

Spatial Analysis of Air Pollution Data Based on Linked Micromap Plots in Korea

Jeong Yong Ahn

Department of Statistics (Institute of Applied Statistics), Chonbuk National University, Jeonbuk, 561-756, Korea

Abstract—Air pollution is becoming an increasingly important problem due to the high levels of air pollutants emitted as a result of industrial and other anthropogenic activities. Exposure to air pollution has been linked to various health outcomes, from modest transient changes in the respiratory tract and impaired pulmonary function to increased mortality. Consequently, analyzing air pollution data and reporting associated statistics have become important tasks for governments and non-governmental organizations. In this study, we analyzed air pollution data and demonstrated how spatial patterns of air pollutants and other variables can be examined visually and simultaneously using linked micromap plots. To explore the potential utility of linked micromap plots, we analyzed and visualized Korean air pollution data for the period 2010-2012. Our analyses revealed more than twice the amount of fine particles in March compared with August, and that ozone levels were highest in May. Regions with high levels of fine particles were located mainly in the northwest, likely influenced by Asian dust, and in the southeast industrialized areas.

Keywords—air pollution; data visualization; linked micromap plots; pollutants; spatial patterns.

I. INTRODUCTION

During the last two centuries, air quality and air pollution have been recognized globally as pressing environmental issues. Since the industrial revolution in the late 18th century in Europe, and through the 19th and 20th centuries, the use of coal in homes and by industry has polluted the air [1]. Smog and ‘pea-soupers’ have become common in cities worldwide, and other kinds of atmospheric pollution have increased greatly. In many places, pollution seriously threatens the lives and activities of people every day.

Air pollution is generated from many sources, including industries, motor vehicles, and heating, and a variety of pollutants are released from each source. Consequently, people are generally exposed to multiple air pollutants simultaneously [2]. Exposure to air pollution has been linked to a number of health outcomes, from modest transient changes in the respiratory tract and impaired pulmonary function, to restricted activity/reduced performance, emergency room visits and hospital admissions, to increased mortality [3]. Therefore, analyzing air pollution and reporting statistics have become important tasks.

Interpreting and understanding air pollution data usually requires considering an appropriate spatial context: most

data come with geographic information, and many studies apply a geographical framework to analyze these data [4-7]. Most of these studies use simple images and maps, such as aerial images and choropleth and isopleth maps to visualize the data. Although these images and maps are commonly used, they also have problems and limitations [8, 9]. One problem is that it is difficult to show more than one variable in an image or a map. Another problem is that it may be difficult and frustrating for a reader to observe trends, relationships, and anomalies that may be present in the data [10].

In this study, we introduce a new approach for visualizing data based on linked micromap plots. To explore the potential for the use of linked micromap plots, we analyzed and visualized South Korean air pollution data for the period 2010-2012. With the plots, we can explore relationships between the variables and geographic patterns that are inherent in the data. First, we investigated temporal variations in pollutants through a statistical analysis. Next, we divided the 16 districts of Korea into two groups and compared the magnitudes of pollutants by region. Finally, we explored geographic patterns and relationships between the variables and pollutants using linked micromap plots.

II. MATERIALS AND METHODS

A. Data sources

To analyze and visualize air pollution data, we collected data for the period 2010-2012 from the Korean Statistical Information Service (<http://kosis.kr>), and aggregated the data by the 16 districts of Korea. The dataset consists of 13 pollutants/variables and region names. Table 1 lists the variables used.

TABLE I. VARIABLES OF DATA SET

Categories	Indicators
Air pollutants	fine particle (PM_{10}), sulphur dioxide (SO_2), ozone (O_3), nitrogen dioxide (NO_2), carbon monoxide (CO)
Heavy metal pollutants	lead (Pb), cadmium (Cd), chrome (Cr), copper (Cu), manganese (Mn), iron (Fe), nickel (Ni), arsenic (As)
Geographical locations	region name

B. Visualization technique

The graphical visualization technique presented in this study is referred to as a linked micromap plot: a statistical graph that combines multiple small maps and statistical data panels. Linked micromap (LM) plots are graphics that link statistical information to an organized set of small maps to explore and communicate patterns in the outcome variable, geographic locations, and the associations among them [11, 12].

Fig. 1 presents a typical template for a LM plot. To visualize the geographically referenced data, LM plots use three or more sequence panels in parallel, linked by location. The types of panels are micromaps, labels, and statistical

summary panels. Fig. 1 shows five parallel sequences of panels: one micromap, one label, and three statistical summary panels. LM plots can sort geographic subregions based on the statistical variables of interest. Sorting improves perception between consecutive panels, from the top to the bottom of the display. To allow the viewer to focus attention on explicit areas, one at a time, LM plots divide the regions into perceptual groups of size five or fewer, and link the locations using colors. These features of LM plots make it possible to visualize specific geographic patterns in the data that are often lost with other types of graphs and maps.

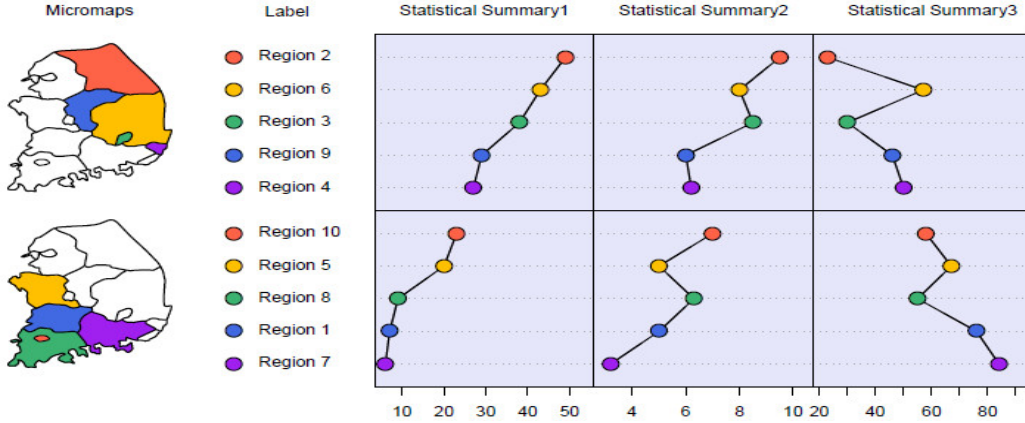


Figure 1. A template of LM plot

III. DATA ANALYSIS AND VISUALIZING

A. Data analysis

To compare pollutants by regional groups, we divided the regions into two groups, metropolitan and rural, according to the extent of urbanization. The degrees of pollution in the groups were compared using the Mann-Whitney U-test, a non-parametric statistical technique.

Table 2 lists the results of the comparison: ozone (O_3), nitrogen dioxide (NO_2), and nickel (Ni) differed significantly by group; other forms of pollution did not. Levels of ozone and nickel were higher in metropolitan areas, while levels of nitrogen dioxide were higher in rural areas. Fig. 2 presents the temporal variations in air pollutants from 2010-2012 in South Korea: carbon monoxide (CO), sulphur dioxide (SO_2), and nitrogen dioxide exhibited little seasonal variation in the study period. In contrast, fine particle (PM_{10}), increased markedly from March to May and then slightly from August to September. Ozone also exhibited some seasonal changes: levels were higher from February to April and then lower from November to January.

TABLE II. RANK COMPARISON BY GROUPS

Categories	Indicators	Group		p-value
		Group A (metropolitan)	Group B (rural areas)	
Air pollutants	PM_{10}	8.88	8.13	0.399
	SO_2	9.81	7.19	0.145
	O_3	6.44	10.56	0.044*
	NO_2	12.13	4.88	0.001**
	CO	8.63	8.38	0.467
Heavy metal pollutants	Pb	7.94	8.07	0.491
	Cd	7.31	7.75	0.438
	Cr	8.06	7.93	0.490
	Cu	7.13	8.00	0.364
	Mn	6.75	9.43	0.140
	Fe	7.25	8.86	0.268
	Ni	6.00	10.29	0.036*
	As	6.56	7.70	0.325

* $p < 0.05$, ** $p < 0.01$

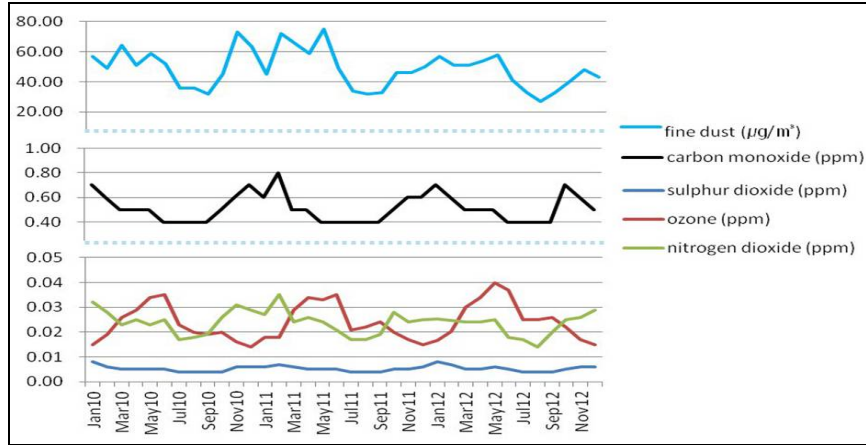


Figure 2. Temporal variation of air pollutants

Table 3 lists the average levels of fine particle for March and August, and ozone in May and December by year. Fine particle levels in March were almost double those in August. Ozone measurements were highest in May, when levels were more than three times those in December.

TABLE III. AVERAGE OF PM_{10} AND O_3 BY YEARS

		2010	2011	2012
PM_{10} ($\mu\text{g}/\text{m}^3$)	March	57.69	72.75	56.38
	August	34.06	30.06	25.75
O_3 (ppm)	May	0.036	0.035	0.041
	December	0.016	0.016	0.016

B. Visualizing data using linked micromap plots

LM plots illustrate relationships between variables and spatial patterns of data while linking regional names to the locations on a map and to the estimates indicated in statistical panels. To represent statistical information in LM plots, we developed an extended version of the “LinkedMicroMap()” function developed in our previous study [13]. In the extended version, we have reformed the function and added various options to establish statistical plots.

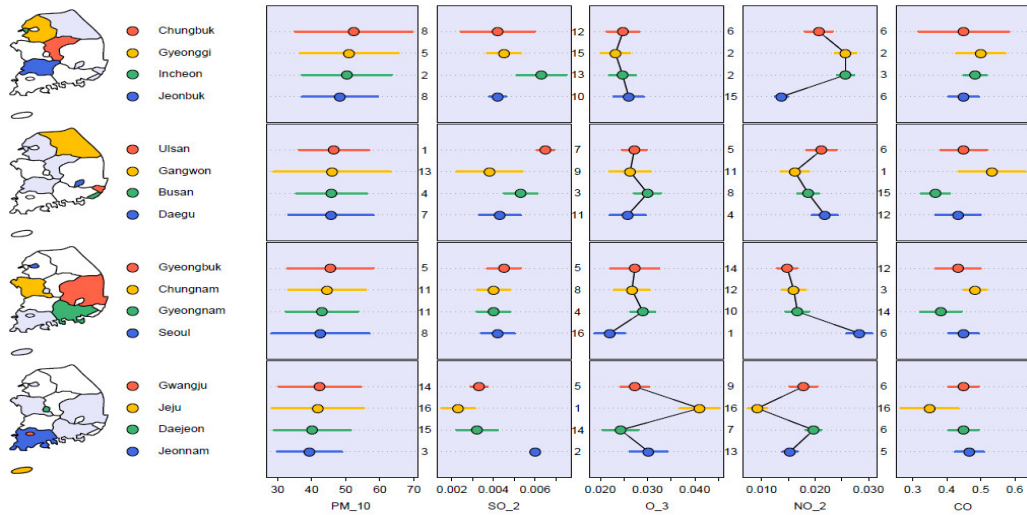


Figure 3. LM plot showing the relationships between air pollutants

Fig. 3 shows how an LM plot can illustrate relationships between air pollutants. The data shown are averages for the period 2010-2012, and regions in the figure are sorted by fine particle levels, from highest to lowest. No spatial patterns were observed in fine particle levels, and no correlations appeared between fine particle

levels and other pollutants. In contrast, substantial inverse correlations appeared between ozone and nitrogen dioxide levels, and between ozone and carbon monoxide levels.

Fine particle values in Korea worsened from January to May, and then slightly in August. Fig. 4 presents trends in fine particle levels by month in 2012. Regions with

high fine particle levels were located mainly in the northwest, likely as the result of a great influx of Asian dust from China and Mongolia. Interestingly, fine particle levels in Ulsan and Busan were quite high despite their

geographic location. Because Ulsan and Busan are quite industrialized, air pollution in these regions was likely caused by industrial processes.

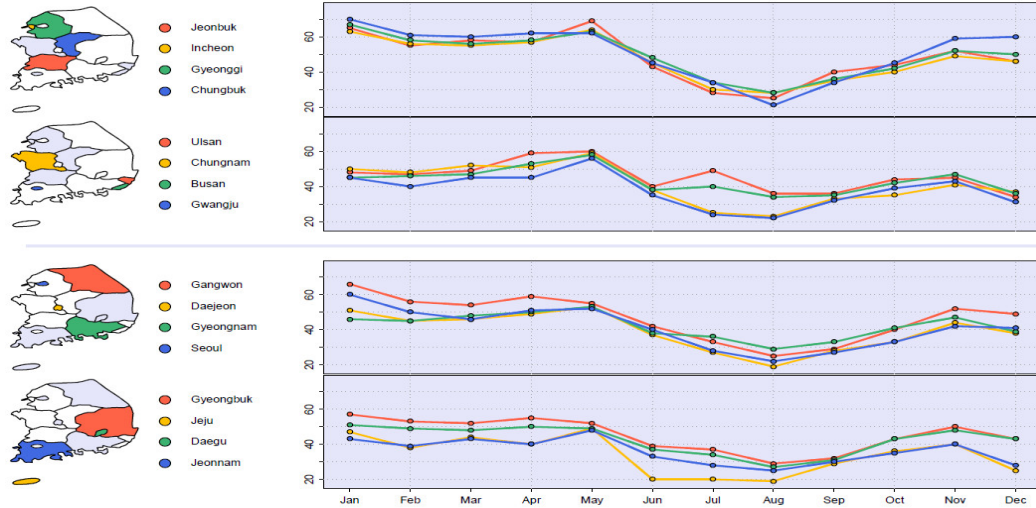


Figure 4. Trends of the fine particles (PM_{10}) by months in 2012

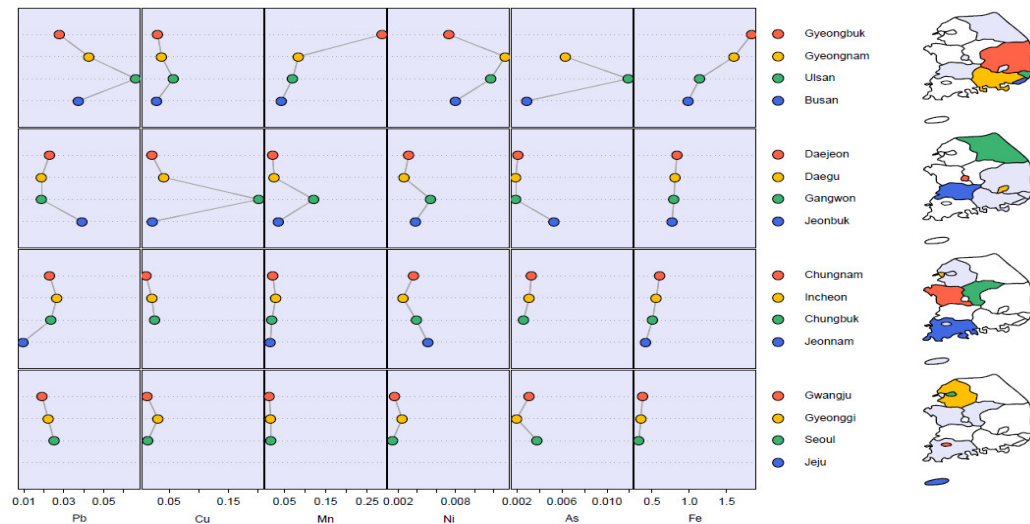


Figure 5. LM plot of heavy metal pollutants

Fig. 5 shows how an LM plot can illustrate relationships between heavy metal pollutants. Regions in the figure are sorted by iron (Fe) level, from highest to lowest. Iron values were highly correlated with manganese (Mn) and nickel (Ni) levels, whereas iron and copper (Cu) levels were not correlated. Lead (Pb) levels were highly correlated with nickel and arsenic (As) levels. The data for copper, manganese, and arsenic levels revealed unusual outliers: specifically, the values for Gangwon, Gyeongbuk, and Ulsan, respectively. These outliers reduced the association between pollutants considerably.

IV. CONCLUSIONS

Air pollution measurements display patterns over space and time, allowing for spatio-temporal modeling, through which pollution concentrations and trends can be analyzed. In this study, we investigated Korean air pollution data for the period 2010-2012 and demonstrated how spatial patterns of the variables can be examined visually and simultaneously using LM plots.

In summary, (1) ozone and nickel levels were higher in metropolitan areas, while nitrogen dioxide levels were higher in rural areas. (2) From 2010-2012, the amount of fine particles measured in March was more than double that

measured in August, and ozone levels were highest in May. (3) The amount of fine particles was not correlated with the level of any other pollutant, but substantial inverse correlations were observed between ozone and nitrogen dioxide and between ozone and carbon monoxide. (4) In 2012, regions with high fine particles were located mainly in the northwest, likely influenced by Asian dust, and in southeast, industrialized areas. (5) Iron levels were highly correlated with manganese and nickel levels, while no correlation was observed between iron and copper levels.

ACKNOWLEDGEMENTS

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2012R1A1A4A01002729).

REFERENCES

- [1] Krivoruchko, K., *Using geostatistical analyst for analysis of California air quality*, White papers, ESRI, New York, 2006.
- [2] Khan, S.U. & Ali, N., Ambient air quality of Aligarh city (U.P., INDIA) at Diwali, *International Journal of Current Engineering and Technology*, 3(5), pp.1730-1732, 2013.
- [3] WHO, *Health Aspects of Air Pollution*, Results from the WHO Project Systematic Review of Health Aspects of Air Pollution in Europe, World Health Organization, 2004.
- [4] Matejicek, L., Spatial modelling of air pollution in urban areas with GIS: a case study on integrated database development, *Advances in Geosciences*, 4, pp.63–68, 2005.
- [5] Jerrett, M., Burnett, R.T., Ma, R., Pope, C.A., Krewski, D., Newbold, K.B., Thurston, G., Shi, Y., Finkelstein, N., Calle, E. & Thun, M.J., Spatial analysis of air pollution and mortality in Los Angeles, *Epidemiology*, 16(6), pp.727-736, 2005.
- [6] Eitana, O., Yuval, Barchana, M., Dubnov, J., Linnd, S., Carmel, Y. & Broday, D.M., Spatial analysis of air pollution and cancer incidence rates in Haifa Bay, Israel, *Science of the Total Environment*, 408, pp.4429–4439, 2010.
- [7] Zammit, L.C., Scerri, K., Attard, M., Bajada, T. & Scerri, M., Spatio-temporal analysis of air pollution data in Malta, *Proceedings of the International Conference of GeoComputation*, pp.99-105, 2011.
- [8] Dent, B.D., *Cartography: Thematic map design*, William C. Brown: Dubuque, IA, 1993.
- [9] Harris, R.L., *Information graphics - A comprehensive illustrated reference*, Oxford University Press: New York, 1999.
- [10] Gebreab, S., Gillies, R.R., Munger, R.G. & Symanzik, J., Visualization and interpretation of birth defects data using linked micromap plots, *Birth Defects Research (Part A)*, 82, pp.110-119, 2008.
- [11] Olsen, A.R., Carr, D.B., Courbois, J.P. & Pierson, S.M., Presentation of data in linked attribute and geographic space, *Proceedings of the Joint Statistical Meetings, American Statistical Association*, 271, 1996.
- [12] Carr, D.B. & Pickle, L.W., *Visualizing data patterns with micromaps*, Chapman & Hall/CRC, FL, 2010.
- [13] Han, K.S., Park, S.J., Mun, G.S., Choi, S.H., Symanzik, J., Gebreab, S. & Ahn, J.Y., Linked micromaps for the visualization of geographically referenced data, *ICIC Express Letters*, 8, pp.443-448, 2014.