MULTI-CITY TIME SERIES ANALYSES OF AIR POLLUTION AND MORTALITY DATA USING GENERALIZED GEOADDITIVE MIXED MODELS

by
Lung-Chang Chien

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirement for the degree of Doctor of Public Health in the Department of Biostatistics, Gillings School of Global Public Health

Chapel Hill
2009

Approved by

Advisor: Shrikant I. Bangdiwala

Reader: John S. Preisser

Reader: Todd A. Schwartz

Reader: Mark A. Weaver

Reader: Jiu-Chiuan Chen

ABSTRACT

LUNG-CHANG CHIEN: Multi-city Time Series Analyses Of Air Pollution And Mortality Data Using Generalized Geoadditive Mixed Models
(Under the direction of Shrikant I. Bangdiwala)

Background Here we introduce the generalized geoadditive mixed model (GGAMM), a combination of generalized additive model and linear mixed model with unified model structure for more flexible applications, to alternatively examine the influence of air pollution to human health.

Methods Extant air pollution and mortality data came from the National Morbidity, Mortality, and Air Pollution Study for 15 U.S. cities in 1991-1995. The PM_{10} main model, distributed lag model and four co-pollutant models used the GGAMM approach to analyze the effect of PM_{10}, lag effects and co-pollutants on several mortalities, adjusting for day-of-week, calendar time and temperature.

Objectives First, the effects of PM_{10} on mortality are preliminarily examined; second, a jackknife-bootstrap method and a principal component analysis are proposed to handle potential convergence problems; third, some missing data imputation methods are evaluated in the GGAMM; fourth, the issues of multicollinearity and concurvity in our models are examined; fifth, comparisons of the GGAMM and 2-stage Bayesian hierarchical model are performed; sixth, three simulations are accomplished for investigating the influence of concurvity, multicollinearity and missing data imputation methods on estimates and smoothing functions.

Results First, the effects of PM_{10} on mortality are preliminarily examined; second, a jackknife-bootstrap method and a principal component analysis are proposed to handle potential convergence problems; third, some missing data imputation methods are evaluated in the GGAMM; fourth, the issues of multicollinearity and concurvity in our models are examined; fifth, comparisons of the GGAMM and 2-stage Bayesian hierarchical model are performed; sixth, three simulations are accomplished for investigating the influence of concurvity, multicollinearity and missing data imputation methods on estimates and smoothing functions.

Conclusions The GGAMM provides an integrate model structure to concern national average estimates, city-specific estimates, smoothing and spatial functions simultaneously. Geographical data can immediately be used in the GGAMM without being affected by missing data, and nation-level smoothing functions can be fitted well by enough valid observations from all cities. These properties are not offered by 2-stage Bayesian hierarchical models, and recommended by using spatio-temporal data.

ACKNOWLEDGEMENTS

The achievement of this dissertation should mainly be attributed to the professional guidance from my advisor, Dr. Shrikant Bangdiwala. His precise judgments and clarified thoughts helped me accomplish the statistical analysis and writing of the dissertation within two years. I also appreciate my master degree's advisor, Dr. Der-Shin Chang. Without his encouragement, the thought of studying abroad never appeared in my mind. In the years of studying in Chapel Hill, many professors, such as Dr. Michael Belyea, Dr. Todd Schwarz, Dr. Mark Weaver, Dr. Sandy Funk, Dr. Chao-Hsing Yeh, Dr. John Preisser and Dr. Jiu-Chiuan Chen, helped me so much in either studying or working. I additionally want to thank Dr. Thomas Kneib, whose unselfish sharing of his experience and knowledge of BayesX helped me significantly in this dissertation.

It's a good memory of studying in the UNC-CH as I made a lot of friends here. They shared my happiness and sadness in these years. When I need any help, they always display indispensible roles. I'll never forget our happy hours when playing volleyball, tennis and softball together, driving several hundreds of miles to watch major league baseball games, or having many great parties for versatile reasons. Lin Lin, Brother Fang, Hung-Chia "Dr. Ha-sue" Hsu, Tung-Cheng "Old Big" Lee, Wei-Ting "The God of Cookery" Lin, Pei-Jung Lin, Chia-Cheng Chen, Charles Liao, Ann Chou, June Cho, Anne Jessup, Susan Rasmussen, Leo Yurek, Yun Zhang, Jen Jen Chang, Anderson Lin, HyunJu Park and Jeongok Logan, I'll miss all of you forever.

Finally, I want to show my greatest appreciation to my parents. To be the first person who gets a doctoral degree in Chien's Family, I will devote the entire honor to my family and hometown.

TABLE OF CONTENTS

LIST OF TABLES……...ix
LIST OF FIGURES…….xi
LIST OF ABBREVIATIONS AND SYMBOLS……..xvi
Chapter 1 1
INTRODUCTION, LITERATURE REVIEW AND OBJECTIVES 1
1.1. Introduction 1
1.2. Literature review 5
1.2.1. Historical events 5
1.2.2. Time series studies 6
1.2.3. Other epidemiological study designs 14
1.2.4. Extended issues in air pollution and mortality studies 18
1.2.5. Spatial regression models 23
1.2.6. Generalized additive mixed models 26
1.2.7. Applications of the GAMM in air pollution and mortality studies 30
1.2.8. Applications of the GGAMM 33
1.3. Motivations and objectives 34
Chapter 2 37
METHODOLOGY 37
2.1. Model formulation and inference 37
2.2. Data collection 46
2.3. Statistical models 46
2.4. Jackknife-bootstrap approach 49
2.5. Multicollinearity and concurvity analysis 50
2.6. The extended distributed lag model and PCA adjusted estimates 51
2.7. Missing data analysis 54
2.8. Simulation 59

2.9. Model diagnostic methods 63
Chapter 3 64
RESULTS 64
3.1. Demographics 64
3.2. Spatial and temporal correlation of air pollutants 68
3.3. Case study of using the GGAMM by BayesX 73
3.4. Convergence problem in the GGAMM and the smoothing parameter λ in smoothing functions 83
3.5. The influence of the number of knots in smoothing functions 96
3.6. The application of jackknife-bootstrap approach 107
3.7. Smoothing functions 111
3.8. Spatial functions 118
3.9. GGAMM v.s. 2-stage Bayesian hierarchical model 123
3.10. Missing data imputation analysis 132
3.11. Extended distributed lag models 137
3.12. Multicollinearity and concurvity 148
3.13. Model diagnostics 152
3.14. Extended applications 160
Chapter 4 166
SIMULATION 166
4.1. Concurvity simulation 166
4.2. Multicollinearity simulation 168
4.3. Missing data imputation simulation 170
Chapter 5 175
DISCUSSION 175
5.1. Summary 175
5.2. Limitations 188
5.3. Future work 193
APPENDIX A Tables of estimates from starting values of smoothing parameters 196
APPENDIX B Tables of estiamtes from numbers of knots in smoothing functions 203

APPENDIX C Smoothing functions of time and temperature from different starting values of smoothing parameters, where (a, b)=($\lambda_{time,}$ λ_{tmean}).....210

APPENDIX D Smoothing functions of time and temperature from different numbers of knots, where (a, b)=($k_{time,}$ k_{tmean}).....222

APPENDIX E Spatial function maps from different starting values of λ_{spat}.....234

APPENDIX F Comparison plots of city-specific PM_{10} effects between the GGAMM and 2-stage Bayesian hierarchial model.....240

APPENDIX G Tables of parameter estimates for cardiovascular, pneumonia, and respiratory mortality in three age categories.....250

APPENDIX H Smoothing function plots of time and temperature and spatial function map in extended applications.....253

REFERENCE.....274

LIST OF TABLES

Table 3.1 Demographics of respiratory mortality, temperature and PM_{10} in 15 U.S. cities from 1991 to 1995 65

Table 3.2 Demographics of CO, NO_2, O_3, and SO_2 in 15 U.S. cities from 1991 to 1995 66

Table 3.3 Correlation coefficient matrix among 15 U.S. cities from 1991 to 1995 67

Table 3.4 The parameter estimates with corresponding estimated standard errors of fixed and random effects in six main models 74

Table 3.5 The adjusted estimates by applying the jackknife-bootstrap approach in model 1 111

Table 3.6 The main estimates with different λ_{spat}s from 10 to 15 in each model 120

Table 3.7 The comparison of estimates from GGAMMs and 2-stage Bayesian hierarchical models 125

Table 3.8 The cities with complete air pollutant data in 108 U.S. cities from 1987 to 2000 133

Table 3.9 The model-fitting results from complete case analysis (CCA), nearest neighbor imputation version 1 and version 2 (NNI1, NNI2) and multiple imputation (MI-MCMC) 135

Table 3.10 The eigenvalues of 7 principal components with corresponding distinguished and cumulative contribute of the total variability over PM_{10} and 1-day~6-day lag PM_{10} effect 138

Table 3.11 The eigenvectors of the first three principal components when condensing current and 1-day~6-day lag PM_{10} effect 139

Table 3.12 The estimated parameter estimates with corresponding estimated standard errors of fixed and random effects in the GGAMM with 3 principal components from current and 1-day~6-day lag PM_{10} effect 140

Table 3.13 The eigenvectors of the first three principal components when condensing current and 1-day~14-day lag PM_{10} effect 142

Table 3.14 The eigenvectors of the first six principal components when condensing

current and 1-day~14-day lag PM_{10} effect.....143

Table 3.15 The estimated parameter estimates with corresponding estimated standard errors of fixed and random effects in the GGAMM with 3 principal components from current and 1-day~6-day lag PM_{10} effect.....144

Table 3.16 The PCA-adjusted PM_{10} estimates in the first extended distributed lag model.....146

Table 3.17 The PCA-adjusted PM_{10} estimates in the second extended distributed lag model.....147

Table 3.18 The multicollinearity and concurvity level in model 1~ model 6.....150

Table 3.19 Goodness-of-fit test of Poisson distribution in 15 U.S. cities from 1991~1995 153

Table 4.1 The main estimates with different concurvity levels and corresponding convergence rate (CR) in the GGAMM.....167

Table 4.2 The main estimates with different multicollinearity levels and corresponding convergence rate (CR) in the GGAMM.....169

Table 4.3 The estimates $\hat{\beta}$, se($\hat{\beta}$), and se($\hat{b}$) of GGAMM with corresponding convergence rates (CR) when applying CCA, NNI1, NNI2 and MI in variable X_1 for missing rates 5%, 10%, 20%, 30%, 40% and 50%.....171

Table 4.4 The main estimates of GGAMM with corresponding convergence rates (CR) when applying CCA, NNI1, NNI2 and MI in variable X_2 for missing rates 5%, 10%, 20%, 30%, 40% and 50% 173

Table 5.1 Comparison of results across studies: estimated percent increase in mortality relative risk per 10 $\mu g/m^3$ increase in PM_{10}.....176

Table 5.2 Comparison of results across studies: estimated percent increase in mortality relative risk per 10 ppb increase in O_3.....177

Table 5.3 The number of days and elders' respiratory mortality average from 1991 to 1995 in 15 U.S. cities on thermal comfort index 183

LIST OF FIGURES

Figure 3.1 City-to-city correlation and separation distance among 15 U.S. cities from 1991 to 1995 69

Figure 3.2 Time trend plot of respiratory mortality, 24-hour average temperature, PM_{10} and co-pollutants over 15 U.S. cities from 1991 to 1995 70

Figure 3.3 Cross-correlation functions of air pollutants vs. temperature 71

Figure 3.4 Cross-correlation functions of co-pollutant vs. PM_{10} 72

Figure 3.5 Smoothing functions of calendar time and 24-hour average temperature and map of spatial effect for 15 U.S. cities from 1991 to 1995 in model 1 75

Figure 3.6 Smoothing functions of calendar time and 24-hour average temperature and map of spatial effect for 15 U.S. cities from 1991 to 1995 in model 2 76

Figure 3.7 Smoothing functions of calendar time and 24-hour average temperature and map of spatial effect for 15 U.S. cities from 1991 to 1995 in model 3 77

Figure 3.8 Smoothing functions of calendar time and 24-hour average temperature and map of spatial effect for 15 U.S. cities from 1991 to 1995 in model 4 78

Figure 3.9 Smoothing functions of calendar time and 24-hour average temperature and map of spatial effect for 15 U.S. cities from 1991 to 1995 in model 5 79

Figure 3.10 Smoothing functions of calendar time and 24-hour average temperature and map of spatial effect for 15 U.S. cities from 1991 to 1995 in model 6 80

Figure 3.11 The estimated PM_{10} fixed effect with corresponding estimated standard errors of fixed and random effect from 25 convergence results using different starting values of λ_{time} and λ_{tmean} in model 1 86

Figure 3.12 The estimated PM_{10} fixed effect with corresponding estimated standard errors of fixed and random effect from 9 convergence results using different starting values of λ_{time} and λ_{tmean} in model 2 87

Figure 3.13 The estimated 1-day lag PM_{10} fixed effect with corresponding estimated standard errors of fixed and random effect from 9 convergence results using different starting values of λ_{time} and λ_{tmean} in model 2 88

Figure 3.14 The estimated 2-day lag PM_{10} fixed effect with corresponding estimated

standard errors of fixed and random effect from 9 convergence results using different starting values of λ_{time} and λ_{tmean} in model 2.....89

Figure 3.15 The estimated PM_{10} fixed effect with corresponding estimated standard errors of fixed and random effect from 20 convergence results using different starting values of λ_{time} and λ_{tmean} in model 4.....90

Figure 3.16 The estimated NO_2 fixed effect with corresponding estimated standard errors of fixed and random effect from 20 convergence results using different starting values of λ_{time} and λ_{tmean} in model 4.....91

Figure 3.17 The estimated PM_{10} fixed effect with corresponding estimated standard errors of fixed and random effect from 19 convergence results using different starting values of λ_{time} and λ_{tmean} in model 5.....92

Figure 3.18 The estimated O_3 fixed effect with corresponding estimated standard errors of fixed and random effect from 20 convergence results using different starting values of λ_{time} and λ_{tmean} in model 5 93

Figure 3.19 The estimated PM_{10} fixed effect with corresponding estimated standard errors of fixed and random effect from 25 convergence results using different starting values of λ_{time} and λ_{tmean} in model 6.....94

Figure 3.20 The estimated SO_2 fixed effect with corresponding estimated standard errors of fixed and random effect from 25 convergence results using different starting values of λ_{time} and λ_{tmean} in model 6.....95

Figure 3.21 The estimated PM_{10} fixed effect with corresponding estimated standard errors of fixed and random effect from 20 convergence results using different number of knots in model 1 97

Figure 3.22 The estimated PM_{10} fixed effect with corresponding estimated standard errors of fixed and random effect from 7 convergence results using different number of knots in model 2.....98

Figure 3.23 The estimated 1-day lag PM_{10} fixed effect with corresponding estimated standard errors of fixed and random effect from 7 convergence results using different number of knots in model 2 99

Figure 3.24 The estimated 2-day lag PM_{10} fixed effect with corresponding estimated standard errors of fixed and random effect from 7 convergence results using different number of knots in model 2 100

Figure 3.25 The estimated PM_{10} fixed effect with corresponding estimated standard errors of fixed and random effect from 18 convergence results using different number of knots in model 4 101

Figure 3.26 The estimated NO_2 fixed effect with corresponding estimated standard errors of fixed and random effect from 18 convergence results using different number of knots in model 4 102

Figure 3.27 The estimated PM_{10} fixed effect with corresponding estimated standard errors of fixed and random effect from 22 convergence results using different number of knots in model 5 103

Figure 3.28 The estimated O_3 fixed effect with corresponding estimated standard errors of fixed and random effect from 22 convergence results using different number of knots in model 5 104

Figure 3.29 The estimated PM_{10} fixed effect with corresponding estimated standard errors of fixed and random effect from 24 convergence results using different number of knots in model 6 105

Figure 3.30 The estimated SO_2 fixed effect with corresponding estimated standard errors of fixed and random effect from 24 convergence results using different number of knots in model 6 106

Figure 3.31 The jackknife estimates in model 3 109

Figure 3.32 The adjusted estimates by applying the jackknife-bootstrap approach in model 3 110

Figure 3.33 The temperature splines of model 1 when (λ_{time}, λ_{tmean})= (11, 15) and (12, 12) 112

Figure 3.34 The non-smoothing splines and 95% CI of temperature in model 5, where (a, b)=(λ_{time}, λ_{tmean}) 113

Figure 3.35 Two regular temperature smoothers in model 1 and model 6 116

Figure 3.36 The temperature smoothers using k_{tmean}=5 with different ktimes in time smoother in model 1 117

Figure 3.37 The temperature smoothers in model 5 with (k_{time}, k_{tmean})=(21, 6), (21, 7) and (21, 9) 118

Figure 3.38 The spatial function maps from λ_{spat}=10 in model 2 and model 4 121

Figure 3.39 The spatial function maps from λ_{spat}=10 in model 5 and model 6 122

Figure 3.40 The city-specific PM_{10} effects of the GGAMM and 2-stage Bayesian hierarchical model in model 1 128

Figure 3.41 The smoothing function plots of time fitting by the GAM in each city in model 1 129

Figure 3.42 The smoothing function plots of temperature fitting by the GAM in each city in model 1 130

Figure 3.43 The city-level residual box plots of the GGAMM and 2-stage Bayesian hierarchical model in model 1 131

Figure 3.44 Smoothing functions of calendar time and 24-hour average temperature and map of spatial effect for 15 U.S. cities from 1991 to 1995 in the GGAMM with 3 principal components from current and 1-day~6-day lag PM_{10} effect 141

Figure 3.45 Smoothing functions of calendar time and 24-hour average temperature and map of spatial effect for 15 U.S. cities from 1991 to 1995 in the GGAMM with 3 principal components from current and 1-day~14-day lag PM_{10} effect 145

Figure 3.46 The city-specific and overall histogram plots of respiratory death count from 1991~1995 154

Figure 3.47 The Q-Q plots for residuals from model 1 to model 6 156

Figure 3.48 The boxplots of residuals in model 1 to model 6 156

Figure 3.49 The histogram plots with normal curves in model 1 to model 6 157

Figure 3.50 The residual plots from model 1 to model 6.....157

Figure 3.51 The city-specific prediction plot and observation plot in model 1 158

Figure 3.52 The level-1 unit (city) residual boxplots in model 1 to model 6 159

Figure 3.53 The percent increase of relative risk with 95% confidence interval in cardiovascular mortality in three age categories 161

Figure 3.54 The percent increase of relative risk with 95% confidence interval in pneumonia mortality in three age categories.....163

Figure 3.55 The percent increase of relative risk with 95% confidence interval in respiratory mortality in three age categories 164

LIST OF ABBREVIATIONS AND SYMBOLS

ASMSE	adjusted sample mean square error
CCA	case complete analysis
CCF	cross-correlation function
CI	confidence interval
CML	conditional marginal likelihood
CO	carbon oxide
CR	convergence rate
DPQL	double penalized quasi-likelihood
DOW	day-of-week variables
EB	empirical Bayes
FB	full Bayes
GAMM	generalized additive mixed models
GAM	generalized additive models
GGAMM	generalized geoadditive mixed models
GIS	geographic information system
GLMM	generalized linear mixed models
i.i.d.	independent and identical distributed
MAR	missing at random
MCAR	missing completely at random
MCMC	Markov chain Monte Carlo
MI-MCMC	multiple imputation with Markov chain Monte Carlo method

NAQQS	National Ambient Air Quality Standards
NMMAPS	National Mortality and Morbidity Air Pollution Study
NNI1	nearest neighbor imputation – version 1
NNI2	nearest neighbor imputation – version 2
NO_2	nitrogen dioxide
O_3	ozone
OR	odds ratio
PCA	principal component analysis
PRIN	principal component vector
REML	restricted maximum likelihood
$PM_{2.5}$	particulate matter of 2.5 micrometers or less
PM_{10}	particulate matter of 10 micrometers or less
ppb	parts per billion
RR	relative risk
SO_2	sulfur dioxide
$TIME$	calendar time
$TMEAN_t$	24-hour mean of temperature at day *t*
VCMM	varying-coefficient mixed models
Y_{ct}	number of deaths from respiratory disease in city *c* at day *t*
β	fixed effect parameter vector
$\hat{\beta}^{(-i)}$	estimated fixed effect parameter without the i^{th} observation
$\hat{\beta}^{b}$	jackknife estimate
$\hat{\beta}^{B}$	bootstrap estimate

b_c	random effect parameter vector
$\hat{\gamma}_j$	PCA-adjusted estimate of fixed effect, $j = 1, \dots, q$
$d(s, s')$	Euclidean distance between site s and site s'
$f_j(.)$	smoothing function of continuous covariate
$f_{spat}(.)$	spatial function
$F^*_{j,k}$	Fisher information matrix, $j, k = 1, \dots, p + 1$
$g(.)$	link function
K_j	precision matrix, $j = 1, \dots, k$
k_{time}	number of knots in time smoothers
k_{tmean}	number of knots in temperature smoothers
l_{ij}	loading of the j^{th} variable in the i^{th} principal component
λ_{time}	smoothing parameter of time smoothers
λ_{tmean}	smoothing parameter of temperature smoothers
λ_{spat}	smoothing parameter of spatial functions
$\mu g/m^3$	microgram per cubic meter
$\hat{\xi}_{kj}$	PCA-adjusted estimate of random effect, $j = 1, \dots, q;\ k = 1, \dots, n$
s_c	$(long, lati)_c$, the longitude and latitude in city c
τ_j^2	variance parameter, $j = 1, \dots, k$
$trun\{.\}$	truncation operator
p_i	proportion of variation of the i^{th} principal component
PM_{ct}	PM_{10} concentration in city c at day t
PM_{ct}^{lagn}	n-day lag PM_{10} concentration in city c at day t
V_{ct}^P	concentration of co-pollutant; P denotes CO, NO_2, O_3 and SO_2

$\omega_{ss'}$	weight inverse proportional to the distance of centroids

Chapter 1
INTRODUCTION, LITERATURE REVIEW AND OBJECTIVES

1.1. Introduction

Air pollution has existed since the first fire was lit, but it is increasing in modern era. There are two main factors contributing to this serious problem around the world. First, the world population is increasing exponentially, especially in urban areas; second, the rapid growth of industries and affluent automobiles have led to a surge in the levels of fossil fuel combustion in developed and developing countries (Nadakavukaren, 2006).

The U.S. Congress passed the Air Pollution Control Act in 1955, the Air Pollution Control Amendments in 1960, the Clean Air Act in 1963, the Air Quality Act in 1967, the Clean Air Act Extension in 1970, and Clean Air Act Amendments in 1990 and 1997. The acts aimed to regulate air pollutant emission and related studies started to analyze air pollution elements, sources, and, most importantly, their influence on human health. Aerologists, environmentalists, ecologists, epidemiologists, and statisticians are collaborating to discover unknown factors that impact health in air pollutants.

Beginning with the London Fog in 1952, surveys, ecological studies, and early time-series studies began addressing air pollution issues in the 1950s and 1960s. Exposure assessment has evolved from the 1970s, and offers informative evidence to statisticians in analyzing air quality data. More significantly, an influential experiment about the health effects of air pollutants in America—the Six Cities Study—started in 1973. The study drew upon 8,000 individuals in six U.S. cities, consistently monitoring health effects and gathering air quality data with sophisticated laboratory equipment and

techniques. The study spanned across 2 decades, establishing an initial standard for related research in air pollution around the world (Dockery, Pope, Xu, Spengler, Ware, Fay, Ferris, & Speizer, 1993).

The U.S. government funds many organizations, allowing them to collect time series data, which offers a complete database for scientists to analyze air pollution issues. A well-known example, the National Morbidity and Mortality Air Pollution Study (NMMAPS), is one of the greatest projects concerning the influence of air pollution to human health in the U.S. It was led by the Department of Biostatistics at the Johns Hopkins Bloomberg School of Public Health, and is still maintained by the Internet-based Health & Air Pollution Surveillance System (iHAPSS), which is funded by the Health Effects Institute (HEI). This is arguably the most organized database as it includes numerous longitudinal mortality and morbidity sources, air pollution monitoring data, and weather condition records in 108 U.S. metropolitan areas from 1987 to 2000. In the NMMAPS, air pollution data is obtained from the AirData database collected by the U.S. Environmental Protection Agency. Daily mortality counts are retrieved from the National Center for Health Statistics, weather data is collected from the National Climatic Data Center, and census data is collected from the U.S. Census Bureau.

As modern statistical tools—especially time series and longitudinal data analysis and computers—were quickly developed in ‘80s and ‘90s, many modern time series studies (Schwartz & Marcus, 1990) and multi-city studies (Daniels, Dominici, Samet, & Zeger, 2000; Dominici, McDermott, Zeger, & Samet, 2003a; Dominici, McDermott, Zeger, & Samet, 2003b; Dominici, McDermott, Daniels, Zeger, & Samet, 2005) of air pollution and human health were published in succession. A third method for the design of air pollution-mortality studies is the case-crossover design. The idea is that the exposure of an individual exposed to an air pollutant immediately prior to some event, such as death, heart attack or stroke, is comparable the exposure of the same individual with the same air

pollutant during control times (Smith, 2007). A case-crossover analysis of particulate matter air pollution and out-of-hospital primary cardiac arrest was conducted in the Seattle region (Levy, Sheppard, Checkoway, Kaufman, Lumley, Koenig, & Siscovick, 2001b). The first two methodologies have represented the main stream of air pollution research since the 1990s.

In the applied statistical world dominated by linear models, more and more researchers discovered that their research purposes can hardly be satisfied by linear models due to not only some special data structures, but also unsatisfying the strict assumptions of model fitting. Hence, the frequency of using semiparametric models is becoming increasingly popular, and areas of application are more widespread than ever before. Semiparametric models have been developed, expanded, and applied to many areas of research and practical problems. Due to the flexibility semiparametric models offer, researchers have tried to adjust the model form to match with some special situations. The development of the generalized additive model (GAM) is an initial adaptation from general semiparametric models. Hastie and Tibshirani (1990) showed a somewhat complete introduction of the transition from additive model to generalized additive model with detailed statistical inference and extensions to other settings, such as the proportional-hazards model, proportional odds model, and seasonal decomposition of time series.

At the same time, Breslow and Clayton (1993) developed the generalized linear mixed model (GLMM), which has dealt with many problems, especially in repeated measurements. However, in application, real data easily violates the assumptions of GLMM, such as non-Gaussian longitudinal data. In addition, the linear assumption may not always be satisfied, so the necessity of a nonlinear approach is increasing as more and more studies require one. Although Wild and Yee (1996) and Berhane and Tibshirani (1998) applied the generalized estimating equations approach (Liang & Zeger, 1986) to

the GAM, a solid framework of dealing with non-Gaussian data and nonlinear association was still unsolved. Lin and Zhang (1999) combined both structures of the GLMM and GAM to create an initial form of the generalized additive mixed model (GAMM) and its statistical inference. They also adapted the spirit of the penalized quasi-likelihood approach of Brewslow and Clayton (1993) to make approximate inferences by the double penalized quasi-likelihood (DPQL) approach. Moreover, a general Bayesian approach via MCMC sampling for inference in the GAMM with structured or unstructured random effects and spatial covariates was presented by Fahrmeir and Lang (2001a), to solve bias problems from binary data and correlated random effects. This Bayesian approach was also applied in multi-categorical response variables in time-space data (Fahrmeir & Lang, 2001b). Besides using full Bayes posterior, the empirical Bayes posterior was later introduced in generalized additive mixed models with penalized splines for space-time data to offer more computationally efficient solutions (Fahrmeir, Kneib, & Lang, 2004). The GAMM can be improved with a spatial function, and was given a new name as the generalized geoadditive mixed model (GGAMM). The boundary data, coordinate data (longitude and latitude), centroid data, and kringing data can all be supported by the spatial function in the GGAMM. The geographical variation can be entirely performed by the spatial function in the GGAMM. In brief, the GAMM/GGAMM is a semiparametric model which can consider non-Gaussian data, linear factors, nonlinear smoothing functions, and spatial function simultaneously. It opened the bottleneck of the methodology of semiparametric modeling at the end of the last decade, and facilitated other researchers extending the theorems to more complicated cases, such as spatial analysis.

The research group of the NMMAPS project primarily used the GAM approach as the preliminary application of time series regression analysis to a single city across a period of time. In order to solve the problems of organizing different model-fitting results

across many cities, researchers have taken a Bayesian approach to combine versatile estimation with the multivariate normal hierarchical model (Everson & Morris, 2000). They named the combination of the generalized additive model and the multivariate normal hierarchical model the 2-stage Bayesian hierarchical model. This process considers not only linear predictors of air pollutants, such as particulate matter (PM_{10} or $PM_{2.5}$) or ozone, but also non-linear smoothing functions of time and weather conditions. Even though the concept is innovative, it is limited by the fact that no overall model structure can easily explain the two-stage Bayesian hierarchical model. The GAMM/GGAMM approach provides a unified model structure that fixed effects and random effects can represent nationwide average effects and marginal county-specific average effects, respectively. The general focus of this research is on how a novel statistical methodology, the GGAMM approach, can be applied in air pollution and human health time series studies. Based on the study design, it is necessary to look for more intuitive methods to describe the relationship of air pollution to human health with clearer interpretations and more advanced applications. More details are shown in the research motivations and objectives section.

1.2. Literature review

1.2.1. Historical events

For centuries, people have gradually become conscious of the influences of air pollution on adverse health effects in response to several earlier dramatic episodes of severe air pollution invasion. These occurrences include Meuse Valley, Belgium in 1930 (Firket, 1936; Fahrmeir, Kneib, & Lang, 2001), Donora, Pennsylvania in 1948 (Ciocco & Thompson, 1961; Davis, 2002; Fahrmeir, Kneib, & Lang, 1949), and London, England in 1952 and 1962 (Brimblecombe, 1987; Logan, 1953). The episodes caused a sudden surge

of illness and death. For example, the four-day fog in 1952 killed approximately 4,000 Londoners (David, 1994). Such events can be ascribed to human activities or industrial development. Some meteorological phenomena appear in another type of air pollution, and also have a severe impact on human beings. Sandstorms or dust storms have caused substantial damage in arid and semi-arid areas, such as Tucson, Arizona in 1971 and Melbourne, Australia in 1983. This meteorological phenomenon occurs regularly in East Asia and Africa, and their frequency is increasing each year. This can be regarded as a branch of air pollution research, and researchers also point out the short-term effects of dust storms on human health, especially in acute diseases (Chang, Hwang, Chan, Wang, & Cheng, 2007; Kwon, Cho, Chun, Lagarde, & Pershagen, 2002).

The improvement of statistical analysis on epidemiological data also provides associated studies with more precision, coherency, and consistency based on different targets (acute or chronic disease), exposures (short-term or lifetime), and study designs (longitudinal study or panel study). Moreover, studies attempting to identify the unique effect of a specific air pollutant are often controlled for many confounding factors and adjusted for simultaneous exposure to a complicated mixture of co-pollutants (Dominici et al., 2003a). In order to consider multiple side-effects in model fitting, advanced statistical methods like Cox proportional hazard models (Cox & Oakes, 1984), generalized linear models for count data and binary time series data (Liang & Zeger, 1986; McCullagh & Nelder, 1989), generalized additive models (Hastie & Tibshirani, 1990), and Bayesian hierarchical models (Lindley & Smith, 1972; Morris & Normand, 1992) have been broadly applied in public health areas and epidemiological issues.

1.2.2. Time series studies

Time series studies can be regarded as the most powerful methodology for

identifying the relationship between time-varying air pollution exposures and time-varying mortality or morbidity counts. Other confounders, such as weather conditions (ex: temperature and humidity) or time variations (ex: day of week and season), can also be included in time series analysis. The discussions about the quantitative method for effects of air pollution, the change of mortality after change of exposure, relationship between age-specific mortality and life expectancy, and association between life expectancy change and mortality after intervention are discussed in Rabl's paper (2006). Most research has focused on mortality or morbidity because that data was collected by related government organizations day by day, corresponding to daily measurements from air pollution monitoring stations. Along with geographical information, it is also available to use time and spatial data to access the within and between subject-specific influences and differences in exposures (Bell, McDermott, Zeger, Samet, & Dominici, 2004a).

The time factor is an important confounder in time series data. Due to its variation by, for example, seasonality, it is explicit that there is almost no purely linear relationship to mortality or morbidity. Hence, one needs to consider using the concept of the smoothing function to adopt a time factor into air pollution analyses. Two statistical tools are frequently used: generalized linear models (GLM) with parametric splines (e.g. natural cubic splines) (McCullagh & Nelder, 1989) and generalized additive models (GAM) with non-parametric splines, (e.g. smoothing splines) (Hastie & Tibshirani, 1990). The GAM is often used in air pollution and mortality studies because its smoothing functions are more flexible than the fully parametric fitting of the GLM (Dominici, Sheppard, & Clyde, 2003c). The parameter of interest, β, in the GAM for mortality counts and air pollution research is usually interpreted as the percentage increase in mortality per 10-unit or 100-unit increase in ambient air pollution levels (e.g. 10 $\mu g/m^3$ in PM_{10} or $PM_{2.5}$). Because of the flexibility of the GAM, both linear and smoothing functions can be easily

included into the analysis. Investigators have already concluded that the following potential confounding effects are important to include in the GAM: (a) the exposure of co-pollutants, (b) weather variations, and (c) long-term (season) and short-term (day of week) time trends (Bell, Samet, & Dominici, 2004b).

Preliminary time series studies in epidemiology on the health effects of air pollution focused mainly on specific events or incidents which often had severe air pollutant invasion over the course of a few days, and caused increased mortality and morbidity (Davis, 2002). The London Fog Event (Bell & Davis, 2001), the Donora Death Fog (Davis, 2002) and the Meuse Valley Fog Disaster (Firket, 1936; Roholm, 1937) are examples of using severe air pollution episodes to examine the health influence of high air pollutant concentrations over several days. Time series analysis in air pollution episodes is an important index to evaluate the impact of highly dense air pollution in a short time; however, their results were not generalized to make conclusions and refer to other non-episodic periods.

As more and more discoveries about air pollution and human health studies were made, researchers learned the necessity of addressing potential confounders to long-term analysis. But, early time series studies did not have overwhelming results due to the lack of suitable statistical models and restricted computational tools. Therefore, most of them only consider the population in a single location to discuss issues based on similar population patterns (Bell et al., 2004b). Fairley (1990) used the Poisson regression model to explore the association between suspended particulate matters and daily mortality in Santa Clara County, CA, from 1980 to 1986, and observed lower influence for health risk at higher concentrations of coefficient of haze (COH) compared with previous studies. Another similar research involving the Steubenville, Ohio, metropolitan area from 1974 to 1984, showed a roughly 4% increase in mortality on the succeeding day along with an increase of 100 $\mu g/m^3$ in particulate matters. However, after controlling for particulate

matters, it was not statistically significant in its association with sulfur dioxide (Schwartz & Dockery, 1992). Moreover, due to its geography, the floor of Utah Valley easily accumulates air pollutants when temperature drops down, especially in the winter season, which can cause high concentrations of PM_{10}. Pope et al. (1992) also used the Poisson regression model with a 5-day moving average to evaluate related data from April 1985 through December 1989. They found the 5-day moving average of PM_{10} concentration has the strongest association. In particular, a 100 $\mu g/m^3$ increase of PM_{10} can cause an increase in death of about 16%, especially in respiratory and cardiovascular diseases. This study was noteworthy as it not only utilized the concept of multiple-day moving averages (up to 7 days), but also simultaneously considered the absence of co-pollutants SO_2 and O_3 (Bell et al., 2004a). Other single location time series studies are the St. Louis and Kingston study (Dockery, Schwartz, & Spengler, 1992), Birmingham study (Schwartz, 1993) and Detroit study (Schwartz, 1991).

While these studies have consistent results, none of them are generally representative due to the heteroskedasticity among locations and consistent research periods. A weighted average approach can roughly combine those results (Ostro, 1993). By organizing those analyses, it was concluded that the percentages of change in the daily mortality for each 10 $\mu g/m^3$ increase in PM_{10} are 1.0%, 3.4%, and 1.4% for total mortality, respiratory disease, and cardiovascular disease, respectively (Bell et al., 2004a). However, this is not a recommended method, especially since more advanced models have appeared.

The criticisms of single location time series did not stop because of a lack of evidence to support the representation of their choices, or ignorance of spatial heterogeneity from either air pollutants or other confounders (Lipfert & Wyzga, 1993; Li & Roth, 1995; Dominici et al., 2003c). In order to improve the deficiencies of single location time series studies, multi-city study design offers a powerful alternative. A European study—Air Pollution and Health: a European Approach (APHEA)—is a

preliminary attempt to merge air pollution and all cause mortality data from 12 European cities to carry out a combined quantitative analysis (Katsouyanni, Touloumi, Spix, Balducci, Medina, Rossi, Wojtyniak, Sunyer, Bacharova, Schouten, Ponka, & Anderson, 1997), but a more concrete method was not presented until 2000 in the report from the National Morbidity, Mortality, and Air Pollution Study (NMMAPS) (Samet, Dominici, Curriero, Coursac, & Zeger, 2000a; Samet, Zeger, Dominici, Dockery, & Schwartz, 2000b). These studies showed a 2-stage Bayesian hierarchical model to manipulate multi-city data, and handled some unsolved problems from single location time series studies. The basic concept was to use the GAM to fit a relevant city-specific model for each city, and then apply a Bayesian method to merge all coefficients of interest, i.e. city-specific air pollution effects, into a nationwide air pollution effect based on some priors. Additionally, they also used the Reversible Jump Markov Chain Monte Carlo (RJMCMC) (Green, 1995) as the criterion for picking knots in smoothing functions (Dominici, Daniels, Zeger, & Samet, 2002a).

Today, this methodology has become the standard for speculating air pollution's impact to human health. Along with the well-organized NMMAPS online database, anyone can conveniently download it from the Internet, which facilitates the application of 2-stage Bayesian hierarchical models to more air pollution studies. An extension of the GAM is the distributed lag model, which is often used to handle the possible lag effects in the air pollution effect to mortality. The distributed lag effect was presented as early as 1965 for capital appropriations and expenditures (Almon, 1965), but in recent years, more and more analyses are using the same concept in air pollution and mortality studies with the GAM (Zanobetti, Wand, Schwartz, & Ryan, 2000; Zanobetti, Schwartz, Samoli, Gryparis, Touloumi, Atkinson, Le Tertre, Bobros, Celko, Goren, Forsberg, Michelozzi, Rabczenko, Ruiz, & Katsouyanni, 2002). Because the entire model structure does not fundamentally change but adds more variables for lag effects, it is also available to

transform to the 2-stage Bayesian hierarchical model for multi-city time series studies.

The number of cities selected in multi-city time series studies varies from 4 to 109 in the U.S. In the 4 U.S. city study, the authors did not use a 2-stage Bayesian hierarchical model, but applied a fixed effect model with weights $W^c=(V^c)^{-1}$ or $W^c=(D+V^c)^{-1}$ to combine the estimated coefficients across cities, where V^c is the covariance matrix in each city, and D is a diagonal between-city covariance matrix (Dominici et al., 2003a). In addition, they also used different kinds of discrete Fourier decompositions (Bloomfield, 1976; Priestley, 1981) for time factors from a high frequency component (less than 3 days) to a low frequency component (more than 2 months). Compared with the logarithm of relative risks among different timescales, they found the lower frequency component (i.e. longer timescale) has a greater effect which reflects a greater biologic impact on chronic exposures than on acute exposures.

The extension to the 20 U.S. city study from 1987 to 1994 considered other possible co-pollutants (O_3, NO_2, SO_2, CO) besides PM_{10} and some specific diseases (Samet et al., 2000a). Researchers used a 2-stage log-linear regression model, fitting a separate log-linear regression of the daily mortality on air pollutant measurements for each city, and pooled all estimates of the relative mortality rates associated with specific air pollutants by a Bayesian statistical approach (Gelman, Carlin, Stern, & Rubin, 1995). In this study, the increase of the estimated relative rate of all death is 0.51% (95% CI: 0.07%, 0.93%) per 10 $\mu g/m^3$ increase in PM_{10}, and slightly rises to 0.68% (95% CI: 0.20%, 1.16%) for cardiovascular and respiratory disease. Besides PM_{10}, the univariate effect of ozone levels during a one-year period was also examined, but strong evidence was not found when ozone levels were the highest during the summer.

A similar case with 20 U.S. cities, 19 of which were included in Samet et al. (2000a) is discussed by Dominici, Samet and Zeger (2000a). Dominici also used a 2-stage Bayesian hierarchical model to analyze the data. The main difference between Dominici

et al. (2000a) and Samet et al.(2000a) is that Dominici and her colleagues used a separate GAM with five smoothing functions in the first stage, and provided a preliminary framework of 2-stage Bayesian hierarchical model that has often been used in following years. Two air pollutants (PM_{10} and O_3) and two lag effects (1-day lag and 2-day lag) were considered in this study. When only PM_{10} was included in the analysis, a 10 $\mu g/m^3$ increase was associated with 0.48% increase in mortality. When adjusted by O_3, this association was slightly increased to 0.52%.

Some multi-city time series studies in ozone have pointed out that ozone has a positive association with mortality. For example, a 95 U.S. city study confirmed statistically significant results that a 10-ppb increase in ozone in the previous week, can cause a 0.52% (95% CI: 0.27%, 0.77%) and a 0.64% (95% CI: 0.31%, 0.98%) increase in daily non-injury-related mortality and cardiovascular/respiratory mortality, respectively the following week (Bell et al., 2004a). A more completed meta-analysis of time series studies of ozone and mortality can be found in Bell, Dominici and Samet (2005).

As related theories matured, researchers not only considered increasing locations in a study, but also used more hierarchical levels to build statistical models. The idea of geographic regions was being implemented to explore their heterogeneity in order to estimate regional air pollution mortality dose-response curves (Dominici et al., 2002a). 88 cities were divided into seven geographic regions (Northwest, Upper Midwest, Industrial Midwest, Northeast, Southern California, Southwest, and Southeast). Although detailed about how the cities were grouped is unknown, the criterion of these geographic divisions is still followed by other studies using the NMMAPS database. It was no surprise that a positive relationship between PM_{10} concentration and total mortality with a 0.5% increase for a 10 $\mu g/m^3$ increase in PM_{10} was found. More innovative discoveries are the differences between regional-adjusted and regional-unadjusted results. Researchers also found the strongest adverse effect of PM_{10} appeared in Northeast region, where the

increase of mortality was double compared to other regions. Moreover, a hierarchical spline model was also constructed to investigate the nonlinear relationship between PM_{10} and mortality by replacing the linear term of PM_{10} with a smoothing function of PM_{10}. The purpose of using the hierarchical spline model was to derive PM_{10} mortality dose-response curves. The reversible jump Markov chain Monte Carlo (RJMCMC) was also used to pool seven regional dose-response curves into a national dose-response curve. The same study design for exploring geographical variation was also showed in Dominici et al. (2003b).

A seasonal analysis using 100 U.S. cities from 1987 to 2000 provided another view point for accessing the variation of PM_{10} in different seasons (Peng, Dominici, Pastor-Barriuso, Zeger, & Samet, 2005). In previous studies, whether with single location or multi-city, all coefficients of target air pollutant were fixed, which means they were not time-varying. However, in Peng et al. (2005), the coefficient of target air pollutant becomes a function of time $\beta^c(t)$. They showed a sine/cosine model for $\beta^c(t)$ for the purpose of estimating smooth seasonal patterns in the city-specific log relative rates, and a pollutant × season interaction model for estimating PM_{10} log relative rates. From their results, the summer season has the highest increment in mortality as a 10 $\mu g/m^3$ increment in PM_{10} with a value of 0.36% (95% CI: 0.11%, 0.61%). When considering geographic regions, the Northeast has the strongest seasonal pattern, especially in the summer.

The distributed lag model can also apply all model modifications in previous papers. Welty and Zeger (2005) used 109 cities in the NMMAPS database to analyze temperature with seasonally and temporally varying coefficients and nonlinear temperature covariates. They concluded that there is a consistency with previous studies in the national average estimates of PM_{10} relative risk and robustness in model specification controlling for weather and seasonal trends.

Today, time series analysis is increasingly popular in air pollution studies in the U.S.

This is mainly because many well-organized databases contain all the data necessary for this analysis. These databases not only provide a convenient data source for researchers, but also effectively prevent a time-consuming data collection process and curb costs. However, time series study is not without limitation. Users need to pay closer attention than they do to other study designs when controlling for season and time trend. The number of lags included in models must also be taken into account as it is difficult to balance considering possible lag effects and avoiding the appearance of a concurvity problem which can result in underestimating the variance of effect estimates (Ramsay, Burnett, & Krewski, 2003a; Morlini, 2006). In addition, the aggregation over the population in time series studies may cause bias in estimated coefficients (Dominic et al., 2003c).

1.2.3. Other epidemiological study designs

Time series study design is not always optimal. When research groups have different research purposes, or restricted funding, time, and data sources, it is advisable to use alternative methods when approaching air pollution and mortality issues. For example, the case-crossover design, a modification of the case-control design (Breslow & Day, 1980; Schlesselman, 1994), was developed by Maclure (1991) from his study of acute transient effects of intermittent exposures, and has been extended to some air pollution and mortality data in Philadelphia (Neas, Schwartz, & Dockery, 1999), Seoul (Lee & Schwartz, 1999), Seoul (Lee & Schwartz, 1999), Australia and New Zealand (Barnett, Williams, Schwartz, Neller, Best, Petroeschevsky, & Simpson, 2005). By definition, the "case or index time" is a hazard period which is the time frame right before the disease or event onset, and the "control or reference time" is a control period which is a specified interval other than the hazard period. Researchers collect individual information from

subjects who had the disease or experienced the event, and test for consistent relationships between the disease and the exposure, while minimizing the possibility of confounding. The target of case-crossover study design is to estimate the odds ratio of effect by dividing the number of subjects who were exposed under specific circumambience during case or index time by those exposed during the control or reference time. Adjusted odds ratios are easily estimated by conditional logistic regression models (Agresti, 1990; Stokes, Davis, & Koch, 2000). For instance, Neas et al. (1999) calculated that a 100 $\mu g/m^3$ increment in the 48-hr mean level of total suspended particulate matter (TSP) was associated with increased all-cause mortality with an odds ratio of 1.056 (95% CI: 1.027, 1.086), with the odds ratio of death due to cardiovascular disease being 1.063 (95% CI: 1.021, 1.107) after adjusting for the same weather variables in Philadelphia. The case-crossover design can not only consider individual information (ex: age, sex, health status and behavior factors) to identify the susceptibility of subjects to the influence of air pollution, but also allow seasonal and secular trends by bi-directional selection of control periods (Jaakkola, 2003). However, it has been proved that this approach has approximately 50% lower power than the time series method (Bateson & Schwartz, 1999), and bias easily arises from time auto-correlated effects (Navidi, 1998; Bateson & Schwartz, 1999; Lumley & Levy, 2000; Levy, Lumley, Sheppard, Kaufman, & Checkoway, 2001a) and overlapped index time periods (Austin, Flanders, & Rothman, 1989; Lumley & Levy, 2000). Some solutions for these problems are conveyed in Navidi (1998), Lee and Schwartz (1999), Neas et al. (1999), Navidi and Weinhandl (2002) and Lumley and Levy (2000).

When it is allowable to collect individual information and track personal exposure measurements consistently for air pollution related research issues over time to investigate changes in repeated outcomes, a panel study is an appropriate design to handle such longitudinal data analysis. In details, a panel study design in air pollution research

often studies N individuals over a well-defined period (length = T), and collects health outcome measurements repeatedly for some time points during this period. The exposure measurement of air pollutants can be from a specific air pollution monitoring station which is close to the location where individuals receive the test, or from personal monitors distributed by well-trained instructors. Modern statistical models of longitudinal data analysis (Diggle, Liang, & Zeger, 1994; Singer & Willett, 2003) have been applied to estimate the effects of air pollutants on human health in panel study design research, including mixed, marginal, and transition models. Generalized linear models with generalized estimating equations (GEE), which can reduce the bias caused from an improperly specified working matrix, becomes an alternative choice for non-normally distributed response variables in panel data (Liang & Zeger, 1986; Zeger, Liang, & Albert, 1988). Along with the progress of Bayesian analysis, Bayesian hierarchical models are also broadly used to handle complicated variance structures induced by panel data. In addition, the advent of the faster Markov chain Monte Carlo (MCMC), has made it possible to do the calculations on these complex models, thus rendering the Bayesian hierarchical approach more practical (Gelman, Carlin, Stern, & Rubin, 2004). In air pollution and human health research, panel study design provides a contribution to investigate chronic diseases, such as asthma symptoms and lung function exacerbation. Some literature includes McCullagh (1980), Heagerty and Zeger (1996), Slaughter, Lumley, Sheppard, Koenig and Shapiro (2002) and Lagorio, Forastiere, Pistelli, Iavarone, Michelozzi, Fano, Marconi, Ziemacki and Ostro (2006).

The panel study design is very popular because it is able to incorporate personal exposure feasibly, estimate within and between-subject effects separately, control for subject-specific covariates, and target subpopulations easily (Dominici et al., 2003c). However, it can be difficult to recruit samples large enough to provide enough power, and subjects often drop out of study in the middle of research period for various reasons.

Another serious problem is that the study design is vulnerable to test threat, which violates internal validity because subjects often become primed to measurement instruments after repeated interviewing, thus making the sample atypical.

In the cohort study, subjects who have certain conditions and share common experiences receive a particular treatment and are followed over time and compared with another group who is thought of as the unexposed, or not affected by the condition under investigation. A cohort group is often defined as a group of individuals who are linked in some way or who have experienced the same significant life event within a given period (Rothman & Greenland, 1998). They can be either retrospective (looking back in time) or prospective (following cohorts over a time period). In air pollution and mortality study, the exposure variable is a measurement of cumulative air pollutants. The parameter of interest is the relative risk of disease incidence or death with high versus low air pollution exposure. The Cox proportional hazards (PH) model is a statistical tool in survival analysis (Cox & Oakes, 1984; Clayton & Hills, 1993), and is often used to assess the relationship between mortality and air pollution, especially to estimate mortality rate ratios for air pollutants by adjusting for other potential confounding factors. This model has been applied to many prominent cohort studies in air pollution research and health research, such as the Harvard Six Cities study (Dockery et al., 1993), American Cancer Society (ACS) study (Pope et al., 1995), and California-based Adventist Health and Smog (AHSMOG) study (Abbey et al., 1999). The participants enrolled in these studies are from 6,000 to 550,000 subjects. A basic assumption is that each subject within the city was assumed to have the same exposure level. In the Harvard Six Cities Study, air pollution has a significant influence in cardiopulmonary and lung disease, and the adjusted mortality-rate ratio for the most polluted city was 1.26 times (95% CI: 1.08, 1.47) higher than the least polluted city. In the ACS study, the adjusted relative risk ratios of all-cause mortality for the most polluted areas compared with the least polluted is 1.15

(95% CI: 1.09, 1.22) and 1.17 (95% CI: 1.09, 1.26) when using sulfate and fine particulate measurements, respectively. The AHSMOG study pointed out that long periods of residence and work location in areas of high ambient air pollution were associated with increased mortality.

1.2.4. Extended issues in air pollution and mortality studies

Even as statistical analyses have improved and progressed more precisely and in sophistication, the challenge of air pollution and human health research never ends because the natural properties of this issue still exist: large populations are exposed at low air pollution levels, so small relative risks should be detected with high statistical power (Dominici et al., 2003c). Scientists are still making efforts to discuss further related topics in this area. Some research opportunities either solved or yet to be solved are discussed below.

a) Multi-pollutant models

From previous literature, most studies have only focused on one main air pollutant effect. Particulate matter (PM_{10} or $PM_{2.5}$) has been identified as the main factor of air pollution contributing to adverse health. Thus, particulate matter has become the primary target of most air pollution research. Ozone is the second factor attracting attention. Some studies using more than one air pollutant in statistical models assigned a main factor (PM or O_3), and then regarded other air pollutants as confounders. We realize that any single air pollutant cannot act in the air without mixing and interacting with other air pollutants. However, little research discusses 2-pollutant models or 3-pollutant models. Chan et al. (2006) applied 2-pollutant models and 3-pollutant models to analyze air pollution influences in hospital emergency room visits for cerebrovascular diseases in Taipei. Many

distributed lag models incorporated with the GAM identified simultaneous mixture effects by 2-pollutant models (O_3+CO, O_3+$PM_{2.5}$, O_3+PM_{10}, CO+$PM_{2.5}$ and CO+PM_{10}) and 3-pollutant models (O_3+CO+$PM_{2.5}$ and O_3+CO+PM_{10}). In detail, they found O_3 and CO were more significantly associated with cerebrovascular admissions than PM_{10} and $PM_{2.5}$ in 2-pollutant models. O_3 had a significantly current effect in cerebrovascular diseases, but PM_{10}'s and $PM_{2.5}$'s effects were deferred to a 3-day lag. Particularly, CO had a significant 2-day lag effect on stroke, but the associations with other air pollutants were very weak. Other multi-pollutant model research can be seen in Wellenius, Schwartz and Mittleman (2005), Tsai, Goggins, Chiu and Yang (2003) and Le Tertre, Medina, Samoli, Forsberg, Michelozzi, Boumghar, Vonk, Bellni, Atkinson, Ayres, Sunyer, Schwartz and Katsouyanni (2002).

b) Multicollinearity and concurvity

Two potential problems recurrent in multi-pollutant models and distributed lag models are multicollinearity and concurvity, which change the direction of estimated parameters and present biases. Multicollinearity appears commonly among highly correlated linear factors, especially in lag effects. Concurvity can be regarded as a nonparametric analogue of multicollinearity, and often shows up among smoothers (Ramsay et al., 2003a). Since regression models were introduced, the task of dealing with multicollinearity problems has never ended. Generally speaking, concurvity in nonparametric and semiparametric models can cause following problems: 1) underestimated standard errors of parameters; 2) confidence intervals that are too narrow; 3) understated p-value; 4) greater type I error (Figueiras, Roca-Pardiñas, & Cadarso-Suárez, 2003). All four problems emanate from the same source. As concurvity exists, underestimated standard errors of parameters cause narrow confidence intervals,

which commit over threefold type I error along with smaller p-values. Concurvity often appears in modern air pollution and mortality research because most of the current studies use the GAM to analyze data. It is especially common for this to happen in the smoothing function of time because smoothing functions are typically used to adjust for possible autocorrelation. If a variable has a temporal trend and a linear effect on mortality simultaneously, the variable becomes a function of time (Ramsay et al., 2003a). Meanwhile, the time factor as the strategy of solving the concurvity problem in this issue cannot be ignored. The concurvity issue must be faced as many studies about air pollution and mortality have already confirmed the existence of concurvity in their analyses (Dominici et al., 2002b; Ramsay et al., 2003a).

So far, there is no strict criterion to identify the level of concurvity which can severely affect model fitting. Ramsay et al. (2003a) suggested using 0.5 to be the cutoff point. If the level of concurvity in nonparametric or semiparametric models is greater than 0.5, it is necessary to seek in order to eliminate or reduce this problem. Currently, researchers pay more attention to concurvity than to multicollinearity in this area. As a result, multicollinearity has been well-addressed in parametric linear models (Farrar & Glauber, 1967), but there is less advanced development in nonparametric or semiparametric models. Zidek, Wong, Le and Burnett (1996) discussed causality, measurement error, and multicollinearity in epidemiology, but only presented the existence of the three problems without further inference or solutions. The method of handling the concurvity problem has not been proposed either. An initial parametric bootstrap method was presented by Ramsay et al. (2003a). However, it failed to have satisfactory results, and returned to suggest replacing smoothers having concurvity by parametric methods with B-splines or natural splines. Nonetheless, the parametric method is not perfect because a parametric smoother cannot always be setup with enough flexibility to consider the possible time trend. Therefore, a nonparametric approach was

created by using a conditional bootstrap method which provides two advantages: 1) calculating bias-free estimated standard error of any linear predictor; and 2) persisting partial reduction of the residual bias when strict convergence criteria are applied (Figueiras et al., 2003, Figueiras, Roca-Pardiñas, & Cadarso-Suárez, 2005). Comparing non-corrected and bootstrap corrected results from simulations, the relative bias of the estimated coefficient reduces 3.6%-12.1% along with the concurvity coefficient from 0.56-0.90 as the bootstrap method is applied. Without correction, the coverage of the confidence intervals slumps from 0.94 to 0.66 as the concurvity coefficient raises from 0 to 0.90; however, there is almost no change in the coverage of the confidence intervals with the bootstrap correction (Figueiras et al. 2003). A reanalysis was published by Figueiras et al. (2005).

The bias from concurvity also appears in spatial air pollution data. Suppose (X_1,X_2) is the spatial location (e.g. longitude and latitude) of a given place, and $f(X_1,X_2)$ is the spatial function along with a linear factor X_3 in a GAM. We can also detect the concurvity coefficient by fitting the model $X_3=h(X_1, X_2)$ to get its correlation coefficient, where $h(.)$ can be any smoothing function (e.g. LOESS). From simulation, when concurvity coefficient is getting worse (i.e. concurvity coefficient > 0.5), Ramsay, Burnett and Krewski (2003b) confirmed the bias of estimated coefficients is raised along with inflated type I error. However, previous research showed the level of bias is up to the size of estimated coefficient. Meanwhile, if the true effect of linear factor is large enough, the bias coming from the concurvity in a GAM with a spatial smoothing function can be ignored (Dominici et al., 2002b).

c) Missing data

It is inevitable that air pollution measurements have some missing data from air

pollution monitoring stations. The levels of missing data from different air pollution monitoring stations also vary. For example, compared among those cities in the NMMAPS database, larger cities generally have less missing data than smaller cities. Among air pollutants, $PM_{2.5}$ tends to have more missing data than other air pollutants because it was not included in the NAQQS air quality standards until later. Sometimes, PM_{10} is measured once or twice per week, resulting in six or five missing PM_{10} measurements within a week. The most immediate impact of the missing data in statistical analysis is that it produces bias in parameter estimation, and several methods have been proposed based on the missing data patterns and mechanisms (Plaia & Bondi, 2006).

After 1970, missing data analysis started to bloom in statistical models or methodologies, and data imputation approaches were also being published at the same time. The methodology of multiple imputation gradually occupied a dominant role, and was preliminarily applied in sample surveys (Rubin, 1978; Rubin, 1980). However, a comprehensive methodology of multiple imputation was not presented until 1987 (Rubin, 1987), and the other multiple imputation approaches were generally developed based on Rubin's theorems.

The problem of missing data was not emphasized much in air pollution and mortality studies. Although most publications used complete case analysis (CCA), which means all missing data would be deleted from model fitting, researchers were not necessarily aware of it—especially in the distributed lag model. Many particulate matter studies have confirmed that daily mortality is associated with air pollutants (e.g. PM_{10} or $PM_{2.5}$) in the previous few days (Schwartz, 2000). Roberts and Martin (2007) created a GAM for imputing the missing PM concentrations:

$$PM_t = s(tmax_t, df = 6) + s(hum_t, df = 3) + s(t, df = 4 \times \#years)$$

where $tmax_t$ and hum_t are maximum temperature and minimum relative humidity on day *t*, respectively. They also suggested the possibility of extending the imputed model with other co-pollutants, but the limitation is that the data completeness of other co-pollutants is also uncertain. Zanobetti and his colleagues (2000) also proposed a similar imputed model to handle the missing data issue in their generalized additive distributed lag model. The imputed model is an additive model:

$$X_t = \theta \times X_{t-1} + f_1(min.temp_t) + f_2(relative.humid_t) + f_3(t) + error_t,$$

which can explain about 70% of the variability of the dependent variable.

Another method which has been used to handle missing air pollutant measurements is data augmentation (Tanner, 1991) within the Gibbs sampler (Geman & Geman, 1984). All missing data was generated and imputed by a conditional distribution of vectors of missing covariates (Z_{cm}), given vectors of observed covariates (Z_{co}) were established as a multivariate normal distribution iteratively (Dominici et al., 2002a). This method was also applied in a discussion of a measurement error model (Dominici, Zeger, & Samet, 2000b).

Besides the aforementioned issues, more research opportunities are still worthy of making efforts to improve, for example, misaligned environmental pollutants and health data which are measured at different scales of temporal and spatial resolution (Zanobetti et al., 2000; Schwartz, 2001), measurement error from using central air pollution monitoring stations' data to indicate individual exposure (McBride, Clyde, & Marcus, 2002; Dominici et al., 2000b; Carroll, Ruppert, & Stefanski, 1995; Zidek et al., 1996), and mortality displacement if the near-death individuals have passed away by the effects or side-effects of other diseases (Zeger, Dominici, & Samet,1999; Schwartz, 2001).

1.2.5. Spatial regression models

A spatial data set consists of a collection of observations that can be spatially located, and the source of spatial data includes maps, census data, air photographs, etc. It generally comes in three basic forms:

1) Map data: A map data set contains points, lines and polygons. Points represent the coordinate (x, y) of any fixed object on the surface of the earth. Lines represent an object with length, such as streets and rivers. Polygons represent natural, political, and administrative boundaries. An advanced map data named topographic map, supplies a detailed depiction of the earth with roads, rivers, buildings and numerous mapped objects (Kraak & Ormeling, 1996).
2) Attribute data: Attribute data also names tabular data, which is the descriptive data that geographic information system (GIS) links to map features. As collecting and compiling data for areas like states and cities, it generates map data packages, and can be implemented in a GIS.
3) Image data: This kind of data is often collected from satellite images and aerial photographs. Images can immediately and efficiently reflect the truest display from the surface of the earth. Image data is also easily cooperated with other map features which support the format of images.

Due to the diversity of map data, many spatial data sources provide enormous and organized databases from governmental units, such as U.S. Geological Survey(USGS), Census Bureau, Environmental Protection Agency, NASA, and many state-level data centers. Some commercial data sources are also available, such as Spatial Insights Inc., Geography Network, and GeoCommunity GIS Data Depot. These spatial data sources store different quantitative geographical data which can be applied in appropriately spatial regression models for the purpose of investigating the geographic impact on select

research targets.

Early developments in the statistical methodology of estimating spatial regression models can be traced back to a pioneer study of stationary processes in the plane (Whittle, 1954). Along with the maturity of statistical methodology, the approaches of estimating spatial interaction terms in modeling were proposed in the mid 70s (Besag, 1974; Ord, 1975). The influence and test of autocorrelation from spatial effects were also introduced during the same decade (Cliff & Ord, 1972; Cliff & Ord, 1973). The maturity and application of spatial modeling showed substantial progress in diagnostics for spatial dependence and heterogeneity, spatial econometrics (Anselin, 1988), advanced Statistics (Griffith, 1988; Cressie, 1993), and social and environmental science (Haining, 1990) in the late 1980s and early 1990s.

Several spatial functions can embed distinguished spatial covariates with different kinds of geographical data, such as Markov random fields (Spitzer, 1971; Preston, 1974), Gaussian random fields, low rank kriging (Nychka, Haaland, O'connell, & Ellner, 1998), and anisotropic spatial effects (Chiles & Delfiner, 1999). In particular, Markov random fields have been widely used because they succeeded the concept of Markov chain to perform an image chain, which can connect a site with its *neighbors* with a conditional probability, but keeps independent to the other sites. Meanwhile, this probability provides the relationship to any couple of geographical sites *s* and *s'* only if they share a common boundary or within a specific distance (Kneib, 2006a). Therefore, a Markov random field is a conditionally specific spatial model, which joint probability of spatial random variables is not constructed immediately, but established by a set of conditional probabilities (Huang, 2000). In the beginning, it has the difficulty that, given a conditional probability, it is not guaranteed that there exists a stochastic process containing the distribution of this conditional probability, but it was solved by Hammersley and Clifford (1971), who found that Markov random fields can be

equivalently characterized by a Gibbs distribution. This is so-called Hammersley-Clifford theorem, and a series of Markov random fields was established by Besag (1974) based on this theorem.

Now spatial analysis has been applied in many areas, such as social science (Goodchild, Anselin, Appelbaum, & Harthorn, 2000), public health (Waller & Gotway, 2004), Ecology (Fortin & Dale, 2005), Econometrics (Arbia, 2006; Anselin & Rey, 1997; Florax & van der Vlist, 2003; LeSage, Pace, & Tiefelsdorf, 2004) and Finance (Pace, Barry, & Sirmans, 1998; Pace & LeSage, 2004). Different study designs with specific data structure can also implemented in different types of spatial regression models, like the spatial econometrics of panel data (Elhorst, 2003; Anselin, Le Gallo, & Jayet, 2006), the analysis of spatial latent variables (Pinkse & Slade, 1998; Fleming, 2002), Bayesian inference of spatial autoregressive models (LeSage, 2000), and spatial generalized linear mixed models (Gotway & Stroup 1997, Zhang 2002, Gotway & Wolfinger 2003).

1.2.6. Generalized additive mixed models

The generalized additive mixed model was not established suddenly, but rather came about through gradually integrating many elements from different research. First, the generalized additive model offered methodology to match the structure of the generalized linear model (Hastie & Tibshirani, 1990). However, both nonparametric regressions and GAMs lack the ability to deal with correlated data, especially time series data. Initially, Hart (1990) showed a failure of cross validation when estimating the smoother parameter when data was positively correlated, and provided a modification in kernel regression estimation. A similar method of estimating the mean and covariance nonparametrically under the assumption that it is smooth with a modified cross-validation was proposed by Rice and Silverman (1991). Until the late 90s, some ideas about using linear mixed

models controlled by a nonparametric time function were finally developed (Zeger & Diggle, 1994; Zhang, Lin, Raz, & Sowers, 1998; Verbyla, Cullis, Kenward, & Welham, 1999). Furthermore, the approach of generalized estimating equations (Liang & Zeger, 1998) was also incorporated within the GAM for non-Gaussian longitudinal data (Wild & Yee, 1996; Berhane & Tibshirani, 1996). However, before the appearance of the GAMM, the research was limited in extending the above theories within the mixed effects model framework.

Initially, there were some difficulties regarding the development of the GAMM. For example, less work had been done on how to choose a good estimator of smoothing parameters and bandwidth parameters (Green & Silverman, 1994). Moreover, cross-validation was often time-consuming in computation, and the influence on the correlation parameters was hard to realize. Cross designs also failed to be used based on lack of contemporarily existing methodology (Lin & Zhang, 1999). Besides, there were some other problems that many researchers frequently encountered in practical regression application, such as spatially and temporally correlated observations, insufficient description of the heterogeneity among subjects by covariates, and complex interactions between covariates (Kneib, 2006a).

These challenges were preliminarily solved by Lin and Zhang (1999), in the spirit of the GAM and GLMM, by using nonparametric additive smoothing functions and adding random effects to the additive predictors for modeling covariate effects along with considering over-dispersion and correlation. In other words, the GAMM overcomes two main shortages of inability to estimate random effects, and within-subject variations in the GAM and the weakness of estimating nonparametric smoothing functions for nonlinear variables in the GLMM. Therefore, besides correlated data, clustered, hierarchical, and spatial data can also be used in the GAMM. Nested and crossed designed are also available (Lin & Zhang, 1999).

The work was not completed because Lin and Zhang (1999) pointed out the existence of bias, especially in binary data or special correlated random effects. Hence, a Bayesian approach via Markov Chain Monte Carlo (MCMC) methods can be used as an alternative to address this problem (Fahrmeir & Lang, 2001a) as MCMC can provide samples from all posteriors which can derive the necessary posterior distribution's profile without any approximate normality assumption. Similar approaches in other issues can refer Wong and Kohn (1996) for Gibbs sampling with Gaussian data in additive models. Another advantage of this method is that it is more efficient in computation than other methods for drawing from posterior distributions of spatial factors. Besides the GAMM, Fehrmeir and Lang also extended this methodology to the varying-coefficient mixed model (VCMM), which allows incorporating smoothing functions with coefficients. Their inference successfully applied to forest damage data by the VCMM with tree-specific random effects by using an unstructured covariance structure and duration of unemployment data by a GAMM with spatial random effects for different districts in Germany.

The basic assumption of random effects in either the GAMM or GLMM is that its distribution should be multivariate normal (Lin & Zhang, 1999; Brewslow & Clayton, 1993). However, this assumption may not be realistic in application. Zhang and Davidian (2004) focused on the violation of the normality assumption in random effects in the GAMM for clustered data, and proposed a conditional marginal likelihood (CML) inference based on a conditional estimation procedure. The central thought of the CML is to use the conditional distribution from the response variable given a sufficient and complete statistic for the random effect, and regarding it as a nuisance parameter to make robustness in any random effect's distribution. In simulation, five different distributions (Normal, Normal mixture, t, chi-square and Bernoulli distribution) were compared in the average of the estimated smoothing parameter based on their inference and double

penalized quasi-likelihood (DPQL) from Lin and Zhang (1999), and more robust results were achieved than those from DPQL estimates. They also applied CML to a case study from a multicenter AIDS cohort study (Kaslow, Ostrow, Detels, Phair, Polk, & Rinaldo, 1987) with the results from DPQL approach, and had consistent findings as simulation results.

Another area worthy of further research is extending the smoothing function to more than one variable in the GAMM. For example, using $f(X_1, X_2)$ instead of $f(X_1)$. This kind of smoothing offers more sophisticated nonlinear relationships between X_1 v.s. Y and X_2 v.s. Y, simultaneously, which can also be regarded as the nonparametric interaction term. So far, three methods are able to handle this special sort of smoothing function. First, a low-rank approximation approach to thin plate splines was developed in geoadditive model (Kammann & Wand, 2003). Second, a tensor product P-spline with the single penalty given by the Kronecker product of the penalties associated with the marginal bases was also constructed (Fahrmeir & Lang, 2001a). Third, a Bayesian type of tensor products B-splines was conducted with spatially symmetric priors on the B-spline coefficients (Lang & Brezler, 2004). A general form for deriving scale-invariant tensor product smoothers from low-rank penalized regression smoothers in the GAMM was also proposed by Wood (2006). The purposes of emphasizing scale-invariant in tensor product smoothers are that the covariates with unequal scales in the same tensor product smoother could lead to poor results. The *gamm* function of *mgcv* package in R software constructed by Wood can handle tensor product smoothers, but models like $f(X) + f(Z) + f(X, Z)$ are not supported in current version of *mgcv* package.

Based on the previous work of the GAMM, an extension of penalized spline generalized additive models for analyzing spatio-temporal data with a Bayesian perspective was proposed (Fahrmier et al., 2004; Kneib, 2006b) to be the prototype of the generalized geoadditive mixed model (GGAMM). The correlated spatial effects were

assumed to use a Gaussian Markov random field (Rue & Follestad, 2003) prior two dimensional P-splines. The inference was performed either with a full Bayes (FB) or an empirical Bayes (EB) approach. The MCMC technique and REML algorithm were still using the FB and EB approach, respectively. With the same framework, a general class of structured additive regression models for multinomial responses in either unordered or ordered categorical spatio-temporal data was developed (Kneib & Fahrmeir, 2006). This structured additive model family also extends to hazard regression models (Kneib, 2006b; Kneib & Fahrmeir, 2007). Note that the computational programming in these models was implemented in the statistical package BayesX, (see http://www.stat.uni-muenchen.de/~bayesx). It was developed at the Department of Statistics, University of Munich, for Bayesian inference in structured additive regression models (Brezger, Kneib, & Liang, 2005). This freeware is powerful to fit the GAM, GAMM, GGAMM, Gaussian directed acyclic graphs (DAGs) models (Fronk & Giudici, 2004) and the VCMM with either MCMC techniques or mixed model based techniques, and appropriate for any spatio-temporal analysis.

1.2.7. Applications of the GAMM in air pollution and mortality studies

As an innovative and outstanding model in the modern era, the GAMM seems not to be widely applied to most areas because of the inconvenience of software support. Lin and Zhang (1999) used the %GLIMMIX macro in SAS to propose simulation and case study in their paper. Right now, SAS Inc. has released the PROC GLIMMIX procedure to replace the function of %GLIMMIX. The *gamm* package in R is also available. BayesX is another freeware product which can fit the GAMM with the Bayesian method.

Some panel studies in air pollution research tried to use the GAMM to explain epidemiological issues. A reanalysis case study (Coull, Schwartz, & Wand, 2001)

collected data from 41 Utah Valley schoolchildren for 109 consecutive days. The main purpose of the study was to investigate the influence of air pollution on peak expiratory flow (PEF) in schoolchildren. In their model design, daily PM_{10} represents air pollution exposure, and it was also the only linear predictor. Two smoothing functions were included for daily lowest temperature and time effect. One random intercept and one random slope for daily PM_{10} to specific subject-level variation were the random effects. They concluded that the PEF decreases 0.7 units for a 10 $\mu g/m^3$ increase for PM_{10}, which is consistent with the initial analysis with the GAM in Pope et al. (1991). The large estimated variance of random intercept with a value of 2,630 implied large variability in the average of PEF among those subjects. The significance of estimated variance in the random slope demonstrated the existence of heterogeneity of PM_{10} effects among those schoolchildren as well.

Another example is the AIRGENE study which is a multi-center epidemiological study to assess inflammatory responses in association with ambient air pollution concentrations in myocardial infarction survivors and define susceptible subgroups of myocardial infarction survivors based on genotyping. The EU (European Union) funded study conducted data from six European cities with 1,000 myocardial infarction survivors, and measured their three inflammatory blood markers (CRP, Fibrinogen and InterLeukin-6) every 4 weeks. The logarithm of CRP is a time-invariant response variable, and a stepwise procedure for allowing either cubic terms or smoothers in all continuous predictors helps to organize this model. Note that they used an alternative model fitting programming by PROC MIXED procedure in SAS proposed by Ngo and Wand (2004) to derive AICs (Akaike, 1974). However, this model selection criterion has not been supported by any statistical inference in GAMM. It is concluded that Hba1c (diabetes indicator), alcohol intake, BMI, BNP (heart failure indicator), packs per year smoked, blood pressure, and cholesterol have linear effects on log(CRP), while age has a cubic

trend. Additionally, a dose-response curve for the air pollution effects with penalized splines was also explored. In this part, an exact likelihood ratio test for penalized splines (Crainiceanu, Ruppert, Claeskens, & Wand, 2005) was applied to test nonlinearity against linearity as well as no effect in the dose-response curve. To sum up, they concluded that ambient particles were associated with increases in IL-6 and fibrinogen concentration, and current treatment of myocardial infarction survivors with lipid-lowering medication may protect them from the adverse effects of air pollutants due to no association between air pollutants and CRP. In regards to those classical risk factors (BMI, age, sex, hand high cholesterol levels), they are correlated with surging inflammatory markers, which also implied the genetic variation in the inflammatory markers genes can determine their levels. Finally, inflammation in myocardial infarction survivors is determined by a number of time-invariant and time-varying environmental factors, which increases the risk for following events in the high-risk group of the population (Greven, Küchenhoff, Picciotto, Pekkanen, Bellander, Leonard, Chalamandaris, Kulmala, & Peters, 2005).

The latest research using the GAMM is a study of looking for the potential mechanism of PM-induced ischemia and connecting particulate matter to cardiovascular diseases (Chuang, Coull, Zanobetti, Suh, Schwartz, Stone, Litonjua, Speizer, & Gold, 2008). The subjects were gathered from Brigham and Women's Hospital at Harvard Medical School with 48 patients who have 4 repeated measurements by a percutaneous intervention for myocardial infarction, acute coronary syndrome without infarction, and stable coronary artery disease without acute coronary syndrome. The association between previous 24-hour mean black carbon levels and the risk of ST-segment depression $\geq$ 0.1mm is positive, but adversely negative in 0.5-hour averaged ST-segment level. Generally speaking, the ST-segment depression is correlated with increased $PM_{2.5}$ and black carbon in cardiac patients. They also identified that the first month after the cardiac event could have the greatest risk of air pollution related ST-segment depression in

myocardial injury. Note that Chuang et al. only assigned a random intercept in this GAMM, and the function of the GAMM is just for diagnosing the possible outliers in samples. In other words, it is no way to see subject-specific effect of $PM_{2.5}$ to cardiovascular outcomes.

1.2.8. Applications of the GGAMM

The GGAMM has not been widely used in air pollution and human health studies, but some research gradually accepted the features introduced by this model. For example, the car insurance claims data from a German insurance company containing detailed information on metrical and geographical covariates was hard to be fitted by an appropriate parametric model to concern their highly nonlinear relationships; but the generalized geoadditive model with MCMC technique had been successful analyzing the amount of loss and claim frequency (Fahrmeir, 2003). A Malawi study for evaluating the geographical location (districts) variation in the prevalence of cough among children ≤ 5 years old also applied the GGAMM, cooperating with many nonlinear categorical covariates (Kandala, 2006). Another similar children's health study covering more symptoms (diarrhea, cough, and fever) applied the GGAMM to investigate the impact of spatial and other potential risk factors in Nigeria (Kandala, Ji, Stallard, Stranges, & Cappuccio, 2007). A geoadditive hazard regression for interval censored survival data was also applied in the same area (Kneib, 2006b). The antenna of its application also reached the most popular human health research—AIDS/HIV studies. The prevalence of AIDS/HIV was pretty epidemic in Zambia throughout several decades, and researchers examined the association between the prevalence, age, gender, and districts location. The GGAMM found that the two districts, Lusaka and Copperbelt, had the first and highest prevalence of AIDS/HIV with marginal odds ratios of 3.24 and 2.88, respectively

(Kandala, Cappuccio, & Stones, 2008).

1.3. Motivations and objectives

After the new millennium, the development of methodology in air pollution and mortality studies gradually summarized many previous considerations with the 2-stage Bayesian hierarchical model (Dominici et al., 2000a; Dominici et al., 2002a; Dominici et al., 2003b, Dominici et al., 2005), and thus far less improvement is presented. Here, I proposed the generalized geoadditive mixed model (GGAMM) with an empirical Bayesian approach and mixed model based estimation (Fehrmeir & Lang, 2001a; Fehrmeir & Lang, 2001b; Kneib, 2006a) instead of the 2-stage Bayesian hierarchical model for the purpose of presenting a more concise model structure which can not only evaluate nation-level air pollutant effects, but also assess marginal city-level air pollution effects simultaneously. This unified model structure provides some advantages (Kneib, 2006a):

1. Both fixed and random effects are random variables distinguished by different priors with adopting a Bayesian perspective.
2. The priors of smoothing functions and spatially correlated data can be embedded into one general frame.
3. The general frame of the priors is able to be used for more general and unified estimating procedures, which is facilitated to be implemented and described.

Because no previous research used the GGAMM in multi-city time series air pollution and mortality studies, some objectives are also anticipated to be done in the following:

1. Our preliminary motivation comes from whether the GGAMM can offer more

intuitive and flexible device for applying multi-city time series air pollution and mortality data. We will pioneer a PM_{10} mortality model with corresponding time, weather, and location factors to see how the GGAMM works on spatio-temporal data.

2. The current effect of air pollution exposure along with its lag effects over a few days on daily mortality will be analyzed by the GGAMM.
3. The co-pollutant model will identify the influence of some co-pollutants, such as O_3, SO_2, NO_2, and CO along with PM_{10} effect.
4. Discussing the influence of starting values of smoothing parameter and the number of knots on handling potential convergence problems, twisted splines, and diminished spatial effect in the GGAMM.
5. Some missing data imputation methods will be implemented with their influence being identified on estimates in the GGAMM.
6. Two classic problems, multicollinearity and concurvity, are always troublesome in any semiparametric model; therefore, we will include the model evaluation in the level of multicollinearity among linear main factors, especially in corresponding lag effects, and the level of concurvity among smoothing functions. Our target is not only in original model, but also in models with imputed data because we wonder whether missing data imputation methods may raise the level of multicollinearity and concurvity.
7. An extended distributed lag model approach will be presented to handle technical problem when including too many lag effects in the GGAMM. The principal component analysis (PCA) will cooperate in this model, and the PCA-adjusted estimates will be developed to transform the estimates of principal component variables to the estimates of original air pollutant variables.
8. A comparison of GGAMMs and 2-stage Bayesian hierarchical models will be

performed with the same data using in case studies.

9. Three simulations will be accomplished for the sake of discussing the influence of concurvity, multicollinearity, and missing data imputation on estimates and smoothing functions.

The NMMAPS database will be used as the main data source of the whole analysis in the GGAMM. The general purpose of this study is to gather more concise and comprehensive analysis in air pollution and mortality studies via the GGAMM, and use the findings to apply it to more related research.

Chapter 2
METHODOLOGY

2.1. Model formulation and inference

Consider longitudinal data for subjects *n*=1,…, N, observed at time points $t \in \{t_1, t_2, \dots\}$. With corresponding responses, covariates and smoothing functions, a generic structure of the GGAMM can be shown as

$$g(y_{nt}) = u_{nt}^{'}\gamma + w_{nt}^{'}b + f_1(x_{nt1}) + \cdots + f_k(x_{ntk}) + f_{spat}(s_{nt}), \qquad (2.1)$$

where y_{nt} are predictor for subjects *n* at time *t* with corresponding covariates u_{nt} in linear fixed effects, $w_{nt} = (w_{nt1}, \dots, w_{ntq})'$ in linear random effects and $x_{nt1}, \dots, x_{ntk}$ in smoothing functions. The unknown parameter γ is a vector for unknown regression coefficients of fixed effects, and $b = (b_{1n}, \dots b_{qn})'$ contains *q* i.i.d. random intercepts or random slopes. Moreover, $g(.)$ is a link function, $f_j(.)$ are smoothing functions of continuous covariates x_{ntj}, and $f_{spat}(.)$ is a spatially correlated effect of the location s_{nt}, which can be boundary, contour and coordinated data. For notational simplicity, we initially exclude spatial function, and then equation (2.1) can be subsumed into a concise model structure with a semiparametric form

$$\eta_i = u_i^{'}\gamma + f_1(v_{i1}) + \cdots + f_p(v_{ip}) \qquad (2.2)$$

by defining $i = (n, t)$, $v_{i1} = u_{ntj}$, $j = 1, \dots k$, $v_{i,k+h} = w_{nth}$, $h = 1, \dots, q$, $f_{k+h}(v_{i,k+h}) = b_{hn}w_{nth}$, and $\eta_i = g(y_i)$. As a result, this model framework a GAM-like structure, which can facilitate advanced Bayesian inference based on maximum

likelihood estimations.

For Bayesian inference, we can express smoothing functions $f_j(v_{ij})$ to be a product of a design vector v_{ij} and a vector of unknown parameters ξ_j, i.e.

$$f_j(v_{ij}) = v_{ij}^{'}\xi_j. \tag{2.3}$$

Hence, (2.2) can be rewritten as

$$\eta_i = u_i^{'}\gamma + v_{i1}^{'}\xi_1 + \cdots + v_{ip}^{'}\xi_p. \tag{2.4}$$

A prior for a smoothing function $f_j(.)$ can be defined by an appropriate design vector v_{ij} and a prior distribution for the vector of unknown parameter ξ_j. Therefore, we can define the general form of the prior of ξ_j is a multivariate Gaussian distribution with density

$$p\left(\xi_j|\tau_j^2\right) \propto \exp\left(-\frac{1}{2\tau_j^2}\xi_j^{'}K_j\xi_j\right), \tag{2.5}$$

where K_j is a precision matrix which can be regarded as a penalty matrix to shrink parameter towards zero. It can also penalize too hasty jumps between adjacent parameters. Because most of K_js are rank deficient, it's no doubt that the prior of ξ_j is partially improper. Some other choices of priors of ξ_j can be found in Gamerman (1997) and Besag and Kooperberg (1995), but we will only restrict to use (2.5) in the following model estimation. For the prior of fixed effect vector γ, we can use either diffuse prior $p(\gamma) \propto const$ or multivariate Gaussian prior $N(\gamma_0, \Sigma_{\gamma_0})$. Here we only consider the noninformative prior because it can emphasize a close link of empirical Bayes approach to maximum likelihood estimation (Kneib, 2006a).

In order to estimate semiparametric models with smoothing functions, we need to have model representation to convert semiparametric models into the frame of parametric

models. A similar representation has already been applied in investigating the respiratory mortality of school children in Utah Valley (Coull et al., 2001). First of all, we need to consider the estimation of unknown parameter ξ_j in smoothing function (2.3). Based on original parameterization, the joint posterior distribution of ξ_j is

$$p(\xi_1, \dots, \xi_p, \gamma|y) \propto L(y, \xi_1, \dots, \xi_p, \gamma) \prod_{j=1}^{p} p(\xi_j|\tau_j^2), \tag{2.6}$$

where $L(.)$ is the likelihood function and $p(\xi_j|\tau_j^2)$ is the prior for ξ_j given in (2.5). However, (2.6) is hard to get ξ_j's maximum likelihood function, so a reparameterization is necessary to apply for estimating ξ_j and the other unknown parameters simultaneously. Generally, a ξ_j can be decomposed in to a penalized and an unpenalized component by this form

$$\xi_j = \tilde{X}_j \beta_j + \tilde{Z}_j b_j, \tag{2.7}$$

where $\tilde{X}_j$ is a $d_j \times (d_j - k_j)$ matrix, and $\tilde{Z}_j$ is a $d_j \times k_j$ matrix. The two matrices should be full rank and orthogonal. In addition, $\tilde{X}_j' K_j \tilde{X}_j$=0 and $\tilde{X}_j' K_j \tilde{X}_j = I_{k_j}$(Kneib, 2006a). Let $x_{ij}' = v_{ij}' \tilde{X}_j$ and $z_{ij}' = v_{ij}' \tilde{Z}_j$, we can import (2.7) to (2.4) as

$$\begin{aligned} \eta_i &= \sum_{j=1}^{p} u_i' \gamma + v_{ij}' \xi_j \\ &= \sum_{j=1}^{p} u_i' \gamma + (v_{ij}' \tilde{X}_j \beta_j + v_{ij}' \tilde{Z}_j b_j) \\ &= \sum_{j=1}^{p} u_i' \gamma + (x_{ij}' \beta_j + z_{ij}' b_j) \\ &= x_i' \beta + z_i' b, \end{aligned} \tag{2.8}$$

where $x_i = \left(u_i', x_{i1}', \ldots, x_{ip}'\right)'$, $\beta = \left(\gamma', \beta_1', \ldots, \beta_p'\right)'$, $z_i = \left(z_{i1}', \ldots, z_{ip}'\right)'$ and $b = \left(b_1', \ldots, b_p'\right)'$. The matrix notation of (2.8) can be presented as

$$\eta = \sum_{j=1}^{p} U\gamma + V_j\xi_j = \sum_{j=1}^{p} U\gamma + \left(X_j\beta_j + Z_j b_j\right) = X\beta + Zb, \qquad (2.9)$$

where the matrices *X* and *Z* are the vector-formatted presentation of x_i and z_i.

Eventually, we transform equation (2.4) to a GLMM structure with fixed effects β and random effects $b \sim N(0, Q)$, where $Q = blockdiag(\tau_1^2 I_{k_1}, \ldots, \tau_p^2 I_p)$. This model structure can allow us to apply GLMM methodology to estimate smoothing function f_i and variance parameters τ_j^2 simultaneously. With a flat prior of β, posterior (2.6) becomes

$$p(\beta, b|y) \propto L(y, \beta, b)\exp\left(-\frac{1}{2} b' Q^{-1} b\right), \qquad (2.10)$$

and the log-posterior is given by

$$l_p(\beta, b|y) = l(y, \beta, b) - \sum_{j=1}^{p} \frac{1}{2\tau_j^2} b_j' b_j, \qquad (2.11)$$

where $l(y, \beta, b) = logL(y, \beta, b)$. Hence, we can use (2.11) to derive a Fisher score algorithm with score function and the expected Fisher information matrix. For the score function, we obtain

$$s(\beta, b) = \frac{\partial l_p(\beta, b|y)}{\partial(\beta, b)} = \begin{pmatrix} s_\beta(\beta, b) \\ s_b(\beta, b) \end{pmatrix}, \qquad (2.12)$$

where

$$s_\beta(\beta, b) = \frac{\partial l_p(\beta, b|y)}{\partial \beta} = X' D S^{-1}(y - \mu), \qquad (2.13)$$

$$s_b(\beta, b) = \frac{\partial l_p(\beta, b|y)}{\partial b} = Z^{'} D S^{-1}(y - \mu) - Q^{-1} b, \tag{2.14}$$

$$D = diag(D_i) = diag\left(\frac{\partial g^{-1}(\eta_i)}{\partial \eta}\right), \tag{2.15}$$

and

$$S = var(y|\beta, b) = diag(\sigma^2) = diag\left(\frac{\phi v(\mu_i)}{\omega_i}\right). \tag{2.16}$$

In particular, $v(\mu_i)$ is the variance function determined by exponential family of response variable, ϕ is the scale parameter of corresponding exponential family and ω_i is a positive weight.

Similarly, the expected Fisher information can be expressed by

$$F(\beta, b) = \begin{pmatrix} F_{\beta\beta}(\beta, b) & F_{\beta b}(\beta, b) \\ F_{b\beta}(\beta, b) & F_{bb}(\beta, b) \end{pmatrix}, \tag{2.17}$$

where

$$F_{\beta\beta}(\beta, b) = X^{'} D S^{-1} D X, \tag{2.18}$$

$$F_{\beta b}(\beta, b) = F_{b\beta}(\beta, b)^{'} = X^{'} D S^{-1} D Z, \tag{2.19}$$

$$F_{bb}(\beta, b) = Z^{'} D S^{-1} D Z + Q^{-1}. \tag{2.20}$$

Based on (2.12) and (2.17), the regression coefficients can be estimated by iterating

$$\begin{pmatrix} \hat{\beta}^{(k+1)} \\ \hat{b}^{(k+1)} \end{pmatrix} = \begin{pmatrix} \hat{\beta}^{(k)} \\ \hat{b}^{(k)} \end{pmatrix} + (F^{(k)})^{-1} s^{(k)}. \tag{2.21}$$

Besides, we can also use an equivalent estimation process to solve parameters by using iteratively weighted least squares procedure to handle this linear system

$$\begin{pmatrix} X'WX & X'WZ \\ Z'WX & Z'WZ + Q^{-1} \end{pmatrix} \begin{pmatrix} \hat{\beta}^{(k+1)} \\ \hat{b}^{(k+1)} \end{pmatrix} = \begin{pmatrix} X'W\tilde{y} \\ Z'W\tilde{y} \end{pmatrix}, \tag{2.22}$$

where

$$\tilde{y} = X\hat{\beta}^{(k)} + Z\hat{b}^{(k)} + D^{-1}(y - \mu), \tag{2.23}$$

and

$$W = diag(\omega_i) = DS^{-1}D. \tag{2.24}$$

Equation (2.22) also provides a fundament for deriving credible intervals of estimated smoothing function (Lin & Zhang, 1999). According to (2.8), estimated smoothing function can be expressed by

$$\hat{f}_j = X_j\hat{\beta}_j + Z_j\hat{b}_j. \tag{2.25}$$

Hence, the covariance matrix of $\hat{f}_j$ is given by

$$cov(\hat{f}_j) = (X_j, Z_j)cov(\hat{\beta}_j', \hat{b}_j')(X_j, Z_j)', \tag{2.26}$$

and then the pointwise credible intervals can be constructed based on the diagnoal elements in (2.26) with assuming approximate normality of the estimated parameters (Kneib, 2006a).

Moreover, when concerning spatial effect, the spatial function f_{spat} can be divided into a structured spatial function f_{str} and an unstructured spatial function f_{unstr}. The importance of a spatial function is that it can explain many unobserved influence that other factors can hardly reflect. The unstructured spatial function can be regarded as a part of random intercept, and assumed to be i.i.d. Gaussian distribution. Moreover, structured spatial function can be estimated by Markov random fields (Kindermann &

Snell, 1980), Kriging (Journal & Huijbregts, 1978), and so on. Here we just introduce the Bayesian approach applying in Markov random fields. Suppose the spatial index $s \in \{1, \dots, S\}$ indicates some nearby locations in a geographical area, and assume two neighboring sites are more alike than any couple of arbitrary sites. A general prior for spatial smoothers for function, for example, $f_{spat}(\mathrm{s}) = \zeta_s$, is

$$\zeta_s | \zeta_{s'}, s' \neq s, \tau^2 \sim N\left(\sum_{s' \in \partial_s} \frac{\omega_{ss'}}{\omega_{s+}} \zeta_{s'}, \frac{\tau^2}{\omega_{s+}} \right), \tag{2.27}$$

where $\omega_{ss'} = c \times \exp(-d(s, s'))$ is the weight inverse proportional to the distance of centroids, and $\omega_{s+} = \sum_{s' \in \partial_s} \omega_{ss'}$ is the sum of weights of those sites *s'* who is a neighbor of site *s*, i.e., ∂_s. Additionally, $d(s, s')$ indicates the Euclidean distance between the centroids of site *s* and site *s'*, and *c* is a normalizing constant. Besides, when using boundary data, the weights can be defined as the proportion of the length in the common boundary of regions *s* and *s'*.

We can obtain maximum likelihood (ML) estimations of fixed effects and random effects by iterating (2.21) or solving (2.22), but the loss of degree of freedom while estimating β is not considered when deriving maximum likelihood estimation of variance parameter τ^2, and makes estimators are biased toward zero. Hence, it's necessary to adopt a restricted maximum likelihood (REML) to overcome this disadvantage (Patterson & Thompson, 1971).

Define $u = A'y$, where its expectation is equal to zero. The distribution of $u = A'y$ is independent of β, and it likelihood fits better in Bayesian model formulation. The matrix *A* can be derived from this decomposition

$$AA' = X(X'X)^{-1}X', \tag{2.28}$$

where A is a matrix with dimension $n \times (n - dim(\beta))$ and full column rank. In

addition, the A derived from (2.28) can satisfy $E(u) = 0$. As a result, we can obtain u with a marginal density

$$p(u) = (\frac{1}{2\pi})^{\frac{n-dim\,(\beta)}{2}} |X'X|^{\frac{1}{2}} |\Sigma|^{-\frac{1}{2}} |X'\Sigma^{-1}X|^{-\frac{1}{2}} exp\left\{-\frac{1}{2}\left(y - X\hat{\beta}\right)'\Sigma^{-1}\left(y - X\hat{\beta}\right)\right\}, \quad (2.29)$$

and the restricted maximum likelihood estimators of τ^2 and σ^2 can be obtained by maximizing

$$l^*(\tau^2, \sigma^2) = -\frac{1}{2}log(|\Sigma|) - \frac{1}{2}log(|X'\Sigma^{-1}X|) - \frac{1}{2}\left(y - X\hat{\beta}\right)'^{\Sigma^{-1}}\left(y - X\hat{\beta}\right), \quad (2.30)$$

where $\Sigma = W^{-1} + ZQZ'$ is the approximation of the marginal covariance matrix of working observation $\tilde{y} = X\hat{\beta}^{(k)} + Z\hat{b}^{(k)} + D^{-1}(y - \mu)$.

Then, we can use Newton-Raphson algorithm to get REML estimations of those variance parameters from (2.29). Another easier approach is using a modification of the Newton-Raphson algorithm given by Fisher score functions (Kneib, 2006a). Suppose the score function of variance parameter τ^2 and dispersion parameter ϕ is

$$s^*(\tau^2, \phi) = \left(s_1^*, \dots, s_{p+1}^*\right)' = \left(\frac{\partial l^*(\tau^2, \phi)}{\partial \tau_1^2}, \dots, \frac{\partial l^*(\tau^2, \phi)}{\partial \tau_p^2}, \frac{\partial l^*(\tau^2, \phi)}{\partial \phi}\right)', \quad (2.31)$$

where each element is the derivative of log-likelihood (2.30) with respect to each variance parameter τ_j^2 and dispersion parameter ϕ. Note that, it is allowable to ignore dispersion parameter, particular in Poisson and Binomial data because ϕ is fixed in several models (Kneib, 2006a). Moreover, the score functions of $\left(s_1^*, \dots, s_p^*\right)$ are not applicable when the number of observations is over 3000 because they handle much computation of several $n \times n$ inverse matrices Σ^{-1}, and possibly need huge amount of total memory to calculate these score functions (Kneib, 2006a). In order to avoid process Σ^{-1}, an adjusted score function derived by Lin and Zhang (1999) is given by

$$s_j^* = -\frac{1}{2}tr\left(Z_j'WZ_j\right) + \frac{1}{2}tr\left(Z_j'W(X\ Z)H^{-1}(X\ Z)'Z_j\right)$$
$$+\frac{1}{2}\left(\tilde{y} - X\hat{\beta} - Z\hat{b}\right)'WZ_jZ_j'W\left(\tilde{y} - X\hat{\beta} - Z\hat{b}\right), \quad (2.32)$$

where

$$H = \begin{pmatrix} X'WX & X'WZ \\ Z'WX & Z'WZ + Q^{-1} \end{pmatrix}. \quad (2.33)$$

Finally, if observations are not Poisson or Binomial data, then the numerical details of deriving score function of dispersion parameter ϕ can be found in Kneib (2006a).

When score functions are prepared, we can derive the expected Fisher information $F^*(\tau^2, \phi) = (F_{j,k}^*)$, $j,k = 1,\ldots, p+1$. By definition, $F_{j,k}^* = -\mathrm{E}\left(\frac{\partial^2 l^*(\tau^2,\phi)}{\partial\tau_j^2\,\partial\tau_k^2}\right)$, where

$$F_{j,p+1}^* = \frac{1}{2\phi}\Big[tr\left(Z_j'WZ_j\right) - 2tr\left((X\ Z)'WZ_jZ_j'W(X\ Z)H^{-1}\right)$$
$$+ tr\left((X\ Z)'W(X\ Z)H^{-1}(X\ Z)'WZ_jZ_j'W(X\ Z)\right)\Big],\ \ j = 1, \ldots, p \quad (2.34)$$

and

$$F_{p+1,p+1}^* = \frac{1}{2\phi^2}[n - 2tr((X\ Z)'W(X\ Z)H^{-1})$$
$$+ tr((X\ Z)'W(X\ Z)H^{-1}(X\ Z)'W(X\ Z)H^{-1})]. \quad (2.35)$$

The detailed proves can refer Kneib (2006a).

To sum up, we can construct a procedure to estimate fixed effects, random effects and variance parameters with the following two steps:

1. Obtain $\hat{\beta}^{(k)}$ and $\hat{b}^{(k)}$ given the current variance parameters by solving

$$\begin{pmatrix} X'WX & X'WZ \\ Z'WX & Z'WZ + Q^{-1} \end{pmatrix}\begin{pmatrix} \hat{\beta}^{(k)} \\ \hat{b}^{(k)} \end{pmatrix} = \begin{pmatrix} X'W\tilde{y} \\ Z'W\tilde{y} \end{pmatrix}.$$

2. Update variance parameters $\vartheta = (\tau^2, \phi)$ by

$$\vartheta^{(k+1)} = \vartheta^{(k)} + F^*\left(\vartheta^{(k)}\right)^{-1} s^*\left(\vartheta^{(k)}\right).$$

The two steps are performed iteratively and stopped until convergence.

2.2. Data collection

Daily time series of mortality, weather, and air pollution data is gathered from the NMMAPS database (Peng & Welty, 2004), and the time period is from 1991 to 1995 (calendar time is from day 1 to day 1,826). In order to reduce potential biases from losing information by missing data, 15 U.S. cities are chosen by missing rate of PM_{10}≤60%. The daily mortality is using the cardiovascular mortality, pneumonia mortality and respiratory mortality with three age categories (<65 years old, 65−74 years old and ≥75 years old), but the case study in the beginning is only adopting respiratory mortality≥65 years old. The ambient 24-h concentration of PM_{10} is the main air pollutant measurement in this study because it is the most influential agent to adverse human health. The 1-day and 2-day lag PM_{10} effects are also concerned for the purpose of evaluating particulate matter's distributed lag influence. Four main co-pollutants, CO, NO_2, O_3 and SO_2, are using in 2-pollutant models with PM_{10} for investigating potential compound effects between them. The 24-h average temperature is used to be the weather factor. Besides, a set of spatial data with longitude and latitude from selected cities is adopted.

2.3. Statistical models

Six GGAMMs will be showed in this section. We first introduce a single pollutant model fitting by the GGAMM to only consider PM_{10} adjusted by two smoothing confounders (calendar time and 24-h average temperature) and a spatial function. Define

Y_{ct} as the outcome variable for the number of deaths from respiratory diseases in city *c* on a particular calendar day *t*, and $Y_{ct} \sim Poisson(\mu_{ct})$. Its corresponding exposure PM_{10} concentration is PM_{ct}. Two confounders in respective smoothing functions are $TIME$ for calendar time and $TMEAN_t$ for 24 hourly mean of temperature. The day-of-week variable DOW contains six dummy variables to indicate Monday to Saturday. The spatial data $s_c = (long, lati)_c$ denotes the longitude and latitude in city *c*. Hence, the single pollutant model can be constructed as

$$g(\mu_{ct}) = \alpha + b_{0c} + \delta'(DOW) + (\beta + b_{1c})PM_{ct} + f(TIME) + f(TMEAN_t) + f_{spat}(s_c), \quad (2.36)$$

where *c*=1,…,15 and *t*=1,…,1826. Parameters α and β are unknown fixed intercept and slope. $\delta = (\delta_{Mon}\ \delta_{Tue}\ \delta_{Wed}\ \delta_{Thur}\ \delta_{Fri}\ \delta_{Sat}\ )'$ denotes the slopes of each day-of-week dummy variable. The random intercept b_{0c} and random slope b_{1c} follow a multivariate distribution, which

$$b = \begin{pmatrix} b_0 \\ b_1 \end{pmatrix} \sim MN\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma_0^2 & 0 \\ 0 & \sigma_1^2 \end{pmatrix}\right).$$

The function $g(.)$ is a log link function, and μ_{ct} is the expected value of Y_{ct}. The smoothing function of $f(TIME)$ was used to take into account the long-term variation in the mortality over several years, and the short-term effect of weather on the risk of death by including temperature smoother $f(TMEAN_t)$. In the interpretation, we mainly use $(\exp(10 \times \beta) - 1) \times 100$ to present the nation-level increase percent of relative risk in mortality rate per 10 μg/m^3 PM_{10} increase, and $(\exp(10 \times b_1^c) - 1) \times 100$ can be described as the marginal city-level increase percentage. Meanwhile, if we would like to investigate the PM_{10} effect in a specific city *m*, it is straightforward to calculate $(\exp(10 \times (\beta + b_1^m)) - 1) \times 100$ to be the answer.

The second model is the distributed lag model extended from model 1 with two additional lag effects PM_{ct}^{lag1} and PM_{ct}^{lag2}. The model form is:

$$g(\mu_{ct}) = \alpha + b_{0c} + \delta'(DOW) + (\beta_1 + b_{1c})PM_{ct} + (\beta_2 + b_{2c})PM_{ct}^{lag1} + (\beta_3 + b_{3c})PM_{ct}^{lag2} + f(TIME) + f(TMEAN_t) + f_{spat}(s_c), \quad (2.37)$$

where β_2 and β_3 are 1-day lag and 2-day lag PM_{10} effect at nation-level, respectively. The distribution of random effects becomes

$$b = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{pmatrix} \sim MN\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma_0^2 & 0 & 0 & 0 \\ 0 & \sigma_1^2 & 0 & 0 \\ 0 & 0 & \sigma_2^2 & 0 \\ 0 & 0 & 0 & \sigma_3^2 \end{pmatrix} \right).$$

Four co-pollutant models include CO, NO_2, O_3 and SO_2 to adjust PM_{10}, respectively. To simplify the notation, we construct a general form of them. Suppose V_{ct}^P denotes the concentration of co-pollutant, and *P* represents any of the above four co-pollutants. The co-pollutant model can be shown as

$$g(\mu_{ct}) = \alpha + b_{0c} + \delta'(DOW) + (\beta_1 + b_{1c})PM_{ct} + (\beta_2 + b_{2c})V_{ct}^P + f(TIME) + f(TMEAN_t) + f_{spat}(s_c). \quad (2.38)$$

Here β_2 and b_{2c} represent the national level and marginal city-specific co-pollutant effects, and the attitudes are the same as previous models. Note that $f(TIME)$ is defined as a P-spline with first order random walk penalty for calendar time, and $f(TMEAN_t)$ is defined as a quadratic P-spline with second order random walk penalty for 24-hour average temperature. The initial number of knots in $f(TIME)$ and $f(TMEAN_t)$ is 31 and 7, respectively.

2.4. Jackknife-bootstrap approach

In order to overcome a potential convergence problem in the estimated standard errors of fixed and random effects in the GGAMM, a traditional jackknife approach can offer reasonable se($\hat{\beta}$) and se($\hat{b}$), but there is an inconsistent problem when drawing different numbers of jackknife estimates, which both of se($\hat{\beta}$) and se($\hat{b}$) increase along with the increase of the number of jackknife estimates. Therefore, we embed the concept of bootstrapping in the jackknife approach to robust se($\hat{\beta}$) and se($\hat{b}$). The modification is that, instead of calculating standard errors from jackknife estimates, the bootstrap standard errors take the place of jackknife standard errors. In addition, the chosen jackknife estimates are not drawn without replacement but with replacement at all. The detailed steps are as follows:

Step 1: fitting the GGAMM from real data without the i^{th} observation in each city, where i = 1,…,1826. Defined these estimated fixed effects as $\hat{\beta}^{(-1)}$, $\hat{\beta}^{(-2)}$, …, $\hat{\beta}^{(-1826)}$.

Step 2: Drawing b jackknife estimates from $\hat{\beta}^{(-1)}$, $\hat{\beta}^{(-2)}$, …, $\hat{\beta}^{(-1826)}$ with replacement. Define these selected jackknife estimates as $\hat{\beta}^{1}$, $\hat{\beta}^{2}$, …, $\hat{\beta}^{b}$

Step 3: Calculating bootstrap estimate $\hat{\beta}^{B}$ by taking the average from $\hat{\beta}^{B} = \sum_{j=1}^{b} \hat{\beta}^{j}/b$.

Step 4: Calculating bootstrap standard error by $se(\hat{\beta}^{B}) = \sum_{j=1}^{b}\left(\hat{\beta}^{j} - \hat{\beta}^{B}\right)^{2}/(b-1)$.

Step 5: Repeating step 2 to step 4 with k times, and then taking the average of $\hat{\beta}^{B}$ and $se(\hat{\beta}^{B})$.

Here m will be assigned values with 100, 200, …, 1800 in chapter 3 to review its stability to jackknife-bootstrap estimates. The value of k is flexible. Theoretically, larger k can reduce the possibility of bias from drawing. In case study, we will use k=10,000 to complete the last step of jackknife-bootstrap procedure.

2.5. Multicollinearity and concurvity analysis

The criteria of evaluating the multicollinearity and concurvity level rely on calculating the correlation coefficient

$$\rho = corr(PM_{ct}, E(PM_{ct})), \tag{2.39}$$

where $E(PM_{ct})$ is the prediction of PM_{ct} by the fitted model

$$E(PM_{ct}) = (\hat{\eta}_0 + \hat{\varphi}_{0c}) + (\hat{\eta}_1 + \hat{\varphi}_{1c})PM_{ct}^{lag1} + (\hat{\eta}_2 + \hat{\varphi}_{2c})PM_{ct}^{lag2} \tag{2.40}$$

or

$$E(PM_{ct}) = (\hat{\eta}_0 + \hat{\varphi}_{0c}) + (\hat{\eta}_1 + \hat{\varphi}_{1c})V_{ct}^{P} \tag{2.41}$$

for multicollinearity level in distributed lag model or co-pollutant models, and

$$E(PM_{ct}) = \hat{f}(TIME) + \hat{f}(TMEAN_{ct}) + \hat{f}_{spat}(s_c) \tag{2.42}$$

for concurvity level. The concurvity influence on co-pollutants can also be determined by

$$\rho = corr(V_{ct}^{P}, E(V_{ct}^{P})), \tag{2.43}$$

where

$$E(V_{ct}^{P}) = \hat{f}(TIME) + \hat{f}(TMEAN_{ct}) + \hat{f}_{spat}(s_c). \tag{2.44}$$

The conclusion of existing severe multicollinearity and concurvity is based on weather ρ is larger than 0.7. In our six models, we will diagnose the following items:

(1) The multicollinearity level between PM_{10} and corresponding 2 lag effects in the distributed lag model.

(2) The multicollinearity level between PM_{10} and CO, NO_2, O_3, SO_2 in the

co-pollutant model, respectively.

(3) The concurvity level between PM_{10} and the equation of smoothing functions $f(TIME)$ and $f(TMEAN)$ with spatial function $f_{spat}(S)$ in each model.

(4) The concurvity level between 2 lag PM_{10} effects and the equation of smoothing functions $f(TIME)$ and $f(TMEAN)$ with spatial function $f_{spat}(S)$ in the distributed lag model.

(5) The concurvity level between co-pollutant CO, NO_2, O_3, SO_2 and the equation of smoothing functions $f(TIME)$ and $f(TMEAN)$ with spatial function $f_{spat}(s)$ in four co-pollutant models.

2.6. The extended distributed lag model and PCA adjusted estimates

The extended distributed lag model is using for handle the convergence and technical problems of multivariate lag effects included in the GGAMM. Theoretically, equation (2.36) can be implemented more number of lag effects, but may make BayesX crash. The multicollinearity level may become higher as well. In order to solve the two issues simultaneously, the most intuitive idea is doing variable condensation. Principal component analysis (PCA) is a popular multivariate analysis for data condensation, and uses an orthogonal linear transformation to convert the data into a new coordinate system such that the greatest variance by any projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on (Jolliffe, 2002; Johnson & Wichern, 2007). The most significant property of PCA is that all principal components are orthogonal with each other; meanwhile, they are totally independent. The basic concept of this transformation is remaining most of information among data by some linear combinations to explain most of its variance. An expectation is that using some ‘functions’ to represent the entire

lag effect series– just like the function of smoothers in time and temperature; therefore, a linear combination organized from PCA would be the best way to merge different lag effects together.

Suppose *n* variables can generate at most *n* principal components. Assume *PRIN* denotes the principal component vector, and vector *X* denotes original PM_{10} variables. They can be represented by a matrix form from PCA as

$$PRIN = \begin{pmatrix} PRIN_1 \\ PRIN_2 \\ \vdots \\ PRIN_n \end{pmatrix} = AX = A \begin{pmatrix} PM_{ct} \\ PM_{ct}^{lag1} \\ \vdots \\ PM_{ct}^{lag(n-1)} \end{pmatrix}, \tag{2.45}$$

$$A = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{2n} & \cdots & a_{nn} \end{pmatrix}, \tag{2.46}$$

where *A* is a loading matrix calculated from the eigenvectors (i.e., a_1, a_2,…) of variance-covariance matrix of $(PM_{ct}, \ldots, PM_{ct}^{lag(n-1)})$. The order of these eigenvalues $(\lambda_1, \ldots, \lambda_n)$ is usually ranked from maximum to minimum. The element of loading matrix can be solved from

$$\boldsymbol{S} \begin{pmatrix} a_{i1} \\ a_{i2} \\ \vdots \\ a_{in} \end{pmatrix} = \lambda_i \begin{pmatrix} a_{i1} \\ a_{i2} \\ \vdots \\ a_{in} \end{pmatrix}, \tag{2.47}$$

where $\boldsymbol{S}$ is the sample covariance matrix of *X*, and each principal component can account for the proportion of variance among all original variables by $(\lambda_i / \sum_{j=1}^{n} \lambda_j) \times 100\%$, for *i*=1,…,*n*. Suppose the first *m* principal components have cumulative proportion of variance≥70%, and then each principal component can be transformed to

$$PRIN_{ctq} = a_{k1} PM_{ct} + a_{k2} PM_{ct}^{lag1} + \cdots + a_{kn} PM_{ct}^{lag(n-1)} \tag{2.48}$$

for q=1,…,m. Meanwhile, the distributed lag model becomes

$$g(\mu_{ct}) = \alpha + b_{0c} + \delta'(DOW) + (\beta_1 + b_{1c})PRIN_{ct1} + \cdots + (\beta_m + b_{mc})PRIN_{ctm} + f(TIME) + f(TMEAN_t) + f_{spat}(S_c), \quad (2.49)$$

which is so called the extended distributed lag model.

Here we propose an approach to transform the estimates $(\beta_1, \ldots, \beta_m)$ in (2.49) to original variables' estimates, say, PCA-adjusted estimates, which can immediately reflect the associate between original particulate matter effect and mortality rate. Define l_{ij} as the loading of the j^{th} variable in the i^{th} principal component. The relative proportion of the loading in the j^{th} variable and the i^{th} principal component can be calculated by $l_{ij}/\sum_{j=1}^{q}|l_{ij}|$. Suppose p_i is the proportion of variation that the i^{th} principal component can account for, so the increase of relative rate per j^{th} original variable increase is

$$\hat{\gamma}_j = \sum_{i=1}^{m} \frac{l_{ij}}{\sum_{j=1}^{q}|l_{ij}|} p_i \hat{\beta}_i. \quad (2.50)$$

The standard error of $\hat{\gamma}_j$ can be derived from

$$se(\hat{\gamma}_j) = \left[\sum_{i=1}^{m} \left(\frac{l_{ij}}{\sum_{j=1}^{q}|l_{ij}|} p_i\right)^2 Var(\hat{\beta}_i)\right]^{\frac{1}{2}}. \quad (2.51)$$

The subject-specific effect in the k^{th} subject in the j^{th} variable can be calculated by

$$\hat{\xi}_{kj} = \sum_{i=1}^{m} \frac{l_{ij}}{\sum_{j=1}^{q}|l_{ij}|} p_i \hat{b}_{ki}, \quad (2.52)$$

so its standard error is

$$se(\hat{\xi}_j) = \left[\sum_{i=1}^{m}\left(\frac{l_{ij}}{\sum_{j=1}^{q}|l_{ij}|}p_i\right)^2 Var(\hat{b}_i)\right]^{\frac{1}{2}}. \qquad (2.53)$$

Therefore, the $\hat{\gamma}_j$, $se(\hat{\gamma}_j)$ and $se(\hat{\xi}_j)$ the PCA-adjusted estimates of $\hat{\beta}_j$, $se(\hat{\beta}_j)$ and $se(\hat{b}_j)$.

2.7. Missing data analysis

Most of published literatures used complete case analysis to handle severe missing data problem by just deleting missingness whatever other corresponding variables are complete (Dominici et al., 2000a; Dominici et al., 2003a). We are interested in the efficiency of using some missing data imputation methods to compensate possible information losing, especially in smoothing functions. Suppose the mechanism of missing data in the NMMAPS follows MAR, the following missing data imputation methods were applied to make these data sets back to completeness:

1. Nearest neighbor imputation – version I (NNI1)
2. Nearest neighbor imputation – version II (NNI2)
3. Multiple imputation by MCMC (MI-MCMC)

In details, the NNI1 is a kind of hot deck imputation with a long story, and has been used in many surveys conducted by Statistics in Canada, the U.S. Bureau of Labor Statistics, and the U.S. Census Bureau. Its statistical properties had not been derived until Chen and Shao (2001), which gave a detailed inference over several issues to get asymptotically unbiased and consistent estimated variances. Generally speaking, suppose the data structure with m missing values for the row indices i=n-m+1,…,n can be re-expressed by

$$\text{observed}\begin{cases} X_1 \\ . \\ . \\ . \\ X_{n-m} \end{cases} \qquad \text{observed}\begin{cases} Y_1 \\ . \\ . \\ . \\ Y_{n-m} \end{cases}$$

$$\text{missing}\begin{cases} X_{n-m+1} \\ . \\ . \\ . \\ X_n \end{cases} \qquad \begin{cases} Y_{n-m+1} \\ . \\ . \\ . \\ Y_n \end{cases}$$

A missing value $X_j,\ j = n - m + 1, \dots, n$, is imputed by choosing that value X_i, $i = 1, \dots, n - m$, which is corresponding to its closest value Y_j to Y_i. This is also the true meaning of *nearest neighbor*. The definition of *closest* is determined by the distance between any two response values. In other words, the distance of the nearest neighborhood is calculated with all observed values for Y by

$$\left|Y_i - Y_j\right| = min_{1 \leqq k \leqq n-m}\left|Y_k - Y_j\right| \tag{2.54}$$

When we find the response value $Y_k,\ k = 1, \dots, n - m$, which is the closest one to $Y_j,\ j = 1, \dots, n$, we can impute its corresponding X_k to the missing value. If there are more than one X_k whose corresponding response values Y_k has the same minimum distance to Y_j among others, then the mean of those *X* values is imputed.

However, the classical NNI has a potential disadvantage because a smoothing function relating x and y can lead to substitutes being far away from the 'true' value (Nittner, 2002). Hence, the NNI2 has been modified from NNI1 to handle possible imprecision in this situation. Consider a neighborhood of y_j contains a pre-determined number of neighbors k. A key concept in this method is to control the range of the fixed neighborhood, and impute data based on different principles after comparing with a percentage p of the length of the data interval. Suppose k=3 and p=0.05, the ordered values $x_{[s]}$ for *s*=1,2,3 satisfying equation (2.54). Then, the range *R* and interval *I* can be expressed as

$$R = x_{[3]} - x_{[1]} \tag{2.55}$$

$$I = \left(x_{[max]} - x_{[min]}\right) \times 0.05 \tag{2.56}$$

Then, a concise step-by-step procedure in NNI2 based on the above assumptions is:

Step 1. If $R \leq I$, then impute a random number generated from $U(x_{[1]}, x_{[3]})$

Step 2. If $R > I$, then compute $z = R - I$, and generate a random number u from $U(0, z_{[max]})$, where $z_{[max]} = 0.95 \times R$

Step 2.1. If $u > z$, then impute $(x_{[1]} + x_{[2]} + x_{[3]})/3$

Step 2.2. If $u \leq z$, then compute an empirical distribution $N(\bar{X}, S^2)$ from observed x and three probabilities $P\left(X \leq x_{[1]}\right)$, $P\left(x_{[1]} < X \leq x_{[3]}\right)$, and $P\left(X > x_{[3]}\right)$. After ordering them, imputing a value satisfying the condition of the maximum probability and satisfying $x_{min} \leq X \leq x_{max}$.

Note that if there are more than one $x_{[1]}$ and $x_{[3]}$, the average $x_{[1]}$s and $x_{[3]}$s can be used in the procedure. The efficiency of NNI1 and NNI2 has been confirmed in missing data in the independent variable when fitting additive model (Nittner, 2002). So far, there is no existing package in any statistical software, so two self-made SAS macros %NNI1 and %NNI2 were used to handle the two imputation procedures.

In original methodology of NNI1 and NNI2, there is no special restriction. Both continuous and categorical variable can apply it. However, even though there is no study to support how large data set it can support, too small sample size may cause somehow imprecision. In addition, NNI1 and NNI2 can be immediately applied in one independent variable with one dependent variable, but impossibly used in multivariate imputation. A compromised way is making a correlation matrix among them, and picking the complete variables with the highest correlation with another variable containing missing data. Nonetheless, this modification is not scientific proven, and loss too many information from other variables which aren't used in NNI1 and NNI2. That is the reason why

multiple imputation is popular in this situation, even though it needs more assumptions from data itself.

The multiple imputation method can easily handle large number of variables simultaneously, whatever variables themselves are complete or incomplete. Among the categories of multiple imputation methods, the Monte Carlo Markov Chain (MCMC) method can simulate the joint posterior distribution of unknown values and estimate simulation-based posterior parameters. Considering general regression model with outcomes $\boldsymbol{Y}$ and a vector of predictors $\boldsymbol{X}$. For a given subjects, these variables are either observed or partially missing. We define $\boldsymbol{Z} = (\boldsymbol{Z}^{obs}, \boldsymbol{Z}^{mis})$, where $\boldsymbol{Z}^{obs} = (\boldsymbol{Z}^{obs}, \boldsymbol{X}^{obs})$ and $\boldsymbol{Z}^{mis} = (\boldsymbol{Y}^{mis}, \boldsymbol{X}^{mis})$, and $\boldsymbol{R}$ as a set of indicator variables, where $R_j = 1$ if the j^{th} element of $\boldsymbol{Z}$ is observed, and $R_j = 0$ otherwise. The appropriate situation of using multiple imputation method is that the data should follow either missing completely at random (MCAR) mechanism

$$P(\boldsymbol{R}|\boldsymbol{Z}) = P\left(\boldsymbol{R}\middle|\boldsymbol{Z}^{obs}, Z^{mis}\right) = P(\boldsymbol{R}|\phi), \tag{2.57}$$

which means the missing data is not related to any variable, whatever known or unknown, or missing at random (MAR) mechanism

$$P(\boldsymbol{R}|\boldsymbol{Z}) = P(\boldsymbol{R}|\boldsymbol{Z}^{obs}) = P(\boldsymbol{R}|\phi), \tag{2.58}$$

which indicates the missing data is only related to observed quantities of variables. Note that ϕ is presumed parameter set. Use of MCAR or MAR allows the analyst to generate imputations $(\boldsymbol{Z}^{\{1\}}, \boldsymbol{Z}^{\{2\}}, \dots, \boldsymbol{Z}^{\{m\}})$ from the conditional distribution $f(\boldsymbol{Z}^{mis}, \phi|\boldsymbol{Z}^{obs})$ iteratively. The whole procedure can be implemented using the IP algorithm (Schafer, 1997), which two steps can be defined at the t^{th} iteration as:

Imputation step: Draw $\boldsymbol{Z}^{mis,(t+1)}$ from $f(\boldsymbol{Z}^{mis}|\boldsymbol{Z}^{obs}, \phi^{(t)})$.

Parameter step: Draw $\phi^{(t+1)}$ from $f(\phi|\boldsymbol{Z}^{obs}, \boldsymbol{Z}^{mis,(t+1)})$.

In imputation step, suppose $\boldsymbol{\mu} = (\boldsymbol{\mu}_1{}', \boldsymbol{\mu}_2{}')'$ is a partitioned mean vector of $\boldsymbol{Z}^{obs}$ and $\boldsymbol{Z}^{mis}$, and a partitioned covariance matrix of $\boldsymbol{Z}^{obs}$ and $\boldsymbol{Z}^{mis}$ is

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{12}' & \boldsymbol{\Sigma}_{22} \end{pmatrix},$$

where $\boldsymbol{\Sigma}_{11}$ and $\boldsymbol{\Sigma}_{22}$ are covariance matrices of $\boldsymbol{Z}^{obs}$ and $\boldsymbol{Z}^{mis}$, respectively, and $\boldsymbol{\Sigma}_{12}$ is the covariance matrix between $\boldsymbol{Z}^{obs}$ and $\boldsymbol{Z}^{mis}$. Hence, the conditional distribution of $\boldsymbol{Z}^{mis}$ given $\boldsymbol{Z}^{obs} = \boldsymbol{z_1}$ is a multivariate normal distribution with the men vector

$$\boldsymbol{\mu}_{2.1} = \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{12}'\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{z_1} - \boldsymbol{\mu}_1), \tag{2.59}$$

and the conditional covariance matrix

$$\boldsymbol{\Sigma}_{22.1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{12}'\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}. \tag{2.60}$$

In Bayesian theorem, suppose that a $n \times p$ matrix $\boldsymbol{Y} = (y_1', y_2', \ldots, y_n')'$ is distributed with a multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, the posterior distributions of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are

$$\boldsymbol{\Sigma}|\boldsymbol{Z} \sim W^{-1}(n + m,\ (n-1)\mathbf{S} + \boldsymbol{\Psi} + \frac{n\tau}{n+\tau}(\bar{z} - \mu_0)(\bar{z} - \mu_0)'\Big), \tag{2.61}$$

$$\boldsymbol{\mu}|\boldsymbol{\Sigma}, \boldsymbol{Z} \sim MN\left(\frac{1}{n+\tau}(n\bar{z} + \tau\mu_0),\ \frac{1}{n+\tau}\boldsymbol{\Sigma}\right), \tag{2.62}$$

where $W^{-1}(a, \boldsymbol{b})$ means an inverted Wishart distribution with the degree of freedom *a* and a precision matrix $\boldsymbol{b}$; *n* is the total number of observations in $\boldsymbol{Z}$; *m* and $\boldsymbol{\Psi}$ are the mean and precision matrix of prior distribution of $\boldsymbol{\Sigma}$; $(n-1)\mathbf{S}$ is the corrected sum of squares and crossproducts (CSSCP) matrix; μ_0 and τ are the mean and the denominator of variance-covariance matrix in the prior distribution of $\boldsymbol{\mu}|\boldsymbol{\Sigma}$, respectively (Anderson,

1984).

Based on (2.61) and (2.62), we can derive the posterior distribution of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ from their prior information in posterior step, and here we only use a noninformative prior, i.e., the Jeffery prior, to obtain

$$\boldsymbol{\Sigma}^{(t+1)}|\boldsymbol{Z}\sim W^{-1}(n-1,(n-1)\mathbf{S}), \tag{2.63}$$

$$\boldsymbol{\mu}^{(t+1)}|\boldsymbol{\Sigma}^{(t+1)},\boldsymbol{Z}\sim MN\left(\bar{z},\frac{\boldsymbol{\Sigma}^{(t+1)}}{n}\right). \tag{2.64}$$

The two steps construct a Markov Chain to simulate draws $\{\boldsymbol{Z}^{(1)},\phi^{(1)}\}$, $\{\boldsymbol{Z}^{(2)},\phi^{(2)}\},\dots,\{\boldsymbol{Z}^{(t+1)},\phi^{(t+1)}\}$ from the posterior distribution of $f(\boldsymbol{Z}^{mis},\phi|\boldsymbol{Z}^{obs})$, and this Markov Chain can converge to this posterior distribution as well. After replicating the above procedure m times to generate m imputed data sets, we can fit the GGAMM for each imputed data sets, and get m model-fitting results. Finally, m results should be integrated into a final result in pooling step. The purpose of this step is providing robust estimates of the parameters and their standard errors. Some extensive papers concerning the asymptotic behavior of multiple imputation methods can refer Barnard and Rubin (1999), Meng and Rubin (1992), and Robins and Wang (2000).

2.8. Simulation

In order to identify more concrete and concise impact of multicollinearity and concurvity and the influence of missing data imputation methods in the GGAMM, a series of simulated data sets are generated to process the three simulations. The concurvity simulation and missing data imputation simulation are using artificial data, but the simulated data in the multicollinearity simulation is generated from real data.

The artificial data generating procedure is similar as the simulation method proposed by Lin and Zhang (1999). The first step is constructing a simulated model to generate

random data repeatedly. 1,000 data sets are generated with 10 subjects in each data set, and each subject has 100 repeated measurements. The framework of the simulated GGAMM is constructed by

$$g(Y_{ij}) = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})X_{1ij} + f(X_{2ij}) + f_{spat}(u_i, v_i) + e_{ij} \qquad (2.65)$$

for *i*=1,…, 10 and *j*=1,…, 100, where X_{1ij} is an independent random variable generated from a normal distribution $N(0, 0.16)$; the variable X_{2ij} is supposed to be a covariate changing within each subject with equally 100 knots in [0, 1], and define it follows a normal distribution

$$N\left(\frac{trun\left\{\frac{i+4}{5}\right\}}{100} + 0.01(j-1), 0.0001\right),$$

where *trun*{.} indicates a truncation operator, which only remains the integer part of any number in it. The between-subject error term e_i is generated from a normal distribution $N(0, 0.09)$, and the within-subject error term has autoregressive correlation by $e_{ij} = \rho e_{i,j-1} + e_{ij}$ with $\rho = 0.2$. The smoothing function *f* is a bimodal function

$$f(x) = \frac{1}{10}\{6F_{30.17}(x) + 4F_{3,11}(x)\} - 1, \qquad (2.66)$$

where $F_{p,q}(.)$ is the probability distribution function of beta distribution

$$F_{p,q}(x) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} x^{p-1}(1-x)^{q-1}, \qquad (2.67)$$

and $\Gamma(.)$ is a gamma function. The constant 1 used in $f(x)$ is for the purpose of centering smoothing function. The spatial function is simply defined as

$$f_{spat}(u_i, v_i) = 0.0001 \times (u_i + v_i), \qquad (2.68)$$

where u_i and v_i are generated from a uniform distribution $U(0, 10)$. It can be regarded as a monotone increasing linear function from south-west to north-east.

The true values of fixed intercept and slop (β_0, β_1) are defined as (0.1, 0.1), and the data of random intercept and slope (b_{0i}, b_{1i}) are simultaneously generated from a multivariate normal distribution

$$MN\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}0.7^2 & 0\\0 & 0.6^2\end{pmatrix}\right).$$

Finally, the response variable Y_{ij} can be simulated from a Poisson distribution with mean parameter

$$\mu_{ij} = exp\left((\beta_0 + b_{0i}) + (\beta_1 + b_{1i})X_{1ij} + f(X_{2ij}) + f_{spat}(u_i, v_i) + e_{ij}\right). \quad (2.69)$$

The artificial data simulated from above steps can be immediately use in missing data imputation simulation. When 1,000 simulated data sets are prepared, each data set randomly drops out linear predictor X_{1ij} and covariate X_{2ij} in smoothing function by different missing rates. In order to clarify the efficiency of each missing data imputation method, X_{1ij} and X_{2ij} are independently dropped out to make two scenarios in this simulation. Note that dependent variable Y_{ij} is always complete in both scenarios. This procedure is strongly based on that the missing data mechanism is missing completely at random (MCAR) or, at least, missing at random (MAR). The missing rates are varied from 5%, 10%, 20%, 30%, 40% and 50%.

We are going to investigate targeted estimates $\hat{\beta}_1$ and estimated smoothing function $\hat{f}$. Two different $\hat{\beta}_1$s are estimated from the simulated data with missing X_{1ij} and missing X_{2ij}, respectively. An adjusted sample mean square error is modified from initial sample mean square error in Nittner's paper (2002), and applied to be the criterion of assessing estimated smoothing functions. It follows

$$\text{ASMSE}\left(\hat{f}(y_j), f(y_j)\right) = \frac{1}{\kappa}\left\{\sum_{j=1}^{\kappa} \widehat{Var}(\hat{f}(y_j)) + \left[\widehat{\text{B}}(\hat{f}(y_j), f(y_j))\right]^2\right\}, \quad (2.70)$$

where κ is the number of valid *y* in the smoothing function, and $\widehat{\text{B}}$ is the bias between $\hat{f}(y_j)$ and $f(y_j)$.

By using the same data generating procedure from (2.65) to (2.69), a set of concurvity data is able to be generated based on initial X_{1ij} and $f(X_{2ij})$. Suppose a new variable defined by

$$NEWX_{1ij} = X_{1ij} + K \times f(X_{2ij}), \quad (2.71)$$

and *K* is a numeric value which can control the concurvity level. When assigning *K*=0, 0.02, 0.05, 0.09, 0.13, 0.17, 0.22, 0.30, 0.41 and 0.64, the concurvity level between $NEWX_{1ij}$ and $f(X_{2ij})$ is 0.03, 0.10, 0.19, 0.31, 0.41, 0.50, 0.59, 0.70, 0.80 and 0.90, respectively. Each scenario with a specific concurvity level contains 1,000 simulated data set, and the average of $\hat{\beta}_1$, se($\hat{\beta}_1$), se($\hat{b}_1$) and ASMSE will be evaluated.

Some previous air pollution studies used to simulate data from real observations (Dominici, McDermott, Zeger, & Samet, 2002b; He, Mazumdar, & Arena, 2006), and we use a similar procedure to generate data from original PM_{10} and SO_2 concentrations. In order to facilitate the velocity of simulation, we restrict the study period in only 1991 from our database. Suppose a couple of principal component variables $Z = (PRIN_1\ PRIN_2)$ are calculated from original 1-year data of PM_{10} and SO_2 by PCA. Define a covariance matrix as

$$V = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix},$$

and a Cholesky decomposition can make it as $V = R'R$, where *R* is a upper triangular matrix. As a result, two correlated variables $W = (W_1\ W_2)$ with correlation coefficient

ρ can be generated by $W = Z \times R$. When using the two new variables to fit

$$g(\mu_{ct}) = \alpha + b_{0c} + \delta'(DOW) + (\beta_1 + b_{1c})W_{1,ct} + (\beta_2 + b_{2c})W_{2,ct} + f(TIME)$$

$$+f(TMEAN_t) + f_{spat}(s_c), \quad (2.72)$$

the prior prediction $\tilde{\mu}_{ct}$ can be estimated from (2.72). The number of 1,000 new responses $Y_{ct}^1, \dots, Y_{ct}^{1000}$ can be generated from a Poisson distribution $Poi(\tilde{\mu}_{ct})$. Each scenario repeats the above steps to generate its own simulated data, and the corresponding estimates can be evaluated from taking the average.

2.9. Model diagnostic methods

The model diagnostic methods of the GGAMM have still undeveloped, and related practical literature also did not discuss this part. We succeeded most of the applicable theorems from generalized linear mixed model here. A goodness-of-fit test for whether the respiratory mortality rate is Poisson-distributed data is identified by Pearson's chi-square test. Some preliminary model diagnostic methods, such as residual analysis, box plots of the residuals by level-1 units, scatter plots of standardized residuals versus fitted predictions and normal plot of residuals are provided for the fundamental detections of model fitting.

Chapter 3
RESULTS

3.1. Demographics

Table 3.1 and table 3.2 present the demographics tables of 15 U.S. cities from 1991 to 1995. Each city had a total 1826 records from respiratory mortality, 24-hour average temperature and PM_{10}. In details, there was no missing data in daily respiratory mortality, and the range of daily mortality average was from 0.23 persons (Huntsville, SD=0.48) to 7.75 persons (Chicago, SD=3.16). In 24-hour average temperature, most of the cities had no missing data besides Colorado Springs (34 missing data), and Detroit (1 missing datum). Las Vegas had the highest 24-hour average temperature with mean of 71.06°F (SD=18.06), and the lowest city was in Minneapolis/St. Paul (Mean=47.36°F, SD=22.3). Compared with respiratory mortality and 24-hour average temperature, PM_{10} relatively had more missing data over those cities during the study period. Pittsburg had most complete PM_{10} data with missing rate only 0.38%, but Las Vegas had almost 70% missing PM_{10}. The highest level of average PM_{10} concentration occurred in Cleveland (Mean=37.23$\mu g/m^3$, SD=20.04) and the lowest one was Minneapolis/St. Paul (Mean=22.68 $\mu g/m^3$, SD=13.1).

The profile of co-pollutants was also very different, especially in CO. The highest average daily concentration of CO was in Spokane, which was 3.23-fold the concentration of Lexington, the city with the lowest mean of CO with only 602.36 ppb daily. Moreover, the data collection of CO was more completed than the other co-pollutants, and 14 cities had less than 3% missing CO. In O_3, six cities had missing

Table 3.1

Demographics of respiratory mortality, temperature and PM_{10} in 15 U.S. cities from 1991 to 1995.

City	Respiratory mortality (person)				Temperature (°F)				PM_{10} ($\mu g/m^3$)			
	N	Missing %	Mean	SD	N	Missing %	Mean	SD	N	Missing %	Mean	SD
Chicago	1826	0.00	7.75	3.16	1826	0.00	50.97	19.85	1778	2.63	33.26	19.17
Cincinnati	1826	0.00	1.71	1.37	1826	0.00	56.12	18.35	1252	31.43	33.16	15.42
Cleveland	1826	0.00	2.28	1.57	1826	0.00	52.38	18.91	1705	6.63	39.8	19.80
Colorado Springs	1826	0.00	0.64	0.81	1792	1.86	49.41	17.24	1747	4.33	24.48	15.64
Detroit	1826	0.00	2.77	1.72	1825	0.05	52.18	19.36	1732	5.15	34.36	20.51
El Paso	1826	0.00	0.60	0.81	1826	0.00	69.36	15.59	1691	7.39	37.23	20.04
Huntsville	1826	0.00	0.23	0.48	1826	0.00	61.58	15.18	1051	42.44	23.35	10.61
Las Vegas	1826	0.00	1.43	1.23	1826	0.00	71.06	18.06	562	69.22	35.86	23.19
Lexington	1826	0.00	0.39	0.63	1826	0.00	57.66	17.79	1127	38.28	25.16	12.17
Minneapolis/St. Paul	1826	0.00	2.61	1.68	1826	0.00	47.36	22.30	1774	2.85	22.68	13.10
Nashville	1826	0.00	0.94	0.97	1826	0.00	61.18	16.53	1499	17.91	30.97	13.78
Pittsburgh	1826	0.00	2.91	1.84	1826	0.00	54.35	18.99	1819	0.38	31.51	20.25
Salt Lack City	1826	0.00	0.88	0.96	1826	0.00	54.59	19.41	1811	0.82	34.71	24.17
Seattle	1826	0.00	2.40	1.66	1826	0.00	54.67	11.10	1782	2.41	24.34	14.13
Spokane	1826	0.00	0.86	0.95	1826	0.00	49.82	17.15	1587	13.09	31.53	30.35

Table 3.2

Demographics of CO, NO_2, O_3, and SO_2 in 15 U.S. cities from 1991 to 1995.

City	CO (ppb)				NO_2 (ppb)				O_3 (ppb)				SO_2 (ppb)			
	N	Missing %	Mean	SD	N	Missing %	Mean	SD	N	Missing %	Mean	SD	N	Missing %	Mean	SD
Chicago	1826	0.00	783.26	316.09	1826	0.00	24.95	7.88	1826	0.00	19.38	9.62	1826	0.00	4.79	3.30
Cincinnati	1826	0.00	1044.62	358.64	1826	0.00	24.63	6.88	1070	41.40	25.09	9.44	1810	0.88	11.11	7.70
Cleveland	1826	0.00	850.09	381.70	1826	0.00	24.76	8.23	1070	41.40	27.96	11.64	1826	0.00	9.68	5.73
Colorado Springs	1826	0.00	1195.20	532.06	0	100.00	-	-	1826	0.00	23.57	9.48	0	100.00	-	-
Detroit	1826	0.00	629.50	321.25	1823	0.16	22.16	7.75	946	48.19	24.65	9.98	1826	0.00	6.79	4.16
El Paso	1826	0.00	1071.07	585.01	1813	0.71	17.8	8.67	1826	0.00	26.18	9.48	1817	0.49	8.33	8.74
Huntsville	1813	0.72	566.68	382.05	1090	40.31	13.33	5.43	1477	19.11	30.37	11.36	194	89.38	4.33	2.18
Las Vegas	1399	30.52	1382.4	1006.75	1459	20.10	27.59	12.52	1826	0.00	31.23	12.59	0	100.00	.	.
Lexington	1782	2.47	602.36	412.11	1790	0.19	16.49	6.71	1067	41.57	31.68	10.92	1811	0.82	7.04	4.78
Minneapolis/St. Paul	1822	0.22	1182.22	359.18	1770	3.07	19.30	7.50	0	100.00	-	-	1822	0.22	3.06	1.64
Nashville	1826	0.00	949.78	411.58	1719	5.86	13.96	7.51	1826	0.00	19.11	9.28	1818	0.44	9.80	6.46
Pittsburgh	1826	0.00	1038.55	531.36	1812	0.77	27.65	8.13	1803	1.26	21.71	11.50	1826	0.00	13.95	8.27
Salt Lack City	1826	0.00	1446.14	817.59	1174	35.71	26.80	11.90	1104	39.54	32.20	10.14	1826	0.00	5.69	5.84
Seattle	1826	0.00	1567.35	621.35	296	83.79	20.15	7.11	1071	41.35	21.92	7.60	0	100.00	-	-
Spokane	1826	0.00	1948.05	770.06	0	100.00	-	-	974	46.66	32.86	7.96	0	100.00	-	-

Table 3.3

Correlation coefficient matrix among 15 U.S. cities from 1991 to 1995.

Correlation coefficient	Respiratory mortality	PM_{10}	PM_{10}-lag1	PM_{10}-lag2	CO	NO_2	O_3	SO_2	Time	Temperature
Respiratory mortality	1.0000									
PM_{10}	0.0198	1.0000								
PM_{10}-lag1	0.0180	0.5376	1.0000							
PM_{10}-lag2	0.0168	0.3335	0.5375	1.0000						
CO	-0.0578	0.3450	0.1774	0.0784	1.0000					
NO_2	0.1915	0.4804	0.2464	0.0935	0.5838	1.0000				
O_3	-0.2045	0.1058	0.0981	0.0641	-0.2634	-0.1574	1.0000			
SO_2	-0.0684	0.3598	0.1894	0.0767	0.2758	0.3277	-0.1361	1.0000		
Time	0.0206	-0.0408	-0.0412	-0.0411	-0.0641	0.0672	0.0152	-0.1871	1.0000	
Temperature	-0.1967	0.2089	0.1769	0.1380	-0.1882	-0.0605	0.5768	-0.0151	0.0302	1.0000

rates over 40%, and Minneapolis/St. Paul was even completely missing. The situation became very extreme in SO_2 because it was either complete or total missing. Five cities had missing data over 89%, and four of them (Colorado Springs, Las Vegas, Seattle and Spokane) had SO_2 entirely missing. Even two cities (Colorado Springs and Spokane) have 100% missing rate in SO_2 and NO_2. The data collection of CO was quite good, and 14 cities had missing data less than 3%.

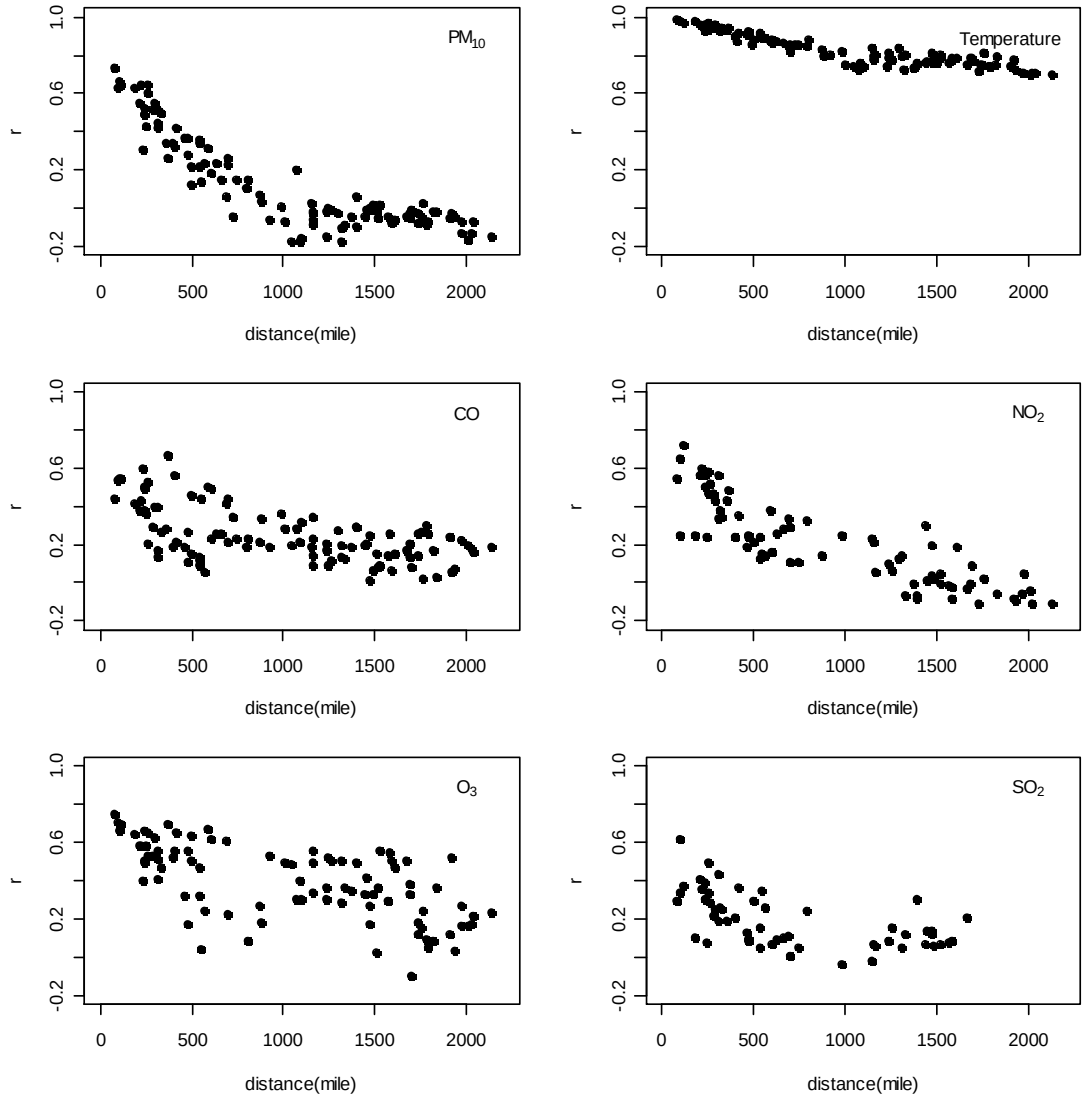
The correlation table is shown in table 3.3. From the relationship between PM_{10} with corresponding two lag effects and respiratory mortality, it shows that current PM_{10} effect has slightly more positive correlation with mortality than lag effects. In co- pollutants, NO_2 was the only co-pollutant having positive relationship to mortality. It was no doubt that PM_{10} and its lag effects were moderately correlated with each other, and provided the evidence that the desirability of using lag distributed models to identify the influence of lag effects to human health. PM_{10} also had slight correlation with O_3, SO_2, and CO, and had moderate correlation to NO_2 with the value of 0.4804. In particular, temperature had much higher correlation with O_3 that the other factors with correlation coefficient 0.5768.

3.2. Spatial and temporal correlation of air pollutants

Besides identifying linear relationship among these variables, we were also concerned with the spatial correlations of variables between any two cities and temporal correlations between any couple of time series variables. Figure 3.1 presents the plots of city-to-city correlation versus corresponding separation distance for each air pollutant and 24-hour average temperature. The definition of strong spatial correlation is that the city-to- city correlation has significant decrease along with the increase of distance

Figure 3.1

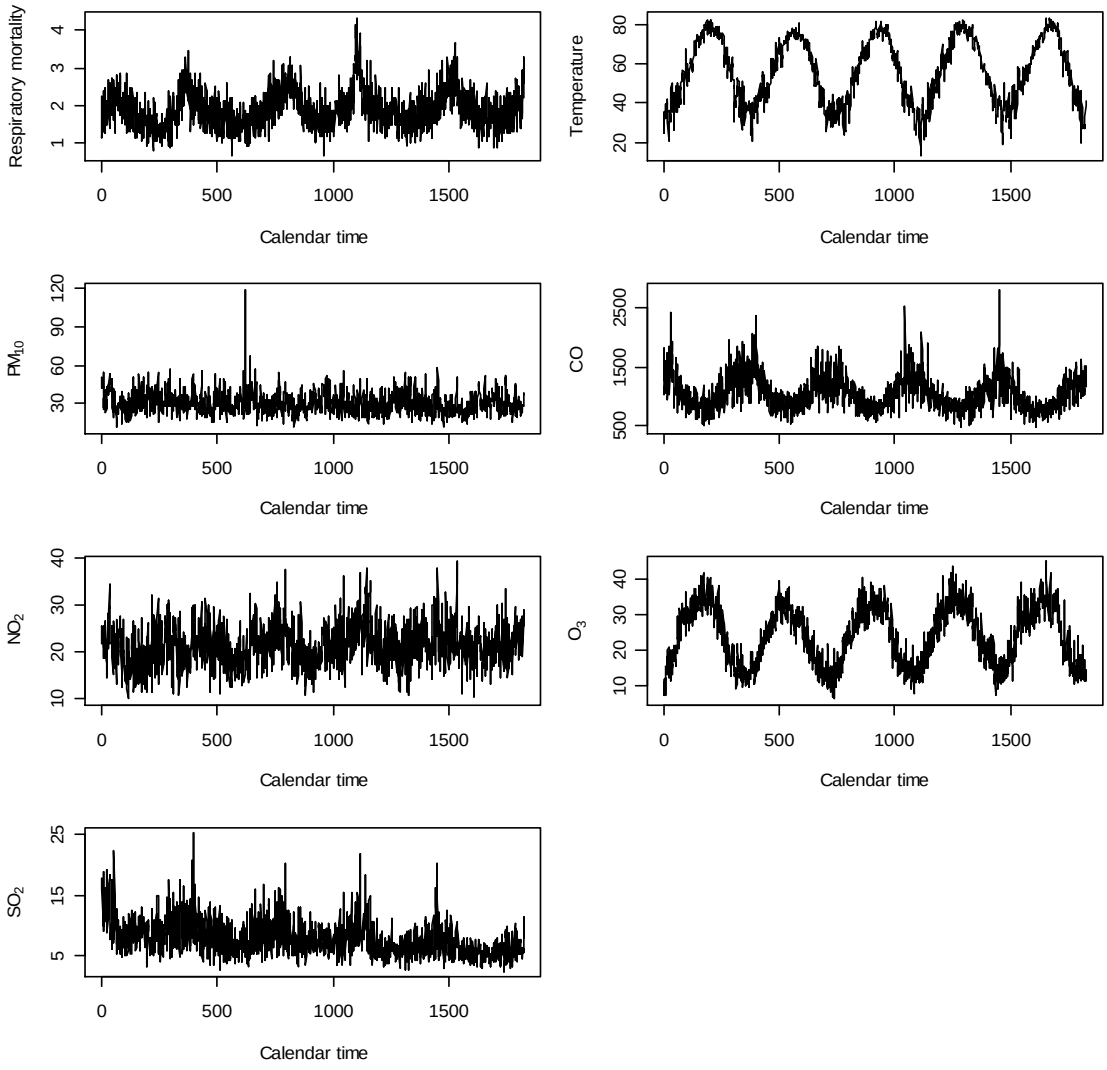
City-to-city correlation and separation distance among 15 U.S. cities from 1991 to 1995.



between two cities. PM_{10} showed the strongest city-to-city correlation. The shortest distance between two cities had the highest correlation coefficient with value of 0.73 (Cincinnati vs. Lexington), and the correlation declined quickly as the distance was getting longer. As long as the distance between two cities was longer than 1,000 miles, their correlations in PM_{10} were closed to zero or tiny negative. In 24-hour average temperature, the correlation was also decreasing along with the distance, but the falling speed was not as fast

Figure 3.2

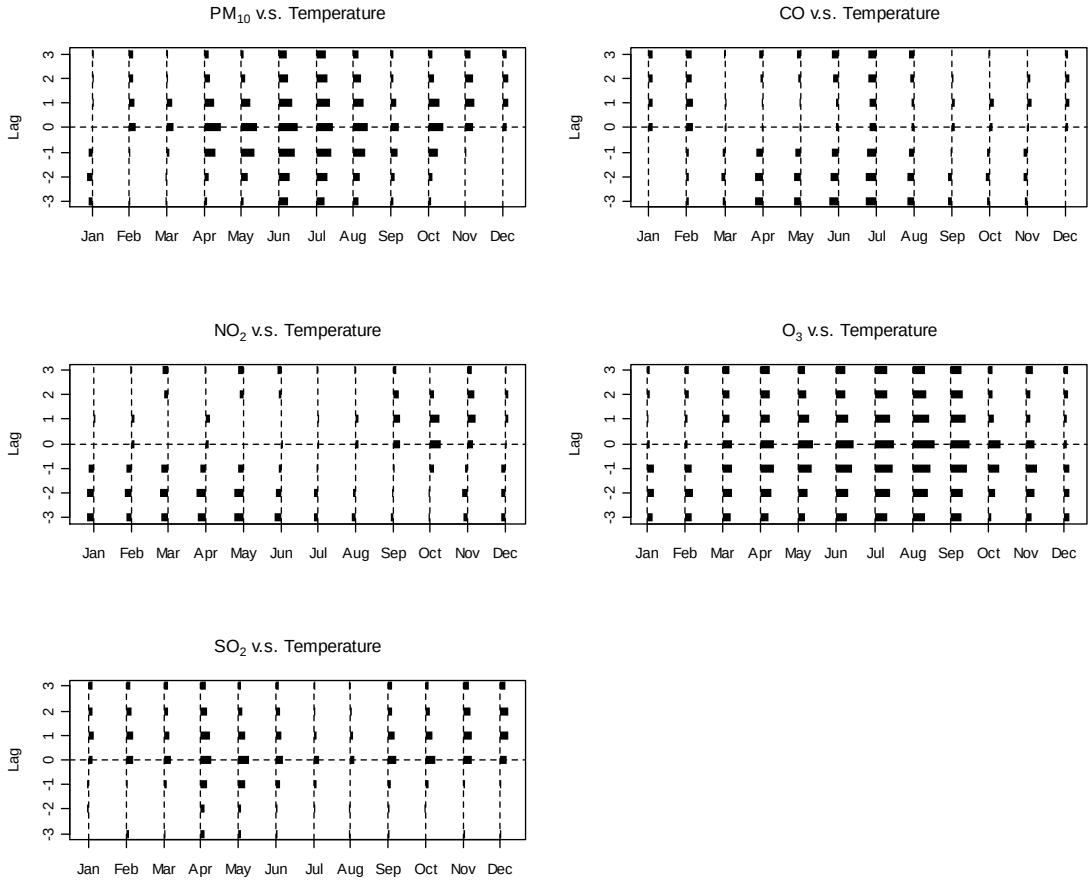
Time trend plot of respiratory mortality, 24-hour average temperature, PM_{10} and co-pollutants over 15 U.S. cities from 1991 to 1995.



as PM_{10}. The smallest correlation of temperature between two cities was still 0.70 when the distance was over 2,000 miles (Pittsburg vs. Seattle). The general declining trends of co-pollutants were similar as PM_{10}, but they were not as obvious as PM_{10}, especially in O_3. The highest city-to-city correlation (r=0.75) happened in Cincinnati and Lexington that the distance between these two cities was only 75.88 miles, but correlations were still as high as 0.6 even though the distance was over 1,500 miles.

Figure 3.3

Cross-correlation functions of air pollutants vs. temperature.



Moreover, we used the cross-correlation function (CCF) to handle the correlation coefficient between two time series variables. In table 3.2, we found that the correlation coefficients between time and the other variables were very small, but these values only indicate the other variables had no strong linear relationships with time. In reviewing these variables' time series plots in figure 3.2, we found that most of them showed an obvious seasonal trend over time. In details, both respiratory mortality and 24- hour average temperature showed significant seasonal variation, and co-pollutants also showed somehow regular fluctuation over time, such as CO and O_3.

Figure 3.4

Cross-correlation functions of co-pollutant vs. PM_{10}.

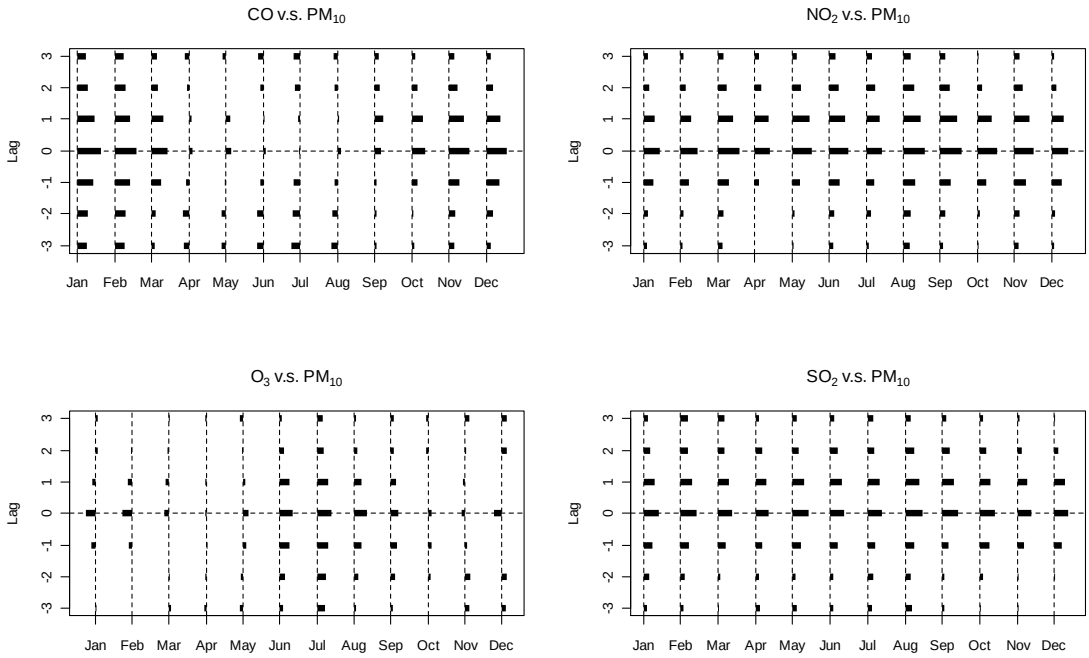


Figure 3.3 showed the CCFs for temperature versus air pollutants. The lag structure of association was generally not symmetric besides PM_{10} vs. temperature and O_3 vs. temperature in summer seasons. CO, NO_2, and SO_2 had relative smaller cross-correlation of temperature than PM_{10} and O_3, but SO_2 generally had more explicitly positive association with the following days' 24-hour average temperature than NO_2 and CO. PM_{10} obviously was positively associated with both previous and following days' 24-hour average temperature, especially from April to October, but its relationship with previous days' 24-hour average temperature almost disappeared in November and December. On the contrary, O_3 almost had both associations whenever any month it is, but the level decreases significantly in cold seasons (October to February).

Figure 3.4 presents the CCFs for co-pollutants and PM_{10}. The lag structure of association was almost symmetric in NO_2 vs. PM_{10} and SO_2 vs. PM_{10} in each month, but NO_2

vs. PM_{10} showed higher positive association than SO_2 vs. PM_{10}. However, both of them did not display stronger or weaker associations in specific months. Moreover, CO vs. PM_{10} and O_3 vs. PM_{10} did not have similar significant performance over months. In CO vs. PM_{10}, the stronger positive associations appeared in cold seasons, and weak negative associations were distributed from April to August. In particular, current associations from June to July were almost disappeared. Additionally, O_3 vs. PM_{10} showed positive association from June to September, but the other associations were very weak in the other months besides current association.

These results suggested that the correlation among air pollution and temperature had varying lag structure of associations, and the association can differ across air pollutants and seasons. Besides, these results also gave researchers a general sense that some potential problems could probably exist among those correlated factors in the GGAMM, such as multicollinearity and concurvity.

3.3. Case study of using the GGAMM by BayesX

Six model-fitting results are presented in table 3.4. In model 1, as the concentration of PM_{10} increased 10 $\mu g/m^3$, the relative risk of respiratory mortality in elders increased around 0.11% (95% CI: -0.46%, 0.67%). The virtualized smoothing and spatial functions are shown in figure 3.5. The time smoother went down in summer and fall seasons and rose in winter and spring seasons. The temperature smoother also reflected that extreme cold and hot weather can increase higher mortality rate. The lowest mortality rate happened at 16°F. As temperature decreased below 11°F or increased above 20°F, the mortality rate started to climb up, and the highest mortality rate appeared at 104°F.

Table 3.4

The parameter estimates with corresponding estimated standard errors of fixed and random effects in six main models.

Model	Variable	$\hat{\beta}$	se($\hat{\beta}$)	se($\hat{b}$)
Model 1	PM_{10}	0.000105	0.000287	0.000194
Model 2	PM_{10}	0.000675	0.196834	0.762327
	PM_{10}-lag1	-0.00053	0.133549	0.517211
	PM_{10}-lag2	-0.001911	0.063516	0.245960
Model 3	PM_{10}	0.000196	0.088258	0.341813
	CO	-0.000005	0.081648	0.316223
Model 4	PM_{10}	0.000163	0.000441	0.000638
	NO_2	0.001224	0.000846	0.000795
Model 5	PM_{10}	0.000227	0.000385	0.000531
	O_3	0.001855	0.000815	0.000650
Model 6	PM_{10}	0.000388	0.000391	0.000419
	SO_2	-0.000281	0.001830	0.004000

Moreover, BayesX can generate a map to reflect how spatial functions worked at different locations. In the map of figure 3.5, cities located around Northeast U.S. had higher geographical influence in respiratory mortality for elders. In addition, cities with higher altitude also had higher mortality rate, such as Salt Lake City (average altitude = 4,333 feet), Colorado Springs (average altitude =6,009 feet) and Spokane (average altitude=2,020 feet). On the contrary, some cities near to coast, lake and desert had lower influence, such as Seattle, Minneapolis/St. Paul and Las Vegas.

Considering the distributed lag model, the increases of relative risk in respiratory mortality in elders were 0.68%, -0.05% and -1.89% per 10 $\mu g/m^3$ increase in PM_{10}

Figure 3.5

Smoothing functions of calendar time and 24-hour average temperature and map of spatial effect for 15 U.S. cities from 1991 to 1995 in model 1.

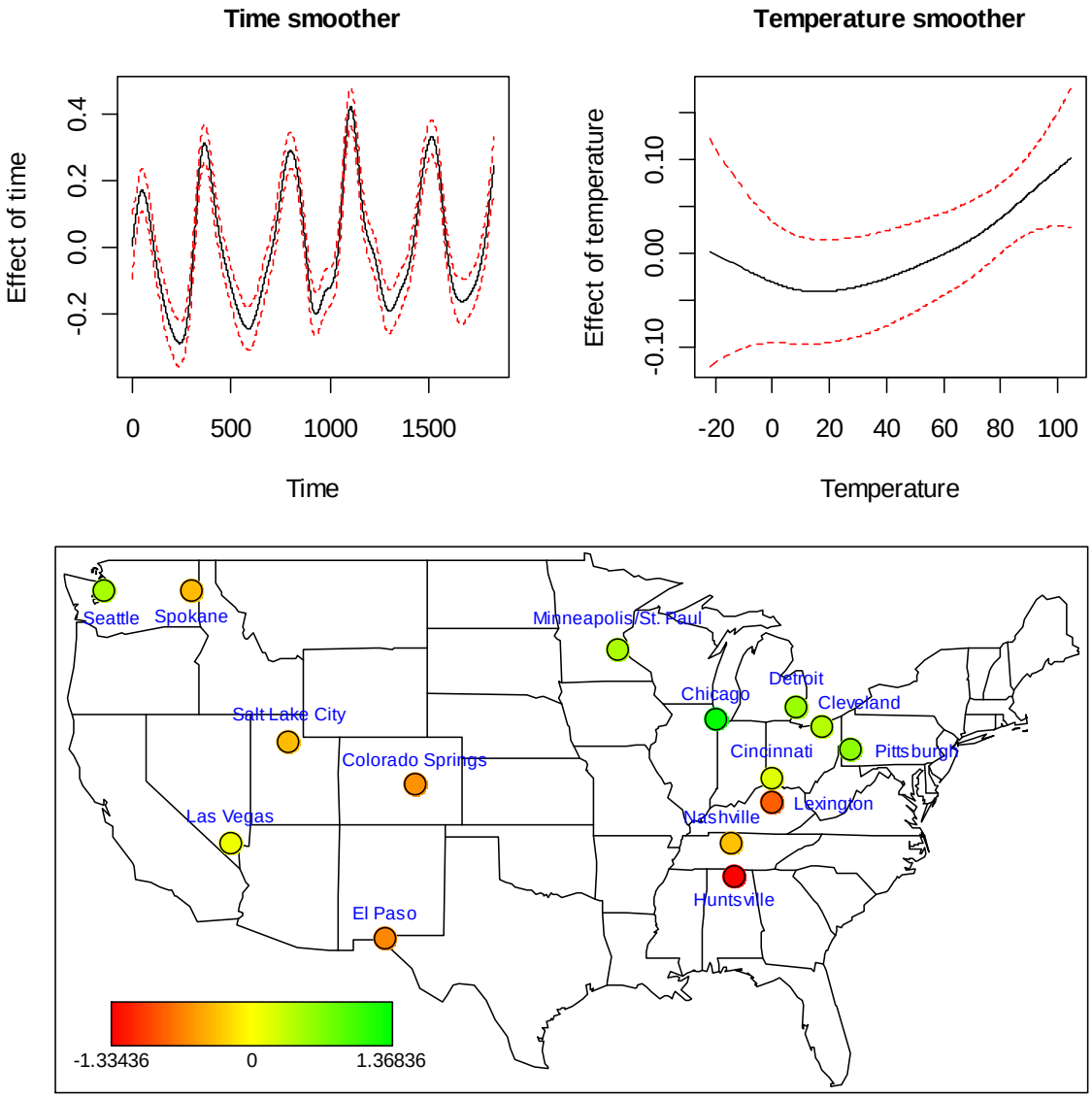


Figure 3.6

Smoothing functions of calendar time and 24-hour average temperature and map of spatial effect for 15 U.S. cities from 1991 to 1995 in model 2.

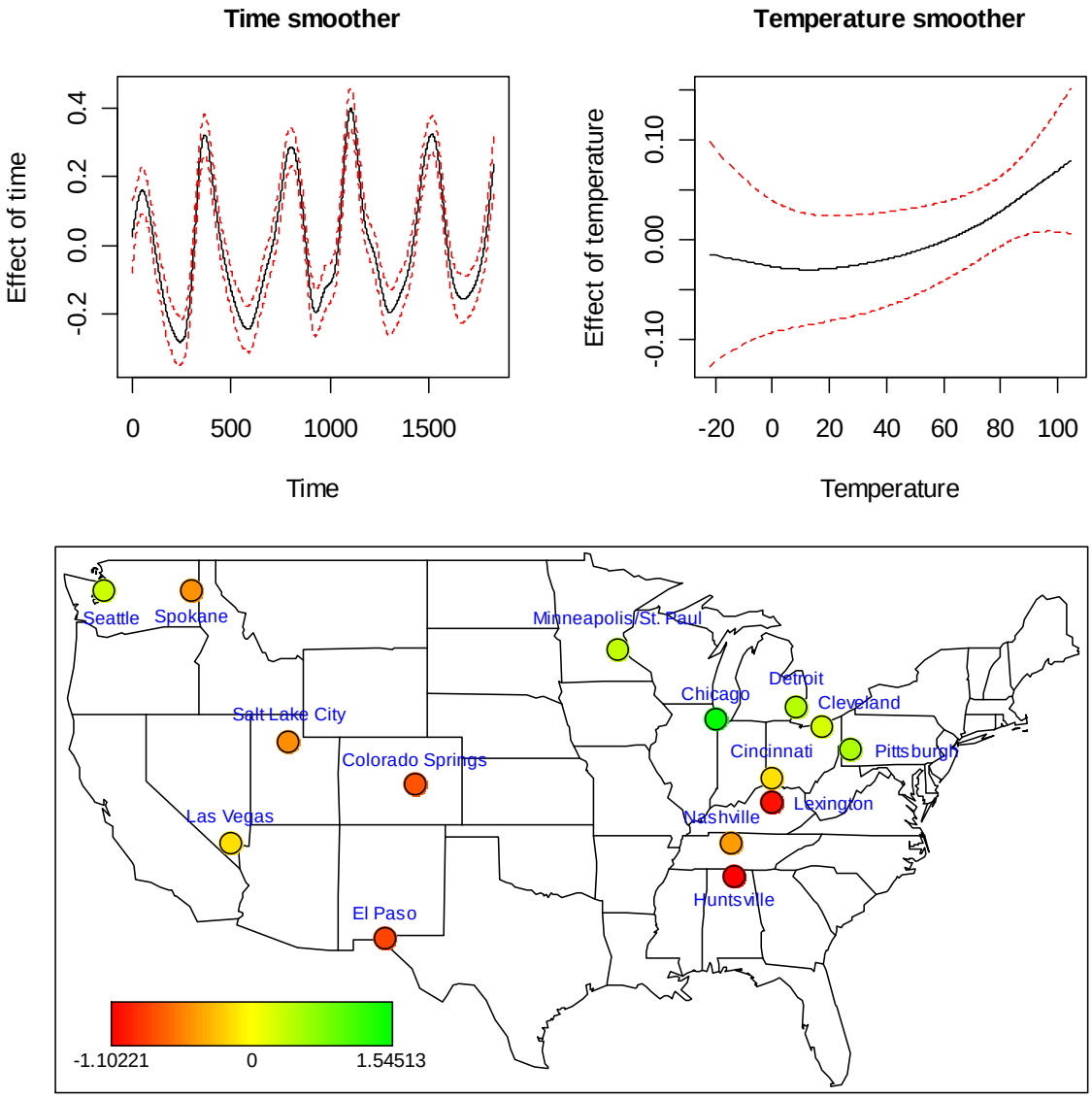


Figure 3.7

Smoothing functions of calendar time and 24-hour average temperature and map of spatial effect for 15 U.S. cities from 1991 to 1995 in model 3.

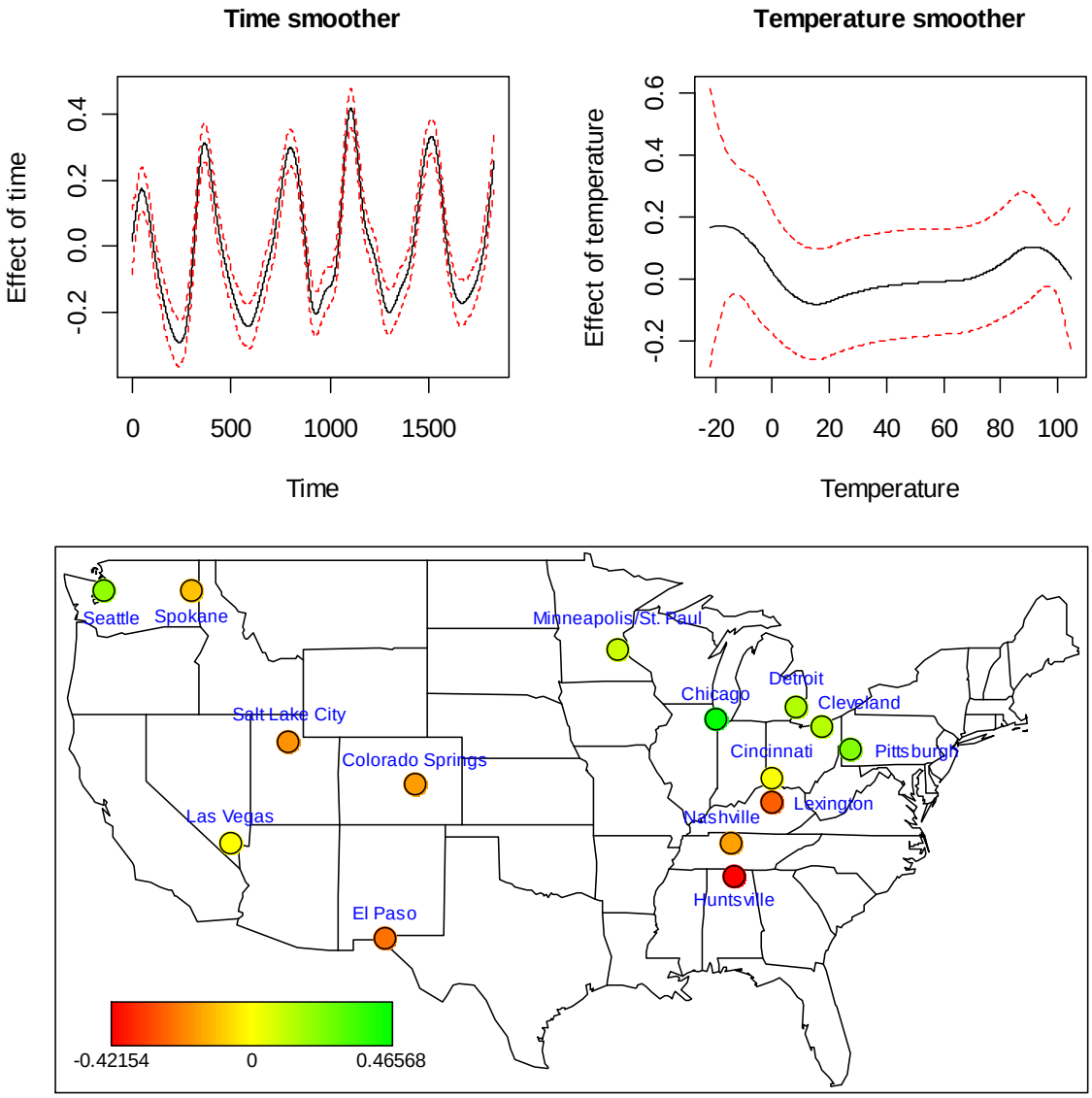
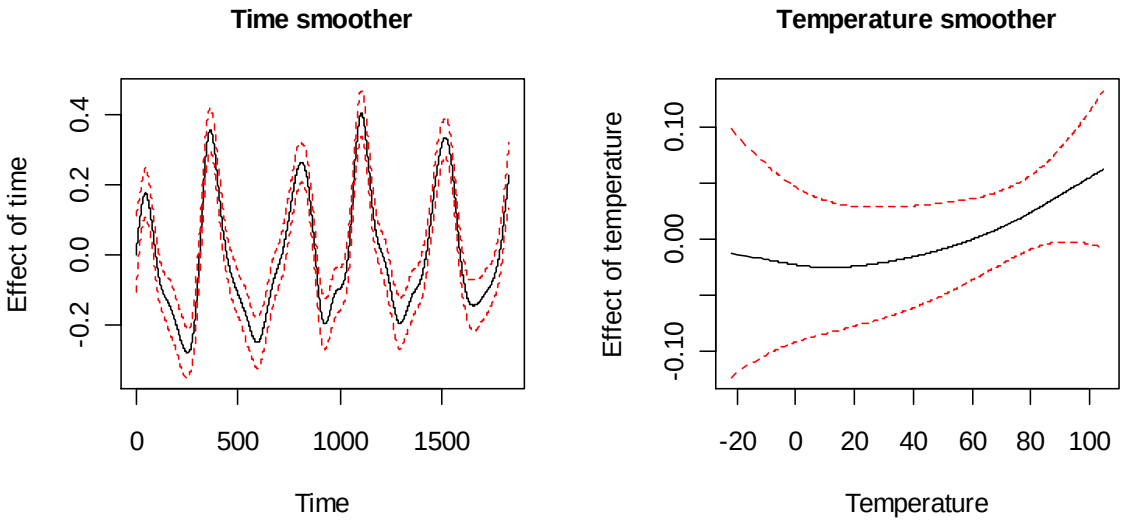


Figure 3.8

Smoothing functions of calendar time and 24-hour average temperature and map of spatial effect for 15 U.S. cities from 1991 to 1995 in model 4.



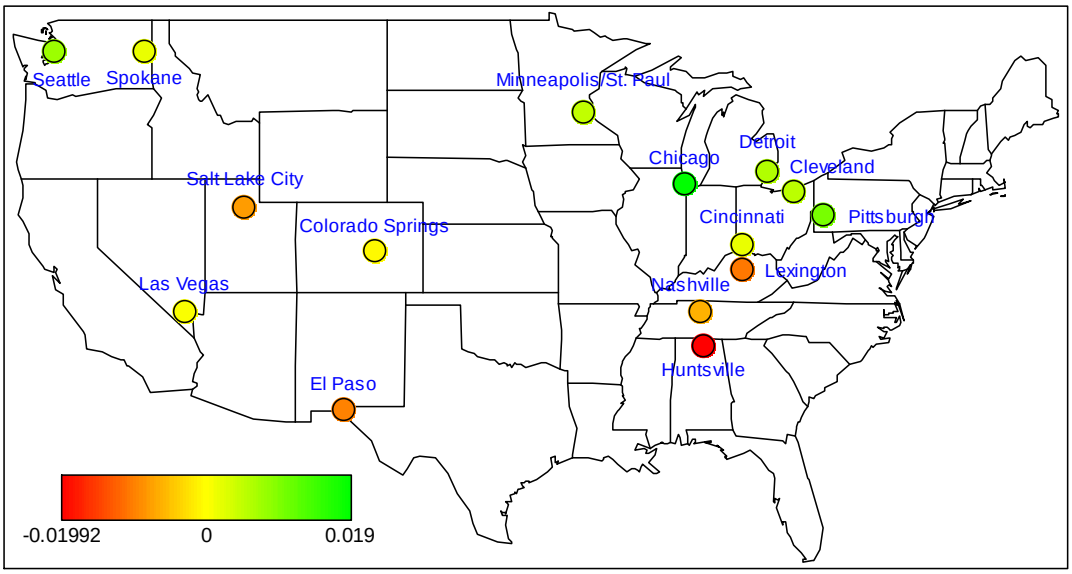


Figure 3.9

Smoothing functions of calendar time and 24-hour average temperature and map of spatial effect for 15 U.S. cities from 1991 to 1995 in model 5.

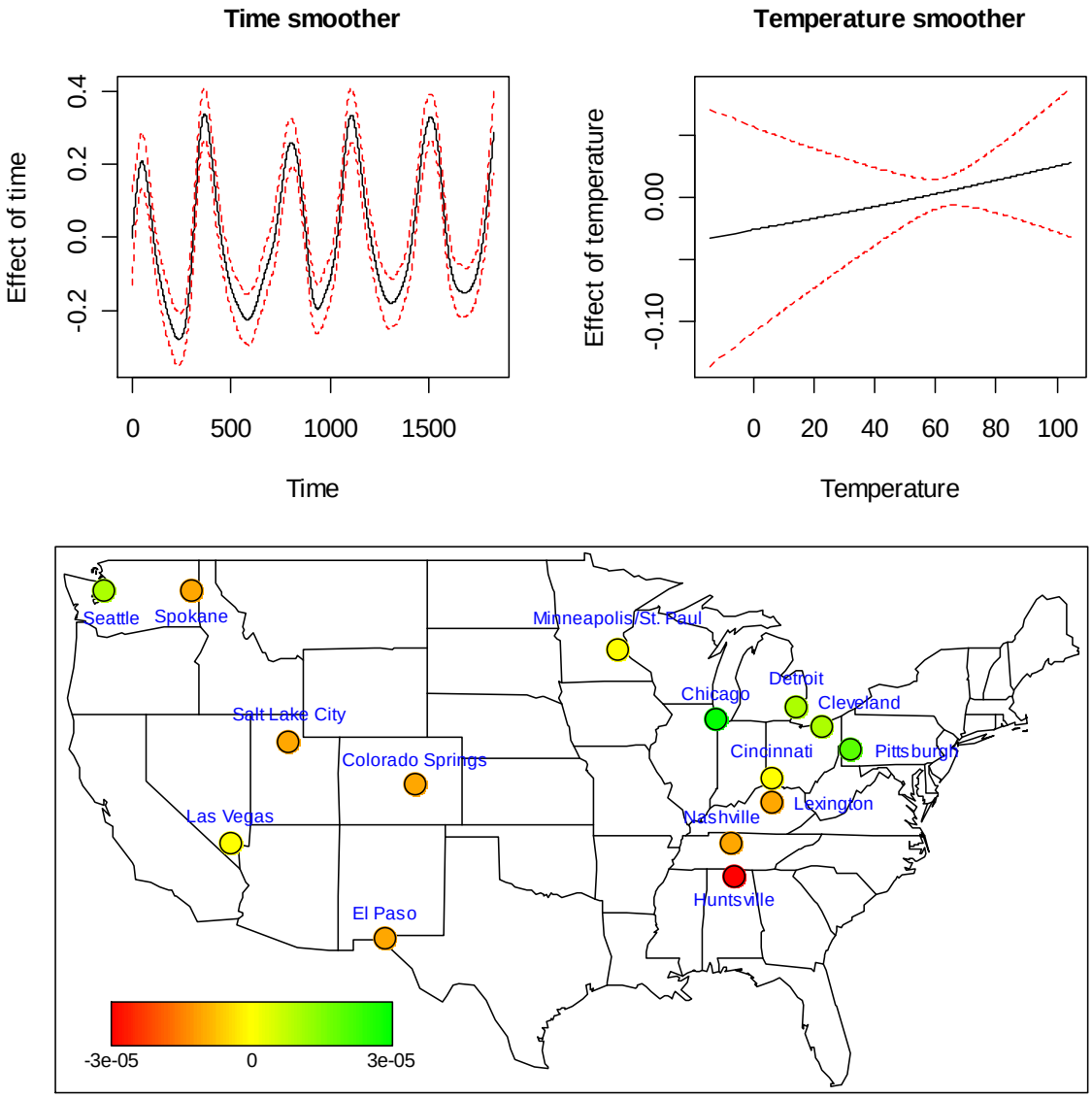
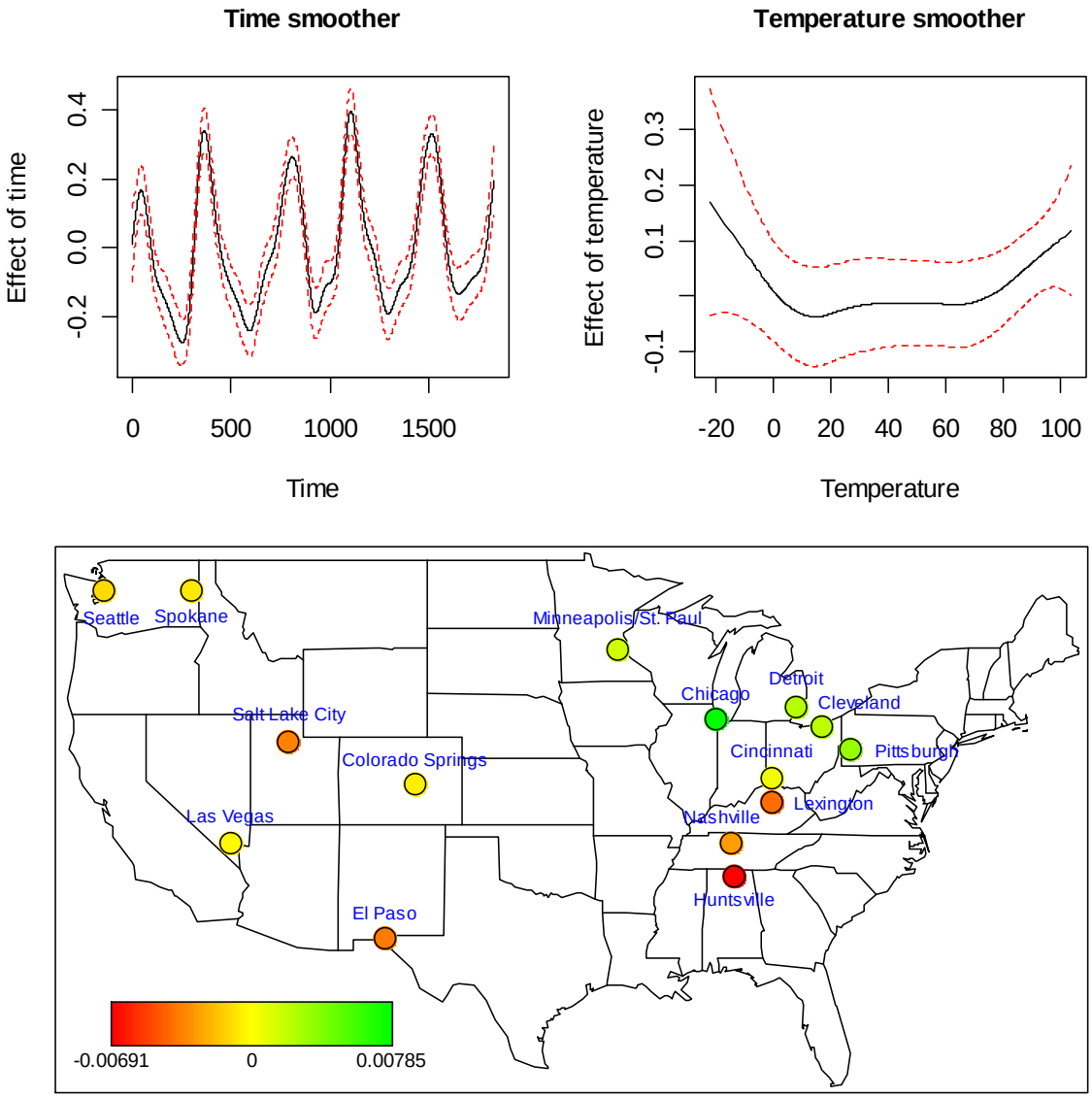


Figure 3.10

Smoothing functions of calendar time and 24-hour average temperature and map of spatial effect for 15 U.S. cities from 1991 to 1995 in model 6.



concentration at current day, 1-day lag and 2-day lag, respectively. Nonetheless, we found that the estimated standard errors of fixed and random effects (se($\hat{\beta}$) and se($\hat{b}$)) were exaggerated 221~600 times larger than those in model 1, and made confidence intervals too wide. In fact, this was a convergence problem especially happening in the GGAMM fitting by BayesX, and we will discuss more details and solutions about this problem in fixed and random effects in section 3.4~3.6.

Regardless of the problem of overestimated standard errors, we still can review the time and temperature smoothers and spatial function map in model 2. Its time smoother still had regular fluctuation as that in model 1, and temperature smoother displayed a slightly similar performance. However, temperature smoother did not have significant rise in left tail. The spatial function also displayed similar pattern between model 1 and model 2. Meanwhile, the relative spatial effects were still the same. Larger spatial effects in model 1 were still large in model 2, and vice versa. The highest spatial effect was still located in Chicago, and the lowest spatial effect was still located in Huntsville. However, based on the existence of convergence problem, this result was still uncertain. It was questionable that why the distributed lag model with similar profile of smoothing and spatial function had problematic estimates in standard errors of fixed and random effects. We will have an advanced analysis to remove this problem in section 3.11, and see how these estimates, smoothing functions and spatial map will perform.

In co-pollutant models, the percent relative risks of respiratory mortality increased 0.20% (95% CI: -172.82%, 173.22%), 0.16% (95% CI: -0.70%, 1.03%), 0.23% (95% CI: -0.53%, 0.98%) and 0.39% (95% CI: -0.38%, 1.15%) per 10 $\mu g/m^3$ PM_{10} increment for model 3, model 4, model 5 and model 6, respectively. Note that the results of estimated

standard errors in model 3 also reflected that this model existed a convergence problem, which also affected the confidence interval of the CO effect. The national effect of CO was associated with mortality rate with relative risk -0.01% when it increased per 10 ppb, but its 95% confidence interval was questionably over-wide as well. It will be combined with model 2 and discussed in section 3.4~3.6. Besides, each 10 ppb increase in the concentration of NO_2 can raise 1.23% (95% CI: -0.43%, 2.92%) relative risk of respiratory mortality in elders. O_3 also had a positive relationship with respiratory mortality rate in elders, and its relative risk significantly increased 1.87% (95% CI: 0.26%, 3.51%) as the concentration of ozone increased per 10 ppb. However, the national effect of SO_2 was negative to mortality rate. Compared with model 1, when adjusting by co-pollutants, the PM_{10} national effect increased 0.55-fold (model 4) to 2.70-fold (model 6).

The smoothing functions in calendar time in co-pollutant models shown in figure 3.7~figure 3.10 also presented very stable fluctuation, but the smoothing functions in 24-hour average temperature was quite varied among the four models. Except for problematic model 3, the temperature smoother of model 4 was almost identical as model 2's temperature smoother with a slight curve. Moreover, the temperature smoother in model 5 shown in figure 3.9 was no longer a curve but a straight line which was monotonic increasing from cold to hot. Its 95% confidence interval also tended to diverge toward the two ends of the tails. In model 6, extreme hot and cold weather had much influence on mortality rate, and presented a bowl-shape over temperature.

Considering geographical influence, in order to keep the same Markov random field and make these spatial function maps comparable, we retained the complete spatial data regardless that some cities contained 100% missing data in co-pollutants, such as Colorado

Springs (O_3, SO_2, NO_2), Las Vegas (SO_2), Minneapolis/St. Paul (O_3), Seattle (SO_2) and Spokane (SO_2, NO_2). This was a special property of BayesX in that missing data in any variable will not affect spatial function because geographical data and air pollutant/mortality/weather data were not stored in the same data set. They belonged to two separated files and imported in BayesX independently. Reviewing the general pattern of these maps, some cities with stronger spatial effects in model 1 and model 2 still had relatively stronger spatial effect in the other models, such as Pittsburg, Cleveland, Salt Lake City, Detroit and Cincinnati. Moreover, comparing the four co-pollutant models with each other, the distributions of spatial effect had significant shrinkage in model 4, model 5 and model 6, and concentrated to zero. Note that there was no test to identify whether a spatial function is statistically significant in the GGAMM, so no evidence can diagnose whether the spatial functions were significant or non-significant; nonetheless, we were still wondering whether the diminish of spatial effect in model 4, model 5 and model 6 was abnormal. To sum up, the virtualization of spatial function in BayesX still provides obvious and straightforward idea to present geographical versatility in spatio-temporal analysis. More detailed investigations of the spatial function will be discussed in section 3.8.

3.4. Convergence problem in the GGAMM and the smoothing parameter λ in smoothing functions

From those results in previous sections, we found there were some irrational values appearing in estimated standard errors of either fixed effects or random effects. The characteristic of those irrational estimated standard errors was that they were overestimated with relative huge values compared to normal values. In air pollution research, reasonable

estimated standard errors of fixed and random effects were less than 0.01, but sometimes they can be estimated over 0.1 abnormally, even over 1.

By checking the log file, we found the main reason was that the iteration of estimating smoothing parameter λ did not reach convergence or reach convergence too earlier. By the way, the default number of iteration in BayesX was 400. When the number of iteration reached 400, BayesX showed a warning message to alert users that this estimation does not reach convergence, and followed with results stopping by the last iteration. Unfortunately, there was no way to increase the number of iteration in BayesX, so the only way was setting up another starting value of λ to expect reaching convergence before 400 iterations appropriately. The default setting of starting value of λ was 10 in BayesX, so users can define any value which is larger than 10 in programs. The modified starting value of λ can be enlarged to 1,000 if necessary. The purpose of adjusting the smoothing function in penalized splines is to facilitate the speed of convergence in iterations of estimating unknown parameters; however, it was not guaranteed that larger smoothing functions can absolutely facilitate iterations. Sometimes, larger smoothing functions also had worse efficacy. Besides, there was no efficient approach to determine the most efficient smoothing functions in the GGAMM, so users have to try manually.

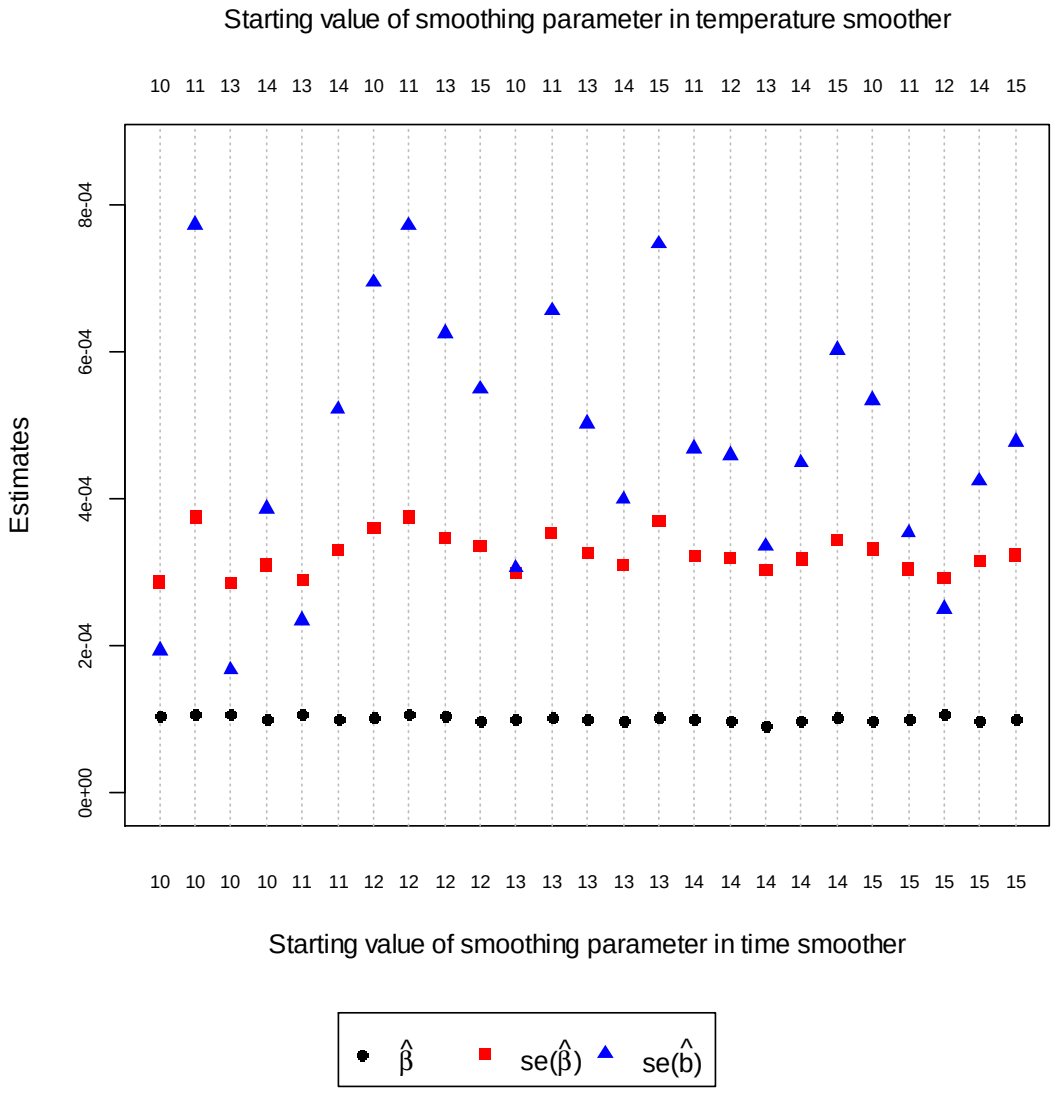
In table 3.4, the estimated standard errors of fixed effects and random effects in model 2 and model 3 had irrationally huge values, so we used λ=11, 12, 13, 14 and 15 in both the time smoother and temperature smoother. The entire results are presented in Appendix A. We found that, if a GGAMM can reach convergence with initial starting value of λ (i.e. λ=10), the probability of reaching convergence with other starting values of λ was higher. For example, each model had 36 trials with 6 different starting values of λ_{time} and 6 starting

values of λ_{tmean}. Model 1 and three co-pollutant models (PM_{10}+NO_2, PM_{10}+O_3 and PM_{10}+SO_2) reached convergence in initial value (λ_{time}, λ_{tmean})=(10,10), so they had 25, 20, 19 and 25 successfully convergent results out of 36 trials, respectively. However, model 2 only had 9 successfully convergent results out of 36 trials, and model 3 even still had no convergent results. Figure 3.11 was constructed from 25 successfully convergent results in model 1. When the user-defined starting values of λ_{time} and λ_{tmean} made iteration convergent, the estimated fixed effect of PM_{10} ($\hat{\beta}$) ranged from 0.0000951 to 0.0001066, and the differences of the effect to the relative risk of mortality over these 25 estimates were only around 0.01%. The se($\hat{\beta}$)s were also estimated stably, especially when λ_{time} = 12, 13, 14 and 15 and λ_{tmean}=10, 11, 13, 14 and 15. Compared with initial se($\hat{\beta}$)(=0.000287), most of the other se($\hat{\beta}$)s were a little bit higher, but the differences were no more than 0.00001. se($\hat{b}$) had much instability than $\hat{\beta}$ and se($\hat{\beta}$). The initial se($\hat{b}$) was close to 0.0002, but it also can be estimated more than 3-folds of initial se($\hat{b}$), especially when starting value of was λ_{time} 12 or 13.

Figure 3.12 to figure 3.14 showed all estimates from different starting values of λ_{time} and λ_{tmean} in model 2. Actually, we did not see any specific λ_{time} or λ_{tmean} which had higher probability to reach convergence. Among these nine convergence results, the estimated PM_{10} fixed effects $\hat{\beta}_1$ were close to 0.0001, which was much smaller than the other non-convergence results. Their corresponding se($\hat{\beta}_1$)s and se($\hat{b}_1$)s were also estimated consistently with reasonable values from 0.000345 to 0.000545. Comparing with the other non-convergence results, we found that the two estimated standard errors can almost reach 3.9466, which was over 10,000-folds than convergence estimates. In addition, the estimated 1-day lag PM_{10} fixed effects $\hat{\beta}_2$ from convergence results were from -0.0000628 to

Figure 3.11

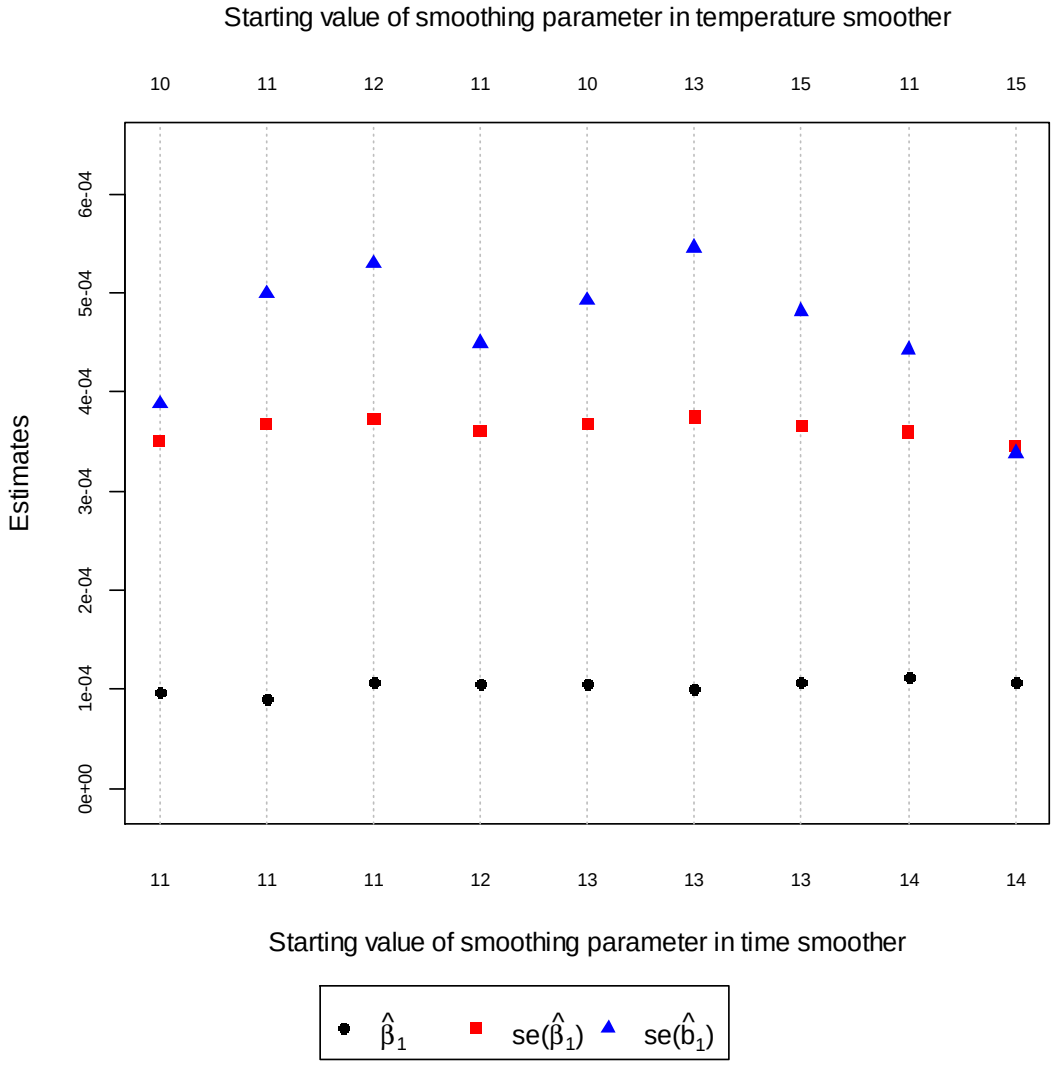
The estimated PM_{10} fixed effect with corresponding estimated standard errors of fixed and random effect from 25 convergence results using different starting values of λ_{time} and λ_{tmean} in model 1.



-0.0000314. From those non-convergence results, $\hat{\beta}_2$ could be much more underestimated than convergent results. Convergent se($\hat{\beta}_2$)s and se ($\hat{b}_2$) were close to convergent se($\hat{\beta}_1$) and se($\hat{b}_1$). An interesting finding was in estimates of 2-day lag PM_{10} effect. If the starting values of either λ_{time} or λ_{tmean} can hardly make iteration reach convergence, the 2-day lag PM_{10} fixed effect $\hat{\beta}_3$ would be estimated negatively, but when it was estimated with convergence,

Figure 3.12

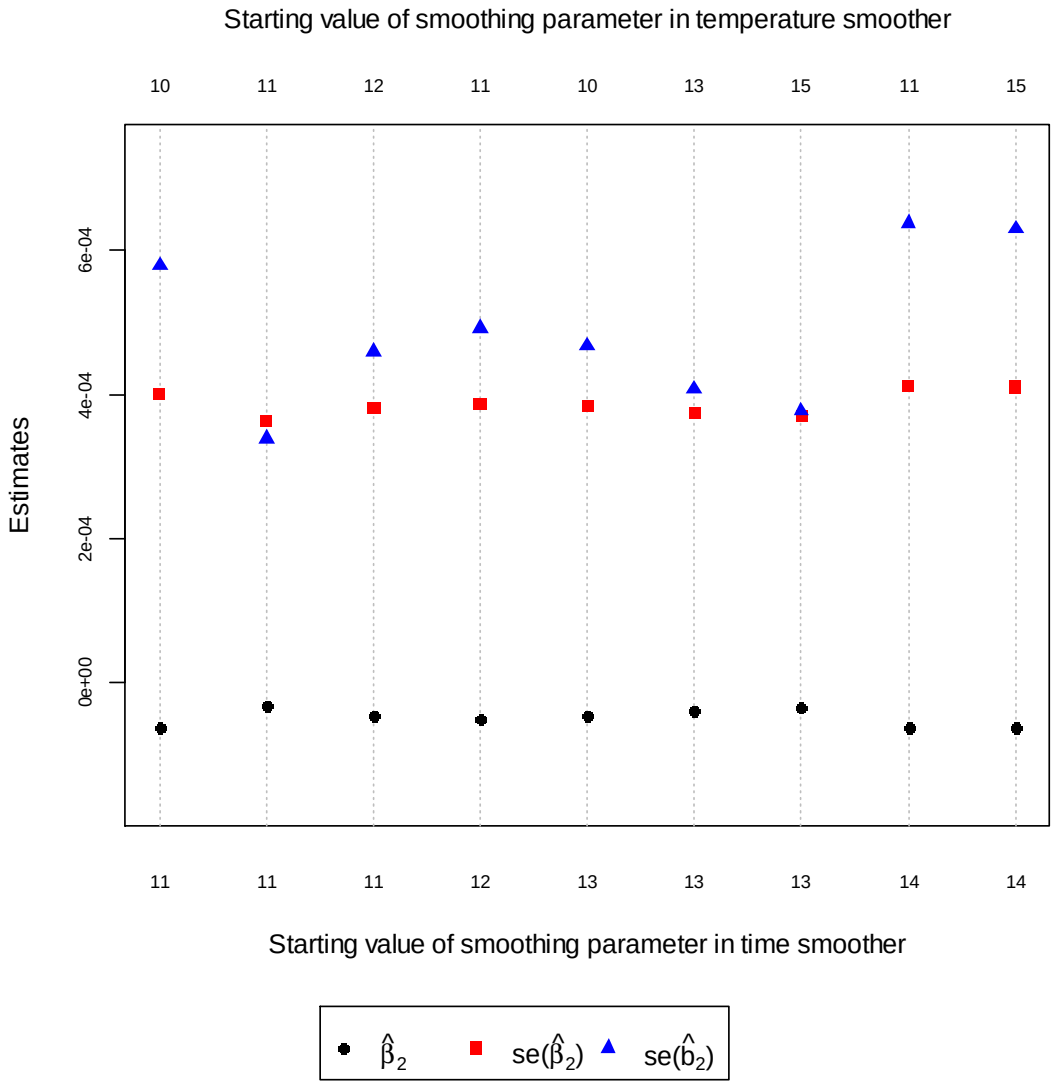
The estimated PM_{10} fixed effect with corresponding estimated standard errors of fixed and random effect from 9 convergence results using different starting values of λ_{time} and λ_{tmean} in model 2.



$\hat{\beta}_3$ would be estimated positively with value from 0.000114 to 0.000157. Its standard error se($\hat{\beta}_3$) was also reasonable and consistently estimated around 0.000318 to 0.000388, but se($\hat{b}_3$) was widely distributed from 0.000244 to 0.000684. Note that larger estimated standard errors in random effect can reflect the higher versatility level of city-level effect.

Figure 3.13

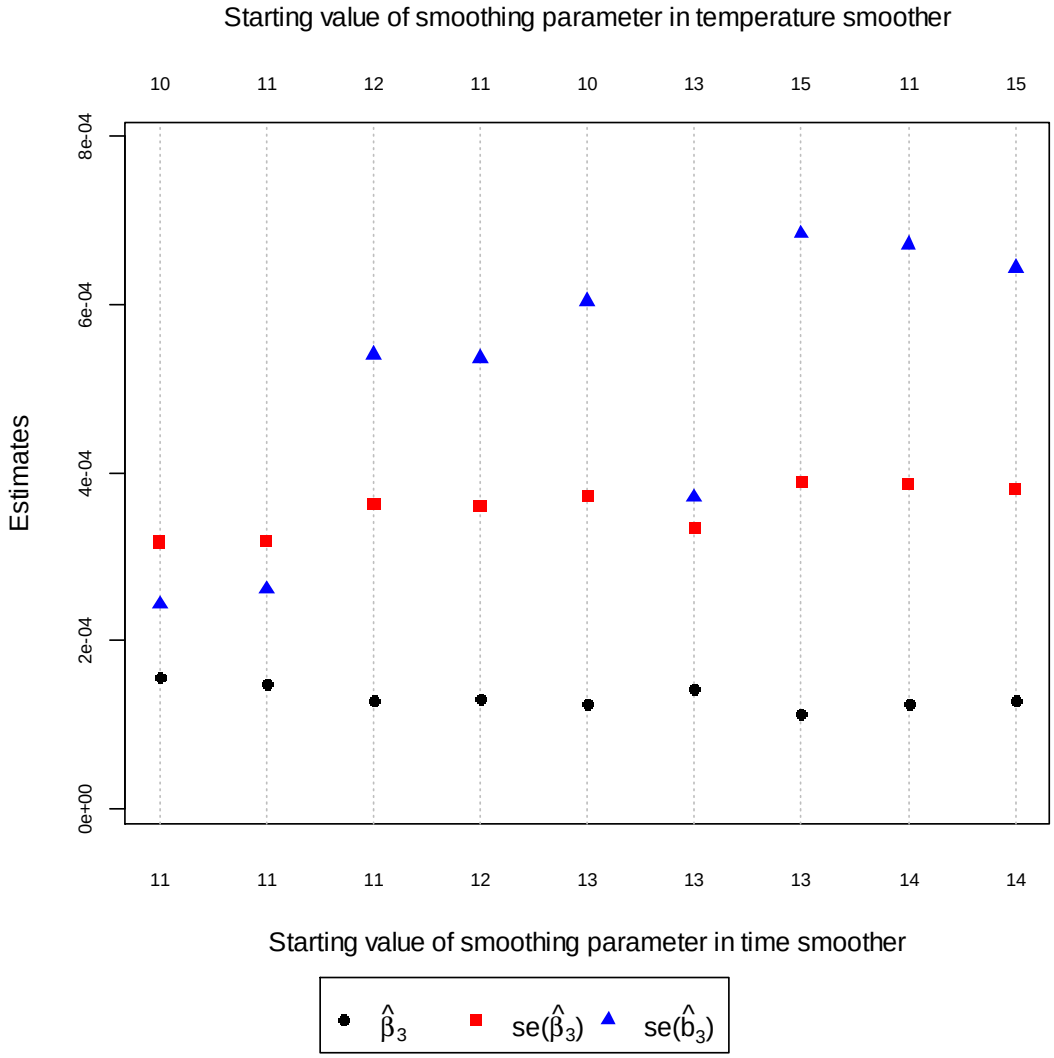
The estimated 1-day lag PM_{10} fixed effect with corresponding estimated standard errors of fixed and random effect from 9 convergence results using different starting values of λ_{time} and λ_{tmean} in model 2.



When considering co pollutant NO_2 in model 4, all estimates had higher opportunities to reach convergence when λ_{time} was 11 and 13, or λ_{tmean} was 12 and 13. As long as iteration cannot end with convergence, most estimated PM_{10} fixed effects $\hat{\beta}_1$s were underestimated with values around 0.05. Sometimes $\hat{\beta}_1$ could be overestimated to 0.4((λ_{time}, λ_{tmean})=(12, 10), (12, 14), (13, 11) and (14, 14)), but from those convergence results, as what figure 3.15

Figure 3.14

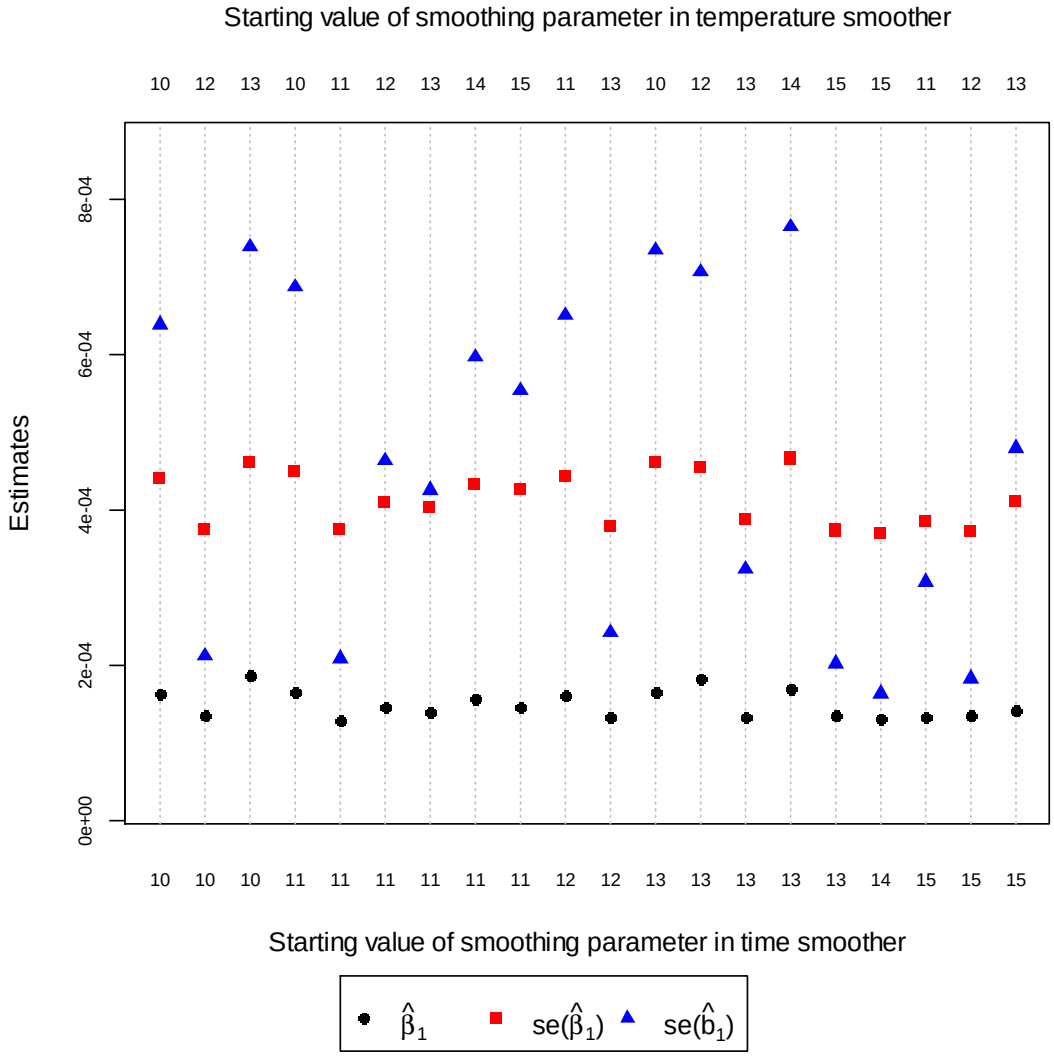
The estimated 2-day lag PM_{10} fixed effect with corresponding estimated standard errors of fixed and random effect from 9 convergence results using different starting values of λ_{time} and λ_{tmean} in model 2.



showed, a reasonable $\hat{\beta}_1$ should be around 0.000130 to 0.000188. The convergent se($\hat{\beta}_1$)s were close to 0.00041 ± 0.00005, but convergent se($\hat{b}_1$)s were not as consistent as se($\hat{\beta}_1$). The largest convergent se($\hat{b}_1$) can be 4.65-fold of the smallest convergent se($\hat{b}_1$). The pattern of estimates related to NO_2 was similar as the pattern of estimates related to PM_{10} (figure 3.16). Comparing with $\hat{\beta}_1$ and $\hat{\beta}_2$, we found the fixed effect of NO_2 would be much stronger that

Figure 3.15

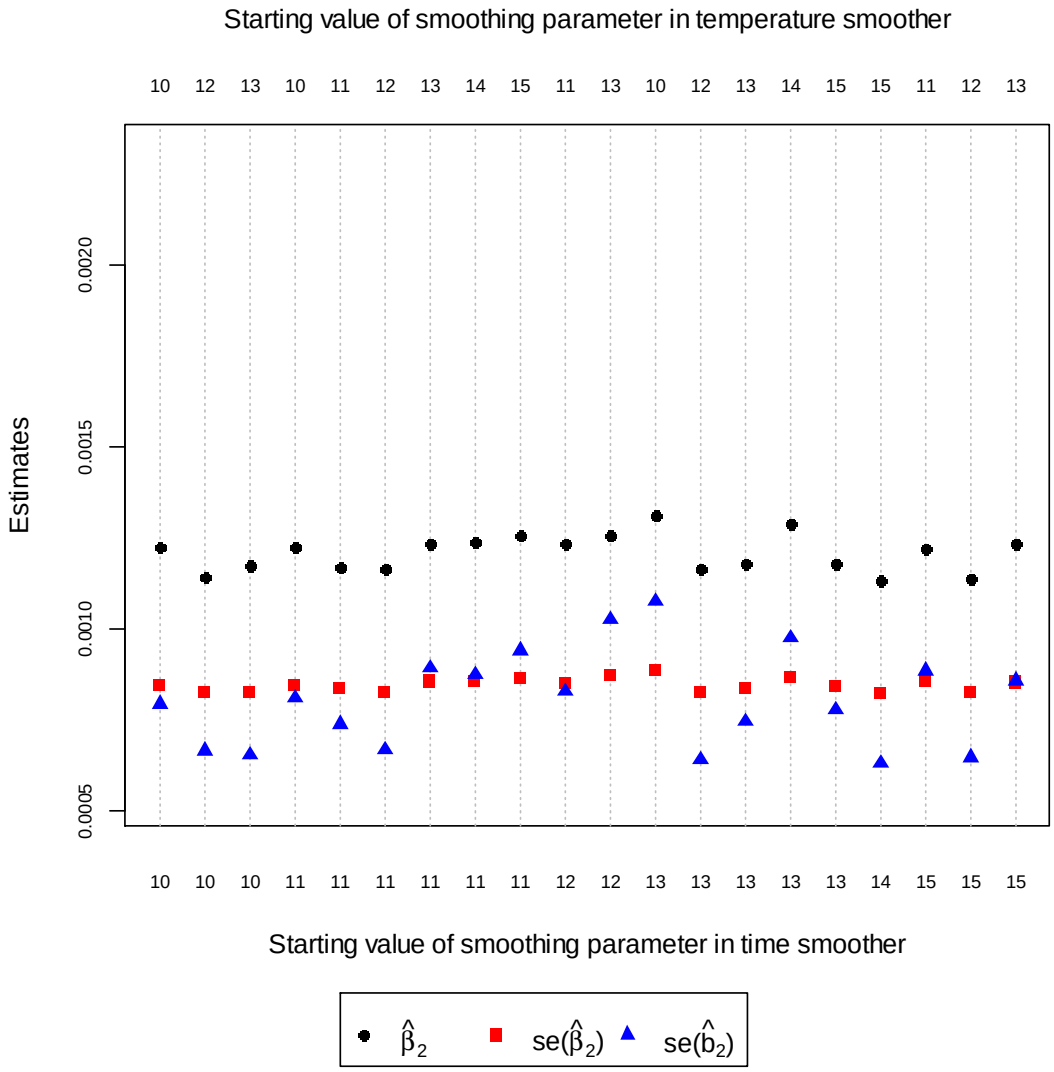
The estimated PM_{10} fixed effect with corresponding estimated standard errors of fixed and random effect from 20 convergence results using different starting values of λ_{time} and λ_{tmean} in model 4.



the fixed effect of PM_{10} when they did not reach convergence. For example, when the starting value of λ_{time} was 10, the convergent $\hat{\beta}_2$s (λ_{tmean}=10, 12 and 13) were averagely 7.39-fold of the convergent $\hat{\beta}_1$s; on the contrary, the non-convergent $\hat{\beta}_2$s (λ_{tmean}=11, 14 and 15) were averagely 27.06-fold of the convergence $\hat{\beta}_1$s. Both convergent se($\hat{\beta}_2$)s and se($\hat{b}_2$)s were concentrated on 0.0008, but se($\hat{\beta}_2$)s were more consistent with se($\hat{b}_2$)s.

Figure 3.16

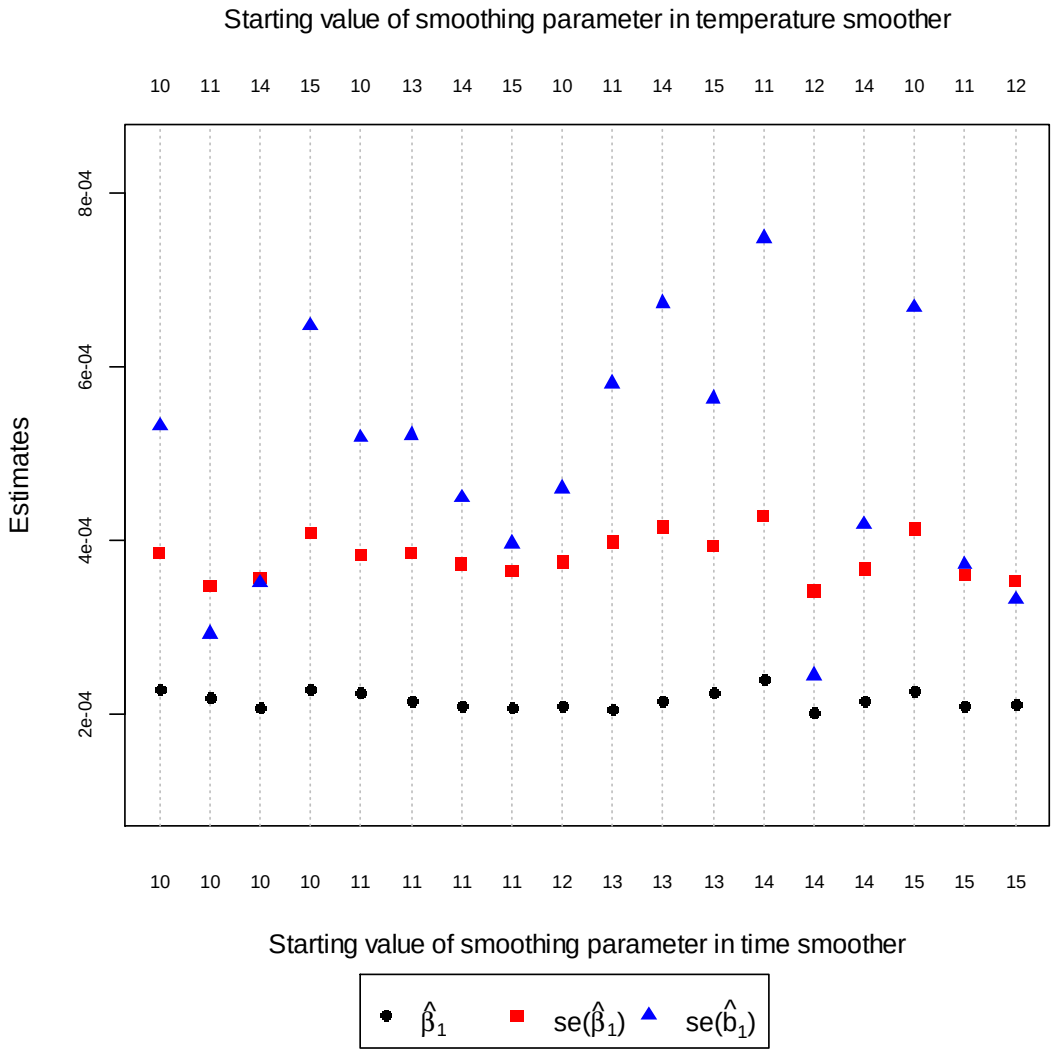
The estimated NO_2 fixed effect with corresponding estimated standard errors of fixed and random effect from 20 convergence results using different starting values of λ_{time} and λ_{tmean} in model 4.



The probability of reaching convergence in model 5 was similar as model 4 over 36 trials. When convergence existed, both of PM_{10} and O_3 fixed effects ($\hat{\beta}_1$ and $\hat{\beta}_2$) were estimated around 0.0002 and 0.0017, respectively. In particular, when this model cannot be fitted with convergence, the value of $\hat{\beta}_1$ became negative, and the value of $\hat{\beta}_2$ shrunk to half of convergent $\hat{\beta}_2$. The convergent se($\hat{\beta}_1$)s and se($\hat{\beta}_2$) were located consistently, where

Figure 3.17

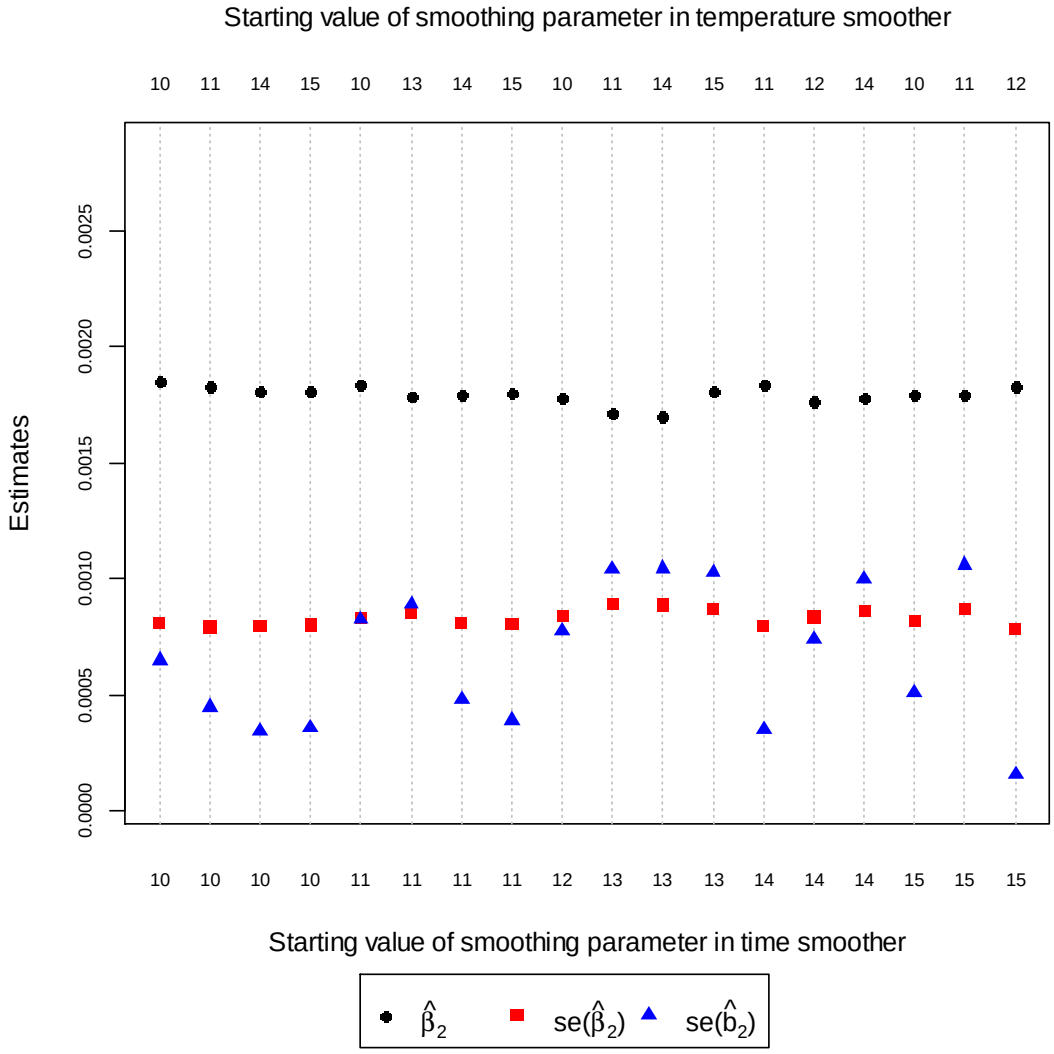
The estimated PM_{10} fixed effect with corresponding estimated standard errors of fixed and random effect from 19 convergence results using different starting values of λ_{time} and λ_{tmean} in model 5.



se($\hat{\beta}_1$) was 0.00038$\pm$0.00005, and se($\hat{\beta}_2$) was 0.00083$\pm$0.00005, but the convergent se($\hat{b}_1$) and se($\hat{b}_2$) changed relatively significantly. However, the relative large se($\hat{b}_1$) and se($\hat{b}_2$) should not be over 0.0008 and 0.0012, respectively when model fitting reached convergence. Meanwhile, as long as model fitting cannot reach convergence, both se($\hat{b}_1$) and se($\hat{b}_2$) would be definitely estimated with values much over 0.0008 and 0.0012.

Figure 3.18

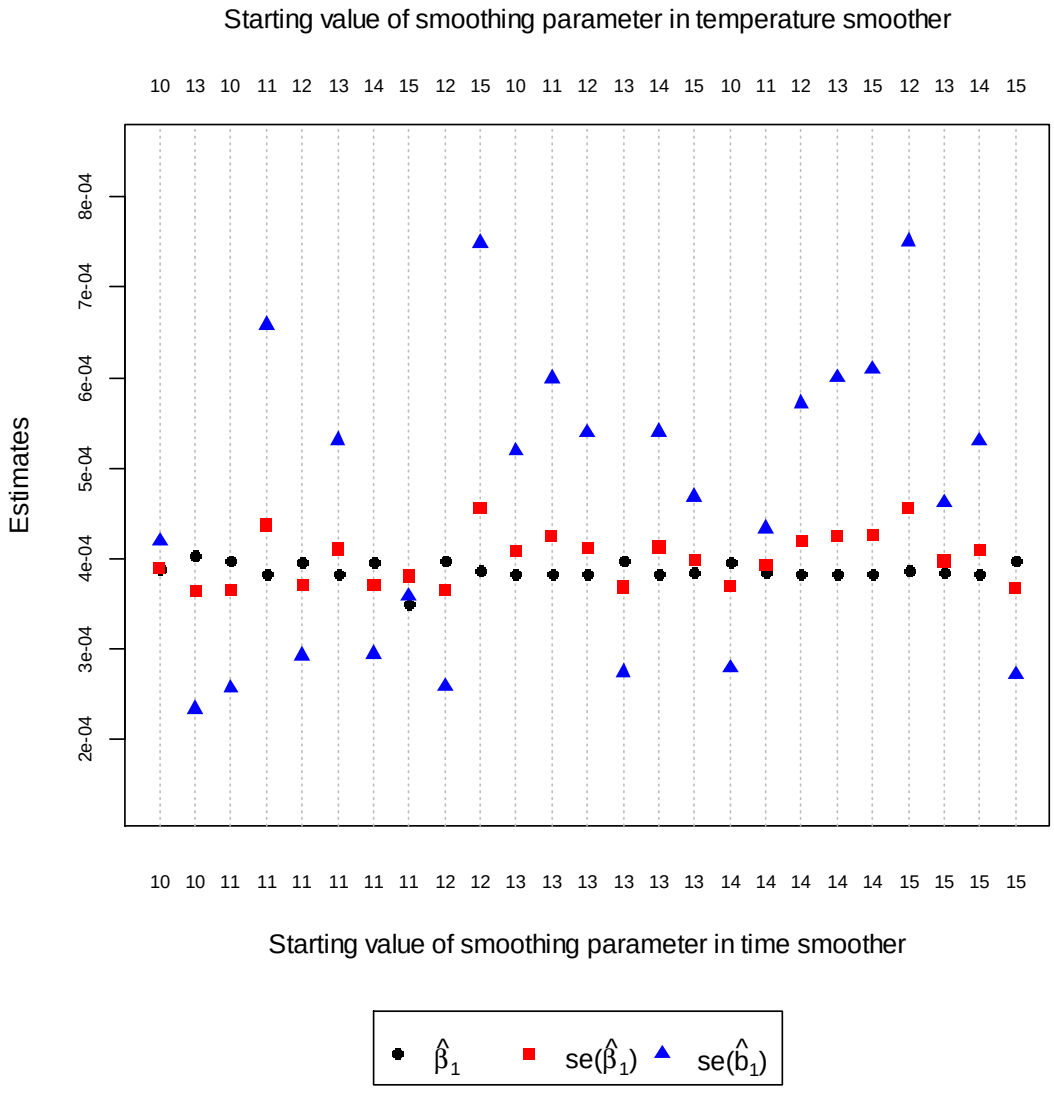
The estimated O_3 fixed effect with corresponding estimated standard errors of fixed and random effect from 20 convergence results using different starting values of λ_{time} and λ_{tmean} in model 5.



SO_2 was the only negative effect in co-pollutants, and all non-convergent results did not happen in SO_2's estimates. Meanwhile, the convergence problem only appeared in PM_{10} in this model, especially in its standard error of fixed and random effect. If model-fitting was convergent, $\hat{\beta}_1$ and $\hat{\beta}_2$ were close to 0.0004 and -0.0003, respectively. In addition, convergent se($\hat{\beta}_2$)s were larger than the other convergent se($\hat{\beta}_2$) in the other co-pollutant

Figure 3.19

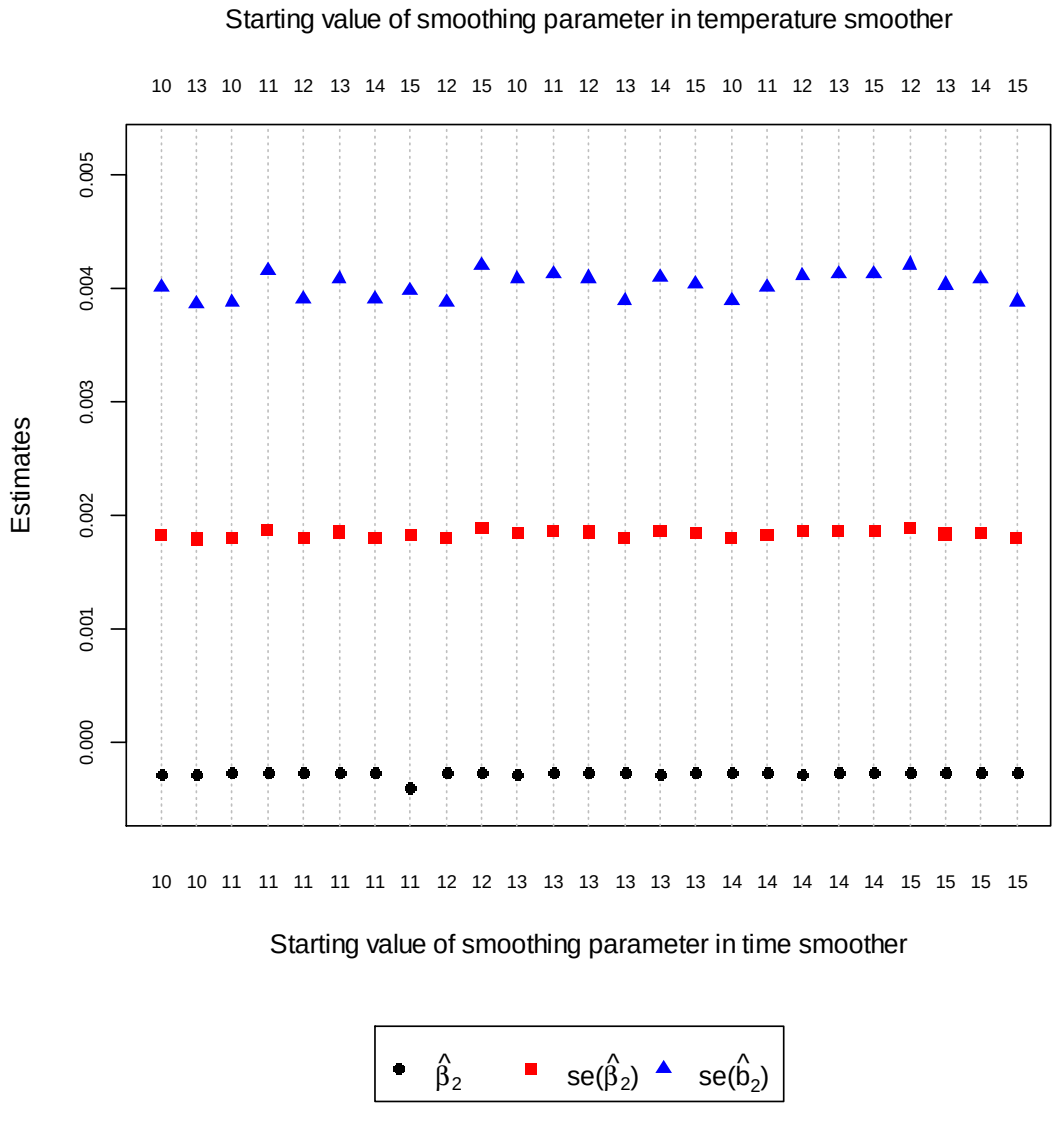
The estimated PM_{10} fixed effect with corresponding estimated standard errors of fixed and random effect from 25 convergence results using different starting values of λ_{time} and λ_{tmean} in model 6.



models, whatever any starting value of λ_{time} and λ_{tmean} were used in smoothing functions. Moreover, the convergent se($\hat{b}_2$)s were more consistent with different starting values of λ_{time} and λ_{tmean}, and the difference between the largest se($\hat{b}_2$) and the smallest se($\hat{b}_2$) was only 0.000348. This stability did not appear in the other co-pollutants' estimated standard errors of random effect. Generally speaking, we found larger starting value of λ_{time} and λ_{tmean} can have

Figure 3.20

The estimated SO_2 fixed effect with corresponding estimated standard errors of fixed and random effect from 25 convergence results using different starting values of λ_{time} and λ_{tmean} in model 6.



better chance to reach convergence easily, and this co-pollutant model showed much consistence by applying different starting values of λ_{time} and λ_{tmean}.

Unfortunately, model 3 did not have convergence results while using 36 different combinations of λ_{time} and λ_{tmean}. It was anticipated that there should be some good starting values of λ_{time} and λ_{tmean} to make the entire model-fitting reach convergence, but in case the

"good" values were really hard to find out, there were another two approaches which can offer opportunities to reach convergence in iterations and get reasonable estimates. We will discuss in chapter 3.5 and 3.6.

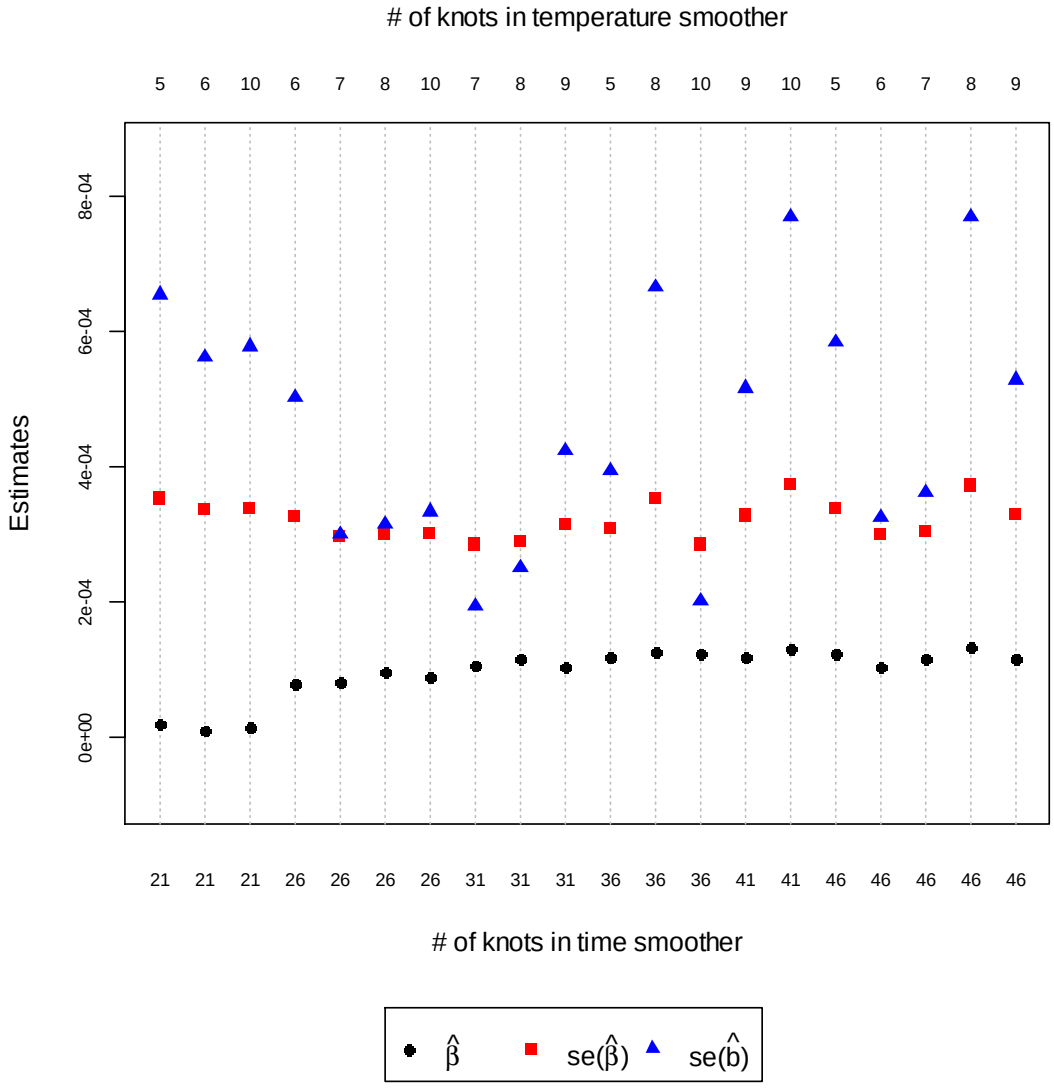
3.5. The influence of the number of knots in smoothing functions

Besides using different starting values of smoothing parameter λ to look for higher opportunity of reasonable estimates, BayesX also allows modifying the number of knots used in the smoothing functions, and claims the number of knots do not affect the result of model-fitting too significantly. However, we found that it indeed has influence on estimations, and sometimes can be used for an alternative solution as long as the approach mentioning in chapter 3.4 can hardly find out reasonable results in the GGAMM. Based on the initial numbers of knots used in our six models (k_{time}=31 and k_{tmean}=7), we also used additional five knots in time smoother (21, 26, 36, 41 and 46) and temperature smoother (5, 6, 8, 9 and 10) to refit all models. The entire results are shown in Appendix B, and all reasonable results are presented in figure 3.21 to figure 3.30. As the findings in smoothing parameter λ, we found model 1, model 4, model 5 and model 6 had higher probability to reach convergence and get reasonable estimates. 6 different knots in either time smoother or temperature smoother can generate 36 different combinations, and these 4 models got 20, 18, 22 and 24 reasonable results over 36 trials, respectively. On the contrary, model 2 and model 3 only had 7 and 0 reasonable results.

Among 20 reasonable results of model 1 shown in figure 3.21, the PM_{10} fixed effect $\hat{\beta}$ and its corresponding standard error se($\hat{\beta}$) had stable estimates around 0.0001 and 0.0003, respectively. The $\hat{\beta}$ had a tiny decrease when k_{time}=21, and we generally found that fixed

Figure 3.21

The estimated PM_{10} fixed effect with corresponding estimated standard errors of fixed and random effect from 20 convergence results using different number of knots in model 1.

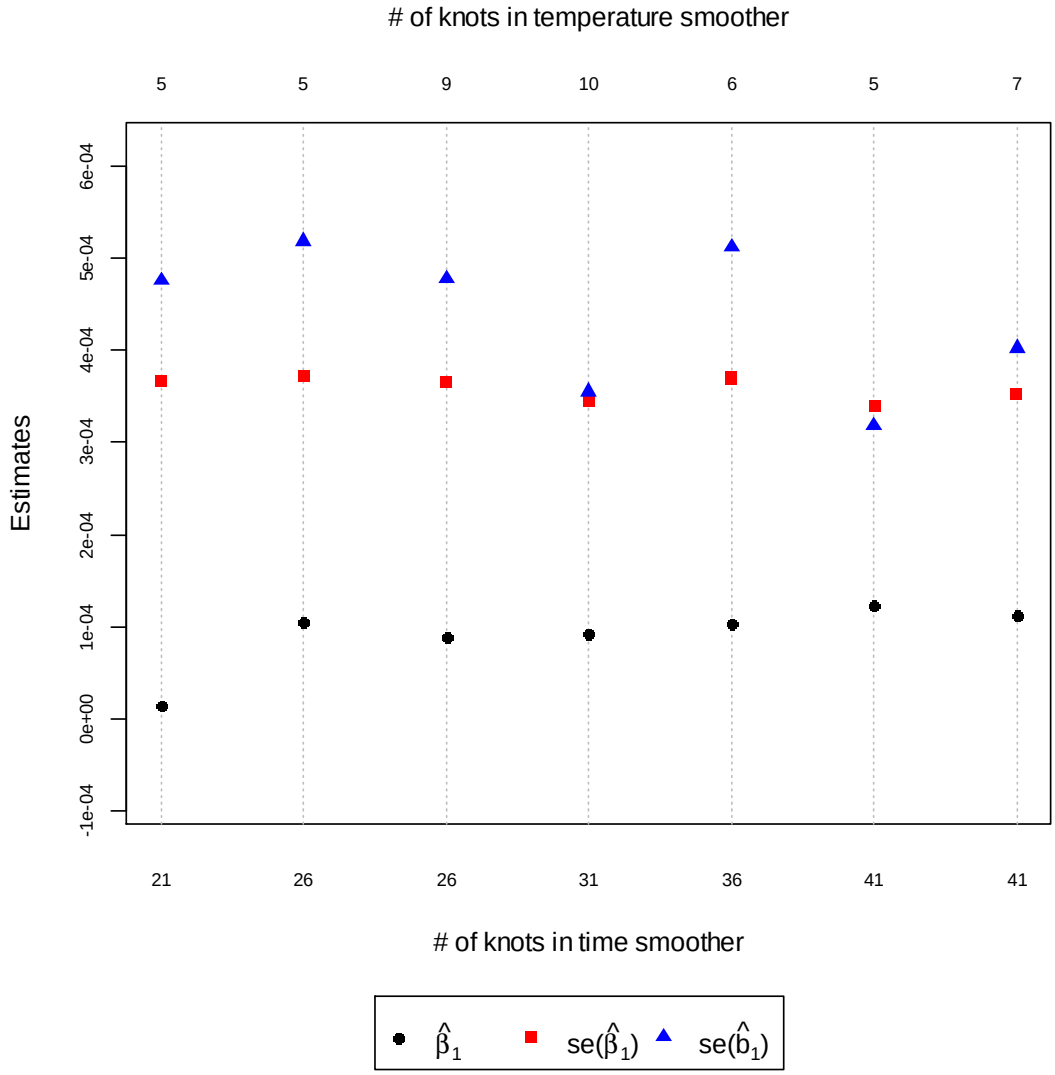


effects were always relative smaller as long as k_{time} was less than 26 in the time smoother. The estimated standard error of random effect se($\hat{b}$) had more fluctuation than se($\hat{\beta}$), and it was smaller especially in k_{time}=31 and 36 in time smoother. In addition, we never saw any tendency in these estimates while k_{tmean} changed in temperature smoother.

From 7 reasonable and convergent results in model 2 (figure 3.22, 3.23, 3.24), all fixed

Figure 3.22

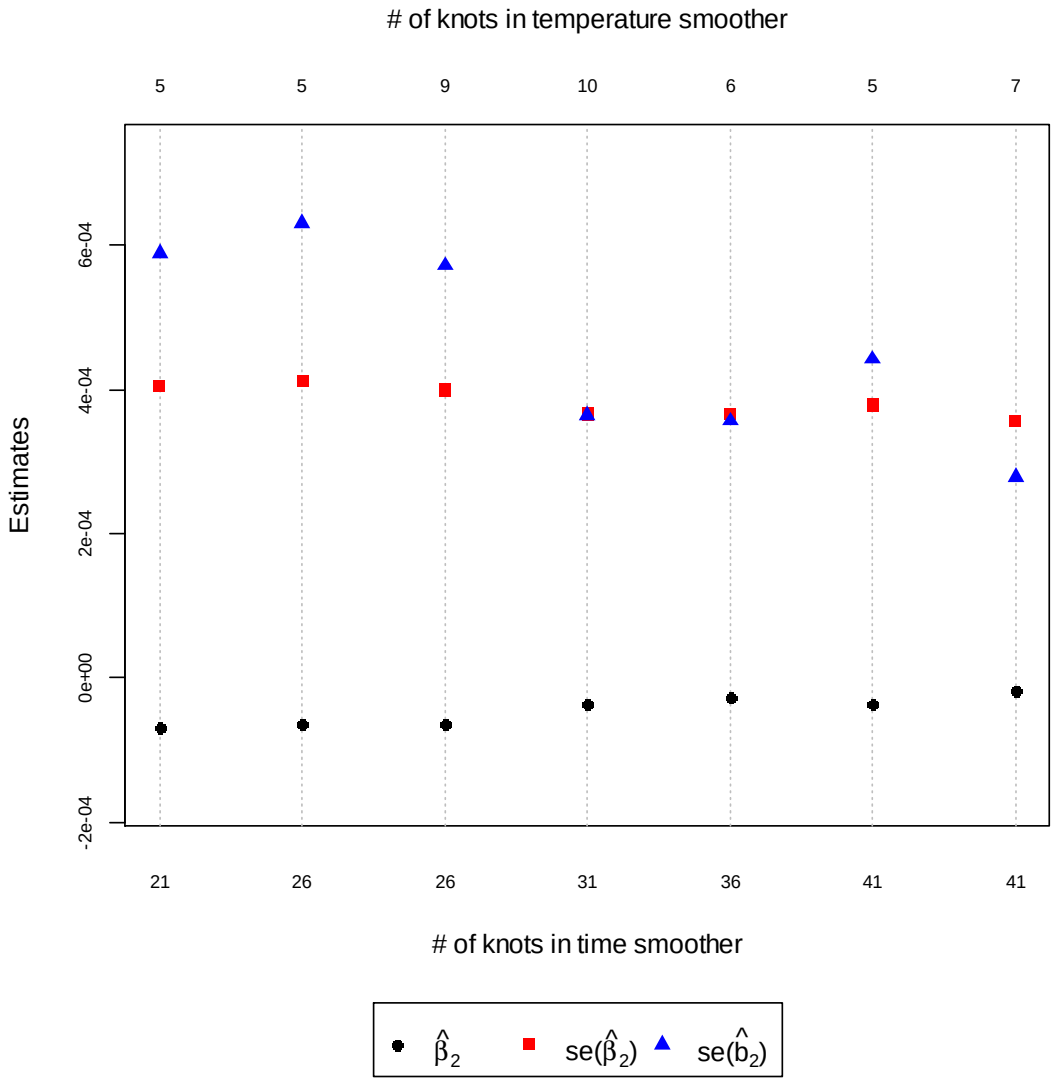
The estimated PM_{10} fixed effect with corresponding estimated standard errors of fixed and random effect from 7 convergence results using different number of knots in model 2.



effects and their standard errors had consistent estimates as the number of knots changed in two smoothing functions. We still found, when model-fitting had convergence problem, $\hat{\beta}_3$ could become negative around -0.0019 with huge overestimated se($\hat{\beta}_3$) and se($\hat{b}_3$). As long as this model was adjusted by suitable number of knots, both of $\hat{\beta}_1$ and $\hat{\beta}_2$ were decreased no more than 0.000123 and -0.000017. This reflected that problematic results would also lead to

Figure 3.23

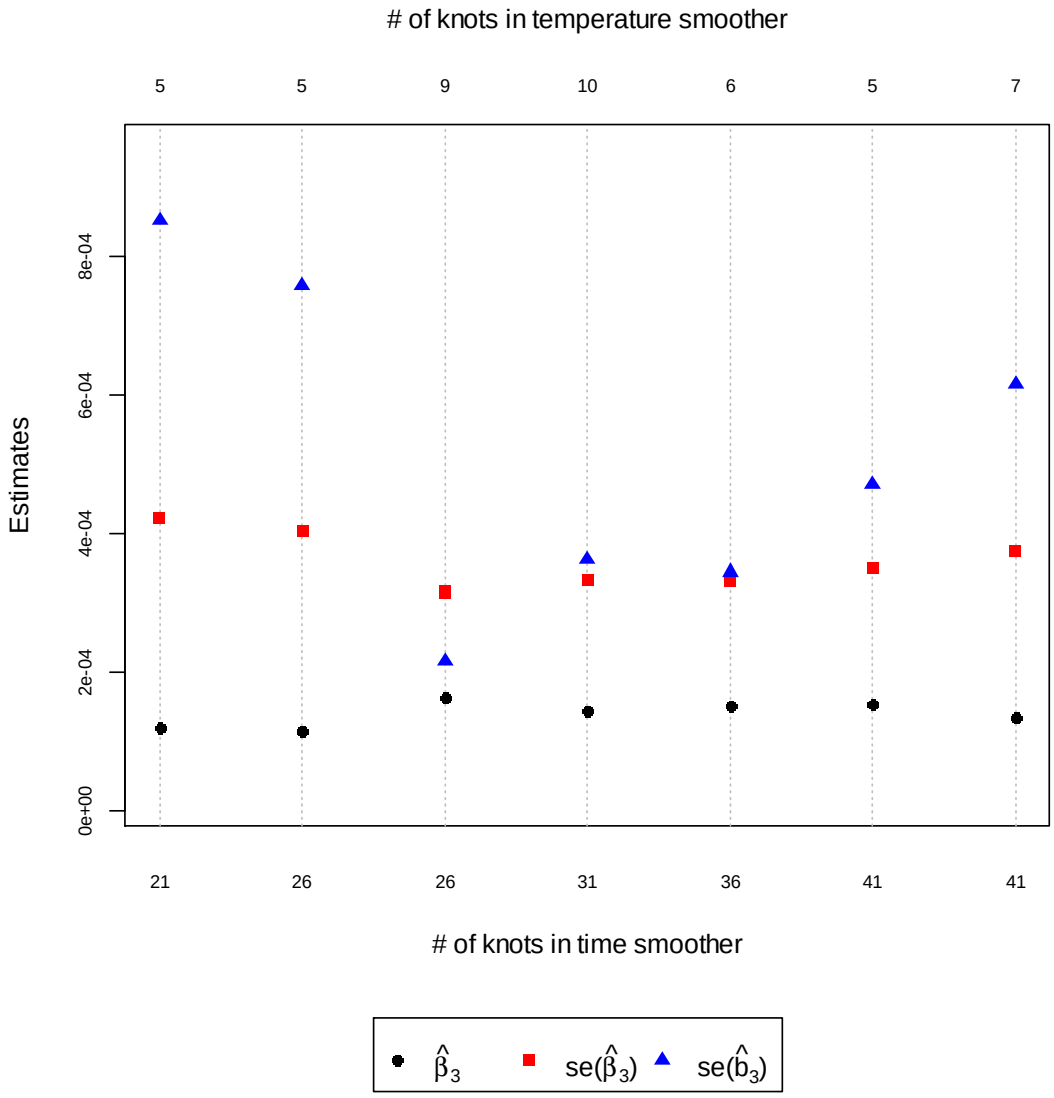
The estimated 1-day lag PM_{10} fixed effect with corresponding estimated standard errors of fixed and random effect from 7 convergence results using different number of knots in model 2.



overestimated parameter estimates. When k_{time}=31, 36 and 46 or k_{tmean}=5, 7 and 8, model 4 had higher opportunities to obtain reasonable results. Figure 3.25 showed that $\hat{\beta}_1$s were estimated between 0.000142 and 0.000208, but it decreased over a half when k_{time}=21. The se($\hat{\beta}_1$)s were relative consistent around 0.0004. The highest se($\hat{b}_1$) could reach 0.000771, and the lowest one was as low as 0.000210. There was no special pattern in se($\hat{b}_1$) in model 4, but

Figure 3.24

The estimated 2-day lag PM_{10} fixed effect with corresponding estimated standard errors of fixed and random effect from 7 convergence results using different number of knots in model 2.



at least over a half of se($\hat{b}_1$)s were higher than 0.000392. Unlike PM_{10}, all estimates related with NO_2 in model 4 were estimated very steady (figure 3.26). An interesting finding was higher $\hat{\beta}_2$ when k_{time}= 21 because this situation did not happen in the other co-pollutants.

While replacing NO_2 by O_3, the PM_{10} relative estimates ($\hat{\beta}_1$, se($\hat{\beta}_1$) and se($\hat{b}_1$)) in model 5 were similar as those in model 4, but se($\hat{b}_1$) appeared higher estimates more frequently. In

Figure 3.25

The estimated PM_{10} fixed effect with corresponding estimated standard errors of fixed and random effect from 18 convergence results using different number of knots in model 4.

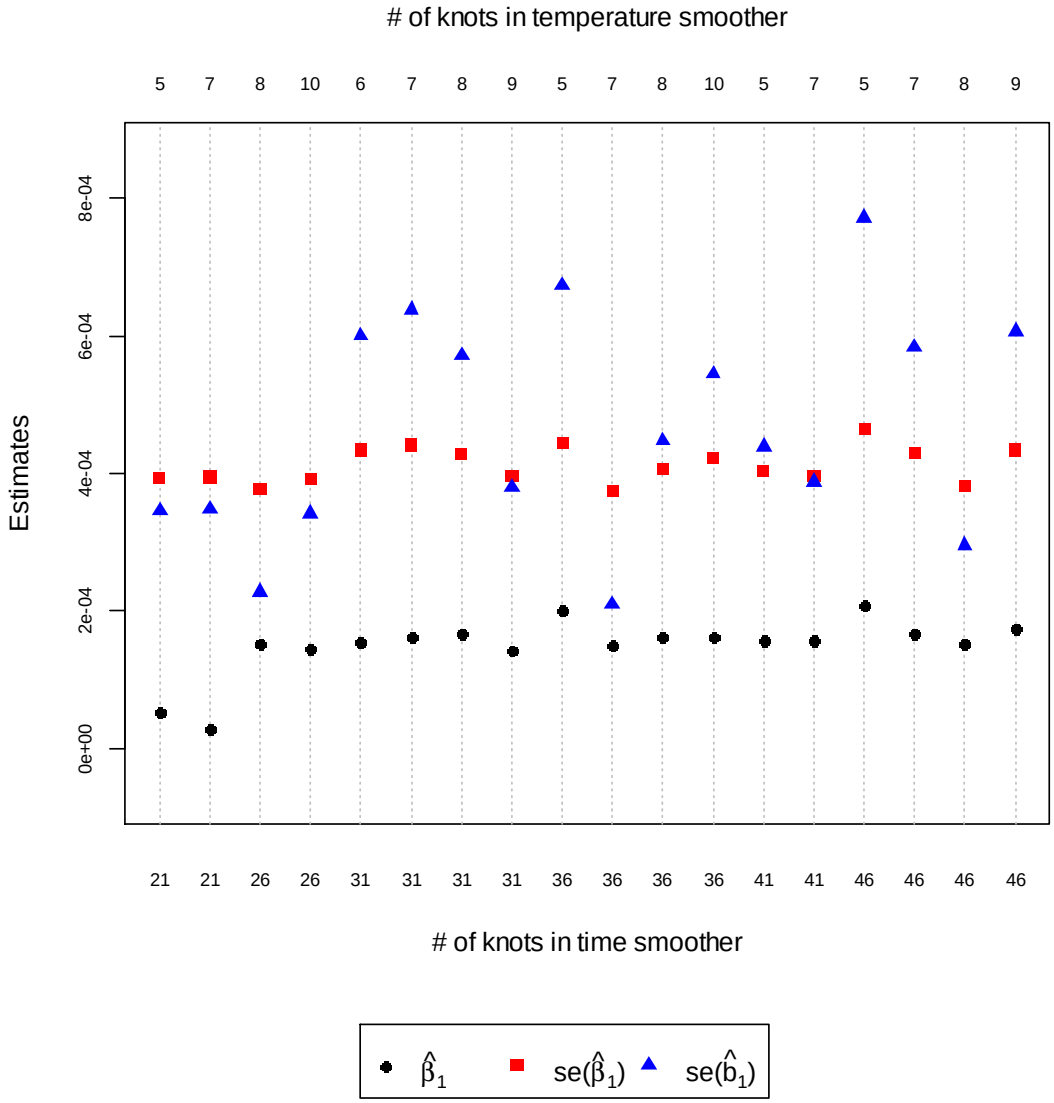
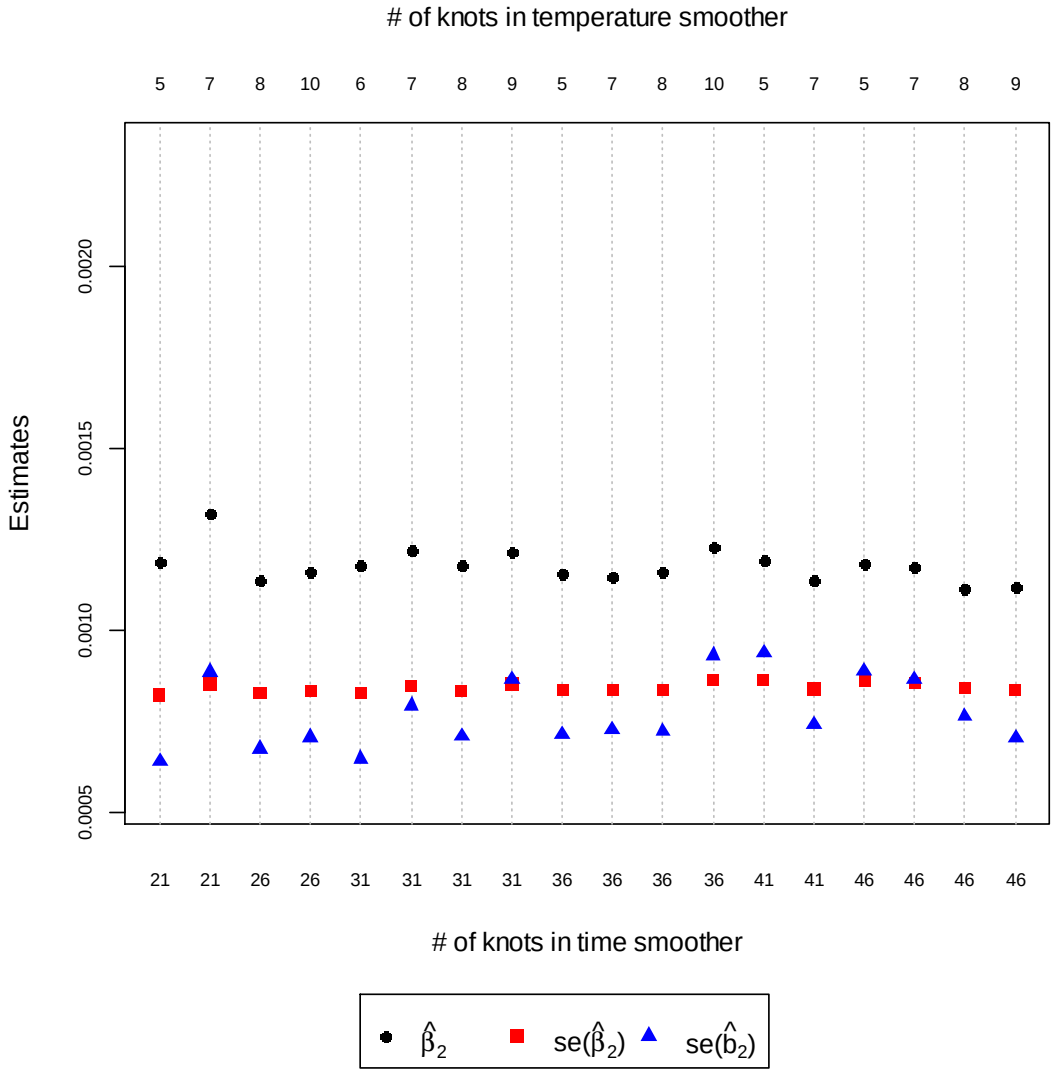


figure 3.27, the se($\hat{b}_1$)s estimated from model 5 with 31 and 36 knots in time smoother had higher probability to have relative smaller values. Generally speaking, most se($\hat{b}_1$)s were larger than 0.000531 in convergent model 5. Moreover, the convergent se($\hat{b}_2$)s were all smaller than $\hat{\beta}_2$, which was the fixed effect of co-pollutant O_3. Figure 3.28 implied that, no matter any number of knots using in time smoother or temperature smoother, the variation of

Figure 3.26

The estimated NO_2 fixed effect with corresponding estimated standard errors of fixed and random effect from 18 convergence results using different number of knots in model 4.

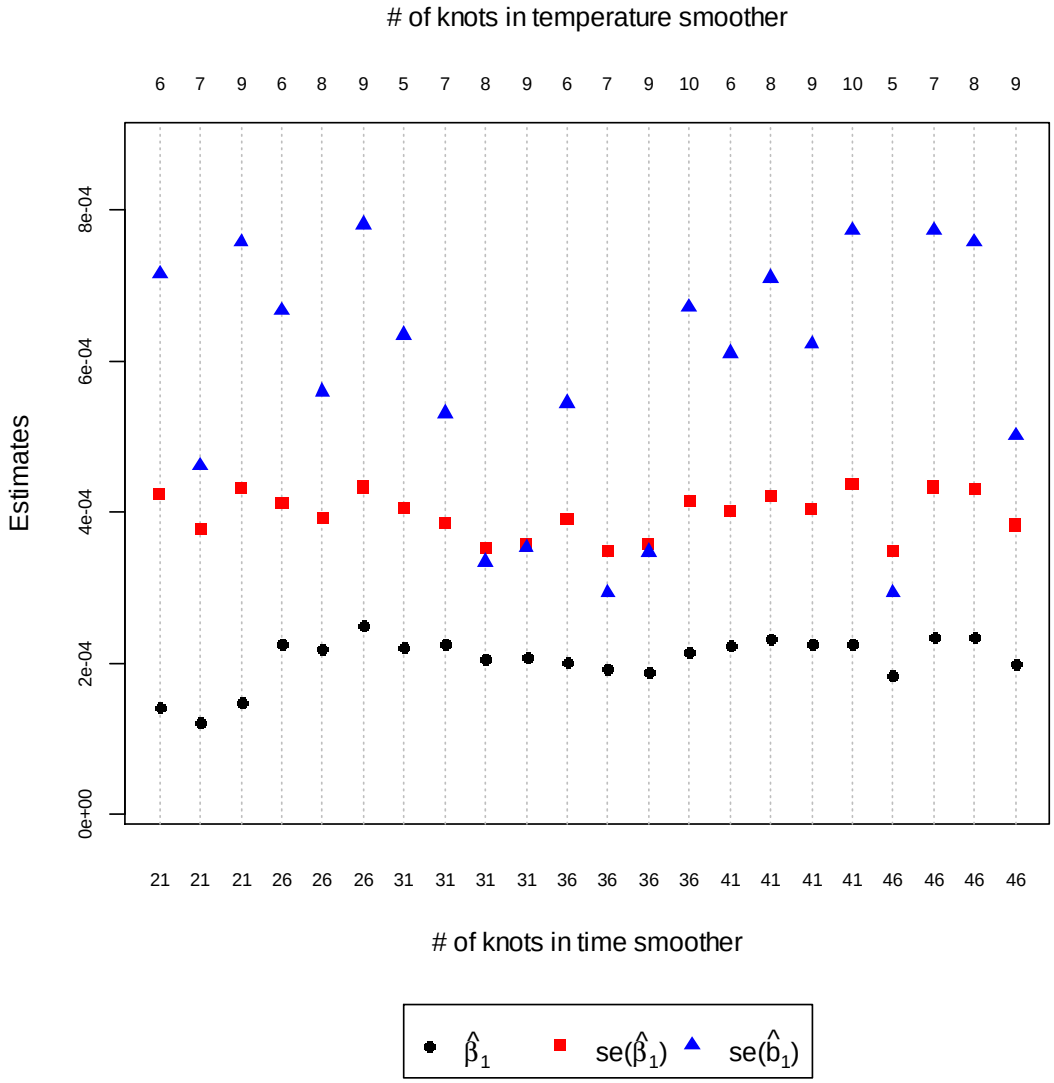


city-level effect of ozone was not as significant as the variation of city-level effect of PM_{10}. The initial se($\hat{b}_2$) from 31 knots in time smoother and 7 knots in temperature smoother (0.000650) was close to the average of 22 convergent se($\hat{b}_2$) (0.000601), so it was concluded that initial result was credible.

Not similar as the estimations of PM_{10} fixed effect from previous models, the $\hat{\beta}_1$ of

Figure 3.27

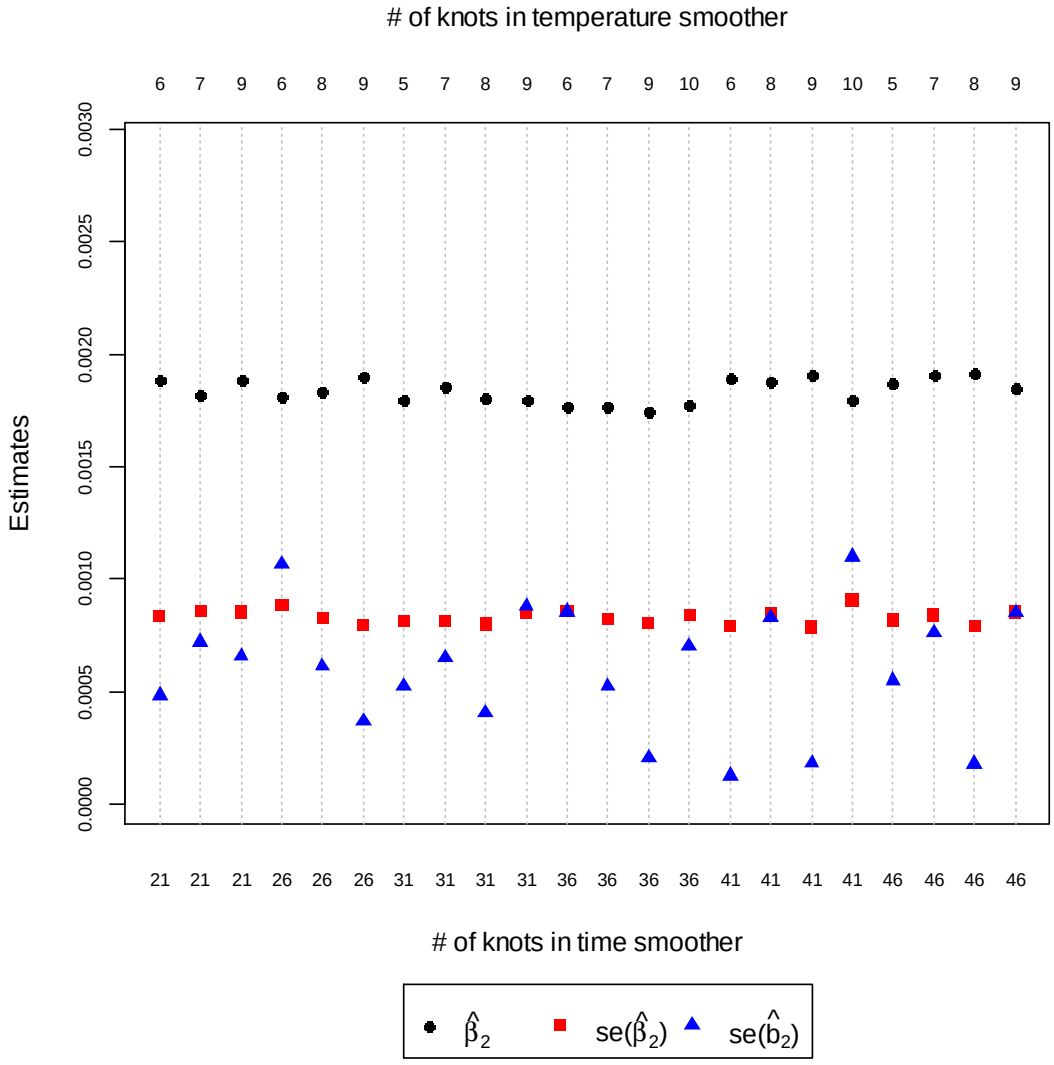
The estimated PM_{10} fixed effect with corresponding estimated standard errors of fixed and random effect from 22 convergence results using different number of knots in model 5.



model 6 had relative higher estimates when k_{time}=26, and the average $\hat{\beta}_1$ was 0.000459, while the other $\hat{\beta}_1$s had average value 0.000378. Figure 3.29 also presented $\hat{\beta}_1$ had static estimates with k_{time}=21 and 46, but there was no consistent $\hat{\beta}_1$ in any specific number of knots in temperature smoother. The se($\hat{\beta}_1$) still showed more consistency than se($\hat{b}_1$). Compared with estimated standard errors of random effect in the other models, the se($\hat{b}_2$)

Figure 3.28

The estimated O_3 fixed effect with corresponding estimated standard errors of fixed and random effect from 22 convergence results using different number of knots in model 5.

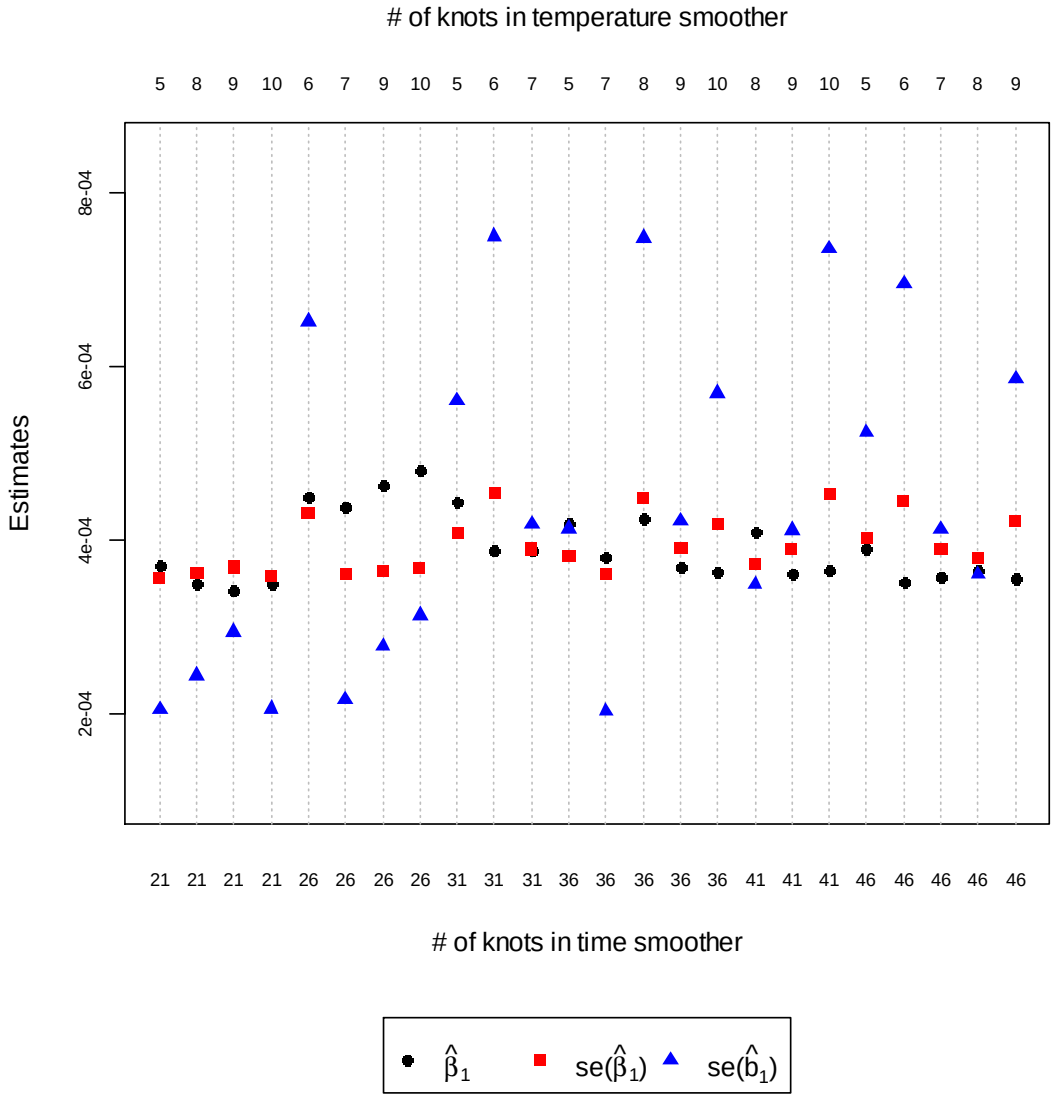


presented pretty high steadiness in model 6, as well as $\hat{\beta}_2$ and se($\hat{\beta}_2$), which implied that SO_2 can have much robust estimates as long as some specific numbers of knots setting in smoothing functions (figure 3.30).

Section 3.4 and section 3.5 showed that the starting values of smoothing parameter and the number of knots had similar influence on the probability of reaching convergence and the

Figure 3.29

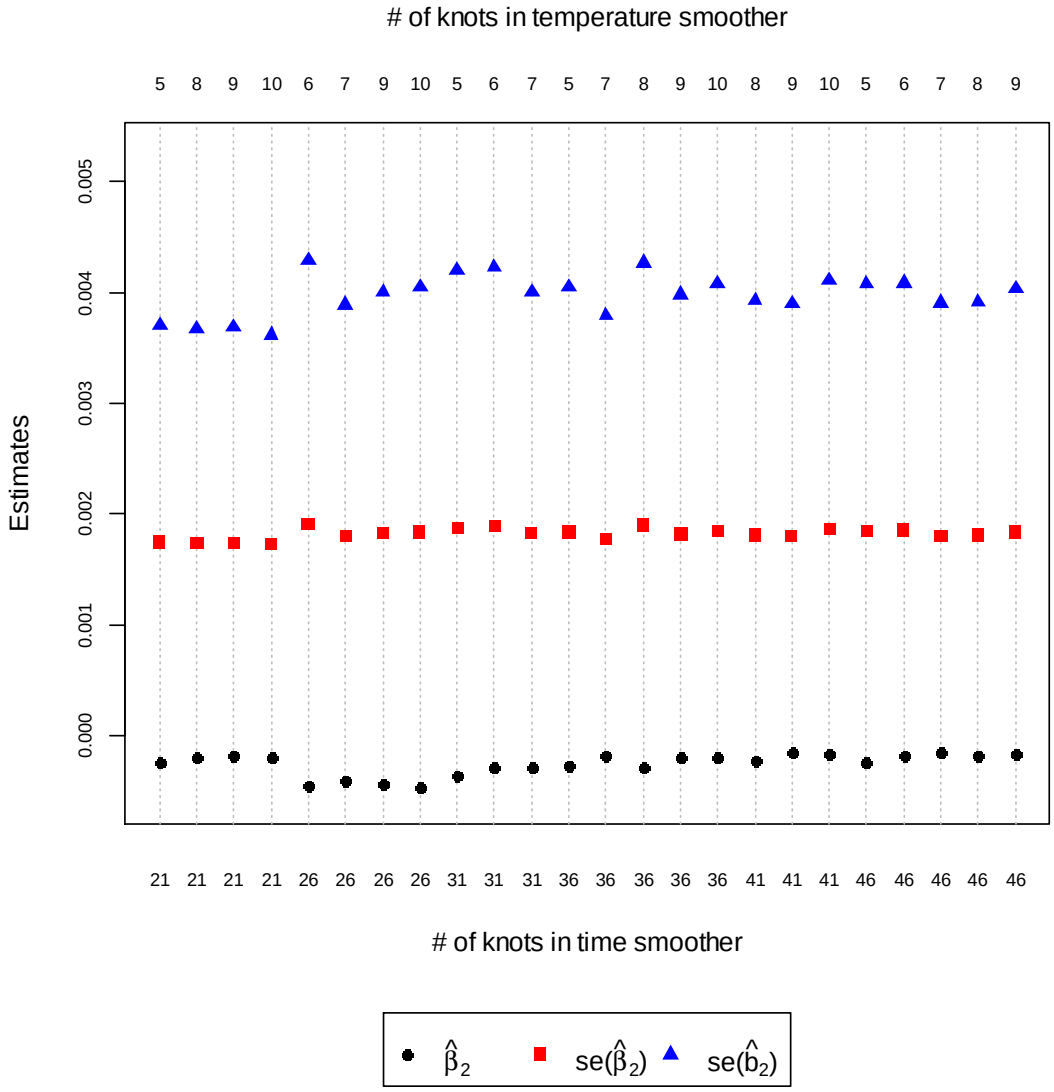
The estimated PM_{10} fixed effect with corresponding estimated standard errors of fixed and random effect from 24 convergence results using different number of knots in model 6.



stability of the estimated parameters. Even though the true source of convergent problem in the GGAMM has not been theoretically identified, the two modifications had been confirmed that the possibility of eliminating convergent problem is existing. Same as the choose of starting values of smoothing parameters, there is also no criterion to determine the best number of knots which can optimize estimated parameters, or immediately determine which

Figure 3.30

The estimated SO_2 fixed effect with corresponding estimated standard errors of fixed and random effect from 24 convergence results using different number of knots in model 6.



number of knots can absolutely help parameter estimations reach convergence. Users should try any possible number of knots by hand in programs, and review all results from BayesX's outputs. However, a principle should be followed in the time smoother. The default number of knots in the time smoother was setup from $(n \times (\#\ of\ years) + 1)$, where *n* is the average number of knots in each circle equal to 6 and the (*# of years*) is equal to 5. Therefore,

when (*# of years*) is fixed, we can only adjust *n* for the other number of knots. The reason is that the time smoother is a regular fluctuated spline which the number of knots in each circle is assigned averagely. Hence, when adding or deducting the number of knots, the average number of knots in each circle (*n*) should be retained the same. Besides, either the starting values of smoothing parameters and the number of knots are immediately operating in smoothing functions, so both time smoother and temperature smoother are also affected from the two modifications directly. This evaluation will be presented in section 3.7.

3.6. The application of jackknife-bootstrap approach

So far we found that the use of starting values of smoothing parameter or numbers of knots had similar results, but model 3 was still lacking reasonable estimates. We expected that there should be a specific starting value of smoothing parameter or number of knots in either time smoother or temperature smoother which can lead model 3 to find convergent values in all six estimates, but the way to precisely locate the exact setting values was unpredictable. In addition, fitting the GGAMM was a very time-consuming procedure, and much relied on considerable time and powerful computer hardware. In order to prevent endless and aimless search for starting values of smoothing parameters and the numbers of knots in smoothing functions, an empirical alternative named the jackknife-bootstrap approach is proposed here.

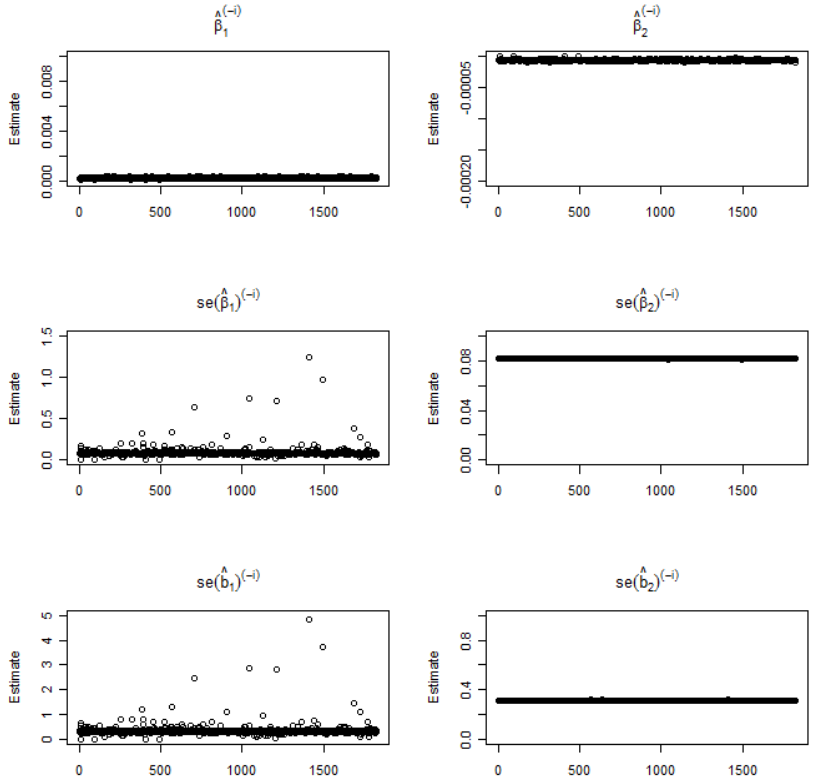
Before applying the jackknife-bootstrap approach, we first analyzed whether there is influential data in a specific date affecting estimations in model 3. Figure 3.31 presents six estimates ($\hat{\beta}_1$, se($\hat{\beta}_1$), se($\hat{b}_1$), $\hat{\beta}_2$, se($\hat{\beta}_2$), se($\hat{b}_2$)) in model 3 by re-fitting 1,826 times with dropping out the i^{th} observation in each city, where i=1, …, 1826. These estimates are kind of

estimates from a jackknife approach, and defined as $\hat{\beta}_1^{(-i)}$, se$(\hat{\beta}_1)^{(-i)}$, se$(\hat{b}_1)^{(-i)}$, $\hat{\beta}_2^{(-i)}$, se$(\hat{\beta}_2)^{(-i)}$ and se$(\hat{b}_2)^{(-i)}$. It was obvious that $\hat{\beta}_1^{(-i)}$, $\hat{\beta}_2^{(-i)}$, se$(\hat{\beta}_2)^{(-i)}$ and se$(\hat{b}_2)^{(-i)}$ had very consistent estimates whatever any date was dropped out in each city. Through 1,826 estimates in each parameter, almost all of $\hat{\beta}_1^{(-i)}$, $\hat{\beta}_2^{(-i)}$, se$(\hat{\beta}_2)^{(-i)}$ and se$(\hat{b}_2)^{(-i)}$ were nearby their averages, but some se$(\hat{\beta}_1)^{(-i)}$s and se$(\hat{b}_2)^{(-i)}$s were significantly away from their averages. We found that, if se$(\hat{\beta}_1)^{(-i)}$ had values explicitly higher or lower than their averages, the corresponding se$(\hat{b}_1)^{(-i)}$ would also have explicitly higher or lower values. For instance, se$(\hat{\beta}_1)^{(-1413)}$ had the highest value (1.246450) among se$(\hat{\beta}_1)^{(-i)}$s, and its corresponding se$(\hat{b}_1)^{(-1413)}$ = 4.827463 was also the highest value among se$(\hat{b}_1)^{(-i)}$s. However, dropping out some observations which can affect se$(\hat{\beta}_1)$ and se$(\hat{b}_1)$ will not improve se$(\hat{\beta}_2)$ and se$(\hat{b}_2)$. In fact, figure 3.31 had proved that se$(\hat{\beta}_1)$ and se$(\hat{b}_1)$ did not significantly decrease to reasonable estimates. In previous example, when dropping out all observations in calendar time t_{1413}, the se$(\hat{\beta}_2)^{(-1413)}$ and se$(\hat{b}_2)^{(-1413)}$ were 0.082051 and 0.317786, which were pretty close to initial estimates.

Therefore, we first applied the traditional jackknife approach to look for improved estimations, but we encountered a problematic situation. The procedure found that the estimated standard errors of fixed and random effects would rise along with the number of repeated drawings. In order to solve the inconsistency of traditional jackknife approach in this case, we modified this approach by embedding a bootstrap concept to randomly draw some samples from these jackknife estimates B times (B=100, 200,…, 1800) with replacement, and repeating this step 10,000 times to take the average. Those averages were shown in figure 3.32, and through B=100 to 1800, all estimates were changed slightly and

Figure 3.31

The jackknife estimates in model 3.

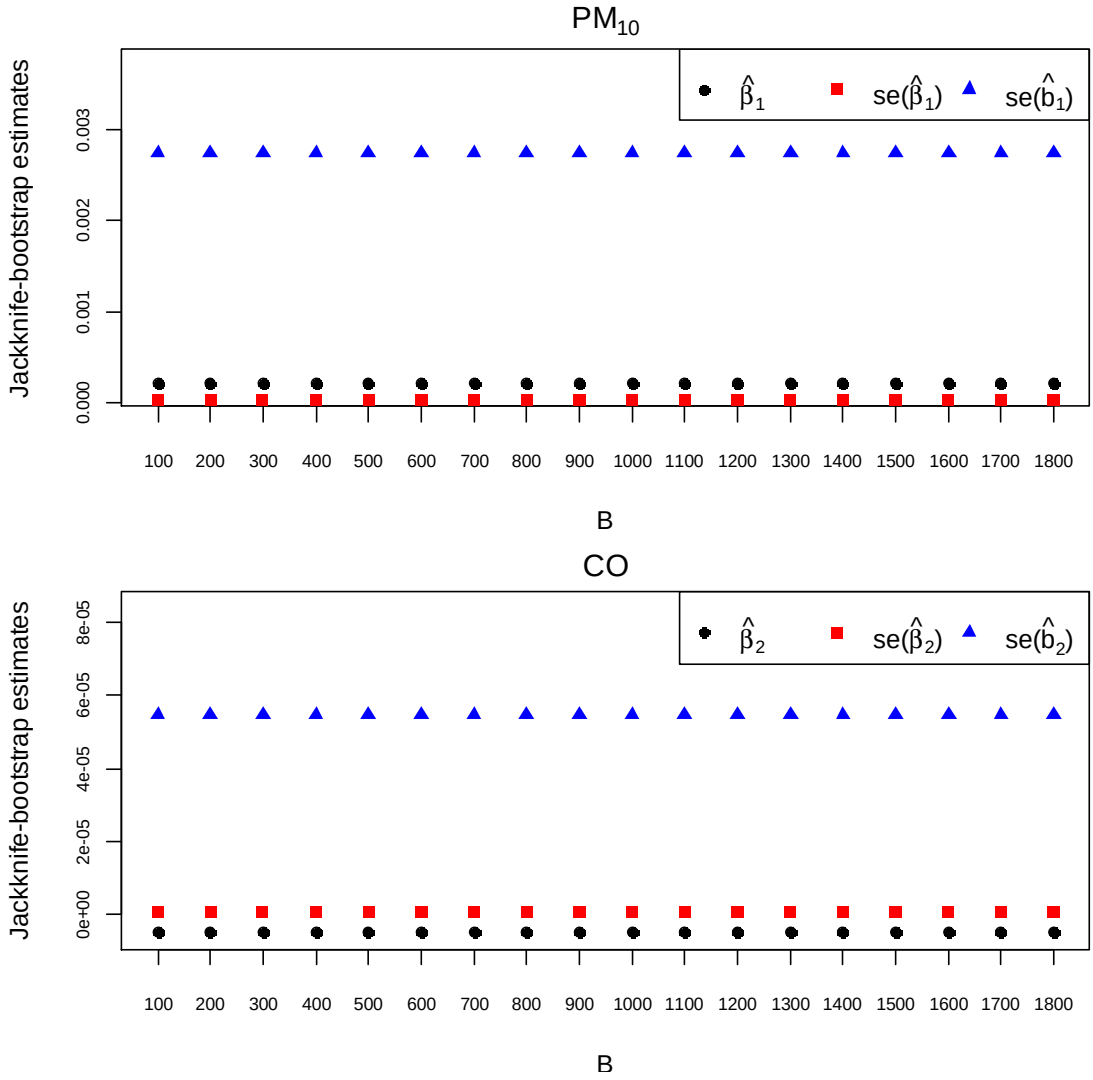


the differences could be almost ignored. Most importantly, the huge overestimated se($\hat{\beta}_1$), se($\hat{\beta}_2$), se($\hat{b}_1$) and se($\hat{b}_2$) in initial results shown in table 3 were reduced to acceptable estimates. Because the number of drawings with replacement did not differ too much, we can choose the result from B=100, where the jackknife-bootstrap adjusted ($\hat{\beta}_1$, $\hat{\beta}_2$)=(0.000206, -0.000005), (se($\hat{\beta}_1$), se($\hat{\beta}_2$))=(0.000027, 0.000001) and (se($\hat{b}_1$),se($\hat{b}_2$))=(0.002740, 0.000055).

However, this approach cannot be applied on any situation that model-fitting can get reasonable and convergent estimates as long as the estimates can be fitted initially well, or can be modified by the starting values of smoothing parameters or the number of knots in smoothing functions. For example, model 1 can be fitted reasonably in the initial settings: starting values of smoothing parameters were 10 and the numbers of knots were 31

Figure 3.32

The adjusted estimates by applying the jackknife-bootstrap approach in model 3.



and 7 in time smoother and temperature smoother, respectively. As applying the jackknife-bootstrap approach in model 1 with the same design using in mode 3, three main estimates $\hat{\beta}$, se($\hat{\beta}$) and se($\hat{b}$) all showed their consistency regardless of the number of resamplings. Nonetheless, as compared with initial estimates shown in the bottom of table 3.5, both of $\hat{\beta}$ and se($\hat{\beta}$) were underestimated, and se($\hat{b}$) was overestimated.

Table 3.5

The adjusted estimates by applying the jackknife-bootstrap approach in model 1.

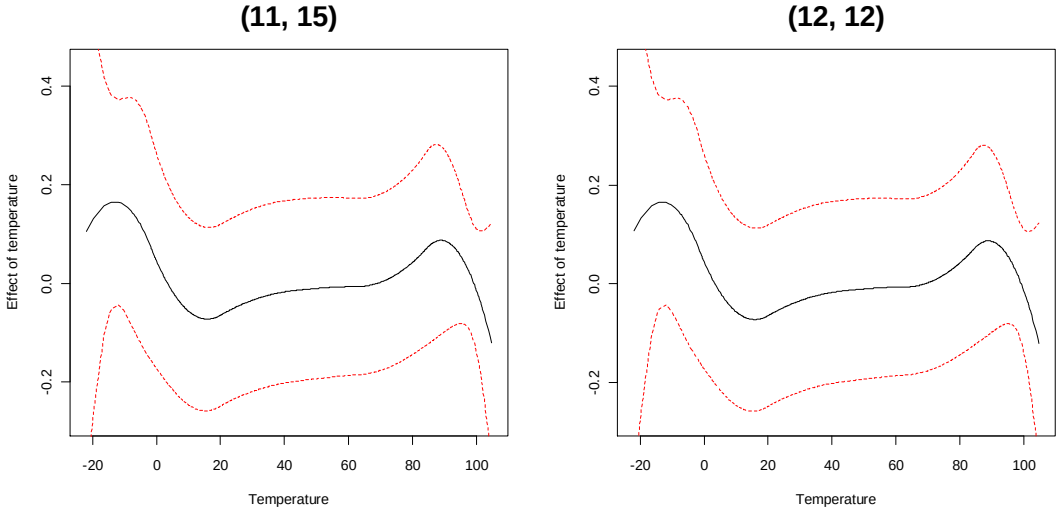
B	$\hat{\beta}$	se($\hat{\beta}$)	se($\hat{b}$)
100	0.000074149	0.000049864	0.000713967
200	0.000074145	0.000050791	0.000713998
300	0.000074088	0.000051179	0.000715293
400	0.000074135	0.000051356	0.000713957
500	0.000074140	0.000051621	0.000714157
600	0.000074107	0.000051787	0.000714489
700	0.000074115	0.000051821	0.000714280
800	0.000074102	0.000051795	0.000714571
900	0.000074115	0.000052060	0.000714666
1000	0.000074117	0.000052076	0.000714519
1100	0.000074103	0.000052170	0.000714849
1200	0.000074113	0.000052043	0.000714353
1300	0.000074145	0.000052127	0.000714040
1400	0.000074088	0.000052041	0.000714875
1500	0.000074108	0.000052217	0.000714660
1600	0.000074096	0.000052120	0.000714626
1700	0.000074105	0.000052176	0.000714732
1800	0.000074092	0.000052117	0.000714643
Initial estimates	0.000105264	0.000286481	0.000193690

3.7. Smoothing functions

In section 3.4, we have showed that the starting value of the smoothing parameter had different influences on estimates, and we also found that it can also affect smoothing functions, especially the temperature smoother. The entire 36 pictures of time smoothers and 36 pictures of temperature smoothers from the combinations of 6 starting values of λ_{time} and λ_{tmean} are in Appendix C. Among six models, time smoothers came out less influenced from different starting values of λ_{time} and λ_{tmean}. The fluctuation frequency in time smoother can be maintained by different starting values of λ_{time} or λ_{tmean}, but sometimes the 95% confidence

Figure 3.33

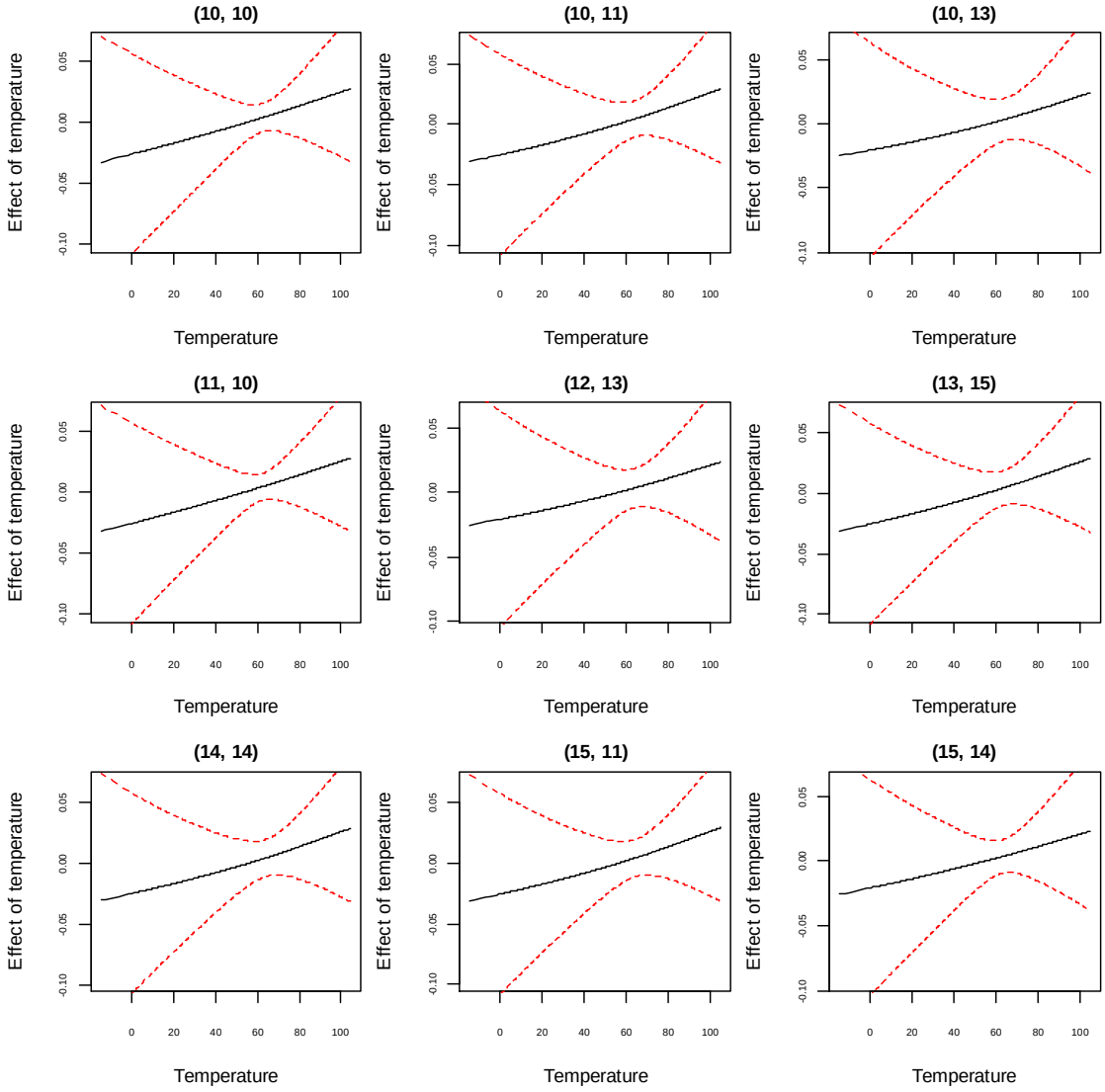
The temperature splines of model 1 when (λ_{time}, λ_{tmean})= (11, 15) and (12, 12).



interval became wider as starting values of (λ_{time}, λ_{tmean})= (13, 10) and (13, 14) in model 3 and (11, 15), (12, 13) and (15, 11) in model 6. In temperature smoothers, different starting values of λ_{time} and λ_{tmean} may have relatively higher probabilities to cause smoothing functions not look as usual. For example, most temperature smoothers in model 1 presented a hook shape, except for (λ_{time}, λ_{tmean}) is (11, 15) and (12, 12). The two exceptions were like a basin, and the temperature effect went down even when temperature reached 100°F and -20°F (figure 3.33). By tracking their results in estimates, it's obvious that both of them did not complete the convergence which also caused their $\hat{\beta}$s to tend to negative value -0.000046. The basin-like temperature smoother also appeared in model 2 when starting values of (λ_{time}, λ_{tmean})=(13, 11), (14, 13), (15, 14) and (15, 15). The basin-like temperature smoothers were much often happening in co-pollutant models, especially in PM_{10}+CO (model 3). Besides, we also took note of that sometimes temperature smoother was not smooth any more, and presented a straight line with divergent tendency at the end of two sides of 95% confidence interval. This situation often happened in modeling PM_{10} adjusted by O_3.

Figure 3.34

The non-smoothing splines and 95% CI of temperature in model 5, where (a, b)=(λ_{time}, λ_{tmean}).



Among 36 temperature smoother plots, 9 of them had non-smoothing results, shown in figure 3.34. These non-smoothing smoothers were accompanied with either convergent or non-convergent results. From the other smoothing temperature smoothers in model 3, most of them were still not smoothing enough, and the widths of 95% confidence intervals were also getting broader in two tails, such as plots with starting values of (λ_{time}, λ_{tmean})=(10, 12) and

(10, 14). It was concluded that, after adjusting co-pollutant O_3, the linear effect of temperature was getting stronger and it was not necessary to enforce it in a smoothing function. The temperature smoother in model 6 with initial starting value of (λ_{time}, λ_{tmean})=(10, 10) was not alike a hook, but the middle segment from 10 °F and 70 °F were flatter, but the spline climbing in two tails was as usual. There were still two temperature smoothers appearing in a basin-like shape, but the others were very robust. Both abnormal smoothers happened in starting value of λ_{time}=12, and their estimates were also absolutely not reaching convergence.

Comparing with the results in section 3.4, it was explicit that the starting values of smoothing function had much influence on estimates than on smoothing functions themselves. When finding abnormal results in estimates which possibly came out from incomplete convergence, we still suggest modifying the starting value of smoothing function to search for reasonable answers with convergence findings. Even though it had lower probability of affecting smoothing functions, an unusual smoothing function can highly affect unknown parameters' estimation. According to our 36 trials in each model, there were 2, 5, 20, 1, 13 and 2 temperature smoothers showed basin-like smoothers or straight lines from model 1 to model 6, respectively. Among these 43 temperature smoothers, there were 37 splines coming out with abnormal estimates with convergence problem. As a result, when looking for suitable starting values of smoothing parameters to speed up the velocity of iteration and avoid convergence problem, it's necessary to also check the smoothing functions ancillary.

The plots of time smoother and temperature smoother with different numbers of knots (k_{time} and k_{tmean}) in each model are gathered in Appendix D. The number of knots should have

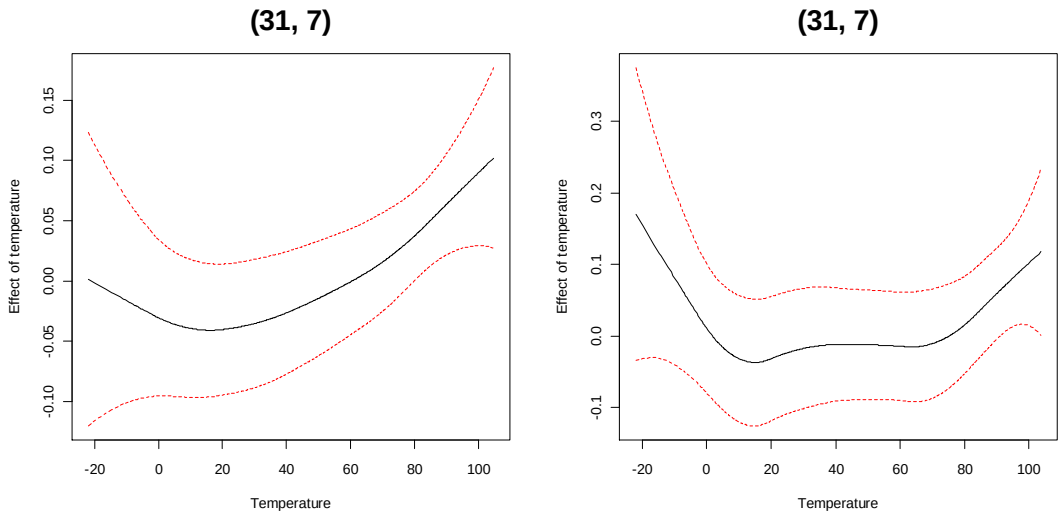
more immediate influence than the starting value of smoothing parameter because the entire structure of smoothing curve was initially decomposed into pieces by these knots and combined with random effects to estimate simultaneously. This part had been demonstrated in chapter 2.1. Using too many knots in a smoothing function with fewer turnovers could cause the estimated smoothing function some twisted segments in unexpected positions. On the contrary, a smoothing function, especially seasonal time series curve, should not use too fewer knots such that the estimated curve has too less variation. The number of knots using in our time smoothers were appropriately enough from 26 to 46, so we did not find poor results in each model. Either k_{time} and k_{tmean} did not make time smoothers unusual, but wider 95% confidence intervals appeared in model 1, model 3 and model 4. In particular, model 3 displayed the situation of 5-fold wider 95% confidence intervals, and 4 of them happened in k_{tmean}=6. In fact, k_{tmean} should only affect temperature smoother itself, but the corresponding temperature smoothers shown in Appendix D did not present abnormal curves.

As what we had mentioned before, temperature smoothers were generally a pure curve with possibly one or two turnovers located at some particular degrees of temperature, such as 20°F in model 1 or 10°F and 70°F in model 6 when (k_{tmean}, k_{tmean})=(31, 7) (figure 3.35). The reason for a regular temperature smoother in model 6 to have two turnovers is when the first significant turnover occurred around 10°F which was earlier than model 1, the curve increased slowly and became flat until 70°F. Another significant increase happened after 70°F which makes the temperature smoother look like a bowl in model 6, but the only difference was whether the second turnover really showed up before the final climb-up until the end.

In addition, it was surprising that more knots using in temperature smoother actually did not produce twisted curve, but alternatively suppressed the effect of cold weather, especially

Figure 3.35

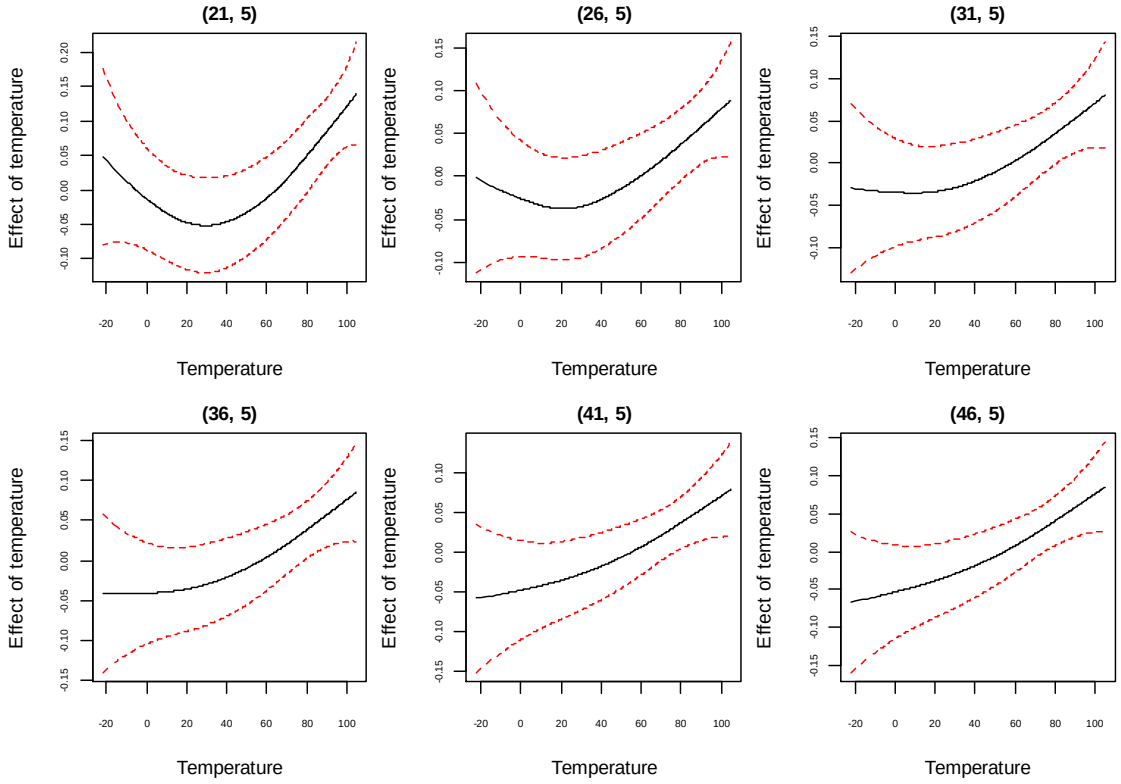
Two regular temperature smoothers in model 1 and model 6.



when time smoothers used more than 31 knots. This situation frequently happened in model 1, model 2 and model 4. In details, the temperature smoothers in model 1 showed nice curves under all numbers of knots in temperature smoothers and 21 knots in time smoothers. When k_{time} increased, the left tail of temperature smoothers gradually lay down, and even became lower than the lowest level of temperature effect in other hook-like temperature smoothers. For example, figure 3.36 presents a series of temperature smoothers' changing when k_{time}=21, 26, 31, 36, 41 and 46 conditional on a fixed k_{tmean}=5. The turnover point in plot (21, 5) was explicitly located between 20 °F and 40 °F, but became nonsignificant when k_{time} increased. Moreover, the extreme point of left tail $f(TEMP = -20°\text{F})$ was 0.05, and the lowest level in the same spline $f(TEMP = 30°\text{F})$ was -0.05 when k_{time}=21. Nonetheless, as k_{time} increased to 31, $f(TEMP = -20°\text{F})$ decreased to -0.03 and $f(TEMP = 30°\text{F})$ slightly increased to -0.03 as well. The $f(TEMP = -20°\text{F})$ started to below $f(TEMP = 30°\text{F})$ when k_{time} kept increasing, and the turnover point at $TEMP = 30°\text{F}$ also became disappeared. We think this situation came out from a possible reason: when using two or

Figure 3.36

The temperature smoothers using k_{tmean}=5 with different k_{time}s in time smoother in model 1.

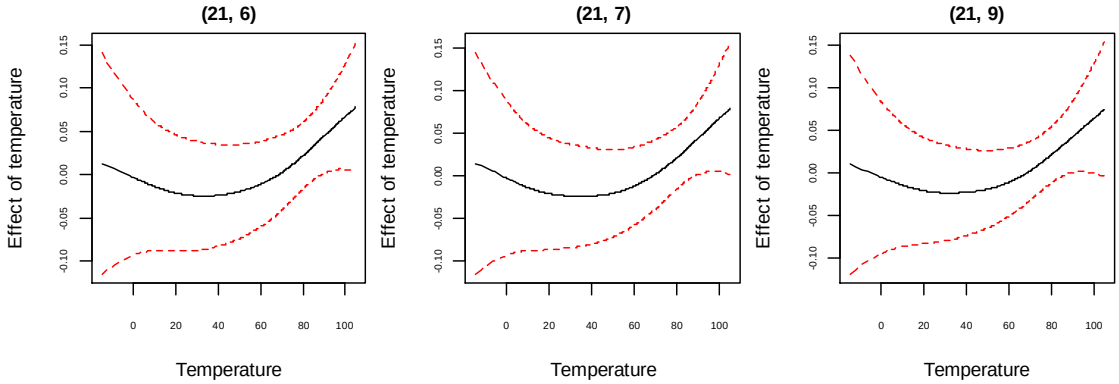


more smoothing functions, all of them were decomposed into pieces by knots, combined with random effects, and estimated simultaneously. When a smoothing function used more knots, it would share more effects than the others.

If smoothing functions were not smooth any more, such as the temperature smoother in model 5 by using (k_{time}, k_{tmean})=(31, 7), we had found that changing the starting values of smoothing parameters did not improve it and still presented straight line. By reducing k_{time} to 21, the temperature smoother would become a hook-like curve accompanied with appropriate k_{tmean} (figure 3.37). Their estimates were consistent (Appendix B) and would be credible to present. By checking all corresponding estimates in linear factors, we found some basin-like or weird twisted temperature smoothers almost happened to accompany with

Figure 3.37

The temperature smoothers in model 5 with (k_{time}, k_{tmean})=(21, 6), (21, 7) and (21, 9).



estimates having convergence problem. Even though well-fitted smoothing functions were not guaranteed for helping linear factors' estimations, they were still indispensable because results with convergence problem always did not have normal temperature smoothers.

3.8. Spatial functions

Besides the investigation of starting values of smoothing parameters in time and temperature smoother, we also examine how the starting value of smoothing parameter λ_{spat} affect in spatial function. For the sake of simplifying analysis, the adjustment of λ_{spat} was only based on convergence results with corresponding λ_{time} and λ_{tmean} in smoothing functions in each model. Meanwhile, model 1, model 4, model 5 and model 6 were using (λ_{time}, λ_{tmean}) = (10, 10), and model 2 was using (λ_{time}, λ_{tmean}) = (11, 10) with λ_{spat} = 10 (default value), 11, 12, 13, 14 and 15. The only exception was model 3 because there was no reasonable result based on the same λ_{time}s and λ_{tmean}s used in the other models. Therefore, we kept (λ_{time}, λ_{tmean}) = (10, 10) in this case. All spatial function maps with corresponding smoothing function plots are included in Appendix E.

Because spatial function uses the relative distances among these cities, the relative effects

theoretically should not be changed with different λ_{spat}s. For example, in model 1, the highest spatial effect $f_{spat}(s) =$ 1.3684 appeared in Chicago with λ_{spat} = 10, and the lowest one with the same λ_{spat} was -1.3344 in Huntsville. Whatever using any λ_{spat} from 11 to 15, the two cities still had the highest and lowest spatial effect, respectively. This comparative relationship was also followed by any couple of cities, and never changed by adjusting different λ_{spat}s. Nonetheless, the magnitude of spatial effect in single city could be changed with the use of different λ_{spat}s. Table 3.6 showed the initial results of adjusting λ_{spat} from model 1 to model 6. There was no doubt to get problematic estimates in some λ_{spat}s, but the situation was similar as our previous investigations in smoothing functions. That is, if a GGAMM can obtain reasonably convergent estimates in default λ, it has higher probability to get reasonably convergence estimates in other λs, whatever smoothing function or spatial function. This principle had been proved again in spatial function. For results having convergence estimates with default λ_{spat} in model 1, model 4, model 5 and model 6, besides initial estimates with λ_{spat} = 10, they obtained additional 3, 3, 4 and 2 convergence results, respectively. In particular, model 3 still did not have convergence results over these λ_{spat}s.

The main estimates from different λ_{spat}s in table 3.6 proved that convergent results still had robust estimates. In model 1, the range of convergent estimates $\hat{\beta}_1$ was from 0.000100 to 0.000105, and two problematic $\hat{\beta}_1$ with large se($\hat{\beta}_1$) significantly decreased to 0.000019 (λ_{spat} = 11 and 12). Considering lag effects, we got another convergent result using λ_{spat} = 13, and 1-day PM_{10} lag effect was slightly raised from -0.000062 to -0.000037. Its 2-day PM_{10} lag effect contrarily decreased from 0.000157 to 0.000142, but the difference was very tiny. Their estimated standard errors were also highly consistent compared with initial result of λ_{spat} = 10. Co-pollutant effects, besides CO, had similar performance as PM_{10} effect. The

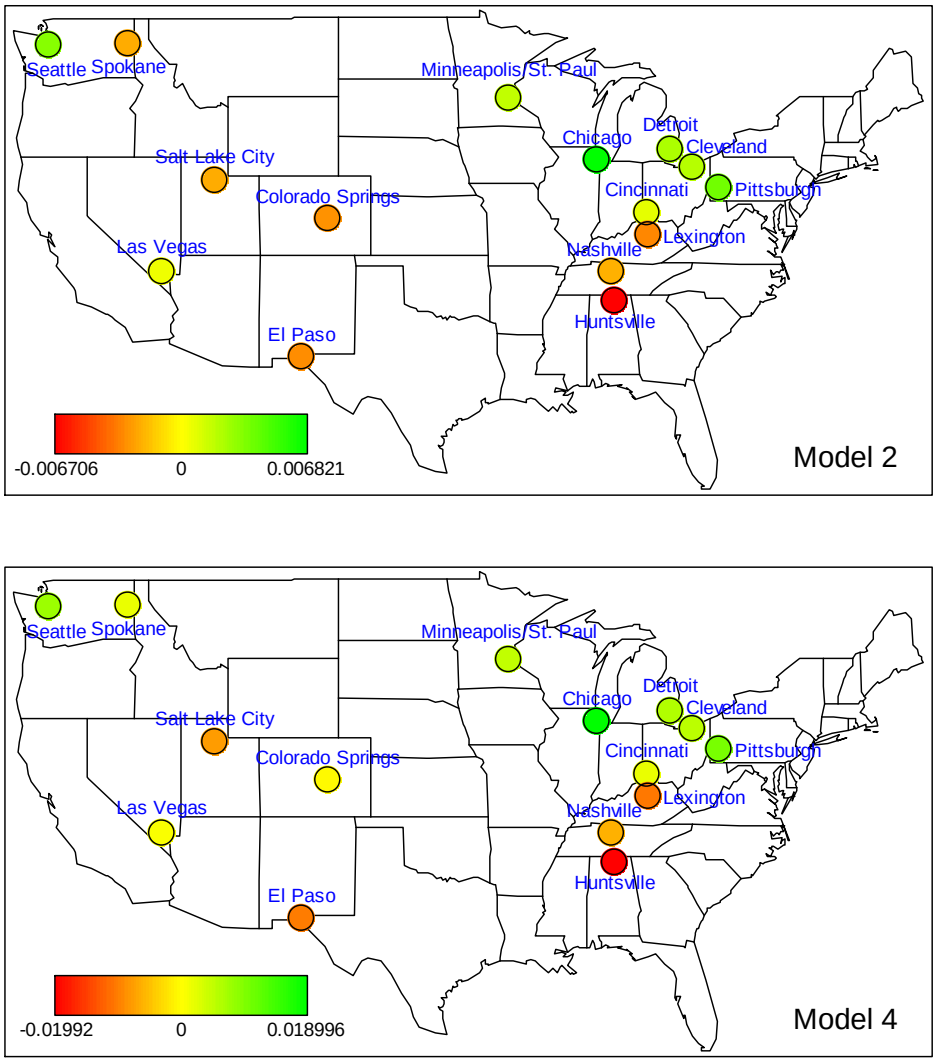
Table 3.6

The main estimates with different λ_{spat}s from 10 to 15 in each model.

		λ_{spat}					
Model #	Parameter	10	11	12	13	14	15
1	$\hat{\beta}_1$	0.000105	0.000019	0.000100	0.000019	0.000101	0.000101
	$\text{se}(\hat{\beta}_1)$	0.000286	0.088380	0.000311	0.094020	0.000297	0.000300
	$\text{se}(\hat{b}_1)$	0.000194	0.342288	0.000403	0.364130	0.000292	0.000315
2	$\hat{\beta}_1$	0.000097	0.000394	0.000738	0.000093	0.000738	0.000457
	$\text{se}(\hat{\beta}_1)$	0.000351	0.366932	0.091934	0.000355	0.101896	0.142061
	$\text{se}(\hat{b}_1)$	0.000388	1.421119	0.356037	0.000414	0.394621	0.550190
	$\hat{\beta}_2$	-0.000062	-0.000040	-0.001929	-0.000037	-0.001917	-0.000019
	$\text{se}(\hat{\beta}_2)$	0.000400	0.000394	0.109107	0.000370	0.120712	0.000367
	$\text{se}(\hat{b}_2)$	0.000579	0.000507	0.422553	0.000385	0.467500	0.000340
	$\hat{\beta}_3$	0.000157	-0.002202	0.000175	0.000142	0.000159	-0.002168
	$\text{se}(\hat{\beta}_3)$	0.000318	0.517645	0.000343	0.000337	0.000410	0.130741
	$\text{se}(\hat{b}_3)$	0.000244	2.004827	0.000394	0.000389	0.000757	0.506347
3	$\hat{\beta}_1$	0.000196	0.000221	0.000221	0.000195	0.000198	0.000197
	$\text{se}(\hat{\beta}_1)$	0.088258	0.081425	0.081336	0.080944	0.080557	0.085143
	$\text{se}(\hat{b}_1)$	0.341813	0.315346	0.315005	0.313484	0.311985	0.329745
	$\hat{\beta}_2$	-0.000005	-0.000005	-0.000005	-0.000005	-0.000005	-0.000005
	$\text{se}(\hat{\beta}_2)$	0.081648	0.081647	0.081651	0.081651	0.081649	0.081650
	$\text{se}(\hat{b}_2)$	0.316223	0.316219	0.316233	0.316231	0.316226	0.316231
4	$\hat{\beta}_1$	0.000163	0.000054	0.000162	0.000455	0.000141	0.000132
	$\text{se}(\hat{\beta}_1)$	0.000441	0.510938	0.000444	0.129984	0.000413	0.000381
	$\text{se}(\hat{b}_1)$	0.000638	1.842208	0.000652	0.468654	0.000480	0.000263
	$\hat{\beta}_2$	0.001224	0.002162	0.001256	0.001162	0.001256	0.001215
	$\text{se}(\hat{\beta}_2)$	0.000846	0.475760	0.000863	0.000835	0.000866	0.000855
	$\text{se}(\hat{b}_2)$	0.000795	1.715366	0.000919	0.000673	0.000951	0.000877
5	$\hat{\beta}_1$	0.000227	0.000214	-0.000384	0.000223	0.000208	0.000216
	$\text{se}(\hat{\beta}_1)$	0.000385	0.000352	0.109500	0.000414	0.000376	0.000382
	$\text{se}(\hat{b}_1)$	0.000531	0.000332	0.409701	0.000672	0.000472	0.000509
	$\hat{\beta}_2$	0.001855	0.001831	0.001875	0.001772	0.001778	0.001818
	$\text{se}(\hat{\beta}_2)$	0.000815	0.000832	0.000847	0.000837	0.000831	0.000796
	$\text{se}(\hat{b}_2)$	0.000650	0.000806	0.000645	0.000670	0.000663	0.000297
6	$\hat{\beta}_1$	0.000388	0.000473	0.000400	0.000392	0.001248	0.001250
	$\text{se}(\hat{\beta}_1)$	0.000391	0.000410	0.000365	0.000378	0.156022	0.160617
	$\text{se}(\hat{b}_1)$	0.000419	0.000505	0.000245	0.000343	0.517428	0.532669
	$\hat{\beta}_2$	-0.000281	0.000687	-0.000267	-0.000266	-0.000296	-0.000295
	$\text{se}(\hat{\beta}_2)$	0.001830	0.707918	0.001793	0.001813	0.001977	0.001977
	$\text{se}(\hat{b}_2)$	0.004000	2.347782	0.003860	0.003936	0.004538	0.004536

Figure 3.38

The spatial function maps from λ_{spat}=10 in model 2 and model 4.

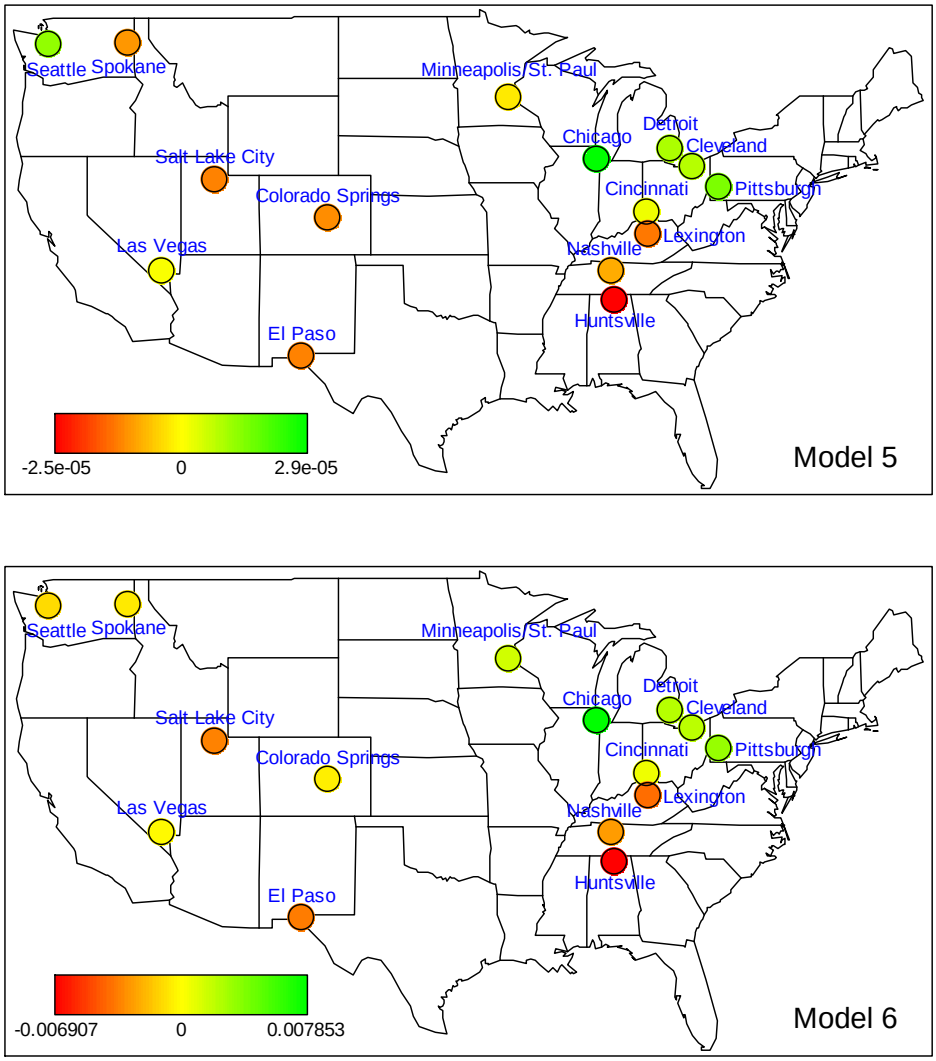


ranges of the other three co-pollutants' estimates under convergence results were (0.001215, 0.001256), (0.001772, 0.001855) and (-0.000281, -0.000266) in NO_2, O_3 and SO_2, respectively. We can have the same conclusion from chapter 3.5 and chapter 3.6 that λ_{spat} can reach robust results based on convergence, and the stability of $\hat{\beta}$ and se($\hat{\beta}$) was higher than that of se($\hat{b}$).

Therefore, the remaining concern is whether spatial function can maintain its

Figure 3.39

The spatial function maps from λ_{spat}=10 in model 5 and model 6.



consistence, and we compared spatial function maps from convergence results in each model. A new investigation is that some spatial effects almost disappeared. This situation can be easily identified by reviewing the palette located in the left-bottom corner of each spatial function map. Each palette was labeled by minimum, zero and maximum spatial effect, and the log relative rate of each city was distributed within the range of band. In model 1, convergent estimates with λ_{spat} = 10, 12 and 14 had spatial effects located in (-1.33, 1.37),

(-1.08, 1.12) and (-0.91, 0.94), respectively, but another convergence result from λ_{spat} = 15 had small spatial effect with range of (-0.02, 0.02), which made the differences of spatial influence in any couple of cities almost diminished. Of course, this kind of spatial effect was not desired anymore. Besides model 1, the other models (excluding model 3) had the same situation, and they had a common point: all diminished spatial effects appeared in λ_{spat} = 10. Figure 3.38 and figure 3.39 gather the initial spatial function maps from model 2, model 4, model 5 and model 6, which shared the same characteristic: very narrow bandwidth of the range of spatial effect in the palette. After using other λ_{spat}s, model 2 found another convergence estimates result with wider spatial effect range (-0.51, 0.66). Model 4 still had even shrinkage spatial effect from -0.0006 o 0.0005 with λ_{spat} = 12, but its geographical variation became more obvious when as model 4, but finally explicit spatial influence appeared in λ_{spat} = 13, 14 and 15. Finally, we also obtained larger spatial function ranged from (-0.77, 0.81) and (-1.39, 1.39) in λ_{spat} = 12 and 13, compared with initial range (-0.007, 0.007) based on default λ_{spat}.

To sum up, λ_{spat} may not be as efficient as λ_{time} and λ_{tmean}, and it's better to regard λ_{spat} as a sort of double-checking criterion when the other estimates and smoothing functions have been fitted well. In the other words, the consideration of λ_{spat} can be ignored temporarily before obtaining convergent and reasonable estimates and smoothing functions. After checking each city's spatial effect, we suggest process the second model-fitting by adjusting λ_{spat} if extremely small spatial effects were discovered.

3.9. GGAMM v.s. 2-stage Bayesian hierarchical model

For the first glance of applying the GGAMM in air pollution research, we are interested

in comparing how different are the GGAMM and traditional 2-stage Bayesian hierarchical model. From model structure, a 2-stage Bayesian hierarchical model was composed by several distinguished GAMs from cities, and then included the estimates of main factors (air pollutants) from those GAMs into a two-level normal model with a Bayesian approach to make a national estimate. Actually, the models from two stages cannot be shown together, or say, they even do not have an organized and unified model form to present. Rather than a mathematical model, as it is a conceptual model. The GGAMM uses a single model form to fit both of national and city-specific effect simultaneously, and realizes that all estimates are based on a statistical model. More importantly, the estimates from GGAMMs showed somehow difference from the estimates from 2-stage Bayesian hierarchical models.

Table 3.7 shows both parameter estimates fitted from GGAMMs and 2-stage Bayesian hierarchical models, and the PM_{10} effects in a 2-stage Bayesian hierarchical model were generally stronger than those in the GGAMM, except for 1-day lag effect in model 2 and model 6. Considering their ratios, the current PM_{10} effect of model 2 fitted from 2- stage Bayesian hierarchical model has the highest ratio with value of 4.34 to that from the GGAMM, but its 1-day lag effect was much weaker than GGAMM's 1-day lag effect with ratio of 4.03. Model 1's PM_{10} in the 2-stage Bayesian hierarchical model was also 3.54 times higher than the same effect in the GGAMM. When controlling for co-pollutants, the ratios of PM_{10} effect between two models were reduced to 0.78~2.78. The estimated standard errors of PM_{10} effects fitted by the 2-stage Bayesian hierarchical model were slightly higher than those fitted by the GGAMM, except for model 3. Actually, from this comparison, we can confirm that the estimated standard errors of fixed effect modified by jackknife-bootstrap method in model 3 were underestimated.

Table 3.7

The comparison of estimates from GGAMMs and 2-stage Bayesian hierarchical models.

		GGAMM		2-stage BH model	
Model #	Parameter	$\hat{\beta}$	se($\hat{\beta}$)	$\hat{\beta}$	se($\hat{\beta}$)
Model 1	PM_{10}	0.000105	0.000287	0.000235	0.000461
Model 2	PM_{10}	0.000093	0.000355	0.000060	0.000537
	PM_{10}-lag1	-0.000037	0.000370	-0.000126	0.001391
	PM_{10}-lag2	0.000142	0.000337	0.000117	0.000563
Model 3	PM_{10}	0.000206	0.000027	0.000397	0.000666
	CO	-0.000005	0.000001	-0.000008	0.000027
Model 4	PM_{10}	0.000141	0.000413	0.000038	0.000779
	NO_2	0.001256	0.000866	0.001313	0.001392
Model 5	PM_{10}	0.000223	0.000414	0.000176	0.000607
	O_3	0.001772	0.000837	0.001405	0.001448
Model 6	PM_{10}	0.000392	0.000378	0.000541	0.000770
	SO_2	-0.000266	0.001813	0.000406	0.002085

* Model 2 fitting by the GGAMM was modified by starting values (λ_{time}, λ_{tmean}, λ_{spat})=(11, 10, 13).
* Model 4 fitting by the GGAMM was modified by starting values (λ_{time}, λ_{tmean}, λ_{spat})=(11, 10, 14).
* Model 5 fitting by the GGAMM was modified by starting values (λ_{time}, λ_{tmean}, λ_{spat})=(10, 10, 13).
* Model 6 fitting by the GGAMM was modified by starting values (λ_{time}, λ_{tmean}, λ_{spat})=(10, 10, 13).
* Model 3 fitting by the GGAMM was modified by jackknife-bootstrap approach.

Comparing with PM_{10} effect, the co-pollutant effects fitted by the 2-stage Bayesian hierarchical model were reversely weaker than those fitted by the GGAMM. In particular, NO_2 effect in the 2-stage Bayesian hierarchical model was 2.33 times smaller than the NO_2 fixed effect in the GGAMM. The ratio became 4.16 in O_3 effect. However, both estimated standard errors in the 2-stage Bayesian hierarchical model were larger than those in the GGAMM, which implied that the confidence intervals of NO_2 and O_3 effect in the GGAMM were narrower. In particular, the fixed effect of SO_2 became positive as long as using the

2-stage Bayesian hierarchical model, but their estimated standard errors were close. Moreover, comparing four co-pollutant models, when using the 2-stage Bayesian hierarchical model and controlling NO_2, O_3 and SO_2, the PM_{10} estimates were changed slightly from 0.000038, 0.000176 and 0.000541, respectively, but it increased to 0.000397 as controlling for CO. In the GGAMM, the PM_{10} fixed effects were robust when controlling for CO, O_3 and SO_2; nonetheless, when controlling for SO_2, the PM_{10} fixed effect significantly increased to 0.000392.

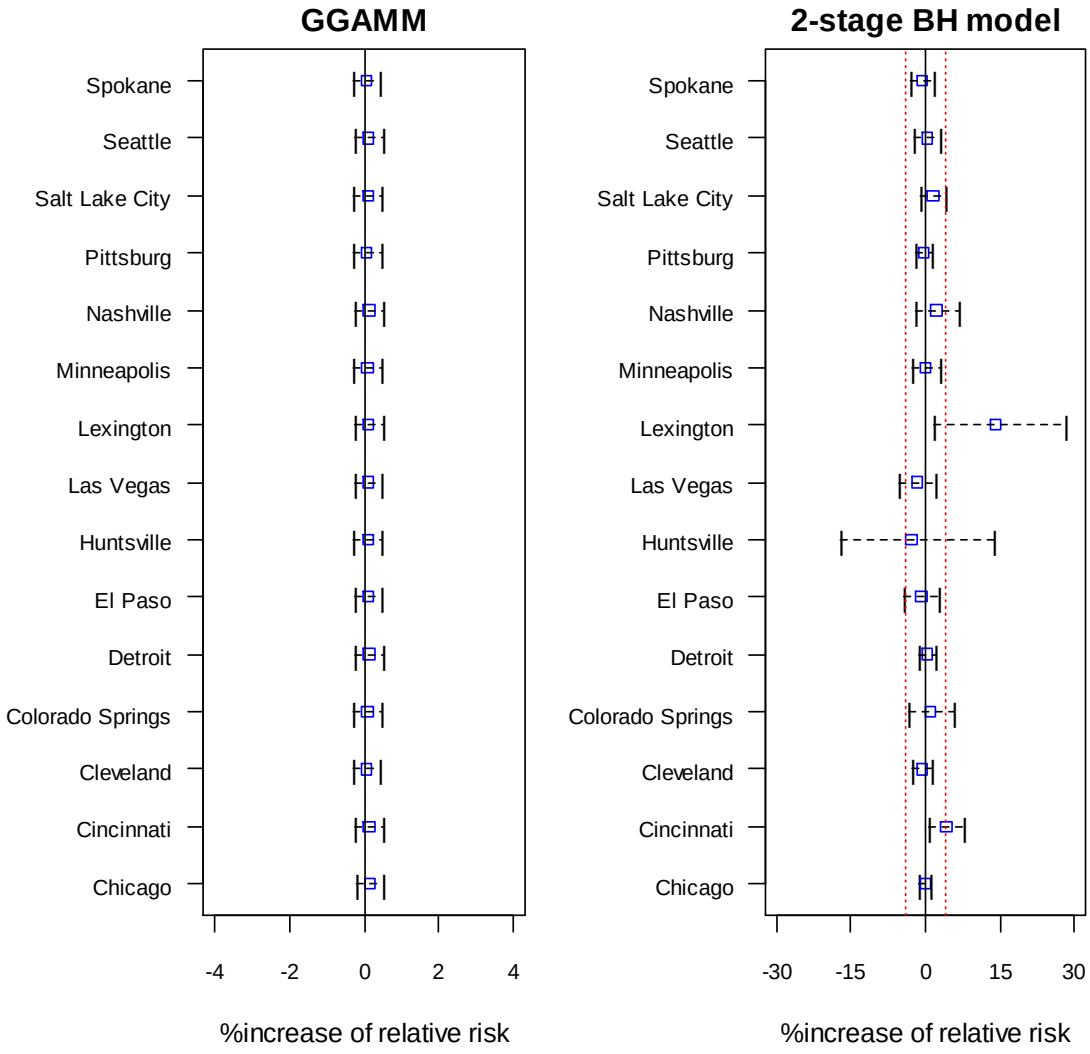
Reviewing the city-specific effects, many differences appeared between two models. Theoretically, the city-specific effect in the 2-stage Bayesian hierarchical model was constructed with several independent GAMs. In figure 3.1, we had proved that air pollutants and temperature were highly correlated when the distance between two cities was short, so geographical correlation cannot be ignored. Even though two-level normal models linked those coefficients from GAMs by two-level normal independent sampling method, the spatial correlation did not contain any real geographical information. The GGAMM can fit several kinds of spatial functions from real geographical data to explain and control most spatial correlation. From this framework, the city-specific effects had different profiles from two models. In the 2-stage Bayesian hierarchical model, the city-specific effect was fitted from each independent GAM in each city. The linear factors in each GAM were controlled by its smoothing functions. Therefore, different city-specific smoothing functions can make city-specific effects more versatile with corresponding confidence intervals. In the GGAMM, the city-specific effects were estimated on a unified framework, so the variation of city-specific effects was also controlled by nation-level smoothers and spatial function, and was finally centralized on fixed effect. In addition, because of being controlled by local

smoothing functions, some city-specific effects in the 2-stage Bayesian hierarchical model were stronger than those in the GGAMM. This property also implied wider 95% confidence intervals of city-specific effects in the 2-stage Bayesian hierarchical model. Besides, we also found that some city-specific effects from the 2-stage Bayesian hierarchical model also had convergence problem with huge estimates and 95% confidence intervals, such as PM_{10} effect at 1-day lag in Cincinnati and PM_{10} effect at 2-day lag in Huntsville (Appendix F). Meanwhile, a well-convergence result in the GGAMM can simultaneously guarantee any estimate at nation-level and city-level.

There were some common characteristics shared from the comparison between two models in each effect. First, the city-specific PM_{10} effects of the GGAMM tended to be positive more than those of the 2-stage Bayesian hierarchical model. In details, as fixed effects were positive, the city-specific effects had higher chance to be positive, and vice versa. Second, the versatility level of city-specific effects of the GGAMM was less than that of the 2-stage Bayesian hierarchical model. Meanwhile, the range of city-specific effects of the GGAMM was narrower than that of the 2-stage Bayesian hierarchical model. For example, in figure 3.40, the range of percent increase in relative risk of mortality per 10 $\mu g/m^3$ increase of PM_{10} concentration in the GGAMM was from 0.06%(Spokane) to 0.15%(Chicago), but the range in the 2-stage Bayesian hierarchical model was wider from -2.94%(Huntsville) to14.19%(Cincinnati). This situation not only appeared in PM_{10} effect, but also happened in PM_{10} lag effects and co-pollutants. It was implied that, in the 2-stage Bayesian hierarchical model, each city-specific effect was fitted from a GAM, and according to our data, each city was easily damaged by missing data because there was only 1,826 observations in each city. Due to applying CCA, missing data in PM_{10} or co-pollutants would also make valid

Figure 3.40

The city-specific PM_{10} effects of the GGAMM and 2-stage Bayesian hierarchical model in model 1.

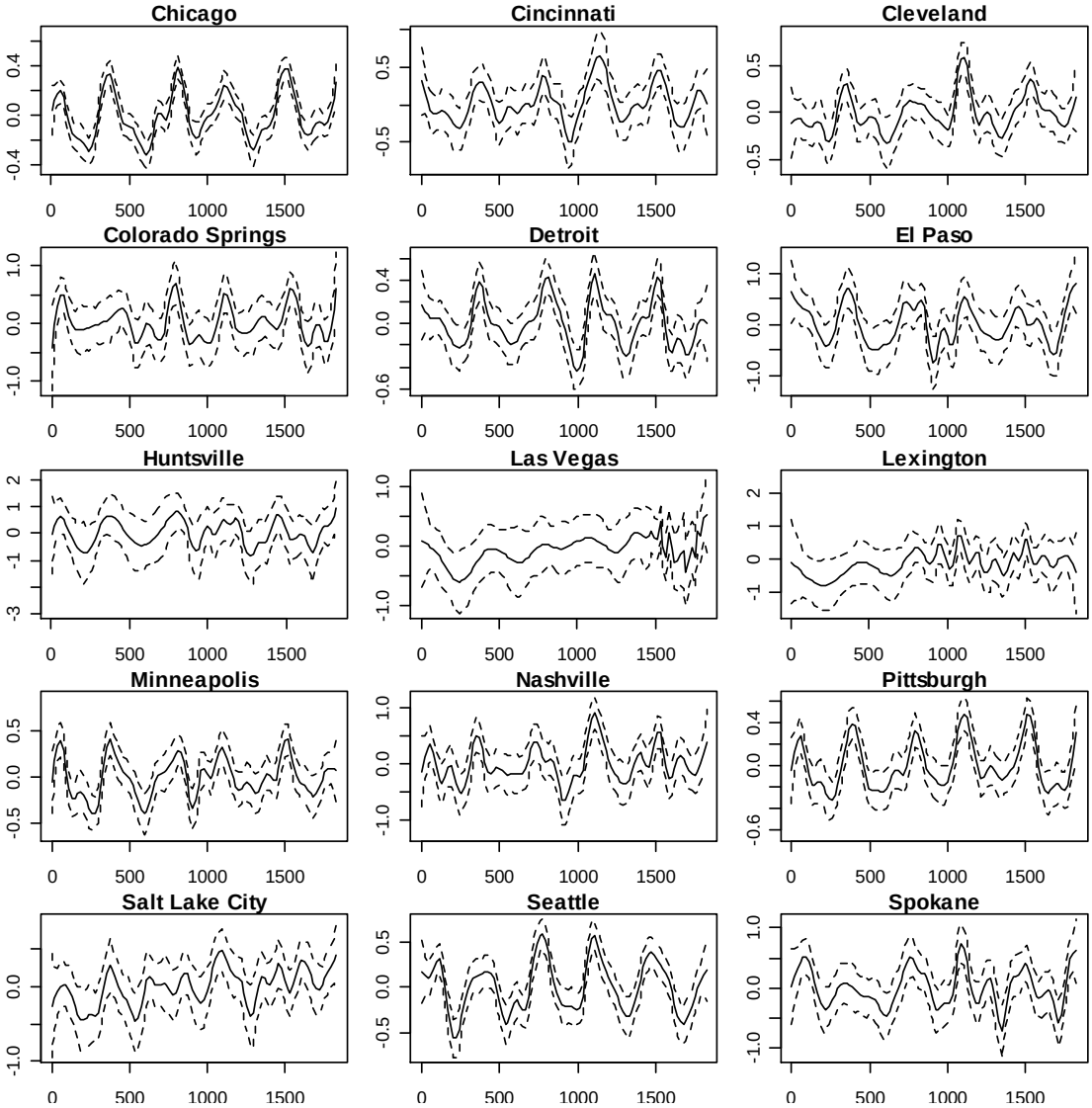


*The red vertical lines in 2-stage Bayesian Hierarchical model plot are the range of x-axis in the GGAMM plot.

temperature or the other weather condition variables deleted, which will make irrational smoothing functions and affect parameter estimation. Figure 3.41 and figure 3.42 are the time smoother plot and temperature smoother plot in model 1 fitted by the GAM. More missing data made time smoother not vary seasonally any more, such as Huntsville, Las Vegas and Lexington. Temperature functions also displayed twisted shapes in most cities, and cannot be

Figure 3.41

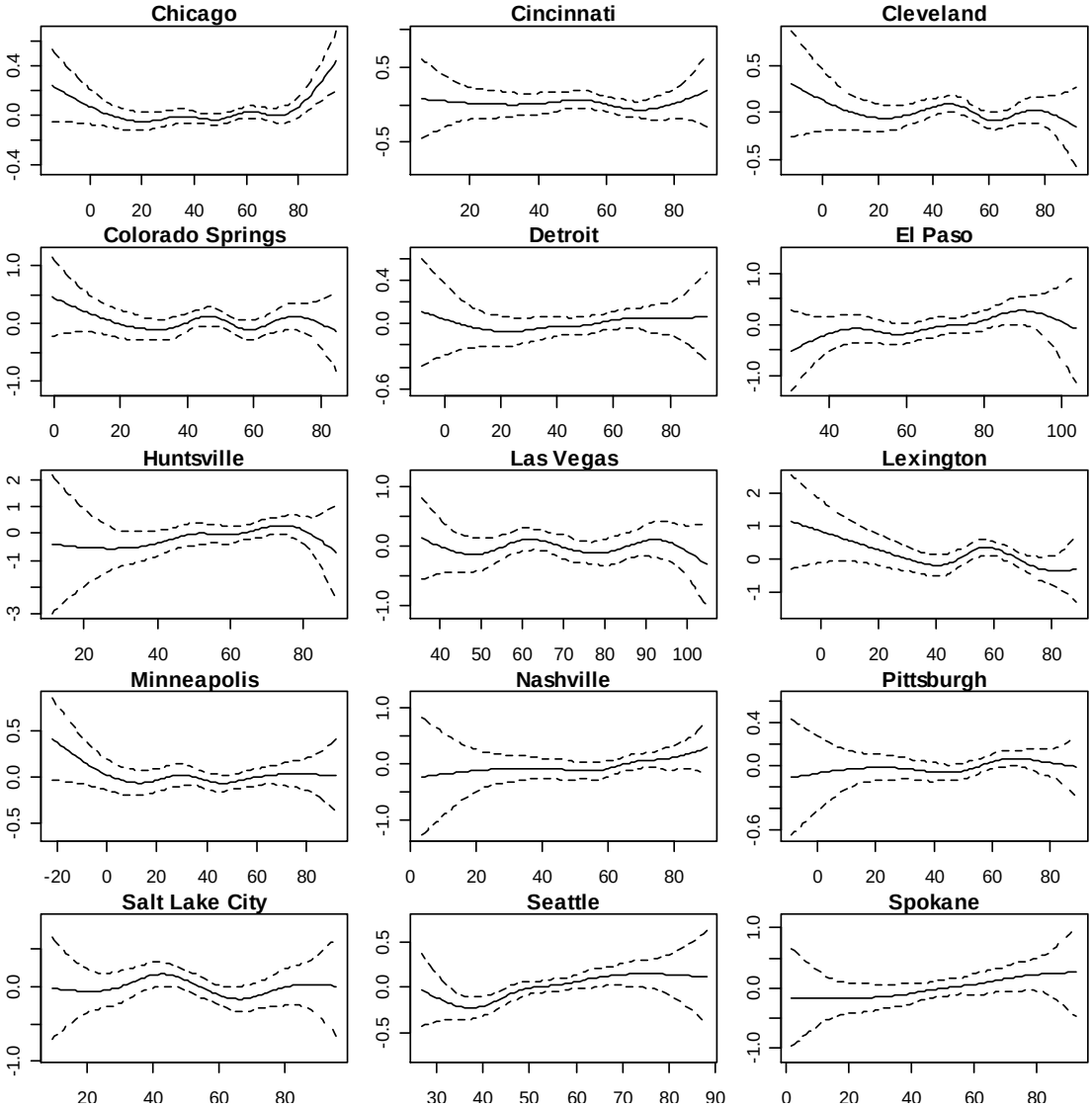
The smoothing function plots of time fitting by the GAM in each city in model 1.



demonstrated. That's why sometimes the ratio of city-specific effect between two locations was over 50-fold, and many negative effects also were unexplainable. The unified structure of the GGAMM can reduce the possibility of risk from missing data and CCA. For instance, there were 574 missing PM_{10} in Cincinnati, and the corresponding temperature and calendar time data will also be ignored by CCA when fitting the GAM. Nonetheless, they still have chance to be complemented from the other cities as fitting the GGAMM. Third, the 95%

Figure 3.42

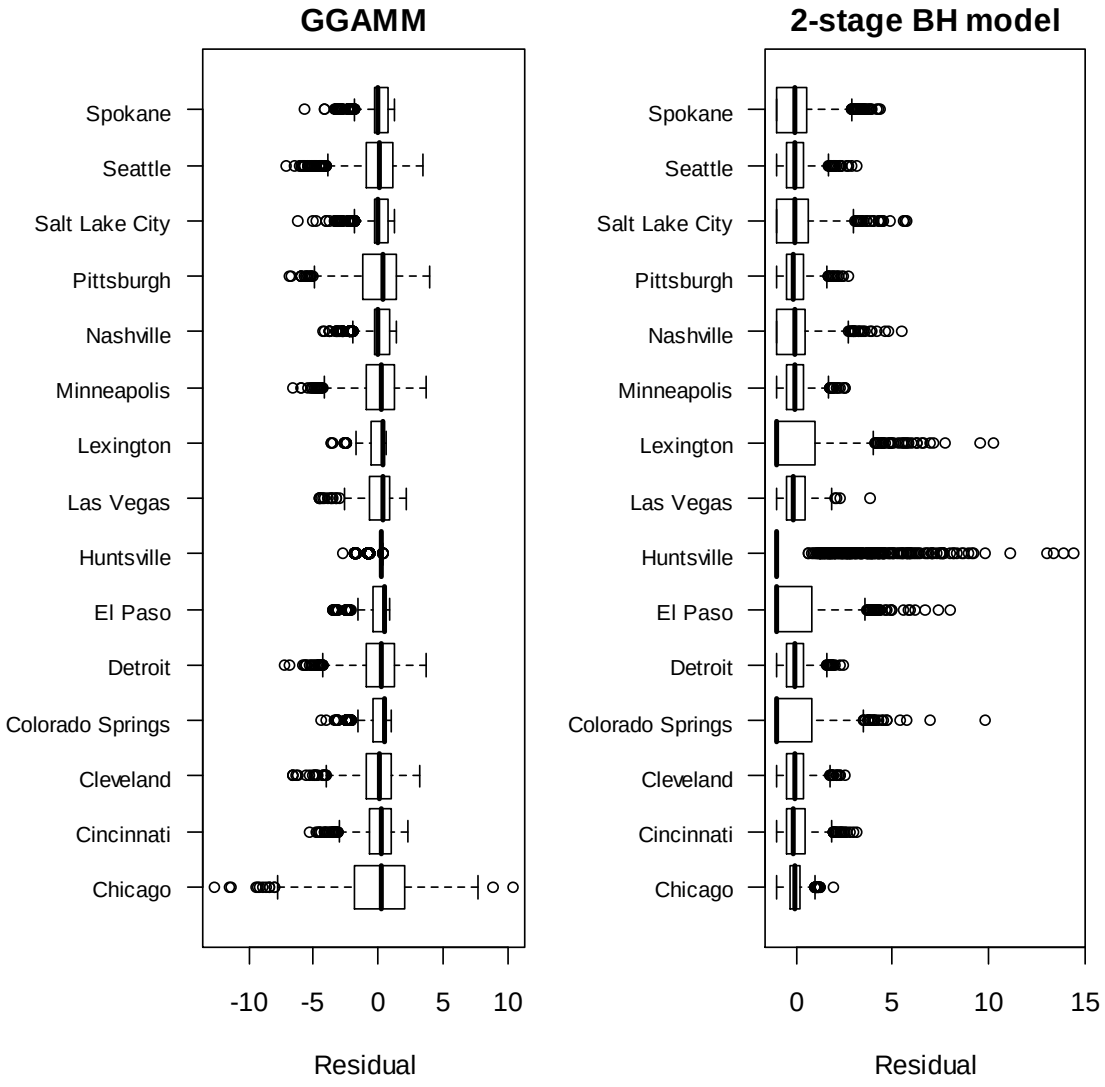
The smoothing function plots of temperature fitting by the GAM in each city in model 1.



confidence interval of city-specific effect in the GGAMM were narrower than the 95% confidence interval of city-specific effect in the 2-stage Bayesian hierarchical model, and the GGAMM can also make less varied 95% confidence intervals among 15 cities. Fourth, due to the city-specific effects in the GGAMM were produced by the summation of fixed effect and random effect, they were all centered around fixed effect, which implied that as long as fixed effects were higher, the city-specific effects were also higher, and vice versa. For example, in

Figure 3.43

The city-level residual box plots of the GGAMM and 2-stage Bayesian hierarchical model in model 1.



model 4, the fixed effects of PM_{10} and NO_2 in the GGAMM were 0.000141 and 0.001256, respectively; hence, all city-specific effects of NO_2 were higher than all city-specific effects of PM_{10}. However, this situation did not happen in 2-stage Bayesian hierarchical models because each city was fitted independently and only controlling its own time smoother and temperature smoother.

To sum up, actually there was no scientific approach or statistical test to evaluate which modeling approach was better or more appropriate; hence, we processed the city-level residual box plot to roughly identify. The two residual box plots from model 1 shown in figure 3.43 had totally different patterns. There was no doubt that the residual means of GGAMMs in each city were very consistently close to zero because of the property of GGAMMs, but some cities had obvious lower residual means in 2-stage Bayesian hierarchical models. Each city had more or less outliers in both models, but the amount of outliers from 2-stage Bayesian hierarchical models was more than the amount of outliers from GGAMMs. Interestingly, the outliers from GGAMMs were mostly negative, but 2-stage Bayesian hierarchical models only produced positive outliers. Considering the distribution of those outliers, GGAMMs had better performance to control those outliers not too far away from residual means, but the locations of outliers in 2-stage Bayesian hierarchical models were more diverse, especially in Colorado Springs, El Paso, Huntsville and Lexington. Those outliers also reflected that predicted values from GGAMMs were often underestimated; nonetheless, over-predicted values always happened in 2-stage Bayesian hierarchical models.

3.10. Missing data imputation analysis

For a long time, analyses of air pollution data were undermined from missing data problem. Even though air pollution monitoring stations were widely built in most large cities, due to the consideration of budget and cost, no air pollution monitoring station can guarantee that they can 100% collect every air pollution materials 24 hours without rest. From the NMMAPS database, we can induce some situations of missing data appearing in air pollutants specifically. For six main air pollutants (CO, NO_2, O_3, SO_2, PM_{10}, $PM_{2.5}$) from

1987 to 2000 among 108 cities, there were only 32, 11, 21, 12 cities with complete data in CO, NO_2, O_3 and SO_2, respectively (Table 3.8). The missing data problem went worse in particulate matters PM_{10} and $PM_{2.5}$, and no city had complete PM_{10} and $PM_{2.5}$ over 14 years at all. In particular, most air pollution monitoring stations started to collect $PM_{2.5}$ data after 1999. If we just look at the valid period of data collection of $PM_{2.5}$, there was still no city with complete $PM_{2.5}$ data from 1999 to 2000. Moreover, the duration was still not enough for a spatio-temporal data analysis with other co-pollutants, that's why $PM_{2.5}$ was not listed in this study.

The missing data pattern was also not totally regular among different air pollutant

Table 3.8

The cities with complete air pollutant data in 108 U.S. cities from 1987 to 2000.

Air pollutant	Cities
CO	Akron, Albuqerque, Boston, Chicago, Cincinnati, Cleveland, Columbus, Dayton, Denver, Dallas/Fort Worth, El Paso, Fresno, Houston, Indianapolis, Kansas City(MO), Los Angeles, Louisville, Memphis, Milwaukee, Nashville, Norfolk, Richmond, Sacramento, Salt Lake City, San Bernardino, Seattle, Spokane, Santa Ana/Anaheim, St. Petersburg, Tampa, Tucson, Wichita
NO_2	Bakersfield, Boston, Chicago, Dallas/Fort Worth, Fresno, Houston, Kansas City(MO), Los Angeles, Oakland, San Bernardino, Santa Ana/Anaheim
O_3	Albuqerque, Bakersfield, Baton Rouge, Chicago, Denver, Dallas/Fort Worth, El Paso, Fresno, Houston, Los Angeles, Little Rock, Nashville, Oakland, Riverside, Sacramento, San Bernardino, Shreveport, Santa Ana/Anaheim, St. Petersburg, Tampa, Tucson
SO_2	Boston, Cleveland, Detroit, Houston, Indianapolis, Kansas City(MO), Los Angeles, Milwaukee, Pittsburg, Providence, St. Petersburg, Tampa
PM_{10}	None
$PM_{2.5}$	None

monitoring stations, even though some stations collected data once per six days (PM_{10} in Los

Angeles) or half-year cycle (O_3 in Akron). In statistical analysis, it is supposed that the missing data scenario was considered with missing completely at random (MCAR), and all missing data was eliminated when fitting models in chapter 3.2. This data management strategy is named as complete case analysis. We were interested in how missing data imputation methods work in the GGAMM, and look for opportunities that missing data imputation methods can repair the damage of missing data along with improving model-fitting.

Table 3.9 was organized by main estimates with CCA, NNI1, NNI2 and MI. When applying missing data imputations, we also encountered the convergence problem to get reasonable estimates in some models, such as model 4 with NNI2 or model 6 with NNI2. imputations have convergent results simultaneously. We tried to search good starting values of smoothing parameter in either time smother or temperature smoother in some imputations within each model, but only model 1, model 2, model 4, model 5 and model 6 can accomplish this. Note that model 3 with CCA, NNI1, NNI2 and MI-MCMC all used initial settings because there was no good result whatever any starting value of smoothing function or the number of knots was used.

Comparing the estimated PM_{10} fixed effect over six models, the NNI1 increased its fixed effect in model 1, model 2 and model 3, and the remaining co-pollutant models did not raise this effect, especially in model 4, where after controlling for NO_2, the PM_{10} fixed effect with NNI1 reduced around 8 times to the same effect with CCA. As imputed by NNI2, the PM_{10} fixed effects were reduced besides model 3. In particular, the PM_{10} fixed effect became negative in model 1 and model 2. The largest decrement appeared in model 5, where PM_{10}

Table 3.9

The model-fitting results from complete case analysis (CCA), nearest neighbor imputation version 1 and version 2 (NNI1, NNI2) and multiple imputation (MI-MCMC).

		$\hat{\beta}$				se($\hat{\beta}$)				se($\hat{b}$)			
Model	Variable	CCA	NNI1	NNI2	MI-MCMC	CCA	NNI1	NNI2	MI-MCMC	CCA	NNI1	NNI2	MI-MCMC
Model 1	PM_{10}	0.000105	0.000186	-0.000044	0.000139	0.000287	0.000345	0.000262	0.000295	0.000194	0.000691	0.000247	0.000434
Model 2	PM_{10}	0.000093	0.000202	-0.000013	0.000127	0.000355	0.000322	0.000320	0.000330	0.000414	0.000417	0.000554	0.000540
	PM_{10}-lag1	-0.000037	-0.000069	-0.000102	0.000008	0.000370	0.000355	0.000319	0.000322	0.000385	0.000540	0.000509	0.000412
	PM_{10}-lag2	0.000142	0.000065	0.000047	0.000061	0.000337	0.000371	0.000283	0.000316	0.000389	0.000767	0.000371	0.000502
Model 3	PM_{10}	0.000196	0.000546	0.000258	0.000183	0.088258	0.076300	0.125322	0.082237	0.341816	0.295501	0.485365	0.318495
	CO	-0.000005	-0.000010	-0.000011	-0.000009	0.081648	0.081645	0.081634	0.081649	0.316223	0.316210	0.316165	0.316226
Model 4	PM_{10}	0.000141	0.000021	0.000114	-0.000234	0.000413	0.000364	0.000301	0.000437	0.000480	0.000474	0.000345	0.000818
	NO_2	0.001256	0.000748	0.000459	0.001658	0.000837	0.000727	0.000663	0.000879	0.000951	0.000835	0.000855	0.001668
Model 5	PM_{10}	0.000223	0.000146	0.000004	0.000027	0.000378	0.000314	0.000266	0.000330	0.000672	0.000435	0.000242	0.000564
	O_3	0.001772	0.001496	0.000546	0.002448	0.000837	0.000652	0.000598	0.000825	0.000670	0.000786	0.000836	0.001663
Model 6	PM_{10}	0.000392	0.000355	0.000373	0.000519	0.000378	0.000394	0.000428	0.000391	0.000343	0.000414	0.000714	0.000417
	SO_2	-0.000266	0.000265	0.000251	0.000027	0.001813	0.001875	0.001846	0.001920	0.003936	0.004269	0.004205	0.004408

fixed effect reduced from 0.000227 to 0.000004. MI-MCMC generally had the same performance as NNI1 in PM_{10}, except for model 6, where PM_{10} fixed effect raised 33.76%.

The fixed estimates of lag effects were all underestimated after using missing data imputation methods, besides 1-day lag effect in NNI1, and the most serious reduced situation happened in MI-MCMC. Both of NNI1 and NNI2 could make co-pollutants CO, NO_2 and O_3 decreased their effect from CCA, but SO_2 was estimated as positive effect with value of 0.000265, 0.000251 and 0.000027 as applying NNI1, NNI2 and MI-MCMC, respectively. In addition, MI-MCMC not only can raise the fixed effect of SO_2, but also increase NO_2 and O_3 effects.

The influence of missing data imputation methods in se($\hat{\beta}$) was not as much as that in $\hat{\beta}$, but most of them were still underestimated, which means that the confidence interval of $\hat{\beta}$ would become narrower. The se($\hat{\beta}$) from NNI2 were all decreased besides SO_2 in model 6 and PM_{10} in model 3. Its decrement was all larger than NNI1 and MI. These missing data imputation methods seemed to have no much influence on se($\hat{\beta}$) in model 6, except for PM_{10} in NNI2. Also, MI had similar se($\hat{\beta}$) with CCA in most models, especial in model 1, model 4 and model 5.

As we had reported, the versatility of city-specific effects in the GGAMM was not too much because it was well controlled on the entire model structure. Sometimes we were wondering whether more valid data can increase the versatility of city-specific effects, so the missing data imputation would be expected to be a good tool to make more variation among cities. The best index was to see whether the estimated standard errors in random effects were increased with imputed data. From our analysis in table 3.9, the PM_{10} random effects indeed had an increase when applying missing data imputation, especially in NNI1, which significantly increased se($\hat{b}$) 3.56 times, but NNI1 and NNI2 cannot keep the same efficacy in co-pollutant models (model 4 & model 5). However, the

se($\hat{b}$)s in co-pollutants were all increased, besides CO, which was almost not varied from CCA. This increment went larger as long as using MI-MCMC. For example, the se($\hat{b}$)s of NO_2 and O_3 in MI-MCMC were 2.10-fold and 2.55-fold to values in CCA. In lag effects, the se($\hat{b}$)s of 1-day lag effect had a slight reduce with missing data imputation methods, but not too far away from the value of CCA. However, the se($\hat{b}$)s of 2-day lag effect had much increase, especially in NNI1, where se($\hat{b}$) was raised from 0.000767 to 0.000244. This implied that the versatility level of city-specific effects in shorter lag effects would be deducted and smaller than the versatility level of city-specific effects in relatively longer lag effects. Note that the relative comparison between two lag effects in model 2 could be changed if more lag effects were included. Nonetheless, this will add some additional problems, so we will discuss distributed lag models with more lag effects of the GGAMM in section 3.11.

3.11. Extended distributed lag models

Theoretically, a GGAMM can include more lag effects, but technically the sophisticated estimation procedure and huge spatial-temporal data will make BayesX crash easily. Suppose we really need more lag effects to control short-term, middle-term or even long-term influence, it is not available in current version of BayesX technically. An alternative way to evaluate longitudinal influence of air pollutant to adverse human health is using a variable reduction approach to reduce the number of variables fitted by the GGAMM. Here we applied principal component analysis (PCA) to solve this problem.

When considering 6 lag effects of PM_{10}, we extended the longitudinal influence of fine particulate to one week. Note that the standardization for each fine particulate variable was not used because they have the same unit. After applying PCA, the

Table 3.10

The eigenvalues of 7 principal components with corresponding distinguished and cumulative contribute of the total variability over PM_{10} and 1-day~6-day lag PM_{10} effect.

Principal component	Eigenvalue	Difference	Proportion	Cumulative
1	3.05540655	1.82111631	0.4365	0.4365
2	1.23429024	0.34165427	0.1763	0.6128
3	0.89263598	0.27240082	0.1275	0.7403
4	0.62023516	0.13884689	0.0886	0.8289
5	0.48138827	0.11210126	0.0688	0.8977
6	0.36928701	0.02253024	0.0528	0.9505
7	0.34675678	--	0.0495	1.0000

eigenvalue and the proportion of explanation over seven variables in each principal component were listed in table 3.10. We defined the minimum cumulative account of the total variability over 7 variables was 70%, and then chose the first three principal components with its own explanation proportion 43.65%, 17.63% and 12.75%, respectively, in the extended lag distributed model. The remaining principal components' contributions were all less than 10%, and were ignored in the following model-fitting.

The loadings of each variable on each principal component were listed in table 3.11. The loadings in the first principal component PRIN1 were averagely and approximately located from 0.32~0.40, and it can be regarded as a measure of moving average of PM_{10} concentration from past one week. The second principal component PRIN2 had positive loadings on 4-day~6-day lag PM_{10} effect, and negative loadings on current to 3-day lag effect. The third principal component PRIN3 alternatively had positive loadings on the most recent PM_{10} effects (current and 1-day lag) and the farthest PM_{10} effects (5-day and 6-day lag), and the middle PM_{10} effect (2-day~4-day lag) were calculated with negative

Table 3.11

The eigenvectors of the first three principal components when condensing current and 1-day~6-day lag PM_{10} effect.

Variable	PRIN1	PRIN2	PRIN3
PM_{10}	0.3304	-0.4157	0.4666
PM_{10}-lag1	0.3805	-0.4778	0.2069
PM_{10}-lag2	0.4054	-0.3099	-0.2907
PM_{10}-lag3	0.4097	-0.0001	-0.5556
PM_{10}-lag4	0.4031	0.3100	-0.9545
PM_{10}-lag5	0.3779	0.4819	0.2079
PM_{10}-lag6	0.3296	0.4177	0.4646

loadings. PRIN2 was measuring the "after-half" PM_{10} effect within the past one week. PRIN3 seemed to measure a fluctuation trend of PM_{10} effect with a cycle of 7 days, and reflected that the influence of PM_{10} concentration on human health was weekly. However, this cycle was not very strong, and can only explain 12.75 percent of the total variance.

The data in each city and each day re-calculated the three principal component variables by the following three equations:

$$PRIN1=0.03304\times X_1+0.3805\times X_2+0.4054\times X_3+0.4097\times X_4+0.4031\times X_5+0.3779\times X_6+0.3296\times X_7$$

$$PRIN2=-0.4157\times X_1-0.4778\times X_2-0.3099\times X_3-0.0001\times X_4+0.3100\times X_5+0.4819\times X_6+0.4177\times X_7$$

$$PRIN3=0.4666\times X_1+0.2069\times X_2-0.2907\times X_3-0.5556\times X_4-0.9545\times X_5+0.2079\times X_6+0.4646\times X_7$$

where X_1~X_7 were original current and 1-day~6-day lag PM_{10} effect. As fitting the GGAMM in BayesX, the results are shown in table 3.12. PRIN1 can be interpreted that when the PM_{10} effect averagely increased per 1 $\mu g/m^3$ in each day of past one week, the relative risk of mortality rate of respiratory diseases in elders generally increased 2.28% (95% CI: -7.95%, 13.64%). The strongest effect was estimated in PRIN3 with value of

Table 3.12

The estimated parameter estimates with corresponding estimated standard errors of fixed and random effects in the GGAMM with 3 principal components from current and 1-day~6-day lag PM_{10} effect.

Variable	$\hat{\beta}$	se($\hat{\beta}$)	se($\hat{b}$)
PRIN1	0.002254	0.005375	0.011141
PRIN2	0.000267	0.012372	0.034817
PRIN3	0.003819	0.008674	0.017478

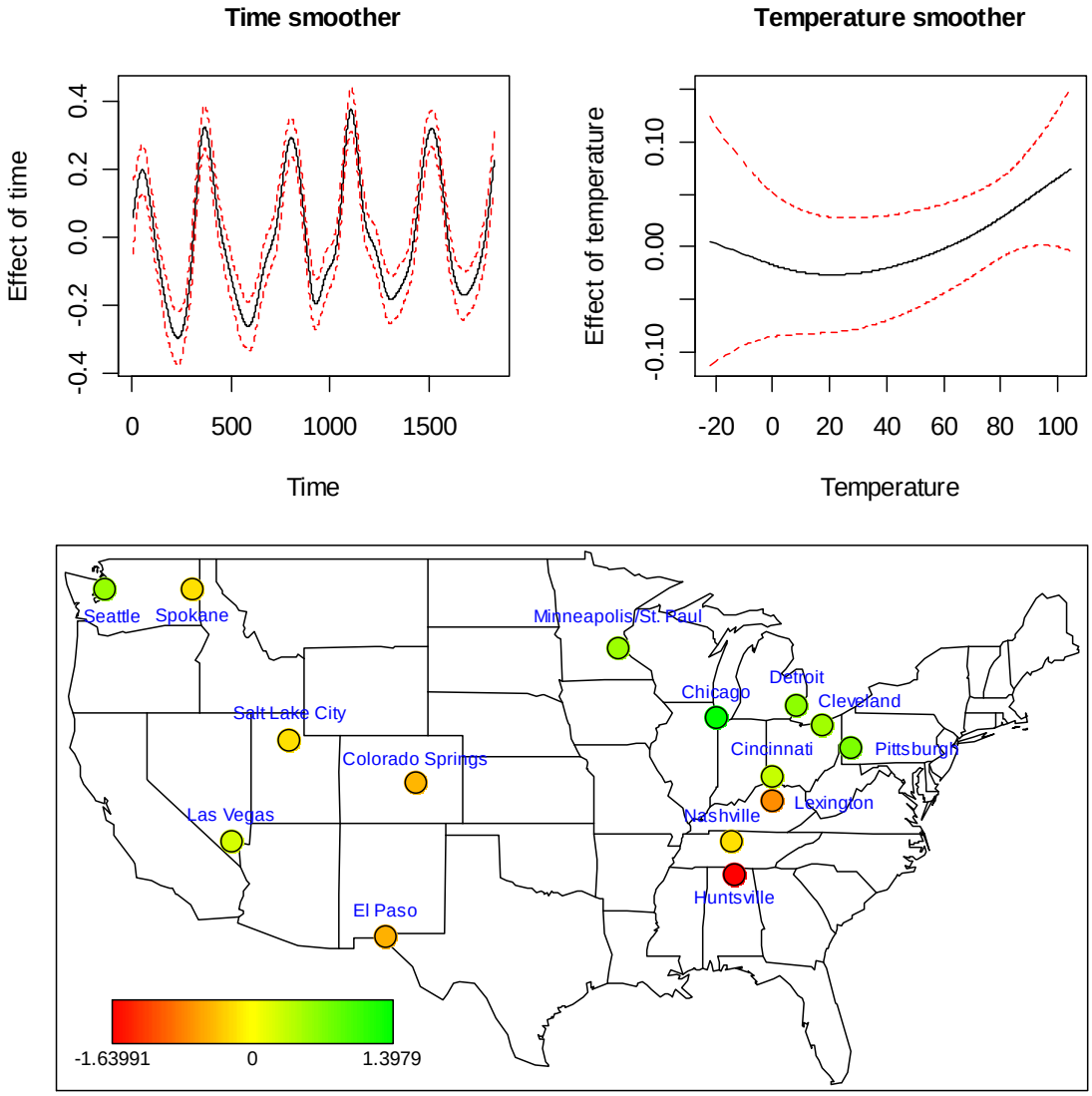
0.003819 (95% CI: -0.013186, 0.020823), and the weakest effect was PRIN2 with value of 0.000267 (95% CI: -0.023987, 0.024521). They cannot be explained as PRIN1, but we can conclude that the PM_{10} week-cycle influence was 14.30 times higher than the PM_{10} "after-half" influence to mortality rate of respiratory diseases in elders. Reviewing the explanation proportion of the entire variance in PRIN2 and PRIN3 (17.63% v.s. 12.75%), there is no huge difference between them, and we can conclude that the PM_{10} effect on respiratory disease in elders reacted much highly in extreme close and farthest day within the past one week.

Smoothing function plots showing in figure3.44 reflected similar patterns as model 2. There was no doubt that time smoother went up in spring and winter season, and went down in summer and fall season. The temperature smoother also displayed a climbing trend when temperature was getting higher and lower. This plot got better performance because the effect went up more significantly when temperature$\leq$0°F compared with the same plot in figure 3.6. The spatial map also pointed out that the highest geographical influence appeared in some heavy-industrial cities, such as Chicago, Pittsburg and Detroit.

The similar study design can be modified by extending the number of lag effect to 14 days if we are interested in the influence of PM_{10} effect in the past a half of month.

Figure 3.44

Smoothing functions of calendar time and 24-hour average temperature and map of spatial effect for 15 U.S. cities from 1991 to 1995 in the GGAMM with 3 principal components from current and 1-day~6-day lag PM_{10} effect.



Using the same criterion in the previous model, we can choose the first 6 principal components which accounted for 73.60% of the total variance over 15 original variables (table 3.13). The loadings in each principal component are shown in table 3.14. We found a similar pattern as using three principal components in the first extended distributed lag model. The first three components in this model performed the same relative loading trends as previous model. The difference was the duration was extended to 15 days.

Table 3.13

The eigenvectors of the first three principal components when condensing current and 1-day~14-day lag PM_{10} effect.

Principal component	Eigenvalue	Difference	Proportion	Cumulative
1	4.8218	3.0983	0.3215	0.3215
2	1.7235	0.3274	0.1149	0.4364
3	1.3961	0.2014	0.0931	0.5294
4	1.1947	0.1209	0.0796	0.6091
5	1.0738	0.2743	0.0716	0.6807
6	0.7995	0.1263	0.0533	0.7340
7	0.6732	0.0860	0.0449	0.7789
8	0.5872	0.0504	0.0391	0.8180
9	0.5369	0.0964	0.0358	0.8538
10	0.4404	0.0336	0.0294	0.8832
11	0.4069	0.0143	0.0271	0.9103
12	0.3925	0.0319	0.0262	0.9365
13	0.3606	0.0534	0.0240	0.9605
14	0.3072	0.0219	0.0205	0.9810
15	0.2854	--	0.0190	1.0000

PRIN1 also measured a moving average over 15 original variables. Because the lag effect was longer, the loading in PRIN1 was diluted and became smaller. In table 3.11, the loading in PRIN1 can reach as high as 0.4, but it did not surpass 0.3 in table 3.14 anymore. The explanation proportion was also decreased from 43.65% to 32.15%. PRIN2 still displayed a quadratic influence, and the turnover point was deferred to 7-day lag. The loadings in PRIN2 were as diluted to around 0.2 and 0.3 approximately, but the explanation proportion was only weakened 6.14%. PRIN3 became a cubic fluctuation, and the peak of loadings also appeared in two extreme side of this period. Due to the

Table 3.14

The eigenvectors of the first six principal components when condensing current and 1-day~14-day lag PM_{10} effect.

Variable	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6
PM_{10}	0.2343	-0.2462	0.2412	-0.2281	0.3383	-0.3854
PM_{10}-lag1	0.2555	-0.3341	0.2701	-0.2926	0.2254	-0.1843
PM_{10}-lag2	0.2660	-0.3581	0.2197	-0.1863	-0.0830	0.2223
PM_{10}-lag3	0.2647	-0.3188	0.1074	0.0512	-0.3616	0.3595
PM_{10}-lag4	0.2667	-0.2534	-0.0597	0.3080	-0.3609	0.0859
PM_{10}-lag5	0.2680	-0.1695	-0.2310	0.4117	-0.1107	-0.2344
PM_{10}-lag6	0.2706	-0.0785	-0.3546	0.2895	0.2071	-0.2611
PM_{10}-lag7	0.2694	0.0132	-0.4114	0.0050	0.3150	0.0054
PM_{10}-lag8	0.2672	0.1146	-0.3647	-0.2679	0.1784	0.2871
PM_{10}-lag9	0.2641	0.2133	-0.2346	-0.3826	-0.1111	0.2204
PM_{10}-lag10	0.2624	0.2909	-0.0441	-0.2998	-0.3119	-0.1459
PM_{10}-lag11	0.2590	0.3271	0.1435	-0.0771	-0.3141	-0.3623
PM_{10}-lag12	0.2523	0.3369	0.2719	0.1668	-0.0778	-0.1874
PM_{10}-lag13	0.2422	0.3001	0.3164	0.2864	0.2184	0.2075
PM_{10}-lag14	0.2254	0.2116	0.2694	0.2303	0.3444	0.3830

period's extension, the duration of positive loading showing in each side was prolonged to four days. PRIN4 and PRIN5 can be regarded as polynomial fluctuation, but the starting time of PRIN4 was earlier than PRIN5 about two days. Both of their explanation proportions were reduced to around 7%. The cycle period was shortened to four days in PRIN6, and it can only account for 5% of the total variance over 15 original variables.

As re-calculating six principal component variables from original data and fitting in a GGAMM by BayesX, the result is shown in table 3.15. Among six parameters, only PRIN4 and PRIN5 had positive estimates, and PRIN4 was 1.48 times stronger than

Table 3.15

The estimated parameter estimates with corresponding estimated standard errors of fixed and random effects in the GGAMM with 3 principal components from current and 1-day~6-day lag PM_{10} effect.

Variable	$\hat{\gamma}$	se($\hat{\gamma}$)	se($\hat{\xi}$)
PRIN1	-0.001057	0.004991	0.010029
PRIN2	-0.008722	0.007329	0.016007
PRIN3	-0.001833	0.007171	0.013993
PRIN4	0.007043	0.008052	0.015637
PRIN5	0.004764	0.009011	0.017917
PRIN6	-0.000252	0.010353	0.021634

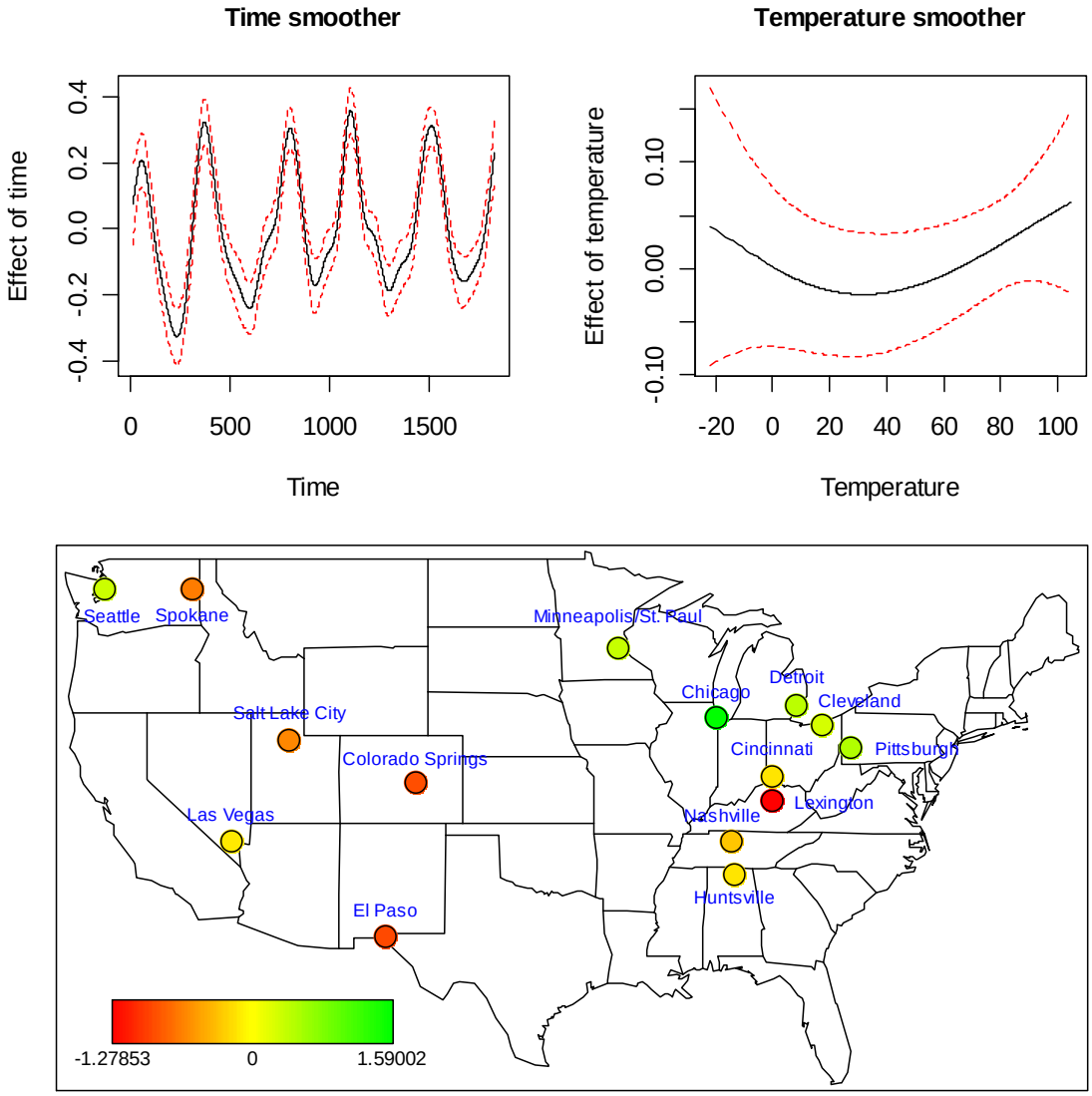
PRIN5. Either one can reflect the previous model with 3 principal component variables from the past one week, and confirm that the week-cycle fluctuation was scientifically the most possible influent period to mortality rate of respiratory diseases in elders.

The smoothing function plots and spatial function map of the second PCA model are shown in figure 3.45. The time smoother was similar as the first PCA model, but the temperature smoother almost became a U-shape. The effect of temperature increased much quicker when temperature was getting lower, and finally reached a similar level as the effect of temperature in extreme hot weather. The relative spatial effects in the map were still the same as previous models, but compared with figure 3.44, both of the highest and lowest spatial effects increased around 0.2~0.3. Chicago still had the strongest spatial effect with value of 1.5900. However, the smallest spatial effect was not located in Huntsville when using 3 principal components, but alternatively located in Lexington. It is worthy of noting that the spatial effect in Huntsville turned from negative to positive (-1.6399→0.0006) when including more principal components from longer lag effect, but the remaining cities generally had the same direction between two models.

Theoretically, the extended distributed lag model can include more principal

Figure 3.45

Smoothing functions of calendar time and 24-hour average temperature and map of spatial effect for 15 U.S. cities from 1991 to 1995 in the GGAMM with 3 principal components from current and 1-day~14-day lag PM_{10} effect.



components derived from more lag effects, but this model design was not recommended because it could raise the difficulty of interpretation. For example, if we consider lag effects with one month, which means 30 lag effect variables should be used in PCA, it can produce 10 candidate principal components to fit a GGAMM. Even though the model-fitting procedure was successful, the result was no longer clarified easily as previous two models.

By using equations (2.51), (2.52) and (2.54), we can transform these principal

Table 3.16

The PCA-adjusted PM_{10} estimates in the first extended distributed lag model.

Variable	$\hat{\gamma}$	se($\hat{\gamma}$)	se($\hat{\xi}$)
PM_{10}	0.000187	0.000504	0.001264
PM_{10}-lag1	0.000165	0.000554	0.001411
PM_{10}-lag2	0.000100	0.000468	0.001106
PM_{10}-lag3	0.000067	0.000414	0.000852
PM_{10}-lag4	0.000009	0.000565	0.001277
PM_{10}-lag5	0.000183	0.000555	0.001418
PM_{10}-lag6	0.000203	0.000505	0.001268

component estimates to obtain original variables' estimates with corresponding fixed and random effects' standard errors. The results are shown in table 3.16 and table 3.17. Reviewing these PCA-adjusted PM_{10} estimates, we found the longitudinal influence of PM_{10} was only extended to one week because the adjusted estimates after 7-day lag in table 3.17 had shown all negative values. Moreover, the strongest effects were not located in the same day. When only considering six lag effects, the pattern of estimates presented a U-shape profile, and current, 1-day lag, 5-day lag and 6-day lag PM_{10} effect displayed relative larger influence to the mortality of respiratory diseases in elders. While considering 14 lag effects, the strongest influence appeared in 4-day and 5-day lag effect. Moreover, the general positive effects in table 3.17 were smaller than corresponding effects in table 3.16. We think it's reasonable because too many lag effects, especially those behind 6-day lag, diluted the influence in each estimate. Also, the se($\hat{\beta}$) and se($\hat{b}$) in table 3.17 were smaller than those in table 3.16, which means the $\hat{\beta}$ in table 3.17 was estimated more conservatively, and the versatility of city-specific effects was also not very distinguished.

Table 3.17

The PCA-adjusted PM_{10} estimates in the second extended distributed lag model.

Variable	$\hat{\beta}$	se ($\hat{\beta}$)	se($\hat{b}$)
PM_{10}	0.000000†	0.000155	0.000102
PM_{10}lag1	0.000011	0.000160	0.000108
PM_{10}lag2	0.000037	0.000154	0.000136
PM_{10}lag3	0.000069	0.000160	0.000156
PM_{10}lag4	0.000100	0.000154	0.000149
PM_{10}lag5	0.000101	0.000153	0.000137
PM_{10}lag6	0.000062	0.000153	0.000128
PM_{10}lag7	-0.000007	0.000148	0.000125
PM_{10}lag8	-0.000081	0.000152	0.000134
PM_{10}lag9	-0.000133	0.000152	0.000147
PM_{10}lag10	-0.000151	0.000153	0.000181
PM_{10}lag11	-0.000134	0.000158	0.000208
PM_{10}lag12	-0.000103	0.000148	0.000206
PM_{10}lag13	-0.000075	0.000155	0.000206
PM_{10}lag14	-0.000055	0.000152	0.000191

†Original value = 0.0000003

The advantages of PCA used in the GGAMM were: 1) resolving the technical problem of BayesX that leads to computer crash when using too many lag effects. From original distributed lag model, we only can estimate up to 2-day lag effect; however, in these extended distributed lag models, we found 3+ principal component variables were still working in BayesX; 2) extended distributed lag model had highly consistent estimating results when the starting values of smoothing parameters changed (detailed results were not shown here); 3) it spent much less time in estimating process compared with fitting original distributed lag effect. Even when including eight principal component

variables, the model-fitting iteration still spent less time than using three original variables. It can contribute to the iteration reaching convergence more quickly; 4) from the benefits of PCA which leads all principal components are independent with each other, the model-fitting of the GGAMM is not affected by the multicollinearity problem.

There are still some disadvantages: 1) compared with previous model 1~model 6, the 95% confidence intervals of estimated parameters in extended distributed lag model were too wide, especially in random effects; 2) it's impossible to exactly identify which lag effect had the strongest influence to the mortality rate of respiratory diseases in elders without advanced calculations. In addition, these estimates cannot be used to conclude a numeric result when each principal component variable increased per unit. Meanwhile, the extended distributed lag model only can explain a general and relative strength over a period of longitudinal trend. An advanced transformation to obtain PCA-adjusted estimates which were calculated by equation (2.51), (2.52) and (2.54) should be considered; 3) the missing data will make severe influence on PCA, and the severity level will also increase along with the number of lag effects. Also, any missing datum appearing in variables using in PCA will not be calculated all of its principal components, so it will immediately affect the valid data using in the model-fitting. For example, a current PM_{10} missing datum appeared in 1991/01/01 made 1-day~6-day lag effect missing in 1991/01/02~1991/01/07, respectively. Therefore, the three principal component variables PRIN1, PRIN2 and PRIN3 were all missing from 1991/01/01~1991/01/07.

3.12. Multicollinearity and concurvity

As what chapter 3.1 and chapter 3.2 had described, some air pollutants were moderately or highly correlated with each other, and they are also connected with time trend and geographical locations. When including them in the GGAMM, it was expected

that the multicollinearity and concurvity problem may exist in model-fitting. The multicollinearity levels were only calculated on model 2~model 6 for evaluating whether this problem was potentially hide between PM_{10} and lag effects or PM_{10} and four co-pollutants. The concurvity levels were examined in all 6 models for identifying how strong relationship between each air pollutant and the combination of time smoother, temperature smoother and spatial function. The two levels were also calculated in data sets applying missing data imputation methods.

In initial data set without dealing with missing data, the highest multicollinearity level appeared between PM_{10} and NO_2 in model 4. Reviewing table 3.2, we can find it is not out of expectation because the correlation between PM_{10} and NO_2 was as high as 0.4804. Even though the strongest relationship to PM_{10} was its 1-day lag effect, the multicollinearity level in model 2 was calculating the relationship between PM_{10} and the combination of its 1-day and 2-day lag effects, and the 2-day lag effects harmonized the high correlation between PM_{10} and its 1-day lag effect. Whatever, the multicollinearity in model 2 was the second highest one over model 2~model 6. Three of them were larger than 0.5, but it's just moderate severity, not high severity. Based on the definition of high multicollinearity level (0.7), these estimates fitted by CCA were not damaged seriously.

In table 3.18, while applying NNI1 and NNI2, we originally doubted that they will increase the multicollinearity level because the two missing data imputation methods were using the nearest datum and its distribution to identify the imputed "candidate" from existing data. Due to such immediate connection from imputed data and reference data which were used for identify "neighbors", our analyses showed the multicollinearity levels from models using NNI1 and NNI2 did not surpass CCA. Compared with NNI1 and NNI2 themselves, NNI2 had lower multicollinearity than NNI1, and NNI2 made the largest reduction in multicollinearity level with value of 0.1205. Moreover, the influence of MI-MCMC on multicollinearity was similar as NNI1, except for model 4, where this

Table 3.18

The multicollinearity and concurvity level in model 1~ model 6.

Model	Multicollinearity	Concurvity		
Model 1		PM_{10}		
CCA	--	0.4515		
NNI1	--	0.4423		
NNI2	--	0.3962		
MI-MCMC	--	0.4267		
Model 2		PM_{10}	1-day lag PM_{10}	2-day lag PM_{10}
CCA	0.5697	0.4534	0.3870	0.3346
NNI1	0.5406	0.4426	0.3834	0.3413
NNI2	0.4773	0.3965	0.3478	0.3142
MI-MCMC	0.5147	0.4269	0.3733	0.3307
Model 3		PM_{10}	CO	
CCA	0.5350	0.4516	0.6632	
NNI1	0.5076	0.4423	0.6531	
NNI2	0.4770	0.3957	0.6428	
MI-MCMC	0.5345	0.4267	0.6479	
Model 4		PM_{10}	NO_2	
CCA	0.5837	0.5136	0.5644	
NNI1	0.5516	0.4856	0.5607	
NNI2	0.4632	0.4209	0.5276	
MI-MCMC	0.6019	0.4660	0.5568	
Model 5		PM_{10}	O_3	
CCA	0.4121	0.4800	0.7777	
NNI1	0.3995	0.4298	0.6861	
NNI2	0.3143	0.3824	0.6324	
MI-MCMC	0.3938	0.4124	0.7511	
Model 6		PM_{10}	SO_2	
CCA	0.4980	0.4944	0.5607	
NNI1	0.4815	0.4921	0.5236	
NNI2	0.4481	0.4199	0.5134	
MI-MCMC	0.4840	0.4779	0.5340	

imputation became worse, and even surpassed the original multicollinearity level (0.5837).

Although we had shown PM_{10} had higher relationship with locations, time and temperature, its concurvity levels in six models were only around 0.45. While applying missing data imputation methods, the concurvity level of PM_{10} did not have too much change in all models. Similar as their performance in multicollinearity, NNI2 had more reduction in concurvity level than NNI1 and MI-MCMC. Note that MI-MCMC made the concurvity levels of PM_{10} exceeded its original value in both of model 4 and model 6, and so did NNI1. Generally speaking, whether using CCA or these missing data imputation methods, they never caused the concurvity level of PM_{10} to be higher than 0.5. As a result, this problem was not severe in fine particulate matter in the GGAMM, but users should pay more attention on that the increase of concurvity level in PM_{10} when more lag effects or 2+ co-pollutants were included in case. The influence of concurvity also did not lead PM_{10}'s lag effects serious. On the contrary, the concurvity level decreased along with the increase of lag day.

The concurvity problem may happen in co-pollutants in the GGAMM because they had higher concurvity level than PM_{10}. The concurvity level of NO_2 and SO_2 were still under control, and only around 0.51~0.56, whether applying missing data imputation methods or not. Reviewing figure 3.3 in chapter 3.2, it's believed that the regular fluctuated time trend of NO_2 and SO_2 contributed most of their concurvity levels. Model 3 had higher concurvity level over 0.6 in CO, and its higher CCFs versus 24-hour average temperature should be the reason that it had higher concurvity level than NO_2 and SO_2 (figure 3.3). Also, missing data imputation methods did not either worsn or improve its concurvity level. The largest concurvity among 4 co-pollutants appeared in O_3. With CCA, its concurvity level was as high as 0.7751, and it can be regarded as severity concurvity problem. It reduced under 0.7 while imputing missing data by NNI1 and NNI2; however,

MI-MCMC made it over 0.7 again. Some evidences shown in figure 3.1~figure 3.3 in chapter 3.2 can explain this situation in O_3. Comparing with the other air pollutants in figure 3.1, the rate of decreasing speed of the city-to-city correlation in O_3 along with the extension of distance was the slowest one. We can set up a horizontal line r=0.6 in all air pollutants in figure 3.1, and find there was no more dot above r=0.6 in PM_{10} as distance≥250 miles. The downward speed of city-to-city correlation in CO, NO_2 and SO_2 was also very quick, and there was almost no dot above r=0.6 as distance≥500 miles. However, there were still some city-to-city correlation of O_3 above or close to r=0.6 as distance≥1,500 miles, so we can reasonably conclude spatial function shall contribute a part of concurvity level of O_3 more than the others. The time trend plot of O_3 almost match up with the time trend plot of temperature in figure 3.2, and the CCFs between O_3 and temperature also confirmed their high relationship over time (figure 3.3).

3.13. Model diagnostics

So far the methodology of model diagnostics based on the GGAMM is still under development because of some unbreakable bottlenecks. These difficulties will be discussed in chapter 5 in details. Here we only applied some existing model diagnostic approaches of general linear model for the GGAMM. Note that these approaches were not guaranteed to appropriately take place in the GGAMM, but we can still take a look at some possible problems existing in the model fitting of the GGAMM, and look for chances to improve in the future as long as relative theorems will be accomplished. Table 3.19 showed separate goodness-of-fit tests of Poisson distribution for death counts from respiratory diseases in elders in each city, respectively. 6 cities (Chicago, Cleveland, Detroit, El Paso, Pittsburg and Seattle) presented that they violated Poisson assumption, with p-values≤0.05, and the overall respiratory disease death count in elders over 15 U.S.

Table 3.19

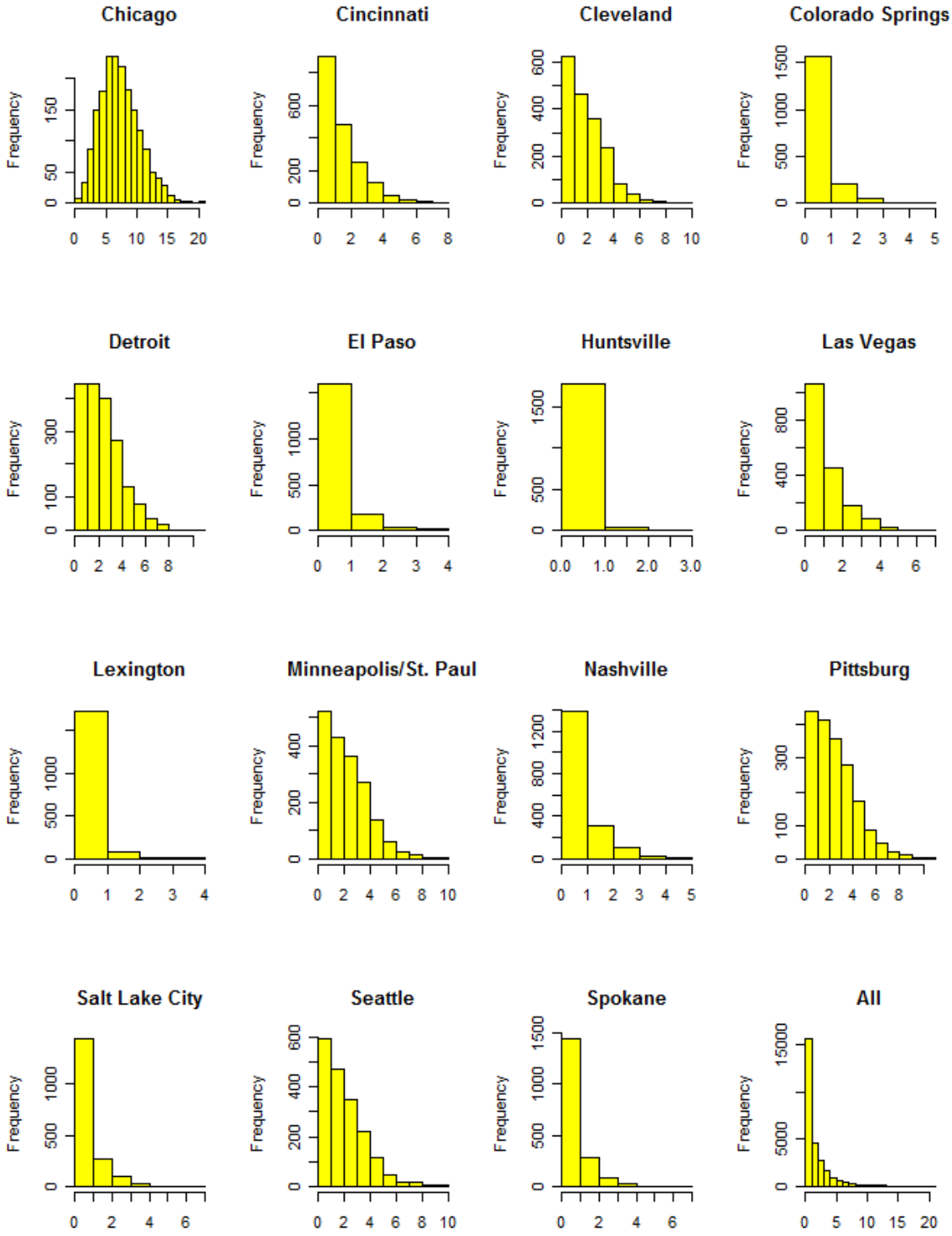
Goodness-of-fit test of Poisson distribution in 15 U.S. cities from 1991~1995.

City	χ^2	d.f.	p-value
Chicago	79.78	20	<.0001
Cincinnati	13.97	7	0.0518
Cleveland	21.18	9	0.0119
Colorado Springs	1.88	4	0.7571
Detroit	21.45	9	0.0108
El Paso	14.89	3	0.0019
Huntsville	0.86	2	0.6508
Las Vegas	0.65	6	0.1910
Lexington	3.61	3	0.3070
Minneapolis/St. Paul	11.39	9	0.2499
Nashville	3.09	4	0.5434
Pittsburg	28.64	10	0.0014
Salt Lake City	9.60	6	0.1427
Seattle	31.91	9	0.0002
Spokane	9.38	5	0.0947

cities also had small p-value to reject Poisson distribution (χ^2=16529.12, d.f.=20, p-value<0.0001). To check the basic properties of Poisson distribution, we found these respiratory death count data had overdispersion problem with mean 1.89 and variance 5.40. The main reason is that there were too many zeros in this mortality data. Among overall 27,390 valid respiratory disease death count data, there were 8,522 zeros, which account for 31.11% of the entire data (figure 3.46). There were several cities with over

Figure 3.46

The city-specific and overall histogram plots of respiratory death count from 1991~1995.



40% zero death count during 1,826 studies days. They are Colorado Springs(53.18%), El Paso(55.91%), Huntsville(79.13%), Lexington (67.96%), Salt Lake City(41.57%) and Spokane(43.10%). Actually, real-life count data are frequently characterized by excess zeros, and immediately encounter the trouble of overdispersion. Even though the zero-inflated count model can provide a more powerful approach to model this type of data and handle overdispersion problem well, so far the GGAMM based on restricted maximum likelihood or marginal likelihood estimation has not support ZIP (zero-inflated Poisson) and ZINB (zero-inflated negative binomial) data by BayesX yet. This situation is not actually unsolvable in the GGAMM, but there are some theoretical and technical difficulties in BayesX. More detailed discussion will be shown in chapter 5.

We also used the Q-Q plot to detect the normality assumption from model 1 to model 6 (figure 3.46). The distributed positions of normal theoretical quantiles and sample residual's quantiles displayed a weak-twisted straight line in each Q-Q plots, so they reflected our residual analyses were a little violating normality assumption. To check the boxplot of residuals shown in figure 3.47, the main reason of having a slight s-shape in Q-Q plots was that there are many outliers in two tails of their distribution, but non-outliers residuals generally presented a symmetric profile in each boxplot. However, their histogram plots compared with the corresponding normal curves displayed that each residual plot had a tiny right-skewness (figure 3.48), but this problem was not too severe because the linear model will not be affected if the residual's distributions are skewed when the number of valid data is not too small.

The residual plots with standardized residuals against predicted values for the purpose of identifying independence assumption are shown in figure 3.50. From model 1 to model 6, they shared two strange situations: 1) predictions were seldom located around 5, and made a gap splitting dots into two parts; 2) in each residual plot, the dots located below prediction equal to 5 displayed a fan shape, but the dots located above prediction

Figure 3.47

The Q-Q plots for residuals from model 1 to model 6.

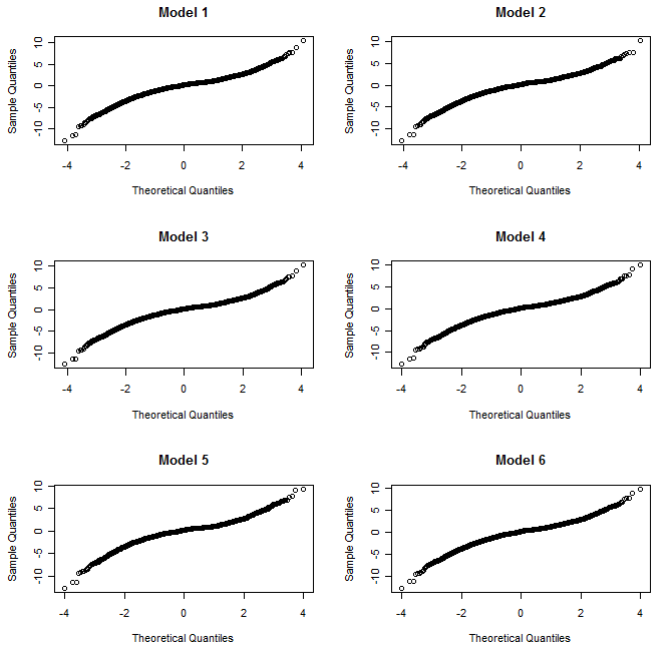


Figure 3.48

The boxplots of residuals in model 1 to model 6.

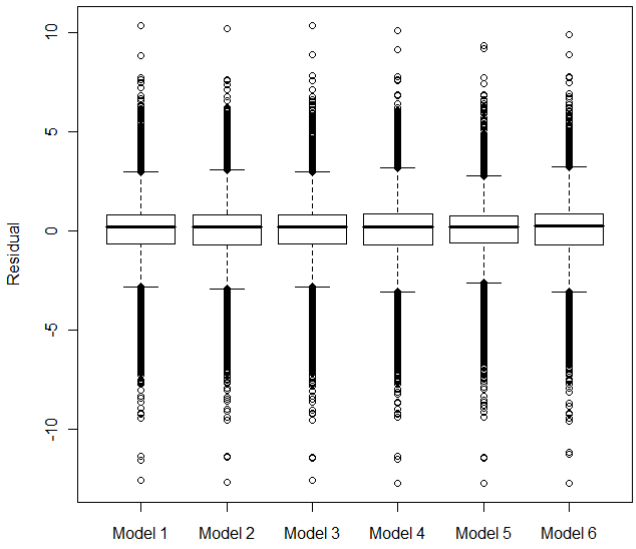


Figure 3.49

The histogram plots with normal curves in model 1 to model 6.

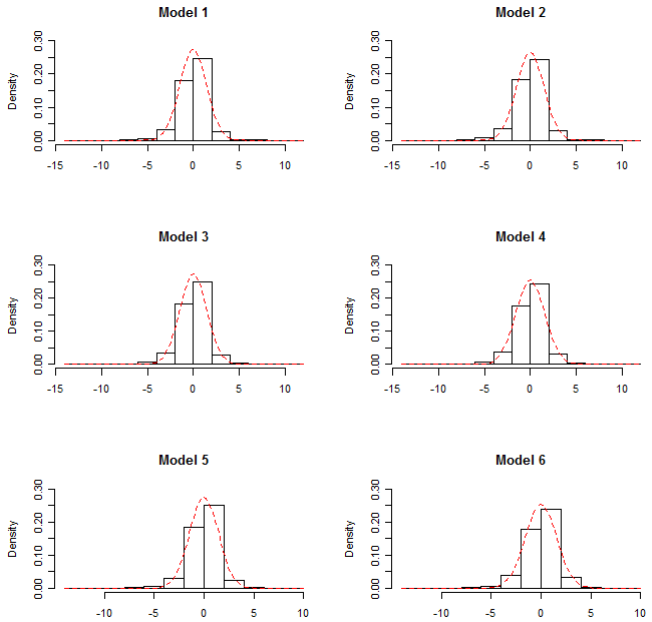


Figure 3.50

The residual plots from model 1 to model 6.

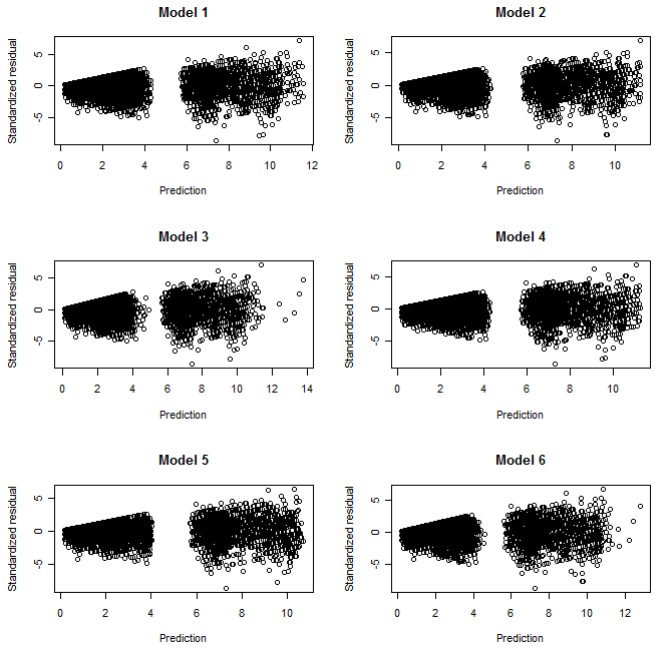
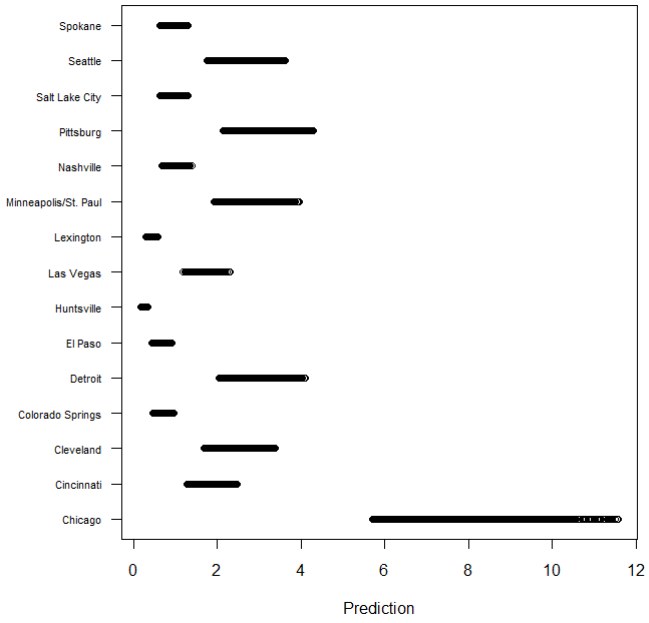


Figure 3.51

The city-specific prediction plot and observation plot in model 1.

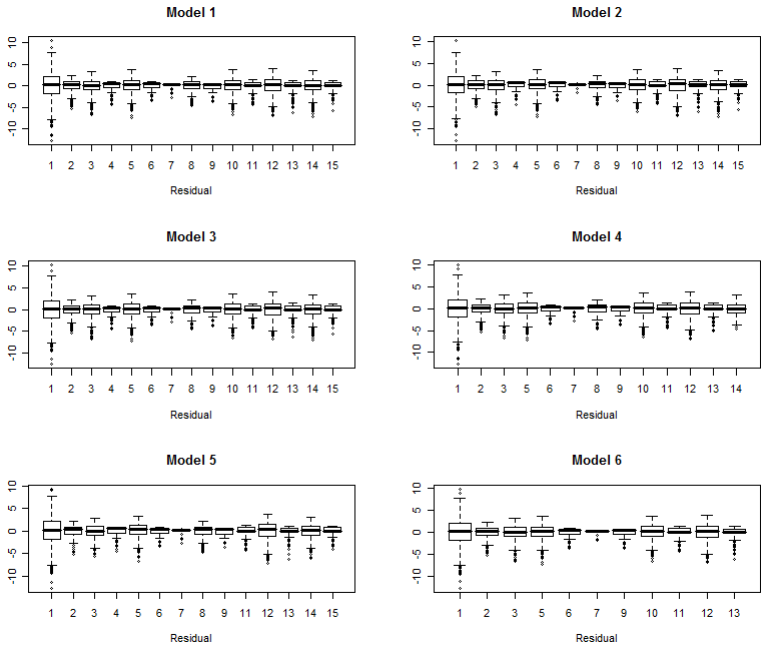


equal to 5 displayed a horizontal band. The first doubtful point can be interpreted by the city-specific prediction plot in figure 3.51. Here we only use predictions in model 1 for this explanation. We found the gap appearing around prediction=5 was because all predictions in Chicago were larger than 5.71, the maximum value of predictions in the other cities were 4.29 in model 1. That's why there is a blank area between 4.29 and 5.71 in figure 3.51. Meanwhile, the dots located in the left part in figure 3.50 can be confirmed that they were all from Chicago. Compared with real observations, the daily respiratory disease death counts less than 5 were all overestimated in Chicago, and that's why there was no prediction less than 5 in that city. As a result, this blank area was produced from the overestimated of predictions with observations less than 5 in Chicago.

About the fan shape in the right part of residual plots, it seemed the correlation among the Y values existed. Actually, the autocorrelation of respiratory disease death count in elders using in this study was only 0.2, and could be ignored. However, we recognize that BayesX does not offer functions of covariance structure to handle potential correlation

Figure 3.52

The level-1 unit (city) residual boxplots in model 1 to model 6.



†City number code: 1-Chicago, 2-Cincinnati, 3-Cleveland, 4-Colorado Springs, 5-Detroit, 6-El Paso, 7-Huntsville, 8-Las Vegas, 9-Lexington, 10-Minneapolis/St. Paul, 11-Nashville, 12-Pittsburg, 13-Salt Lake City, 14-Seattle, 15-Spokane

from observations when fitting the GGAMM with REML estimation. In fact, this function is only compiled in a fully Bayesian interpretation of structure additive regression model using Markov chain Monte Carlo simulation techniques, but our data encountered technical problems based on this model in BayesX. More details will be discussed in chapter 5. Besides, the variance homogeneity assumption may not be followed as well, at least the level-1 unit boxplot (figure 3.52) showed the variance of residuals in Chicago was significantly larger than the variance of residuals in the other cities.

Finally, the model evaluation methodology of the GGAMM is still underdeveloped, so the output from BayesX does not provide any information similar as R^2 or adjusted R^2 in linear regression models. However, we still succeeded in calculating the mean square error (MSE) to show each model's evaluation. The MSE in model 1 was 2.14, and increased to 2.25 as including 1-day and 2-day lag effects. In co-pollutant models, the MSE had a little bit decrease to 2.13 and 2.11 as long as including CO and O_3,

respectively; however, it increased to 2.47 and 2.48 in model 4 (PM_{10}+NO_2) and model 6 (PM_{10}+SO_2).

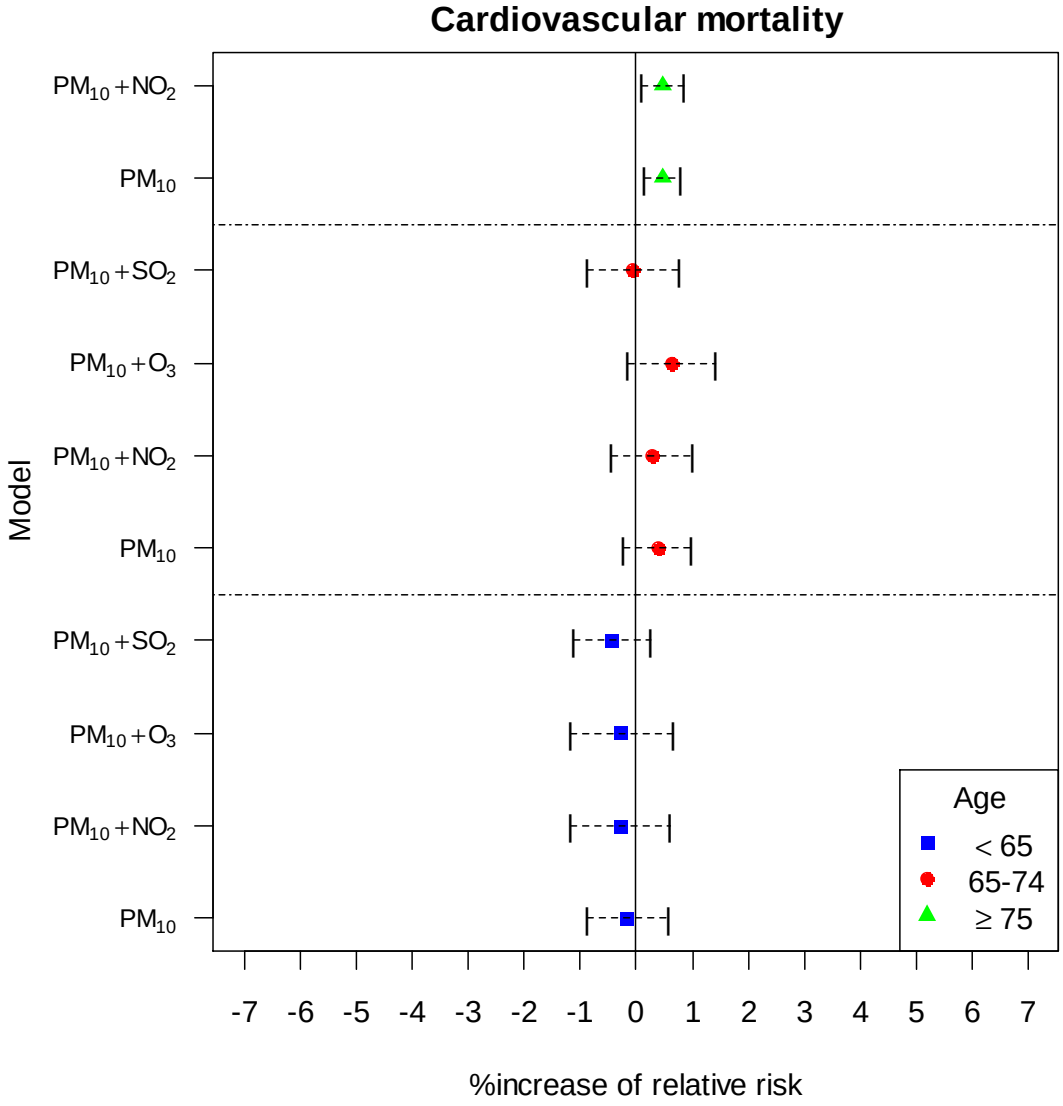
3.14. Extended applications

Based on the entire investigation in previous sections, it was desired to apply the GGAMM in more mortality data from the NMMAPS. Besides respiratory disease, this database also offers daily death counts in cardiovascular disease and pneumonia. These three mortalities were examined in three age categories (<65, 65−74, ≥75) by the GGAMM, and the results of the PM_{10} influence on the percent increase of mortality relative risk are visualized in figure 3.53~figure 3.55. Detailed outputs are contained in Appendix G and Appendix H. Note that all models were modified by appropriate starting values of smoothing parameters from 10 to 15, but some models were ignored without obtaining reasonable results while using up those λs.

The particulate matter has much different influence on cardiovascular diseases. Some negative but non-significant results derived in mortality<65 years old. The impact of PM_{10} turned to positive in the population with age 65−74 years old besides adjusting with SO_2. The reason was that SO_2 has much higher effect on mortality than PM_{10}, and each 10 ppb increase of SO_2 concentration rose 1.77% (95% CI: -0.51%, 4.11%) relative risk in cardiovascular mortality in this age category. Also, SO_2 had stronger consequence than the other co-pollutants. Adjusting with ozone in the same age category, PM_{10} produced the highest 0.63% increase of relative risk in cardiovascular disease, but was still not significant. The first significant result occurred in mortality≥75 years old. We found cardiovascular mortality increased 0.47% (95% CI: 0.14%, 0.81%) relative risk per 10 $\mu g/m^3$ increase in PM_{10}. This significant result also maintained significance and consistence when adjusting by NO_2.

Figure 3.53

The percent increase of relative risk with 95% confidence interval in cardiovascular mortality in three age categories.



The city-specific PM_{10} effect displayed different patterns in different age levels. Without co-pollutant's adjustment, Chicago, Cincinnati, Seattle and Spokane had stronger PM_{10} effect than the other cities to cardiovascular mortality<65 years old; however, when considering co-pollutants, no city showed positive PM_{10} influence. In age from 65−74 years old, most cities had positive PM_{10} influence besides Cincinnati, and this situation generally remained similar after adjusting by NO_2, O_3 and SO_2. In particular, when adjusting by O_3, the city-specific PM_{10} effect increased in Cleveland, Colorado Springs,

Detroit and Nashville where each 10 μg/m^3 PM_{10} can increase over 1% relative risk of cardiovascular mortality. Nonetheless, the location influence did not show high diversity in mortality for age≥75 years old, even though all city-specific effects were positive in both models (PM_{10}, PM_{10}+NO_2).

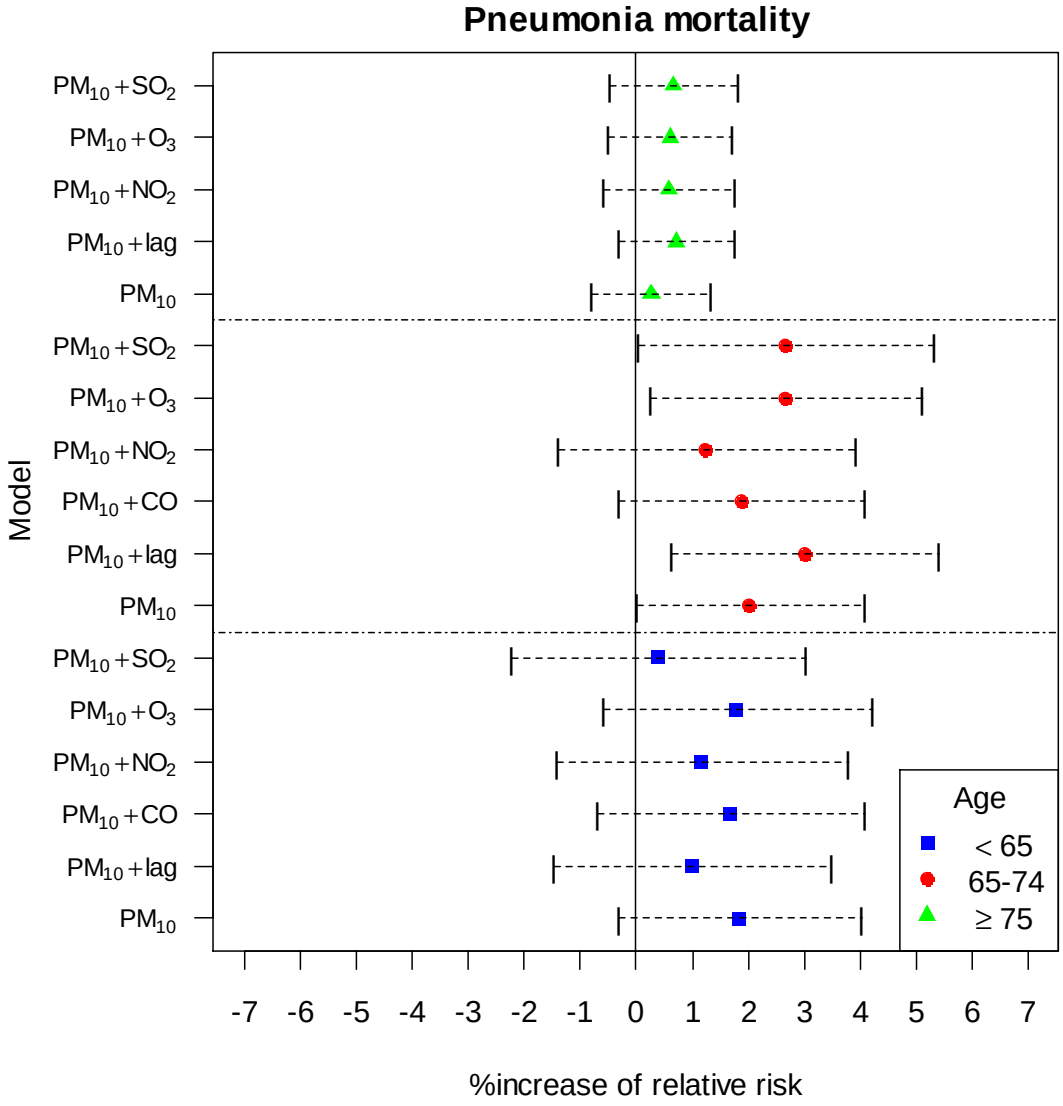
Compared with what we had done in previous sections, the time smoother from cardiovascular mortality was not as regular as respiratory mortality, but still had somewhat seasonal effect which caused negative influence from July to August and positive influence from September to next June in each year. All temperature smoother in cardiovascular disease had the highest impact in the coldest weather, and immediately decreased to the lowest point until 40°F~60°F with a slight rebound to 100°F. In geographical variation, Chicago, Detroit, Cleveland and Pittsburg had stronger influence than the other cities. Among three age levels, the highest spatial effect appeared in age≤65 years old in Chicago, which relative rate was 5.44, and was 15.48-fold of Lexington, the location with the lowest spatial effect.

To be one of the ten leading causes of death in the U.S., pneumonia presented some significant results in our findings of mortality with age 65−74 years old. Without adjusting with lag effects and co-pollutant, each 10 μg/m^3 increase in PM10 concentration can significantly increase 2.02% (95% CI: 0.01%, 4.08%) the relative risk of pneumonia mortality, which was higher than the same model in the other two age levels (<65 years old and ≥75 years old). It was enlarged to 3.00% (95% CI: 0.64%, 5.41%) after adjusting for a 1-day and a 2-day PM_{10} lag effect. Also, controlling by O_3 and SO_2, we got another two same significant results with 2.65%; nonetheless, the effect became non-significance and decreased to 1.87% and 1.23% when controlling by CO and NO_2.

The city-specific PM_{10} effect had enough variation and strong influence on pneumonia mortality in age≤65 years old and 65−74 years old. Detroit and Spokane had higher effect, almost reaching 3% increase of relative risk in pneumonia mortality in

Figure 3.54

The percent increase of relative risk with 95% confidence interval in pneumonia mortality in three age categories.

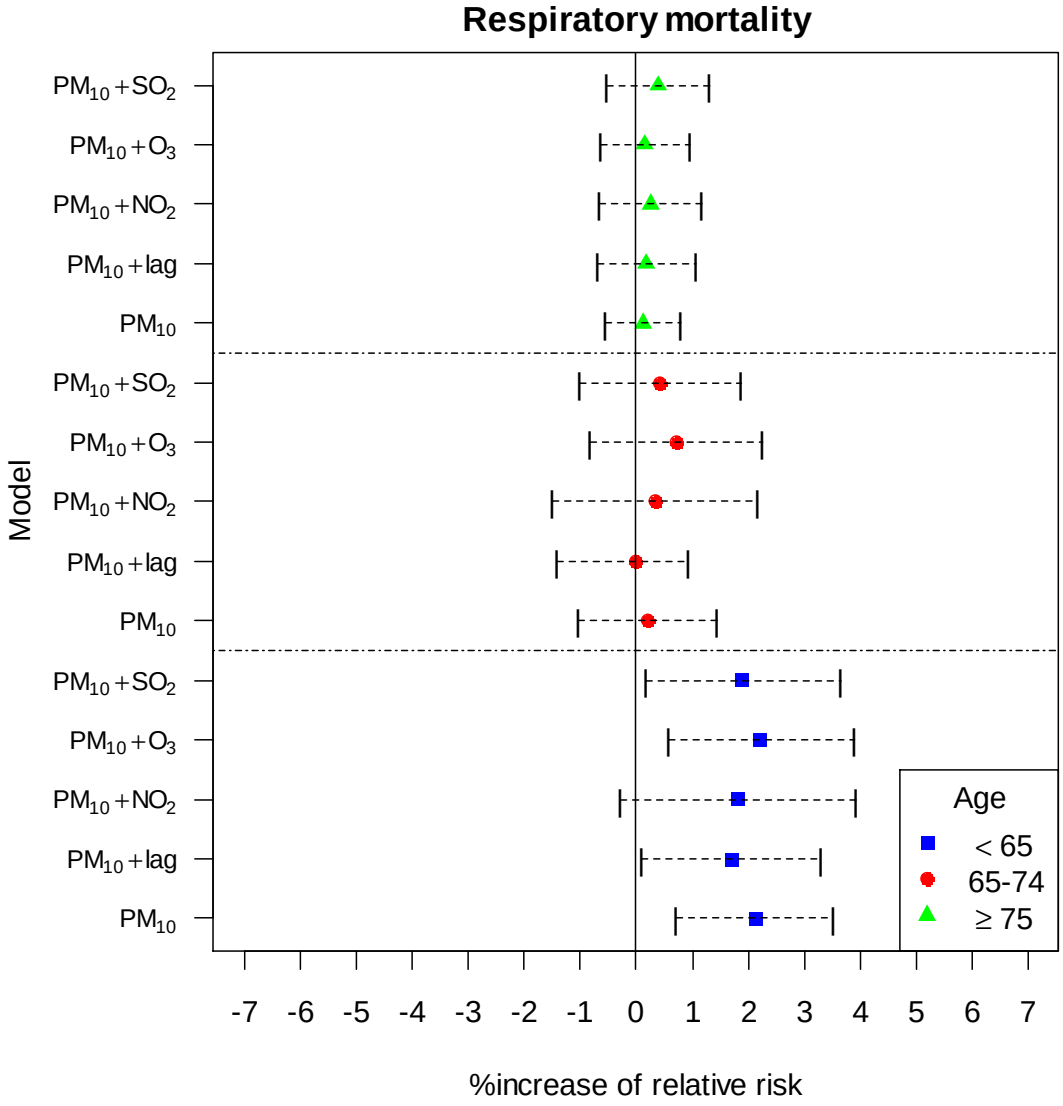


age<65 years old while not adjusting by lag effects and O_3, but no specifically strong PM_{10} city-specific influence occurred in some cities in the other two age levels. In particular, Colorado Springs, El Paso, Minneapolis/St. Paul, Salt Lake City and Spokane had negative city-specific effects in PM_{10} to pneumonia mortality≥75 years old without controlling by lag effects or co-pollutants, but after adjusting with lag effects and co-pollutants, all city-specific PM_{10} effects were modified to positive.

There are three characteristics in smoothing functions in pneumonia models: 1)

Figure 3.55

The percent increase of relative risk with 95% confidence interval in respiratory mortality in three age categories.



seasonal variation was not very significant with slight fluctuation in mortality<65 years old, and the longitudinal influence was decreasing over time; 2) temperature effect was not very nonlinear in all age levels; 3) pneumonia mortality≥75 years old did not have obvious decrease, but presented a slight horizontal line over temperature. The spatial function showed weak geographical variation in most of pneumonia models, but Chicago, Detroit, Cleveland, Pittsburgh, Minneapolis/St. Paul and Seattle showed relative stronger relative rate than the other cities.

Respiratory mortality was examined again with three age levels, and 4 of 5 results in age<65 years old also found significant findings. Each 10 μg/m^3 increase of PM_{10} concentration had at least 1.69% to 2.22% increase of relative risk in respiratory mortality<65 years old. Besides the model controlling for NO_2, the other respiratory models were all significant. A negative PM_{10} influence appeared in an extended distributed respiratory model with 65−74 years old, but corresponding 1-day and 2-day lag effects were all positive with value of 0.26% and 0.41%, respectively. However, both of them were still non-significant. Models with respiratory mortality≥75 years old had affected positively by PM_{10} concentration, but the magnitude had been smaller than the other two age levels.

The seasonal trend in respiratory models became more significant alone with the age level, and temperature smoother also displayed much curved in age≥75 years old than the other two age levels. The spatial effect pattern in respiratory mortality was not as consistent as cardiovascular and pneumonia mortality. In respiratory mortality<65 years old, only Chicago, Detroit and Las Vegas had positive relative rate. While considering mortality in 65−74 years old or ≥75 years old, more cities were included in the positive side, such as Seattle, Cleveland and Pittsburg, but the spatial influence was weaken in Las Vegas in mortality≥75 years old, surprisingly.

Similar as pneumonia mortality, larger national PM_{10} effect made higher city-specific effect in respiratory mortality<65 years old and 65−74 years old. In age<65 years old, most of city-specific PM_{10} effects were estimated between 1%~3% whether controlling by lag effects or co-pollutants. The city-specific PM_{10} effects were generally weakened to zero in respiratory mortality≥75 years old, but all of them were still positive in each model.

Chapter 4
SIMULATION

In order to make more objective conclusions from case studies, we made some simulations to investigate convergence problem and missing data analysis in the GGAMM. The convergence problem in the GGAMM has not been clearly discovered in previous studies using BayesX, and missing data analysis is an interesting issue, that is, how imputation methods work in the GGAMM. Besides, we were also interested in the influence of concurvity in the GGAMM, and found a connection between concurvity and convergence problem.

4.1. Concurvity simulation

As what we have mentioned in chapter 3, BayesX is not similar as other statistical computational tools, and always shows outputs whatever the entire iteration reaches convergence. Even though a warning message reminds users that the model fitting is doubtful without convergence, people would easily ignore this warning message unintentionally. A significant clue to identify the convergence problem in a model fitting is from those overestimated standard errors. For the only statistical software which can handle the GGAMM so far, the source of the convergence problem is still undiscovered, and we found some evidences.

One of possible sources is from concurvity problem. Table 4.1 shows 10 different concurvity levels with corresponding $\hat{\beta}$, se($\hat{\beta}$), se($\hat{b}$) and convergence rate in the GGAMM. The comprehensive analysis can investigate the influence of concurvity in the model fitting of the GGAMM and convergence rate simultaneously. When concurvity

Table 4.1

The main estimates with different concurvity levels and corresponding convergence rate (CR) in the GGAMM.

Concurvity	$\hat{\beta}$	se($\hat{\beta}$)	se($\hat{b}$)	ASMSE	CR(%)
0.03	0.0954	0.1323	0.3482	2.3582	99.8
0.10	0.0895	0.1311	0.3438	2.3603	99.2
0.19	0.0922	0.1274	0.3311	2.3511	99.1
0.31	0.0875	0.1197	0.3026	2.3538	98.5
0.41	0.0801	0.1111	0.2710	2.3521	97.2
0.50	0.0752	0.1032	0.2402	2.3387	96.1
0.59	0.0735	0.0944	0.2042	2.3431	95.3
0.71	0.0661	0.0841	0.1584	2.3356	93.4
0.80	0.0756	0.0755	0.1140	2.3101	90.5
0.90	0.0926	0.0683	0.0692	2.2390	87.9

* True (β, se(β), se(b))=(0.1, 0.15, 0.35)

level was increasing, the convergence rate was decreasing from 99.8% to 87.9%. For 1,000 replicated simulation data sets in each concurvity scenario, the data sets with convergence problem were only around 20 data sets as concurvity level = 0.03, but the number of data sets without convergence increased to around 120. Meanwhile, the increased rate climbed to 6 times as almost-non-concurvity became extreme concurvity in the GGAMM.

In addition, the influence of concurvity in the GGAMM had also been revealed by our simulations. We found that $\hat{\beta}$, se($\hat{\beta}$) and se($\hat{b}$) could be underestimated along with the raise of concurvity level in the GGAMM, and higher concurvity level generally caused more biases than lower concurvity level. Note that, as concurvity level reached 0.8, the decreasing-tendency of $\hat{\beta}$ was curbed, and the average of $\hat{\beta}$ bounced back to its true value 0.1 until concurvity level was 0.9. However, the averages of se($\hat{\beta}$) and se($\hat{b}$) from simulation data with concurvity level=0.9 were still highly underestimated, and compared with the results of concurvity level=0.03, the two estimates slumped around 48.37% and 80.13%, respectively.

The influence of concurvity on smoothing functions was also investigated here. We applied ASMSE from equation (2.71) to quantify smoothing functions, and compared estimated smoothing functions with real smoothing functions. The research question came from a hypothesis. In chapter 2, we had described the methodology of estimating smoothing functions in the GGAMM. Generally speaking, it's very similar with other semiparametric models, such as the GAM or GAMM, but their concept is decomposing smoothing functions with many pieces by knots. Those pieces can be reconstructed with a matrix form and combined with linear factors in a unified design matrix. Meanwhile, the piece parts of smoothing functions and linear factors were estimated simultaneously by the estimating equation. If the concurvity level would be a reason of making biases in linear estimated parameters, its influence on the estimation of smoothing functions should also be inevitable.

In table 4.1, we show the ASMSE of smoothing function in each concurvity level. The trend of ASMSE did not climb along with the increase of concurvity level, which means that concurvity problem only affected the parameter's estimation of linear factors, but there is no much influence of making smoothing functions more biased. Reviewing the concurvity level appearing in case studies in chapter 3, each model had concurvity level with values of around 0.5. If table 4.1 can reflect the true reality, we can conclude that all estimates shown in table 3.4 could be probably underestimated, and it may explain why air pollutant influence to adverse human health in our analyses was less than previous studies.

4.2. Multicollinearity simulation

The biases produced from multicollinearity in the GGAMM mainly appeared in fixed effects and their standard errors. Table 4.2 shows the simulation results containing

Table 4.2

The main estimates with different multicollinearity levels and corresponding convergence rate (CR) in the GGAMM.

Multicollinearity	$\hat{\beta}_1$	se($\hat{\beta}_1$)	se($\hat{b}_1$)	$\hat{\beta}_2$	se($\hat{\beta}_2$)	se($\hat{b}_2$)	CR(%)
0.0	0.0016	0.0162	0.0194	0.0214	0.0279	0.0477	100.0
0.1	-0.0002	0.0167	0.0202	0.0216	0.0272	0.0443	100.0
0.2	-0.0035	0.0173	0.0201	0.0245	0.0264	0.0384	100.0
0.3	-0.0061	0.0183	0.0201	0.0244	0.0262	0.0350	99.9
0.4	-0.0088	0.0197	0.0207	0.0247	0.0266	0.0315	99.9
0.5	-0.0110	0.0217	0.0214	0.0266	0.0276	0.0299	100.0
0.6	-0.0157	0.0245	0.0226	0.0313	0.0293	0.0278	100.0
0.7	-0.0216	0.0284	0.0223	0.0345	0.0322	0.0257	100.0
0.8	-0.0304	0.0354	0.0241	0.0408	0.0382	0.0271	100.0
0.9	-0.0464	0.0504	0.0250	0.0558	0.0520	0.0277	100.0

two correlated variables in the GGAMM. As the multicollinearity level increased from 0 to 0.9, two estimated fixed effects $\hat{\beta}_1$ and $\hat{\beta}_2$ appeared biases toward different directions gradually, where $\hat{\beta}_1$ decreased to negative value, and $\hat{\beta}_2$ increased to 0.0558, which was over 2-fold of 0.0214 from multicollinearity level=0. Both of the estimated standard errors se($\hat{\beta}_1$) and se($\hat{\beta}_2$) did not appeared too significant trends as multicollinearity level $<$ 0.5, and started to have a steep climb-up until multicollinearity level=0.9. In random effects, se($\hat{b}_1$) seemed not be affected too much, but se($\hat{b}_2$) decreased almost a half from 0.0477 to 0.0271.

In this simulation, we actually did not generate data for smoothing functions but used original calendar time and 24-hour average temperature data, so there was no true smoothing function used for calculating the ASMSE. We believed, from the definition of multicollinearity, the smoothing functions, even spatial function, should not be affected, unless they originally contain severe concurvity in the GGAMM. However, this concurvity-confounding situation did not exist because the almost perfect convergence rate in each scenario reflected that this simulation was not disturbed by concurvity.

4.3. Missing data imputation simulation

In missing data analysis, we processed two different analyses to see how missing data affects modeling as long as they exist in either linear factors or smoothing functions. Define two factors X_1 and X_2 as used in a simulated GGAMM, where X_1 was used in linear factor, and X_2 was fitted as a penalized spline. The missing rate is 5%, 10%, 20%, 30%, 40% and 50%.

Table 4.3 displays the simulation results considering missing data only appears in linear factor X_1. Based on MCAR and compared with true β=0.1 , CCA can maintain their consistency and did not make too much bias, and NNI2 also had better performance, even though either one became underestimated when missing rate reached 50%. NNI1 only had better estimates on $\hat{\beta}$ as missing rate was 5% and 10%, and it tended to underestimate while missing rate was over 10%. This result made sense because nonparametric missing data imputation method could only be helpful if missing rate is controlled below 10%. In addition, NNI1 made the range of simulated results of $\hat{\beta}$ getting wider than NNI2 when missing rate of X_1 was increasing, which implied that nonparametric missing data imputation method applying in higher missing rate could possibly cause targeted estimated parameters more unstable (result is not shown here). A surprising result was that MI-MCMC did not work very well in the GGAMM, and $\hat{\beta}$ started to estimate smaller than its true value as a few missing data appeared, even though its stability from 1,000 simulations was always more concentrated than CCA, NNI1 and NNI2, and also did not change too much as missing rate was raised.

The influence of missing data imputation in smoothing functions is also presented in table 4.3. It is obvious that, without using imputation method, the precision of estimated smoothing function was not changing too much along with the increase of missing rate in X_1. The ASMSE began with value of 1.2493 when missing rate of X_1 was 5%, and had a

Table 4.3

The estimates $\hat{\beta}$, se($\hat{\beta}$), and se($\hat{b}$) of GGAMM with corresponding convergence rates (CR) when applying CCA, NNI1, NNI2 and MI in variable X_1 for missing rates 5%, 10%, 20%, 30%, 40% and 50%.

		Missing rate (%)					
Method	Parameter	5	10	20	30	40	50
CCA	$\hat{\beta}$	0.1069	0.1060	0.1034	0.1024	0.1025	0.0941
	se($\hat{\beta}$)	0.3025	0.3035	0.3052	0.3074	0.3109	0.3172
	se($\hat{b}$)	0.9145	0.9153	0.9150	0.9147	0.9148	0.9203
	ASMSE	1.2493	1.2399	1.0690	1.1027	1.0135	1.2410
	CR(%)	99.2	99.6	99.4	99.7	99.9	99.8
NNI1	$\hat{\beta}$	0.1052	0.1014	0.0907	0.0787	0.0687	0.0475
	se($\hat{\beta}$)	0.3022	0.3015	0.2994	0.2961	0.2940	0.2869
	se($\hat{b}$)	0.9145	0.9098	0.8981	0.8807	0.8663	0.8325
	ASMSE	1.3371	1.4733	1.3322	2.0638	2.2029	2.1213
	CR(%)	99.6	98.9	99.0	99.1	99.4	99.7
NNI2	$\hat{\beta}$	0.1102	0.1129	0.1117	0.1109	0.1124	0.0956
	se($\hat{\beta}$)	0.2971	0.2921	0.2812	0.2665	0.2519	0.2338
	se($\hat{b}$)	0.8982	0.8801	0.8404	0.7877	0.7332	0.6664
	ASMSE	1.3358	1.4711	2.0661	2.2034	2.1191	1.9457
	CR(%)	99.5	99.7	99.1	99.4	98.6	99.3
MI-MCMC	$\hat{\beta}$	0.0996	0.0939	0.0746	0.0452	0.0233	0.0005
	se($\hat{\beta}$)	0.2886	0.2722	0.2546	0.2218	0.1986	0.1744
	se($\hat{b}$)	0.8716	0.8174	0.7580	0.6485	0.5699	0.4850
	ASMSE	1.3342	1.4717	2.0649	2.2040	2.1220	1.9439
	CR(%)	97.3	97.4	97.4	98.4	97.9	98.3

* True (β, se(β), se(b))=(0.1, 0.3, 0.7)

tiny decrease when missing rate of X_1 was 30% and 40%. But, finally it came back to 1.2410 when missing rate of X_1 was 50%. It reflected that missing data appearing in linear factor X_1 did not cause damages on smoothing functions' estimation. When missing X_1 data were imputed by NNI1, NNI2 and MI-MCMC, the values of ASMSE were all higher than those in CCA. Three missing data imputation methods generally did not have marked difference in any missing rate of X_1. When missing rate of X_1 was 5% and 10%,

their ASMSE values increased to 1.33~1.47, and raised over 2 hugely along with the increase of missing rate of X_1. To sum up, the efficacy of these missing data imputation methods did not have significant improvement in smoothing functions when missing data only appeared in the linear factor.

We were also interested in the convergence rate among different missing rates in X_1 and missing data imputation. Table 4.3 showed the convergence rate did not have significant change as long as missing rate was increasing. Nonetheless, missing data imputation method also did not guarantee any improvement on convergence rate. Whatever any missing rate is, all missing data imputation methods made tiny decrease on convergence rate, especially in MI-MCMC, which decreased convergence rate as low as 98%, and even lower. A reasonable explanation is that the number of imputation easily caused higher probability of convergence problem. For example, we used 5 imputations in each simulation data, and as long as one of them had convergence problem, the average result from 5 imputations would become biased. Even though this is not totally unsolvable because chapter 3 had shown three methods to adjust it, the process will be still time-consuming for modification in either the starting value of smoothing parameter or the number of knots.

On the other hand, when missing data only appeared in the variable used in smoothing functions, applying missing data imputation method would make serious effects on parameter estimation in linear factors. In table 4.3, compared with true value of β, se(β) and se(b), CCA produced close results in $\hat{\beta}$ and se($\hat{\beta}$), where $\hat{\beta}$ was from 0.1081 to 0.0979 and se($\hat{\beta}$) was from 0.3032 to 0.3160 while missing data rate is from 5% to 50%. The se($\hat{b}$) was a little overestimated, and located around 0.91 under any missing data rate in X_2, but its ASMSE confirmed that higher missing data rate did not make more biases on the estimation of smoothing function. Generally speaking, the result of ASMSE

Table 4.4

The main estimates of GGAMM with corresponding convergence rates (CR) when applying CCA, NNI1, NNI2 and MI in variable X_2 for missing rates 5%, 10%, 20%, 30%, 40% and 50%.

		Missing rate (%)					
Method	Parameter	5	10	20	30	40	50
CCA	$\hat{\beta}$	0.1081	0.1049	0.1044	0.1031	0.1019	0.0979
	se($\hat{\beta}$)	0.3032	0.3040	0.3061	0.3085	0.3113	0.3160
	se($\hat{b}$)	0.9173	0.9172	0.9181	0.9182	0.9165	0.9167
	ASMSE	1.2468	1.2424	1.0637	1.1049	1.0176	1.2398
	CR(%)	99.4	99.4	99.9	99.5	99.5	99.9
NNI1	$\hat{\beta}$	0.1042	0.0987	0.0809	0.0595	0.0381	0.0134
	se($\hat{\beta}$)	0.2954	0.2885	0.2724	0.2563	0.2405	0.2251
	se($\hat{b}$)	0.8941	0.8714	0.8180	0.7645	0.7113	0.6594
	ASMSE	1.3429	1.4676	2.0625	2.2026	2.1188	1.9330
	CR(%)	99.1	99.3	99.2	99.6	99.5	99.6
NNI2	$\hat{\beta}$	0.0883	0.0676	0.0454	0.0132	-0.0219	-0.0386
	se($\hat{\beta}$)	0.2961	0.2898	0.2775	0.2678	0.2598	0.2537
	se($\hat{b}$)	0.8959	0.8753	0.8345	0.8023	0.7755	0.7556
	ASMSE	1.3395	1.4807	2.0568	2.2025	2.1174	1.9307
	CR(%)	99.4	98.5	99.5	99.6	99.3	99.6
MI-MCMC	$\hat{\beta}$	0.0791	0.0412	-0.0086	-0.0495	-0.0841	-0.1076
	se($\hat{\beta}$)	0.3003	0.2971	0.2957	0.2914	0.2861	0.2905
	se($\hat{b}$)	0.9098	0.8995	0.8951	0.8813	0.8641	0.8788
	ASMSE	1.3462	1.4712	2.0647	2.2031	2.1118	1.9479
	CR(%)	96.3	97.4	98.0	99.1	98.9	99.7

* True (β, se(β), se(b))=(0.1, 0.3, 0.7)

in table 4.3 was not better than the result of ASMSE n table 4.2. Even though most values of ASMSE in table 4.3 were smaller than those in table 4.2, the difference was very tiny and can be ignored. This result in table 4.3 reflected that, when missing data appeared in valuables fitting with smoothing functions severely, traditional missing data imputation may not efficiently improve the entire model-fitting, whatever in the estimations of unknown parameters in linear factor or smoothing functions. If the amount of data is large

enough (in this simulation, the sample size of each data set is 1,000), it can stand more missing rate up to 50%. More discussion will be mentioned in chapter 5.

Finally, compared with the results of MI-MCMC in table 4.3 and table 4.4, models with missing data in X_2 had better convergence rate than models with missing data in X_1 in high missing rates. Meanwhile, the impact of convergence problem was relative weak when missing data were not in linear factors.

Chapter 5
DISCUSSION

5.1. Summary

Our main goal was to accomplish comprehensive progress using the GGAMM in air pollution and adverse human health research, and evaluate the reasonability of findings to enhance confidence in this statistical model. Because elders ≥ 65 years old dying from respiratory diseases were our population in the initial case study, it was smaller than other studies containing death from all ages, and thus lead to a lower estimated PM_{10}. Despite the lack of a previous study using the respiratory mortality data from elders ≥ 65 years old, the application results from chapter 3.12 still had reasonable conclusions when referring to published literature. From previous U.S. studies, the national relative risk ranged from 0.09%~0.68% in a 10 $\mu g/m^3$ increase of PM_{10} over 4 to 100 cities (table 5.1). When restricted to cardiovascular and respiratory mortality, the range became 0.13%~0.68%, which covered results from all models with cardiovascular and respiratory mortality in ages 65−74 and ≥ 75 years old, besides PM_{10}+$SO_{2,}$ in with cardiovascular mortality and PM_{10}+O_3 in respiratory mortality in 65−74 years old. Due to technical limitations discussed in chapter 5.2, our extended analysis did not adjust by age group, but was separated into three models in each mortality. Therefore, no age-specific results from previous studies can be compared. In addition, multi-city air pollution study has not investigated pneumonia mortality yet, so our findings provided a possibility to make it examinable.

Table 5.1

Comparison of results across studies: estimated percent increase in mortality relative risk per 10 μg/m^3 increase in PM_{10}.

# of cities	Period	Mortality	Air pollutant	%increase in mortality RR per 10 μg/m^3 increase in PM_{10} (95% PI)	Reference
4 cities	1987-1994	Total	PM_{10}	0.17 (-0.01, 0.34)	Dominici et al. (2003a)
		CVD/RESP	PM_{10}	0.22 (-0.02, 0.46)	
20 cities	1987-1994	Total	PM_{10}	0.48 (0.05, 0.92)	Dominici et al. (2000a)
100 cities	1987-2000	Total	PM_{10}	0.09 (-0.01, 0.19)	Peng et al. (2005)
			PM_{10}-lag1	0.19 (0.10, 0.28)	
			PM_{10}-lag2	0.08 (-0.03, 0.19)	
88 cities	1987-1994	Total	PM_{10}	0.55 (0.10, 0.98)	Dominici et al. (2002a)
90 cities	Unknown	CVD/RESP	PM_{10}	0.13 (-0.05, 0.31)	Dominici et al. (2005)
			PM_{10}-lag1	0.31 (0.13, 0.49)	
			PM_{10}-lag2	0.20 (0.02, 0.38)	
20 cities	1987-1994	Total	PM_{10}	0.51 (0.07, 0.93)	Samet et al. (2000a)
		CVD/RESP	PM_{10}	0.68 (0.20, 1.16)	
88 cities	1987-1994	Total	PM_{10}	0.22 (0.10, 0.38)	Dominici et al. (2003b)
		CVD/RESP	PM_{10}	0.31 (0.15, 0.50)	
100 cities	1987-2000	Total	PM_{10}	0.22 (0.13, 0.31)	Welty and Zeger (2005)

Table 5.2

Comparison of results across studies: estimated percent increase in mortality relative risk per 10 ppb increase in O_3.

# of cities	Period	Mortality	Air pollutant	%increase in mortality RR per 10 μg/m^3 increase in O_3 (95% PI)	Reference
95 cities	1987-2000	Total	Ozone	0.28 (0.13, 0.48)	Bell et al. (2005)
95 cities	1987-2000	Total	Ozone	0.52 (0.27, 0.77)	Bell et al. (2004a)
		CVD/RESP	Ozone	0.64 (0.31, 0.98)	
98 cities	1987-2000	Total	Ozone-lag1	0.21 (-0.06, 0.47)	Bell et al. (2007)
98 cities	1987-2000	Total	Ozone-lag1	0.32 (0.17, 0.46)	Bell et al. (2006)
19 cities	1987-1994	CVD/RESP	Ozone	0.73 (0.27, 1.19)	Huang et al. (2005)
			Ozone-lag1	0.70 (0.26, 1.12)	
			Ozone-lag2	0.64 (0.17, 1.07)	

The GAM was the main statistical model for fitting city-specific effect in the 2-stage Bayesian hierarchical model. From valid numeric record, the city-specific PM_{10} effect was generally from 0.07%~0.35% (Dominici et al., 2002a; Dominici et al., 2003a; Dominici et al., 2003b). Our estimates shown in table 3.7 were from 0.06%~0.15%. In our extended analysis shown in section 3.12, the city-specific estimates from the pure PM_{10} model without controlling lag effects and co-pollutants were more diverse than the other models because they had a stronger PM_{10} effect at a national level and estimated standard errors of random effect. Moreover, the distributed lag model had a negative effect on 1-day lag PM_{10} effect, which was not shown in previous studies. Four studies (Peng et al., 2005; Roberts & Martin, 2007; Welty & Zeger, 2005; Bell et al., 2004a) showed that their PM_{10} effects at a 1-day lag and national level had 0.19%, 0.23%, 0.55% and 0.50%, respectively. Note that the effects in Welty and Zeger (2005) and Dominici et al. (2002a) were estimated from single pollutant models, which means there was no current PM_{10} adjusting in their models simultaneously. Nonetheless, when extending the lag effects to 6 days and 14 days in our models, and adjusting by the principal component analysis, our results derived positive association to mortality rate with values of 0.17% and 0.01%, respectively. Our analysis also found patterns from PCA-adjusted estimates in PM_{10} with longer lag effects that previous multi-city air pollution studies never discovered.

Compared with CO, NO_2 and SO_2, previous multi-city time series studies examined O_3 and human health specifically (table 5.2). Bell et al. (2004a) showed a 0.64% increase in cardiovascular and respiratory mortality per 10-ppb increase in the previous week's ozone by fitting 2-stage Bayesian hierarchical models with NMMAPS data, and got a consistent result with adjustment for PM_{10}. Compared with results shown in section 3.9, a 10-ppb increase in O_3 at national level was significantly associated with 1.79% (95% CI: 0.13%, 3.47%) increase in respiratory mortality with adjustment by PM_{10}, which was

close to the O_3 effect estimated from the 2-stage Bayesian hierarchical model (1.41%, 95% CI: -1.42%, 4.33%). Ozone and human health research was implemented by many meta-analysis studies, and Thurston and Ito's (2001) conclusion was closest to ours with a value of 1.37% (95% CI: 0.78%, 1.96%). Different from the other ozone meta-analysis studies (Stieb, Judek, & Burnett, 2002; Stieb, Judek, & Burnett, 2003; Levy, Carrothers, Tuomisto, Hammitt, & Evans, 2001; Anderson, Atkinson, Paecock, Marston, & Konstantinou, 2004; Bell et al., 2005), Thurstons and Ito's study (2001) was the only one which included a nonlinear relationship between temperature and mortality. This may explain why our results were more consistent with Thurston and Ito's (2001) than the others. Our result showed the PM_{10} effect in nation-level increased from 0.11% to 0.23% as adjusting with O_3, which was consistent with Dominici et al. (2000a). However, the increase amount is only 0.13%. Moreover, different ozone concentrations were also used. In Bell et al. (2004a), daily average, 8-hour maximum, and daily hourly maximum ozone concentration were used in a distributed lag model and the increase in mortality was 0.52% (95% CI: 0.27%, 0.77%), 0.64% (95% CI: 0.41%, 0.86%) and 0.67%(95% CI: 0.42%, 0.92%), respectively. Actually, the NMMAPS database provided detailed measurements for each air pollutant, such as hourly maximum, 2nd~5th hourly maximum, trimmed mean, daily median of 1-year trends, daily mean of 1-year trends, and lag 1~lag 3 trimmed mean. Our study was not as detailed, but is expected to have overwhelming detections.

Seldom had multiple location time series studies specifically focused on CO, SO_2, and NO_2. In most air pollution research, the two co-pollutants often used for identifying sensitivity analysis to adjust main factor are PM_{10} or O_3. For example, Peng et al. (2005) examined the influence of PM_{10} after adjusting with SO_2 in spring, summer, fall, and winter. The range of per 10 $\mu g/m^3$ increase on the national average estimates of season-specific 1-day lag PM_{10} concentration was 0.08%~0.33%. Although our results in

both the initial case study and extended analysis were all higher than 0.33%, we proved that PM_{10} effects were close between GGAMMs and 2-stage Bayesian hierarchical models in section 3.9. In addition, the national level effect of PM_{10} increased 3-fold when adjusting SO_2, but the SO_2 fixed effect was negative, which was contradicted from the result of the 2-stage Bayesian hierarchical model. In fact, a difficulty of investigating the health influence from SO_2 is that the gas is associated with particulate matter as sulfate dioxide is the precursor for fine sulfate particles, making their effects to adverse human health not easy to distinguish. Many cities made efforts to reduce the ambient concentration of SO_2, and it was believed the level was still high in larger cities (WHO, 2005). A Netherlands study concluded that SO_2 did not seem to be a causative factor for particulate matter associated health effects (Buringh, Fischer, & Hoek, 2000), but some studies still proposed the concentration of SO_2 was associated with total mortality for relative risk larger than 1% (Katsouyanni et al., 1997; Sunyer, Castellsague, Saez, Tobias, & Anto, 1996; Touloumi, Samoli, & Katsouyanni, 1996; Dab, Medina, Quenel, Le Moullec, Le Tertre, Thelot, Monteil, Lameloise, Pirard, Momas, Ferry, & Festy, 1996; Zmirou, Barumandzadeh, Balducci, Ritter, Laham, & Ghilardi, 1996). These studies did not consider spatial variation on the heterogeneity of multiple locations; as a result, the conclusion for SO_2 has not yet been clearly determined.

The estimated NO_2 effect at the national level was 1.26% (95% CI: -0.44%, 3.00%) associated with respiratory mortality rate in elders. The estimated PM_{10} effect was also slightly raised from 0.11% to 0.14%. It is similar to Peng et al. (2005) in winter and spring season. Some cross-sectional and cohort studies proved an NO_2 effect in lung function growth of children (Gauderman, McConnell, Gilliland, London, Thomas, vol, Vora, Berhane, Rappaport, Lurmann, Margolis, & Peters, 2000; Gauderman, Gilliland, Vora, Avol, Stram, McConnell, Thomas, Lurmann, Margolis, Rappaport, Berhane, & Peters, 2002), acute bronchitis (Ackermann-Liebrich, Leuenberger, Schwartz, Schindler,

Monn, Bolognini, Bongard, Brandli, Domenighetti, Elsasser, Grize, Karrer, Keller, Keller-Wossidlo, Kunzli, Martin, Medici, Perruchoud, Schoni, Tschopp, Villiger, Wuthrich, Zellweger, & Zemp, 1997), chronic respiratory symptoms (Braun-Fahrländer, Vuille, Sennhauser, Neu, Kunzle, Grize, Gassner, Minder, Schindler, Varonier, & Wuthrich, 1997; Shima & Adachi, 2000), and cough and phlegm symptoms in adults (Forsberg, tjernberg, & Wall, 1997; Zemp, Elsasser, Schindler, Kunzli, Perruchoud, Domenighetti, Medici, Ackermann-Liebrich, Leuenberger, Monn, Bolognini, Bongard, Brandli, Karrer, Keller, Schoni, Tschopp, Villiger, Zellweger, & Team, 1999). These studies did not consider spatial factor seriously, and most of them were proposed before the development of the 2-stage Bayesian hierarchical model. A meta-analysis collected 109 studies published from 1982 to 2000 and concluded that over a 24-hours average NO_2 concentration, the overall effect for all-cause mortality was 2.8%±0.3% (mean±SE) and 0.9%±0.5%(mean±SE) per 44 $\mu g/m^3$ in the single-pollutant model and multi-pollutant model, respectively (Stieb et al., 2002). Limited results about the influence of NO_2 to adverse human health were in previous multiple location time series studies. Samet et al. (2000a) fitted NMMAPS data for NO_2 from 19 cities in the U.S. found little evidence on such an association after adjusting PM_{10} and ozone level. The heterogeneity among cities was also discovered by a European study on short-term exposure to air pollution and mortality and morbidity, APHEA project, and investigated data from 29 cities (Katsouyanni, Touloumi, Samoli, Gryparis, Le Tertre, Monopolis, Rossi, Zmirou, Ballester, Boumghar, Anderson, Wojtyniak, Paldy, Braunstein, Pekkanen, Schindler, & Schwartz, 2001).

The CO estimates were fitted questionably in our analyses, and no improvement occurred after adjusting the starting value of smoothing parameters and the number of knots. Our proposed jackknife-bootstrap approach could force to curb the standard errors of fixed and random effects, but risked the variation of the subject-specific effect being

diminished. In addition, the too small adjusted se($\hat{\beta}$) would increase type I error. The adjusted PM_{10} effect at the national level enlarged after adjusting CO, but the CO effect was too weak to identify its influence. This situation was also happened in the 2-stage Bayesian hierarchical model, but it doubled PM_{10} effect. However, previous studies showed the carbon monoxide had a short-term effect on mortality. Another APHEA project concluded that a 1-$\mu g/m^3$ increase in CO was associated with a 1.20% and 1.25% increase in total deaths and cardiovascular mortality rate, respectively (Samoli, Touloumi, Schwartz, Anderson, Schindler, & Forsberg, 2007). Note that their CO level was the average of current and 1-day lag effect, and the spatial heterogeneity was derived by assigning all city-specific estimates into a random-effects regression model with a variance component estimated by iteratively reweighted least squares, which concept was similar to Dominici et al. (2003a).

The starting value of smoothing parameter and spatial function, as well as the number of knots, displayed new roles in this study that had not been discussed by previous literature. The main function of setting starting a smoothing function starting value was initially to facilitate the speed of estimation. However, we found that at times it can solve the overestimated problem in standard errors of parameters, irrational splines, and diminished geographical variation. The number of knots in smoothing functions theoretically does not have much influence on estimations, but we found it does have some impact, especially being a surrogate if the modification from the starting value of smoothing parameter does not work. Our results concluded that, if the model-fitting can reach convergence successfully, the main estimates $\hat{\beta}$, se($\hat{\beta}$) and se($\hat{b}$) will be robust, but the robust level in se($\hat{\beta}$) and se($\hat{b}$) is smaller than $\hat{\beta}$.

Comparing time smoothers shown in figures 3.5 to 3.10, there was generally no difference, but temperature smoothers reflected some concerns. According to the thermal comfort index (Fanger, 1970) in table 5.3, the trend of mortality average was confirmed to

Table 5.3

The number of days and elders' respiratory mortality average from 1991 to 1995 in 15 U.S. cities on thermal comfort index.

ET and ET_w indices(℃)	ET and ET_w indices(°F)	Thermal sensitivity	Physiological stress	# of days	Mean of death count
<13	<55.4	Very cold	Cold stress	861	32.05
13-16	55.4-60.8	Cold	Chills	181	27.78
16-19	60.8-66.2	Cool	Body cooling	150	25.91
19-22	66.2-71.6	Slightly cool	Vasoconstriction	158	24.96
22-25	71.6-77.0	Comfort zone	Neutral	280	23.72
25-28	77.0-82.4	Warm	Light sweat, vasodilatation	188	24.09
28-31	82.4-87.8	Moderately hot	Moderate sweating	8	28.38
31-34	87.8-93.2	Hot	Profuse sweating	0	0
>34	>93.2	Very hot	Impaired thermoregulation	0	0

†ET: effective temperature; ET_w: wind effective temperature.

present a U-shape, which had largest values in "very cold" and "moderately hot". Note that there was no data during our study period in the thermal comfort index with "hot" and "very hot". When an unreasonable temperature smoother came out, such as the initial plots showing in figure 3.7 and figure 3.9, it was suggested to check corresponding estimates results to determine whether this model fitting had convergence problem. This double-check is very important because we had proved that irrational smoothing functions may affect the convergence of estimated parameters. Although it was not absolutely solvable, most situations could be improved by adjusting the starting values of smoothing parameters and the number of knots in smoothing functions manually. In addition, the statement from BayesX manual about the consistence of the number of knots used in smoothing functions is not entirely guaranteed. A better way to handle it carefully is relying on re-fitting the same model with a different numbers of knots based on a bench-mark. For example, if we got a reasonable smoothing function plot with a fixed

number of knots, it is better to re-fit the same model with some versatile numbers of knots around the initial number of knots. If the new plots are similar to the original one, this plot can be confirmed to be free from the knots effect.

The missing data problem is inevitable in air pollution study, and some imputation methods did not have the anticipated effect, especially in multiple imputation method. As we know, the multiple imputation method needs strong assumptions (Little & Rubin, 1987), and may not be appropriate for imputing time series data. The NNI1 and NNI2 provide convenient ways to impute missing data, but each one has its own flaws. In our case study, three methods encountered the convergence problem, and, in particular, the multiple imputation method which needs repeated imputations, made the entire procedure very inefficient. Comparing the three methods, the NNI2 has the worse relative results because it underestimated most air pollutant effects in model 1, model 2, model 4, and model 5. This situation appeared in simulation, and under purer data structure, the underestimated situation in $\hat{\beta}$ was even worse than the other two methods, which did not occur in case study. Moreover, these imputation methods made severe damages when missing data appeared in smoothing functions. This situation was not significant in our real data, and the simulation implied that imputed data will not improve either estimates or smoothing functions. While MI-MCMC was preferred to CCA because CCA may cause over-fitted models and biased estimates if the data does not follow MAR mechanism (Carder, McNamee, Beverland, Elton, Van Tongeren, Cohen, Boyd, MacNee, & Agius, 2008), our real data did not violate this mechanism. Because the GGAMM is a kind of semiparametric model, some semiparametric imputation methods can be considered (Durrant, 2005; Lipsitz, Zhao, & Molenberghs, 1998). The model-based imputation method (Zanobetti et al., 2000) failed here because it raised the concurvity level, and made the model-fitting harder to reach convergence.

The application of principal component analysis solved the technical problem in the

distributed lag model with the structure of the GGAMM in BayesX, and the estimates from principal component variables were more stable, which indirectly but considerably mark down the probability of a convergence problem. A quantification research for mortality displacement used 45 lags total suspended particulate (TSP) in the generalized additive distributed lag model, and the strongest log relative risk occurred at current effect. The trend decreased to near zero around lag 15, and slightly climb until lag 25 (Zanobetti et al., 2000). This study design was severely affected by missing data, and reduced the power with CCA. Therefore, imputing missing data is always a precursor of fitting distributed lag models. This is a complicated design because more problems will come about with each other in the GGAMM, such as multicollinearity, concurvity and convergence problem. Most studies only used two or three lag effects (Peng et al., 2005; Roberts & Martin, 2007), but our findings showed that lag effects longer than two to six days probably still had an influence on mortality rate.

Most of our models had moderate multicollinearity and concurvity levels around 0.3~0.5. Carbon monoxide and ozone had higher concurvity levels over 0.6 with smoothing and spatial functions. A study found some extreme results in New York City (Lipsitz, Zhao, & Molenberghs, 2007), but a comparison with our findings should still be considered. The concurvity level in this NYC study can be controlled in single pollutant models, and the highest one (0.77) appeared in ozone adjusting by natural splines of same-day temperature, dew-point, and time trend, which was close to our results in model 5. The lowest concurvity level was 0.28 in NO_2, but increased to 0.60 when including day-of-week variables. This level was also close to our result. Generally speaking, their single pollutant models were producing concurvity level≥0.6 when day-of-week variables were considered, and most multiple pollutant models had concurvity level from 0.69 to 0.91. Their findings implied that adding more factors and smoothing functions can increase the risk of more concurvity. Note that this NYC study used 18~30 air pollution

monitoring stations' data for different models, but actually the average values across these monitors were used in the GAM, so the spatial variation among monitoring stations did not consider in model fitting.

Even though some simulations explored biases from the effect of concurvity by using the GAM with air pollution data (Ramsay et al., 2003a; Ramsay et al., 2003b), the conditional bootstrap method and partial regression approach are rarely used in most air pollution studies (Figueiras et al., 2005; He et al., 2006). The conditional bootstrap method was proved to produce more biases from a simulation (results are not shown here), and unable to work in the GGAMM. The partial regression approach needs to have advanced development to correspond with the structure of the GGAMM, but this part is not the main purpose of our study. At least, we not only revealed the influence of concurvity by simulations, but also discovered that it is a potential source of causing a convergence problem in the GGAMM.

Multicollinearity is very common in any linear model, but the 2-stage Bayesian hierarchical model is actually not a unified linear model. It causes the evaluation of multicollinearity in the 2-stage Bayesian hierarchical model to be less intuitive than the traditional generalized linear regression model in air pollution epidemiology (Chen, Chock, & Winkler, 1999). A precursor of the 2-stage regression model discussed the confounding from multicollinearity in air pollution research (Marcus & Kegler, 2001). They suggested that factor analysis can be used as a preprocessing step to evaluate multicollinearity in 2-stage regression model, but did not propose clear analyses to prove it. Our approach succeeded multiple linear regression models, and offered the PCA-adjusted estimates to help us have indirect interpretations for each factor in the GGAMM. The immediate way of handling multicollinearity under the structure of the GGAMM is more anticipated, but generally speaking, some traditional approaches of handling multicollinearity in regression models can theoretically be applied in the

GGAMM, such as dropping one of the variables, obtaining more data, or centralizing the predictor variables.

Compared with the 2-stage Bayesian hierarchical model, the GGAMM revealed some advantages. First, the geographical correlation can be virtualized by spatial map for using Markov random fields with real geographical data. The 2-stage Bayesian hierarchical model cannot use geographical data, but uses another Bayesian model containing coefficients from the first stage to consider between-city variability (Dominici et al., 2002a). An improved 2-stage Bayesian hierarchical model was developed by Dominici et al. (2003b), and the distance between two cities can be considered in the second stage. However, the purpose of using the model in the second stage is still obtaining a weighted overall average of the MLE of PM_{10}, and no spatial function can be independently presented. Second, the model-fitting of the GAM could be easily damaged by missing data, especially in smoothing functions. The confidence intervals of estimated parameters also tended to be much wider, causing unreliable coefficients. This potential risk can be reduced by the GGAMM for two reasons: 1) geographical data will not be affected by missing data, and 2) smoothing functions can be fitted well by enough valid observations from all cities. Therefore, both fixed and random effects are controlled by a nation-level time smoother and temperature smoother, and have more reliable estimates and confidence intervals. Third, Markov random fields offer the ability to explain spatial correlation across cities, and removed possible confounding influence on estimated coefficients. Hence, from the unified model structure of the GGAMM, we can conveniently demonstrate the relationship between nation-level and city-level effects to mortality from fixed and random effect estimates in any air pollutant.

In summary, the general findings from our analyses are: 1) the GGAMM provided an integrate model structure concerning national average estimates, city-specific estimates, smoothing and spatial functions simultaneously, and the results were acceptable in most

cases; 2) when the model fitting encounter a convergence problem in BayesX, users can solve it by adjusting either starting values of smoothing parameter or the number of knots manually to modify estimating results. A jackknife-bootstrap approach can be the alternative in case that the two methods do not work successfully; 3) single the pollutant model and distributed lag model with PM_{10} in the GGAMM presented smaller estimates than the results in the 2-stage Bayesian hierarchical model in national level, but co-pollutants, especially in NO_2 and O_3, had a stronger association to mortality in the GGAMM; 4) the application of missing data imputation methods in case study made each fixed effect much stronger, but easily cause convergence problems, especially in multiple imputation method; 5) the application of principal component analysis solved the technical problem in BayesX and the theoretical problem in multicollinearity; 6) The multicollinearity was not very severe in our case study, but concurvity severity appeared in co-pollutant model PM_{10}+O_3 (model 5); 7) Our models still needs some adjustments to satisfy model diagnostics, although the corresponding methods using in the GGAMM are still undeveloped; 8) the simulation results confirmed that concurvity is one of the reason causing a convergence problem in BayesX, and it also made biases on all estimates. In addition, missing data imputation methods did not improve estimates much, and even made severe damages when missing data appeared in smoothing functions.

5.2. Limitations

In the process of fitting GGAMMs with spatio-temporal air pollution data in BayesX, we actually encountered many problems and difficulties. Some of them have been solved in chapter 3, but for others, there is still no way to address them. The methodology of structured additive models is almost completely developed, and can be accomplished in BayesX, the only statistical software which can fit the GGAMM without writing

complicated programs. Therefore, if any model-fitting failed in BayesX, there is no alternative software to successfully accomplish it. An unnatural limitation in BayesX is its data capacity. Not like SAS and some other statistical software, the amount of data simultaneously importing and fitting in BayesX is restricted. Although the manual of BayesX showed a simple example with a data set containing 100,000 cases and 2 variables, the actual amount of spatio-temporal air pollution data which can be successfully fitted in BayesX was no more than 30,000. This upper bound is the main restriction making the number of U.S. cities 15 and the length of time studied 5 years (1,826 days). As long as it continues to exceed the capacity, BayesX will crash immediately. One possible reason is that there are too many matrix calculations in priors, posteriors, estimating equations, and B-splines, which easily exceeds the maximum of memory in computers. Despite the memory being expanded to 4GB, the amount of data used in our study can no longer increase. We can expect to install more RAM in computers, but according to the official document from Microsoft®, 32-bit Windows with 4GB RAM in any version of Vista is commonly used by most users. These hardware standards are still unaffordable. Using 128GB RAM looks possibly, but it still does not fit common computer hardware standards. A workstation system should be an alternative, but the UNIX version of BayesX remains under development.

Another technical problem of BayesX happened in simulation. It contains two issues. One is that there is no compiled do-loop function in BayesX, and another one is that the number of models which can be fitted in an implemented BayesX is limited. The lack of do-loop function makes users unable to conveniently and simultaneously fit a series of models with a few lines of programs. An alternative method is using external software to generate corresponding syntax in each model, and export them together in a *.prg* file, a format of external text file which can be read in BayesX. The second limitation comes from the lack of a do-loop function. We found the maximum number of models to fit in a

BayesX windows at once to be roughly 25,000. Exceeding this number will also cause BayesX to crash. This disadvantage has a severe influence on huge simulations, especially in those with bootstrap and jackknife methods. Users have to control the amount in a *.prg* file by splitting them into several smaller *.prg* files. Whatever, these external operations will always make advanced research more complicated and inconvenient.

About the issue of software, BayesX is well-compiled software with fixed syntax for fitting models only, and has less flexibility to modify from inside. Compared with other statistical software, BayesX is only for fitting structure additive models. The *gamm()* function in R software can also fit structure additive models, but its algorithm is based on Lin and Zhang (1999), which means it can only fit the GAMM. No Bayesian approach is applied in this package, and only contour data can be used to fit spatial function. While BayesX provides some syntax to make data management and descriptive statistics, there are few functions to perform some adjustments and improvements if we develop new algorithms or theorems based on structure additive models. For example, suppose a new method is developed for solving the concurvity problem in the GGAMM by some elements internally adjusting in estimating equations or iterations—BayesX cannot process this new method unless its programmers write algorithms in source codes, and compile them in a new version of BayesX. As a result, BayesX becomes an entirely closed programming environment which is unable to develop from users, with the exception of original programmers. In addition, the *offset* function still has problems when assigning population data, but it is still doubtful that whether the offset can affect the result with huge population and pretty small death counts.

Besides technical limitations in BayesX , there are still some issues which have not been accomplished in the methodology of the GGAMM. A critical one shall be the model diagnostic method and influential analysis, and of course, BayesX does not support

corresponding syntax either. The model diagnostics shown in chapter 3.11 were borrowed from the general linear mixed model, but the influence diagnostics and outliers tests for original GAMM which was developed by Lin and Zhang (1999) or GAM had been published (Fung, Zhu, Wei, & He, 2002; Kim, Park, & Kim, 2002). Lin's and Zhang's GAMM can be regarded as a protocol of Bayesian generalized structured additive models; therefore, the development of relative methods in the GGAMM can follow Fung et al.'s research to establish its own model diagnostic methods. Nonetheless, the relative theories have not been studied yet. It is expected that some model diagnostic methods developed on Lin's and Zhang's methodology can also be applied in the GGAMM, but subject to restrictions on BayesX, it is still pretty hard to implement.

An extended problem from model diagnostics is the over-dispersion situation from zero-inflated Poisson (ZIP) data. A similar situation also happens in zero-inflated negative binomial (ZINB) data. Air pollution research inevitably often has to use ZIP or ZINB data, but the two types of data can only be fitted by the GGAMM with MCMC approach in BayesX. Compared with MCMC method and REML method in the GGAMM, the fully Bayesian approach implementing in MCMC method is much more efficient than the REML method, but the main reason that the REML method does not support ZIP or ZINB data is that the standard error of parameter is still inestimable. However, our air pollution data cannot be fitted by the GGAMM with MCMC method because BayesX encountered errors from floating calculation. The main reason is still unknown, but according to the statement, the MCMC method can handle more than 1,000 variables and 200,000 observations. Air pollution data fitted by the GGAMM with MCMC method should be technically implemented with some modifications.

Using the options of smoothing functions compiling in BayesX, users cannot define the degree of freedom, but instead estimate the generalized cross-validation (GCV) value for the degree of freedom. Meanwhile, there is no possibility to specify the degrees of

freedom since estimation of the smoothing parameter is an integral part of the model fitting process. There is a data-driven choice of the smoothing parameter leading to a data-driven choice of the degrees of freedom which ultimately avoids user-specified subjective choices of smoothing parameters. As a result, this default setup makes studies unable to perform sensitivity analyses over different degrees of freedoms which are the necessary part in previous air pollution research. In addition, even though BayesX allows users to define different numbers of knots in smoothing functions, it is unlike some packages in R, such as *gamm()*, which can import user-defined knots from non-kriging or non-geokriging data. Meanwhile, the distance between any two nearby knots from non-kriging or non-geokriging data is always equal in BayesX.

Finally, in regards to the convergence problem, we developed a jackknife-bootstrap approach to adjust overestimated standard errors of fixed and random effects. The disadvantage of this method is making 95% confidence intervals conservatively, but it can guarantee the robustness will be held with a different number of jackknife estimates drawing with replacement. However, using this method also indicates that the original estimated smoothing and spatial functions cannot be used immediately. The jackknife method cannot adjust parameters and smoothing/spatial functions simultaneously. As a result, changing the starting values of smoothing parameters or the numbers of knots in smoothing functions is still the first choice to look for reasonable estimates. It is anticipated that the jackknife-bootstrap approach can also apply in each fitted value of splines or spatial effects, but its efficiency should be evaluated further. We propose to proceed from the starting values of smoothing parameters because it can be modified up to the value of 1,000, but the numbers of knots must be limited (20 knots~50 knots), because it could make smoothing functions too twisted. In the event that the three methods are all unavailable, data transformation might be an alternative strategy.

5.3. Future work

The use of the GGAMM in air pollution research with spatio-temporal data is a new application, and the multitudes of analyses done in our study are just a premiere that still leaves a lot of work to be done. First, rather than endlessly expanding computer hardware, we anticipate that reducing the dimension of matrixes used in algorithms could be a more efficient way to use more data in BayesX. In addition, this kind of development could potentially save estimating time, even though BayesX is less time-consuming than WINBUGS or R when fitting some structure additive models. However, in our experience it is still too time-consuming when models include more variables, especially in models with many random effects. Sometimes, assigning more knots in smoothing functions will also extend process time. For example, the process time to fit model 1 shown in section 3.3 was 3 minutes 10 seconds. When immediately including two PM_{10} lag effects (model 2, including of fixed and random effects), without changing any default values of parameters in program, the process time extended to 48 minutes 53 seconds. When just increasing the number of knots of the time smoother from 31 to 46 in model 1, the process time extended to 5 minutes and 31 seconds. This time-consuming estimating procedure is an obstacle of applying some statistical approaches, such as bootstrap and jackknife method. Some complicated simulations will also pose problems. However, as long as the two disadvantages can be overcome, it is definite that the use of the GGAMM or the other structure additive models will become overwhelmingly popular in the future.

Second, the solution approach of concurvity appearing in the GGAMM is still underdeveloped. This issue has been examined in some air pollution studies (Ramsay et al, 2003a; Ito et al, 2007), and a conditional bootstrap method was also proposed in the GAM (Figueiras et al, 2005). Unfortunately, we applied the conditional bootstrap method in the GGAMM and found it failed to solve dwindle concurvity level in a small

simulation. We still believed that the damages of a high concurvity level in the GAM will still occur in the GGAMM, but a new theory should be developed first to handle it in the GGAMM or the other structured additive models.

Third, the model diagnostic methods based on the GGAMM or the other generalized structured additive models with either the MCMC method or REML method should be the next step of methodological establishment. In addition, how to solve the problem of inestimable standard errors in parameters on the link function for ZIP and ZINB distribution could be another challenge. Concerning the application of the GGAMM in air pollution and adverse human health research, there are some opportunities related to enhance or develop advanced research: 1) it is believed that limited or inaccurate environmental exposure data induce huge measurement errors to properly estimate small risks, and researchers are unable to detect small effects that would probably be undetectable (Dominici et al, 2003c). Ambient measurements from personal exposure may be a better surrogate for average population exposure from air pollution monitoring stations. In addition, scientists were previously worried about the misrecognition of spatial variation in these studies (Greenland & Morgenstern, 1989; Sheppard, Prentice, & Rossing, 1996), but now the GGAMM or the other structured additive models have proven that they have the ability to handle this issue; 2) the biases from measurement errors in monitoring data have been confirmed in some air pollution studies using linear regression models (Zeager et al, 2000) and the generalized additive model (Dominici et al, 2000b), so the succeeded research in the GGAMM should also be expected; 3) short-term mortality displacement comes from the near-death individuals who are infected by other diseases, such as cardiovascular disease or cerebrovascular disease, and high concentration air pollution could hasten their deaths within a short period. This issue induces the timescale analysis for identifying the influence of different timescales, and initially has been done by a hierarchical Poisson regression model and Fourier analysis in

the 4 U.S city study (Dominici et al, 2003b). The GGAMM should have the power to integrate more cities to estimate association between air pollution exposure and mortality at different timescales of variation.

Finally, except for the GGAMM, other members in the generalized structured additive models can be implemented in air pollution and human health studies for different designs or purposes, but the application of related models is still very rare. A re-analysis of Utah Valley Study (Coull et al, 2001) with a general additive mixed model is a pioneer of this kind of application, and can be a good example to follow. Some ideas could be accomplished in the future: 1) using boundary, centroid, or kriging data to fit spatial functions; 2) performing versatile advanced interaction terms in models, such as geographically weighted regression, two-dimensional surface function, and time-varying effect in Cox PH models or multi-state models; 3) analyzing panel or clinical trial data by fitting models for continuous time survival analysis based on structured hazard regression (Kneib & Fahrmeir, 2007; Kneib, 2006b) and multi-state model (Kneib & Hennerfeind, 2006).

APPENDIX A

Tables of estimates from starting values of smoothing parameters

Model 1

λ_{time}	Parm	λ_{tmean} 10	11	12	13	14	15
	$\hat{\beta}$	0.000105	0.000107	0.000013	0.000107	0.000100	0.000014
10	se($\hat{\beta}$)	0.000286	0.000375	0.168194	0.000284	0.000309	0.778382
	se($\hat{b}$)	0.000194	0.000773	0.651408	0.000167	0.000387	3.014658
	$\hat{\beta}$	0.000017	0.000019	0.000021	0.000106	0.000100	-0.000046
11	se($\hat{\beta}$)	0.675334	0.096806	1.233190	0.000290	0.000330	0.479905
	se($\hat{b}$)	2.615557	0.374921	4.776118	0.000235	0.000522	1.858661
	$\hat{\beta}$	0.000103	0.000106	-0.000046	0.000104	0.000019	0.000098
12	se($\hat{\beta}$)	0.000360	0.000375	0.126045	0.000348	0.097463	0.000335
	se($\hat{b}$)	0.000694	0.000772	0.488165	0.000626	0.377467	0.000549
	$\hat{\beta}$	0.000101	0.000102	0.000017	0.000100	0.000098	0.000104
13	se($\hat{\beta}$)	0.000298	0.000353	0.050717	0.000327	0.000311	0.000370
	se($\hat{b}$)	0.000306	0.000656	0.196412	0.000503	0.000399	0.000746
	$\hat{\beta}$	0.000014	0.000100	0.000098	0.000091	0.000099	0.000102
14	se($\hat{\beta}$)	0.952007	0.000321	0.000320	0.000303	0.000318	0.000344
	se($\hat{b}$)	3.687113	0.000469	0.000459	0.000336	0.000449	0.000603
	$\hat{\beta}$	0.000099	0.000100	0.000106	0.000017	0.000098	0.000100
15	se($\hat{\beta}$)	0.000332	0.000305	0.000292	0.049699	0.000315	0.000323
	se($\hat{b}$)	0.000535	0.000354	0.000251	0.192470	0.000424	0.000477

Model 2

λ_{time}	Parm	λ_{tmean} 10	11	12	13	14	15
10	$\hat{\beta}_1$	0.000675	0.000158	0.000166	0.000216	0.000676	0.000144
	se($\hat{\beta}_1$)	0.196834	0.000405	0.000398	0.000369	0.266623	0.000356
	se($\hat{b}_1$)	0.762327	0.000693	0.000641	0.000500	1.032623	0.000399
	$\hat{\beta}_2$	-0.000530	-0.000024	-0.000164	-0.000132	-0.000528	-0.001494
	se($\hat{\beta}_2$)	0.133549	0.000412	0.086452	0.085502	0.088435	0.066438
	se($\hat{b}_2$)	0.517211	0.000618	0.334801	0.331121	0.342475	0.257295
	$\hat{\beta}_3$	-0.001911	-0.001984	-0.001913	-0.001908	-0.001913	0.000160
	se($\hat{\beta}_3$)	0.063516	0.177789	0.068344	0.182711	0.113770	0.000419
	se($\hat{b}_3$)	0.245960	0.688567	0.264662	0.707626	0.440610	0.000811
11	$\hat{\beta}_1$	0.000097	0.000091	0.000108	0.000460	0.000676	0.000148
	se($\hat{\beta}_1$)	0.000351	0.000367	0.000373	0.077479	0.838364	0.000368
	se($\hat{b}_1$)	0.000388	0.000499	0.000530	0.300056	3.246968	0.000477
	$\hat{\beta}_2$	-0.000062	-0.000031	-0.000046	-0.000025	-0.000529	-0.001516
	se($\hat{\beta}_2$)	0.000400	0.000364	0.000382	0.000374	0.595126	0.125494
	se($\hat{b}_2$)	0.000579	0.000339	0.000459	0.000387	2.304908	0.486025
	$\hat{\beta}_3$	0.000157	0.000150	0.000129	-0.002165	-0.001912	0.000182
	se($\hat{\beta}_3$)	0.000318	0.000319	0.000362	0.097382	0.096031	0.000346
	se($\hat{b}_3$)	0.000244	0.000261	0.000540	0.377142	0.371903	0.000422
12	$\hat{\beta}_1$	0.000676	0.000105	0.000459	0.000675	0.000136	0.000676
	se($\hat{\beta}_1$)	0.386378	0.000360	0.096833	0.070675	0.000351	0.278259
	se($\hat{b}_1$)	1.496432	0.000449	0.375017	0.273698	0.000360	1.077687
	$\hat{\beta}_2$	-0.000528	-0.000050	-0.000031	-0.000528	-0.001511	-0.000528
	se($\hat{\beta}_2$)	0.241014	0.000387	0.000383	0.416730	0.138876	0.193717
	se($\hat{b}_2$)	0.933432	0.000492	0.000440	1.613983	0.537856	0.750247
	$\hat{\beta}_3$	-0.001913	0.000131	-0.002162	-0.001912	0.000185	-0.001913
	se($\hat{\beta}_3$)	0.123914	0.000361	0.078948	0.132455	0.000335	0.160887
	se($\hat{b}_3$)	0.479898	0.000536	0.305746	0.512978	0.000352	0.623100
13	$\hat{\beta}_1$	0.000105	0.000612	0.000738	0.000100	0.000461	0.000107
	se($\hat{\beta}_1$)	0.000367	0.209067	0.273995	0.000375	0.059710	0.000365
	se($\hat{b}_1$)	0.000493	0.809705	1.061174	0.000545	0.231234	0.000482
	$\hat{\beta}_2$	-0.000046	-0.000536	-0.001920	-0.000040	-0.000026	-0.000035
	se($\hat{\beta}_2$)	0.000383	1.089360	0.171184	0.000374	0.000377	0.000370
	se($\hat{b}_2$)	0.000468	4.219052	0.662980	0.000408	0.000403	0.000378
	$\hat{\beta}_3$	0.000125	-0.001949	0.000163	0.000143	-0.002166	0.000114
	se($\hat{\beta}_3$)	0.000373	0.934210	0.000395	0.000334	0.153929	0.000388

	$se(\hat{b}_3)$	0.000603	3.618176	0.000683	0.000370	0.596153	0.000684
	$\hat{\beta}_1$	0.000458	0.000112	0.000459	0.000611	0.000125	0.000106
	$se(\hat{\beta}_1)$	0.573193	0.000360	0.201665	0.046127	0.000353	0.000345
	$se(\hat{b}_1)$	2.219966	0.000443	0.781038	0.178608	0.000381	0.000339
	$\hat{\beta}_2$	-0.000026	-0.000063	-0.000022	-0.000536	-0.000016	-0.000063
14	$se(\hat{\beta}_2)$	0.000377	0.000412	0.000372	0.223534	0.000382	0.000410
	$se(\hat{b}_2)$	0.000404	0.000638	0.000368	0.865732	0.000450	0.000630
	$\hat{\beta}_3$	-0.002164	0.000125	-0.002167	-0.001948	-0.001973	0.000129
	$se(\hat{\beta}_3)$	0.056609	0.000386	0.194721	0.115720	0.151860	0.000381
	$se(\hat{b}_3)$	0.219217	0.000671	0.754142	0.448163	0.588143	0.000643
	$\hat{\beta}_1$	0.000676	0.000677	0.000676	0.000737	0.000097	0.000610
	$se(\hat{\beta}_1)$	0.189319	0.382686	1.019010	0.112704	0.000356	0.151886
	$se(\hat{b}_1)$	0.733220	1.482134	3.946619	0.436486	0.000373	0.588239
	$\hat{\beta}_2$	-0.000528	-0.000544	-0.000528	-0.001929	-0.001530	-0.000535
15	$se(\hat{\beta}_2)$	0.129039	0.150525	0.635551	0.084190	0.096627	0.131020
	$se(\hat{b}_2)$	0.499745	0.582962	2.461475	0.326041	0.374222	0.507418
	$\hat{\beta}_3$	-0.001913	-0.001888	-0.001914	0.000175	0.000140	-0.001949
	$se(\hat{\beta}_3)$	0.154094	0.023828	0.530478	0.000347	0.000413	0.205771
	$se(\hat{b}_3)$	0.596789	0.092197	2.054529	0.000419	0.000778	0.796937

Model 3

λ_{time}	Parm	λ_{tmean} 10	11	12	13	14	15
10	$\hat{\beta}_1$	0.000196	0.000195	0.000220	0.000221	0.000220	0.000220
	se($\hat{\beta}_1$)	0.088258	0.084556	0.087631	0.082792	0.081822	0.083734
	se($\hat{b}_1$)	0.341813	0.327477	0.339383	0.320643	0.316885	0.324292
	$\hat{\beta}_2$	-0.000005	-0.000005	-0.000005	-0.000005	-0.000005	-0.000005
	se($\hat{\beta}_2$)	0.081648	0.081648	0.081647	0.081651	0.081650	0.081647
	se($\hat{b}_2$)	0.316223	0.316222	0.316216	0.316231	0.316229	0.316217
11	$\hat{\beta}_1$	0.000194	0.000195	0.000221	0.000221	0.000221	0.000221
	se($\hat{\beta}_1$)	0.080495	0.077259	0.079291	0.080210	0.076693	0.080765
	se($\hat{b}_1$)	0.311744	0.299210	0.307084	0.310642	0.297021	0.312790
	$\hat{\beta}_2$	-0.000005	-0.000005	-0.000005	-0.000005	-0.000005	-0.000005
	se($\hat{\beta}_2$)	0.081651	0.081650	0.081650	0.081650	0.081649	0.081651
	se($\hat{b}_2$)	0.316233	0.316228	0.316229	0.316231	0.316224	0.316234
12	$\hat{\beta}_1$	0.000196	0.000198	0.000198	0.000196	0.000198	0.000220
	se($\hat{\beta}_1$)	0.081868	0.082859	0.081704	0.087217	0.086197	0.081964
	se($\hat{b}_1$)	0.317065	0.320900	0.316430	0.337782	0.333828	0.317433
	$\hat{\beta}_2$	-0.000005	-0.000005	-0.000005	-0.000005	-0.000005	-0.000005
	se($\hat{\beta}_2$)	0.081651	0.081651	0.081649	0.081658	0.081651	0.081649
	se($\hat{b}_2$)	0.316233	0.316231	0.316227	0.316261	0.316231	0.316225
13	$\hat{\beta}_1$	0.000221	0.000178	0.000221	0.000221	0.000262	0.000195
	se($\hat{\beta}_1$)	0.071796	0.039088	0.082335	0.111095	0.081963	0.061053
	se($\hat{b}_1$)	0.278054	0.151368	0.318873	0.430262	0.317432	0.236445
	$\hat{\beta}_2$	-0.000005	-0.000007	-0.000005	-0.000005	-0.000007	-0.000005
	se($\hat{\beta}_2$)	0.081648	0.081629	0.081650	0.081648	0.081646	0.081654
	se($\hat{b}_2$)	0.316223	0.316149	0.316227	0.316221	0.316213	0.316244
14	$\hat{\beta}_1$	0.000193	0.000194	0.000192	0.000220	0.000196	0.000194
	se($\hat{\beta}_1$)	0.085497	0.087237	0.073339	0.083137	0.080626	0.079331
	se($\hat{b}_1$)	0.331118	0.337859	0.284029	0.321977	0.312253	0.307236
	$\hat{\beta}_2$	-0.000005	-0.000005	-0.000005	-0.000005	-0.000005	-0.000005
	se($\hat{\beta}_2$)	0.081655	0.081641	0.081617	0.081652	0.081652	0.081663
	se($\hat{b}_2$)	0.316247	0.316194	0.316103	0.316236	0.316236	0.316278
15	$\hat{\beta}_1$	0.000197	0.000196	0.000220	0.000221	0.000194	0.000221
	se($\hat{\beta}_1$)	0.083373	0.082452	0.081541	0.079796	0.104653	0.081438
	se($\hat{b}_1$)	0.322893	0.319324	0.315796	0.309038	0.405310	0.315397
	$\hat{\beta}_2$	-0.000005	-0.000005	-0.000005	-0.000005	-0.000005	-0.000005
	se($\hat{\beta}_2$)	0.081650	0.081643	0.081648	0.081651	0.081628	0.081649
	se($\hat{b}_2$)	0.316231	0.316201	0.316222	0.316233	0.316144	0.316223

Model 4

λ_{time}	Parm	λ_{tmean} 10	11	12	13	14	15
10	$\hat{\beta}_1$	0.000163	0.000056	0.000136	0.000188	0.000093	0.000101
	se($\hat{\beta}_1$)	0.000441	0.033306	0.000375	0.000460	0.000443	0.000414
	se($\hat{b}_1$)	0.000638	0.120042	0.000212	0.000738	0.000627	0.000468
	$\hat{\beta}_2$	0.001224	0.002160	0.001145	0.001173	0.002076	0.002058
	se($\hat{\beta}_2$)	0.000846	0.095700	0.000829	0.000829	0.295381	0.112645
	se($\hat{b}_2$)	0.000795	0.344993	0.000665	0.000655	1.064998	0.406102
11	$\hat{\beta}_1$	0.000166	0.000130	0.000147	0.000139	0.000156	0.000147
	se($\hat{\beta}_1$)	0.000451	0.000375	0.000410	0.000404	0.000433	0.000426
	se($\hat{b}_1$)	0.000686	0.000209	0.000464	0.000426	0.000596	0.000554
	$\hat{\beta}_2$	0.001224	0.001170	0.001164	0.001233	0.001238	0.001259
	se($\hat{\beta}_2$)	0.000848	0.000837	0.000830	0.000858	0.000856	0.000865
	se($\hat{b}_2$)	0.000810	0.000739	0.000668	0.000893	0.000874	0.000942
12	$\hat{\beta}_1$	0.000437	0.000162	0.000051	0.000133	0.000458	0.000051
	se($\hat{\beta}_1$)	0.091350	0.000443	0.000386	0.000379	0.141606	0.572836
	se($\hat{b}_1$)	0.329354	0.000650	0.000221	0.000242	0.510559	2.065386
	$\hat{\beta}_2$	0.001227	0.001235	0.002126	0.001258	0.001150	0.002161
	se($\hat{\beta}_2$)	0.000863	0.000850	0.317771	0.000876	0.000830	0.044137
	se($\hat{b}_2$)	0.000886	0.000828	1.145722	0.001027	0.000630	0.159011
13	$\hat{\beta}_1$	0.000165	0.000447	0.000183	0.000134	0.000170	0.000135
	se($\hat{\beta}_1$)	0.000460	1.259810	0.000454	0.000388	0.000466	0.000374
	se($\hat{b}_1$)	0.000734	4.542323	0.000706	0.000324	0.000764	0.000203
	$\hat{\beta}_2$	0.001314	0.001193	0.001167	0.001179	0.001290	0.001178
	se($\hat{\beta}_2$)	0.000886	0.000848	0.000828	0.000839	0.000871	0.000842
	se($\hat{b}_2$)	0.001076	0.000775	0.000641	0.000747	0.000975	0.000777
14	$\hat{\beta}_1$	0.000055	0.000055	0.000054	0.000137	0.000427	0.000131
	se($\hat{\beta}_1$)	0.123753	0.123179	0.099355	0.000375	0.138230	0.000371
	se($\hat{b}_1$)	0.446186	0.444118	0.358215	0.000179	0.498386	0.000164
	$\hat{\beta}_2$	0.002161	0.002161	0.002159	0.002009	0.001263	0.001135
	se($\hat{\beta}_2$)	0.399323	0.270087	0.251801	0.157348	0.000880	0.000825
	se($\hat{b}_2$)	1.439764	0.973791	0.907857	0.567293	0.001005	0.000631
15	$\hat{\beta}_1$	0.000055	0.000134	0.000136	0.000142	0.000052	0.000052
	se($\hat{\beta}_1$)	0.127711	0.000386	0.000372	0.000412	0.197652	0.061465
	se($\hat{b}_1$)	0.460456	0.000308	0.000183	0.000479	0.712637	0.221592
	$\hat{\beta}_2$	0.002161	0.001220	0.001138	0.001235	0.002162	0.002164
	se($\hat{\beta}_2$)	0.179242	0.000856	0.000827	0.000853	0.301458	0.191393
	se($\hat{b}_2$)	0.646234	0.000886	0.000648	0.000858	1.086904	0.690049

Model 5

λ_{time}	Parm	λ_{tmean} 10	11	12	13	14	15
10	$\hat{\beta}_1$	0.000227	0.000218	-0.000380	-0.000371	0.000207	0.000228
	se($\hat{\beta}_1$)	0.000385	0.000346	0.123656	0.196623	0.000356	0.000408
	se($\hat{b}_1$)	0.000531	0.000291	0.462672	0.735692	0.000350	0.000647
	$\hat{\beta}_2$	0.001855	0.001832	0.001825	0.001937	0.001808	0.001806
	se($\hat{\beta}_2$)	0.000815	0.000795	0.000925	0.000817	0.000798	0.000801
	se($\hat{b}_2$)	0.000650	0.000451	0.001184	0.000495	0.000344	0.000362
11	$\hat{\beta}_1$	0.000224	-0.000176	-0.000176	0.000215	0.000209	0.000208
	se($\hat{\beta}_1$)	0.000383	0.117644	1.851610	0.000385	0.000372	0.000363
	se($\hat{b}_1$)	0.000517	0.440172	6.928088	0.000520	0.000449	0.000396
	$\hat{\beta}_2$	0.001840	0.000933	0.000935	0.001791	0.001792	0.001803
	se($\hat{\beta}_2$)	0.000836	0.308503	3.222230	0.000854	0.000812	0.000803
	se($\hat{b}_2$)	0.000825	1.154301	12.056492	0.000894	0.000480	0.000393
12	$\hat{\beta}_1$	0.000209	0.000151	-0.000060	-0.000371	-0.000392	-0.000175
	se($\hat{\beta}_1$)	0.000374	0.000414	4.430980	0.197671	0.118872	0.058461
	se($\hat{b}_1$)	0.000459	0.000610	16.579174	0.739611	0.444768	0.218719
	$\hat{\beta}_2$	0.001781	0.001034	0.001254	0.001942	0.001822	0.000935
	se($\hat{\beta}_2$)	0.000842	1.472320	6.927030	0.000815	0.000885	0.528293
	se($\hat{b}_2$)	0.000777	5.508920	25.918526	0.000488	0.000872	1.976687
13	$\hat{\beta}_1$	-0.000382	0.000206	-0.000465	-0.000381	0.000215	0.000224
	se($\hat{\beta}_1$)	0.144754	0.000397	0.131895	0.121192	0.000415	0.000392
	se($\hat{b}_1$)	0.541613	0.000580	0.493498	0.453450	0.000673	0.000563
	$\hat{\beta}_2$	0.001844	0.001714	0.001720	0.001910	0.001700	0.001813
	se($\hat{\beta}_2$)	0.000896	0.000889	0.000875	0.000816	0.000888	0.000867
	se($\hat{b}_2$)	0.001003	0.001046	0.000405	0.000392	0.001048	0.001028
14	$\hat{\beta}_1$	-0.000383	0.000240	0.000201	0.000196	0.000214	-0.000512
	se($\hat{\beta}_1$)	0.147136	0.000427	0.000341	0.000359	0.000366	0.138424
	se($\hat{b}_1$)	0.550527	0.000748	0.000243	0.000359	0.000417	0.517930
	$\hat{\beta}_2$	0.001880	0.001836	0.001768	0.001726	0.001777	0.001917
	se($\hat{\beta}_2$)	0.000843	0.000800	0.000837	0.000893	0.000862	0.000797
	se($\hat{b}_2$)	0.000618	0.000353	0.000741	0.001130	0.001000	0.000020
15	$\hat{\beta}_1$	0.000226	0.000209	0.000211	-0.000384	-0.000367	-0.000376
	se($\hat{\beta}_1$)	0.000412	0.000359	0.000352	0.097049	0.109707	0.740317
	se($\hat{b}_1$)	0.000668	0.000372	0.000331	0.363113	0.410477	2.770011
	$\hat{\beta}_2$	0.001796	0.001795	0.001834	0.001903	0.001936	0.001913
	se($\hat{\beta}_2$)	0.000818	0.000871	0.000784	0.000816	0.000834	0.000834
	se($\hat{b}_2$)	0.000510	0.001060	0.000159	0.000362	0.000668	0.000612

Model 6

λ_{time}	Parm	λ_{tmean} 10	11	12	13	14	15
10	$\hat{\beta}_1$	0.000388	0.001246	0.001236	0.000404	0.001249	0.001251
	$se(\hat{\beta}_1)$	0.000391	0.237680	2.236930	0.000364	0.157285	0.343927
	$se(\hat{b}_1)$	0.000419	0.788270	7.419057	0.000234	0.521620	1.140662
	$\hat{\beta}_2$	-0.000281	-0.000296	-0.000296	-0.000283	-0.000296	-0.000295
	$se(\hat{\beta}_2)$	0.001830	0.001977	0.001977	0.001792	0.001977	0.001977
	$se(\hat{b}_2)$	0.004000	0.004537	0.004537	0.003856	0.004536	0.004537
11	$\hat{\beta}_1$	0.000399	0.000384	0.000396	0.000383	0.000396	0.000350
	$se(\hat{\beta}_1)$	0.000366	0.000437	0.000371	0.000411	0.000371	0.000382
	$se(\hat{b}_1)$	0.000258	0.000658	0.000293	0.000531	0.000295	0.000359
	$\hat{\beta}_2$	-0.000267	-0.000267	-0.000267	-0.000265	-0.000268	-0.000401
	$se(\hat{\beta}_2)$	0.001795	0.001871	0.001803	0.001850	0.001803	0.001825
	$se(\hat{b}_2)$	0.003868	0.004154	0.003896	0.004076	0.003899	0.003976
12	$\hat{\beta}_1$	0.001247	0.001245	0.000399	0.001264	0.001236	0.000388
	$se(\hat{\beta}_1)$	0.077443	0.141010	0.000367	0.119366	0.162610	0.000456
	$se(\hat{b}_1)$	0.256774	0.467637	0.000259	0.395846	0.539279	0.000749
	$\hat{\beta}_2$	-0.000296	-0.000272	-0.000267	0.000197	-0.000296	-0.000267
	$se(\hat{\beta}_2)$	0.001977	0.001964	0.001796	0.648557	0.001977	0.001883
	$se(\hat{b}_2)$	0.004537	0.004486	0.003870	2.150891	0.004538	0.004198
13	$\hat{\beta}_1$	0.000384	0.000383	0.000383	0.000398	0.000383	0.000384
	$se(\hat{\beta}_1)$	0.000409	0.000425	0.000413	0.000369	0.000413	0.000399
	$se(\hat{b}_1)$	0.000520	0.000600	0.000539	0.000275	0.000541	0.000468
	$\hat{\beta}_2$	-0.000282	-0.000266	-0.000265	-0.000267	-0.000282	-0.000265
	$se(\hat{\beta}_2)$	0.001849	0.001862	0.001852	0.001799	0.001853	0.001838
	$se(\hat{b}_2)$	0.004073	0.004119	0.004082	0.003883	0.004087	0.004032
14	$\hat{\beta}_1$	0.000397	0.000386	0.000383	0.000383	0.001245	0.000383
	$se(\hat{\beta}_1)$	0.000369	0.000393	0.000419	0.000425	0.126867	0.000427
	$se(\hat{b}_1)$	0.000279	0.000433	0.000571	0.000600	0.420726	0.000610
	$\hat{\beta}_2$	-0.000266	-0.000264	-0.000282	-0.000266	-0.000296	-0.000266
	$se(\hat{\beta}_2)$	0.001800	0.001831	0.001858	0.001862	0.001978	0.001863
	$se(\hat{b}_2)$	0.003885	0.004003	0.004108	0.004120	0.004538	0.004125
15	$\hat{\beta}_1$	0.001252	0.001270	0.000388	0.000385	0.000383	0.000398
	$se(\hat{\beta}_1)$	0.835696	0.193174	0.000457	0.000398	0.000411	0.000368
	$se(\hat{b}_1)$	2.771682	0.640657	0.000750	0.000462	0.000530	0.000272
	$\hat{\beta}_2$	-0.000295	0.000233	-0.000268	-0.000265	-0.000265	-0.000267
	$se(\hat{\beta}_2)$	0.001977	0.495179	0.001884	0.001837	0.001850	0.001798
	$se(\hat{b}_2)$	0.004536	1.642157	0.004202	0.004027	0.004075	0.003880

APPENDIX B

Tables of estimates from numbers of knots in smoothing functions

Model 1

k_{time}	Parm	k_{tmean} 5	6	7	8	9	10
	$\hat{\beta}$	0.000019	0.000011	-0.000087	-0.000078	-0.000083	0.000015
21	se($\hat{\beta}$)	0.000353	0.000338	0.125928	0.371204	0.085468	0.000340
	se($\hat{b}$)	0.000655	0.000562	0.487711	1.437665	0.331008	0.000578
	$\hat{\beta}$	-0.000021	0.000079	0.000083	0.000097	-0.000021	0.000088
26	se($\hat{\beta}$)	0.818302	0.000327	0.000299	0.000299	0.167414	0.000302
	se($\hat{b}$)	3.169274	0.000503	0.000300	0.000315	0.648388	0.000334
	$\hat{\beta}$	0.000025	0.000012	0.000105	0.000115	0.000104	0.000023
31	se($\hat{\beta}$)	0.125015	0.490935	0.000286	0.000291	0.000314	0.176241
	se($\hat{b}$)	0.484175	1.901381	0.000194	0.000252	0.000424	0.682576
	$\hat{\beta}$	0.000119	-0.000024	-0.000020	0.000126	0.000060	0.000123
36	se($\hat{\beta}$)	0.000309	0.412745	3.396110	0.000354	0.128303	0.000287
	se($\hat{b}$)	0.000395	1.598556	13.153060	0.000666	0.496911	0.000202
	$\hat{\beta}$	0.000093	0.000076	0.000004	-0.000075	0.000119	0.000131
41	se($\hat{\beta}$)	0.134636	0.386299	0.206967	0.123937	0.000328	0.000374
	se($\hat{b}$)	0.521436	1.496128	0.801577	0.480001	0.000517	0.000770
	$\hat{\beta}$	0.000123	0.000104	0.000117	0.000133	0.000116	0.000018
46	se($\hat{\beta}$)	0.000339	0.000300	0.000304	0.000373	0.000330	0.111745
	se($\hat{b}$)	0.000584	0.000324	0.000362	0.000769	0.000529	0.432782

Model 2

k_{time}	Parm	k_{tmean} 5	6	7	8	9	10
	$\hat{\beta}_1$	0.000015	-0.000573	-0.000549	0.000067	0.000333	0.000330
	se($\hat{\beta}_1$)	0.000367	0.066817	0.149416	0.000394	0.130682	0.176493
	se($\hat{b}_1$)	0.000476	0.258765	0.578680	0.000611	0.506117	0.683545
	$\hat{\beta}_2$	-0.000069	-0.000088	-0.000088	-0.000206	-0.000043	-0.000040
21	se($\hat{\beta}_2$)	0.000405	0.000398	0.000400	0.135527	0.000382	0.000377
	se($\hat{b}_2$)	0.000589	0.000540	0.000549	0.524879	0.000435	0.000402
	$\hat{\beta}_3$	0.000119	0.000160	0.000127	-0.001905	-0.002165	-0.002169
	se($\hat{\beta}_3$)	0.000422	0.000317	0.000364	0.112373	0.080010	0.180094
	se($\hat{b}_3$)	0.000851	0.000219	0.000540	0.435200	0.309857	0.697492
	$\hat{\beta}_1$	0.000106	0.000646	0.000135	0.000660	0.000088	0.000429
	se($\hat{\beta}_1$)	0.000372	0.076514	0.000367	0.175791	0.000365	0.122638
	se($\hat{b}_1$)	0.000518	0.296311	0.000456	0.680826	0.000478	0.474962
	$\hat{\beta}_2$	-0.000064	-0.000549	-0.000179	-0.000545	-0.000065	-0.000026
26	se($\hat{\beta}_2$)	0.000412	0.123011	0.219379	0.058361	0.000399	0.000379
	se($\hat{b}_2$)	0.000631	0.476398	0.849642	0.225983	0.000572	0.000418
	$\hat{\beta}_3$	0.000116	-0.001921	-0.001921	-0.001919	0.000164	-0.002182
	se($\hat{\beta}_3$)	0.000403	0.104139	0.117528	0.118494	0.000315	0.413553
	se($\hat{b}_3$)	0.000757	0.403309	0.455166	0.458907	0.000216	1.601680
	$\hat{\beta}_1$	0.000172	0.000666	0.000675	0.000130	0.000152	0.000092
	se($\hat{\beta}_1$)	0.000399	0.706279	0.196834	0.000350	0.000373	0.000346
	se($\hat{b}_1$)	0.000649	2.735403	0.762327	0.000365	0.000498	0.000355
	$\hat{\beta}_2$	-0.000161	-0.000529	-0.000530	-0.000008	-0.000151	-0.000035
31	se($\hat{\beta}_2$)	0.105562	0.480065	0.133549	0.000364	0.075206	0.000367
	se($\hat{b}_2$)	0.408819	1.859277	0.517211	0.000328	0.291245	0.000365
	$\hat{\beta}_3$	-0.001914	-0.001916	-0.001911	-0.001974	-0.001915	0.000144
	se($\hat{\beta}_3$)	0.229934	0.449746	0.063516	0.261647	0.201168	0.000332
	se($\hat{b}_3$)	0.890520	1.741852	0.245960	1.013351	0.779109	0.000362
	$\hat{\beta}_1$	0.000775	0.000103	0.000711	0.000506	0.000164	0.000160
	se($\hat{\beta}_1$)	0.116089	0.000370	0.425461	0.259217	0.000370	0.000364
	se($\hat{b}_1$)	0.449596	0.000512	1.647799	1.003937	0.000489	0.000446
	$\hat{\beta}_2$	-0.001878	-0.000027	-0.000507	-0.000016	-0.001483	-0.000114
36	se($\hat{\beta}_2$)	0.099787	0.000367	0.331252	0.000373	0.110981	0.187078
	se($\hat{b}_2$)	0.386451	0.000357	1.282922	0.000377	0.429815	0.724541
	$\hat{\beta}_3$	0.000149	0.000152	-0.001896	-0.002140	0.000193	-0.001898
	se($\hat{\beta}_3$)	0.000471	0.000330	0.326809	0.079032	0.000326	0.209285
	se($\hat{b}_3$)	0.001051	0.000344	1.265721	0.306071	0.000280	0.810545

	$\hat{\beta}_1$	0.000123	0.000147	0.000113	0.000731	0.000149	0.000527
	se($\hat{\beta}_1$)	0.000339	0.000371	0.000352	0.272296	0.000356	0.090195
	se($\hat{b}_1$)	0.000318	0.000480	0.000403	1.054590	0.000389	0.349309
	$\hat{\beta}_2$	-0.000035	-0.000107	-0.000017	-0.000502	-0.000096	-0.000041
41	se($\hat{\beta}_2$)	0.000379	0.287598	0.000357	0.213788	0.057250	0.000411
	se($\hat{b}_2$)	0.000442	1.113854	0.000279	0.827985	0.221691	0.000599
	$\hat{\beta}_3$	0.000154	-0.001878	0.000135	-0.001871	-0.001877	-0.002106
	se($\hat{\beta}_3$)	0.000350	0.135982	0.000374	0.206086	0.189658	0.159508
	se($\hat{b}_3$)	0.000470	0.526641	0.000614	0.798159	0.734529	0.617761
	$\hat{\beta}_1$	0.000681	0.000650	0.000167	0.000799	0.000142	0.000137
	se($\hat{\beta}_1$)	0.403723	4.400490	0.000379	0.317198	0.000362	0.000355
	se($\hat{b}_1$)	1.563608	17.043092	0.000540	1.228499	0.000451	0.000399
	$\hat{\beta}_2$	-0.000510	-0.000511	-0.000096	-0.001879	-0.000014	-0.000009
46	se($\hat{\beta}_2$)	0.323911	2.217000	0.125513	0.327453	0.000397	0.000377
	se($\hat{b}_2$)	1.254492	8.586373	0.486095	1.268215	0.000537	0.000419
	$\hat{\beta}_3$	-0.001879	-0.001905	-0.001869	0.000184	-0.001896	-0.001895
	se($\hat{\beta}_3$)	0.157244	2.212080	0.119002	0.000372	0.175126	0.115730
	se($\hat{b}_3$)	0.608989	8.567339	0.460875	0.000562	0.678254	0.448210

Model 3

k_{time}	Parm	k_{tmean} 5	6	7	8	9	10
21	$\hat{\beta}_1$	0.000114	0.000078	0.000079	0.000091	0.000118	0.000090
	se($\hat{\beta}_1$)	0.082894	0.083943	0.081465	0.106514	0.082598	0.079446
	se($\hat{b}_1$)	0.321039	0.325100	0.315502	0.412520	0.319891	0.307684
	$\hat{\beta}_2$	-0.000007	-0.000005	-0.000005	-0.000004	-0.000005	-0.000004
	se($\hat{\beta}_2$)	0.081643	0.081650	0.081650	0.081654	0.081650	0.081652
	se($\hat{b}_2$)	0.316201	0.316227	0.316227	0.316244	0.316229	0.316237
26	$\hat{\beta}_1$	0.000289	0.000192	0.000177	0.000277	0.000225	0.000254
	se($\hat{\beta}_1$)	0.095123	0.079127	0.084949	0.074416	0.084485	0.081688
	se($\hat{b}_1$)	0.368402	0.306446	0.328994	0.288201	0.327197	0.316365
	$\hat{\beta}_2$	-0.000004	-0.000005	-0.000006	-0.000004	-0.000004	-0.000004
	se($\hat{\beta}_2$)	0.081641	0.081649	0.081651	0.081648	0.081650	0.081652
	se($\hat{b}_2$)	0.316195	0.316226	0.316236	0.316220	0.316228	0.316237
31	$\hat{\beta}_1$	0.000262	0.000179	0.000196	0.000203	0.000232	0.000230
	se($\hat{\beta}_1$)	0.082433	0.078968	0.088258	0.087620	0.083747	0.079721
	se($\hat{b}_1$)	0.319252	0.305832	0.341813	0.339343	0.324342	0.308747
	$\hat{\beta}_2$	-0.000006	-0.000007	-0.000005	-0.000004	-0.000005	-0.000005
	se($\hat{\beta}_2$)	0.081650	0.081649	0.081648	0.081650	0.081652	0.081651
	se($\hat{b}_2$)	0.316228	0.316226	0.316223	0.316228	0.316236	0.316234
36	$\hat{\beta}_1$	0.000289	0.000192	0.000209	0.000277	0.000225	0.000254
	se($\hat{\beta}_1$)	0.095123	0.079127	0.079955	0.074416	0.084485	0.081688
	se($\hat{b}_1$)	0.368402	0.306446	0.309653	0.288201	0.327197	0.316365
	$\hat{\beta}_2$	-0.000004	-0.000005	-0.000004	-0.000004	-0.000004	-0.000004
	se($\hat{\beta}_2$)	0.081641	0.081649	0.081651	0.081648	0.081650	0.081652
	se($\hat{b}_2$)	0.316195	0.316226	0.316231	0.316220	0.316228	0.316237
41	$\hat{\beta}_1$	0.000314	0.000253	0.000221	0.000297	0.000238	0.000238
	se($\hat{\beta}_1$)	0.082777	0.078146	0.103654	0.081360	0.082587	0.041331
	se($\hat{b}_1$)	0.320584	0.302648	0.401441	0.315097	0.319847	0.160054
	$\hat{\beta}_2$	-0.000004	-0.000003	-0.000003	-0.000003	-0.000003	-0.000002
	se($\hat{\beta}_2$)	0.081651	0.081647	0.081648	0.081650	0.081652	0.081607
	se($\hat{b}_2$)	0.316233	0.316216	0.316221	0.316228	0.316236	0.316061
46	$\hat{\beta}_1$	0.000322	0.000215	0.000227	0.000233	0.000282	0.000235
	se($\hat{\beta}_1$)	0.082036	0.063927	0.080571	0.081140	0.081961	0.084935
	se($\hat{b}_1$)	0.317715	0.247577	0.312038	0.314244	0.317426	0.328942
	$\hat{\beta}_2$	-0.000003	-0.000003	-0.000002	-0.000002	-0.000003	-0.000002
	se($\hat{\beta}_2$)	0.081649	0.081659	0.081650	0.081656	0.081651	0.081661
	se($\hat{b}_2$)	0.316227	0.316263	0.316227	0.316251	0.316234	0.316270

Model 4

k_{time}	Parm	k_{tmean}					
		5	6	7	8	9	10
21	$\hat{\beta}_1$	0.000054	0.000336	0.000029	0.000353	-0.000047	-0.000023
	se($\hat{\beta}_1$)	0.000393	0.399171	0.000395	0.295355	0.167047	0.000445
	se($\hat{b}_1$)	0.000346	1.439229	0.000347	1.064913	0.602286	0.000620
	$\hat{\beta}_2$	0.001190	0.001234	0.001324	0.001209	0.002271	0.002220
	se($\hat{\beta}_2$)	0.000823	0.000829	0.000854	0.000826	0.095360	1.154620
	se($\hat{b}_2$)	0.000641	0.000646	0.000887	0.000619	0.343764	4.163040
26	$\hat{\beta}_1$	0.000416	0.000399	0.000506	0.000153	0.000395	0.000145
	se($\hat{\beta}_1$)	0.187388	0.096328	0.373232	0.000377	0.346531	0.000392
	se($\hat{b}_1$)	0.675629	0.347301	1.345704	0.000228	1.249432	0.000342
	$\hat{\beta}_2$	0.001156	0.001175	0.001137	0.001137	0.001235	0.001160
	se($\hat{\beta}_2$)	0.000834	0.000838	0.000833	0.000829	0.000866	0.000833
	se($\hat{b}_2$)	0.000677	0.000708	0.000673	0.000676	0.000916	0.000708
31	$\hat{\beta}_1$	0.000053	0.000154	0.000163	0.000168	0.000142	0.000447
	se($\hat{\beta}_1$)	0.000434	0.000434	0.000441	0.000428	0.000396	0.134948
	se($\hat{b}_1$)	0.000580	0.000600	0.000638	0.000571	0.000379	0.486554
	$\hat{\beta}_2$	0.002015	0.001181	0.001224	0.001179	0.001218	0.001209
	se($\hat{\beta}_2$)	0.167224	0.000828	0.000846	0.000835	0.000854	0.000856
	se($\hat{b}_2$)	0.602905	0.000649	0.000795	0.000711	0.000865	0.000834
36	$\hat{\beta}_1$	0.000201	0.000467	0.000151	0.000161	0.000060	0.000163
	se($\hat{\beta}_1$)	0.000446	0.121271	0.000373	0.000406	2.896330	0.000423
	se($\hat{b}_1$)	0.000674	0.437239	0.000210	0.000447	10.442844	0.000545
	$\hat{\beta}_2$	0.001156	0.001168	0.001147	0.001163	0.002194	0.001230
	se($\hat{\beta}_2$)	0.000837	0.000847	0.000837	0.000837	10.562100	0.000865
	se($\hat{b}_2$)	0.000714	0.000763	0.000727	0.000725	38.081623	0.000934
41	$\hat{\beta}_1$	0.000158	-0.000071	0.000158	0.000473	0.000063	0.000064
	se($\hat{\beta}_1$)	0.000404	0.103523	0.000396	0.131680	0.255017	0.062080
	se($\hat{b}_1$)	0.000439	0.373244	0.000388	0.474769	0.919471	0.223808
	$\hat{\beta}_2$	0.001195	0.002340	0.001139	0.001208	0.002217	0.002217
	se($\hat{\beta}_2$)	0.000867	0.122648	0.000841	0.000880	0.650972	0.129046
	se($\hat{b}_2$)	0.000941	0.442165	0.000743	0.000987	2.347107	0.465236
46	$\hat{\beta}_1$	0.000208	0.000364	0.000168	0.000153	0.000175	0.000496
	se($\hat{\beta}_1$)	0.000465	0.175408	0.000430	0.000382	0.000434	0.308258
	se($\hat{b}_1$)	0.000771	0.632436	0.000583	0.000296	0.000607	1.111436
	$\hat{\beta}_2$	0.001184	0.001252	0.001174	0.001116	0.001122	0.001109
	se($\hat{\beta}_2$)	0.000861	0.000852	0.000858	0.000843	0.000837	0.000841
	se($\hat{b}_2$)	0.000890	0.000763	0.000867	0.000764	0.000706	0.000697

Model 5

k_{time}	Parm	k_{tmean}					
		5	6	7	8	9	10
21	$\hat{\beta}_1$	-0.000470	0.000143	0.000122	0.000025	0.000150	-0.000255
	se($\hat{\beta}_1$)	0.089339	0.000424	0.000377	0.000371	0.000432	0.609019
	se($\hat{b}_1$)	0.334263	0.000715	0.000461	0.000356	0.000758	2.278741
	$\hat{\beta}_2$	0.001985	0.001890	0.001817	0.000954	0.001886	0.000895
	se($\hat{\beta}_2$)	0.000878	0.000836	0.000860	0.702301	0.000854	0.555564
	se($\hat{b}_2$)	0.000773	0.000484	0.000722	2.627767	0.000660	2.078726
26	$\hat{\beta}_1$	-0.000383	0.000226	-0.000097	0.000219	0.000250	-0.000391
	se($\hat{\beta}_1$)	0.095490	0.000413	0.287002	0.000392	0.000434	0.104591
	se($\hat{b}_1$)	0.357279	0.000667	1.073858	0.000560	0.000780	0.391332
	$\hat{\beta}_2$	0.001936	0.001809	0.001232	0.001837	0.001902	0.001977
	se($\hat{\beta}_2$)	0.000885	0.000885	1.100490	0.000827	0.000798	0.000812
	se($\hat{b}_2$)	0.000989	0.001065	4.117657	0.000611	0.000368	0.000348
31	$\hat{\beta}_1$	0.000223	-0.000379	0.000227	0.000206	0.000208	-0.000395
	se($\hat{\beta}_1$)	0.000406	4.484990	0.000385	0.000353	0.000356	0.169782
	se($\hat{b}_1$)	0.000635	16.781269	0.000531	0.000335	0.000353	0.635261
	$\hat{\beta}_2$	0.001798	0.001912	0.001855	0.001802	0.001795	0.001795
	se($\hat{\beta}_2$)	0.000816	0.000819	0.000815	0.000803	0.000848	0.000906
	se($\hat{b}_2$)	0.000522	0.000447	0.000650	0.000407	0.000879	0.000993
36	$\hat{\beta}_1$	-0.000020	0.000202	0.000193	-0.000361	0.000188	0.000216
	se($\hat{\beta}_1$)	0.066078	0.000391	0.000349	0.114272	0.000358	0.000415
	se($\hat{b}_1$)	0.247221	0.000544	0.000294	0.427556	0.000348	0.000671
	$\hat{\beta}_2$	0.001371	0.001770	0.001766	0.001914	0.001745	0.001778
	se($\hat{\beta}_2$)	0.187260	0.000858	0.000819	0.000870	0.000808	0.000844
	se($\hat{b}_2$)	0.700649	0.000853	0.000526	0.000901	0.000203	0.000703
41	$\hat{\beta}_1$	-0.000351	0.000223	-0.000126	0.000233	0.000227	0.000226
	se($\hat{\beta}_1$)	0.119205	0.000402	7.243420	0.000421	0.000404	0.000436
	se($\hat{b}_1$)	0.446016	0.000610	27.102546	0.000710	0.000622	0.000773
	$\hat{\beta}_2$	0.001984	0.001898	0.001176	0.001882	0.001907	0.001800
	se($\hat{\beta}_2$)	0.000808	0.000788	0.152673	0.000848	0.000788	0.000906
	se($\hat{b}_2$)	0.000318	0.000128	0.571232	0.000827	0.000185	0.001099
46	$\hat{\beta}_1$	0.000185	-0.000353	0.000235	0.000234	0.000199	-0.000352
	se($\hat{\beta}_1$)	0.000349	0.185156	0.000434	0.000430	0.000383	0.080523
	se($\hat{b}_1$)	0.000294	0.692786	0.000773	0.000757	0.000501	0.301276
	$\hat{\beta}_2$	0.001871	0.001986	0.001910	0.001913	0.001851	0.001973
	se($\hat{\beta}_2$)	0.000816	0.000830	0.000839	0.000790	0.000854	0.000873
	se($\hat{b}_2$)	0.000549	0.000436	0.000760	0.000180	0.000849	0.000790

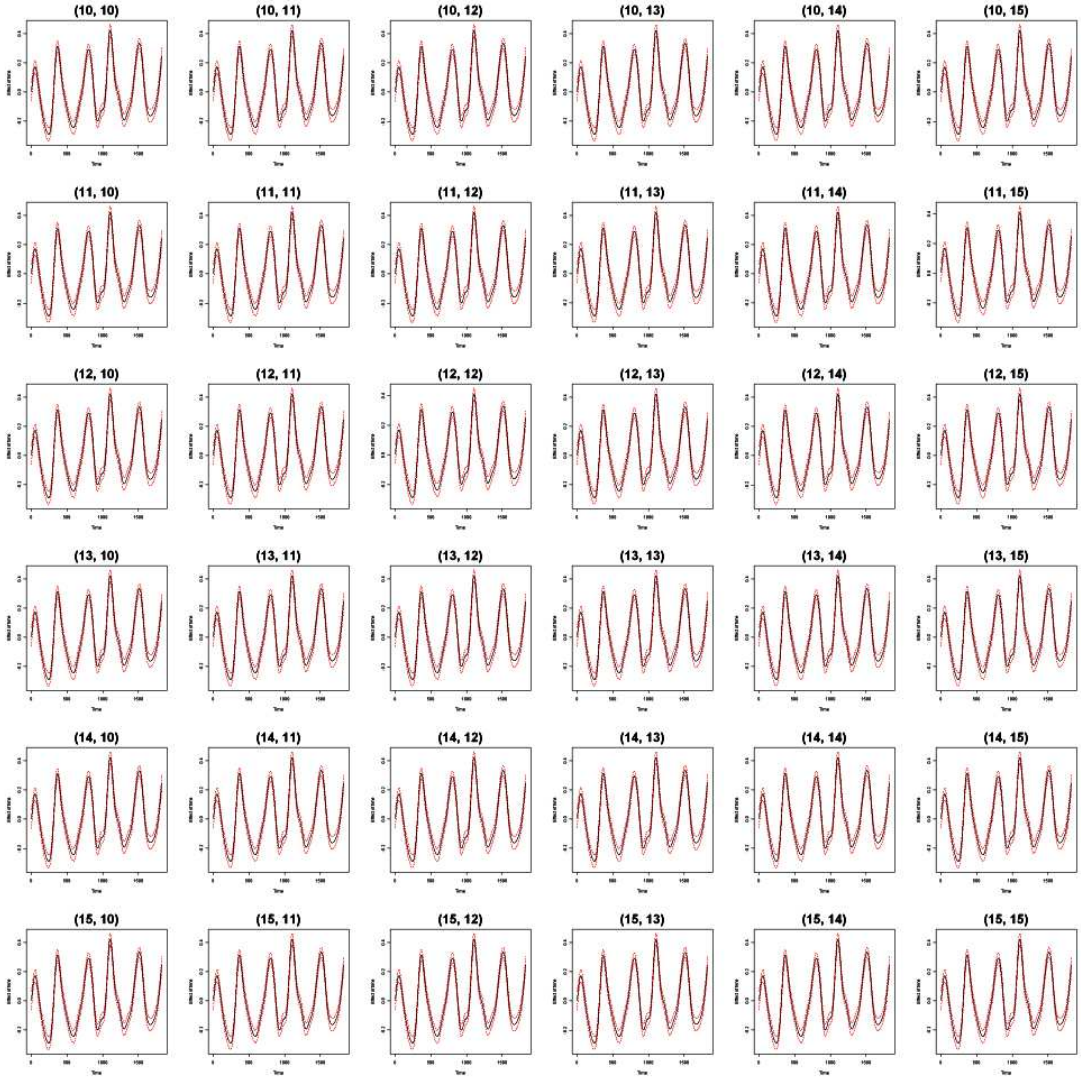
Model 6

k_{time}	Parm	k_{tmean}					
		5	6	7	8	9	10
21	$\hat{\beta}_1$	0.000371	0.001147	0.001155	0.000350	0.000343	0.000351
	se($\hat{\beta}_1$)	0.000357	0.233861	2.586480	0.000363	0.000369	0.000359
	se($\hat{b}_1$)	0.000205	0.775606	8.578397	0.000245	0.000295	0.000206
	$\hat{\beta}_2$	-0.000231	-0.000202	-0.000189	-0.000195	-0.000180	-0.000198
	se($\hat{\beta}_2$)	0.001750	0.001922	0.001922	0.001743	0.001747	0.001730
	se($\hat{b}_2$)	0.003699	0.004335	0.004333	0.003670	0.003684	0.003619
26	$\hat{\beta}_1$	0.001311	0.000449	0.000439	0.001307	0.000465	0.000482
	se($\hat{\beta}_1$)	0.129040	0.000432	0.000361	0.080172	0.000365	0.000368
	se($\hat{b}_1$)	0.427932	0.000652	0.000218	0.265830	0.000278	0.000314
	$\hat{\beta}_2$	-0.000496	-0.000444	-0.000402	-0.000495	-0.000439	-0.000459
	se($\hat{\beta}_2$)	0.002023	0.001906	0.001800	0.002022	0.001830	0.001840
	se($\hat{b}_2$)	0.004715	0.004294	0.003890	0.004713	0.004007	0.004049
31	$\hat{\beta}_1$	0.000445	0.000389	0.000388	0.001308	0.001313	0.001254
	se($\hat{\beta}_1$)	0.000409	0.000456	0.000391	0.167843	0.073811	0.088550
	se($\hat{b}_1$)	0.000561	0.000750	0.000419	0.556638	0.244725	0.293621
	$\hat{\beta}_2$	-0.000357	-0.000284	-0.000281	0.000365	0.000425	-0.000319
	se($\hat{\beta}_2$)	0.001880	0.001892	0.001830	0.905552	1.103570	0.001991
	se($\hat{b}_2$)	0.004198	0.004234	0.004000	3.003283	3.660055	0.004588
36	$\hat{\beta}_1$	0.000420	0.001254	0.000382	0.000425	0.000369	0.000364
	se($\hat{\beta}_1$)	0.000382	0.128542	0.000361	0.000450	0.000391	0.000419
	se($\hat{b}_1$)	0.000414	0.426282	0.000204	0.000748	0.000422	0.000570
	$\hat{\beta}_2$	-0.000263	-0.000230	-0.000181	-0.000288	-0.000191	-0.000191
	se($\hat{\beta}_2$)	0.001842	0.001977	0.001776	0.001900	0.001826	0.001853
	se($\hat{b}_2$)	0.004051	0.004537	0.003791	0.004267	0.003983	0.004083
41	$\hat{\beta}_1$	0.001317	0.001280	0.001255	0.000409	0.000362	0.000366
	se($\hat{\beta}_1$)	0.175718	0.124692	0.123943	0.000373	0.000390	0.000454
	se($\hat{b}_1$)	0.582759	0.413512	0.411027	0.000350	0.000412	0.000735
	$\hat{\beta}_2$	-0.000257	-0.000192	0.000452	-0.000218	-0.000152	-0.000161
	se($\hat{\beta}_2$)	0.001983	0.001960	0.001138	0.001811	0.001805	0.001862
	se($\hat{b}_2$)	0.004558	0.004469	0.000380	0.003927	0.003899	0.004114
46	$\hat{\beta}_1$	0.000391	0.000352	0.000358	0.000367	0.000356	0.001375
	se($\hat{\beta}_1$)	0.000403	0.000445	0.000391	0.000380	0.000422	0.062549
	se($\hat{b}_1$)	0.000525	0.000695	0.000413	0.000361	0.000586	0.207362
	$\hat{\beta}_2$	-0.000242	-0.000179	-0.000153	-0.000179	-0.000163	0.000276
	se($\hat{\beta}_2$)	0.001852	0.001856	0.001807	0.001810	0.001842	1.390010
	se($\hat{b}_2$)	0.004080	0.004088	0.003904	0.003916	0.004035	4.610087

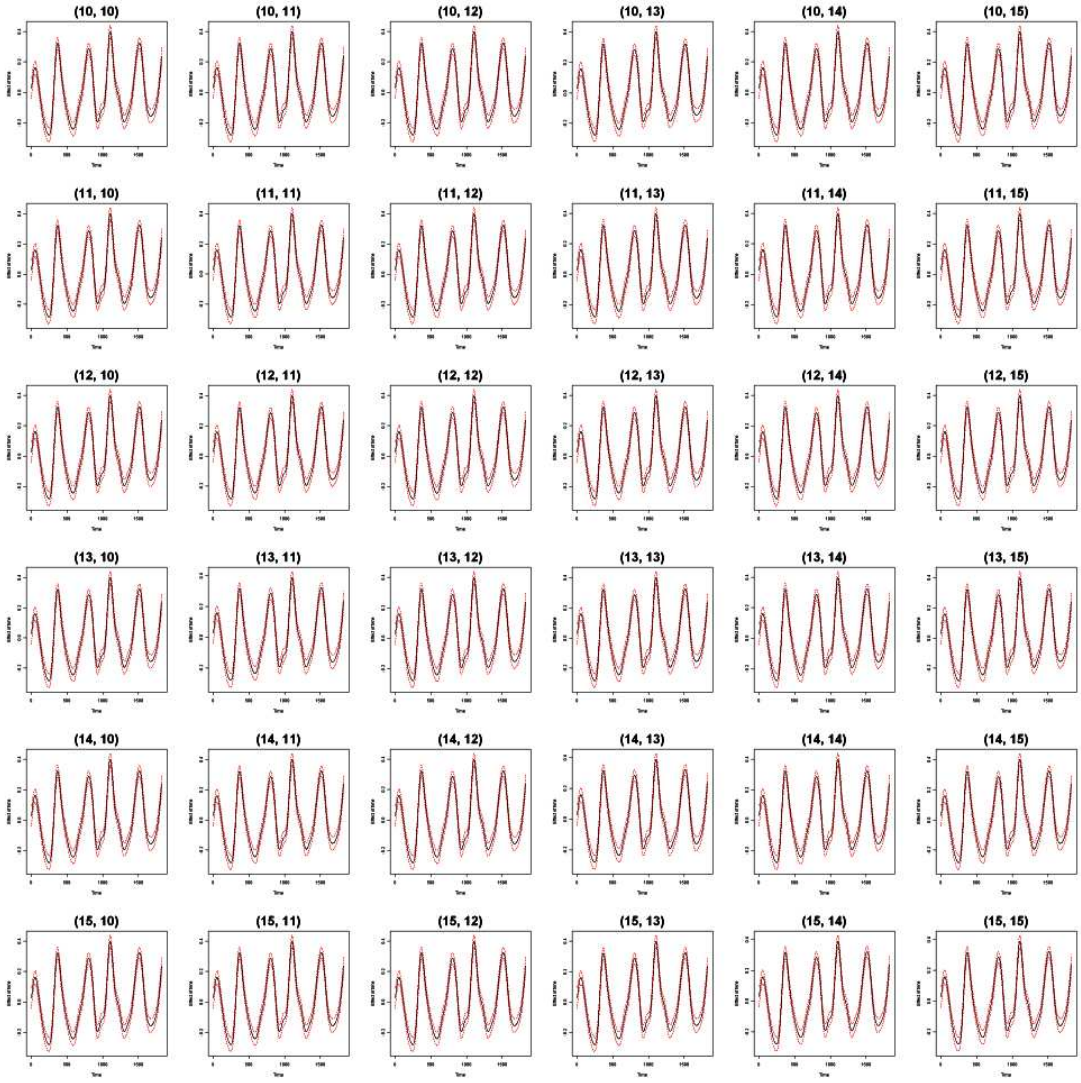
APPENDIX C

Smoothing functions of time and temperature from different starting values of smoothing parameters, where (a, b)=($\lambda_{time,}$ λ_{tmean}).

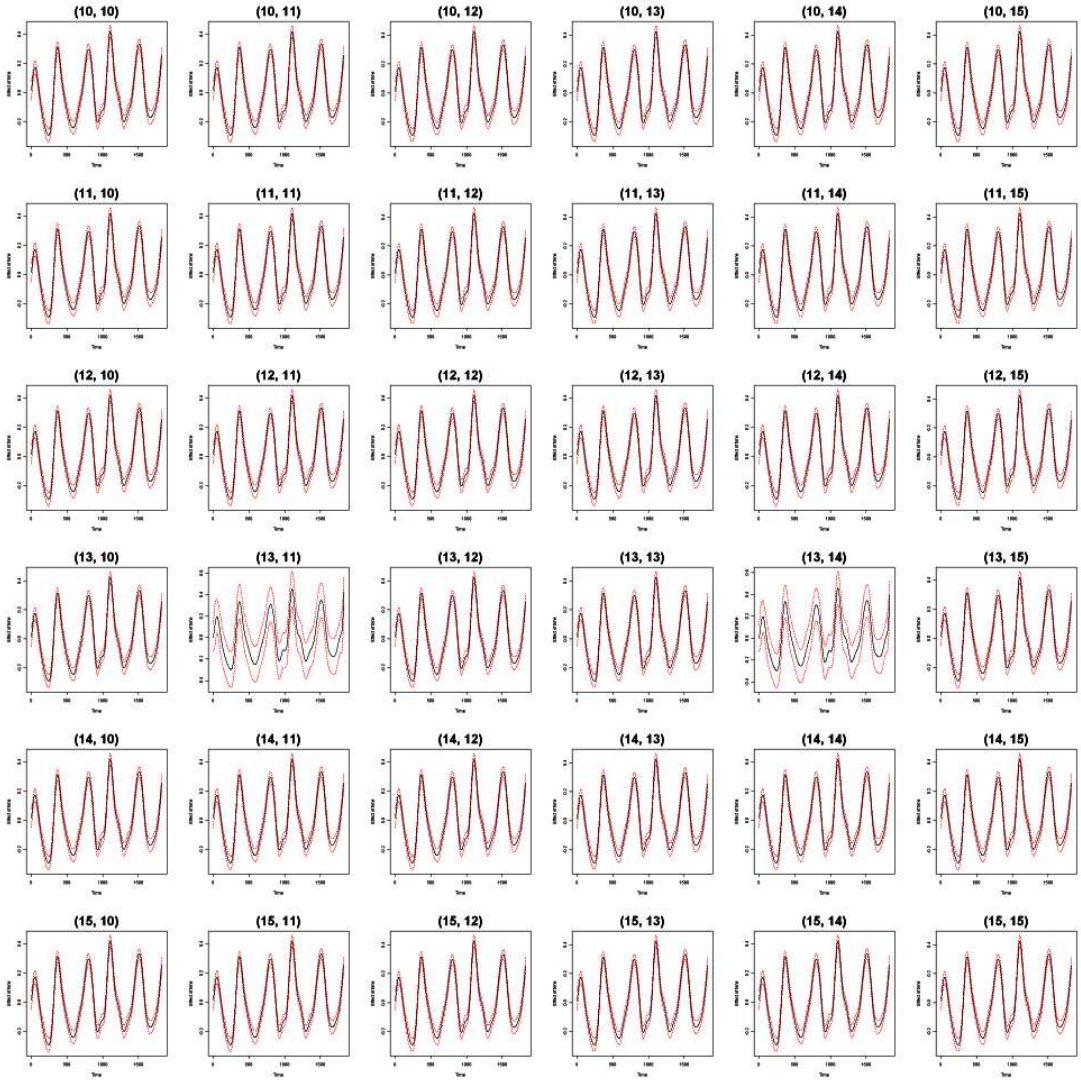
Model 1: Time



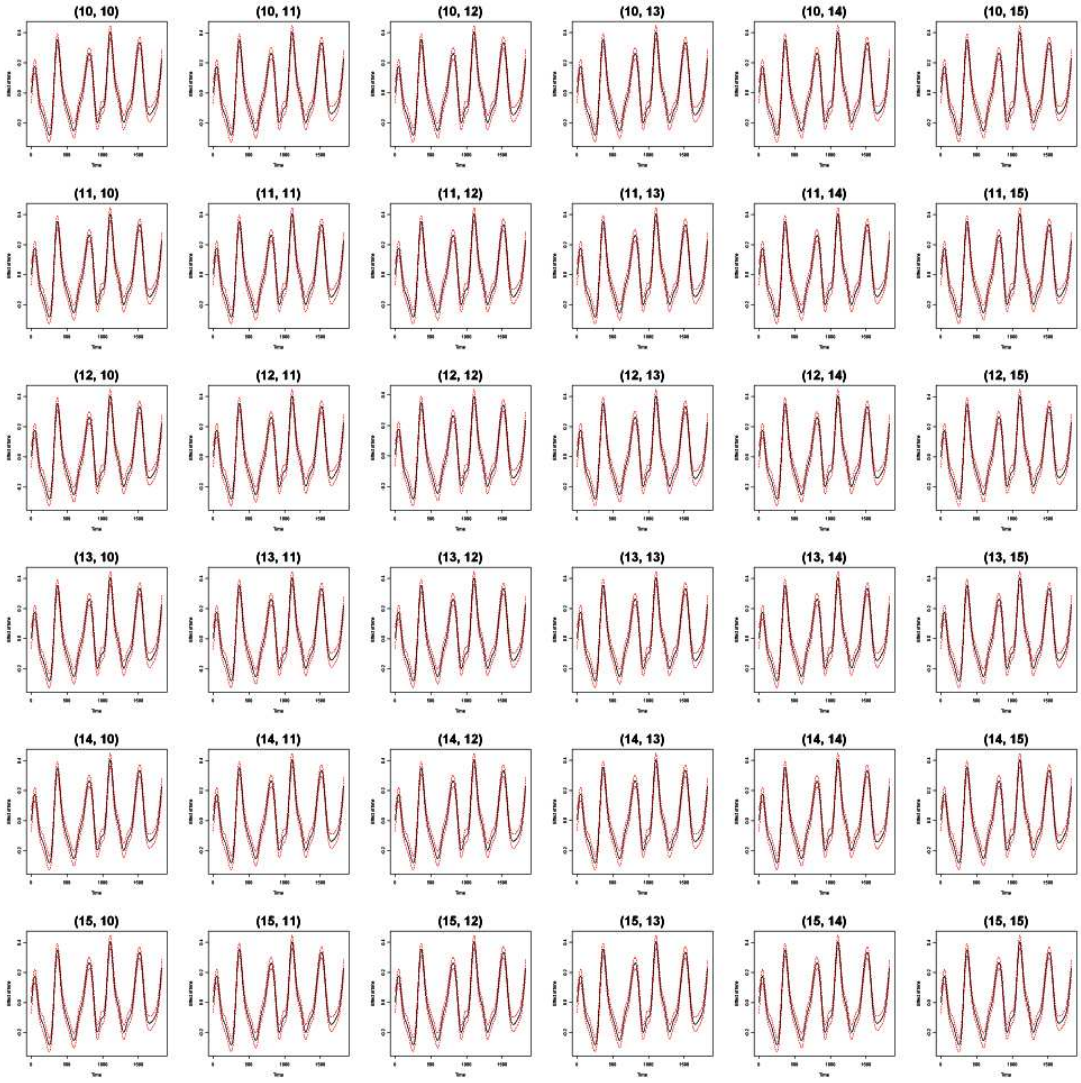
Model 2: Time


(10, 10)
(10, 11)
(10, 12)
(10, 13)
(10, 14)
(10, 15)
(11, 10)
(11, 11)
(11, 12)
(11, 13)
(11, 14)
(11, 15)
(12, 10)
(12, 11)
(12, 12)
(12, 13)
(12, 14)
(12, 15)
(13, 10)
(13, 11)
(13, 12)
(13, 13)
(13, 14)
(13, 15)
(14, 10)
(14, 11)
(14, 12)
(14, 13)
(14, 14)
(14, 15)
(15, 10)
(15, 11)
(15, 12)
(15, 13)
(15, 14)
(15, 15)

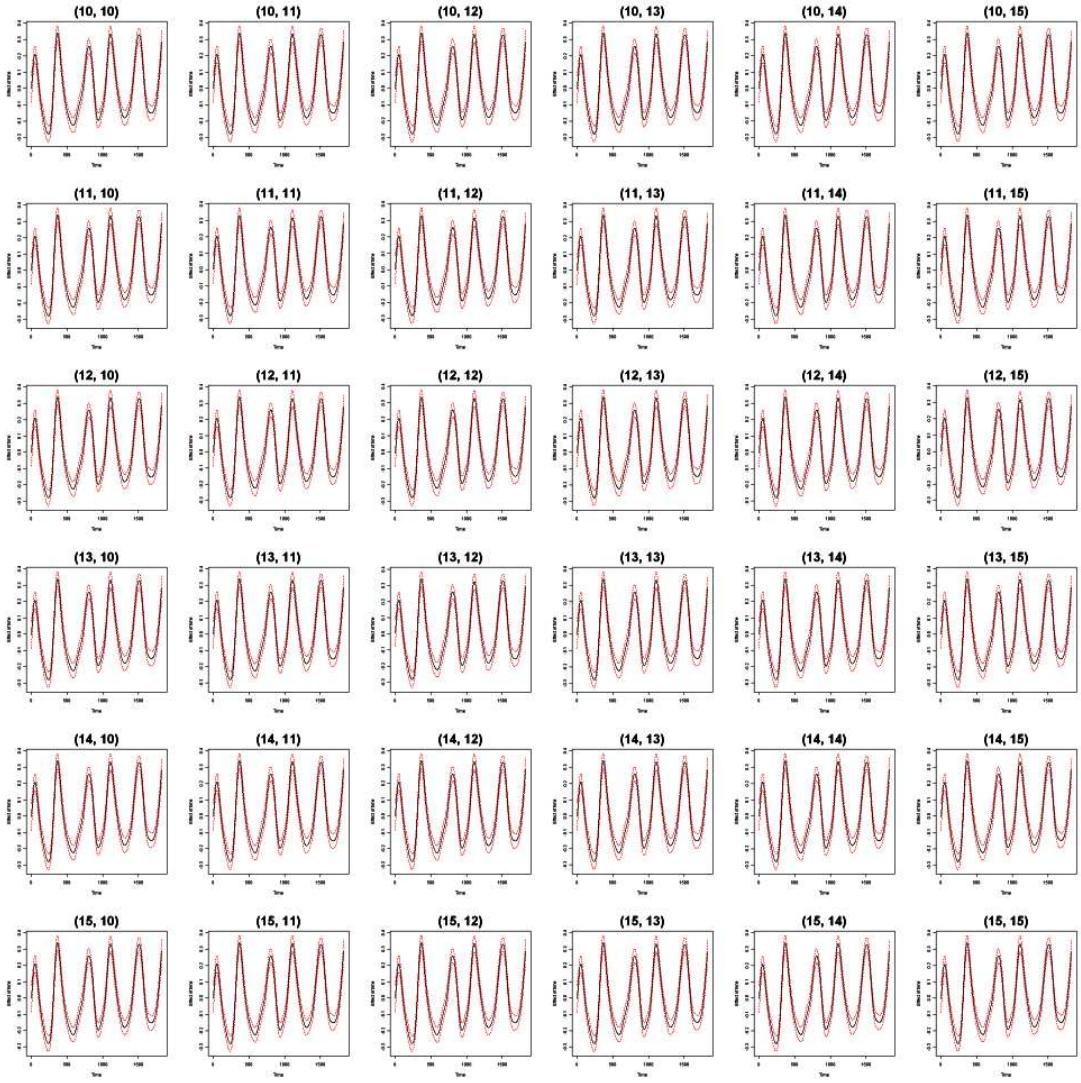
Model 3: Time


(10, 10)
(10, 11)
(10, 12)
(10, 13)
(10, 14)
(10, 15)
(11, 10)
(11, 11)
(11, 12)
(11, 13)
(11, 14)
(11, 15)
(12, 10)
(12, 11)
(12, 12)
(12, 13)
(12, 14)
(12, 15)
(13, 10)
(13, 11)
(13, 12)
(13, 13)
(13, 14)
(13, 15)
(14, 10)
(14, 11)
(14, 12)
(14, 13)
(14, 14)
(14, 15)
(15, 10)
(15, 11)
(15, 12)
(15, 13)
(15, 14)
(15, 15)

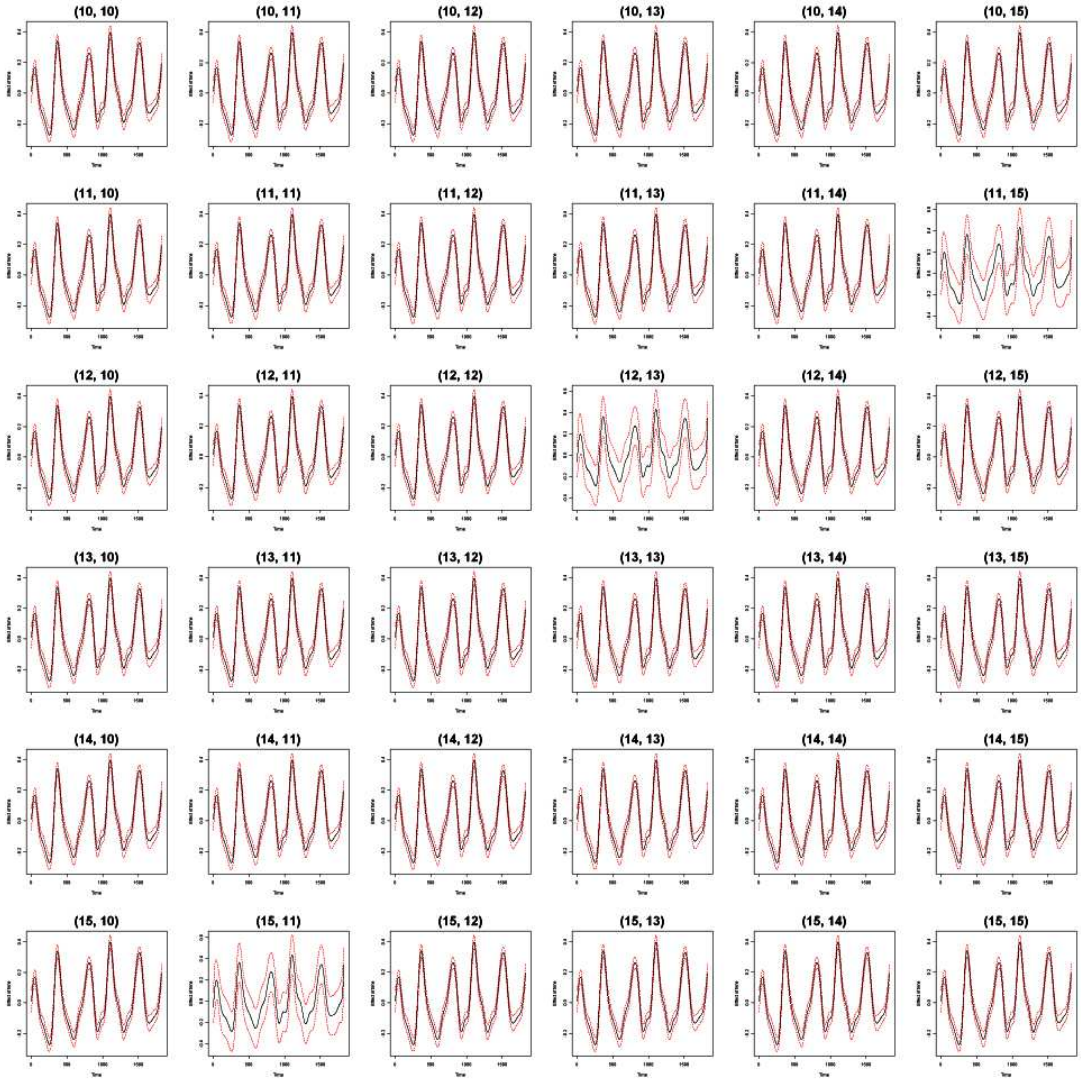
Model 4: Time


(10, 10)
(10, 11)
(10, 12)
(10, 13)
(10, 14)
(10, 15)
(11, 10)
(11, 11)
(11, 12)
(11, 13)
(11, 14)
(11, 15)
(12, 10)
(12, 11)
(12, 12)
(12, 13)
(12, 14)
(12, 15)
(13, 10)
(13, 11)
(13, 12)
(13, 13)
(13, 14)
(13, 15)
(14, 10)
(14, 11)
(14, 12)
(14, 13)
(14, 14)
(14, 15)
(15, 10)
(15, 11)
(15, 12)
(15, 13)
(15, 14)
(15, 15)

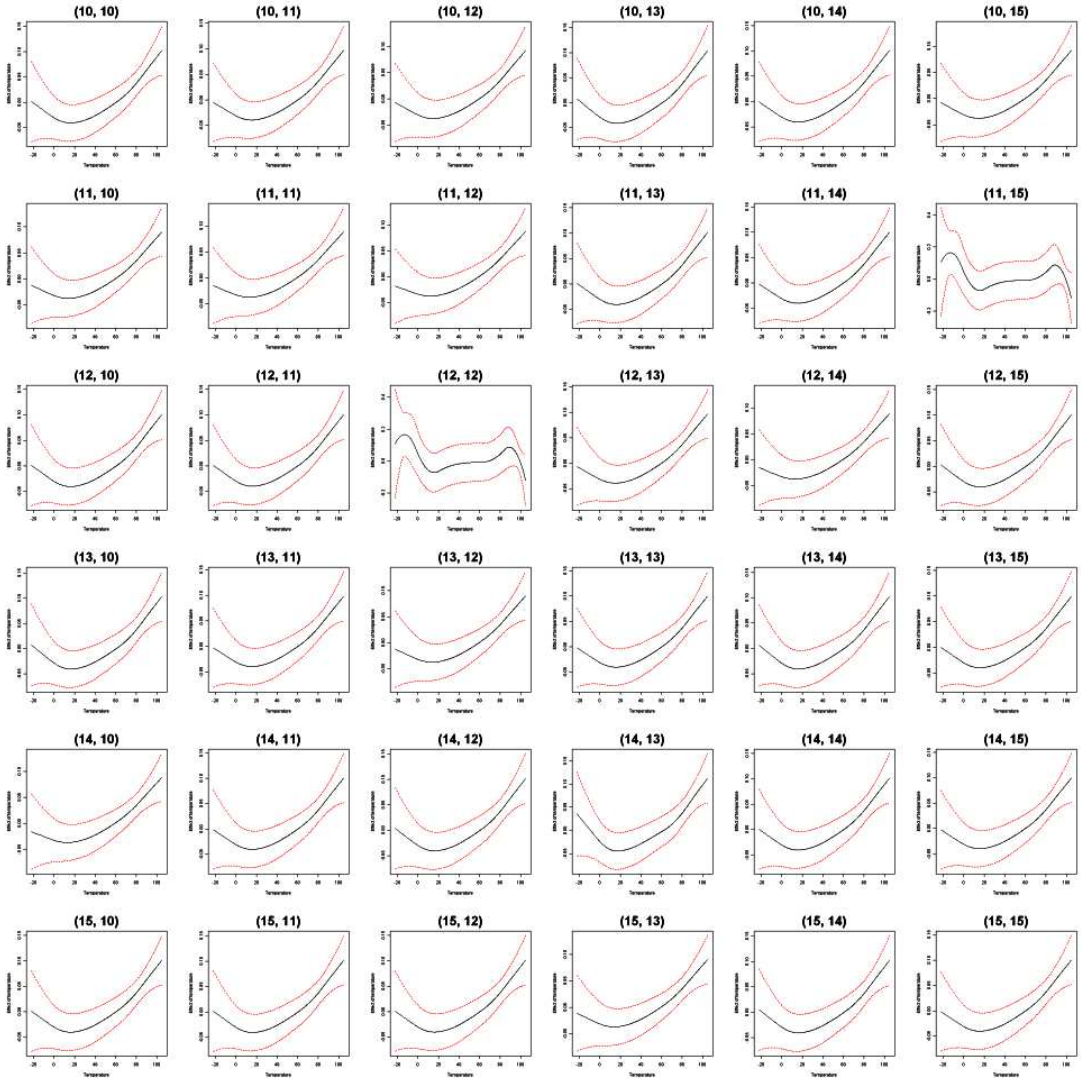
Model 5: Time


(10, 10)
(10, 11)
(10, 12)
(10, 13)
(10, 14)
(10, 15)
(11, 10)
(11, 11)
(11, 12)
(11, 13)
(11, 14)
(11, 15)
(12, 10)
(12, 11)
(12, 12)
(12, 13)
(12, 14)
(12, 15)
(13, 10)
(13, 11)
(13, 12)
(13, 13)
(13, 14)
(13, 15)
(14, 10)
(14, 11)
(14, 12)
(14, 13)
(14, 14)
(14, 15)
(15, 10)
(15, 11)
(15, 12)
(15, 13)
(15, 14)
(15, 15)

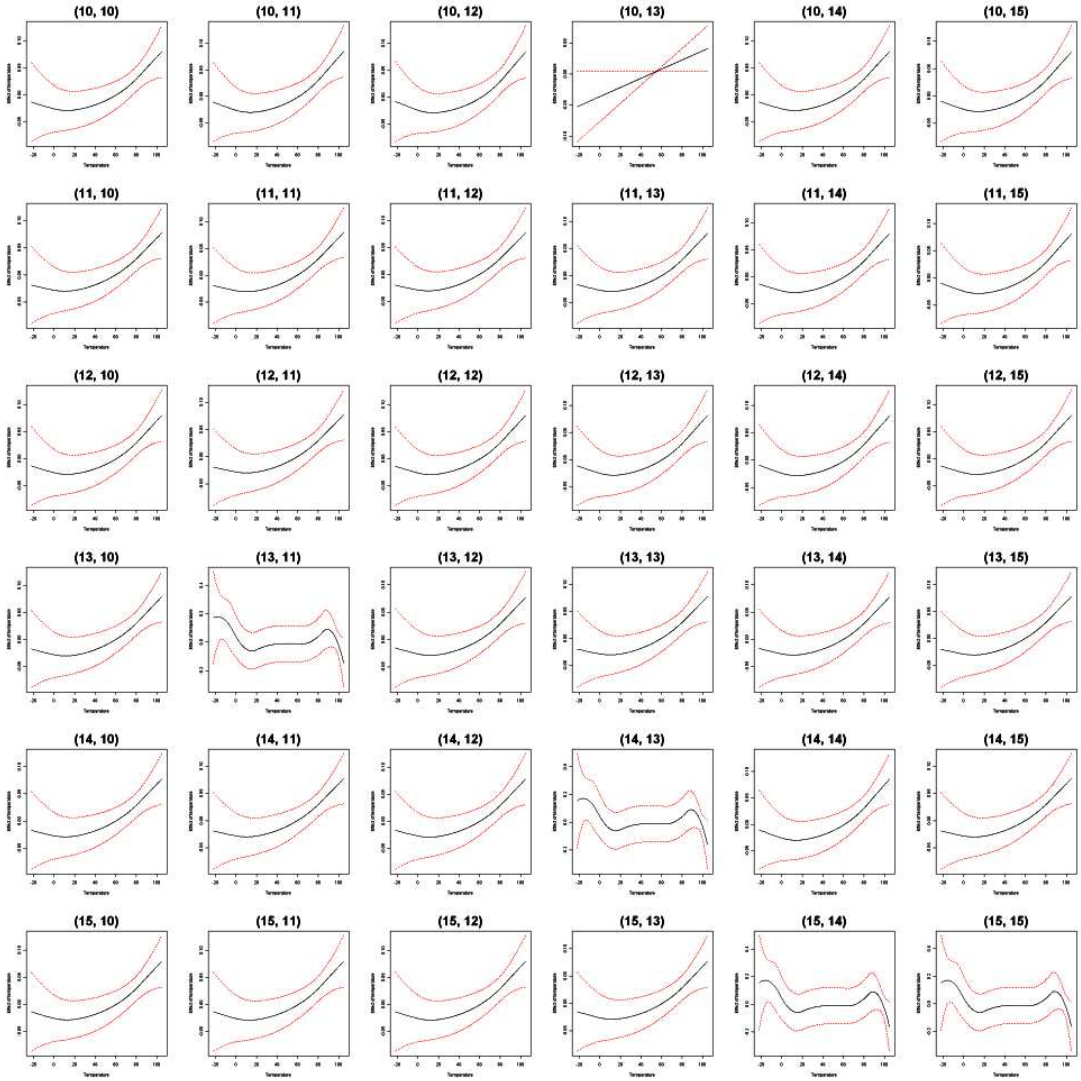
Model 6: Time


(10, 10)
(10, 11)
(10, 12)
(10, 13)
(10, 14)
(10, 15)
(11, 10)
(11, 11)
(11, 12)
(11, 13)
(11, 14)
(11, 15)
(12, 10)
(12, 11)
(12, 12)
(12, 13)
(12, 14)
(12, 15)
(13, 10)
(13, 11)
(13, 12)
(13, 13)
(13, 14)
(13, 15)
(14, 10)
(14, 11)
(14, 12)
(14, 13)
(14, 14)
(14, 15)
(15, 10)
(15, 11)
(15, 12)
(15, 13)
(15, 14)
(15, 15)

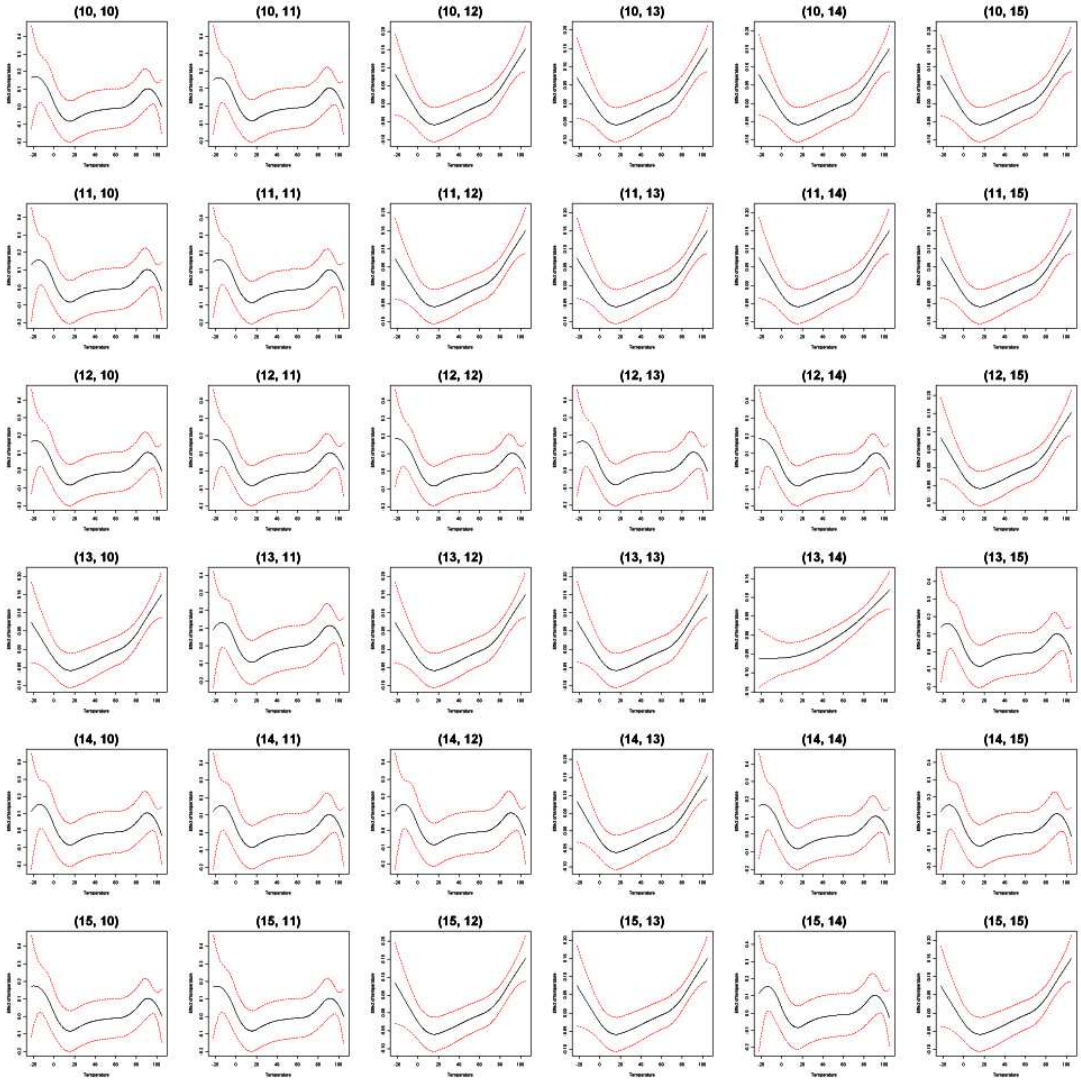
Model 1: Temperature



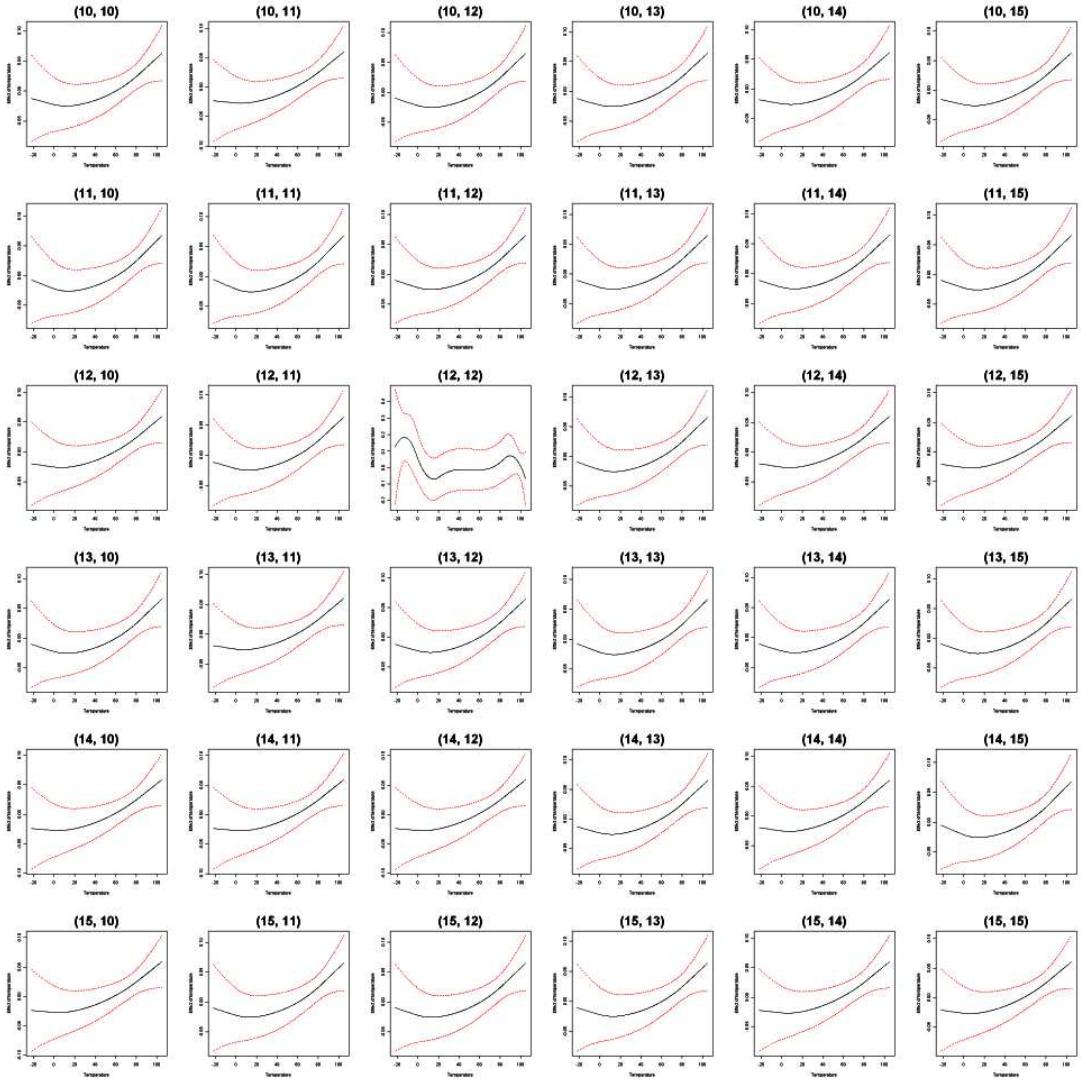
Model 2: Temperature



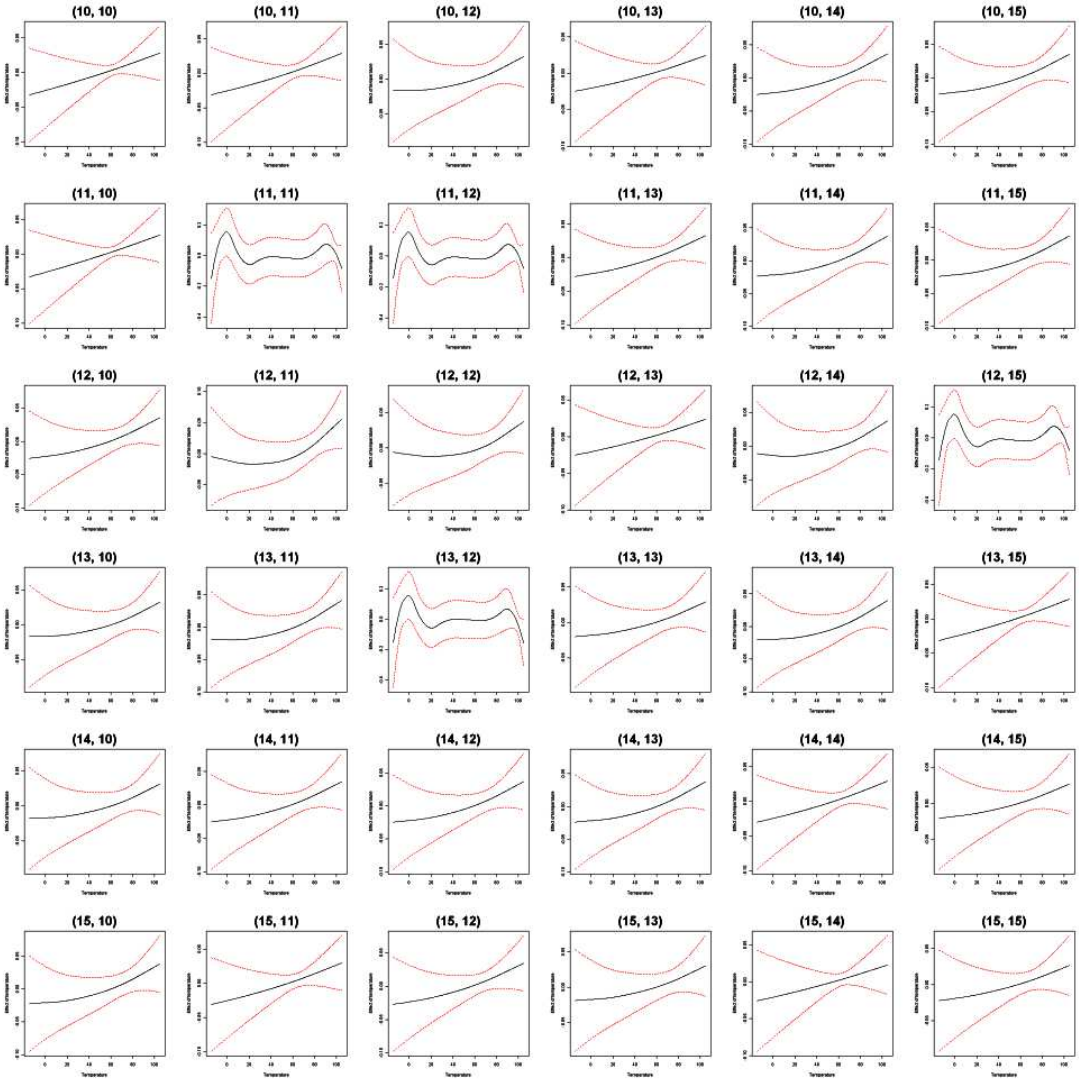
Model 3: Temperature



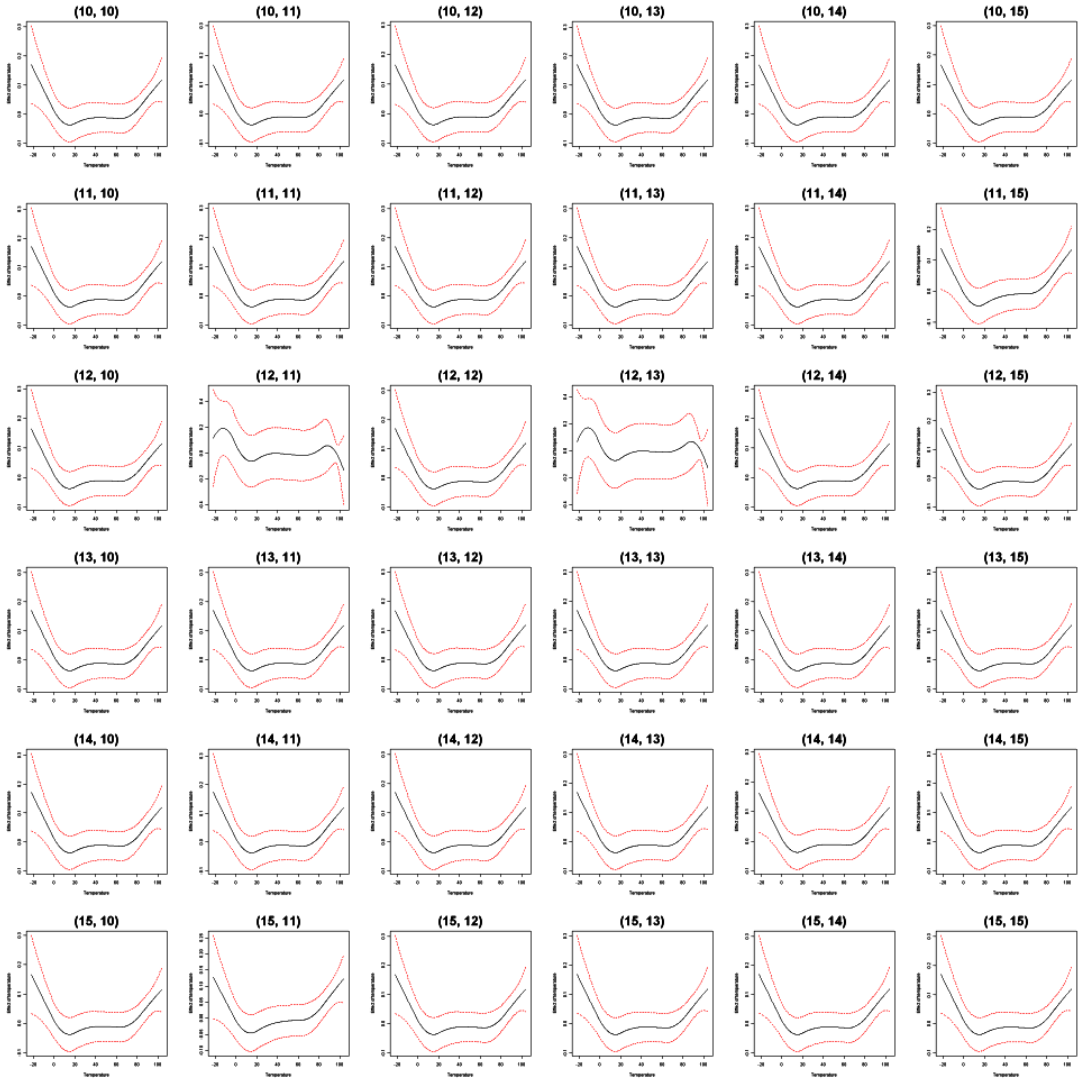
Model 4: Temperature



Model 5: Temperature



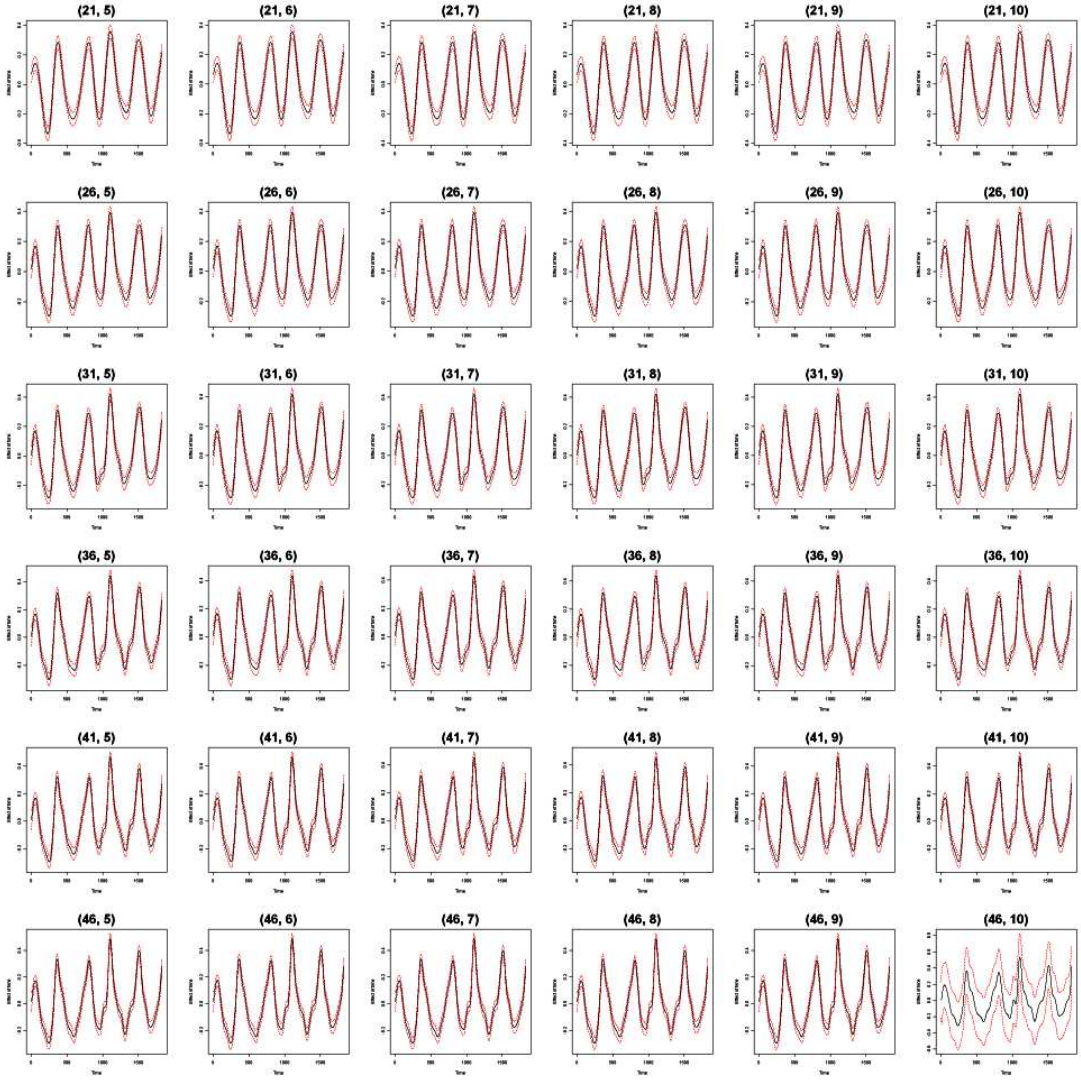
Model 6: Temperature



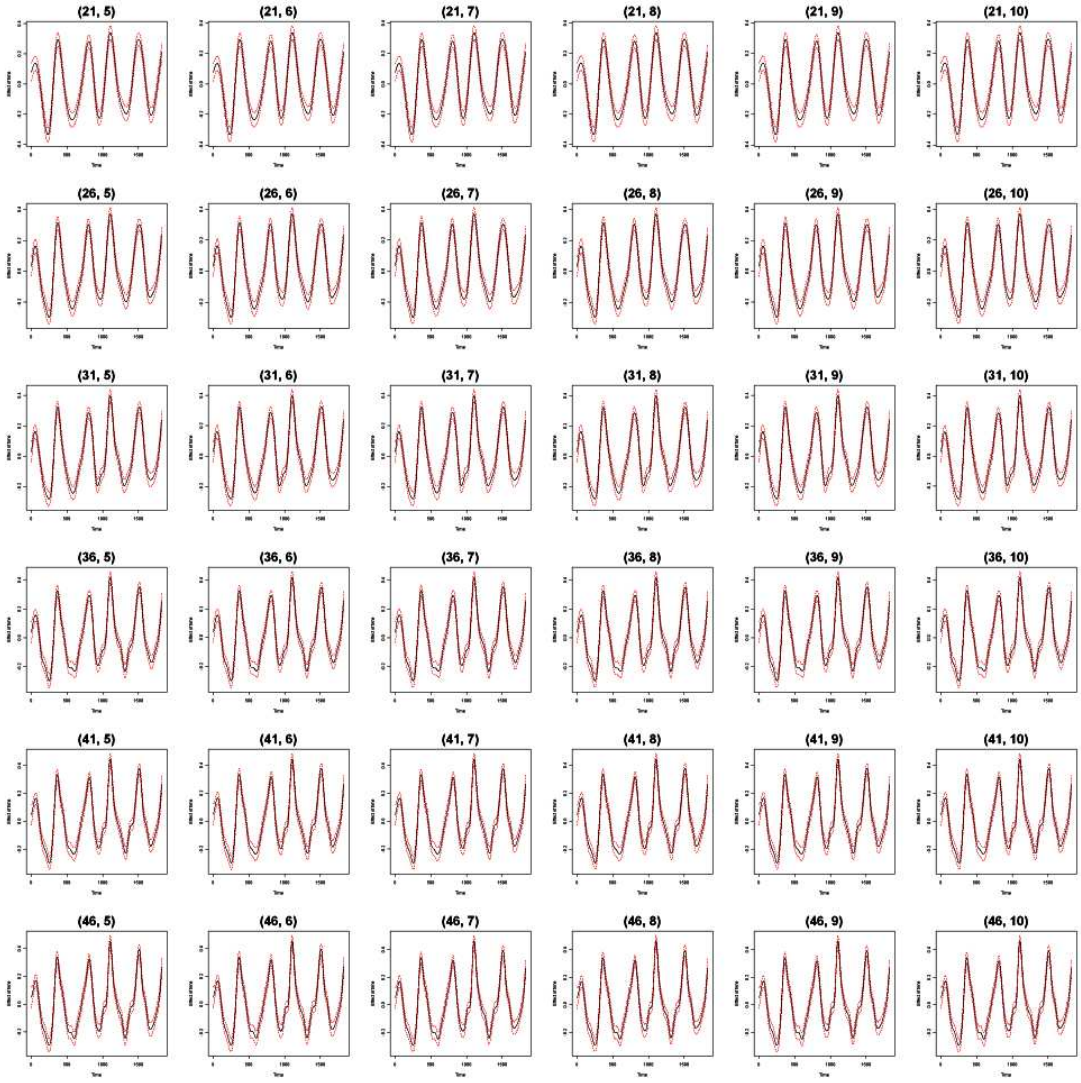
APPENDIX D

Smoothing functions of time and temperature from different numbers of knots, where (a, b)=(k_{time}, k_{tmean}).

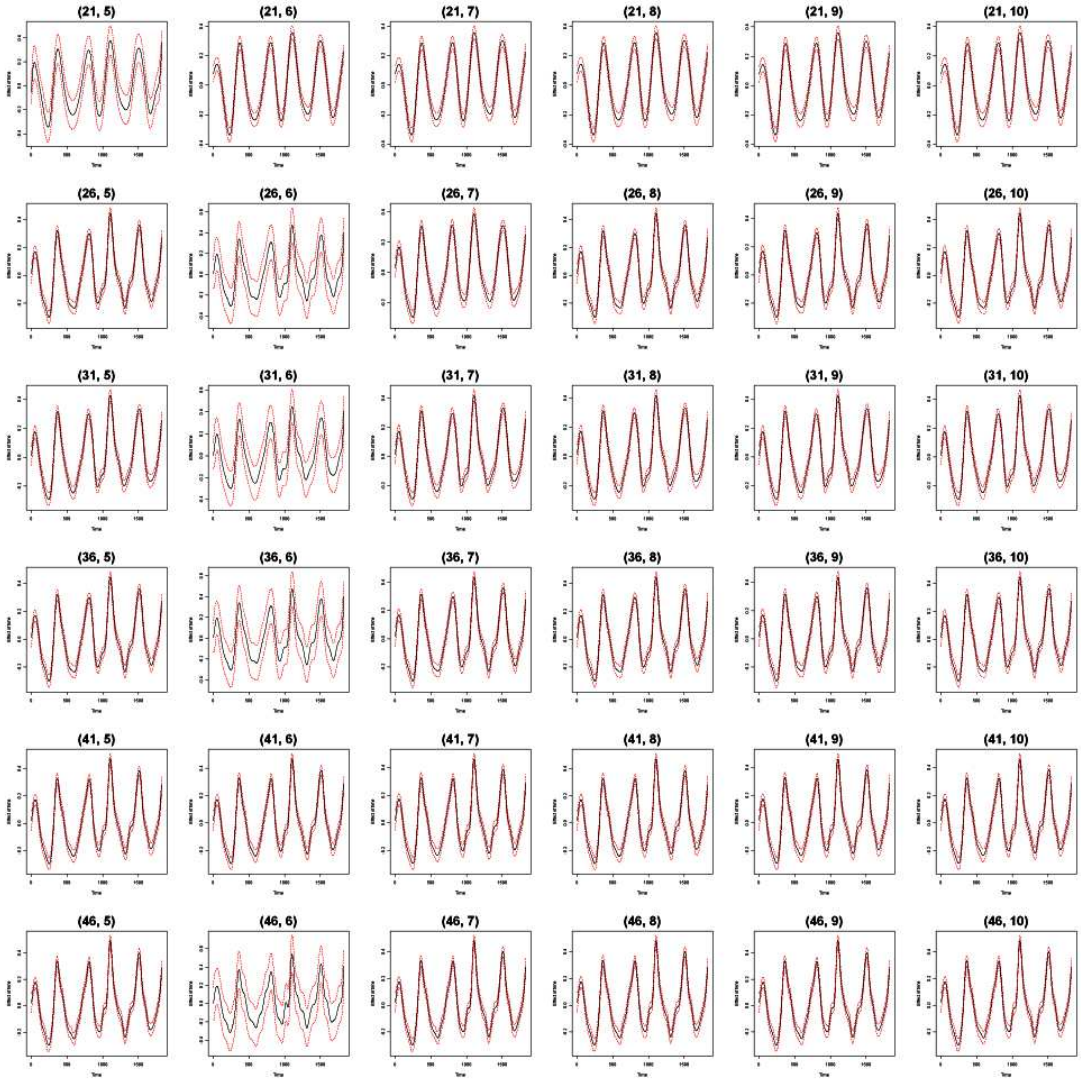
Model 1: Time



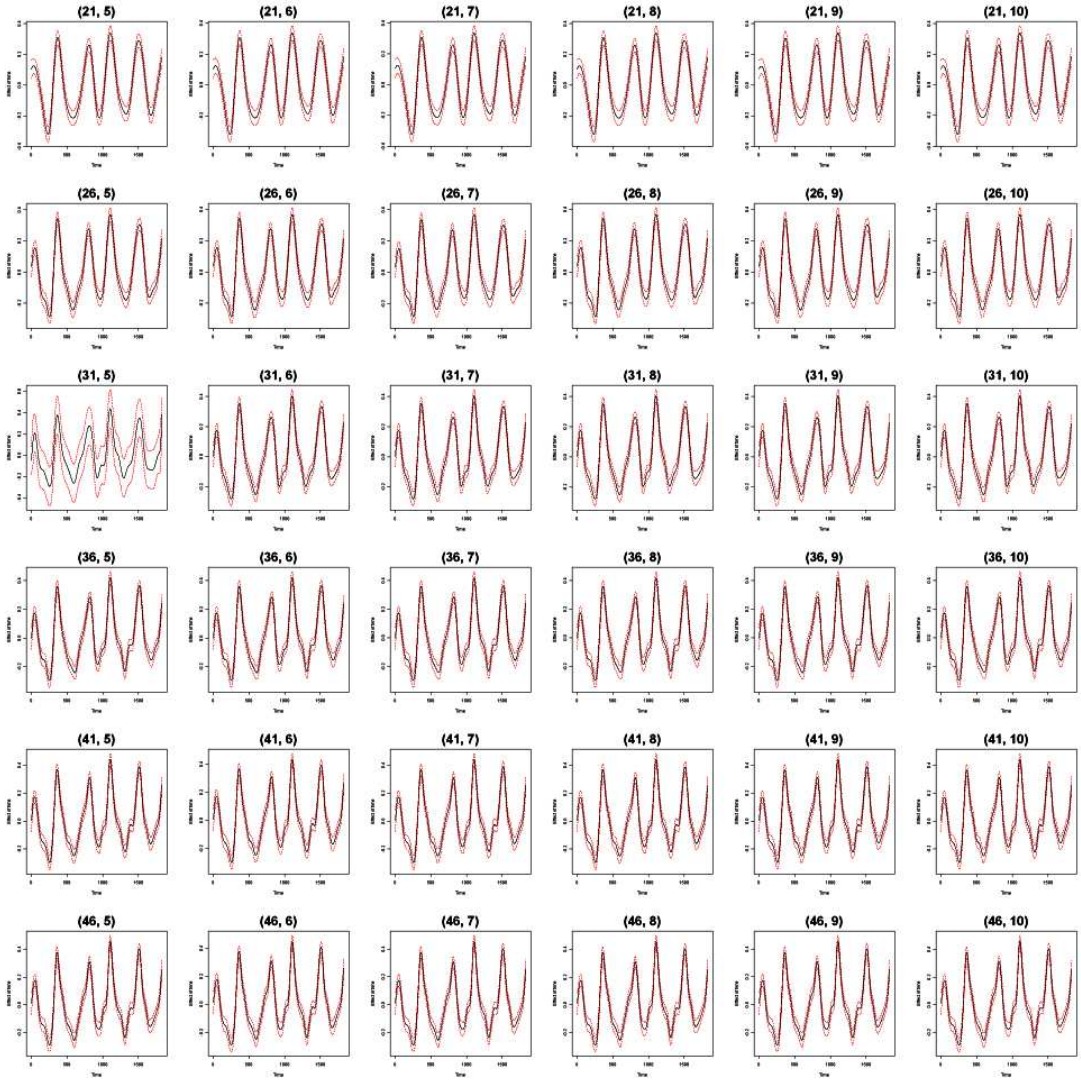
Model 2: Time


(21, 5)
(21, 6)
(21, 7)
(21, 8)
(21, 9)
(21, 10)
(26, 5)
(26, 6)
(26, 7)
(26, 8)
(26, 9)
(26, 10)
(31, 5)
(31, 6)
(31, 7)
(31, 8)
(31, 9)
(31, 10)
(36, 5)
(36, 6)
(36, 7)
(36, 8)
(36, 9)
(36, 10)
(41, 5)
(41, 6)
(41, 7)
(41, 8)
(41, 9)
(41, 10)
(46, 5)
(46, 6)
(46, 7)
(46, 8)
(46, 9)
(46, 10)

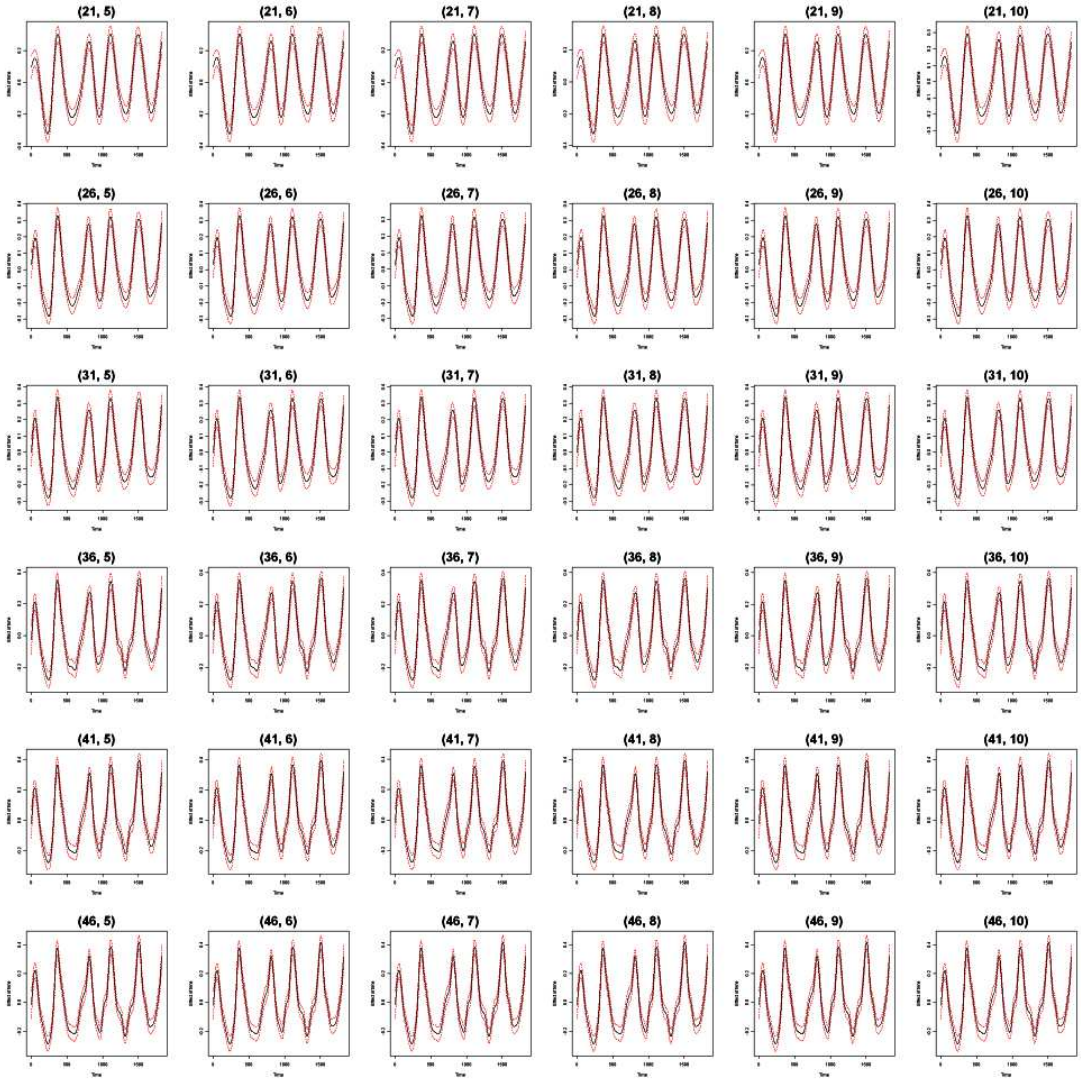
Model 3: Time


(21, 5)
(21, 6)
(21, 7)
(21, 8)
(21, 9)
(21, 10)
(26, 5)
(26, 6)
(26, 7)
(26, 8)
(26, 9)
(26, 10)
(31, 5)
(31, 6)
(31, 7)
(31, 8)
(31, 9)
(31, 10)
(36, 5)
(36, 6)
(36, 7)
(36, 8)
(36, 9)
(36, 10)
(41, 5)
(41, 6)
(41, 7)
(41, 8)
(41, 9)
(41, 10)
(46, 5)
(46, 6)
(46, 7)
(46, 8)
(46, 9)
(46, 10)

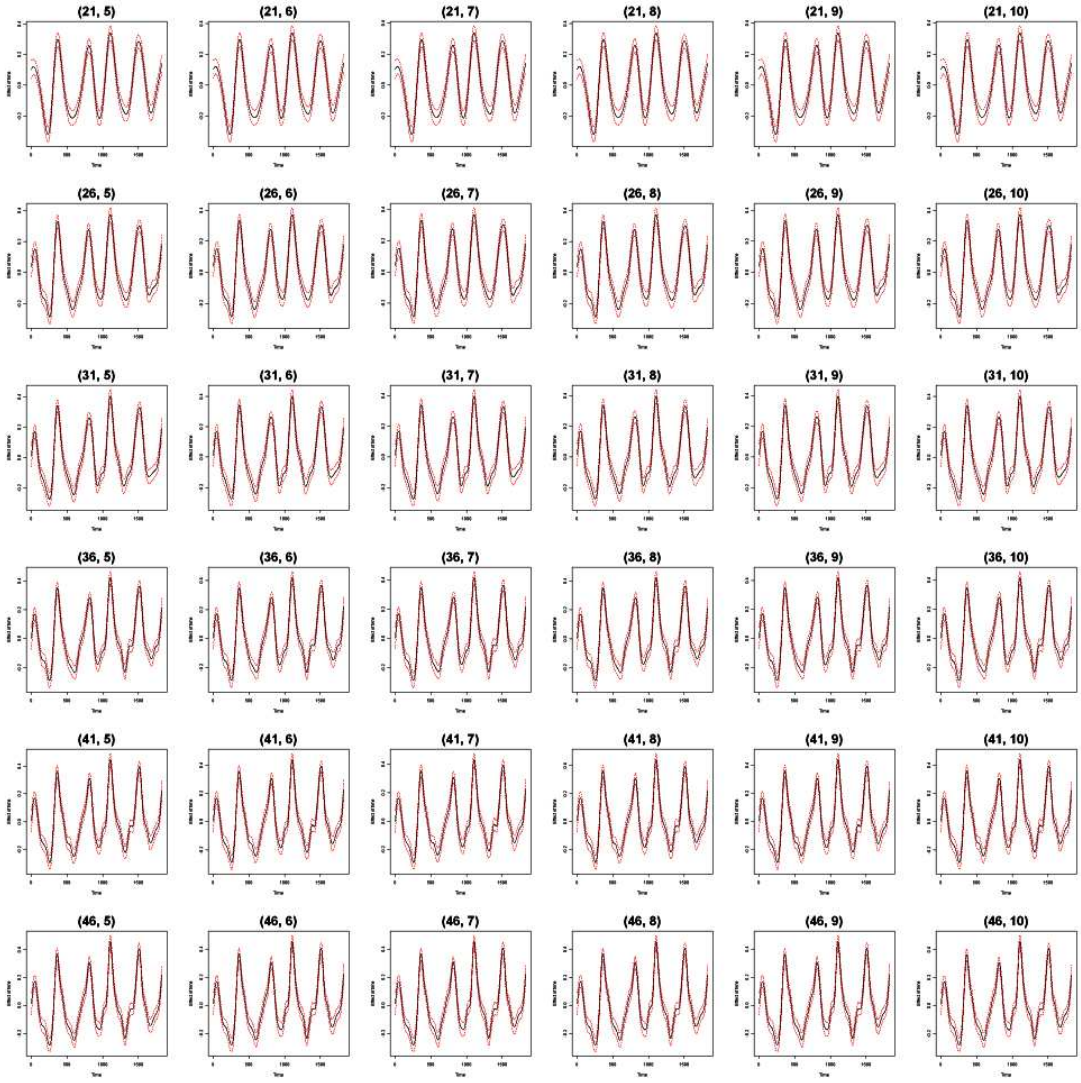
Model 4: Time


(21, 5)
(21, 6)
(21, 7)
(21, 8)
(21, 9)
(21, 10)
(26, 5)
(26, 6)
(26, 7)
(26, 8)
(26, 9)
(26, 10)
(31, 5)
(31, 6)
(31, 7)
(31, 8)
(31, 9)
(31, 10)
(36, 5)
(36, 6)
(36, 7)
(36, 8)
(36, 9)
(36, 10)
(41, 5)
(41, 6)
(41, 7)
(41, 8)
(41, 9)
(41, 10)
(46, 5)
(46, 6)
(46, 7)
(46, 8)
(46, 9)
(46, 10)

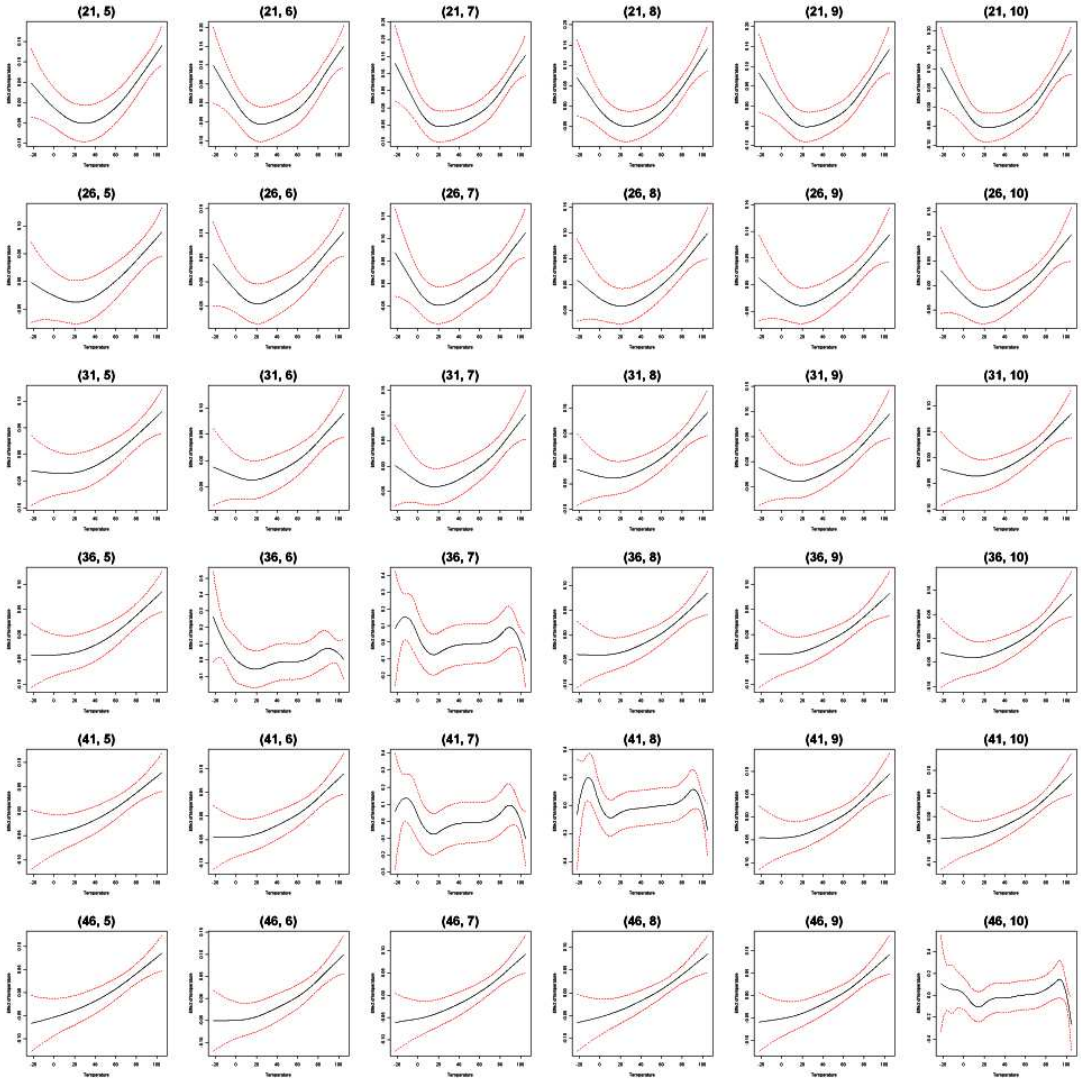
Model 5: Time


(21, 5)
(21, 6)
(21, 7)
(21, 8)
(21, 9)
(21, 10)
(26, 5)
(26, 6)
(26, 7)
(26, 8)
(26, 9)
(26, 10)
(31, 5)
(31, 6)
(31, 7)
(31, 8)
(31, 9)
(31, 10)
(36, 5)
(36, 6)
(36, 7)
(36, 8)
(36, 9)
(36, 10)
(41, 5)
(41, 6)
(41, 7)
(41, 8)
(41, 9)
(41, 10)
(46, 5)
(46, 6)
(46, 7)
(46, 8)
(46, 9)
(46, 10)

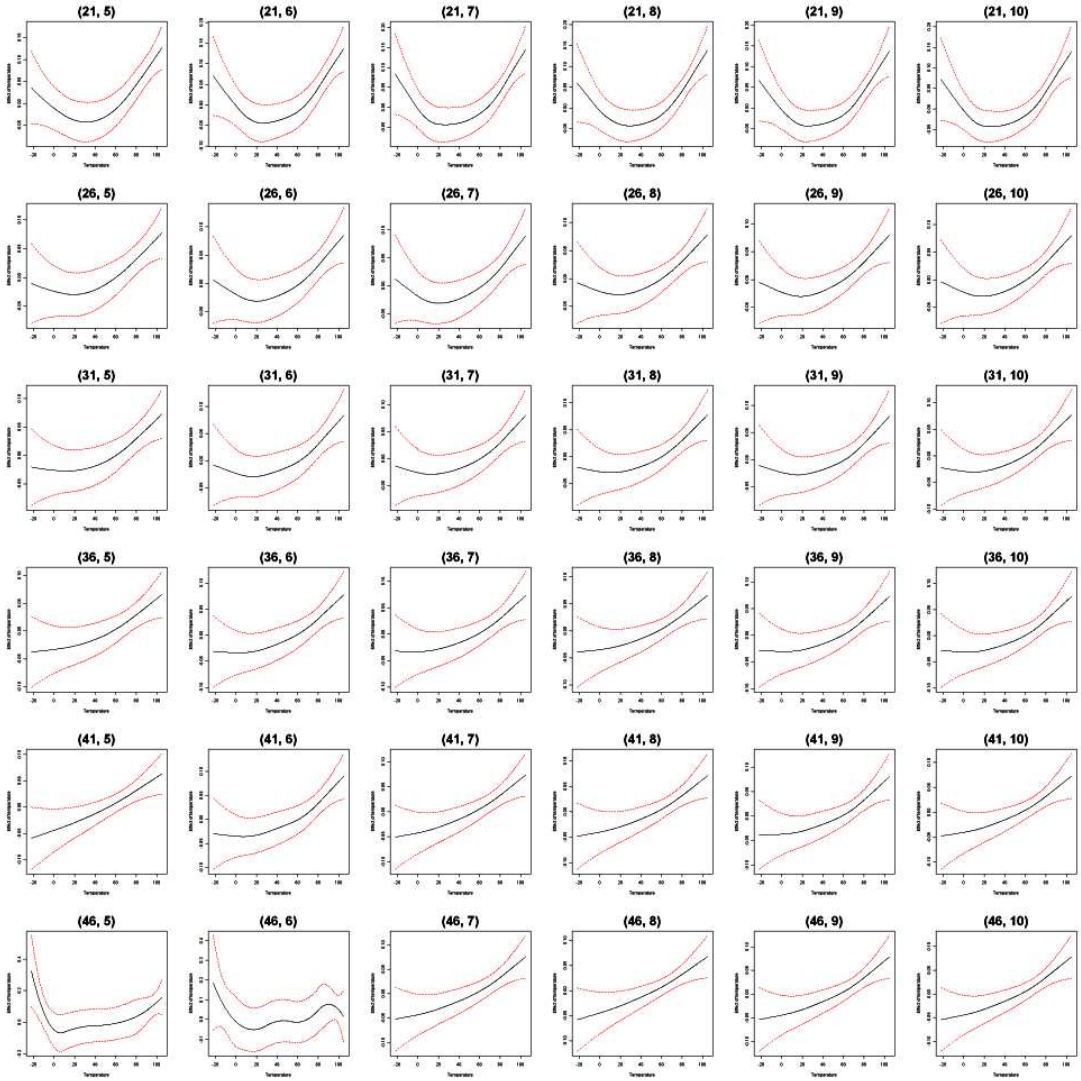
Model 6: Time


(21, 5)
(21, 6)
(21, 7)
(21, 8)
(21, 9)
(21, 10)
(26, 5)
(26, 6)
(26, 7)
(26, 8)
(26, 9)
(26, 10)
(31, 5)
(31, 6)
(31, 7)
(31, 8)
(31, 9)
(31, 10)
(36, 5)
(36, 6)
(36, 7)
(36, 8)
(36, 9)
(36, 10)
(41, 5)
(41, 6)
(41, 7)
(41, 8)
(41, 9)
(41, 10)
(46, 5)
(46, 6)
(46, 7)
(46, 8)
(46, 9)
(46, 10)

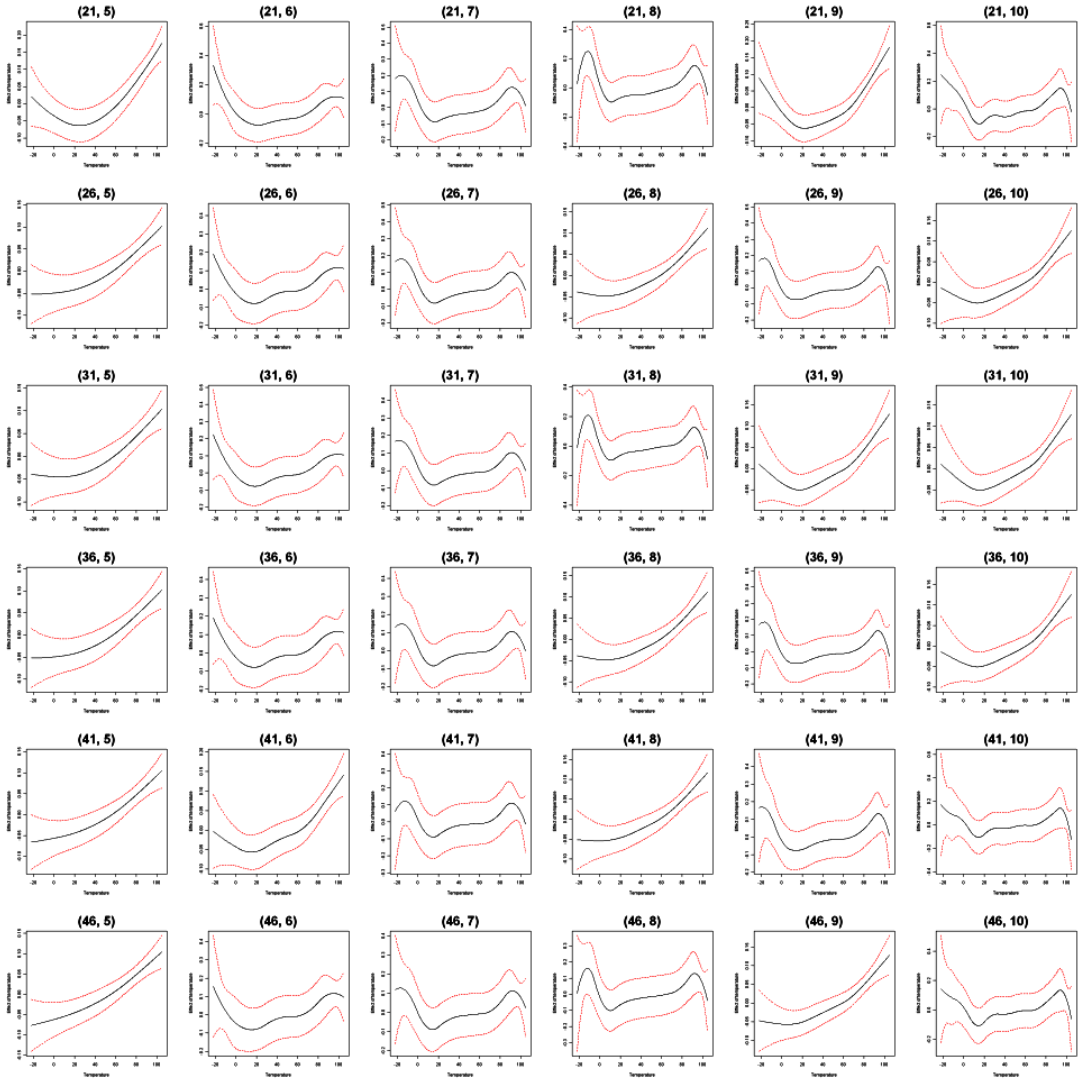
Model 1: Temperature



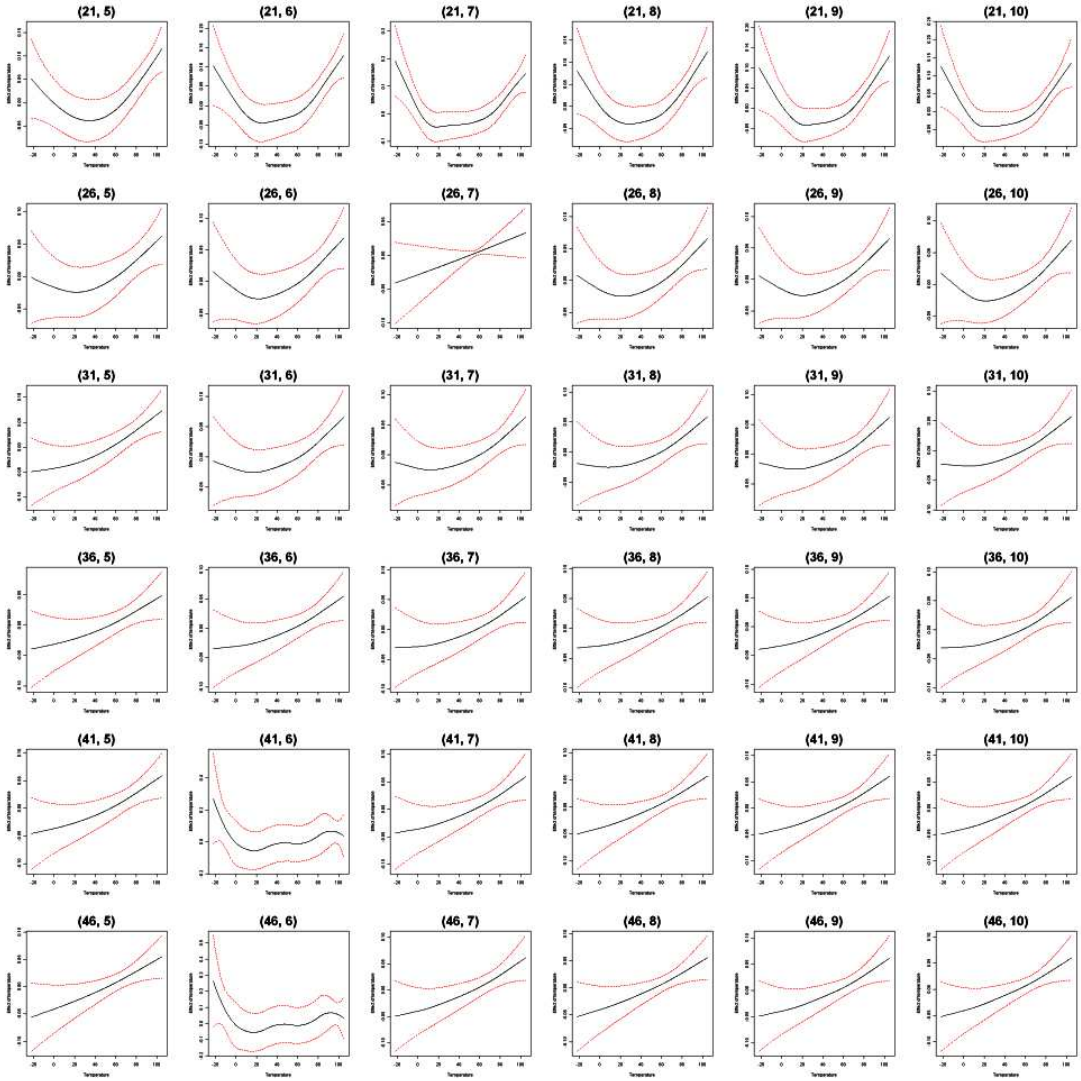
Model 2: Temperature



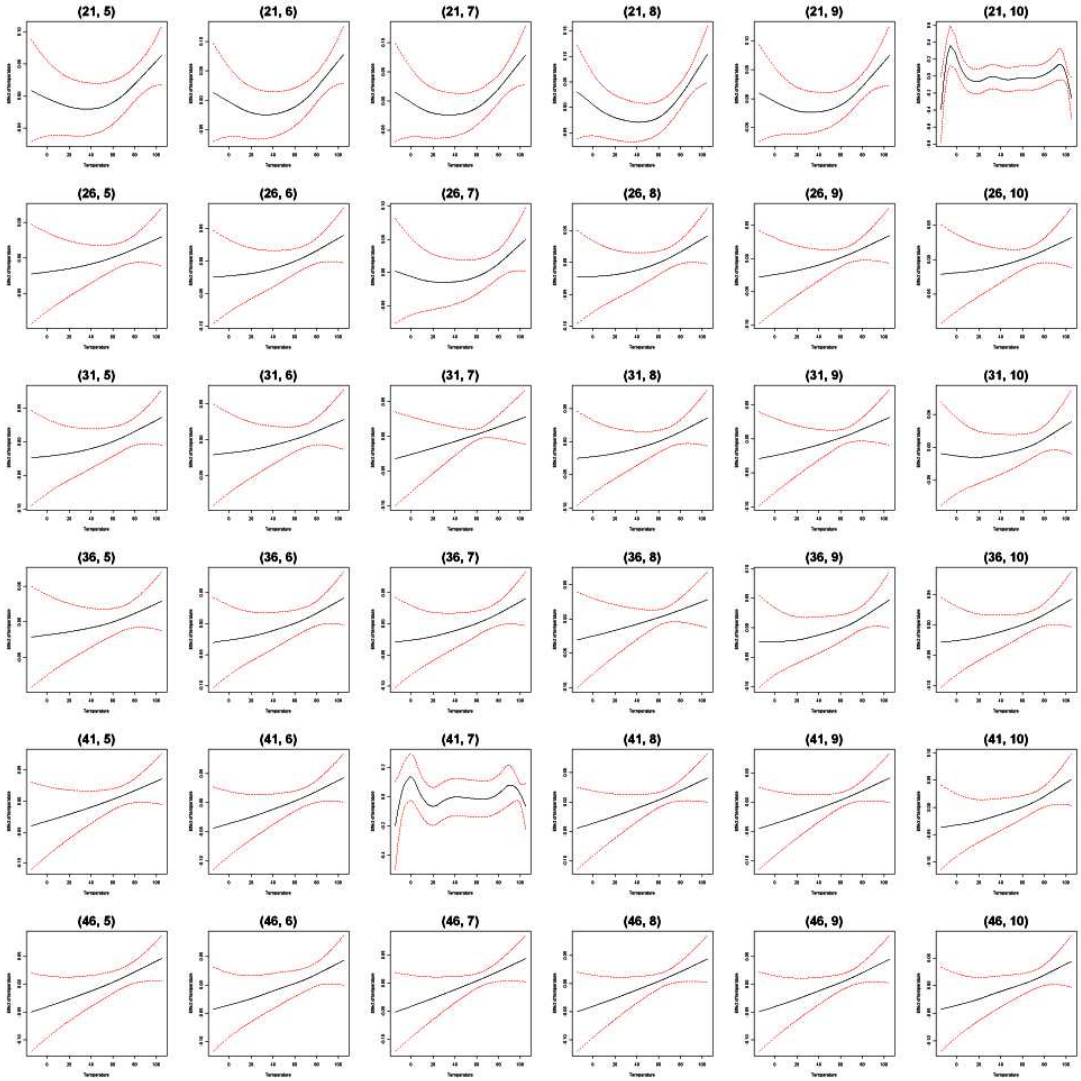
Model 3: Temperature



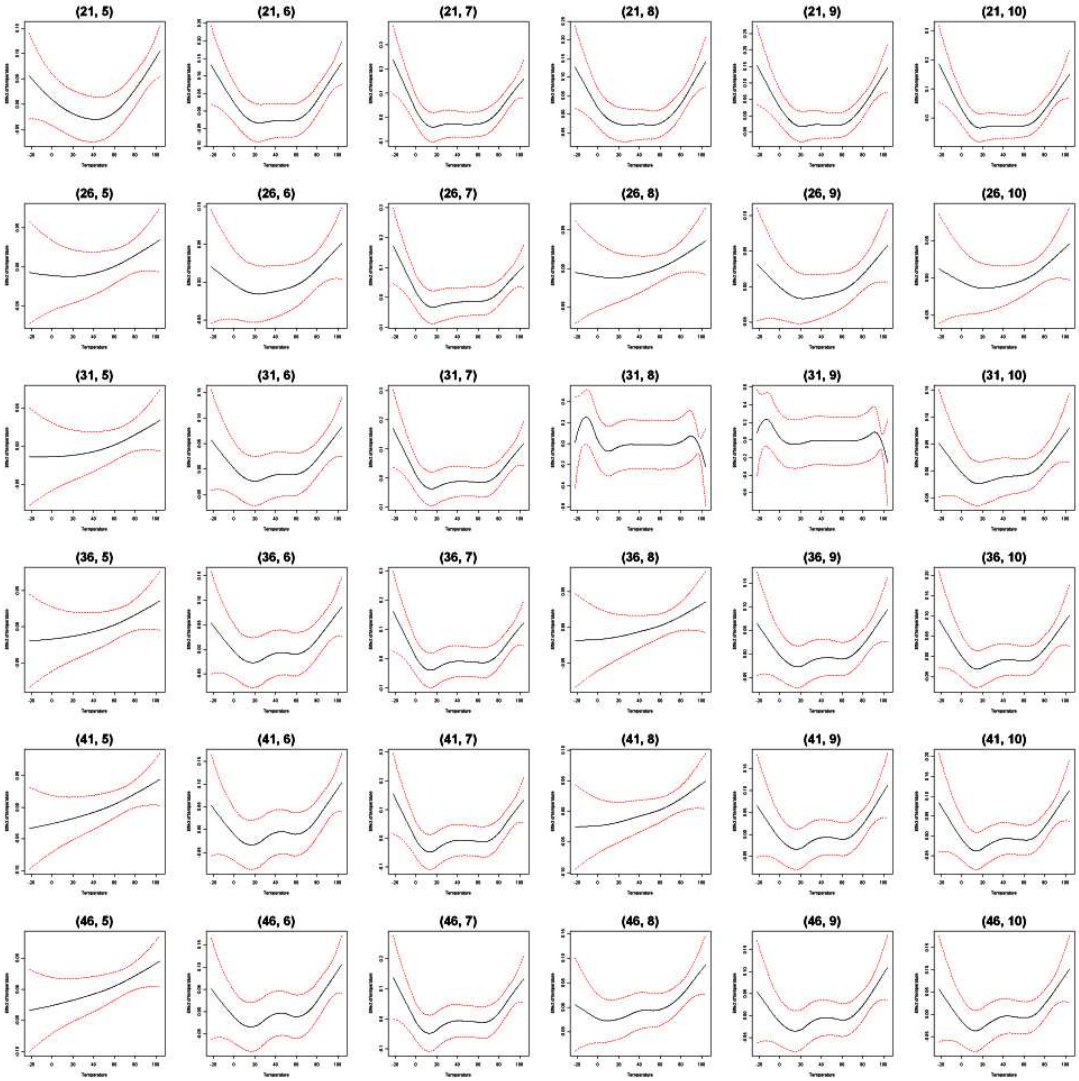
Model 4: Temperature



Model 5: Temperature



Model 6: Temperature

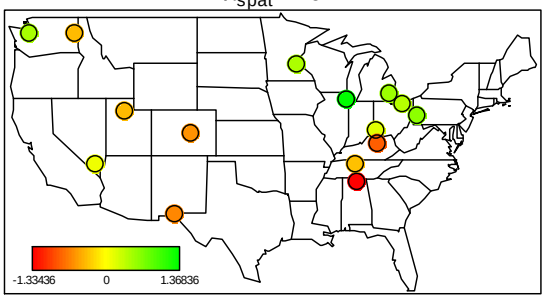


APPENDIX E

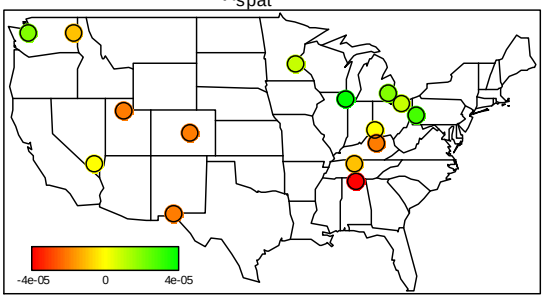
Spatial function maps from different starting values of λ_{spat}.

Model 1

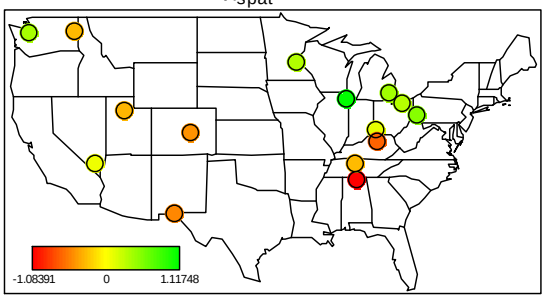
$\lambda_{spat} = 10$



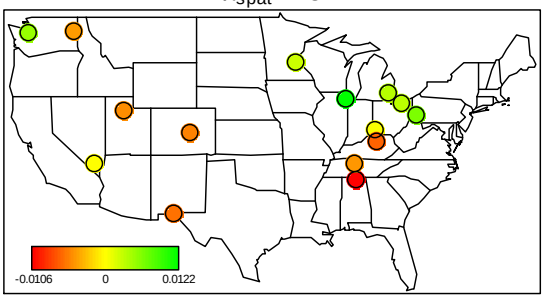
$\lambda_{spat} = 11$



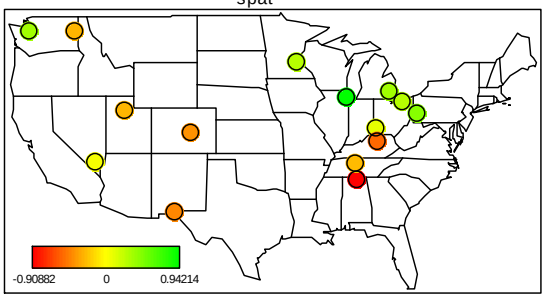
$\lambda_{spat} = 12$



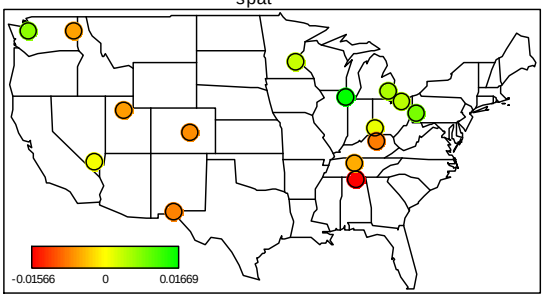
$\lambda_{spat} = 13$



$\lambda_{spat} = 14$

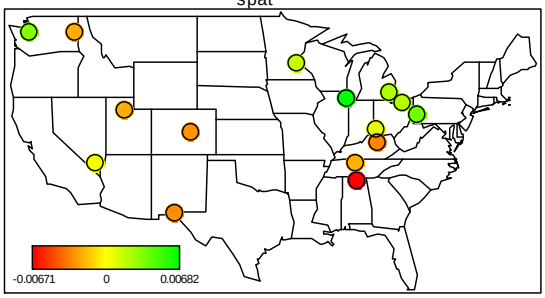


$\lambda_{spat} = 15$

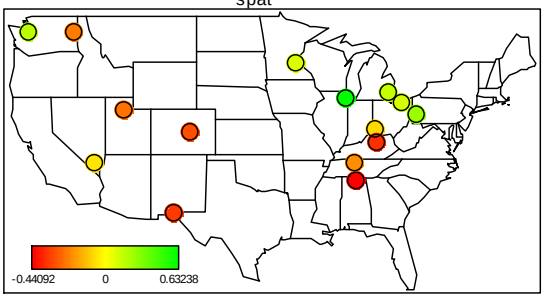


Model 2

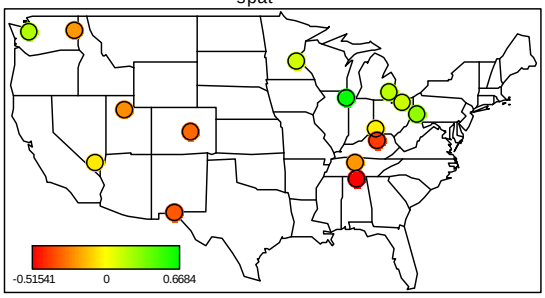
$\lambda_{spat} = 10$



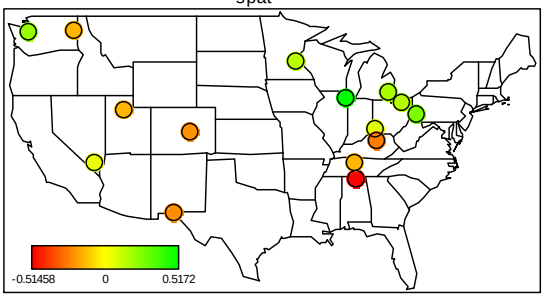
$\lambda_{spat} = 11$



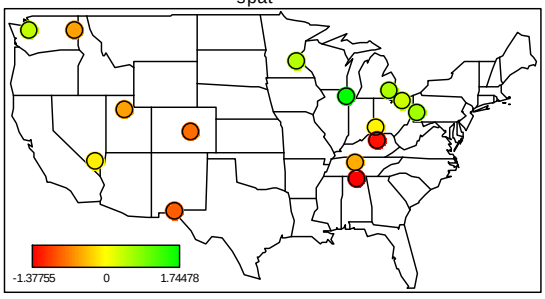
$\lambda_{spat} = 12$



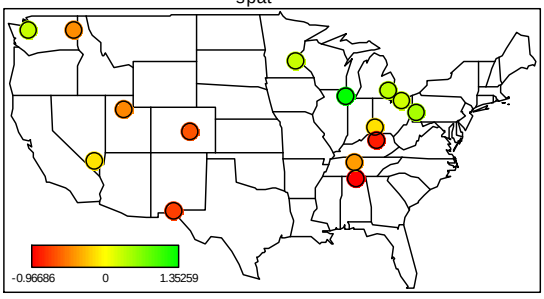
$\lambda_{spat} = 13$



$\lambda_{spat} = 14$

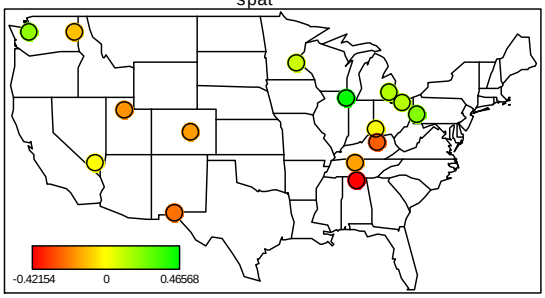


$\lambda_{spat} = 15$

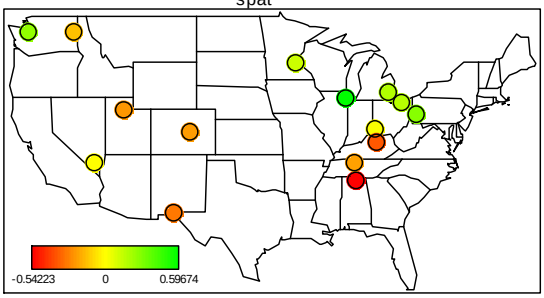


Model 3

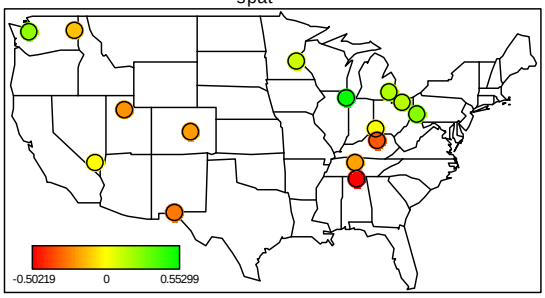
$\lambda_{spat} = 10$



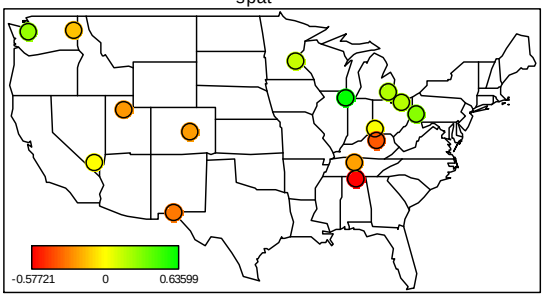
$\lambda_{spat} = 11$



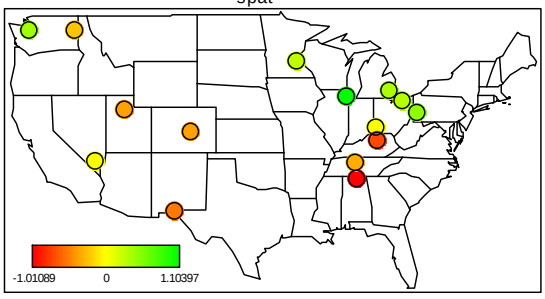
$\lambda_{spat} = 12$



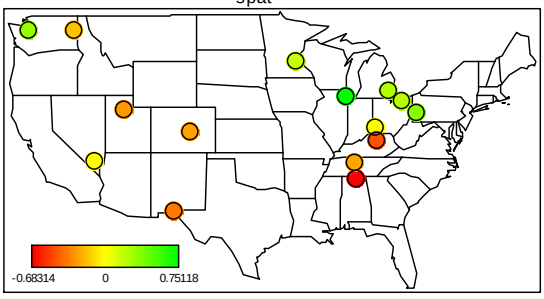
$\lambda_{spat} = 13$



$\lambda_{spat} = 14$

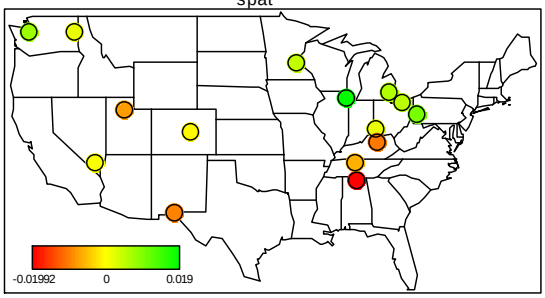


$\lambda_{spat} = 15$

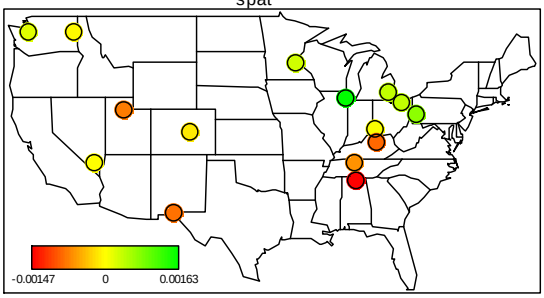


Model 4

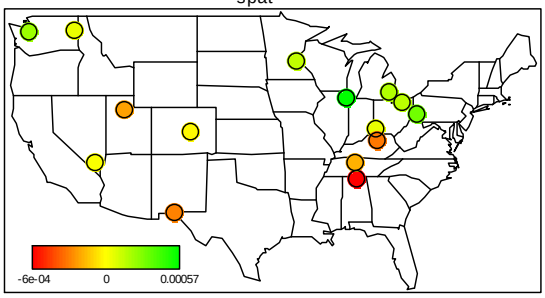
$\lambda_{spat} = 10$



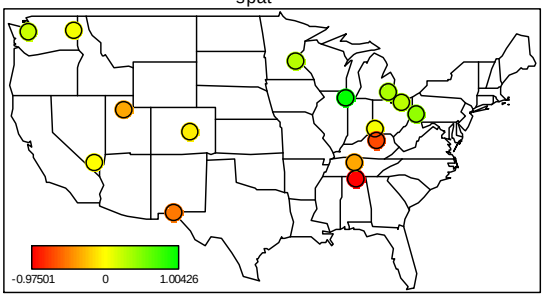
$\lambda_{spat} = 11$



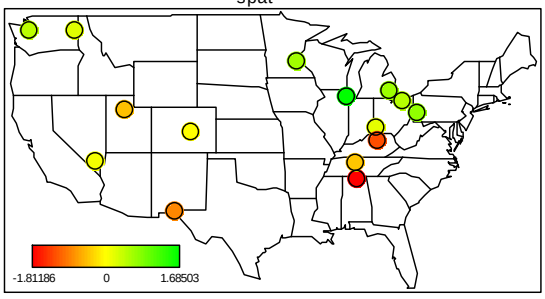
$\lambda_{spat} = 12$



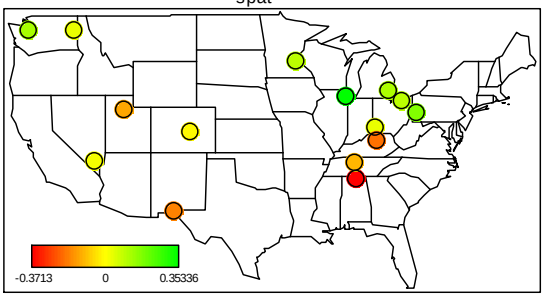
$\lambda_{spat} = 13$



$\lambda_{spat} = 14$

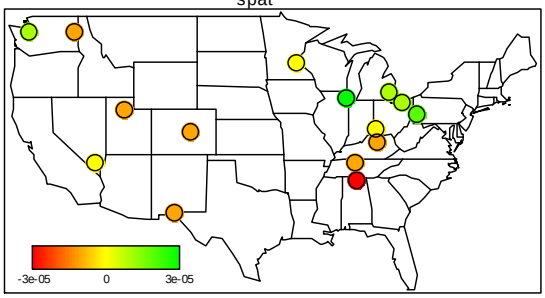


$\lambda_{spat} = 15$

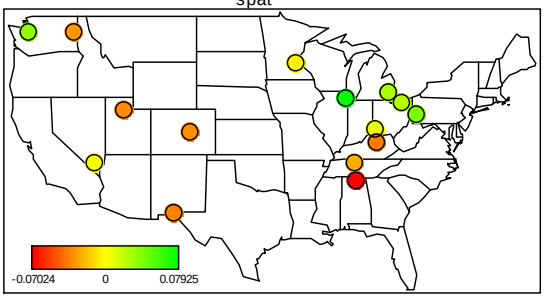


Model 5

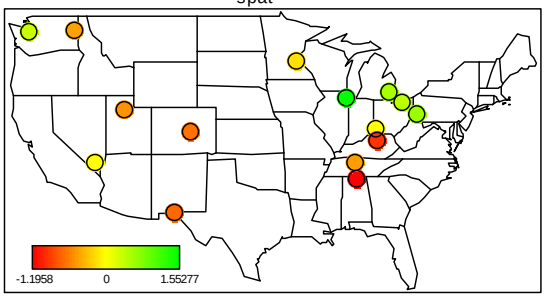
$\lambda_{spat} = 10$



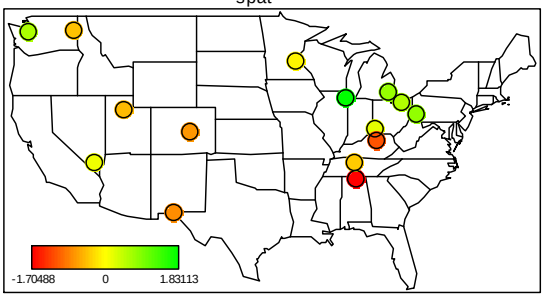
$\lambda_{spat} = 11$



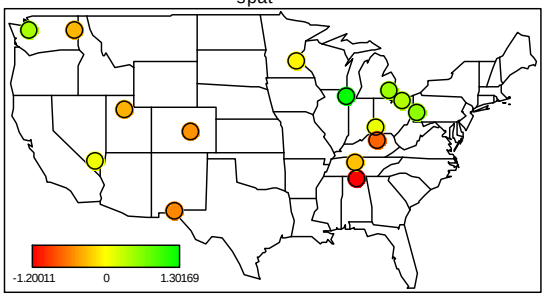
$\lambda_{spat} = 12$



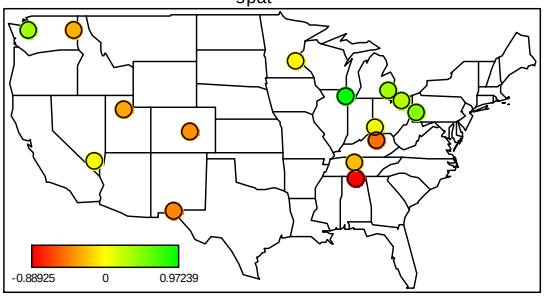
$\lambda_{spat} = 13$



$\lambda_{spat} = 14$

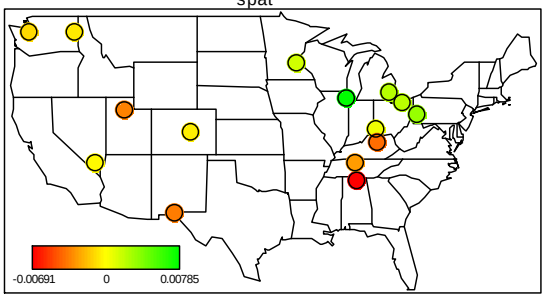


$\lambda_{spat} = 15$

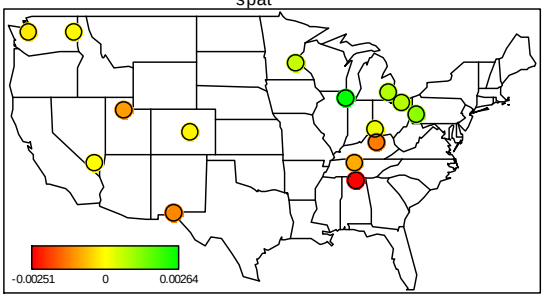


Model 6

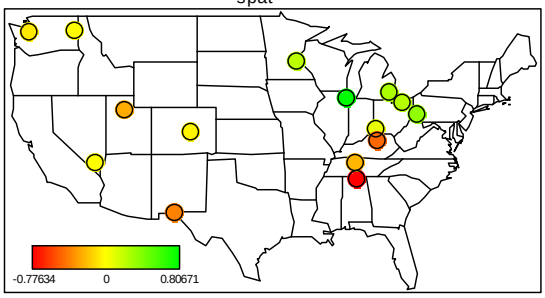
$\lambda_{spat} = 10$



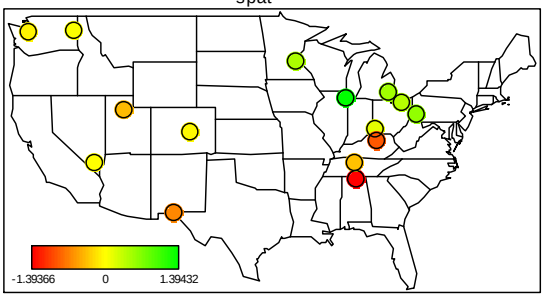
$\lambda_{spat} = 11$



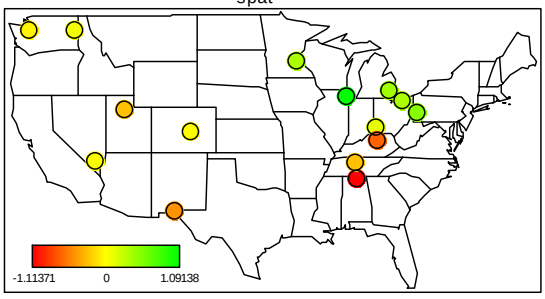
$\lambda_{spat} = 12$



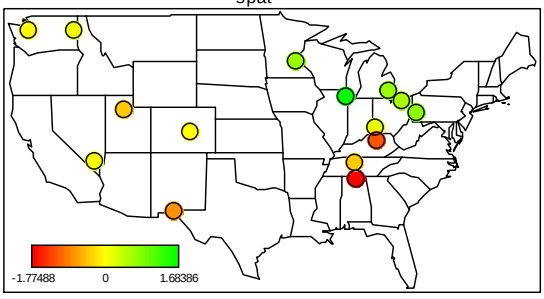
$\lambda_{spat} = 13$



$\lambda_{spat} = 14$



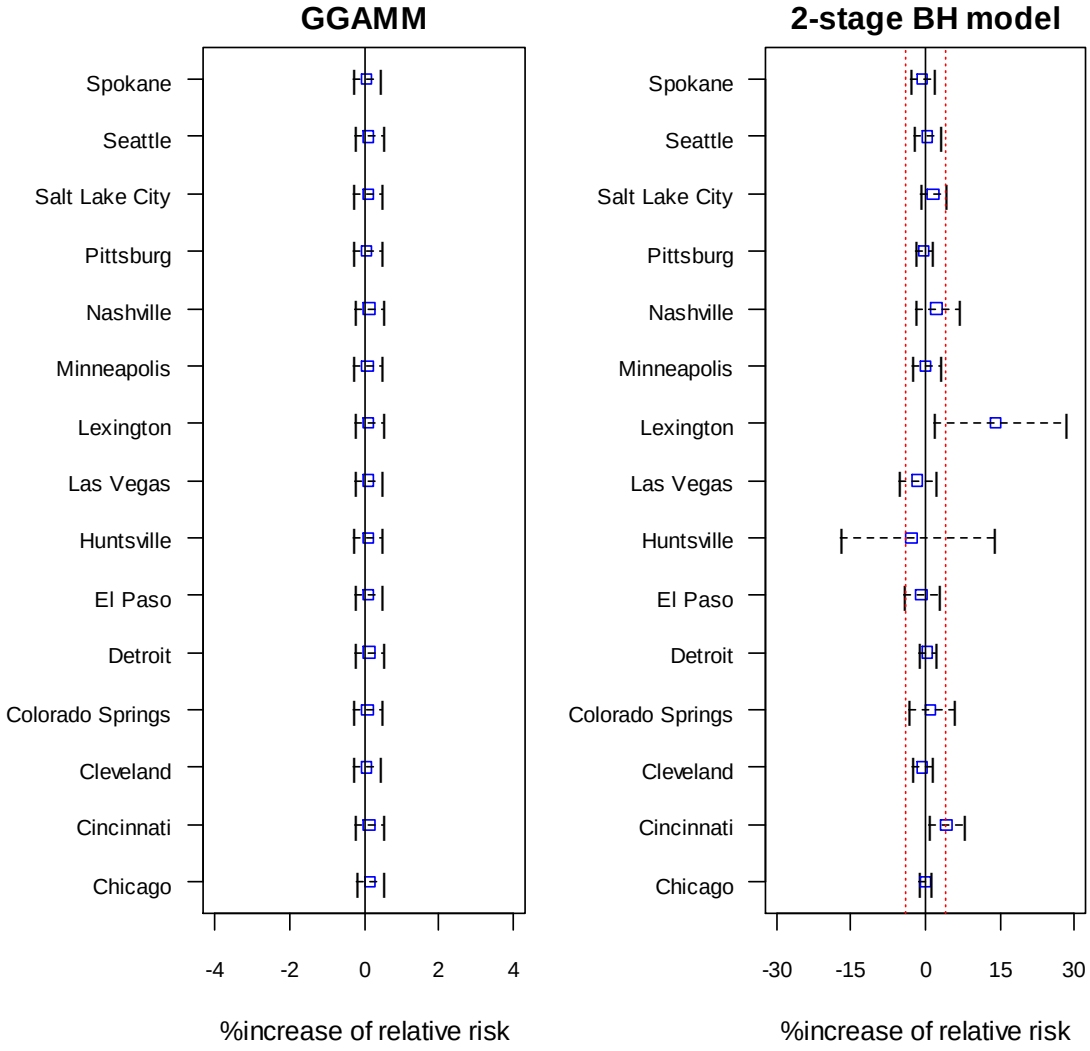
$\lambda_{spat} = 15$



APPENDIX F

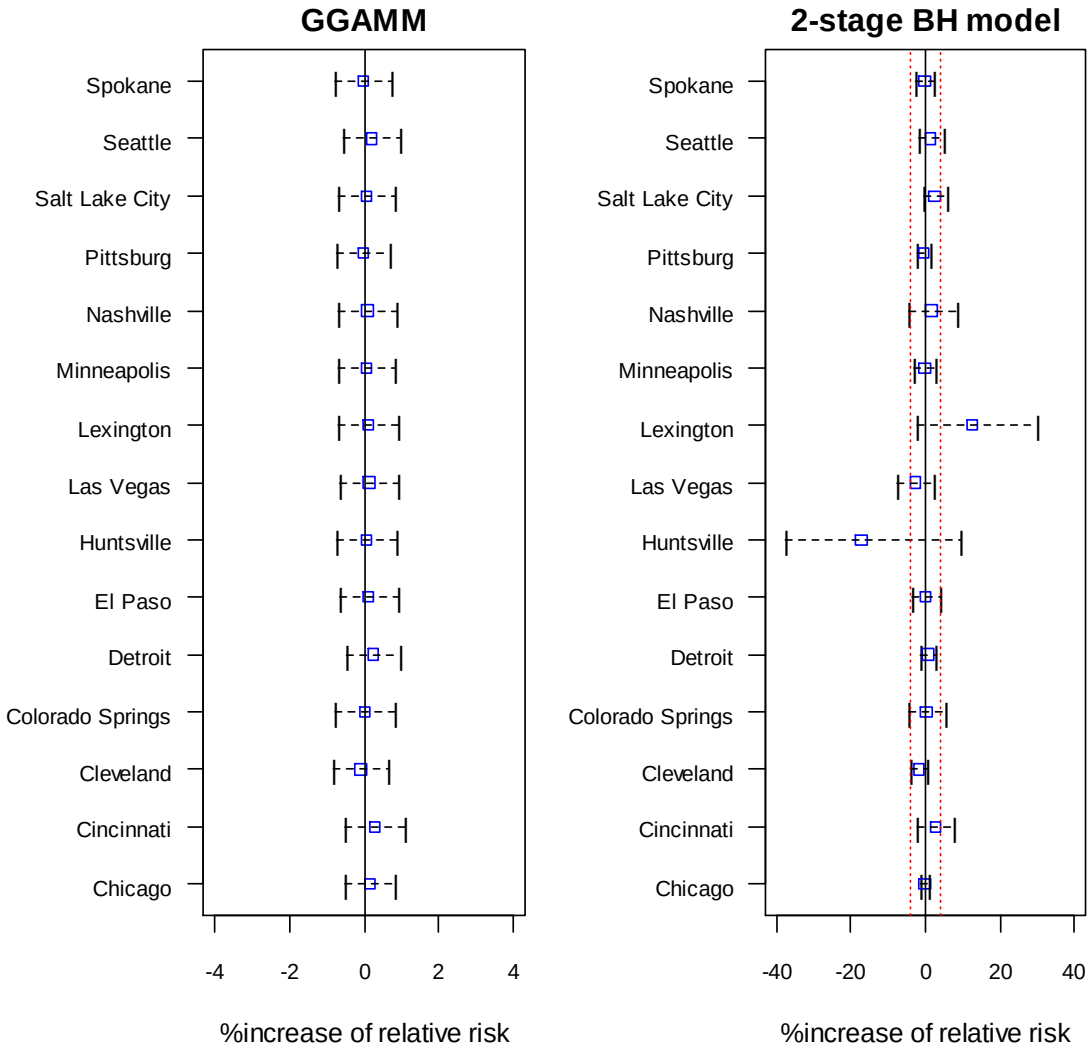
Comparison plots of city-specific PM_{10} effects between the GGAMM and 2-stage Bayesian hierarchical model.

Model 1: PM_{10}



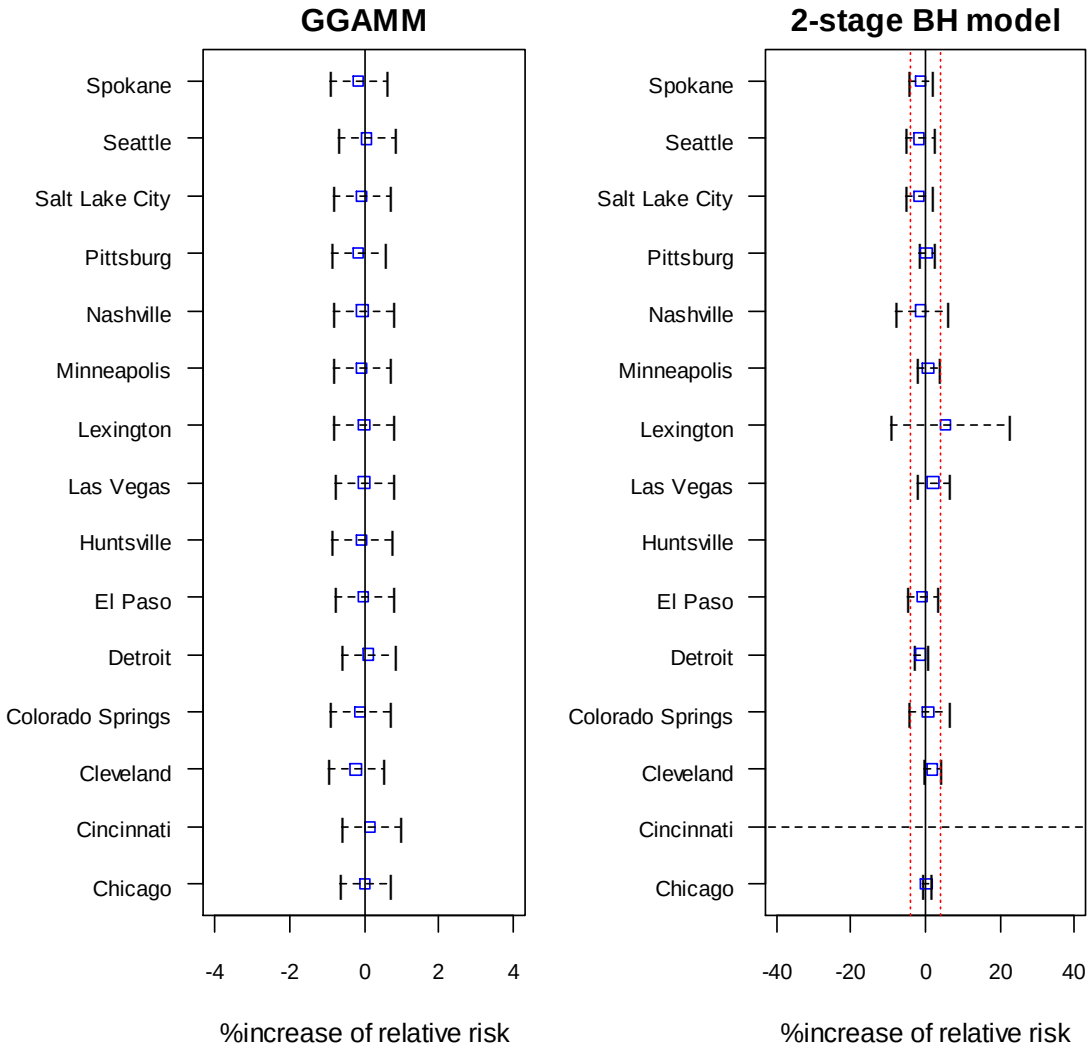
*The red vertical lines in 2-stage Bayesian Hierarchical model plot are the range of x-axis in the GGAMM plot.

Model 2: PM_{10}



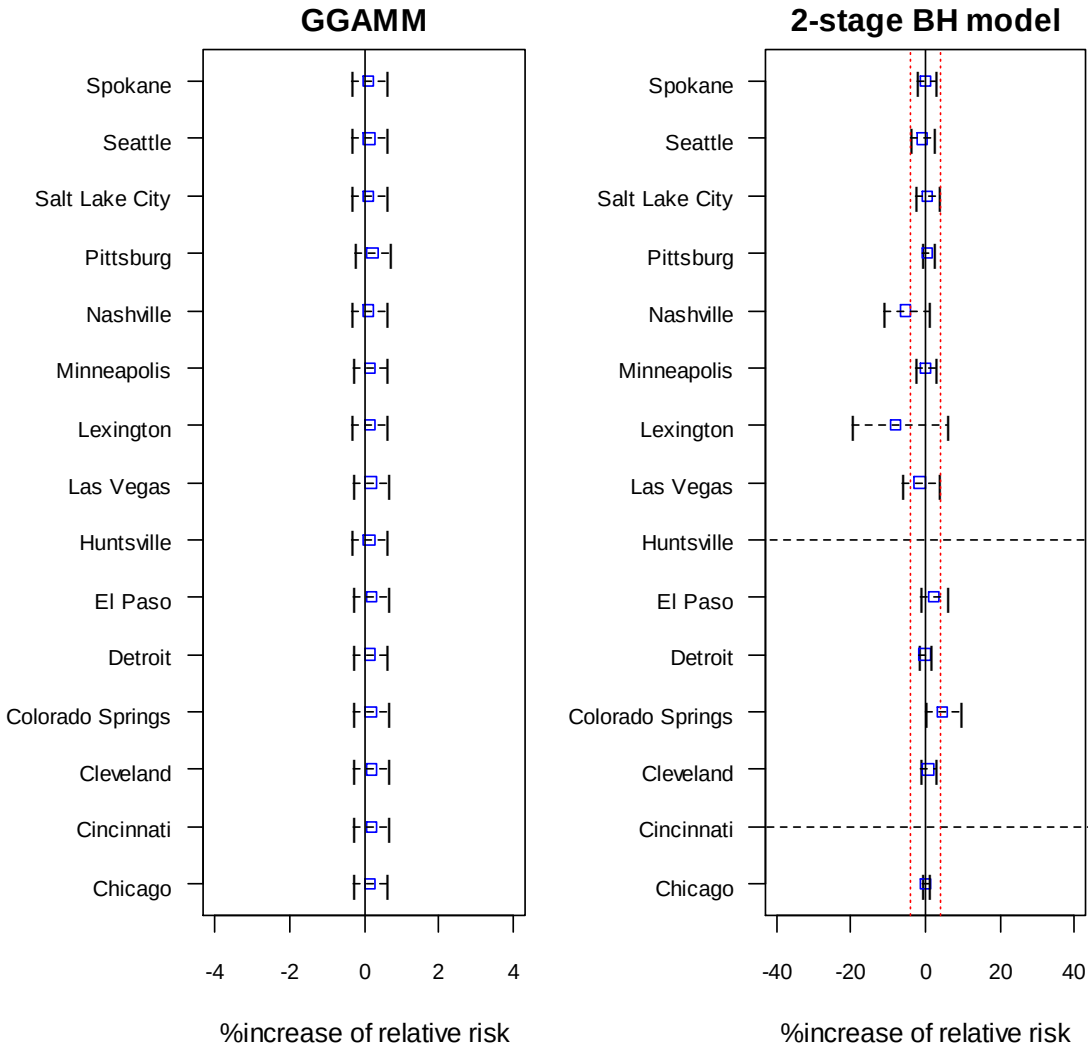
*The red vertical lines in 2-stage Bayesian Hierarchical model plot are the range of x-axis in the GGAMM plot.

Model 2: PM_{10}-lag1



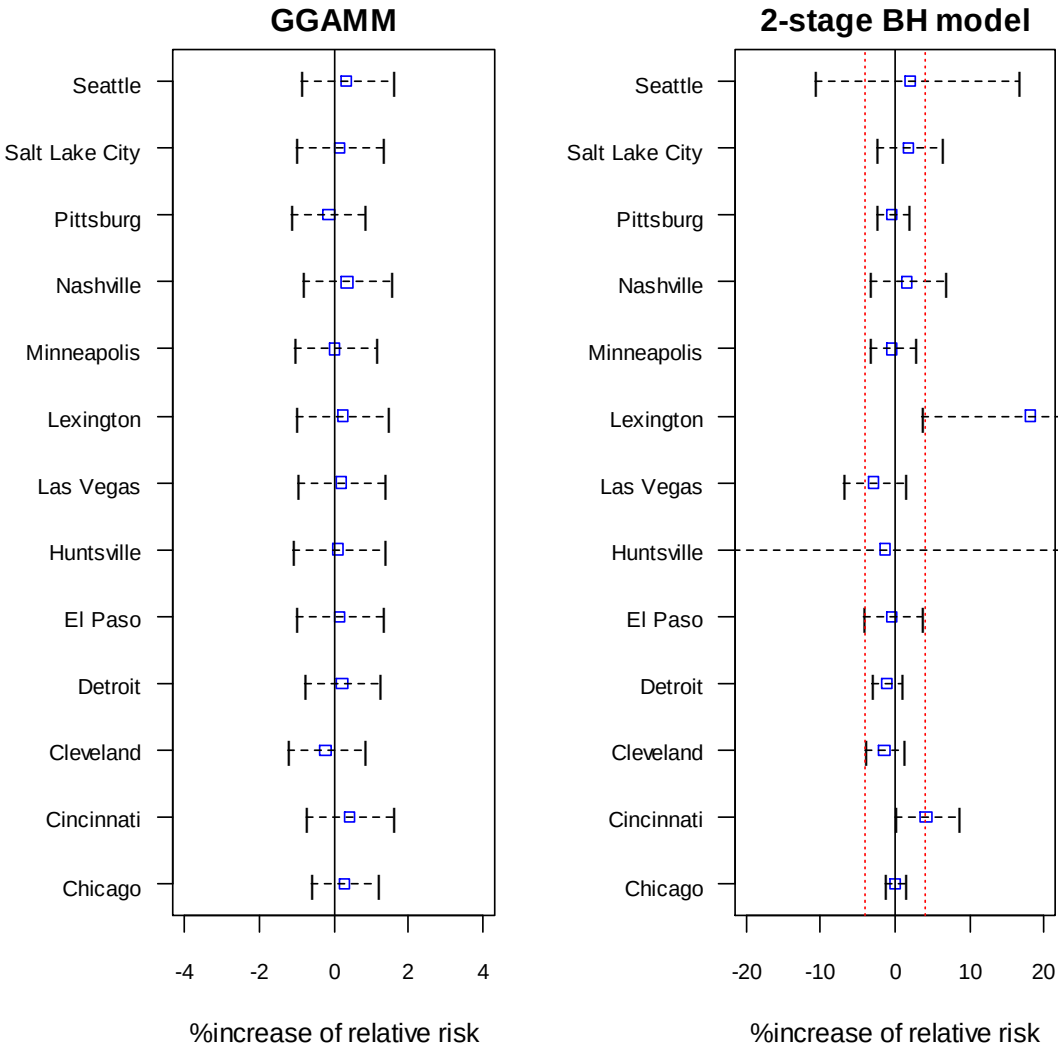
*The red vertical lines in 2-stage Bayesian Hierarchical model plot are the range of x-axis in the GGAMM plot.

Model 2: PM_{10}-lag2



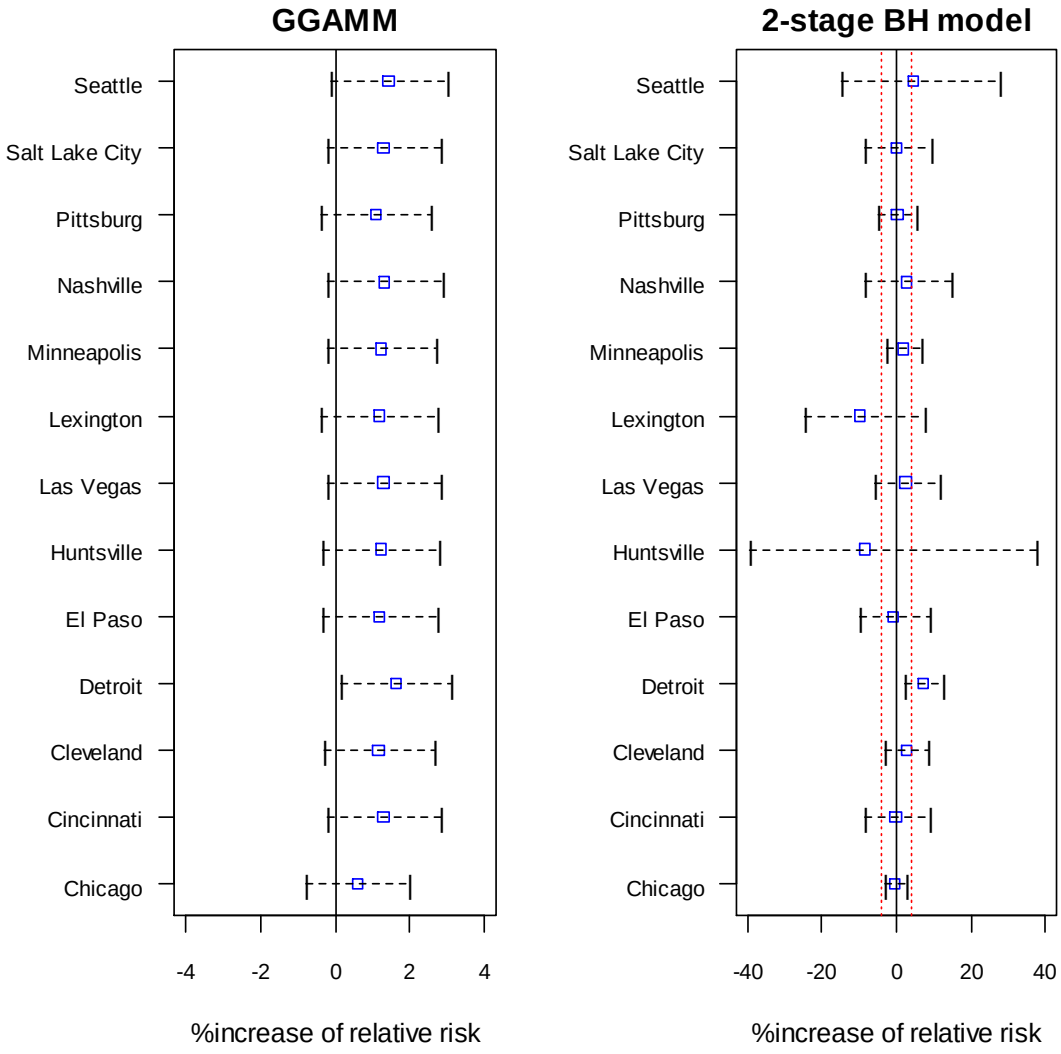
*The red vertical lines in 2-stage Bayesian Hierarchical model plot are the range of x-axis in the GGAMM plot.

Model 4: PM_{10}



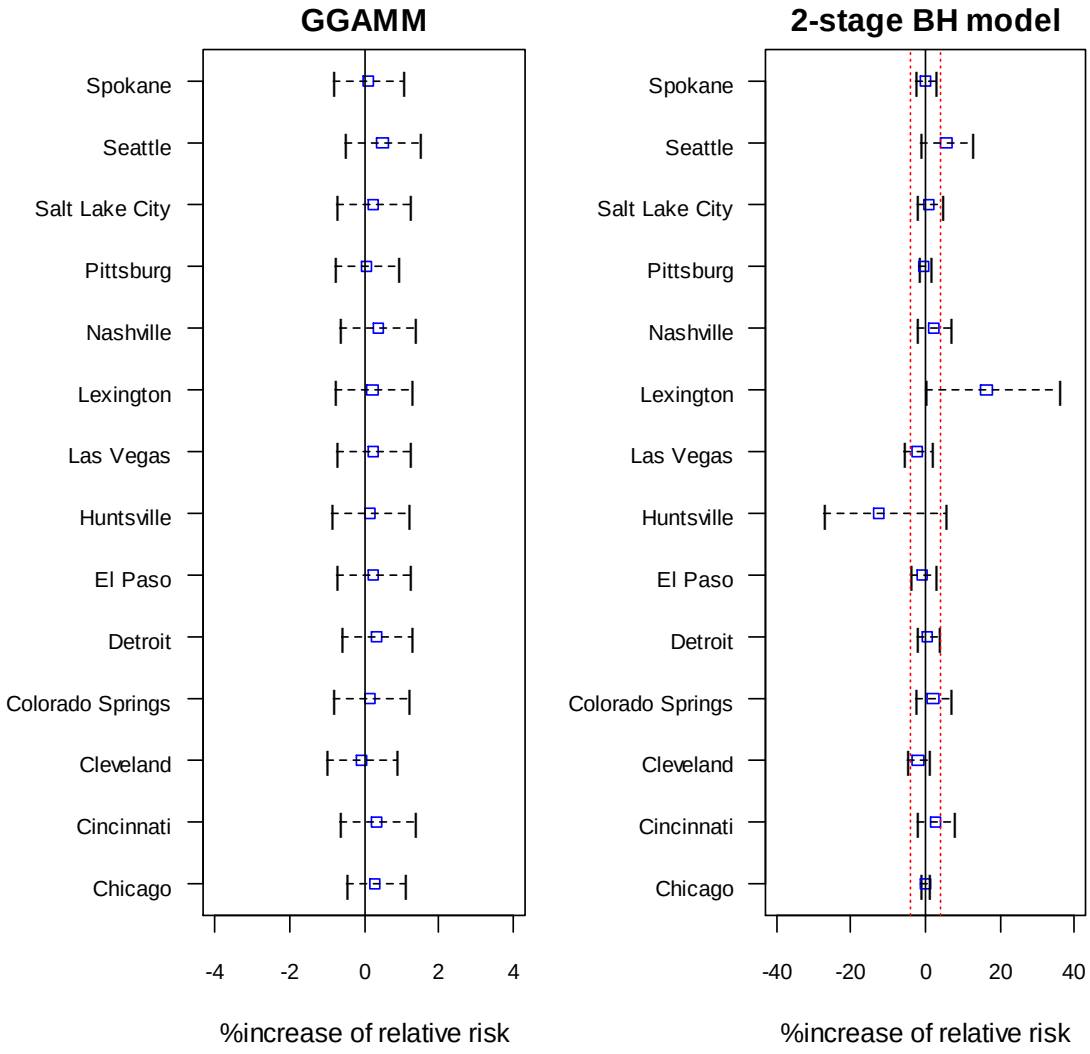
*The red vertical lines in 2-stage Bayesian Hierarchical model plot are the range of x-axis in the GGAMM plot.

Model 4: NO_2



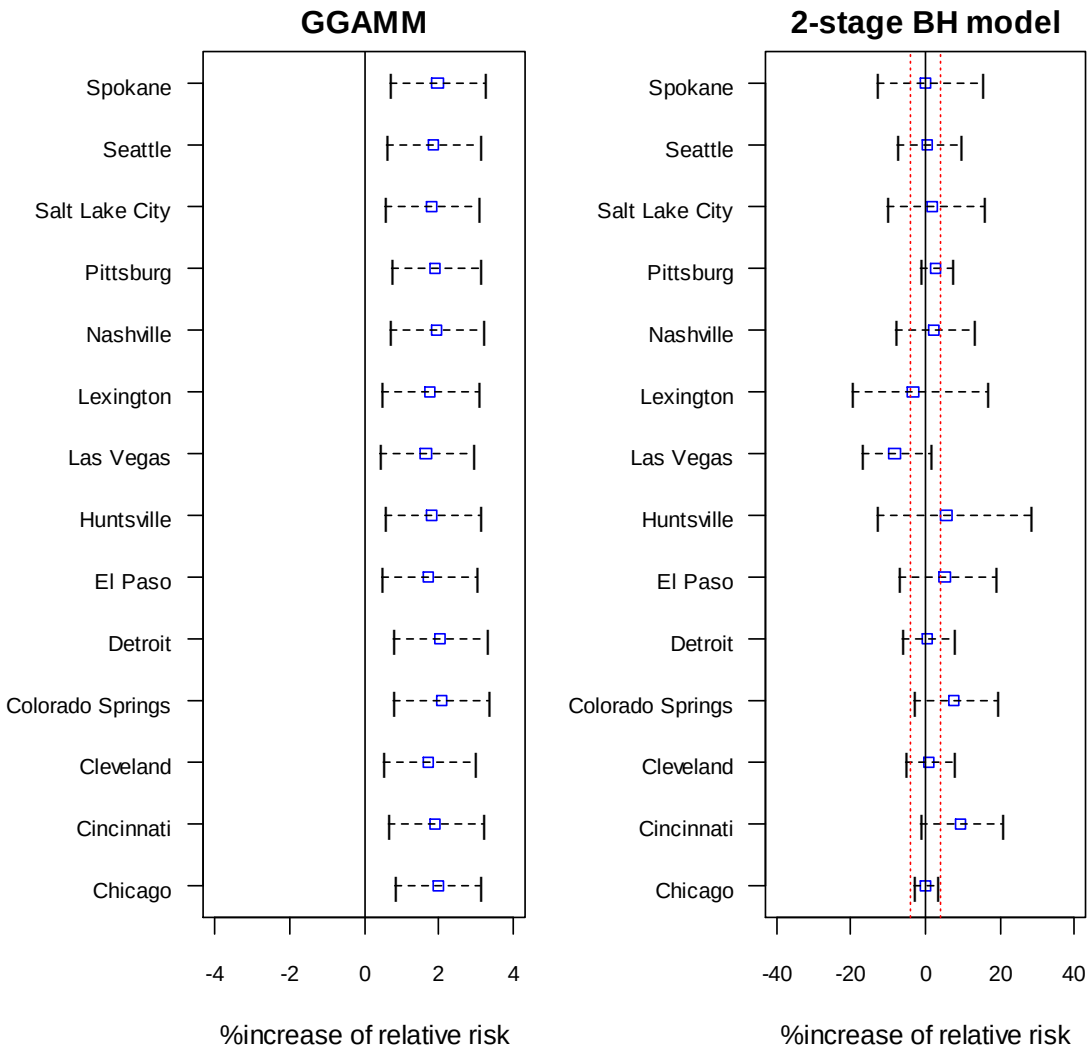
*The red vertical lines in 2-stage Bayesian Hierarchical model plot are the range of x-axis in the GGAMM plot.

Model 5: PM_{10}



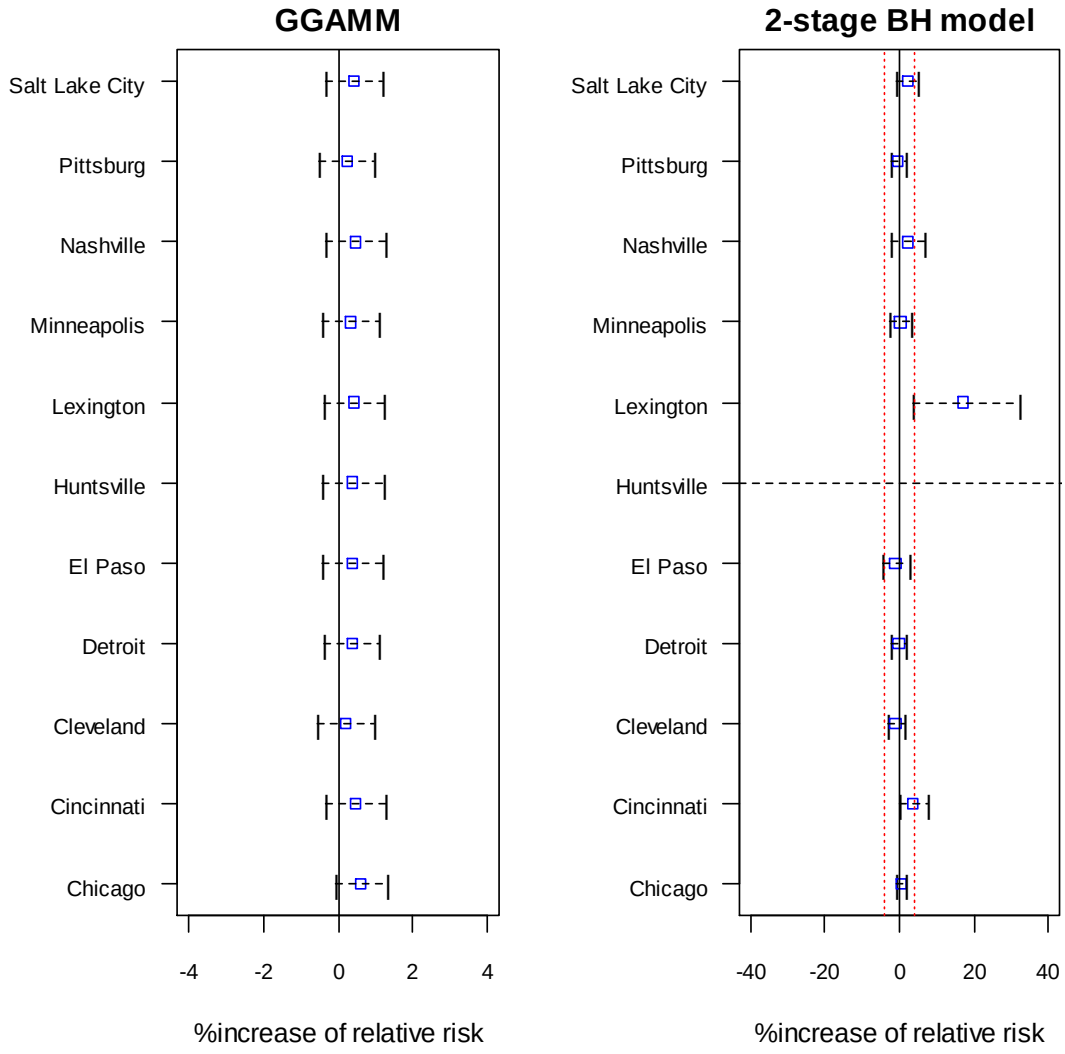
*The red vertical lines in 2-stage Bayesian Hierarchical model plot are the range of x-axis in the GGAMM plot.

Model 5: O_3



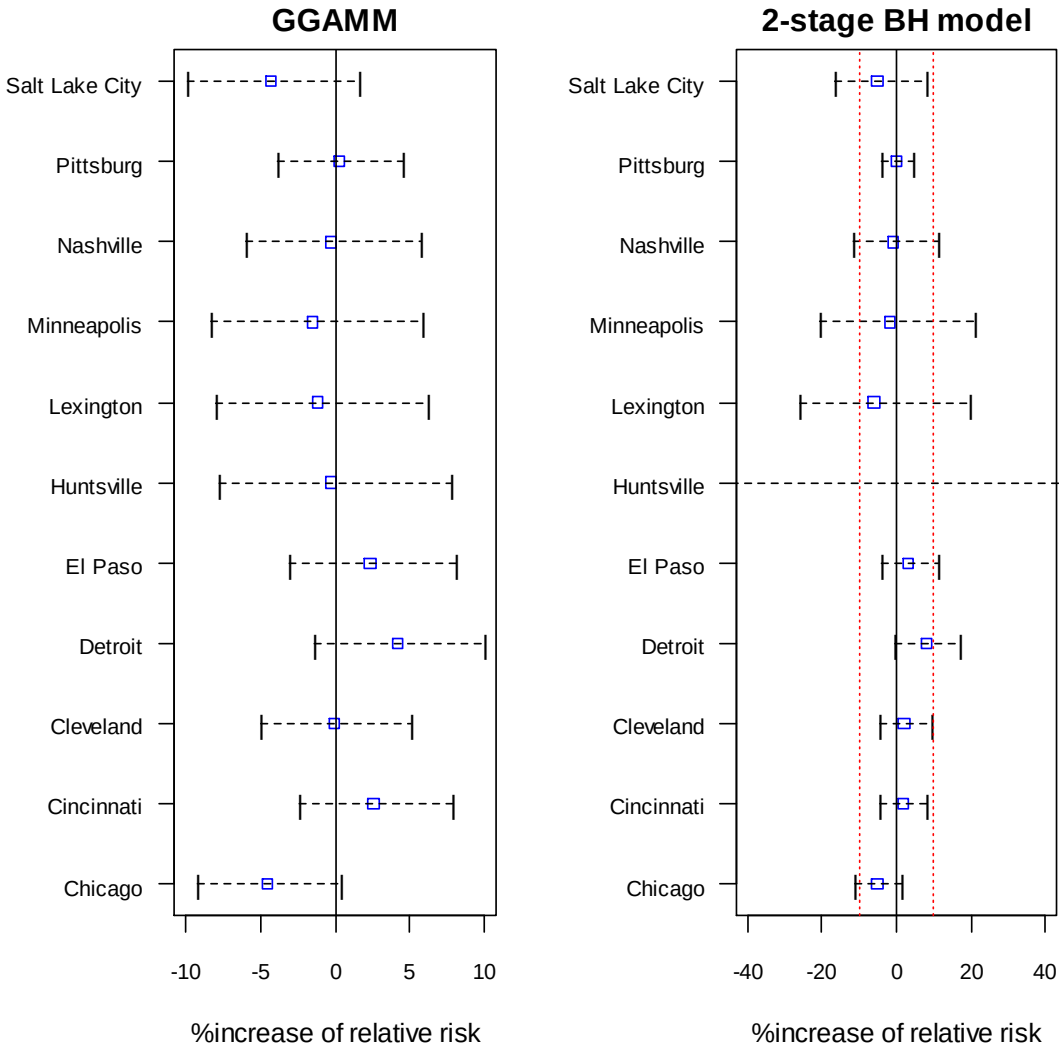
*The red vertical lines in 2-stage Bayesian Hierarchical model plot are the range of x-axis in the GGAMM plot.

Model 6: PM_{10}



*The red vertical lines in 2-stage Bayesian Hierarchical model plot are the range of x-axis in the GGAMM plot.

Model 6: SO_2



*The red vertical lines in 2-stage Bayesian Hierarchical model plot are the range of x-axis in the GGAMM plot.

APPENDIX G

Tables of parameter estimates for cardiovascular, pneumonia, and respiratory mortality in three age categories

Cardiovascular mortality

Age	Model	Variable	$\hat{\beta}$	se($\hat{\beta}$)	se($\hat{b}$)
<65	Model 1	PM_{10}	-0.000157	0.000370	0.000707
	Model 4	PM_{10}	-0.000276	0.000458	0.000740
		NO_2	0.000039	0.000799	0.000606
	Model 5	PM_{10}	-0.000266	0.000468	0.000929
		O_3	0.000892	0.000998	0.001801
	Model 6	PM_{10}	-0.000433	0.000352	0.000288
		SO_2	0.001292	0.001506	0.002752
65−74	Model 1	PM_{10}	0.000375	0.000310	0.000555
	Model 4	PM_{10}	0.000284	0.000374	0.000512
		NO_2	-0.000013	0.000984	0.002018
	Model 5	PM_{10}	0.000632	0.000403	0.000804
		O_3	0.000735	0.000903	0.001750
	Model 6	PM_{10}	-0.000061	0.000427	0.000787
		SO_2	0.001755	0.001157	0.001667
≥75	Model 1	PM_{10}	0.000472	0.000171	0.000237
	Model 4	PM_{10}	0.000480	0.000193	0.000042
		NO_2	0.000342	0.044061	0.158832

Pneumonia mortality

Age	Model	Variable	$\hat{\beta}$	se($\hat{\beta}$)	se($\hat{b}$)
<65	Model 1	PM_{10}	0.001823	0.001091	0.001699
	Model 2	PM_{10}	0.000981	0.001250	0.001759
		PM_{10}-lag1	0.001484	0.001356	0.001702
		PM_{10}-lag2	-0.001443	0.001218	0.001180
	Model 3	PM_{10}	0.001659	0.001193	0.001652
		CO	0.000069	0.104924	0.316837
	Model 4	PM_{10}	0.001144	0.001315	0.000847
		NO_2	0.002744	0.002886	0.002037
	Model 5	PM_{10}	0.001774	0.001205	0.001693
		O_3	-0.004638	0.002870	0.002997
	Model 6	PM_{10}	0.000382	0.001339	0.001394
		SO_2	0.005537	0.004722	0.006357
65−74	Model 1	PM_{10}	0.002004	0.001018	0.001283
	Model 2	PM_{10}	0.002957	0.001182	0.001600
		PM_{10}-lag1	-0.002252	0.001256	0.000827
		PM_{10}-lag2	0.001095	0.001306	0.002056
	Model 3	PM_{10}	0.001848	0.001108	0.000961
		CO	0.000026	0.081491	0.315612
	Model 4	PM_{10}	0.001225	0.001336	0.001268
		NO_2	0.005534	0.002675	0.000887
	Model 5	PM_{10}	0.002619	0.001205	0.001667
		O_3	0.004249	0.002574	0.002219
	Model 6	PM_{10}	0.002614	0.001317	0.001543
		SO_2	-0.002958	0.004472	0.005302
≥75	Model 1	PM_{10}	0.0002662	0.000545	0.001013
	Model 2	PM_{10}	0.0007194	0.000530	0.000506
		PM_{10}-lag1	-0.001241	0.000800	0.001686
		PM_{10}-lag2	-0.000091	0.000743	0.001654
	Model 4	PM_{10}	0.000586	0.000590	0.000490
		NO_2	0.001298	0.001272	0.001085
	Model 5	PM_{10}	0.000603	0.000558	0.000567
		O_3	0.003231	0.001355	0.001287
	Model 6	PM_{10}	0.000664	0.000582	0.000652
		SO_2	-0.000427	0.002657	0.005582

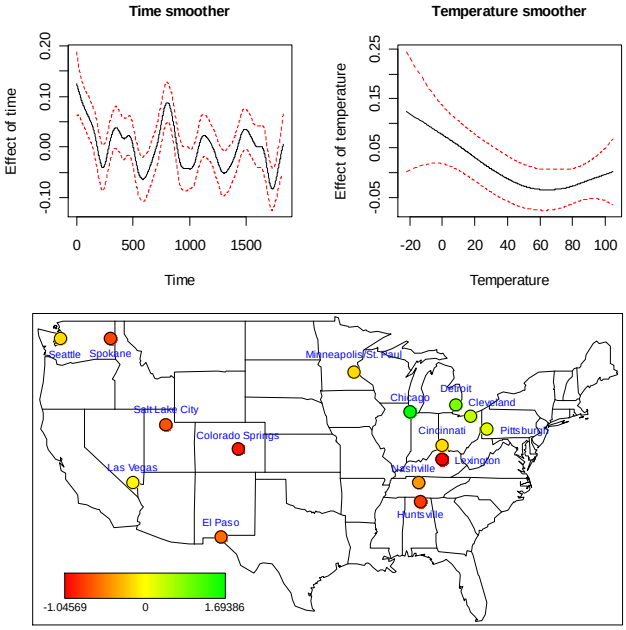
Respiratory mortality

Age	Model	Variable	$\hat{\beta}$	se($\hat{\beta}$)	se($\hat{b}$)
<65	Model 1	PM_{10}	0.002095	0.000703	0.000947
	Model 2	PM_{10}	0.001672	0.000802	0.000949
		PM_{10}-lag1	0.000818	0.000912	0.001130
		PM_{10}-lag2	-0.000177	0.000841	0.001019
	Model 4	PM_{10}	0.001797	0.001055	0.001566
		NO_2	0.000183	0.002081	0.002362
	Model 5	PM_{10}	0.002191	0.000828	0.001154
		O_3	-0.007584	0.002745	0.006312
	Model 6	PM_{10}	0.001886	0.000871	0.000837
		SO_2	-0.000951	0.003108	0.003716
65−74	Model 1	PM_{10}	0.000199	0.000630	0.000953
	Model 2	PM_{10}	-0.000009	0.000728	0.001030
		PM_{10}-lag1	0.000262	0.000728	0.000870
		PM_{10}-lag2	0.000405	0.000658	0.000806
	Model 4	PM_{10}	0.000329	0.000933	0.001670
		NO_2	0.002703	0.001603	0.001242
	Model 5	PM_{10}	0.000706	0.000779	0.001221
		O_3	0.000332	0.002128	0.004678
	Model 6	PM_{10}	0.000416	0.000736	0.000767
		SO_2	-0.000182	0.002840	0.004784
≥75	Model 1	PM_{10}	0.000133	0.000343	0.000311
	Model 2	PM_{10}	0.000184	0.000443	0.000654
		PM_{10}-lag1	-0.000317	0.000545	0.001060
		PM_{10}-lag2	0.000074	0.000412	0.000550
	Model 4	PM_{10}	0.000258	0.000462	0.000445
		NO_2	0.000703	0.000954	0.000670
	Model 5	PM_{10}	0.000155	0.000408	0.000360
		O_3	0.002087	0.000964	0.000760
	Model 6	PM_{10}	0.000397	0.000467	0.000564
		SO_2	-0.000082	0.001924	0.003835

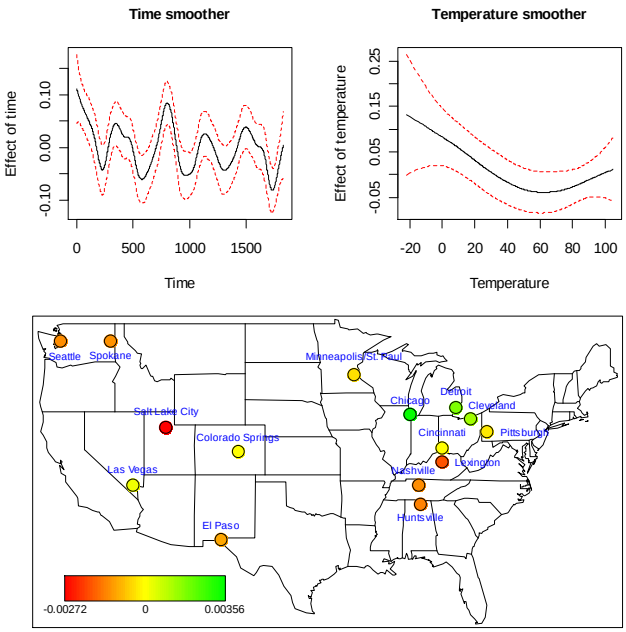
APPENDIX H

Smoothing function plots of time and temperature and spatial function map in extended applications

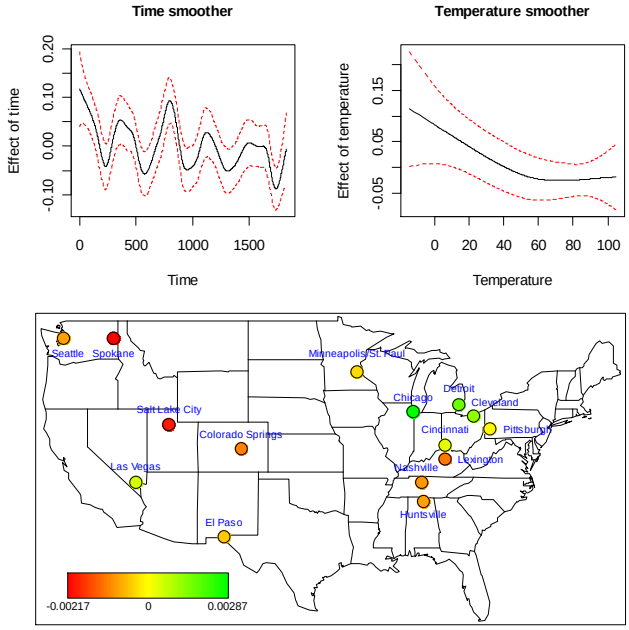
Cardiovascular disease＜65 years old / Model: PM_{10}



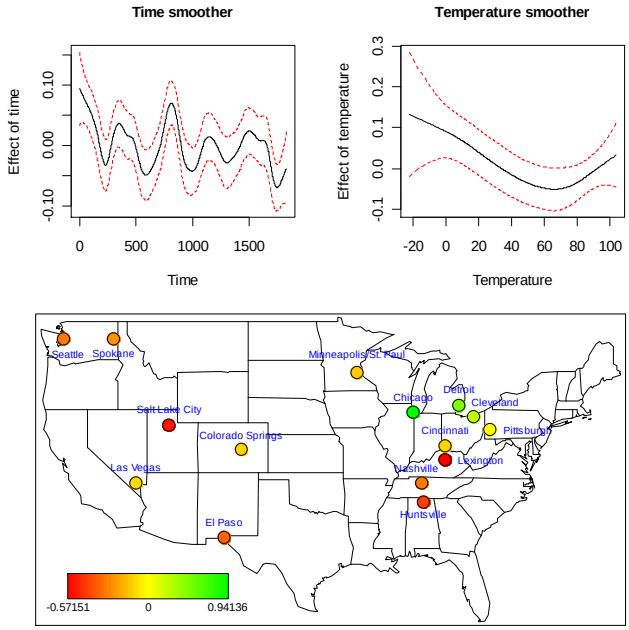
Cardiovascular disease＜65 years old / Model: PM_{10}+NO_2



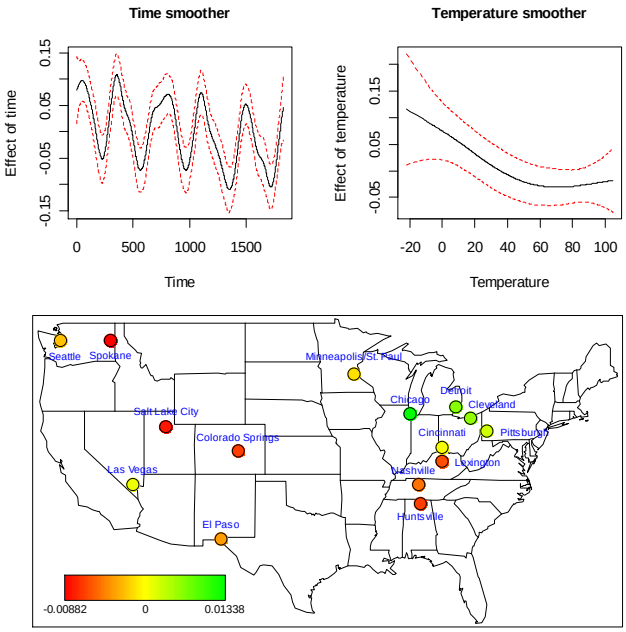
Cardiovascular disease < 65 years old / Model: $PM_{10}+O_3$



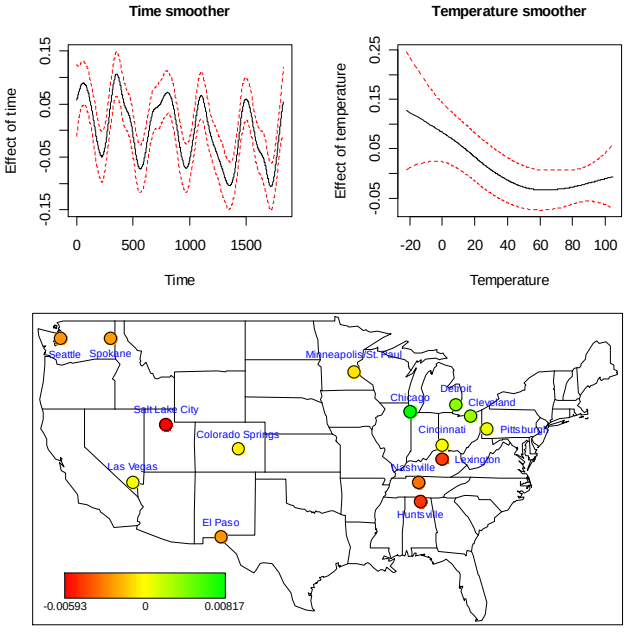
Cardiovascular disease < 65 years old / Model: $PM_{10}+SO_2$



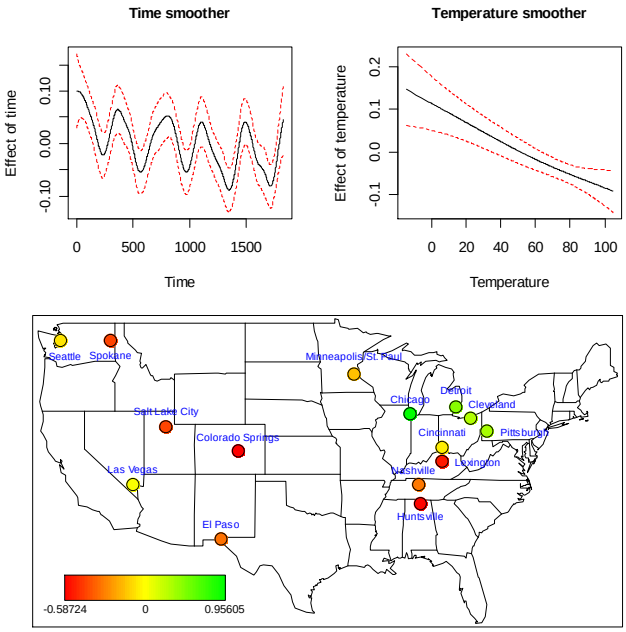
Cardiovascular disease 65−74 years old / Model: PM_{10}



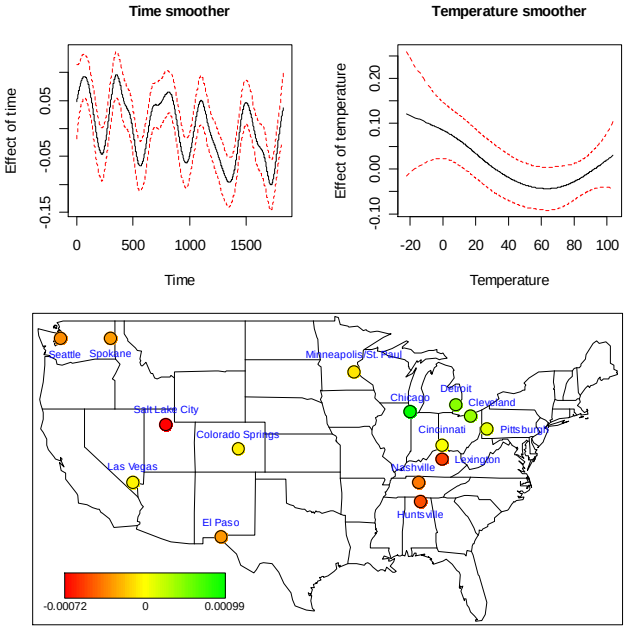
Cardiovascular disease 65−74 years old / Model: PM_{10}+NO_2



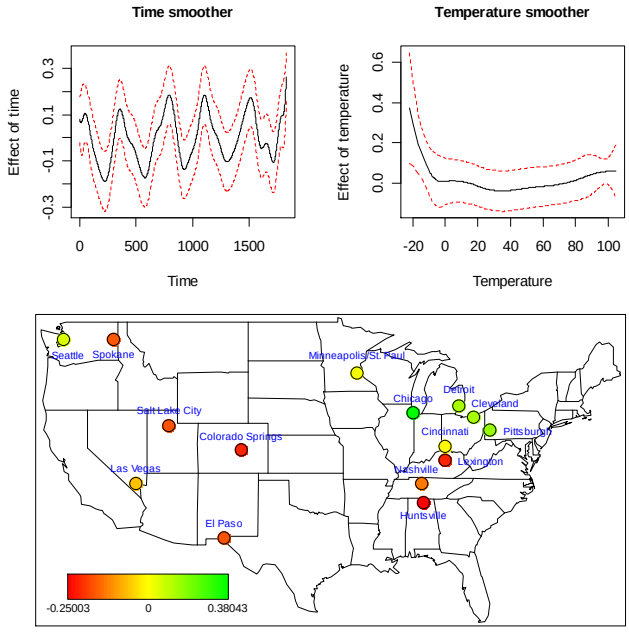
Cardiovascular disease 65−74 years old / Model: PM_{10}+O_3



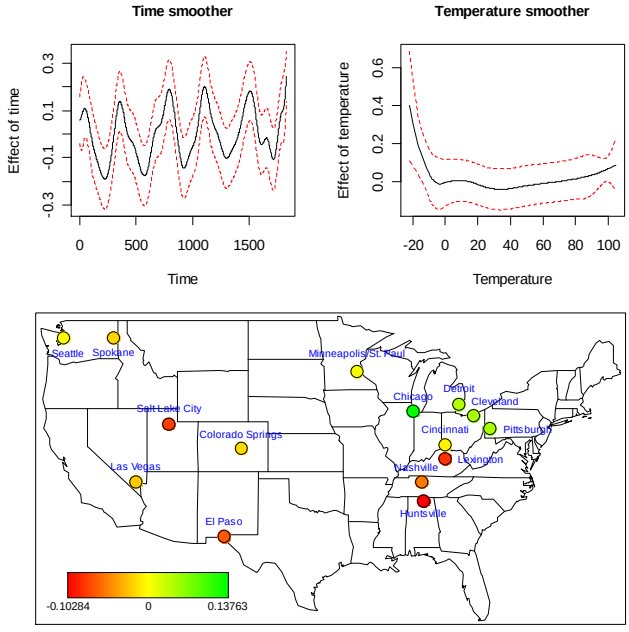
Cardiovascular disease 65−74 years old / Model: PM_{10}+SO_2



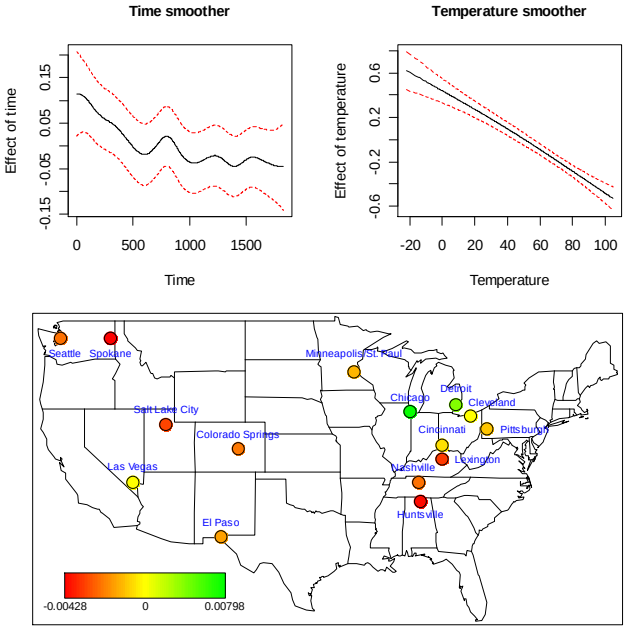
Cardiovascular disease≥75 years old / Model: PM_{10}



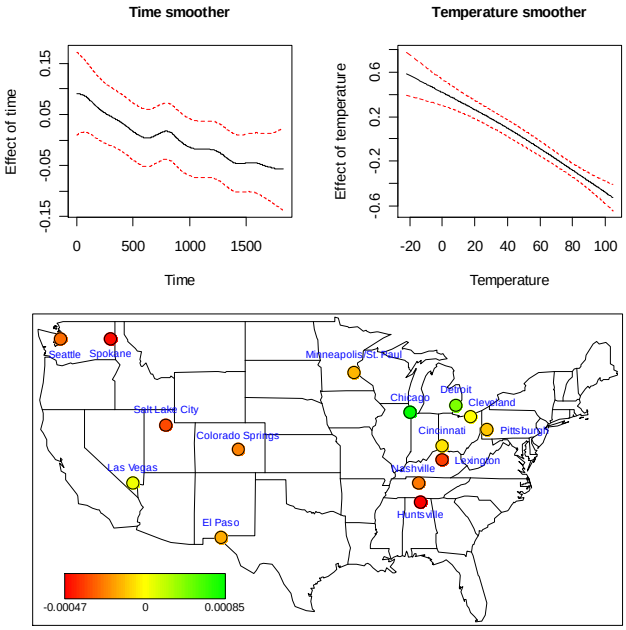
Cardiovascular disease≥75 years old / Model: PM_{10}+NO_2



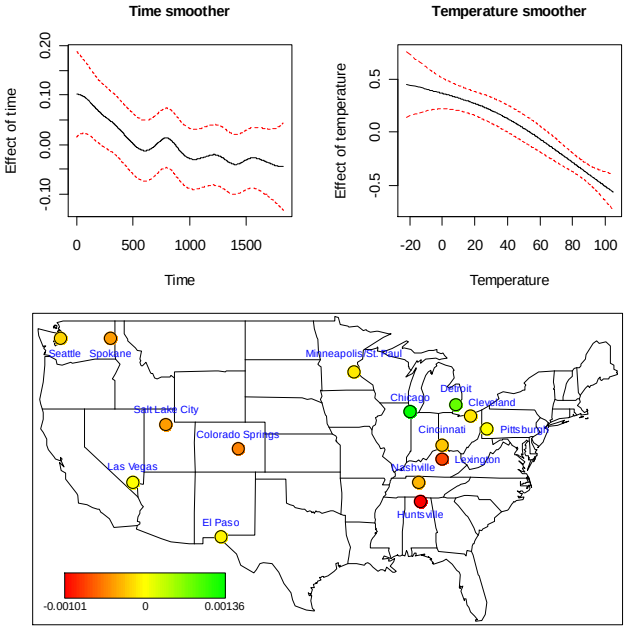
Pneumonia < 65 years old / Model: PM_{10}



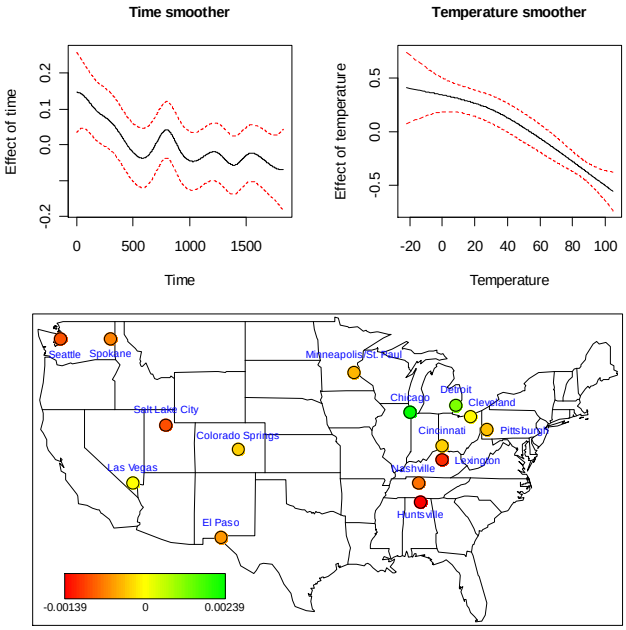
Pneumonia < 65 years old / Model: PM_{10}+lag



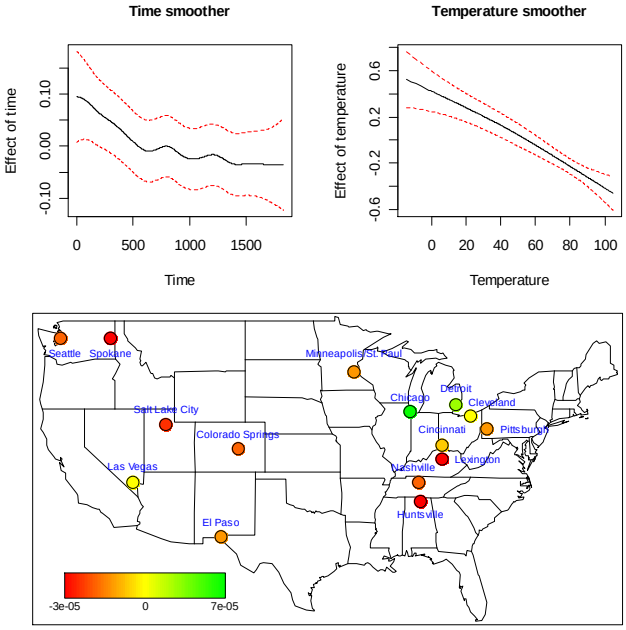
Pneumonia < 65 years old / Model: PM_{10}+CO



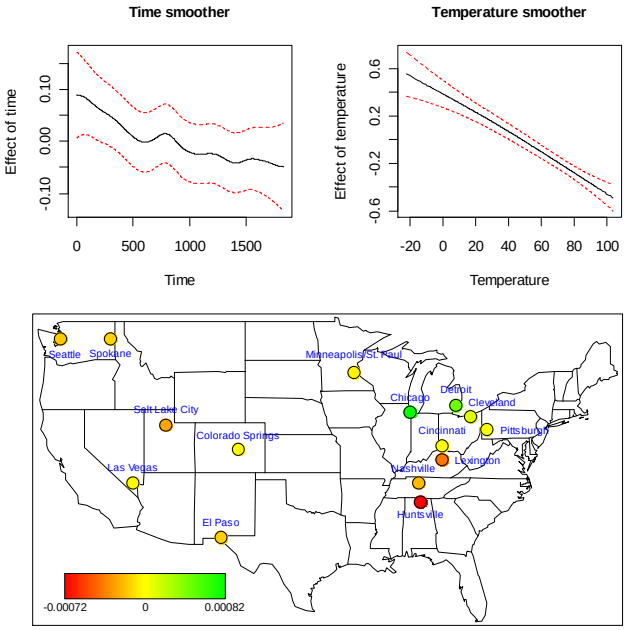
Pneumonia < 65 years old / Model: PM_{10}+NO_2



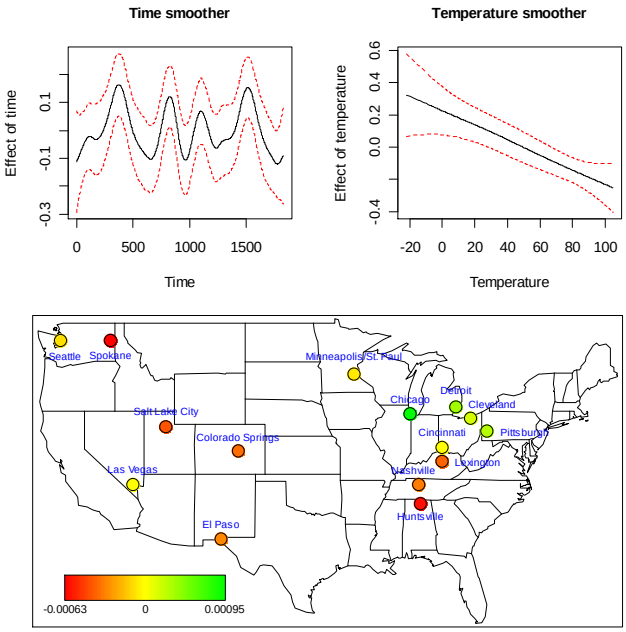
Pneumonia < 65 years old / Model: PM_{10}+O_3



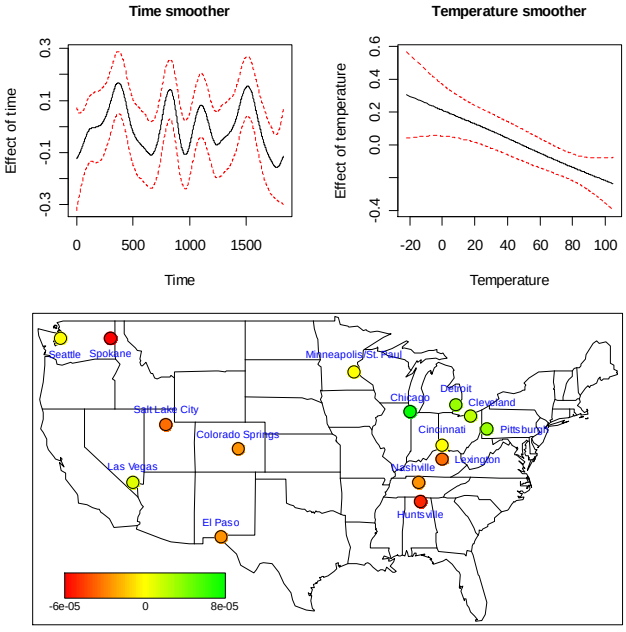
Pneumonia < 65 years old / Model: PM_{10}+SO_2



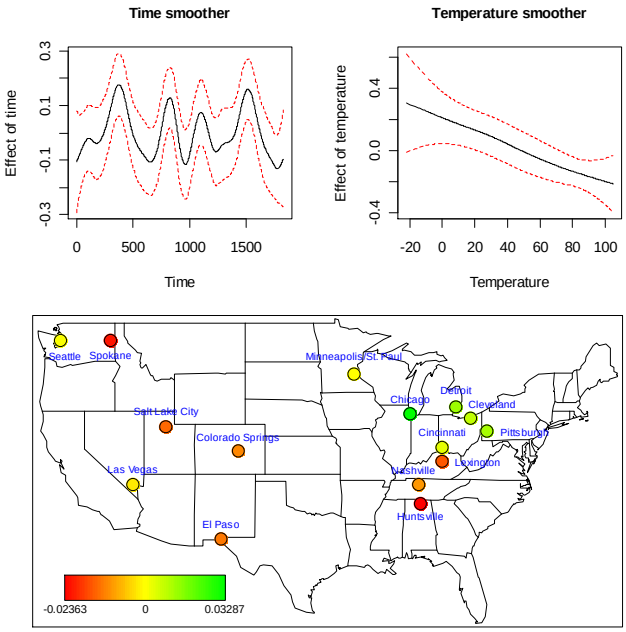
Pneumonia 65−74 years old / Model: PM_{10}



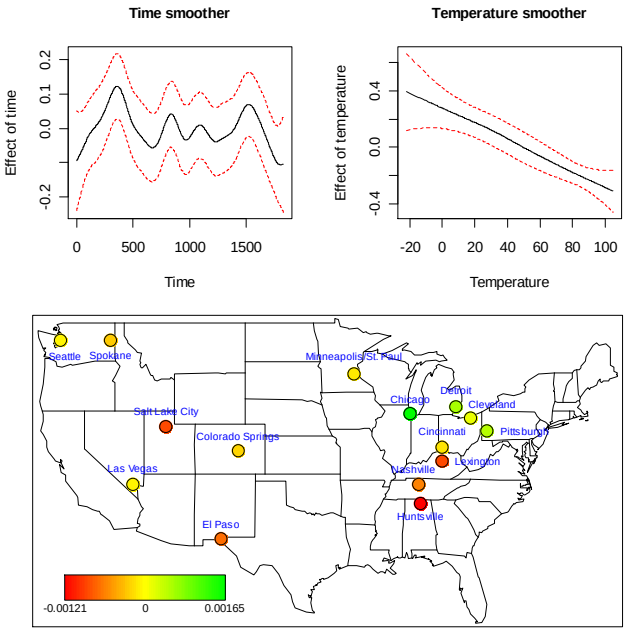
Pneumonia 65−74 years old / Model: PM_{10}+lag



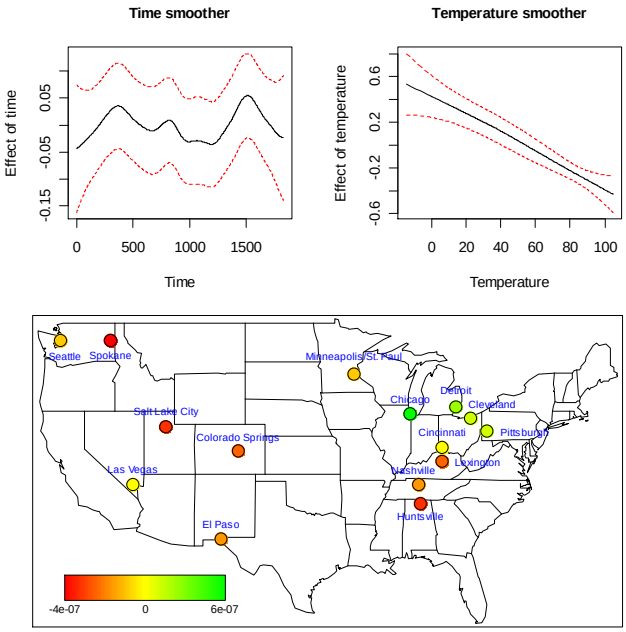
Pneumonia 65−74 years old / Model: PM_{10}+CO



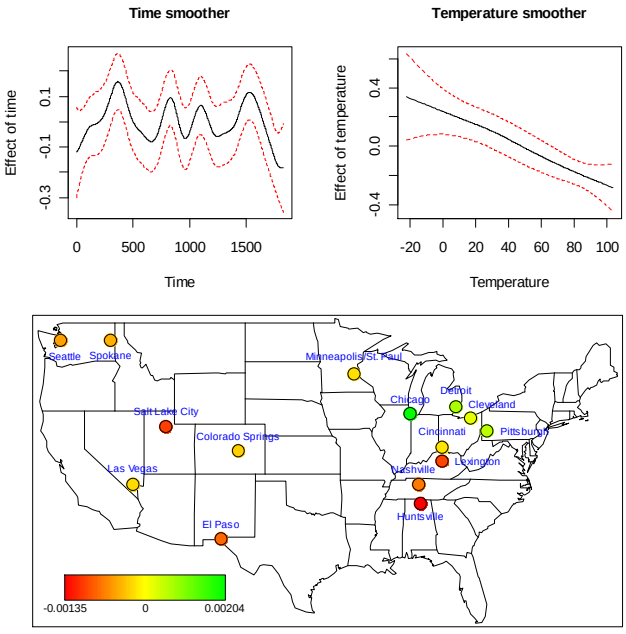
Pneumonia 65−74 years old / Model: PM_{10}+NO_2



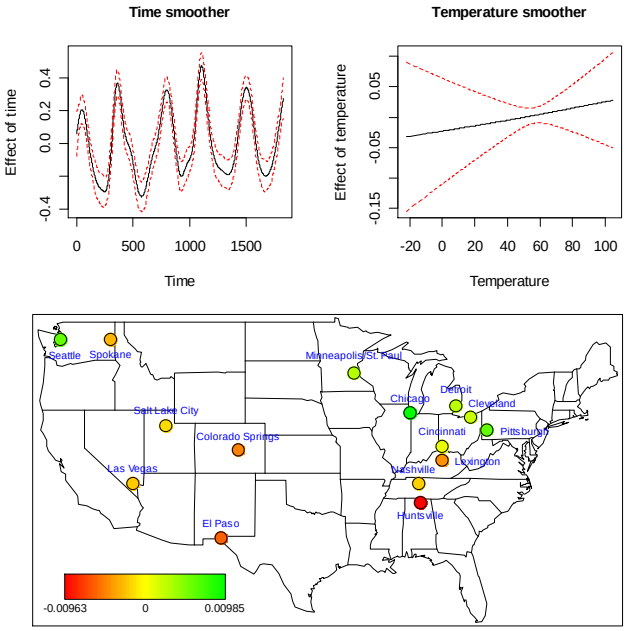
Pneumonia 65−74 years old / Model: PM_{10}+O_3



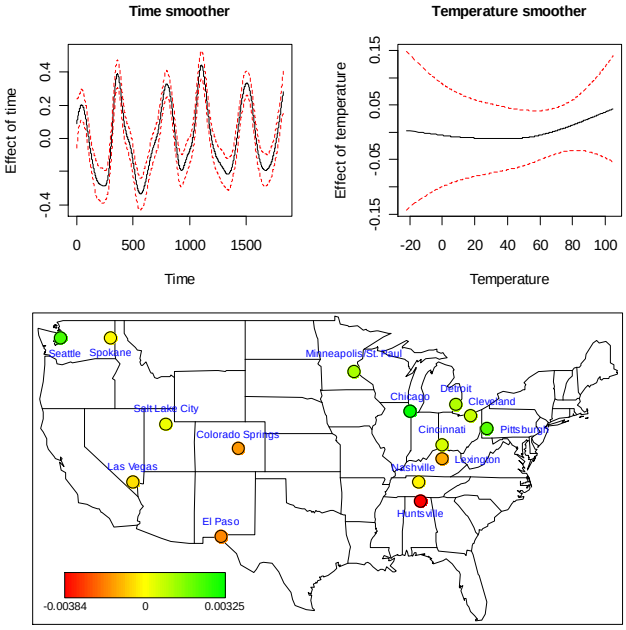
Pneumonia 65−74 years old / Model: PM_{10}+SO_2



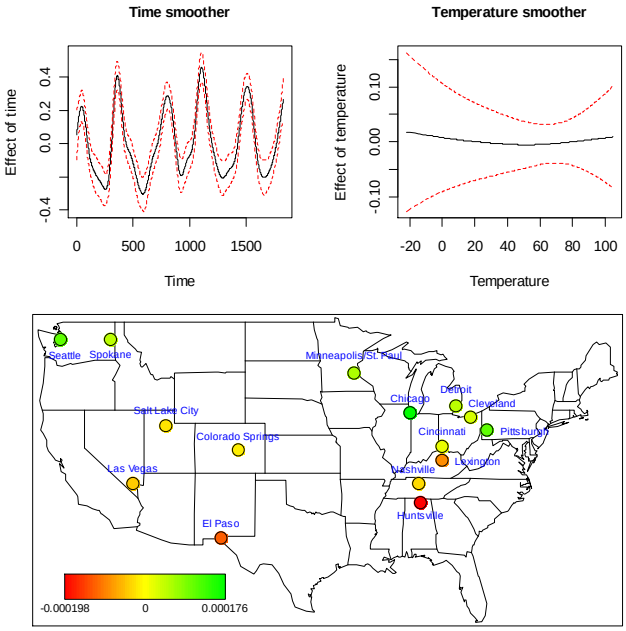
Pneumonia≥75 years old / Model: PM_{10}



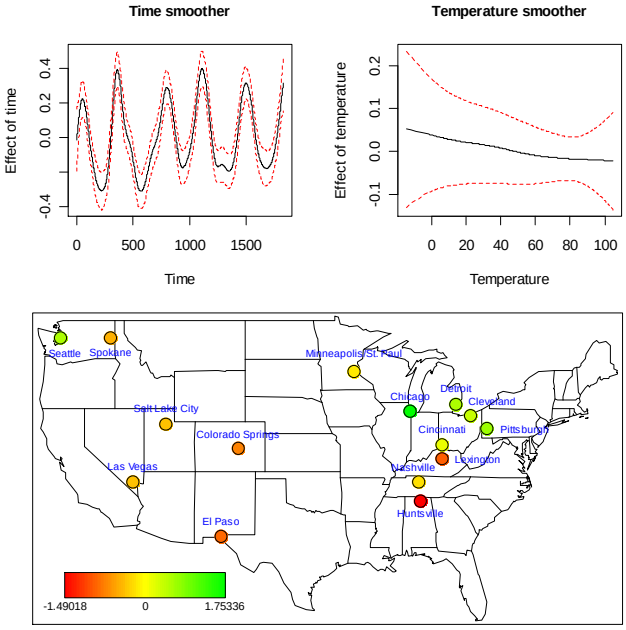
Pneumonia≥75 years old / Model: PM_{10}+lag



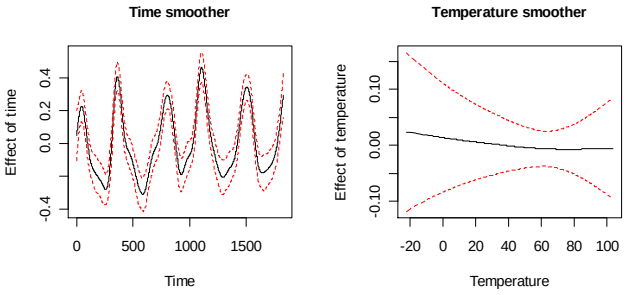
Pneumonia≥75 years old / Model: PM_{10}+NO_2

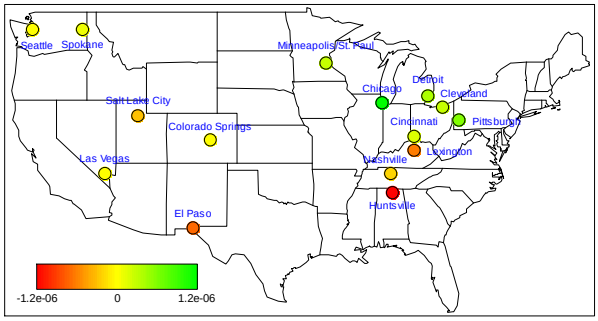


Pneumonia≥75 years old / Model: PM_{10}+O_3

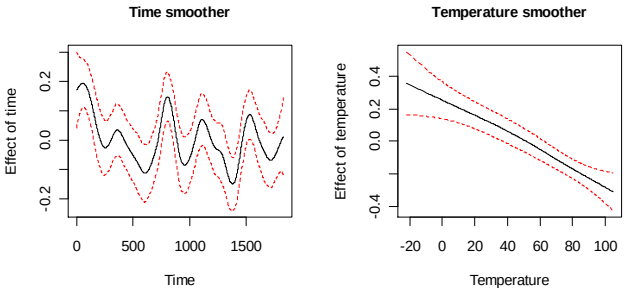


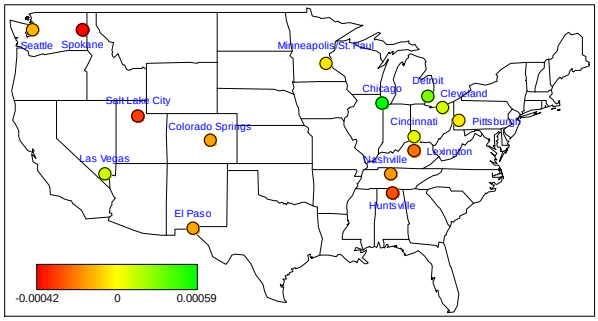
Pneumonia≥75 years old / Model: PM_{10}+SO_2



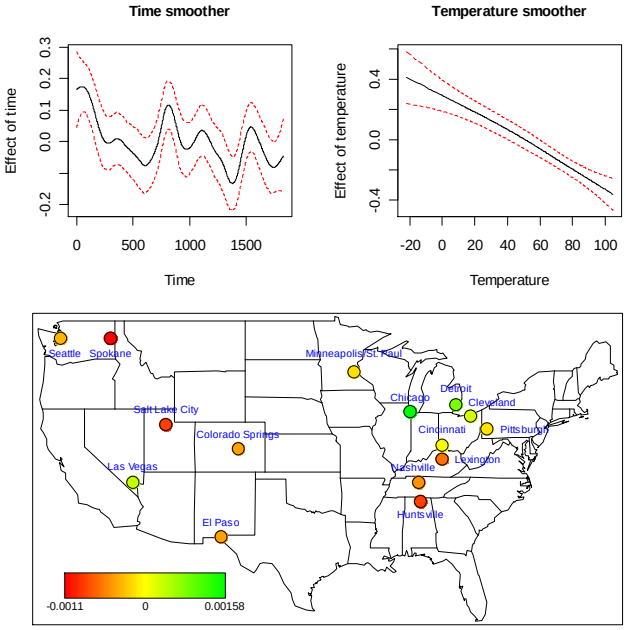


Respiratory disease＜65 years old / Model: PM_{10}

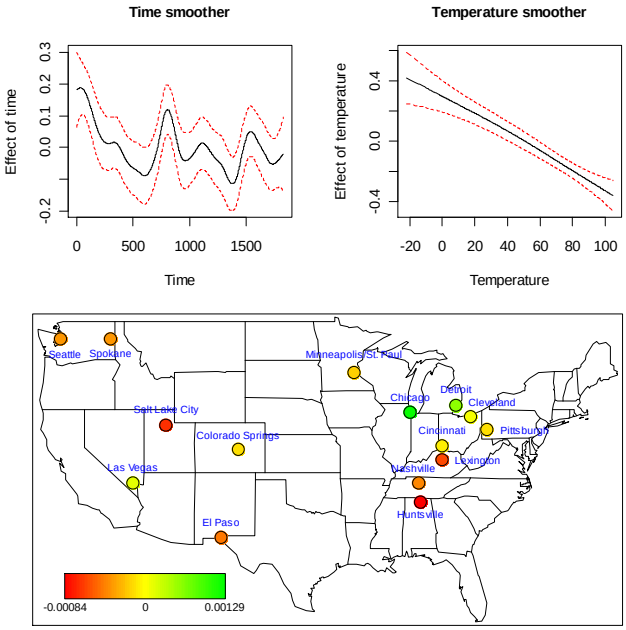




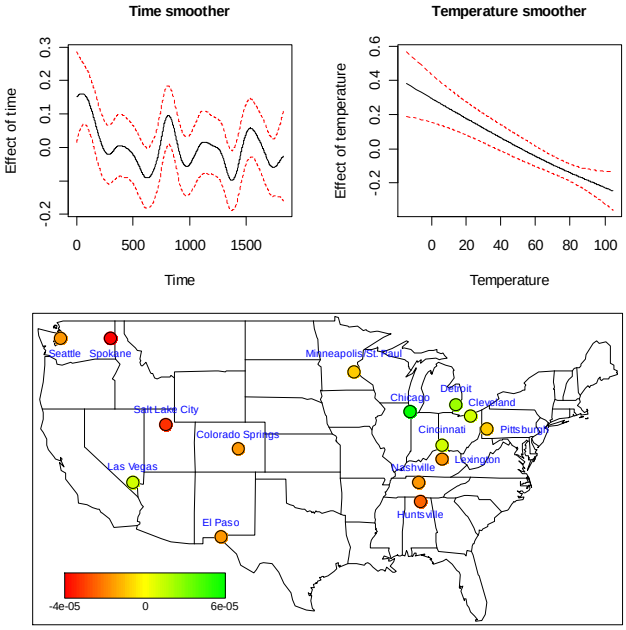
Respiratory disease＜65 years old / Model: PM_{10}+lag



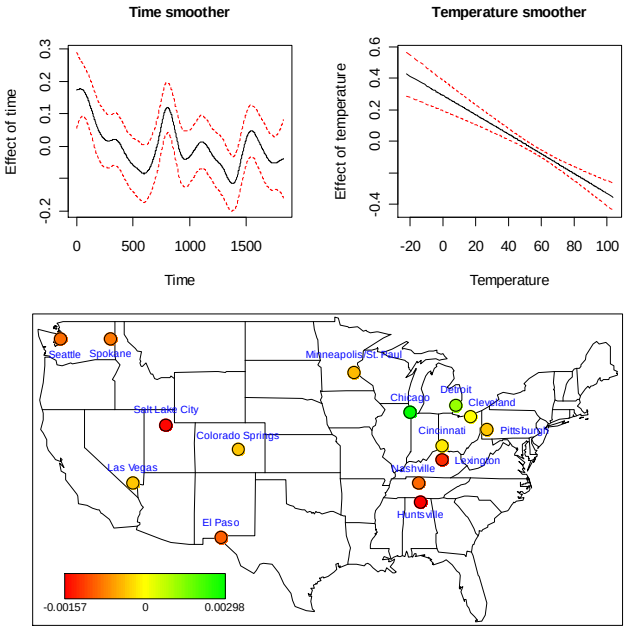
Respiratory disease＜65 years old / Model: PM_{10}+NO_2



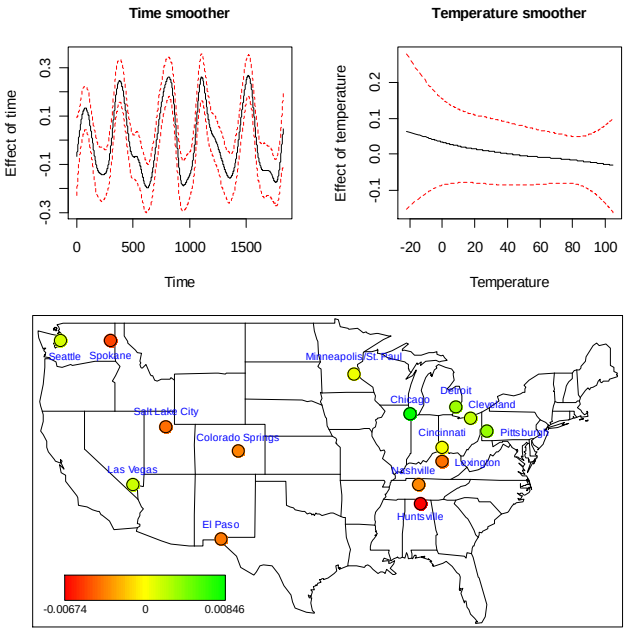
Respiratory disease＜65 years old / Model: PM_{10}+O_3



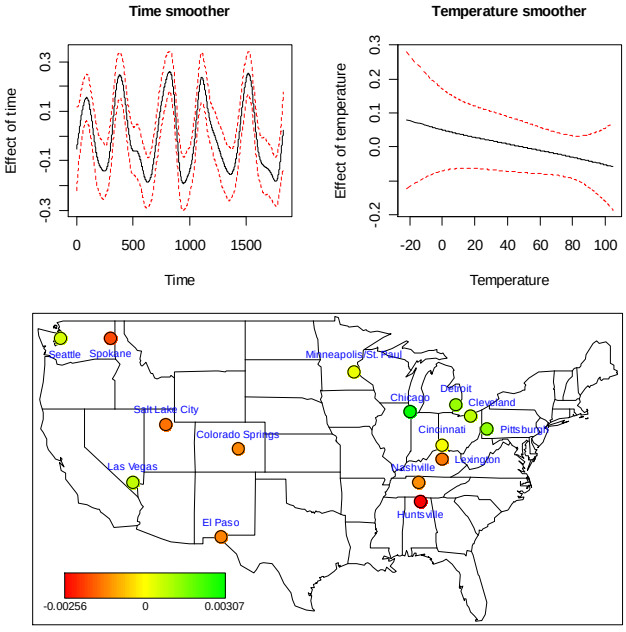
Respiratory disease＜65 years old / Model: PM_{10}+SO_2



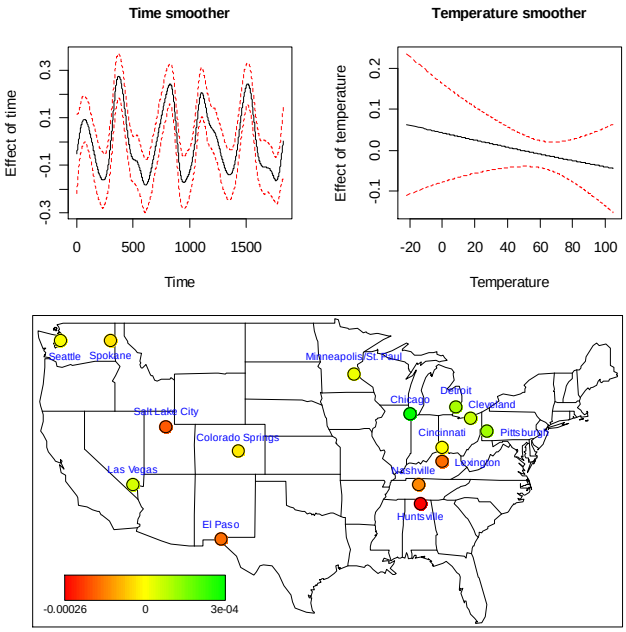
Respiratory disease 65−74 years old / Model: PM_{10}



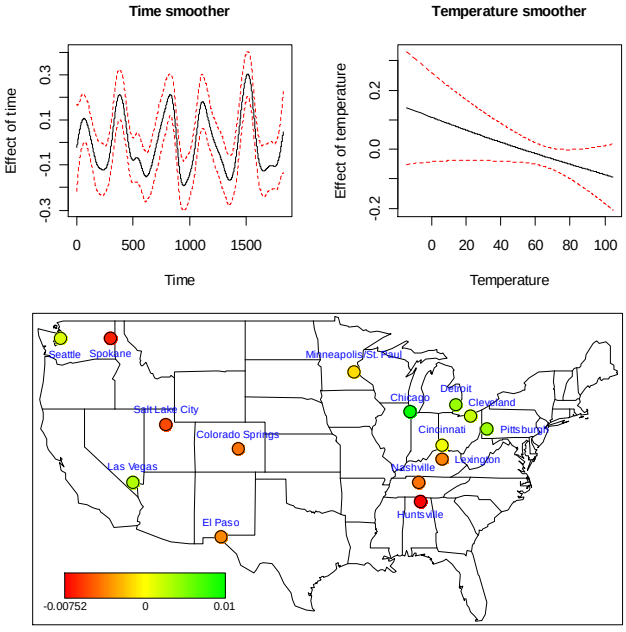
Respiratory disease 65−74 years old / Model: PM_{10}+lag



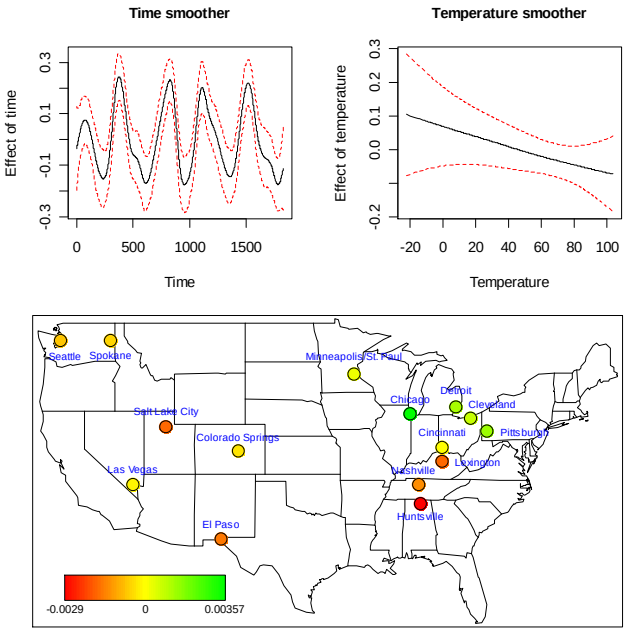
Respiratory disease 65−74 years old / Model: PM_{10}+NO_2



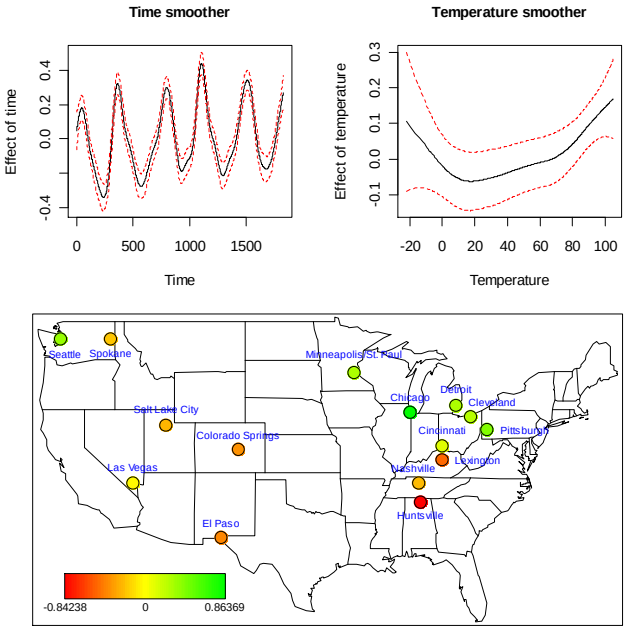
Respiratory disease 65−74 years old / Model: PM_{10}+O_3



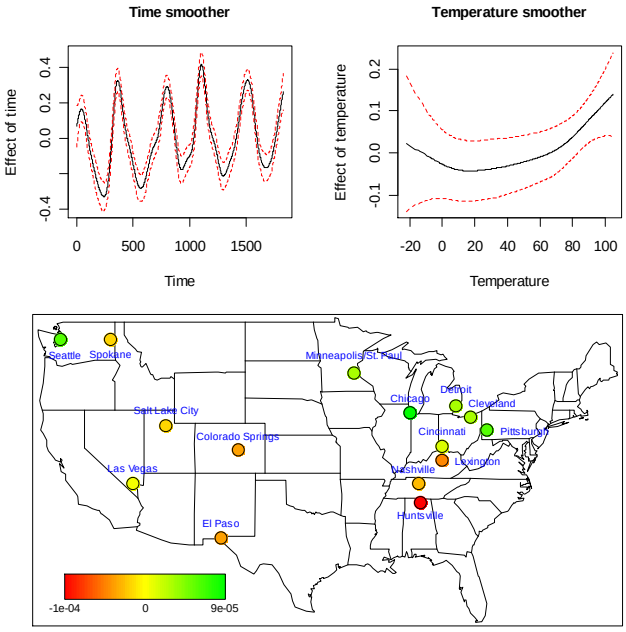
Respiratory disease 65−74 years old / Model: PM_{10}+SO_2



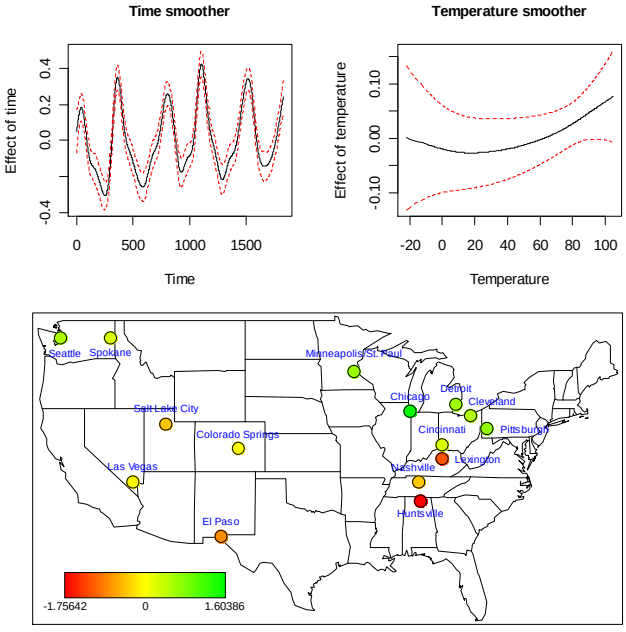
Respiratory disease≥75 years old / Model: PM_{10}



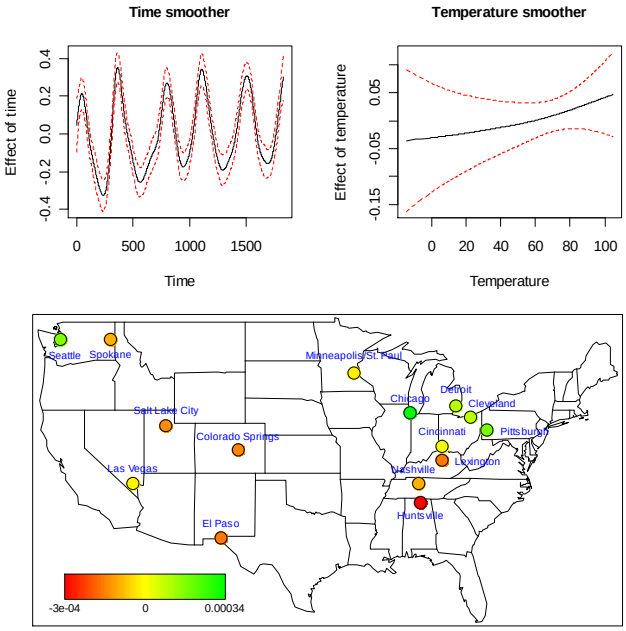
Respiratory disease≥75 years old / Model: PM_{10}+lag



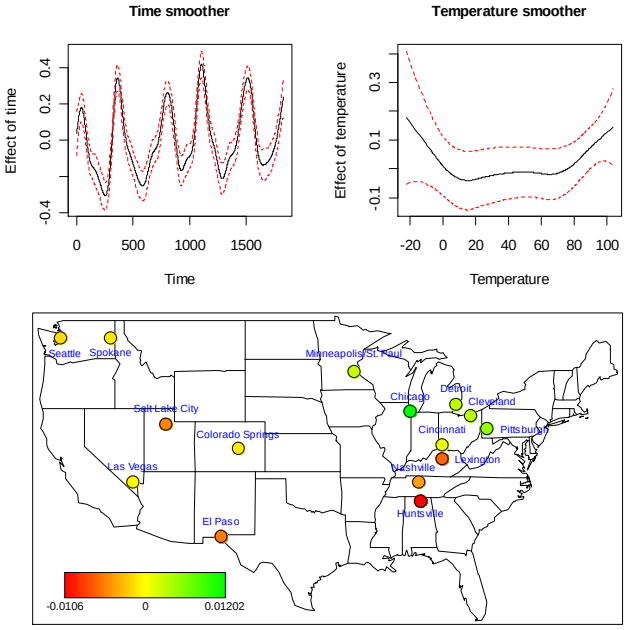
Respiratory disease≥75 years old / Model: PM_{10}+NO_2



Respiratory disease≥75 years old / Model: PM_{10}+O_3



Respiratory disease≥75 years old / Model: PM_{10}+SO_2



REFERENCE

Abbey, D. E., Nishino, N., McDonnell, W. F., Burchette, R. J., Knutsen, S. F., Lawrence, B. W., & Yang, J. X. (1999). Long-term inhalable particles and other air pollutants related to mortality in nonsmokers. *American Journal Respiratory and Critical Care Medicine*, **159(2)**, 373-382.

Ackermann-Liebrich, U., Leuenberger, P., Schwartz, J., Schindler, C., Monn, C., Bolognini, C., Bongard, J. P., Brandli, O., Domenighetti, G., Elsasser, S., Grize, L., Karrer, W., Keller, R., Keller-Wossidlo, H., Kunzli, N., Martin, B. W., Medici, T. C., Perruchoud, A. P., Schoni, M. H., Tschopp, J. M., Villiger, B., Wuthrich, B., Zellweger, J. P., & Zemp, E. (1997). Lung function and long term exposure to air pollutants in Switzerland. *American Journal of Respiratory and Critical Care Medicine*, **155(1),** 122-129.

Agresti, A. (1990). *Categorical Data Analysis*. New York: John Wiley & Sons, Inc.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19(6),** 716–723.

Almon, S. (1965). The distributed lag between capital appropriations and expenditures. *Econometrica*, **33(1),** 178-196.

Anderson, R., Atkinson, A., Peacock, J. L., Marston, L., & Konstantinou, K. (2004). Meta-analysis of time-series and panel studies on Particulate Matter and Ozone (O_3). *WHO Task Group. WHO Regional Office for Europe, Copenhagen 2004 (EUR/04/5042688)*. Retrieved from http://www.euro.who.int/document/E82792.pdf.

Anderson, T. W. (1984). *An introduction to multivariate statistical analysis* (Second Edition). New York: John Wiley & Sons, Inc.

Anselin, L. (1988). *Spatial Econometrics, methods and models.* Boston: Kluwer Academic.

Anselin, L., & Rey, S. (1997). Introduction to the special issue on spatial econometrics. *International Regional Science Review*, **20,** 1-7.

Anselin, L., Le Gallo, J., & Jayet, H. (2006). *The Econometrics of panel data, fundamentals and recent developments in theory and practice* (3rd ed.). Springer Berlin Heidelberg.

Arbia, G. (2006). *Spatial Econometrics: Statistical Foundations and Applications to Regional Convergence.* Berlin: Springer-Verlag.

Austin, H., Flanders, W. D., & Rothman, K. J. (1989). Bias arising in case-control studies from selection of controls from overlapping groups. *International Journal of Epidemiology*, **18(3),** 713-716.

Barnard, J., & Rubin, D. B. (1999). Small sample degrees of freedom with multiple

imputation. *Biometrika,* **86,** 948-955.

Barnett, A. G., Williams, G. M., Schwartz, J., Neller, A. H., Best, T. L., Petroeschevsky, A. L., & Simpson, R. W. (2005). Air pollution and child respiratory health: A case-crossover study in Australia and New Zealand. *American Journal of Respiratory and Critical Care Medicine, 171*(11), 1272-1278.

Bateson, T. F., & Schwartz, J. (1999). Control for seasonal variation and time trend in case-crossover studies of acute effects of environmental exposures. *Epidemiology,* **10(5),** 539-44.

Bell, M. L., McDermott, A., Zeger, S. L., Samet, J. M., & Dominici, F. (2004a). Ozone and short-term mortality in 95 US urban communities, 1987-2000. *Journal of the American Medical Association,* **292(19),** 2372-2378.

Bell, M. L., Samet, J. M., & Dominici, F. (2004b). Time-series studies of particulate matter. *Annual Review of Public Health*, **25,** 247-280.

Bell, M. L., Dominici, F., & Samet, J. M. (2005). A meta-analysis of time-series studies of ozone and mortality with comparison to the national morbidity, mortality, and air pollution study. *Epidemiology*, **16(4),** 436-45.

Bell, M. L., Peng, R. D., & Dominici, F. (2006). The exposure-response curve for ozone and risk of mortality and the adequacy of current ozone regulations. *Environmental Health Perspectives*, **114(4),** 532-536.

Bell, M. L., Kim, J. Y., & Dominici, F. (2007). Potential confounding of particulate matter on the short-term association between ozone and mortality in multisite time-series studies. *Environmental Health Perspectives,* **115(11),** 1591-1595.

Berhane, K., & Tibshirani, R.J. (1996). Generalized additive models for longitudinal data. *The Canadian Journal of Statistics,* **26(4),** 517-535.

Besag, J. (1974). Spatial interaction and statistical-analysis of lattice systems. *Journal of the Royal Statistical Society Series B-Methodological*, **36(2),** 192-236.

Besag, J., & Kooperberg, C. (1995). On conditional and intrinsic autoregressions. *Biometrika*, **82,** 733-746.

Bloomfield, P. (1976). *Fourier analysis of time series: an introduction*. New York: John Wiley and Sons, Inc.

Braun-Fahrlander, C., Vuille, J. C., Sennhauser, F. H., Neu, U., Kunzle, T., Grize, L., Gassner, M., Minder, C., Schindler, C., Varonier, H. S., & Wuthrich, B. (1997). Respiratory health and long-term exposure to air pollutants in Swiss schoolchildren. *American Journal of Respiratory and Critical Care Medicine*, **155(3),** 1042-1049.

Breslow, N. E., & Day, N. E. (1980). Statistical methods in cancer research. Volume I - The

analysis of case-control studies. *IARC Scientific Publication,* **32,** 335-338.

Breslow, N. E., & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association,* **88,** 9-25.

Brezger, A., Kneib, T., & Lang, S. (2005). BayesX: Analyzing Bayesian structured additive regression models. *Journal of Statistical Software,* **14(11).** Retrieved from http://www.jstatsoft.org/v14/i11/paper.

Brimblecombe, P. (1987). *The big smoke: A history of air pollution in London since medieval times.* New York: Methuen.

Buringh, E., Fischer, P., & Hoek, G. (2000). Is SO_2 a causative factor for the PM-associated mortality risks in the Netherlands? *Inhalation Toxicology,* **12(Supplement 1),** 55-60.

Carder, M., McNamee, R., Beverland, I., Elton, R., Van Tongeren, M., Cohen, G. R., Boyd, J., MacNee, W., & Agius, R. M. (2008). Interacting effects of particulate pollution and cold temperature on cardiorespiratory mortality in Scotland. *Occupational and Environmental Medicine*, **65(3),** 197-204.

Carroll, R. J., Ruppert, D., & Stefanski, L. (1995). *Measurement Error in Nonlinear Models.* London: Chapman & Hall.

Chan, C. C., Chuang, K. J., Chien, L. C., Chen, W. J., & Chang, W. T. (2006). Urban air pollution and emergency admissions for cerebrovascular diseases in Taipei, Taiwan. *European Heart Journal,* **27,** 1238-1244.

Chang, C. C., Hwang, J. S., Chan, C. C., Wang, P. Y., & Cheng, T. J. (2007). Effects of concentrated ambient particles on heart rate, blood pressure, and cardiac contractility in spontaneously hypertensive rats during a dust storm event. *Inhalation Toxicology*, **19(11),** 973-978.

Chen, C., Chock, D. P., & Winkler, S. L. (1999). A simulation study of confounding in generalized linear models for air pollution epidemiology. *Environmental Health Perspectives,* **107(3),** 217-222.

Chen, J., & Shao, J. (2000). Nearest Neighbor Imputation for Survey Data. *Journal of Official Statistics,* **16(2),** 113-131.

Chiles, J. P., & Delfiner, P. (1999). *Geostatistics. Modeling spatial uncertainty*. New York: John Wiley & Sons.

Chuang, K. J., Coull, B. A., Zanobetti, A., Suh, H., Schwartz, J., Stone, P. H., Litonjua, A., Speizer, F. E., & Gold, D. R. (2008). Particulate air pollution as a risk factor for st-segment depression in patients with coronary artery disease. *Circulation*, **118(13),** 1314-1320.

Ciocco, A., & Thompson, D. (1961). A follow-up of Donora ten years after: methodology and

findings. *American Journal of Public Health,* **51,** 155-164.

Clayton, D., & Hills, M. (1993). *Statistical Models in Epidemiology*. Oxford: Clarendon Press.

Cliff, A., & Ord, J. K. (1972). Testing for spatial autocorrelation among regression residuals. *Geographical Analysis,* **4(3),** 267-284.

Cliff, A., & Ord, J. K. (1973). *Spatial Autocorrelation.* London: Pion.

Coull, B. A., Schwartz, J., & Wand, M. P. (2001). Respiratory health and air pollution: Additive mixed model analyses. *Biostatistics,* **2(3),** 337-49.

Cox, D. R., & Oakes, D. O. (1984). *Analysis of Survival Data*. New York: Chapman & Hall.

Crainiceanu, C., Ruppert, D., Claeskens, G., & Wand, M. P. (2005). Exact likelihood ratio tests for penalized splines. *Biometrika,* **92(1),** 91-103.

Cressie, N. (1993). *Statistics for Spatial Data.* New York: Wiley.

Dab, W., Medina, S., Quenel, P., Le Moullec, Y., Le Tertre, A., Thelot, B., Monteil, C., Lameloise, P., Pirard, P., Momas, I., Ferry, R., & Festy, B. (1996). Short term respiratory health effects of ambient air pollution: results of the APHEA project in Paris. *Journal of Epidemiology and Community Health***, 50,** S42-S46.

Daniels, M. J., Dominici, F., Samet, J. M., & Zeger, S. L. (2000). Estimating particulate matter-mortality dose-response curves and threshold levels: An analysis of daily time-series for the 20 largest US cities. *American Journal of Epidemiology*, **152(5),** 397-406.

Davis, D. L. (2002). *When Smoke Ran Like Water: Tales of Environmental Deception and the Battle Against Pollution*. New York: Basic Books.

Diggle, P. J., Liang, K. Y., & Zeger, S. L. (1994). *Analysis of Longitudinal Data*. Oxford: Clarendon Press.

Dockery, D. W., Schwartz, J., & Spengler, J. D. (1992). Air-pollution and daily mortality - associations with particulates and acid aerosols. *Environmental Research,* **59(2),** 362-373.

Dockery, D. W., Pope, C. A., Xu, X. P., Spengler, J. D., Ware, J. H., Fay, M. E., Ferris, B. G., & Speizer, F. E. (1993). An association between air-pollution and mortality in six US cities. *New England Journal of Medicine*, **329(24),** 1753-1759.

Dominici, F., Samet, J. M., & Zeger, S. L. (2000a). Combining evidence on air pollution and daily mortality from the 20 largest US cities: A hierarchical modelling strategy. *Journal of the Royal Statistical Society Series a-Statistics in Society*, **163,** 263-284.

Dominici, F., Zeger S. L., & Samet J. M. (2000b). A measurement error model for time-series studies of air pollution and mortality. *Biostatistics*, **2,** 157-175.

Dominici, F., Daniels, M., Zeger, S. L., & Samet, J. M. (2002a). Air pollution and mortality: Estimating regional and national dose-response relationships. *Journal of the American Statistical Association*, **97(457),** 100-111.

Dominici, F., McDermott, A., Zeger, S. L., & Samet, J. M. (2002b). On the use of generalized additive models in time-series studies of air pollution and health. *American Journal of Epidemiology*, **156(3),** 193-203.

Dominici, F., McDermott, A., Zeger, S. L., & Samet, J. M. (2003a). Airborne particulate matter and mortality: Timescale effects in four US cities. *American Journal of Epidemiology*, **157(12),** 1055-1065.

Dominici, F., McDermott, A., Zeger, S. L., & Samet, J. M. (2003b). National maps of the effects of particulate matter on mortality: Exploring geographical variation. *Environmental Health Perspectives*, **111(1),** 39-43.

Dominici, F., Sheppard, L., & Clyde, M. (2003c). Health effects of air pollution: A statistical review. *International Statistical Review*, **71(2),** 243-276.

Dominici, F., McDermott, A., Daniels, M., Zeger, S. L., & Samet, J. M. (2005). Revised analyses of the national morbidity, mortality, and air pollution study: Mortality among residents of 90 cities. *Journal of Toxicology and Environmental Health-Part a-Current Issues*, **68(13-14),** 1071-1092.

Durrant, G. B. (2005). A semi-parametric multiple imputation data augmentation procedure. *Proceedings of the Survey Research Methods Section, American Statistical Association. American Statistical Association.* Retrieved from http://www.amstat.org/Sections/Srms/Proceedings/y2005/Files/JSM2005-000425.pdf.

Elhorst, J. P. (2003). Specification and estimation of spatial panel data models. *International Regional Science Review*, **26(3),** 244-268.

Everson, P. J., & Morris, C. N. (2000). Inference for multivariate normal hierarchical models. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, **62,** 399-412.

Fahrmeir, L., & Lang, S. (2001a). Bayesian inference for generalized additive mixed models based on Markov random field priors. *Journal of the Royal Statistical Society Series C-Applied Statistics*, **50,** 201-220.

Fahrmeir, L., & Lang, S. (2001b). Bayesian semiparametric regression analysis of multicategorical time-space data. *Annals of the Institute of Statistical Mathematics*, **53,** 10-30.

Fahrmeir, L., Lang, S., & Spies, F. (2003). Generalized geoadditive models for insurance

claims data. *Blätter der DGVFM,* **26(1),** 7-23.

Fahrmeir, L., Kneib, T., & Lang, S. (2004). Penalized structured additive regression for space-time data: A Bayesian perspective. *Statistica Sinica,* **14(3),** 731-761.

Fairley, D. (1990). The relationship of daily mortality to suspended particulates in Santa Clara County 1980-1986. *Environmental Health Perspectives,* **89,** 159-168

Fanger, P. O. (1970). *Thermal Comfort*. Copenhagen: Danish Technical Press.

Farrar, D. E., & Glauber, R. R. (1967). Multicollinearity in regression analysis - problem revisited. *Review of Economics and Statistics,* **49(1),** 92-107.

Figueiras, A., Roca-Pardiñas, J., & Cadarso-Suárez, C. (2003). Avoiding the effect of concurvity in generalized additive models in time-series studies of air pollution. *The ISI International Conference on Environmental Statistics and Health*. Available from http://isi-eh.usc.es/trabajos/i10_70_fullpaper.pdf.

Figueiras, A., Roca-Pardiñas, J., & Cadarso-Suárez, C. (2005). A bootstrap method to avoid the effect of concurvity in generalised additive models in time series studies of air pollution. *Journal of Epidemiology and Community Health*, **59(10),** 881-884.

Firket J. (1936). Fog along the Meuse Valley. *Transactions of the Faraday Society,* **32,** 1192-1197.

Fleming, M. (2002). *Techniques for estimating spatially dependent discrete choice models.* In Anselin, L., Florax, R. J., and Rey, S. J., editors, Advances in Spatial Econometrics. Springer-Verlag, Heidelberg. Forthcoming.

Florax, R., & Van der Vlist, A. J. (2003). Spatial econometric data analysis: Moving beyond traditional models. *International Regional Science Review,* **26(3),** 223-243.

Forsberg, B., Stjernberg, N., & Wall, S. (1997). Prevalence of respiratory and hyperreactivity symptoms in relation to levels of criteria air pollutants in Sweden. *European Journal of Public Health,* **7(3),** 291-296.

Fortin, M. J., & Dale, M. (2005). *Spatial Analysis: A Guide for Ecologists.* Cambridge: Cambridge University Press.

Fronk, E. M., & Giudici, P. (2004). Markov chain Monte Carlo model selection for DAG models. *Statistical Methods and Applications*, **13,** 259-273

Fung, W.-K., Zhu, Z.-Y., Wei, B.-C., & He, X. (2002). Influence diagnostics and outlier tests for semiparametric mixed models. *Journal of the Royal Statistical Society,Series B,* **64,** 565-579.

Gamerman, D. (1997). *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*. New York: Chapman & Hall.

Gauderman, W. J., McConnell, R., Gilliland, F., London, S., Thomas, D., Avol, E., Vora, H., Berhane, K., Rappaport, E. B., Lurmann, F., Margolis, H. G., & Peters, J. (2000). Association between air pollution and lung function growth in southern California children. *American Journal of Respiratory and Critical Care Medicine*, **162(4),** 1383-1390.

Gauderman, W. J., Gilliland, G. F., Vora, H., Avol, E., Stram, D., McConnell, R., Thomas, D., Lurmann, F., Margolis, H. G., Rappaport, E. B., Berhane, K., & Peters, J. M. (2002). Association between air pollution and lung function growth in southern California children - results from a second cohort. *American Journal of Respiratory and Critical Care Medicine*, **166(1),** 76-84.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). *Bayesian Data Analysis.* London: Chapman & Hall.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian Data Analysis.* Second Edition. Boca Raton: Chapman & Hall/CRC.

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6,** 721-741.

Goodchild, M. F., Anselin, L., Appelbaum, R. P., & Harthorn, B. H. (2000). Toward spatially integrated social science. *International Regional Science Review,* **23(2),** 139-159.

Gotway, C. A., & Stroup, W. W. (1997). A generalized linear model approach to spatial data analysis and prediction. *Journal of Agricultural, Biological and Environmental Statistics,* **2(2),** 157-178.

Gotway, C. A., & Wolfinger, R. D. (2003). Spatial prediction of counts and rates. *Statistics in Medicine,* **22(9),** 1415-1432.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, **82(4),** 711-732.

Green, P. J., & Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models*. London: Chapman and Hall.

Greenland, S., & Morgenstern, H. (1989). Ecological bias, confounding, and effect modification. *International Journal of Epidemiology*, **18(1),** 269-274.

Greven, S., Küchenhoff, K., Koenig, W., Picciotto, S., Pekkanen, J., Bellander, T., Leonard, B. J., Chalamandaris, A., Kulmala, M., & Peters A. (2005). Statistical Aspects in Additive Mixed Models for the AIRGENE Study. Available from http://www.stat.uni-muenchen.de/~greven/presentations/mcv.pdf

Griffith, D. A. (1988). *Advanced Spatial Statistics.* Dor-drecht: Kluwer Academic.

Haining, R. (1990). *Spatial data analysis in the social and environmental sciences*. Cambridge: Cambridge University Press.

Hammersley, J. M., & Clifford, P. (1971). *Markov fields on finite graphs and lattices*. Unpublished.

Hart, J. D. (1991). Kernel regression estimation with time-series errors. *Journal of the Royal Statistical Society Series B-Methodological*, **53(1),** 173-187.

Hastie, T. J., & Tibshirani, R. J. (1990). *Generalized additive models.* New York: Chapman and Hall.

He, S., Mazumdar, S., & Arena, V. C. (2006). A comparative study of the use of gam and glm in air pollution research. *Environmetrics,* **17(1),** 81-93.

Heagerty, P. J., & Zeger, S. L. (1996). Marginal regression models for clustered ordinal measurements. *Journal of the American Statistical Association,* **91(435),** 1024-1036.

Huang, Y., Dominici, F., & Bell, M. L. (2005). Bayesian hierarchical distributed lag models for summer ozone exposure and cardio-respiratory mortality. *Environmetrics,* **16(5),** 547-562.

Ito, K., Thurston, G. D., & Silverman, R. A. (2007). Characterization of PM2.5, gaseous pollutants, and meteorological interactions in the context of time-series health effects models. *Journal of Exposure Science and Environmental Epidemiology*, **17,** S45-S60.

Jaakkola, J. J. K. (2003). Case-crossover design in air pollution epidemiology. *European Respiratory Journal*, **21,** 81S-85S.

Johnson, R. A., & Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis*. New Jersey: Prentice Hall.

Jolliffe, I. T. (2002). *Principal Component Analysis*. 2nd Edition. New York: Springer.

Journel, A. G., & Huijbregts, C. J. (1978). *Mining Geostatistics*. London: Academic Press.

Kammann, E. E., & Wand, M. P. (2003). Geoadditive models. *Journal of the Royal Statistical Society Series C-Applied Statistics*, **52,** 1-18.

Kandala, N. B. (2006). Bayesian geo-additive modelling of childhood morbidity in Malawi. *Applied Stochastic Models in Business and Industry,* **22(2),** 139-154.

Kandala, N. B., Ji, C., Stallard, N., Stranges, S., & Cappuccio, F. P. (2007). Spatial analysis of risk factors for childhood morbidity in Nigeria. *American Journal of Tropical Medicine and Hygiene,* **77(4),** 770-778.

Kandala, N. B., Ji, C., Cappuccio, P. F., & Stones, R. W. (2008). The epidemiology of HIV infection in Zambia. *Aids Care-Psychological and Socio-Medical Aspects of Aids/Hiv,*

20(7), 812-819.

Kaslow, R., Ostrow, D., Detels, R., Phair, J., Polk, F., & Rinaldo, C. (1987). The multicenter AIDS cohort study: rationale, organization and selected characteristics of participants. *American Journal of Epidemiology,* **126(2),** 310-318.

Katsouyanni, K., Touloumi, G., Spix, C., Schwartz, J., Balducci, F., Medina, S., Rossi, G., Wojtyniak, B., Sunyer, J., Bacharova, L., Schouten, J. P., Ponka, A., & Anderson, H. R. (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(7095),** 1658-1663.

Katsouyanni, K., Touloumi, G., Samoli, E., Gryparis, A., Le Tertre, A., Monopolis, Y., Rossi, G., Zmirou, D., Ballester, F., Boumghar, A., Anderson, H. R., Wojtyniak, B., Paldy, A., Braunstein, R., Pekkanen, J., Schindler, C., & Schwartz, J. (2001). Confounding and effect modification in the short-term effects of ambient particles on total mortality: Results from 29 European cities within the APHEA2 project. *Epidemiology*, **12(5),** 521-531.

Kim, C., Park, B. U., & Kim, W. (2002). Influence diagnostics in semiparametric regression models. *Statistics and Probability Letters,* **60(1),** 49-58.

Kindermann, R., & Snell, J. L. (1980). *Markov Random Fields and Their Applications*. American Mathematical Society. Retrieved from http://www.ams.org/online_bks/conm1/conm1-whole.pdf

Kneib, T. (2006a). *Mixed model based inference in structured additive regression. Retrieved from Munich University Library*.

Kneib, T. (2006b). Mixed model-based inference in geoadditive hazard regression for interval-censored survival times. *Computational Statistics & Data Analysis*, **51,** 777-792.

Kneib, T., & Fahrmeir, L. (2006). Structured additive regression for categorical space-time data: A mixed model approach. *Biometrics*, **62(1),** 109-118.

Kneib, T., & Hennerfeind, A. (2006). Bayesian semiparametric multi-state models. *Statistical Modelling*, **8(2),** 169-198.

Kneib, T., & Fahrmeir, L. (2007). A mixed model approach for geoadditive hazard regression. *Scandinavian Journal of Statistics*, **34(1),** 207-228.

Kraak, M.-J., & Ormeling, F. (1996). *Cartography: Visualization of Spatial Data.* New Jersey: Prentice Hall.

Kwon, H. J., Cho, S. H., Chun, Y., Lagarde, F., & Pershagen, G. (2002). Effects of the Asian dust events on daily mortality in Seoul, Korea. *Environmental Research,* **90(1),** 1-5.

Lagorio, S., Forastiere, F., Pistelli, R., Iavarone, I., Michelozzi, P., Fano, V., Marconi, A., Ziemacki, G., & Ostro, B. D. (2006). Air pollution and lung function among susceptible adult subjects: a panel study. *Environmental Health: A Global Access Science Source*, **5(11).** Retrieved from http://www.ehjournal.net/content/5/1/11.

Lang, S., & Brezger, A. (2004). Bayesian p-splines. *Journal of Computational and Graphical Statistics*, **13(1),** 183-212.

Le Tertre, A., Medina, S., Samoli, E., Forsberg, B., Michelozzi, P., Boumghar, A., Vonk, J. M., Bellini, A., Atkinson, R., Ayres, J. G., Sunyer, J., Schwartz, J., & Katsouyanni, K. (2002). Short-term effects of particulate air pollution on cardiovascular diseases in eight European cities. *Journal of Epidemiology and Community Health*, **56(10),** 773-779.

Lee, J. T., & Schwartz, J. (1999). Reanalysis of the effects of air pollution on daily mortality in Seoul, Korea: a case-crossover design. *Environmental Health Perspectives*, **107,** 633–636.

LeSage, J. P. (2000). Bayesian estimation of limited dependent variable spatial autoregressive models. *Geographical Analysis*, **32(1),** 19-35.

LeSage, J. P., Pace, R. K., & Tiefelsdorf, M. (2004). Methodological developments in spatial econometrics and statistics. *Geographical Analysis*, **36(2),** 87-89.

Levy, D., Lumley, T., Sheppard, L., Kaufman, J., & Checkoway, H. (2001a). Referent selection in case-crossover analyses of acute health effects of air pollution. *Epidemiology*, **12(2),** 186-192.

Levy, D., Sheppard, L., Checkoway, H., Kaufman, J., Lumley, T., Koenig, J., & Siscovick, D. (2001b). A case-crossover analysis of particulate matter air pollution and out-of-hospital primary cardiac arrest. *Epidemiology*, **12(2),** 193-199.

Levy, J. I., Carrothers, T. J., Tuomisto, J. T., Hammitt, J. K., & Evans, J. S. (2001). Assessing the public health benefits of reduced ozone concentrations. *Environmental Health Perspectives*, **109,** 1215-1226.

Li, Y. Z., & Roth, H. D. (1995). Daily mortality analysis by using different regression-models in Philadelphia county, 1973-1990. *Inhalation Toxicology*, **7(1),** 45-58.

Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data-analysis using generalized linear-models. *Biometrika*, **73(1),** 13-22.

Lin, X. H., & Zhang, D. W. (1999). Inference in generalized additive mixed models by using smoothing splines. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, **61,** 381-400.

Lindley, D. V., & Smith, A. F. M. (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society - Series B*, **34,** 1–41.

Lipfert, F. W., & Wyzga, R. E. (1995). Air-pollution and mortality - issues and uncertainties. *Journal of the Air & Waste Management Association,* **45(12),** 949-966.

Lipsitz, S. R., Zhao, L. P., & Molenberghs, G. (1998). A semiparametric method of multiple Imputation. *Journal of the Royal Statistical Society, Series B, Statistical Methodology,* **60(1),** 127-144.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical Analysis with Missing Data*. New York: John Wiley & Sons.

Logan, W. P. D. (1953). Mortality in London fog incident, 1952. *Lancet.* **264(1),** 336-338.

Lord, F. M. (1955). Estimation of parameters from incomplete data. *Journal of the American Statistical Association,* **50(271),** 870-876.

Lumley, T., & Levy, D. (2000). Bias in the case-crossover design: Implications for studies of air pollution. *Environmetrics*, **11(6),** 689-704.

Maclure, M. (1991). The case-crossover design - a method for studying transient effects on the risk of acute events. *American Journal of Epidemiology*, **133(2),** 144-153.

Marcus, A. H., & Kegler, S. R. (2001). Confounding in air pollution epidemiology: When does two-stage regression identify the problem? *Environmental Health Perspectives,* **109(12),** 1193-1196.

McBride, S. J., Clyde, M. A., & Marcus, A. (2002). Modeling Airborne Particulate Matter Concentrations by Combining Data from Multiple Monitors. *Technical report, Institute of Statistics and Decision Sciences.*

McCullagh, P. (1980). Regression-models for ordinal data. *Journal of the Royal Statistical Society Series B-Methodological*, **42(2),** 109-142.

McCullagh, P., & Nelder, J. A. (1983). *Generalized Linear Models*. London: Chapman and Hall.

Meng, X. L., & Rubin, D. B. (1992). Performing likelihood ratio tests with multiply-imputed data sets. *Biometrika*, **79(1),** 103-111.

Morlini, I. (2006). On Multicollinearity and Concurvity in Some Nonlinear Multivariate Models. *Statistical Methods and Applications,* **15,** 3-26.

Morris, C. N., & Normand, S. L. (1992). Hierarchical models for combining information and for meta-analysis. *Bayesian Statistics 4* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.). Oxford: University Press, 321–344.

Nadakavukaren, A. (2006). *Our Global Environment – A Health Perspective.* Illinois: Waveland Press.

Navidi, W. (1998). Bidirectional case-crossover designs for exposures with time trends. *Biometrics*, **54(2),** 596-605.

Navidi, W., & Weinhandl, E. (2002). Risk set sampling for case-crossover designs. *Epidemiology*, **13(1),** 100-105.

Neas, L. M., Schwartz, J., & Dockery, D. (1999). A case-crossover analysis of air pollution and mortality in Philadelphia. *Environmental Health Perspectives*, **107(8),** 629-631.

Nemery, B., Hoet, P. H. M., & Nemmar, A. (2001). The Meuse valley fog of 1930: An air pollution disaster. *Lancet*, **357(9257),** 704-708.

Ngo, L., & Wand, M. P. (2004). Smoothing with mixed model software. *Journal of Statistical Software, 9*(1). Retrieved from http://www.jstatsoft.org/v09/i01/paper.

Nittner, T. (2004). The additive model affected by missing completely at random in the covariate. *Computational Statistics*, **19,** 261-282.

Nychka, D., Haaland, P., O'connell, M., & Ellner, S. (1998). FUNFITS, data analysis and statistical tools for estimating functions. In Nychka, D., Piegorsch, W. W. and Cox, L. H. (eds), *Case Studies in Environmental Statistics*, New York: Springer.

Ord, K. (1975). Estimation methods for models of spatial interaction. *Journal of the American Statistical Association*, **70(349),** 120-126.

Ostro, B. (1993). The association of air-pollution and mortality - examining the case for inference. *Archives of Environmental Health*, **48(5),** 336-342.

Pace, R. K., Barry, R., & Sirmans, C. F. (1998). Spatial statistics and real estate. *Journal of Real Estate Finance and Economics*, **17(1),** 5-13.

Pace, R. K., & Lesage, J. P. (2004). Spatial statistics and real estate. *Journal of Real Estate Finance and Economics*, **29(2),** 147-148.

Patterson, H. D., & Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, **58,** 545-554.

Peng, R.D., & Welty, L. J. (2004). The NMMAPSdata Package. *R News*, *4*, 10–14. Retrieved from http://cran.r-project.org/doc/Rnews/.

Peng, R. D., Dominici, F., Pastor-Barriuso, R., Zeger, S. L., & Samet, J. M. (2005). Seasonal analyses of air pollution and mortality in 100 US cities. *American Journal of Epidemiology*, **161(6),** 585-594.

Pinkse, J., & Slade, M. E. (1998). Contracting in space: An application of spatial statistics to discrete-choice models. *Journal of Econometrics*, **85(1),** 125-154.

Plaia, A., & Bondi, A. L. (2006). Single imputation method of missing values in

environmental pollution data sets. *Atmospheric Environment,* **40(38),** 7316-7330.

Pope, C. A., Dockery, D. W., Spengler, J. D., & Raizenne, M. E. (1991). Respiratory health and PM_{10} pollution: a daily time series analysis. *American Review of Respiratory Disease,* **144,** 668-674.

Pope, C. A., Schwartz, J., & Ransom, M. R. (1992). Daily mortality and PM_{10} pollution in Utah valley. *Archives of Environmental Health,* **47(3),** 211-217.

Pope, C. A., Thun, M. J., Namboodiri, M. M., Dockery, D. W., Evans, J. S., Speizer, F. E., & Health, C. W., Jr. (1995). Particulate air pollution as a predictor of mortality in a prospective study of U.S. adults. *American Journal of Respiratory and Critical Care Medicine,* **151,** 669-674.

Preston, C. J. (1974). *Gibbs States on Countable Sets*. Cambridge: Cambridge University Press.

Priestley, M. B. (1981). *Spectral analysis and time series.* New York: Academic Press, Inc.

Rabl, A. (2006). Analysis of air pollution mortality in terms of life expectancy changes: relation between time series, intervention, and cohort studies. *Environmental Health: A Global Access Science Source,* **5(1),** 1-11.

Ramsay, T. O., Burnett, R. T., & Krewski, D. (2003a). The effect of concurvity in generalized additive models linking mortality to ambient particulate matter. *Epidemiology,* **14(1),** 18-23.

Ramsay, T. O., Burnett, R. T., & Krewski, D. (2003b). Exploring Bias in a Generalized Additive Model for Spatial Air Pollution Data. *Environmental Health Perspectives,* **111(10),** 1283-1288.

Rice, J. A., & Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society Series B-Methodological,* **53(1),** 233-243.

Roberts, S., & Martin, M. A. (2007). Methods for bias reduction in time-series studies of particulate matter air pollution and mortality. *Journal of Toxicology and Environmental Health-Part a-Current Issues,* **70(8),** 665-675.

Robins, J. M., & Wang, N. S. (2000). Inference for imputation estimators. *Biometrika*, **87(1),** 113-124.

Roholm, K. (1937). The Fog Disaster in the Meuse Valley, 1930: A Fluorine Intoxication. *The Journal of Industrial Hygiene and Toxicology,* **19,** 126-137

Rothman, K., & Greenland, S. (1998). *Modern Epidemiology* (2nd ed.). Philadelphia: Lippincott-Raven.

Rubin, D. B. (1978). Multiple imputation in sample surveys – a phenomenological Bayesian approach to nonresponse. Proceedding of the Section on Survey Research Methods. Alexandra, VA: American Statistical Association.

Rubin, D. B. (1980). Illustrating the use of multiple imputations to handle nonresponse in sample surveys, *42nd Session of the International Statistical Institute, 1979, Book 2*, 517-532.

Rubin, D. B. (1987). *Multiple Imputatin for Nonresponse in Surveys.* New York: Wiley.

Rue, H., & Follestad, T. (2003). *Gaussian markov random fields models with application to spatial statistics*. Report no. 5-2003. Trondheim, Norway, Norvegian University of Science and Technology.

Samet, J. M., Dominici, F., Curriero, F. C., Coursac, I., & Zeger, S. L. (2000a). Fine particulate air pollution and mortality in 20 US cities, 1987-1994. *New England Journal of Medicine,* **343(24),** 1742-1749.

Samet, J. M., Zeger, S. L., Dominici, F., Dockery, D., & Schwartz, J. (2000b). *The National Morbidity, Mortality, and Air Pollution Study Part I: Methods and Methodological Issues*. Health Effects Institute, Cambridge, MA.

Samoli, E., Touloumi, G., Schwartz, J., Anderson, H. R., Schindler, C., & Forsberg B. (2007). Short-term effects of carbon monoxide on mortality: an analysis within the APHEA project. *Environmental Health Perspectives,* **115,** 1578-1583.

Schlesselman, J. J. (1994). *Case Control Studies: Design, Conduct, Analysis.* New York: Oxford University Press.

Schwartz, J., & Marcus, A. (1990). Mortality and air pollution in London: A time series analysis. *American Journal of Epidemiology,* **131,** 185-194.

Schwartz, J. (1991). Particulate air-pollution and daily mortality in Detroit. *Environmental Research,* **56(2),** 204-213.

Schwartz, J., & Dockery D.W. (1992). Particular air pollution and daily mortality in Steubenville, Ohio. *American Journal of Epidemiology,* **135(1),** 12-19.

Schwartz, J. (1993). Air pollution and daily mortality in Birmingham, Alabama. *American Journal of Epidemiology,* **137,** 1136-1147.

Schwartz, J. (2000). The distributed lag between air pollution and daily deaths. *Epidemiology,* **11(3),** 320-326.

Schwartz, J. (2001). Is there harvesting in the association of airborne particles with daily deaths and hospital admissions? *Epidemiology,* **12(1),** 55-61.

Scott, J. A. (1963). The London fog of December, 1962. *Medical Officer,* **109,** 250-252.

Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data*. London: Chapman & Hall.

Sheppard, L., Prentice, R. L., & Rossing, M. A. (1996). Design considerations for estimation of exposure effects on disease risk, using aggregate data studies. *Statistics in Medicine*, **15(17-18),** 1849-1858.

Shima, M., & Adachi, M. (2000). Effect of outdoor and indoor nitrogen dioxide on respiratory symptoms in schoolchildren. *International Journal of Epidemiology*, **29(5),** 862-870.

Shrenk, H. H., Heimann, H., Clayton, G. D., Gafafer, W. M., & Wexler, H. (1949). *Air pollution in Donora, PA: Epidemiology of the unusual smog episode of October 1948.* Washington, DC: Public Health Service.

Singer, J. D., & Willett, J. B. (2003). *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence.* New York: Oxford University Press.

Slaughter, J. C., Lumley, T., Sheppard, L., Koenig, J. Q., & Shapiro, G. S. (2002). Effects of ambient air pollution on symptoms severity and medication usage in asthmatic children. *Annals of Allergy, Asthma, and Immunology*, **91,** 346-353.

Smith, R. L. (2007). Air pollution risk. Retrieved 05-17-07, from http://www.stat.unc.edu/postscript/rs/airpollutionrisk.pdf.

Spitzer, F. (1971). Markov random fields and Gibbs ensembles. *American Mathematical Monthly*, **78(2),** 142-154.

Stieb, D. M., Judek, S., & Burnett, R. T. (2002). Meta-analysis of time-series studies of air pollution and mortality: Effects of gases and particles and the influence of cause of death, age, and season. *Journal of the Air & Waste Management Association*, **52(4),** 470-484.

Stieb, D. M., Judek, S., & Burnett, R. T. (2003). Meta-analysis of time-series studies of air pollution and mortality: Update in relation to the use of generalized additive models. *Journal of the Air & Waste Management Association*, **53(3),** 258-261.

Stokes, M. E., Davis, C. S., & Koch, G. G. (2000). *Categorical Data Analysis Using The SAS System* (Second Ed.). Cary: SAS Inc.

Sunyer, J., Castellsague, J., Saez, M., Tobias, A., & Anto, J. (1996). Air pollution and mortality in Barcelona. *Journal of Epidemiology and Community Health*, **50,** S76-S80.

Tanner, M. A. (1991). *Tools for statistical inference-observed data and data augmentation methods lecture notes in Statistics 67*. New York: Springer.

Thurston, G. D., & Ito, K. (2001). Epidemiological studies of acute ozone exposures and mortality. *Journal of Exposure Analysis and Environmental Epidemiology*, **11(4),**

286-294.

Touloumi, G., Samoli, E., & Katsouyanni, K. (1996). Daily mortality and "winter type" air pollution in Athens, Greece - a time series analysis within the APHEA project. *Journal of Epidemiology and Community Health,* **50,** S47-S51.

Tsai, S. S., Goggins, W. B., Chiu, H. F., & Yang, C. Y. (2003). Evidence for an association between air pollution and daily stroke admissions in Kaohsiung, Taiwan. *Stroke*, **34(11),** 2612-2616.

Verbyla, A. P., Cullis, B. R., Kenward, M. G., & Welham, S. J. (1999). The analysis of designed experiments and longitudinal data by using smoothing splines (with discussion), *Applied Statistics,* **48,** 269-311.

Waller, L. A., & Gotway, C. A. (2004). *Applied spatial statistics for public health data.* Hoboken, NJ: John Wiley.

Wellenius, G. A., Schwartz, J., & Mittleman, M. A. (2005). Air pollution and hospital admissions for ischemic and hemorrhagic stroke among Medicare beneficiaries. *Stroke,* **36(12),** 2549-2553.

Welty, L. J., & Zeger, S. L. (2005). Are the acute effects of particulate matter on mortality in the national morbidity, mortality, and air pollution study the result of inadequate control for weather and season? A sensitivity analysis using flexible distributed lag models. *American Journal of Epidemiology*, **162(1),** 80-88.

Whittle, P. (1954). On stationary processes in the plane. *Biometrika*, **41,** 434-449.

Wild, C. J., & Yee, T. W. (1996). Additive extensions to generalized estimating equation methods. *Journal of the Royal Statistical Society Series B-Methodological,* **58(4),** 711-725.

Wong, C. M., & Kohn, R. (1996). A Bayesian approach to additive semiparametric regression. *Journal of Econometrics,* **74,** 209-235.

WHO. (2005). *Air quality guidelines. Global update 2005. Particulate matter, ozone, nitrogen dioxide and sulfur dioxide*. Geneva: World Health Organization. Retrieved from http://www.euro.who.int/Document/E90038.pdf.

Wood, S. N. (2006). Low-rank scale-invariant tensor product smooths for generalized additive mixed models. *Biometrics*, **62(4),** 1025-1036.

Zanobetti, A., Wand, M., Schwartz, J., & Ryan, L. (2000). Generalized additive distributed lag models. *Biostatistics,* **1,** 279-292.

Zanobetti, A., Schwartz, J., Samoli, E., Gryparis, A., Touloumi, G., Atkinson, R., Le Tertre, A., Bobros, J., Celko, M., Goren, A., Forsberg, B., Michelozzi, P., Rabczenko, D., Ruiz, E. A., & Katsouyanni, K. (2002). The Temporal Pattern of Mortality Responses

to Air Pollution. *Epidemiology,* **13,** 87-93.

Zeger, S. L., Liang, K. Y., & Albert, P. S. (1988). Models for longitudinal data - a generalized estimating equation approach. *Biometrics*, **44(4),** 1049-1060.

Zeger, S. L., & Diggle, P. J. (1994). Semiparametric models for longitudinal data with application to cd4 cell numbers in HIV seroconverters. *Biometrics*, **50(3),** 689-699.

Zeger, S. L., Dominici, F., & Samet, J. (1999). Harvesting-resistant estimates of air pollution effects on mortality. *Epidemiology*, **10(2),** 171-175.

Zeger, S. L., Thomas, D., Dominici, F., Samet, J. M., Schwartz, J., Dockery, D., & Cohen, A. (2000). Exposure measurement error in time-series studies of air pollution: Concepts and consequences. *Environmental Health Perspectives*, **108(5),** 419-426.

Zemp, E., Elsasser, S., Schindler, C., Kunzli, N., Perruchoud, A. P., Domenighetti, G., Medici, T., Ackermann-Liebrich, U., Leuenberger, P., Monn, C., Bolognini, G., Bongard, J. P., Brandli, O., Karrer, W., Keller, R., Schoni, M. H., Tschopp, J. M., Villiger, B., Zellweger, J. P., & Team, S. (1999). Long-term ambient air pollution and respiratory symptoms in adults (sapaldia study). *American Journal of Respiratory and Critical Care Medicine*, **159(4),** 1257-1266.

Zhang, D., Lin, X. H., Raz, J., & Sowers, M. F. (1998). Semiparametric stochastic mixed models for longitudinal data. *Journal of the American Statistical Association*, **93(442),** 710-719.

Zhang, D., & Lin, X. (1999). Inference in generalized additive mixed models by using smoothing splines. Journal *of Royal Statistical Society B,* **61,** 381-400.

Zhang, D., & Davidian, M. (2004). Likelihood and conditional likelihood inference for generalized additive mixed models for clustered data. *Journal of Multivariate Analysis*, **91(1),** 90-106.

Zhang, H. (2002). On estimation and prediction for spatial generalized linear mixed models. *Biometrics*, **58(1),** 129-136.

Zidek, J. V., Wong, H., Le, N. D., & Burnett, R. (1996). Causality, measurement error and multicollinearity in epidemiology. *Environmetrics*, **7(4),** 441-451.

Zmirou, D., Barumandzadeh, T., Balducci, F., Ritter, P., Laham, G., & Ghilardi, J. P. (1996). Short term effects of air pollution on mortality in the city of Lyon, France, 1985-90. *Journal of Epidemiology and Community Health*, **50,** S30-S35.