppl 2), 3–13.
- Burnett, R., R. Dales, M. Raizenne, D. Krewski, P. Summers, G. Roberts, M. Raad-Young, T. Dann, and J. Brook (1994). Effects of low ambient levels of ozone and sulfates on the frequency of respiratory admissions to Ontario hospitals. *Environmental Research* 65, 172–94.
- Burton, P., L. Gurrin, and P. Sly (1998). Extending the simple linear regression model to account for correlated responses: an introduction to generalized estimating equations and multi-level modelling. *Statistics in Medicine* 11, 1825–1839.
- Carroll, R., R. Chen, T. Li, H. Newton, H. Schmiediche, H. Wang, and E. George (1997). Modeling ozone exposure in Harris County, Texas. *Journal of the American Statistical Association* 92, 392–413.
- Carroll, R., D. Ruppert, and L. Stefanski (1995). *Measurement Error in Nonlinear Models*. Chapman and Hall, London.
- Carstairs, V. and R. Morris (1989). Deprivation: explaining differences between mortality between Scotland and England. *British Medical Journal* 299, 886–889.
- Carter, C. and R. Kohn (1994). On Gibbs sampling for state space models. *Biometrika* 81, 541–53.
- Chambers, J. and T. Hastie (1993). *Statistical Models in S*. Chapman and Hall. London.

- Chatfield, C. (1995). Model uncertainty, data mining and statistical inference. *Journal of the Royal Statistical Society, Series A* 158, 419–466.
- Chatfield, C. (1996). *The analysis of time series An Introduction*. Chapman and Hall. London.
- Chatfield, C. (2000). *Time-series forecasting*. Chapman and Hall/CRC. London.
- Chatfield, C. and A. Collins (1980). *Introduction to multivariate analysis*. Chapman and Hall. London.
- Chestnut, L., J. Schwartz, D. Savitz, and C. Burchfiel (1991). Pulmonary function and ambient particulate matter: epidemiological evidence from NHANES I. *Archives of Environmental Health* 46, 135–144.
- Chinn, S., C. d. V. Florey, I. Baldwin, and M. Gorgol (1981). The relation of mortality in england and wales 1969-73 to measurements of air pollution. *Journal of Epidemiology and Community Health* 35, 174–179.
- Clayton, D. and M. Hills (1993). *Statistical Models in Epidemiology*. Oxford Scientific Publications.
- COMEAP (1995). *Committee on the Medical Effects of Air Pollutants. Asthma and Outdoor Air Pollution*. London, HMSO.
- COMEAP (2001). *Committee on the Medical Effects of Air Pollutants. Statement and report on long-term effects of particles on mortality*. London, HMSO.
- Cressie, N. (1986). Kriging nonstationary data. *Journal of the American Statistical Association* 81(396), 625–634.
- Cressie, N. (1993). *Statistics for Spatial Data, Revised edition*. John Wiley, New York.
- Cressie, N. (1997). Discussion of Carroll, R.J. et al. ‘Modeling ozone exposure in Harris county, Texas’. *Journal of the American Statistical Association* 92, 392–413.
- Cressie, N. and C. Wilke (1998). Discussion of Mardia et al., the kriged Kalman filter. *Test* 7, 257–263.
- Curwin, M. and T. Devis (1988). Winter mortality, temperature and influenza: has the relationship changed in recent years? *Population Trends* 54, 17–20.
- Daly, C. (1959). Air pollution and causes of death. *British Journal of Preventive and Social Medicine* 13, 14–27.
- Damian, D., P. Sampson, and P. Guttorp (2000). Bayesian estimation of semi-parametric non-stationary spatial covariance structures. *Environmetrics* 12(2), 161–178.

- DeGroot, M. (1986). *Probability and Statistics*. Addison-Wesley.
- Denison, D., B. Mallick, and A. Smith (1998). Automatic Bayesian curve fitting. *Journal of the Royal Statistical Society, Series B* 60(2), 333–350.
- DETR (1998). Review and assessment: Monitoring air quality. Technical Paper LAQM.TG1(00), Department of the Environment, Transport and the Regions, HMSO, London.
- Dewanji, A., M. Goddard, D. Krewski, and S. Moolgavkar (1999). Two stage model for carcinogenesis: number and size distributions of premalignant clones in longitudinal studies. *Mathematical Bioscience* 155, 1–12.
- Diggle, P. (1991). *Time Series, an Biostatistical Introduction*. Oxford University Press.
- Diggle, P., K.-Y. Liang, and S. Zeger (1994). *Analysis of longitudinal data*. Oxford Science.
- Diggle, P., J. Tawn, and R. Moyeed (1998). Model-based geostatistics. *Journal Royal Statistical Society C* 47(3), 299–350.
- Dockery, D. and C. I. Pope (1994). Acute respiratory effects of particulate air pollution. *Annual Review of Public Health* 15, 107–132.
- Dockery, D., C. I. Pope, X. Xiping, J. Spengler, J. Ware, M. Fay, B. Ferris, and F. Speizer (1993). An association between air pollution and mortality in six U.S. cities. *New England Journal of Medicine* 329, 1753–1759.
- Dockery, D. and J. Schwartz (1995). Air pollution and daily mortality the authors reply. *Epidemiology* 6, 213–214.
- Dockery, D., F. Speizer, D. Stram, J. Ware, J. Spengler, and B. Ferris (1989). Effects of inhalable particles on respiratory health of children. *American Review of Rrespiratory Disease* 139, 587–594.
- Dolk, H., G. Shaddick, P. Walls, C. Grundy, B. Thakrar, I. Kleinschmidt, and P. Elliott (1997). Cancer incidence near radio and television transmitters in Great Britain I. Sutton Coldfield transmitter. *American Journal of Epidemiology* 145, 1–9.
- Domininci, F., J. M. Samet, and S. Zeger (2000). Combining evidence on air pollution and daily mortality from the 20 largest cities: a hierarchical modelling strategy. *Journal of the Royal Statistical Society, Series A* 163(3), 263–302.
- Donaldson, G. (1997). Mortality relating to cold weather in elderly people in southeastern England, 1979-94. *British Medical Journal* 315, 1055–1056.

- Donaldson, G., S. Ermakov, Y. Komarov, C. MacDonald, , and W. Keatinge (1998). Winter mortality and cold stress in Yekaterinburg, Russia: interview survey. *British Medical Journal* 317, 978–982.
- Donaldson, G., V. Tchernjavskii, S. Ermakov, and W. Keating (1998). Winter mortality and cold stress in Yekaterinburg, Russia: interview survey. *British Medical Journal* 316, 514–518.
- Draper, D. (1995). Assessment and propagation of model uncertainty (with discussion). *Journal of the Royal Statistical Society, Series B* 57, 45–97.
- Draper, N., I. Guttman, and L. Lapczak (1979). Actual rejection levels in a certain stepwise test. *Communications in Statistics A8*, 99–105.
- Elliott, P., M. Hills, J. Beresford, I. Kleinschmidt, D. Jolley, S. Pattenden, L. Rodrigues, A. Westlake, and G. Rose (1992). Incidence of cancer of the larynx and lung near incinerators of waste solvents and oils in Great Britain. *Lancet* 339, 854–58.
- Elliott, P., G. Shaddick, I. Kleinschmidt, D. Jolley, P. Walls, J. Beresford, and C. Grundy (1996). Cancer incidence near municipal solid waste incinerators in Great Britain. *British Journal of Cancer* 73, 702–710.
- Enquselassie, F., A. Dobson, H. Alexander, and P. Steele (1993). Seasons, temperature and coronary disease. *International Journal of Epidemiology* 22(4), 632–6.
- Euler, G., D. Abbey, A. Magie, and J. Hodgkin (1997). Chronic obstructive pulmonary disease symptom effects of long-term cumulative exposure to ambient levels of total suspended particulates and sulfur dioxide in California Seventh-Day Adventist residents. *Archives of Environmental Health* 42, 213–222.
- Eurowinter Group (1997). Cold exposure and winter mortality from ischaemic heart disease, cerebrovascular disease, respiratory disease, and all causes in warm and cold regions of Europe. *Lancet* 349, 1341–46.
- Fahrmeir, L. and L. Knorr-Held (2000). Dynamic and semiparametric models. In M. Schimek (Ed.), *Smoothing and Regression: Approaches, Computation and Application*. Wiley. To be published.
- Fahrmeir, L. and G. Tutz (1994). *Multivariate statistical modelling based on generalized linear models*. Springer.
- Fairley, D. (1990). The relationship of daily mortality to suspended particulates in Santa Clara County, 1980–86. *Environmental Health Perspectives* 89, 159–68.
- Farr, W. (1840). Letter to the Registrar general. *2nd report of the Registrar general, 1838–1839*, HMSO.

- Farr, W. (1841). Letter to the Registrar General. *3rd report of the Registrar General, 1839-1840, HMSO*.
- Firket, J. (1936). Fog along the Meuse valley. *Trans Faraday Soc* 32, 1191–1194.
- Gamerman, D. (1997). *Markov Chain Monte Carlo: Stochastic simulation for Bayesian inference*. London: Chapman and Hall.
- Gamerman, D. and H. Migon (1993). Dynamic hierarchical models. *Journal Royal Statistical Society B* 55, 629–642.
- Gamerman, D. and A. Smith (1996). Bayesian analysis of longitudinal data studies. In J. Bernardo, J. Berger, A. Dawid, and A. Smith (Eds.), *Fifth Valencia international meeting on Bayesian statistics*, London. Oxford University Press.
- Gardiner, M. and M. Crawford (1969). Patterns of mortality in middle and early old age in the county boroughs of england and wales. *British Journal of Preventive and Social Medicine* 23, 133–140.
- Gelman, A., J. Carlin, H. Stern, and D. Rubin (1995). *Bayesian Data Analysis*. Chapman and Hall, New York.
- Gelman, A. and D. Rubin (1992). Inference from iterative simulation using multiple sequences. *Statistical Science* 7, 457–511.
- Gilks, W., S. Richardson, and D. Spiegelhalter (1995). *Markov Chain Monte Carlo in Practice*. Chapman and Hall, London.
- Goldstein, H. (1987). *Multilevel models in educational and social research*. Charles Griffin. London.
- Gore, A. and C. Shaddick (1958). Atmospheric pollution and mortality in the county of London. *British Journal of Preventative and Social Medicine* 12, 104–113.
- Gouveia, N. and T. Fletcher (2000). Respiratory diseases in children and outdoor air pollution in Sao Paulo, Brazil: a time series analysis. *Occupational and Environmental Medicine* 57, 477–483.
- Greenland, S. (1992). Divergent biases in ecologic and individual level studies. *Statistics in Medicine* 11, 1209–1223.
- Guttorp, P., W. Meiring, and P. Sampson (1994). A space-time analysis of ground-level ozone data. *Environmetrics* 5, 241–254.
- Hamilton, J. (1994). *Time series analysis*. Princeton University Press.

- Handcock, M. and J. Wallis (1994). An approach to statistical spatial-temporal modelling of meteorological fields (with discussion). *Journal of the American Statistical Association* 89, 368–390.
- Harvey, A. (1981). *Time series models*. Philip Allan. Oxford.
- Haslett, J. and A. Raftery (1989). Space-time modelling with long-memory dependence: assessing Ireland's wind power. *Applied Statistics* 38, 1–50.
- Hastie, T. and R. Tibshirani (1990). *Generalized Additive Models*. Chapman and Hall, London.
- Hocking, R. (1976). The analysis and selection of variables in linear regression. *Biometrics* 38, 1–49.
- Holland, D., V. De Oliveria, L. Cox, and R. Smith (2000). Estimation of regional trends of sulfur dioxide over the eastern united states. *Environmetrics* 11, 373–393.
- Kass, E. and L. Wasserman (1995). A reference bayesian test for nested hypothesis and its relationship to the Schwarz criterion. *Journal of the American Statistical Association* 90(431), 928–934.
- Kass, R. and A. Raftery (1995). Bayes factors. *Journal of the American Statistical Association* 90, 773–795.
- Katsouyanni, K. (1996). The APHEA project. Short term effects of air pollution on health: a European approach using epidemiological time series data. *Journal of Epidemiology and Public Health* 50 (Suppl 1).
- Katsouyanni, K., J. Schwartz, C. Spix, G. Touloumi, D. Zmirou, A. Zanobetti, B. Wojtniak, J. Wonk, A. Tobias, A. Ponka, S. Medina, L. Bacharova, and H. Anderson (1995). Short term effects of air pollution on health: a European approach using epidemiologic time series data: the APHEA protocol. *Journal of Epidemiology and Public Health* 50 (Suppl 1), S12–S18.
- Katsouyanni, K., G. Touloumi, C. Spix, J. Schwartz, F. Balducci, S. Medina, G. Rossi, B. Wojtniak, J. Sunyer, L. Bacharova, J. Schouten, A. Ponka, and H. Anderson (1997). Short term effects of ambient sulphur dioxide and particulate matter on mortality in 12 European cities: results from time series data from the APHEA project. *British Medical Journal* 314, 1658–1663.
- Keatinge, W., G. Donaldson, E. Cordoli, M. Martinelli, K. Kunst, J. Mackenbach, S. Nayha, and I. Vuori (2000). Heat related mortality in warm and cold regions of Europe: observational study. *British Medical Journal* 321, 670–673.
- Kelsall, J., J. Samet, S. Zeger, and J. Xu (1997). Air pollution and mortality in philadelphia, 1974–88. *American Journal of Epidemiology* 146(9), 750–762.

- Kinney, P. and H. Ozkaynak (1991). Associations of daily mortality and air pollution in los angeles county. *Environmental Research* 54, 99–120.
- Kleinschmidt, I., M. Hills, and P. Elliott (1995). Smoking behaviour can be predicted by neighbourhood deprivation measures. *Journal of Epidemiology and Community Health* 49 (Suppl 2), S72–7.
- Krewski, D., R. Burnett, M. Goldberg, K. Hoover, J. Siemiatycki, M. Jerret, M. Abrahamowicz, and W. White (2000). Reanalysis of the Harvard six cities study and the american cancer society study of particulate air pollution and mortality. Technical report, Health Effects Institute.
- Laird, N. and J. Ware (1982). Random-effects models for longitudinal data. *Biometrics* 38, 963–974.
- Le, N., W. Sun, and J. Zidek (1997). Bayesian multivariate spatial interpolation with data missing by design. *Journal of the Royal Statistical Society, Series B* 59, 501–510.
- Levy, J., J. Hammitt, and J. Spengler (2000). Estimating the mortality impacts of particulate matter: What can be learned from between-study variability. *Environmental Health Perspectives* 2, 109–117.
- Liang, K.-Y. and S. L. Zeger (1986). Longitudinal data analysis using generalized linear models. *Biometrika* 73, 13–22.
- Lippmann, M. and K. Ito (1995). Separating the effects of temperature and season on daily mortality from those of air pollution in London: 1965-71. *Inhalation Toxicology* 7, 85–97.
- Macfarlane, A. (1977). Daily mortality and environment in english conurbations 1: Air pollution, low temperature, and influenza in Greater London. *British Journal of Preventative and Social Medicine* 31, 54–61.
- Mackenbach, J., A. Kunst, and C. Looman (1992). Seasonal variation in mortality in the Netherlands. *Journal of Epidemiology and Community Health* 46, 261–265.
- Mackenbach, J., C. Looman, and A. Kunst (1993). Air pollution, lagged effects of temperature, and mortality: the Netherlands, 1979-87. *Journal of Epidemiology and Community Health* 47, 121–126.
- Mardia, K., C. Goodall, E. Redfern, and F. Alonso (1998). The kriged Kalman filter. *Test* 7, 217–276.
- Martin, A. and W. Bradley (1960). Mortality, fog, and atmospheric pollution: An investigation during the winter of 1958-59. *Monthly Bulletin of the Ministry of Health* 19, 56–72.

- Mazumdar, S., H. Schimmel, and I. Higgins (1982a). Relation of daily mortality to air pollution: an analysis of 14 London winters, 1958/59-1971/72. *Archives of Environmental Health* 37, 213-20.
- Mazumdar, S., H. Schimmel, and I. Higgins (1982b). Relation of daily mortality to air pollution: an analysis of 14 London winters, 1958/9-1971/72. *Archives of Environmental Health* 37, 213-220.
- McCullagh, P. and J. Nelder (1989). *Generalized Linear Models, 2nd Edition*. Chapman and Hall.
- Meinhold, R. and N. Singpurwalla (1983). Understanding the Kalman filter. *American Statistician* 37, 123-127.
- Miller, A. (1990). *Subset Selection in Regression*. Chapman and Hall.
- Ministry of Health (1954). *Mortality and morbidity during the London fog of December 1952*. London: HMSO.
- Moolgavkar, S., E. Luebeck, T. Hall, and E. Anderson (1995). Air pollution and daily mortality in Philadelphia. *Epidemiology* 6, 476-484.
- Nyberg, F., P. Gustavsson, L. Jarup, T. Bellender, N. Berglund, R. Jakobsson, and G. Pershagen (2000). Urban air pollution and lung cancer in Stockholm. *Epidemiology* 11, 487-95.
- O'Hagan, A. (1995). Fractional Bayes factors for model comparison. *Journal of the Royal Statistical Society, Series B* 57(1), 99-138.
- Omar, R., E. Wright, R. Turner, and S. Thompson (1999). Analysing repeated measurements data: a practical comparison of methods. *Statistics in Medicine* 18, 1587-1603.
- Onursal, B. and S. Gautan (1997). Vehicular air pollution. experiences from seven latin american urban centers. Technical Paper 373, World Bank, The World Bank, Washington, D.C.
- Ostro, B. (1993). The association of air pollution and mortality: examining the case for inference. *Archives of Environmental Health* 48, 336-42.
- Ostro, B., J. Sanchez, C. Aranda, and G. Eskeland (1996). Air pollution and mortality: results from a study of Santiago, Chile. *Journal of Exposure Analysis & Environmental Epidemiology* 6, 97-114.
- Pole, A., M. West, and J. Harrison (1994). *Bayesian Forecasting and Time Series Analysis*. Chapman and Hall.
- Pope, C. I., D. Dockery, and J. Schwartz (1995). Review of epidemiological evidence of health effects of particulate air pollution. *Inhalation Toxicology* 7, 1-18.

- Pope, C. I., J. Schwartz, and M. Ransom (1992). Daily mortality and PM₁₀ pollution in Utah Valley. *Archives of Environmental Health* 47, 211–217.
- Pope, C. I., M. Thun, M. Namboodiri, D. Dockery, J. Evans, F. Speizer, and C. Heath (1995). Particulate air pollution as a predictor of mortality in a prospective study of U.S. adults. *American Journal of Respiratory Critical Care Medicine* 151, 669–674.
- Raftery, A. and S. Richardson (1996). *Bayesian Biostatistics*, Chapter 12: Model Selection for Generalised Linear Models via GLIB: Application to Nutrition and Breast Cancer, pp. 321–353. Marcel-Dekker, New York.
- Raftery, E., D. Madigan, and J. Hoeting (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association* 92(437), 179–191.
- Rahlenbeck, S. and H. Kahl (1996). Air pollution and mortality in East Berlin during winters of 1981–1989. *International Journal of Epidemiology* 25(6), 1220–1225.
- Raizenne, M., L. Neas, A. Damokosh, D. Dockberg, J. Spengler, P. Koutrakis, J. Ware, and F. Speizer (1996). Health effects of acid aerosols on North American children: pulmonary function. *Environmental Health Perspectives* 104, 506–514.
- Richter, D. and W. Williams (1998). Assessment and management of urban air quality in europe. EEA Monograph 5, European Environmental Agency, EEA, Copenhagen.
- Rooney, C. and T. Devis (1996). Mortality trends by cause of death in England and Wales 1980–94: the impact of introducing automated cause coding and related changes in 1993. *Population Trends* 86, 29–35.
- Rothman, K. and S. Greenland (1998). *Modern Epidemiology*. Lippencott-Raven.
- Saez, M., J. Sunyer, A. Tobias, F. Ballester, and J. Anto (2000). Ishaemic heart disease mortality and weather temperature in Barcelona, Spain. *European Journal of Public Health* 10, 58–63.
- Samet, J. (1995). Particulate air pollution and mortality: the Philadelphia story. *Epidemiology* 6, 471–3.
- Samet, J., S. Zeger, F. Dominici, F. Curriro, I. Coursac, D. Dockery, J. Schwartz, and A. Zanobetti (2000). *The National Morbidity, Mortality, and Air pollution study Part II: Morbidity, Mortality, and Air pollution in the United States*. Cambridge, MA: Health Effects Institute.
- Sampson, P. and P. Guttorp (1992). Nonparametric estimation of nonstationary spatial covariance structure. *Journal of the American Statistical Association* 87, 108–119.

- Schmidt, A. M. and A. O'Hagan (2000). Bayesian inference for nonstationary spatial covariance structure via spatial deformations. Research report, Department of Probability and Statistics, University of Sheffield.
- Schwartz, J. (1989). Lung function and chronic exposure to air pollution: a cross-sectional analysis of NHANES II. *Environmental Research* 50, 309–321.
- Schwartz, J. (1991). Particulate air pollution and daily mortality in Detroit. *Environ. Res.* 56, 204–213.
- Schwartz, J. (1993). Air pollution and daily mortality in Birmingham, Alabama. *American Journal of Epidemiology* 137, 1136–1147.
- Schwartz, J. (1994a). Air pollution and daily mortality: a review and meta analysis. *Environmental Research* 64, 36–52.
- Schwartz, J. (1994b). Air pollution and hospital admissions for the elderly in Detroit. *American Journal of Respiratory Critical Care Medicine* 150, 648–655.
- Schwartz, J. (1994c). Nonparametric smoothing in the analysis of air pollution and respiratory health. *Canadian Journal of Statistics* 22(4), 471–487.
- Schwartz, J. (1994d). PM₁₀, ozone, and hospital admissions for the elderly in Minneapolis-St.Paul. *Archives of Environmental Health* 49, 366–374.
- Schwartz, J. (1994e). Total suspended particulate matter and daily mortality in Cincinnati, Ohio. *Environmental Health Perspectives* 102, 186–189.
- Schwartz, J. (1995). Short term fluctuations in air pollution and hospital admissions of the elderly for respiratory disease. *Thorax* 50, 531–538.
- Schwartz, J. (1997). Air pollution and hospital admissions. *Epidemiology* 8(4), 371–377.
- Schwartz, J. (2000). The distributed lag between air pollution and daily deaths. *Epidemiology* 11, 320–326.
- Schwartz, J. and D. Dockery (1992a). Increased mortality in Philadelphia associated with daily air pollution concentrations. *American Review of Respiratory Disease* 145, 600–604.
- Schwartz, J. and D. Dockery (1992b). Particulate air pollution and daily mortality in Steubenville, Ohio. *American Journal of Epidemiology* 135, 12–29.
- Schwartz, J., D. Dockery, and L. Neas (1996). Is daily mortality associated with fine particles ? *Journal of the Air and Waste Management Association* 46, 927–939.

- Schwartz, J. and A. Marcus (1990). Mortality and air pollution in London: A time series analysis. *American Journal of Epidemiology* 131, 185–194.
- Schwartz, J., D. Slater, T. Larson, W. Pierson, and J. Koenig (1993). Particulate air pollution and emergency hospital visits for asthma in Seattle. *American Rev Respir Dis* 147, 826–831.
- Sheppard, L., D. Levy, G. Norris, T. Larson, and J. Keonig (1999). Effects of ambient air pollution on nonelderly asthma hospital admissions in Seattle, Washington, 1987-1994. *Epidemiology* 10, 23–30.
- Shumway, R., R. Tai, and Y. Pawitan (1983). *Statistical analysis of daily London mortality and associated weather and pollution effects*. Sacramento, CA: California air resources board,.
- Silverman, B. W. (1986). *Density estimation*. London: Chapman and Hall.
- Smith, A. and G. Roberts (1993). Bayesian computation via the Gibbs sampler and other related Markov Chain Monte Carlo methods. *Journal of the Royal Statistical Society, Series B* 55, 3–23.
- Spiegelhalter, D., A. Thomas, and N. Best (1996). Computation on Bayesian graphical models. In J. Bernardo, J. Berger, A. Dawid, and A. Smith (Eds.), *Sixth Valencia international meeting on Bayesian statistics*, London. Oxford University Press.
- Spiegelhalter, D., A. Thomas, and N. Best (1998). WinBUGS user manual, version 1.1.1. Cambridge, UK.
- Spix, C., H. Anderson, and J. Schwartz (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, 54–57.
- Sullivan, L., K. Dukes, and E. Losina (1999). An introduction to heirarchical linear modelling. *Statistics in Medicine* 18, 855–888.
- Sun, L., J. Zidek, N. Le, and H. Ozkaynak (2000). Interpolating Vancouver’s daily ambient pm_{10} field. *Environmetrics* 11, 651–663.
- Sunyer, J., C. Spix, and P. Quenel (1997). Urban air pollution and emergency admissions for asthma in four European cities: The APHEA project. *Archives of Environmental Health* 52, 760–5.
- Thurston, G. (1996). A critical review of PM_{10} mortality time-series studies. *Journal of Exposure Analysis & Environmental Epidemiology* 6, 3–21.
- Thurston, G. and K. Ito (1999). *Epidemiological studies of ozone exposure effects*. San Diego: Academic Press.

- Thurston, G. and P. Kinney (1995). Air pollution epidemiology: considerations in time-series modelling. *Inhalation Toxicology* 7, 71–83.
- Tonellato, S. (2001). A multivariate time series model for the analysis and prediction of carbon monoxide atmospheric concentrations. *Journal of the Royal Statistical Society C* 50(2), 187–200.
- Touloumi, G., K. Katsouyanni, D. Zmirou, J. Schwartz, C. Spix, A. Ponce de Leon, A. Tobias, P. Quennel, D. Rabczenko, L. Bacharova, L. Bisanti, J. Vonk, and A. Ponka (1997). Short-term effects of ambient oxidant exposure on mortality: a combined analysis within the APHEA project. *American Journal of Epidemiology* 146(2), 177–185.
- Wakefield, J., N. Best, and L. Waller (2000). Bayesian approaches to disease mapping. In P. Elliott, J. Wakefield, N. Best, and D. Briggs (Eds.), *Spatial Epidemiology: Methods and Applications*, Oxford. Oxford University Press.
- Wakefield, J. and R. Salway (2001). A statistical framework for ecological and aggregate studies. *Journal of the Royal Statistical Society, Series A* 164, 119–137.
- Walters, S., R. Griffiths, and J. Ayres (1994). Temporal association between hospital admissions for asthma in Birmingham and ambient levels of sulphur dioxide and smoke. *Thorax* 49, 133–140.
- Walters, S., M. Phupinyokul, and J. Ayres (1995). Hospital admission rates for asthma and respiratory disease in the West Midlands: their relationship to air pollution levels. *Thorax* 50, 948–954.
- Ware, J., L. Thibodeau, F. Speizer, S. Colome, and B. Ferris (1981). Assessment of the health effects of atmospheric sulfur oxides and particulate matter. *Environmental Health Perspectives* 41, 255–276.
- Waternaux, C., N. Laird, and J. Ware (1989). Methods for analysis of longitudinal data: blood lead concentrations and cognitive development. *Journal of the American Statistical Association* 84, 33–41.
- West, M., J. Harrison, and H. Migon (1985). Dynamic generalized linear models and Bayesian forecasting. *Journal of the American Statistical Association* 80, 73–90.
- Wilke, C., L. Berliner, and N. Cressie (1998). Hierarchical Bayesian space-time models. *Environmental and Ecological Statistics* 5, 117–154.
- Wilke, C. and N. Cressie (1999). A dimension reduction approach to space-time Kalman filtering. *Biometrika* 86, 815–829.

- Wilson, R. and J. Spengler (1996). *Particles in our air: Concentrations and health effects*. Harvard University Press.
- Zanobetti, A., M. Wand, J. Schwartz, and L. Ryan (2000). Generalized additive distributed lag models: quantifying mortality displacement. *Biostatistics* 1(3), 279–292.
- Zeger, S. and K.-Y. Liang (1992). An overview for the analysis of longitudinal data. *Statistics in Medicine* 11, 1825–1839.
- Zeger, S., K.-Y. Liang, and P. Albert (1988). Models for longitudinal data: A generalized estimating approach. *Biometrika* 44(405), 1049–1060.
- Zeger, S., D. Thomas, F. Dominici, J. Samet, J. Schwartz, D. Dockery, and A. Cohen (2000). Exposure measurement error in time-series studies of air pollution: concepts and consequences. *Environmental Health Perspectives* 108, 419–426.
- Zmirou, D., J. Schwartz, M. Saez, A. Zanobetti, B. Wojtyniak, G. Touloumi, C. Spix, A. Ponce de Leon, Y. Le Moullec, L. Bacharova, J. Schouten, A. Ponka, and K. Kaysouyanni (1998). Time series analysis of air pollution and cause-specific mortality. *Epidemiology* 9, 495–503.