
Contents

Preface	V
1 Studies of Air Pollution and Health	1
1.1 Introduction	1
1.2 Time Series Studies	2
1.3 Case-Crossover Studies	2
1.4 Panel Studies	3
1.5 Cohort Studies	4
1.6 Design Comparisons	5
2 Introduction to R and Air Pollution and Health Data	7
2.1 Starting Up R	7
2.2 The National Morbidity, Mortality, and Air Pollution Study ..	9
2.3 Organization of the NMMAPlite Package	9
2.3.1 Reading city-specific data	10
2.3.2 Pollutant data detrending	11
2.3.3 Mortality age categories	12
2.3.4 Metadata	13
2.3.5 Configuration options	14
2.4 MCAPS Data	14
3 Reproducible Research Tools	19
3.1 Introduction	19
3.2 Distributing Reproducible Research	20
3.3 Getting Started	21
3.4 Exploring a Cached Analysis	22
3.5 Verifying a Cached Analysis	25
3.6 Caching a Statistical Analysis	28
3.7 Distributing a Cached Analysis	29
3.8 Summary	30

4	Statistical Issues in Estimating the Health Effects of Spatial–Temporal Environmental Exposures	31
4.1	Introduction	31
4.2	Time-Varying Environmental Exposures	32
4.3	Estimation Versus Prediction	33
4.4	Semiparametric Models	35
4.4.1	Overdispersion	36
4.4.2	Representations for f	36
4.4.3	Estimation of β	37
4.4.4	Choosing the degrees of freedom for f	38
4.5	Combining Information and Hierarchical Models	39
5	Exploratory Data Analyses	41
5.1	Introduction	41
5.2	Exploring the Data: Basic Features and Properties	41
5.2.1	Pollutant data	41
5.2.2	Mortality data	46
5.3	Exploratory Statistical Analysis	50
5.3.1	Timescale decompositions	50
5.3.2	Example: Timescale decompositions of PM_{10} and mortality	51
5.3.3	Correlation at different timescales: A look at the Chicago data	53
5.3.4	Looking at more detailed timescales	57
5.4	Exploring the Potential for Confounding Bias	60
5.5	Summary	65
5.6	Reproducibility Package	65
5.7	Problems	65
6	Statistical Models	69
6.1	Introduction	69
6.2	Models for Air Pollution and Health	69
6.3	Semiparametric Models	71
6.3.1	GAMs in R	73
6.4	Pollutants: The Exposure of Interest	73
6.4.1	Single versus distributed lag	74
6.4.2	Mortality displacement	77
6.5	Modeling Measured Confounders	77
6.6	Accounting for Unmeasured Confounders	82
6.6.1	Using GAMs for air pollution and health	84
6.6.2	Computing standard errors for parametric terms in GAMs	88
6.6.3	Choosing degrees of freedom from the data	88
6.6.4	Example: Semiparametric model for Detroit	90
6.6.5	Smoothers	92

6.7	Multisite Studies: Putting It All Together	93
6.8	Summary	93
6.9	Reproducibility Package	95
6.10	Problems	95
7	Pooling Risks Across Locations and Quantifying Spatial Heterogeneity	99
7.1	Hierarchical Models for Multisite Time Series Studies of Air Pollution and Health	99
7.1.1	Two-stage hierarchical model	102
7.1.2	Three-stage hierarchical model	104
7.1.3	Spatial correlation model	107
7.1.4	Sensitivity analyses to the adjustment for confounders ..	110
7.2	Example: Examining Sensitivity to Prior Distributions	112
7.3	Reproducibility Package	114
7.4	Problems	114
8	A Reproducible Seasonal Analysis of Particulate Matter and Mortality in the United States	117
8.1	Introduction	117
8.2	Methods	121
8.2.1	Combining information across cities	123
8.3	Results	123
8.3.1	Sensitivity analyses	127
8.4	Comments	130
8.5	Reproducibility Package	131
	References	133
	Index	143