Assessment of Areal Average Air Quality Level over Irregular Areas: a Case Study of PM10 Exposure Estimation in Taipei (Taiwan)

Hwa-Lung Yu¹, Shang-Chen Ku¹, Chiang-Hsing Yang², Tsun-Jen Cheng³ and Likwang Chen⁴
¹Dept of Bioenvironmental Systems Engineering, National Taiwan University, Taipei,
²Dept of Health Care Management, National Taipei University of Nursing and Health Sciences
³Institute of Occupational Medicine and Industrial Hygiene, National Taiwan University
⁴Center for Health Policy Research and Development, National Health Research Institutes, Miaoli, Taiwan

1. Introduction

Assessment of public health risk from the exposure of ambient pollutants has been heavily based upon the analyses of the associations between the pollutant exposure and its potential consequence, e.g. disease occurrences (Abbey et al., 1995; AckermannLiebrich et al., 1997; Beeson et al., 1998; Jerrett et al., 2005; Pope et al., 2002). In many air pollution epidemiologic investigations, individual health datasets at nationwide or regional scales are available to assess the subtle risks of pollution exposure. Among them, the understanding of the spatial or spatiotemporal distribution of ambient pollutants is essential due to their prevalent heterogeneity across space and/or time. In these cases, governmental agencies has established ambient air-quality monitoring networks, such as the Air Quality System (AQS) operated by the U.S. Environmental Protection Agency (EPA), regularly recording important and useful environmental data sources concerning the acute and chronic effects of ambient pollutants (TWEPA, 2006; USEPA, 1992). Based upon these databases, an ideal exposure assessment is performed by applying techniques of spatial or/and spatiotemporal analysis for the estimation of the pollutant level at the locations of individuals in health dataset. However, due to the raising concerns of privacy and confidentiality of personal information, the sensitive personal information, such as residential addresses, of health dataset is usually not allowed to be accessed. As a result, individual information of health dataset obtained from institutes or governmental agencies is often removed or degraded. Among them, the spatial locations of individuals of health dataset are usually aggregated into a larger spatial unit, e.g. higher administrative level, in which no personal identities can
Spatial interpolation techniques have been increasingly used by environmental epidemiologists to estimate the spatiotemporal distribution of air quality levels. Among them, the deterministic approaches such as nearest-neighbor method (NN) and inverse distance weighted method (IDW) are the most widely used methods (Hendryx et al., 2010; Hoek et al., 2001; Michelozzi et al., 2002; Sohel et al., 2010), in which the estimation results are derived from certain assumed functional relationship of geographical distances between the observation and estimation locations without considering the stochastic associations among the air quality measurements. The stochastic techniques, in particular kriging method, have been applied with increasing frequency in exposure assessment studies to address the spatial heterogeneity among the air quality data as well as the estimation uncertainty of predictions (Brauer et al., 2003; Brimicombe, 2000; Buzzelli and Jerrett, 2004; Hoek et al., 2001). Most spatial interpolators are originally developed for the estimations with geographical support of estimation and observation locations at point scale. As a result, most studies of exposure assessment apply spatial techniques for point estimation at the centroid of the geographical unit of interest to characterize its average air pollution level (Chen and Schwartz, 2008; Chen and Schwartz, 2009; Lertxundi-Manterola and Saez, 2009; Maheswaran and Elliott, 2003; Miller et al., 2007). The inconsistency of geographical support between the exposure estimations and aggregated health dataset can potentially distort the results of their associated epidemiological studies (Young et al., 2009; Young et al., 2008).

This study investigates and compares the estimation results of areal-averaged air quality level by several popular spatial mapping techniques, i.e. NN, IDW, ordinary kriging (OK) and block kriging method (BK). Among them, BK is a kriging-based upscaling method which can perform spatiotemporal estimation with the consideration of the stochastic dependence among the locations with irregular sizes and shapes. This comparison is performed on the spatiotemporal PM10 estimation of the townships over Taipei area (Taiwan) during 2004-2006 on the basis of data during 1997-2007.

2. Materials and methods

2.1 Study area

Taipei is the largest metropolitan area in Taiwan including Taipei city and Taipei county with the vehicle density as high as over 6000 vehicles per km$^2$. Except for traffic emissions, the three incineration plants are the major stationary emission sources in the area (Chang and Lee, 2007). The Taipei area is bounded by mountains, i.e. Yangming mountains to the north, Linkou mesa to the west, and ridge of Snow mountains to the southeast which forms the second largest basin of the island (see Figure 1). Because of the significant variation of the basin topography, the air convection and circulation are generally degraded in this area. In addition, the alluvial plains at basin floor materialized the highly urbanization area of Taipei. As a result, the ambient pollutant concentration across the basin floor is generally higher than that over its surrounding mountain areas. Taiwan Environmental Protection Agency (TWEPA) has established the air quality monitoring network which regularly records the ambient pollutants, including PM10 and other criteria pollutants (TWEPA, 2006), and meteorological covariates throughout the island since September, 1993. Figure 2
shows the eighteen TWEPA monitoring stations in Taipei area. As it can be noticed, the
distribution of monitoring locations is spatially imbalanced that most of the stations in
Taipei are located at the high urbanized area. In this study, all the PM10 observations from
these stations are aggregated into monthly data following the procedure suggested by
USEPA (USEPA, 2004). The areal concentration estimations are performed at each of the
townships shown in Figure 2 by various spatial interpolation techniques.

Fig. 1. The highways, rivers and topography in Taipei metropolitan area
2.2 Methods
This study investigates several common used spatial interpolation methods for ambient pollutant exposure estimation. Among them, the nearest neighbor method (NN), also called polygon method and Theissen method, is the simplest method for spatial estimation. It assumes the air quality level at an ungauged location is completely determined by its closest observations so that spatial distribution of pollutants is composed of a set of polygon across space. The inverse distance weighted method (IDW) is a weighted average of neighboring observation values. The weights given to each observation is a function of distances between observation and estimation locations (Waller and Gotway, 2004) as below

$$\hat{z}(s_k, t_k) = \frac{\sum_{i=1}^{N} z(s_i, t_i) \frac{1}{d_{ki}^p}}{\sum_{i=1}^{N} d_{ki}^p}$$  \hspace{1cm} (1)
where $d_{kl}$ is the space-time distance between the estimation location $(s_k,t_k)$ to the observation locations $z(s_i,t_i)$. $p$ is the modeler-specified degree. The increase of $p$ can decreases the weights of distant observations. The space-time distance is assumed to be the function of the geographical distance and time interval (Christakos, 2000; Christakos et al., 2000). It can be expressed into the form of $d_{kl} = |s_k - s_i| + m_{st}|t_k - t_i|$, where $m_{st}$ is the space-time metric to account for the relationship between spatial and temporal metric (Christakos et al., 2002). Both NN and IDW methods are deterministic methods which provide the estimations without the information of estimation uncertainty.

Stochastic techniques, on the other hand, can account for the uncertainty of the space-time data under the framework of random field theory which forms a multivariate joint distribution among the space-time attribute of interest. Among them, kriging method is the most popular technique in space-time mapping which is so-called best linear unbiased estimator (BLUE) in the sense of providing the minimum estimation uncertainty (Olea, 1999). Various types of kriging method have been developed in terms of different assumptions and analytical goals. Among them, ordinary kriging (OK) is the most widely used one which assumes the unknown mean among the space-time dependent data. The basic equations of ordinary kriging are shown below (Goovaerts, 1997)

$$
\sum_{i=1}^{n(p_k)} \lambda_{i,p_k} C(p_i,p_j) + \mu_{OK,p_k} = C(p_k,p_j)
$$

(2)

where $\lambda_{i,p_k}$ are the OK weights for the observation at $p_i = (s_i,t_i)$; $C(p_i,p_j)$ denotes the covariance of the attribute between $p_i$ and $p_j$, $n(p_k)$ is the number of observations used for the kriging estimation at $p_k$, and $\mu_{OK,p_k}$ is the Lagrange multiplier. The OK estimation at $p_k$ can be the linear combination of observations expressed as form of $\hat{z}(p_k) = \sum_{i=1}^{n(p_k)} \lambda_{i,p_k} z(p_i)$ . Block kriging (BK) inherits the framework of kriging method and considers the geographical support of observation and estimation locations, i.e. their size and shape. The average value of attribute of concern over a block $V$ can be represented as $\hat{z}_V(p_k) = \int_{V(p_k)} du' z(u')/|V|$, where $|V|$ is the areal size of block $V$ and $u'$ is the point locations within the block $V$ . The block kriging system is very similar to Eq. (2) with the covariance $C(p_k,p_j)$ replaced by the covariance between point observations and estimation block $\bar{C}(V(p_i),p_j)$ (Goovaerts, 1997), where $\bar{C}(p_k,p_j) = \int_{V(p_k)} du' C(u',p_j)/|V|$.

This study compares the township-based estimations of PM10 level by the methods above, i.e. NN, IDW, OK and BK. Among them, three different spatial resolutions are used in BK method, i.e. dividing the irregular areas into 5x5, 10x10 and 40x40 grids. The estimations are performed at all townships in Taipei during 2006 with the support of space-time PM10 observations during 1997-2007. Among them, the township-level PM10 estimations by NN, IDW and OK methods follow the conventional approach that uses the estimations at
geographical centroid to characterize the areal average level of pollutant concentration (Chen and Schwartz, 2009; Lertxundi-Manterola and Saez, 2009; Maheswaran and Elliott, 2003). In addition, to assure the comparison is performed at the same basis, the nonstationarity of PM10 process in space and time is removed by locally weighted smoothing regression method (LOESS) (Cleveland, 1979; Cleveland and Devlin, 1988) in advance to the applications of the spatial-time interpolation techniques.

3. Results

The observed PM10 levels across the stations in Taipei vary significantly as shown in Figure 3. The average and variance of monthly PM10 observations at the stations located in highly urbanized areas are generally higher than those in the city’s surrounding areas. Temporal variation of averaged PM10 across stations is shown in Figure 4 in which the increased PM10 variability happens at the time of higher average PM10 value, i.e. proportional effect. In general, the average of observed PM10 is higher during the seasons of spring and winter. In this study, the spatiotemporal processes of monthly PM10 is decomposed into a nonstationary trend in space and time to account for the general pattern of PM10 observations, e.g. PM10 variation resulting from the changes of seasons and distribution of emissions, and stationary residuals which account for the spatiotemporal dependence among the PM10 transport process. The spatiotemporal trend is estimated by using LOESS method which applies low-order polynomials to obtain the general pattern of PM10 variation at each space-time locality (Cleveland, 1979; Cleveland and Devlin, 1988). The estimation of spatiotemporal trend is performed by R software package. To characterize the spatiotemporal dependence, a space-time separable function is used as shown below (see Figure 5)

\[
c(h, r) = c_0 \exp\left(-\frac{3h}{a_r}\right) \exp\left(-\frac{3r}{a_t}\right)
\]

where \(c_0 = 150\) and \([a_r, a_t] = [2000\text{ m}, 3 \text{ month}]\). Eq. (3) is used in kriging methods, i.e. OK and BK, to account for the space-time covariance at point scale in space and time. In addition, the covariance function identifies the influential ranges of PM10 data in space and time, and therefore characterizes the space-time metric which is used to estimate the space-time distance for IDW and kriging methods as discussed above.

The estimations of the areal PM10 level during the period of 2004-2006 at all townships of irregular sizes and shapes are compared in this study and their differences are shown in Figure 6 in which BK1, BK2, and BK denote the results of block kriging with spatial resolution of 5 by 5, 10 by 10, and 40 by 40 elements of each township. Results show that the spatiotemporal dependence is important to the estimations, i.e. significant differences between the results from deterministic methods, i.e. NN and IDW, and stochastic methods, OK and BK. The spatial distribution of the mean squared difference of estimations between the results from BK and other methods are shown in Figure 7. Figures 7 (a) and (b) shows the comparison between BK and the deterministic methods, i.e., NN and IDW, in which the mean squared differences (MSD) are generally high and increase as the estimation location further apart from the cluster of monitoring stations located at city central. The comparisons among spatial patterns of kriging results in Figures 7 (c) and (d), for OK-BK and BK1-BK
Fig. 3. Spatial distribution of (a) mean and (b) standard deviation of PM10 observations.
Fig. 4. Temporal variation of averaged PM10 observations and its associated 95% confidence interval.

Fig. 5. Spatiotemporal covariance of PM10 observations across (upper) space and (bottom) time.
Fig. 6. Boxplot of the estimation differences between the results of block kriging with highest resolution and other methods.

Fig. 7. Spatial distribution of the mean squared differences between the methods of (a) NN-BK, (b) IDW-BK, (c) OK-BK, and (d) BK1-BK
respectively, show the relatively higher MSDs are located at the boundary of study area and townships surrounding by multiple stations with relatively drastic variations of their observations. It should be noted that estimation differences between deterministic methods and BK are generally higher than those in the comparisons between kriging methods. Similar results are shown in the comparisons of over time in Figure 8 in which the seasonality of MSDs is shown in all the comparisons that elevated estimation variability is elevated in spring and winter.

Fig. 8. Temporal distribution of mean squared differences among the methods of NN, IDW, OK, and BK with various spatial resolutions.

4. Discussions

Various methods of different level of complexity have been proposed and used in spatiotemporal estimation of air pollution exposure. Due to the privacy issue of personal data access, the spatiotemporal exposure estimation is often performed on the aggregated dataset with various geographical supports. As a result, certain approximations or assumptions of common techniques of point scale are required to be applied on the estimation of the average level over an area of concern, e.g. administrative division. Among them, the most common approximation is to use the air pollution estimation at the centroid to represent the average pollution level of the entire area of concern (Chen and Schwartz, 2008; Chen and Schwartz, 2009; Lertxundi-Manterola and Saez, 2009; Maheswaran and Elliott, 2003; Miller et al., 2007). The impact of the size and shape of geographical units to the estimation results is seldom considered (Goovaerts, 2008). The assumption of the similar
spatial resolution among observation and estimation areas can be dubious. In addition, the estimation ignores the variability of pollution levels within each of the studied units. Such approximation can be inadequate, especially, in the exposure estimation in metropolitan area where, in addition to its complex topography, various emission sources spreading across space elevates the spatiotemporal variability of air pollution. Though it is theoretical defective, exposure estimation at centroids usually provides a quick and convenient way to assess the air quality level in the corresponding areas of interest. The impact of this assumption of spatial prediction is worthwhile to be examined because the potential biased estimation results of air quality can distort their associated environmental epidemiology studies (Son et al., 2010).

As shown in Figure 6, comparing with the results of BK, the inconsistency is clearly shown between the estimations obtained from the two groups of methods, i.e. deterministic and stochastic, that the variability of the between-group differences are significantly larger than the within group differences. Among them, the results from NN method, i.e. a very common used method for ambient exposure estimation in environmental health studies (Basu et al., 2004; Miller et al., 2007), are significantly variable compared to those from other methods, i.e. its standard deviation is about 20% of common level of PM10. Figure 7 (a) shows that the discrepancy levels of NN results can vary across space and its generally high variability can be reduced at the areas of the higher density of monitoring stations. The periodic feature of the variation in Figure 8 shows that the temporal characteristics of NN results are still distinct from those of the results by other methods. It may resort to the fact that NN method is the only method disregarding the space-time dependence among the PM10 dataset in this study. Despite of its deterministic assumption, IDW accounts for the spatiotemporal dependence by considering the space-time metric among the observation and estimation locations. As a result, IDW reduces the discrepancy in both space and time across Taipei area between its results and those by kriging methods which considering the spatiotemporal dependence by not only space-time metric but the similarity among the observations. The standard deviation of the estimation variability compared to BK is about 10% of common PM10 level in Taipei. In addition, the results of deterministic approach are overestimated, especially at the townships with scarce monitors around, due to the preferential sampling at the highly polluted areas of Taipei central area (see Figure 3), i.e. the observations are mostly sampled at high concentration areas.

Figures 6-8 show the results from kriging methods, i.e. OK and BK with two spatial resolutions, are relatively similar. Among them, the standard deviation of estimation variability is lower than 5% of PM10 level. The distributions of MSDs among kriging methods in both space and time are relatively close to each other shown in Figure 7 (c) and (d). Among them, the areal estimation variability is further reduced by using BK instead of OK that their 95% confidence ranges are 2.8 and 6.8 \( \mu g / m^3 \), respectively. Because of the short spatial influential range of local PM10 transport, i.e. 2000m (see Eq. (3)), it implies the high spatial variability of PM10 distribution within the townships, i.e. point estimation results can vary significantly from location to location within townships. As a result, the consideration of geographical characteristics can effectively improve the understanding of the areal average level of PM10 concentration. This effect can be especially obvious at the townships containing and surrounding by multiple stations, e.g. San-chung township (the
township with two stations), because the nature of kriging methods completely appreciates the observations and therefore result in the higher variability of estimation results among these townships. It should be noted that all the estimations by the present spatial interpolation techniques only depend upon the PM10 observations. The general characteristics of these methods can only provide reasonable estimations within the convex of data locations (Olea, 1999).

5. Conclusions

This study applies several popular spatial techniques to assess the average areal PM10 level of townships in Taipei area. The comparison shows that the importance of the inclusion of space-time metric among the observation and estimation locations as well as the consideration of spatiotemporal dependence among the observations for the estimations. This study shows the consideration of the shape and size of the townships is important to the performance of the estimations of areal average concentration. The MSDs between the estimations by kriging methods with and without considering the geographical characteristics of townships are up to 6% and 2% of the common PM10 level across space and time, respectively. This study provides insights for the impact of geographical support to the exposure estimation in Taipei for environmental epidemiological studies.

6. Acknowledgement

This research was supported by the National Science Council of Taiwan (NSC 99-2625-M-002-007-).

7. References

Abbey DE, Lebowitz MD, Mills PK, Petersen FF, Beeson WL, Burchette RJ. Long-Term Ambient Concentrations of Particulates and Oxidants and Development of Chronic Disease in a Cohort of Nonsmoking California Residents. Inhalation Toxicology 1995; 7: 19-34.


Lertxundi-Manterola A, Saez M. Modelling of nitrogen dioxide (NO2) and fine particulate matter (PM10) air pollution in the metropolitan areas of Barcelona and Bilbao, Spain. Environmetrics 2009; 20: 477-493.


Leading air quality professionals describe different aspects of air pollution. The book presents information on four broad areas of interest in the air pollution field: the air pollution monitoring; air quality modeling; the GIS techniques to manage air quality; the new approaches to manage air quality. This book fulfills the need on the latest concepts of air pollution science and provides comprehensive information on all relevant components relating to air pollution issues in urban areas and industries. The book is suitable for a variety of scientists who wish to follow application of the theory in practice in air pollution. Known for its broad case studies, the book emphasizes an insightful of the connection between sources and control of air pollution, rather than being a simple manual on the subject.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following:
