이학석사 학위논문

Estimating mortality in municipalities of Korea with Bayesian hierarchical models

계층적 베이즈 모형을 이용한 한국 기초자치단체의 사망자수 추정

2019년 2월

서울대학교 대학원

통계학과

김 효 신

Abstract

Hyoshin Kim
The Department of Statistics
The Graduate School
Seoul National University

With increasing concerns about the adverse impacts of air pollution on health, estimating the health effects of poor air quality has become a subject of great interest. We estimated mortality risk from air pollution in municipalities of Korea from 2010 to 2017 and examined how effects differ by aerosol levels. To obtain meteorological measurements of all municipalities in Korea, satellite based aerosol and surface temperature estimates were used to represent daily air pollution levels and temperature, respectively. Poisson regression models with smooth terms to control for potential confounders were fit within each municipality. With the fitted models we compared mortality during days with low and high aerosol levels within each municipality. Bayesian hierarchical modeling was used to combine individual air pollution effect estimates to generate regional and national level effects. Our findings show that overall, total mortality increased 1% during days with high level air pollution, with a 7% increase in Chungcheongbuk-do. Estimated mortality was higher in rural areas (2%) than in urban areas (1%). These results support evidence of increased mortality risk from air pollution.

Keyword :*Bayesian hierarchical modeling, Air pollution, Satellite based estimates, MODIS*

Student Number : 2017-23708

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Air pollution poses a serious threat to human health. According to WHO, approximately 7 million people die from air pollution over the world every year [1]. Exposure to pollutants is a leading cause of cardiovascular and respiratory diseases, heart attack, strokes, lung disease, and cancer [1, 13]. Research shows that even low levels of air pollution can lead to negative health outcomes [4]. In Korea, air pollution has become a matter of great concern during the last few years. In this thesis we investigate the daily mortality effect of poor air quality on the residents in Korea. Distinctive areas of the world have different mixtures and levels of pollutants that vary substantially by region. Studies have shown that pollutant-mortality associations differ worldwide, and thus regional based research is needed [10, 9]. To fully understand the impact of pollutants on health in Korea, we focus on estimating mortality in municipalities of Korea.

For the case of Korea, numerous studies have been conducted to investigate the relationships between air pollution and health, but the scope of most studies has been limited to a single city or at most several major cities [10, 9]; research on rural communities is lacking. In many counties (Gun) and districts (Gu) of Korea, ground recording stations are sparse with one or more stations per county (Gun), and not available for some districts (Gu). Monitoring stations are only densely distributed in populated cities (Si). Thus in most counties (Gun) and districts (Gu) weather data collected at recording stations do not necessarily represent regional variability, and regional characteristics are difficult to validate from ground measurements alone. This has served as a major obstacle for conducting research on rural municipalities in previous studies. Satellite observations provide a means of collecting meteorological information in areas where weather stations are too thinly dispersed. In this paper we focus on the health effects of poor air quality in Korea for all its cities (Si), counties (Gun), and districts (Gu) by utilizing the meteorological data retrieved from satellite.

Since the launch of the National Aeronautics and Space Administration's (NASA) Terra satellite in 1999 and Aqua satellite in 2002, the Moderate-resolution Imaging Spectroradiometer (MODIS) instrument on board Terra and Aqua has produced data of the Earth's atmosphere, land, and ocean for nearly 20 years. The two satellites view the complete surface of the Earth every one to two days. The acquired MODIS products are transferred to ground stations, which are then distributed to the science and applications community. They have been widely applied to various fields of study including climatology, meteorology, ecology, agriculture, and public health [12, 11, 5]. Among these datasets, we used the MODIS Aerosol Optical Depth (AOD) data and the MODIS Land Surface Temperature (LST) data for our analysis. Aerosols are fine solid particles and liquid droplets in the atmosphere, and AOD is the satellite measurement of aerosols. Observations of AOD are used to monitor high aerosol amounts related to fires, dust storms, and most importantly human produced air pollution. Many studies have used AOD to predict atmospheric particulate matter (PM) with statistical regression models, in particular $PM_{2.5}$ concentrations [8, 2]. For the present study we adopted the satellite retrieved AOD as a measure of air pollution. Temperature has a well known relationship to air pollution and health [7, 6, 3]. We used LST as a representation of daily temperature, and controlled for the possible confounding effect of temperature on the air pollution and health relationship.

In this study, we examined the effects of air pollution on mortality in municipalities of Korea from 2010 to 2017. Poisson regression models with splines were fit for each municipality to compare mortality for days with moderate and severe aerosol levels. We also applied a Bayesian hierarchical model to combine individual effects from municipalities, and evaluated the overall effects at the regional and national levels.

Chapter 2

Data

To investigate the association between air pollution and daily mortality, we first obtained daily mortality data for the municipalities in Korea between 2010 to 2017. The mortality data was provided by the Korean National Statistical Office. Previous studies on metropolitan areas stratified deaths by cause (e.g., cardiovascular and respiratory) and individual characteristics (e.g., gender, age, years of education). In contrast, many counties (Gun) and districts (Gu) of Korea have a small population size with low death counts per day, making mortality stratification difficult. Thus we consider only the total mortality count for the present study.

Meteorological data for the study period was collected from the MODIS instrument on board NASA's Aqua satellite. Aerosol values are obtained from the MODIS AOD data (MYD04L2, Collection 6) which is produced at the spatial resolution of 10×10 km. The MODIS LST data (MYD11A1, Collection 6), with a resolution of 1×1 km, provided temperature measurements. Both data sets are available for download in NASA's Distributed Active Archive Centers (DAACs). Daily measurements were extracted from the downloaded data, and imputation was performed for cloud-contaminated samples. Finally, mean values were derived for each municipality of Korea.

Figures 2.1 and 2.2 show the map of Korea with LST and AOD values displayed for a random day on October, 2017, respectively. LST values indicate daily temperature ($°C$) measurements for the municipalities in Korea. In Figure 2.2 the range of AOD has been scaled to 0 and 1 range, where large AOD values indicate high aerosol amounts. AOD values less than 0.1 (dark blue) suggest a clean sky view with high visibility. Values close to 1 (reddish brown), on the contrary, suggest a very hazy sky view with low visibility.

In Figure 2.3, the plots of daily mortality, AOD and LST values in the Seoul capital area (Seoul, Gyeonggi-do, Incheon) are given for the study period 2010 to 2017. We have added a natural cubic spline, with three degrees of freedom per year, in each plot to highlight the over pattern. We see there is an apparent seasonal trend. The mortality data exhibit higher levels during winter than in summer, and vice versa for AOD and LST.

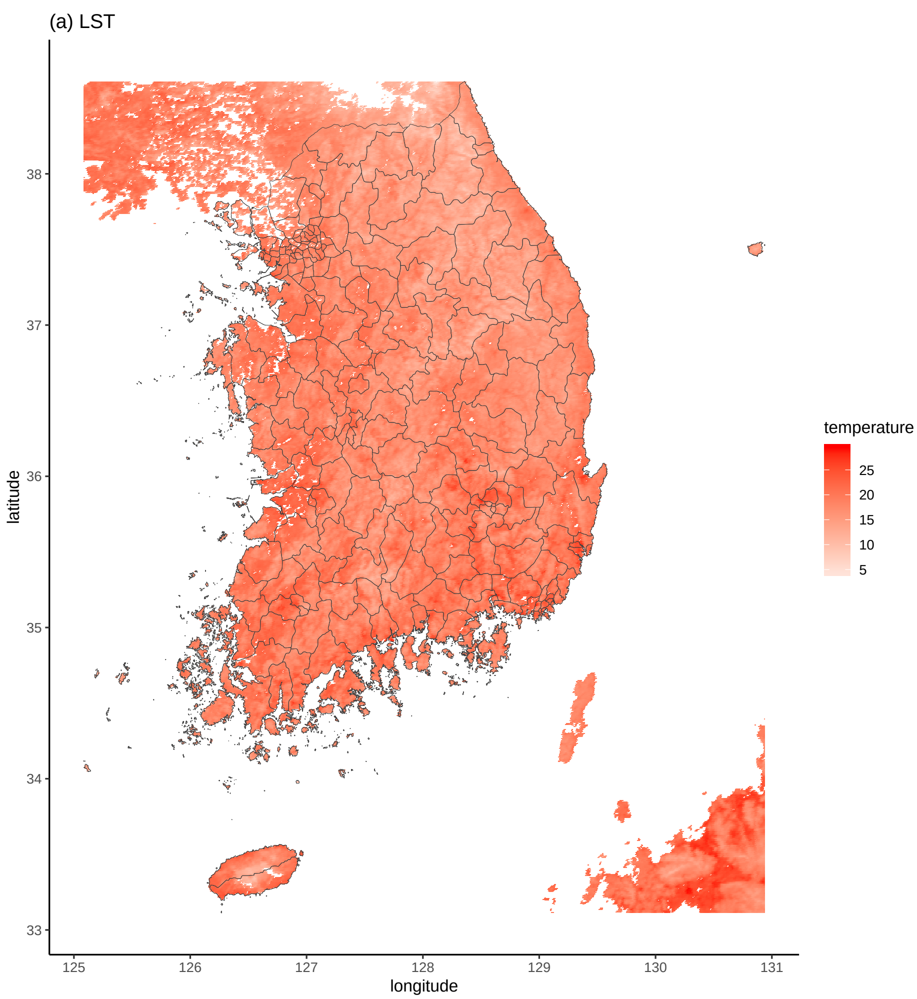

Figure 2.1: LST values of Korea on October 31, 2017

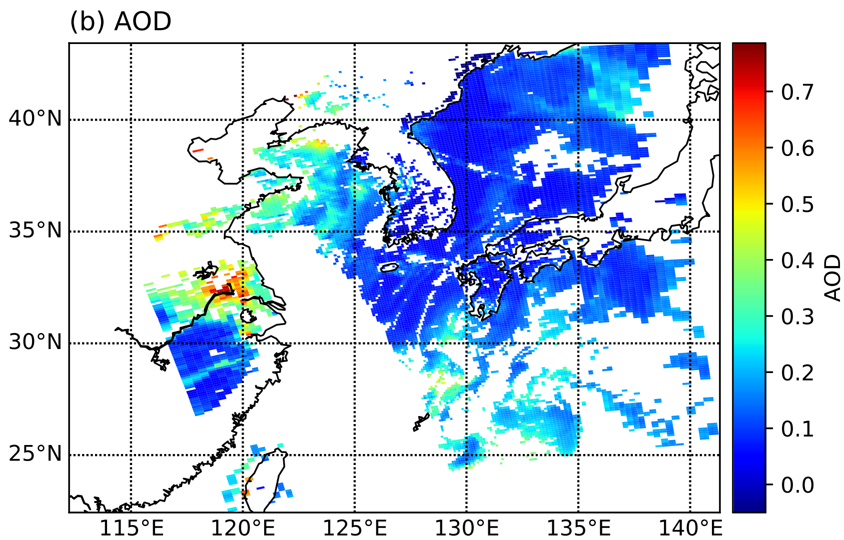

Figure 2.2: AOD values of Korea on October 31, 2017

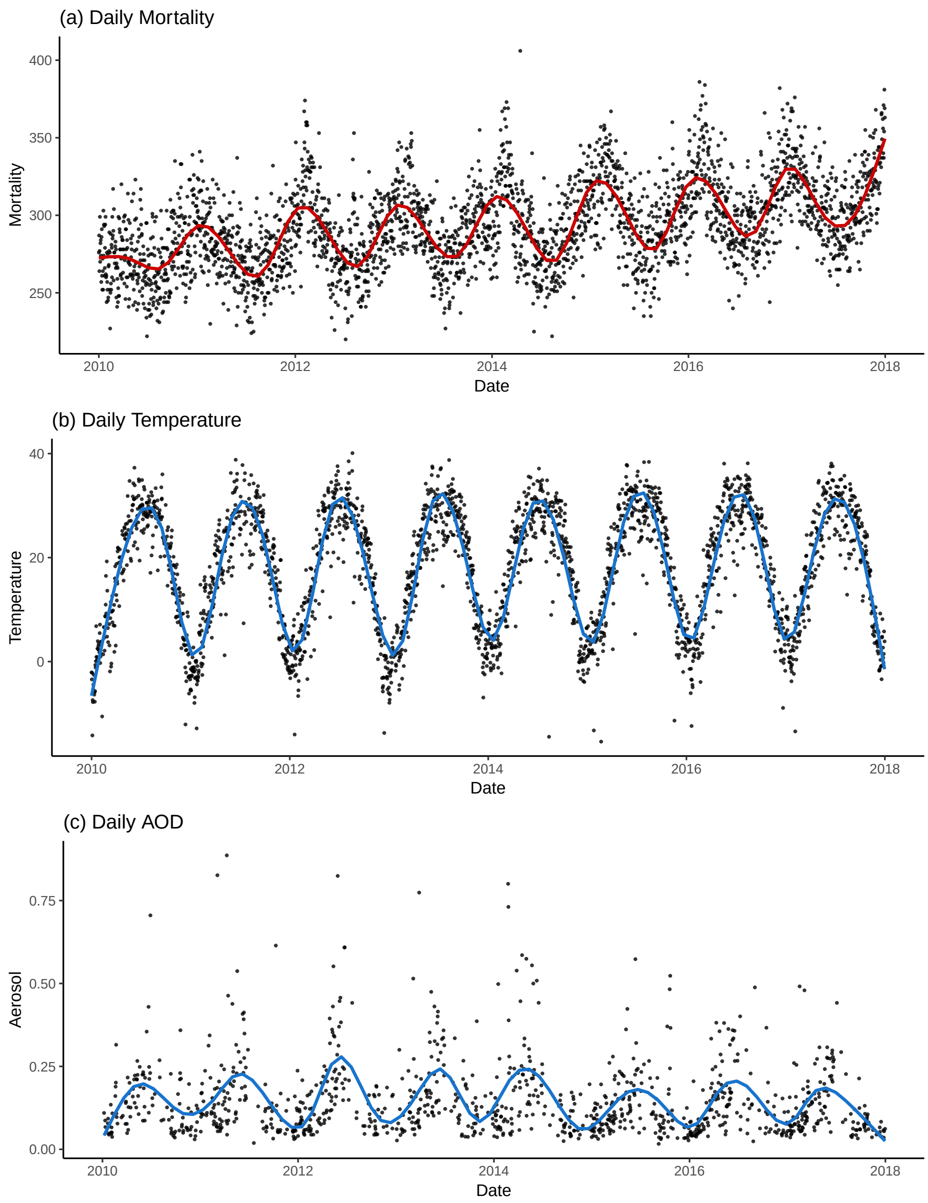

Figure 2.3: Daily mortality, AOD, and LST values of the Seoul capital area

Chapter 3

Methodology

We began with fitting a Poisson regression model for each municipality to estimate the difference in mortality during days with low and high level air pollution. We used an indicator variable of air pollution as the predictor variable, and daily mortality as the response variable. Poisson regression modeling was used as in previous studies [3, 10, 9]. The model is commonly used in cases when the response is a rate or count, as is our setting. Studies have shown that factors such as temperature and season generally have a strong nonlinear relationship with mortality in Korea. To remove the confounding effect of these variables we controlled for daily temperature, along with day of the week and time trends to explain seasonal and yearly time trends. In addition, smooth terms were introduced to account for possible overdispersion as well. Overdispersion is often encountered in Poisson models when the variability of the data set is greater than the mean. The model structure is given as follows.

$$Y_{it} \sim Poisson(\mu_{it})$$
$$log(\mu_{it}) = \beta_{0i} + \beta_{1i}AP_{it} + ns(Temp_{it}, 8) + ns(DOW_t, 3) + ns(Time_t, 8),$$
$$i = 1, \cdots I \quad t = 1, \cdots, T \tag{3.1}$$

Y_{it} denotes the total number deaths in municipality i on on day t, and we assume that Y_{it} follows a Poisson distribution with mean μ_{it}. I is the total number of municipalities in the study, and T is the length of the study period. β_{0i} is an intercept for municipality i, and β_{1i} the coefficient for air pollution in municipality i, describing the log relative risk of mortality associated with pollution. The relative risk is the ratio of the probability of the outcome (death), occuring in the exposed group (population on high level air pollution days) to the probability of the outcome in the unexposed group (population on low level air pollution days). We characterized severity of air pollution based on a municipality specific definition of AOD levels. AP_{it} is the indicator of low and high level air pollution in municipality i on day t. AP_{it} is 0 if day t has low level air pollution, that is, AOD is less than or equal to the 50^{th} percentile of AOD levels within municipality i during the study period. AP_{it} is 1 if day t has high level air pollution with AOD value exceeding the 50^{th} percentile of AOD levels for municipality i during the study period. $ns(Temp_{it}, 8)$ is the natural cubic spline of daily temperature, represented by LST, for municipality i on day t, with eight degrees of freedom per year. $ns(DOW_t, 3)$ is the natural cubic spline, with three degrees of freedom per year, of the categorical variable for day of the week. Finally, $ns(Time_t, 8)$ is the natural cubic spline of time with eight degrees of freedom. A total of

220 municipalities were analyzed with the model, with municipalities with excessive cloud contaminated samples excluded.

In Equation 3.1. we estimated the effect of air pollution on mortality separately for each municipality. Next we estimate the regional and national pollutant - mortality relationship using a two stage Bayesian hierarchical model. Bayesian hierarchical modeling extends the model for analyzing time series data in a single location to analyzing multiple time series data across areas. It can pool information of mortality risk over municipalities, and create overall estimates. This modeling approach is frequently used to combine estimates recorded at different spatial location in public health studies. Our two stage Bayesian hierarchical model assumes the following structure for the estimate of β_{1i}, $i = 1, \cdots, I$ in Equation 3.1.

$$\begin{aligned} \widehat{\beta_{1i}} | \beta_{1i} &\sim N(\beta_{1i}, \widehat{\nu_i}) \\ \beta_{1i} | \mu, \tau^2 &\sim N(\mu, \tau^2) \end{aligned} \tag{3.2}$$

In the first stage of the model, $\widehat{\beta_{1i}}$ is the estimate of air pollution effect in municipality i, β_{1i} is the true air pollution effect, assumed to be independent across municipalities for all $i = 1, \cdots, I$. $\widehat{\nu_i}$ is the within location variance, an estimate of the statistical variance of β_{1i} obtained by fitting Equation 3.1. At the second stage, information from multiple municipalities is combined. μ is the true average air pollution effect over all municipalities, and τ^2 is the between location variance of β_{1i}. The national estimate was generated by considering all municipalities $i = 1, \cdots, I$. Regional estimates were calculated for urban and rural areas separately. We generated estimates for the seven major cities of Korea: Seoul, Busan, Incheon, Daegu, Daejeon, Gwangju, and Ulsan. The urban area estimate was generated by pooling estimates from these cities. Results were obtained for the rural areas by generating estimates from the nine provinces of Korea: Gyeonggi-do, Gangwon-do, Chungcheongbuk-do, Chungcheongnam-do, Jeollabuk-do, Jeollanam-do, Gyeongsangbuk-do, Gyeongsangnam-do, and Jeju-do.

Chapter 4

Results

Table 4.2 shows summary statistics of municipalities within the seven cities and nine provinces in Korea. The mean daily mortality ranged from 2.85 in Ulsan to 4.70 in Seoul. In the case of provinces, mortality was lowest in Jeollanam-do with 2.51, and highest in Gyeonggi-do with 4.97. Average daily AOD values were similar across all regions. AOD ranged from 0.116 in Daegu to 0.214 in Incheon for the seven cities. Gangwon-do showed the lowest AOD of 0.098, and Chungcheongnam-do the highest AOD of 0.152 for the nine provinces. Daily mean LST values ranged from $17.5°C$ in Daegu to $19.0°C$ in Ulsan for the cities, and from $13.4°C$ in Gangwon-do to $19.3°C$ in Jeju-do for the case of provinces.

We estimated a separate effect for each municipality using Equation 3.1, then estimated the regional estimates for the seven cities and nine provinces using regional specific Bayesian hierarchical models, as suggested in Equation 3.2. The estimated mortality effect was similar across all cities (Table 4.3).

Areas	No. of municipalities	Effect [mean ± sd (range)]
Urban	56	0.01 ± 0.15 (-0.14 ∼ 0.16)
Rural	144	0.02 ± 0.09 (-0.07 ∼ 0.11)
Total	200	0.01 ± 0.07 (-0.06 ∼ 0.08)

Table 4.1: Comparison of regional (urban, rural) and national areas regarding estimated increase in mortality during high level air pollution days and low level air pollution days.

Six out of seven cities (Busan, Daegu, Incheon, Gwangju, Daejeon, Ulsan) showed a positive effect ($1\% - 3\%$) of high air pollution days compared with low air pollution days on mortality. We estimated a significant increase in mortality in Chungcheongbuk-do on high air pollution days in comparison with low air pollution days (7%). Jeollanam-do was the province with the lowest mortality risk estimate (-1%).

Bayesian hierarchical models were again applied to calculate overall regional and national pollutant - mortality relationships. The results are given in Table 4.1. Urban area estimates, generated by pooling estimates across the seven cities (Si), showed a positive effect of air pollution on mortality (1%). Estimates from the nine provinces (Gun, Gu) were pooled together to calculate rural area estimates of mortality risk. We observed a stronger effect in rural areas, with a 2% estimated effect of high air pollution days compared with low air pollution days. Overall, for the entire municipalities in Korea, the mortality effect was estimated to be 1% in total.

Cities	2017 population	Mortality (mean ± sd)	Aerosol (mean ± sd)	Temperature (mean ± sd)
Seoul	9,857,426	4.70 ± 2.44	0.130 ± 0.106	18.6 ± 12.1
Busan	3,470,653	3.89 ± 2.26	0.137 ± 0.089	18.9 ± 9.39
Daegu	2,475,231	4.59 ± 2.62	0.116 ± 0.091	17.5 ± 10.7
Incheon	2,948,542	4.16 ± 2.72	0.214 ± 0.127	18.3 ± 11.8
Gwangju	1,463,770	4.21 ± 2.46	0.132 ± 0.089	18.5 ± 10.1
Daejeon	1,502,227	3.81 ± 2.06	0.129 ± 0.095	17.8 ± 10.7
Ulsan	1,165,132	2.85 ± 1.62	0.125 ± 0.086	19.0 ± 9.32

Provinces	2017 population	Mortality (mean ± sd)	Aerosol (mean ± sd)	Temperature (mean ± sd)
Gyeonggi-do	12,873,895	4.97 ± 3.73	0.142 ± 0.110	17.0 ± 11.4
Gangwon-do	1,550,142	2.42 ± 1.83	0.098 ± 0.075	13.4 ± 10.6
Chungcheongbuk-do	1,594,432	3.15 ± 2.88	0.121 ± 0.088	16.2 ± 10.9
Chungcheongnam-do	2,116,770	3.00 ± 2.02	0.152 ± 0.109	17.2 ± 10.4
Jeollabuk-do	1,854,607	3.30 ± 2.62	0.145 ± 0.108	16.9 ± 10.4
Jeollanam-do	1,896,424	2.51 ± 1.69	0.132 ± 0.091	17.3 ± 9.43
Gyeongsangbuk-do	2,691,706	3.05 ± 2.24	0.116 ± 0.082	16.2 ± 9.89
Gyeongsangnam-do	3,380,404	3.57 ± 3.40	0.122 ± 0.083	17.2 ± 9.12
Jeju-do	657,083	4.68 ± 2.63	0.108 ± 0.072	19.3 ± 9.10

Table 4.2: Summaries of municipalities within seven cities and nine provinces in Korea, 2010 - 2017.

Cities	No. of municipalities	Effect [mean ± sd (range)]
Seoul	25	-0.02 ± 0.25 (-0.27 ∼ 0.23)
Busan	4	0.03 ± 0.76 (-0.73 ∼ 0.79)
Daegu	8	0.01 ± 0.50 (-0.49 ∼ 0.51)
Incheon	7	0.02 ± 0.55 (-0.53 ∼ 0.57)
Gwangju	3	0.02 ± 0.91 (-0.89 ∼ 0.93)
Daejeon	5	0.03 ± 0.66 (-0.63 ∼ 0.69)
Ulsan	4	0.03 ± 0.76 (-0.73 ∼ 0.79)

Provinces	No. of municipalities	Effect [mean ± sd (range)]
Gyeonggi-do	31	0.02 ± 0.22 (-0.20 ∼ 0.24)
Gangwon-do	18	0.01 ± 0.31 (-0.30 ∼ 0.32)
Chungcheongbuk-do	11	0.07 ± 0.41 (-0.34 ∼ 0.48)
Chungcheongnam-do	15	0.03 ± 0.34 (-0.31 ∼ 0.37)
Jeollabuk-do	14	-0.002 ± 0.360 (-0.362 ∼ 0.358)
Jeollanam-do	17	-0.01 ± 0.32 (-0.33 ∼ 0.31)
Gyeongsangbuk-do	22	0.02 ± 0.27 (-0.25 ∼ 0.29)
Gyeongsangnam-do	14	0.002 ± 0.359 (-0.357 ∼ 0.361
Jeju-do	2	0.01 ± 1.19 (-1.18 ∼ 1.20)

Table 4.3: Estimated mortality risk on high level air pollution days compared with low level air pollution days for seven cities and nine provinces in Korea, 2010 - 2017.

Chapter 5

Conclusion

Study of health impacts of air pollution is important for the present day as severity of air pollution continues to increase. This study is aims to estimate the effects of air pollution for the municipalities in Korea over a long time frame, by adopting MODIS satellite measurements of daily air pollution and temperature, via AOD and LST values, respectively. Municipality level estimates were than pooled to generate regional and national estimates. We found that overall mortality risk increased on a national level, as well as for individual cities and provinces in general. Rural areas exhibited higher increase in relative rate of mortality compared to urban regions. These findings have implications on increased mortality risk from air pollution across regions in Korea. For future study, further investigation using a mixture of both satellite measurements as well as ground measurements to track the severity of air pollution is needed. While satellite measurements can provide a means to estimate the level of air pollution in areas where ground measurements are unavailable, for areas where ground weather stations are present, a combination of the two sources can be more reliable. Further analysis using $PM_{2.5}$ concentrations, obtained from predictions on AOD values, should be considered as well. Such data could improve air pollution estimates, and thus enhance the accuracy and reliability of the present study.

Bibliography

[1] Albert Einstein. World health statistics 2018: monitoring health for the sdgs, sustainable development goals. geneva: World health organization; 2018. licence: Cc by-nc-sa 3.0 igo.

[2] Donkelaar et al. Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: Development and application. *Environ Health Perspectives*, 118(6):847–855, 2010.

[3] Donkelaar et al. High temperature, heat index, and mortality in 6 major cities in south korea. *Environ Health Perspectives*, 2012.

[4] Dr Aaron JCohenDSc et al. Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the global burden of diseases study 2015. *The Lancet*, 389(10082):1907–1918, 2017.

[5] Duan et al. A framework for the retrieval of all-weather land surface temperature at a high spatial resolution from polar-orbiting thermal infrared and passive microwave data. *Remote Sensing of Environment*, 195:107–117, 2017.

[6] G Maier et al. Assessing the performance of a vulnerability index during oppressive heat across georgia, united states. *American Meteorological Society*, 2014.

[7] GB Anderson et al. Heat waves in the united states: Mortality risk during heat waves and effect modification by heat wave characteristics in 43 u.s. communities. *Environ Health Perspectives*, 119(2), 2011.

[8] Hu et al. Estimating ground-level pm2.5 concentrations in the southeastern united states using maiac aod retrievals and a two-stage model. *Remote Sensing of Environment*, 140:220–232, 2014.

[9] Lee et al. Air pollution and daily mortality in seoul and ulsan, korea. *Environmental Health Perspective*, 107(2):149–154, 1999.

[10] Lee et al. Air pollution and daily mortality in seven major cities of korea, 1991–1997. *Environmental Research*, 84:247–254, 2000.

[11] Li et al. Evaluation of 10 year aqua/modis land surface temperature with surfrad observations. *International Journal of Remote Sensing*, 35:830–856, 2014.

[12] Michaelides et al. Monitoring and forecasting air pollution levels by exploiting satellite, ground-based, and synoptic data, elaborated with regression models. *Advances in Meteorology*, 2017(ID 2954010), 2017.

[13] Zhao D et al. Xie W, Li G. Relationship between fine particulate air pollution and ischaemic heart disease morbidity and mortality. *Heart*, 101:257–263, 2015.

초 록

대기오염이 건강에 미치는 악영향에 관한 우려가 커짐에 따라 그 영향을 정확히 측정하는 일이 크게 주목받고 있다. 본 연구는 2010년에서 2017년 간 한국 기초자치단체에서의 대기오염에 의한 사망위험이 에어로졸 수준에 따라 변화하는 정도를 측정하였다. 인공위성에서 얻은 에어로졸과 지상기온 자료를 이용하여 각 기초자치단체의 대기오염과 기온 기상관측자료를 얻었다. 각 기초자치단체에 포아송회귀모형을 적합하여 에어로졸 수준이 높은 날과 낮은 날의 사망위험을 비교하였다. 포아송회귀모형에 스플라인을 추가하여 교란변수로 인한 효과를 통제하였다. 계층적 베이즈 모형을 이용하여 기초자치단체들의 추정값을 결합하였다. 이를 이용해 지역 별, 그리고 전국적 효과를 추정하였다. 전체적으로 사망위험은 에어로졸 수준이 높은 날에 약 1% 증가하였다. 특히 충청북도는 약 7% 증가세를 보였다. 비도시 지역의 사망위험은 약 2%로, 도시 지역의 사망위험 1%에 비해 높은 경향을 보였다. 이러한 결과는 대기오염으로 인한 사망위험의 증가를 뒷받침해주고 있다.

주요어 :*계층적 베이즈 모형, 대기오염, 인공위성, MODIS*

학번 : 2017-23708