Asian Journal of Atmospheric Environment
Vol. 10-4, pp. 197-207, December 2016
doi: https://doi.org/10.5572/ajae.2016.10.4.197

ISSN (Online) 2287-1160
ISSN (Print) 1976-6912

A Stochastic Approach for Prediction of Partially Measured Concentrations of Benzo[a]pyrene in the Ambient Air in Korea

Yongku Kim, Young-Kyo Seo[1)], Kyung-Min Baek[2)], Min-Ji Kim[2)] and Sung-Ok Baek[2),*]

Department of Statistics, Kyungpook National University, Republic of Korea
[1)]National Institute of Environmental Research, Republic of Korea
[2)]Department of Environmental Engineering, Yeungnam University, Republic of Korea

*Corresponding author. Tel: +82-53-810-2544, E-mail: sobaek@yu.ac.kr

ABSTRACT

Large quantities of air pollutants are released into the atmosphere and hence, must be monitored and routinely assessed for their health implications. This paper proposes a stochastic technique to predict unobserved hazardous air pollutants (HAPs), especially Benzo[a]pyrene (BaP), which can have negative effects on human health. The proposed approach constructs a nearest-neighbor structure by incorporating the linkage between BaP and meteorology and meteorological effects. This approach is adopted in order to predict unobserved BaP concentrations based on observed (or forecasted) meteorological conditions, including temperature, precipitation, wind speed, and air quality. The effects of BaP on human health are examined by characterizing the cancer risk. The efficient prediction provides useful information relating to the optimal monitoring period and projections of future BaP concentrations for both industrial and residential areas within Korea.

Key words: Benzo[a]pyrene, Cancer risk, Correlation function, K-nearest neighbor approach, Stochastic prediction

1. INTRODUCTION

Unlike many conventional air pollutants, such as SO_2, NO_x, CO, and O_3, hazardous air pollutants (HAPs) represent a serious risk to human health due to their carcinogenic properties and non-threshold nature of toxicity. Long-term exposure to HAPs, even at low levels, is known to have significant negative effects on human health (WHO, 2000). Among the many types of HAPs, polycyclic aromatic hydrocarbons (PAHs) have received special attention, as several PAHs are considered to be potential human carcinogens. Benzo[a]pyrene (BaP), a PAH, has been classified as a WHO group 1 carcinogen (WHO, 2010). PAHs are emitted into the atmosphere primarily through the combustion of fossil fuel and wood. Existing knowledge of the occurrence and chemistry of PAHs has previously been outlined in several reports (Suvarapu *et al*., 2012a, b; WHO, 1983).

Air quality management in Korea has been mainly focused on reducing the volumes of emissions of a number of priority pollutants. Recently, reducing risk to the general public has become the primary concern of many countries. Monitoring air quality is a key step for the risk assessment of a specific air pollutant, as it provides direct data relating to the exposure level for the general public (NRC, 1983). An accurate estimation of the exposure level requires a large number of reliable databases, which are available for only typical criteria pollutants such as SO_2, NO_x, and O_3 because these pollutants may be monitored continuously using automatic monitoring instruments. By contrast, it is often difficult to measure HAPs because this measurement requires more sophisticated sampling and analytical skills and because automatic monitoring technologies have not yet been fully established for many HAPs. In this regard, the daily monitoring of HAPs is not practical in many circumstances due to associated cost and labor constraints. As a result, HAPs are usually monitored only partially or intermittently in many countries. Hence, risk assessment based on such partial data may not be reliable, increasing the level of uncertainty relating to human health risk assessments.

A number of studies have attempted to characterize the atmospheric concentrations of PAHs in Korea. Park *et al*. (2002) reported that from 1998 to 1999 PAH concentrations in the Seoul Metropolitan Area (SMA) were significantly affected by fossil fuel usage for residential heating (also see Kim *et al*., 2012; Bae *et al*., 2002). There have also been some efforts in Korea to develop statistical models for describing the dynamic behavior of ambient air pollutants such as O_3 and par-

ticulate matter (PM), but no attempt has been made for any HAPs. Mastral *et al.* (2003) studied spatial and temporal PAH concentrations in Spain. Callén *et al.* (2010) considered a multivariate linear regression model to estimate BaP concentrations in Spain, based on meteorological conditions and PM_{10} concentrations. A statistical model is characterized by simplifying assumptions in the modeling scheme (while preserving the most important dynamic features) and quantifying the level of uncertainty through the probabilistic description of processes (Berliner, 2003). Recently, Kim *et al.* (2013) proposed a flexible statistical method for the estimation of unmeasured BaP concentrations within a given period of time and for a given site, by using an inference model established based on partially measured BaPs data from the site. The model incorporates the linkage between BaPs and meteorological factors, and is specifically formulated to identify meteorological effects and to allow for seasonal trends. The model is used to estimate future temporal fields of BaPs based on observed (or forecasted) meteorological conditions, including temperature, precipitation, wind speed, and air quality. Because observations of BaPs often contain a considerable degree of uncertainty, the model allows for this error of measurement.

This paper proposes a stochastic technique to predict the unobserved hazardous air pollutants (HAPs), especially Benzo[a]pyrene (BaP) which can have negative effects on human health. The proposed approach constructs a nearest-neighbor structure by incorporating the linkage between BaP and meteorological factors. Based on the proposed approach, we can determine the appropriate resolution of measured BaP concentrations to obtain annual BaP concentrations and identify outliers within the measured concentrations. Finally, the paper compares the results of various risk assessments between measured BaP data and estimated annual BaP data using stochastic prediction. By constructing a complete yearly database based on intermittent data through the statistical model, this study estimates more realistic exposure levels, and thus, enhances the accuracy of risk assessments.

2. STOCHASTIC PREDICTION

Kim *et al.* (2013) considered a two-stage statistical model to estimate the relationships between the BaP concentration and other factors such as weather, seasonality, long-term trends, and air pollutants. For $t = 1,\ldots,T$ and $c = 1,\ldots,C$, let Y_t^c denote the BaP concentration for site c on day t in a given year modeled as a gamma distribution, with an annual cycle in the form of a sine wave for mean intensity μ_t^c (see McCullagh and Nelder, 1989). Alternatively, a seasonally changed intercept model can be considered. Here, μ_t^c is modeled using an autocorrelation term with a time lag, air quality variables, and meteorological variables as covariates. The potential effect of weather is controlled for by including temperature, wind speed, and precipitation. In addition, smooth functions of the calendar time (natural cubic splines) are used to adjust for seasonality and long-term trends:

$$\log \mu_t^c = \mu^c + \alpha^c Y_{t-1}^c + \gamma_1^c C_t + \gamma_2^c S_t + \mathbf{x}_t^{c\prime}\boldsymbol{\beta}_1^c + \mathbf{z}_t^{c\prime}\boldsymbol{\beta}_2^c, \quad (1)$$

where

- μ^c is a logarithm of (constant) baseline BaP concentration for site c.
- Y_{t-1}^c is the HAP concentration for site c on the previous day $t-1$.
- $C_t = \cos(2\pi(t-181)/365)$ and $S_t = \sin(2\pi(t-181)/365)$ are cosine and sine waves for the annual cycle.
- $\mathbf{x}_t^c$ indicates air quality variables influencing BaP concentrations for site c on day t (e.g., SO_2, PM_{10}, O_3, NO_2, CO).
- $\mathbf{z}_t^c$ indicates meteorological variables for site c on day t (e.g., temperature, wind speed, wind direction, precipitation amount, precipitation occurrence and dew point). To avoid multicollinearity, ambient temperature are deseasonalized.

One of the limitations of stochastic model is a marked (strong) sensitivity to observed (or forecasted) meteorological conditions, including temperature, precipitation, wind speed, and air quality. Under the worst-case scenario, the statistical model often produces extremely high values which are not observed for partially observed BaP concentrations. That is, the models tend to overestimate the observed variance of BaP concentrations. To reduce this phenomenon, we introduce new approaches toward the prediction of unobserved BaP concentrations.

2.1 Semiparametric Approach

The framework follows a K-nearest neighbor approach. We describe the framework and the implementation algorithm for temporal BaP concentrations. K-nearest neighbor approach queries days similar to a given feature vector and identify a subset of days (K) similar to the feature day. These K days are then weighted using a bi-square weight function and randomly sampled to generate ensembles. Here we incorporate K-nearest neighbor approach to detrended logarithm of BaP concentrations, denoted by ε_t^c. That is,

$$\varepsilon_t^c = \log(Y_t^c) - (\mu^c + \alpha^c Y_{t-1}^c + \gamma_1^c C_t + \gamma_2^c S_t + \gamma_3^c P_t^c), \quad (2)$$

where P_t^c is an index of the precipitation occurrence for site c on day t.

In a simple K-nearest neighbor approach, the nearest neighbors are obtained from the observed data by computing the distance between the unobserved BaP concentration and the observed BaP concentration, and the neighbors are assigned weights based on their distance. The weight function gives more weight to the nearest neighbors and less to the farthest neighbors. In general, the number of nearest neighbors, K is based on the heuristic scheme $K=\sqrt{N}$, where N equals the sample size (Lall and Sharma, 1996), following the asymptotic arguments of Fukunaga (1990). Objective criteria, such as generalized cross validation, can also be used. The proposed approach aims to compute the distance based on the similarity of factors such as air quality variables and meteorological variables. That is, for two distinct time points, t and t' for site c, the distance $d_c(t,t')$ can be defined by

$$d_c(t,t') = w_1\|\mathbf{x}_t^c - \mathbf{x}_{t'}^c\| + w_2\|\mathbf{z}_t^c - \mathbf{z}_{t'}^c\|, \tag{3}$$

where $\|\cdot\|$ is a matric or distance function. Note that air quality variables $\mathbf{x}_t^c$ and meteorological variables $\mathbf{z}_t^c$ need to be standardized before computing distances. The weights w_1 and w_2 plays a role of tuning the proximities between air quality variables and meteorological variables. For a fixed time point t, its neighbors are assigned weights based on the K smallest distances (say $d_1,d_2,\ldots,d_K$) from the time point t, as follows:

$$W(k) = \frac{1/d_k^\beta}{\sum_{j=1}^K 1/d_j^\beta} \text{ or } W(k) = \frac{\exp(-\beta d_k)}{\sum_{j=1}^K \exp(-\beta d_j)} \tag{4}$$

for $k=1,\ldots,K$. Further extensions to (4) are possible, such as extensions to include a more general class of functions of distance. The index for site c is omitted for brevity for now. The unobserved detrended logarithm of BaP concentrations can then be generated by resampling from the observed detrended logarithm of BaP concentrations based on the assigned weights, or alternatively, can be estimated based on the weighted average of the observed detrended logarithm of BaP concentrations. Based on resampled or estimated detrended logarithm of the BaP concentration, $\hat{\varepsilon}_t^c$, we get the resampled or estimated BaP concentration $\hat{Y}_t^c$ by the following equation:

$$\hat{Y}_t^c = \exp(\hat{\mu}^c + \hat{\alpha}^c Y_{t-1}^c + \hat{\gamma}_1^c C_t + \hat{\gamma}_2^c S_t + \hat{\gamma}_3^c P_t^c + \hat{\varepsilon}_t^c), \tag{5}$$

where $\hat{\mu}^c$, $\hat{\alpha}^c$, $\hat{\gamma}_1^c$, $\hat{\gamma}_2^c$ and $\hat{\gamma}_3^c$ are the estimated coefficients of the trend GLM model fitted to the observed BaP concentrations. In principle, we may extend our neighbor structure not only to a temporal scale but also to a spatial scale. That is, other sites can also be used as neighbors. In such cases, a more sophisticated and complex measure of distance is required.

2.2 Parametric Approach

In general, the approach based on resampling has unrealistic properties for extremes. That is, it is not possible to generate value greater than the highest observed. Because those air quality and meteorological factors which are more similar to BaP processes are assigned higher weights, the expression (3) accommodates a spatial dependence structure. In general, BaP concentrations are assumed to be log-normally distributed. Here, the BaP concentration Y_t^c is modeled as a lognormal distribution, as follows:

$$(\log(Y_1^c),\log(Y_2^c),\ldots,\log(Y_T^c))' \sim \text{MVN}(\boldsymbol{\mu}^c, \boldsymbol{\Sigma}^c), \tag{6}$$

where $\boldsymbol{\mu}^c = (\mu_1^c,\mu_2^c,\ldots,\mu_T^c)'$ is a mean trend vector of the logarithm of the BaP concentration and $\boldsymbol{\Sigma}^c$ is a covariance matrix of the logarithm of the BaP concentration process. The mean process μ_t^c is modeled by

$$\mu_t^c = \mu^c + \alpha^c Y_{t-1}^c + \gamma_1^c C_t + \gamma_2^c S_t + \gamma_3^c P_t^c. \tag{7}$$

The temporal dependence among the logarithms of BaP concentrations is incorporated by modeling $T\times T$ covariance matrix $\boldsymbol{\Sigma}^c$ whereas the temporal mean trend is represented by μ_t^c. We can parameterize the covariance matrices by variance terms and correlation functions as $\boldsymbol{\Sigma}^c(t,t') = \sigma^2 r(t,t')$.

To avoid the difficulty in dealing with covariance matrices, we can consider a simple exponential correlation function as follows:

$$\mathrm{r}_\beta(d_c(t,t')) = \exp(-\beta|d_c(t,t')), \tag{8}$$

where $\beta(>0)$ is a function which is used to tune the range of individual neighbors. This function provides a simple covariance matrix, and the resulting matrices can be easily manipulated. Let $\mathrm{r}_\beta(d_c(t,t')) = \rho^{d_c(t,t')}$ (i.e., $\rho = \mathrm{e}^{-\beta} > 0$). The correlation parameter, β, can be empirically estimated from the data. In principle, we can consider a more general correlation function

$$\rho_{\rho,\upsilon}(t,t') = \frac{(d_c(t,t')/\rho)^\upsilon}{2^{\upsilon-1}\Gamma(\upsilon)}\kappa_\upsilon(d_c(t,t')/\rho), \tag{9}$$

which is called the Matérn correlation function (Matérn, 1986) with parameters ρ and ν, evaluated at distance $d_c(t,t')$.

Let $\mathbf{U}$ and $\mathbf{O}$ be the vector of logarithms of unobserved BaP concentrations and the vector logarithms of observed BaP concentrations, respectively. In our scheme, assume that $\mathbf{U}$ and $\mathbf{O}$ are jointly distributed, that is,

$$\begin{pmatrix}\mathbf{U}\\ \mathbf{O}\end{pmatrix} \sim \left(\begin{pmatrix}\boldsymbol{\mu}_u\\ \boldsymbol{\mu}_o\end{pmatrix}, \begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}\right), \tag{10}$$

where Σ_{11} is the marginal covariance matrix of $\mathbf{U}$, Σ_{22} is the marginal covariance matrix of $\mathbf{O}$, and Σ_{12} is the

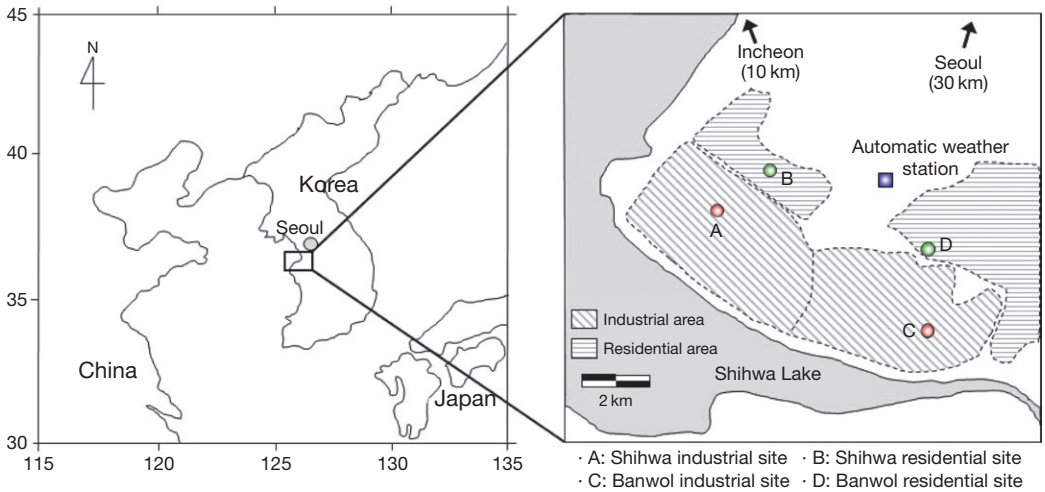

Fig. 1. Location of sampling sites for benzo(a)pyrene.

cross-covariance matrix of $\mathbf{U}$ and $\mathbf{O}$. Note that $\boldsymbol{\mu}_u$, $\boldsymbol{\mu}_o$, Σ_{11}, Σ_{12}, Σ_{21} and Σ_{22} are obtained from the rearrangement of $\boldsymbol{\mu}_c$ and Σ^c. Then,

$$\mathbf{U}|\mathbf{O} \sim (\boldsymbol{\mu}_u - \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{O} - \boldsymbol{\mu}_o), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}. \quad (11)$$

The predictor for logarithms of unobserved BaP concentrations, $\mathbf{U}$, is then the usual best linear unbiased estimator (BLUE) for the prediction problem. Therefore, we may estimate BaP concentration by $\exp(\boldsymbol{\mu}_u - \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{O} - \boldsymbol{\mu}_o))$ with the uncertainty in terms of $\exp(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$. Note that the parametric approach provides prediction with uncertainty.

Because BaP concentrations are typically available on a daily basis, considering various time lag models for $\mathbf{x}_t$ or $\mathbf{z}_t$ allows for greater flexibility in exploring the lag between air quality and meteorological variables and BaP concentrations, as compared to single-lag models. Another topic of special interest lies in the hierarchical modeling of the multivariate or spatial structure of BaP concentrations. Here, this paper focuses on the temporal behavior of BaP concentrations.

3. APPLICATION TO BAP DATA

3.1 Air Quality Data

The BaP and other air quality data sets used in this paper are obtained from a field study conducted in the Sihwa-Banwol national industrial complex, one of the largest industrial areas in Korea (Seo, 2010). The main purpose of the field study was to characterize the occurrence and concentrations of a wide range of HAPs in the atmosphere of the industrial complex. Temporal, spatial, and seasonal variations in HAP concentrations were determined at four sites (two industrial and two residential sites) over a two-year period, from August 2005 to July 2007. Fig. 1 shows the location of the industrial complex as well as the specific locations of air sampling sites. Seasonal HAP monitoring campaigns were carried out throughout the complex, where 12 and 10 consecutive days of monitoring are carried out for each season in the first year and the second year of operation, respectively. Detailed monitoring periods and air quality data for each site are summarized in Table 1. Samples of total suspended particles (TSPs) are collected by high-volume sampling for 24 h, and are used for PAH analysis. A total of 38 different PAH compounds are determined by GC/MS. Sampling, analyses, and quality control for PAHs are carried out in accordance with the US EPA TO-13A protocol (US EPA, 1999) and the ISO method (ISO, 2000). Interpretations of PAH determinations should be based on the objective of PAH measurement, e.g., whether the intention is to monitor for concentrations which exceed reference values, analyzing causes, or conducting epidemiological surveys. In the case of BaP, the PAH substance most frequently measured in air pollution analyses, close attention should be paid to the implications which are associated with these concentrations, as the concentrations themselves can be used to indicate the carcinogenic activity of PAH mixtures that regularly occur in ambient air (ISO, 2000). According to the WHO Air Quality Guidelines (WHO, 2000), BaP is a useful indicator of the carcinogenic potential of total PAH emissions with respect to lung cancer. For these reasons, BaP is selected for this paper's analysis as a good representative PAH.

PAH sampling sites are all located within national

Table 1. Summary of air quality, meteorological, and BaP data.

6/24/2005-7/23/2006		Site A		Site B		Site C		Site D	
Variable	N	Mean	SD	Mean	SD	Mean	SD	Mean	SD
SO_2 (ppb)	364	9.4	6.2	10.8	5.2	9.6	5.4	6.5	3.1
PM_{10} ($\mu g/m^3$)	364	71.7	70.5	90.0	57.4	96.2	76.6	57.3	44.3
O_3 (ppb)	364	20.9	11.5	18.3	9.2	20.8	10.6	17.4	8.7
NO_2 (ppb)	364	29.4	13.2	23.8	10.8	24.3	12.1	30.4	13.2
CO (ppb)	364	620.8	366.6	1013.0	378.1	817.7	546.5	588.0	330.6
Temp (°C)	364	12.7	10.4	12.7	10.4	12.7	10.4	12.7	10.4
Wind speed (m/s)	364	1.8	0.8	1.8	0.8	1.8	0.8	1.8	0.8
Rain (mm)	364	3.1	10.8	3.1	10.8	3.1	10.8	3.1	10.8
Benzo[a]pyrene (ng/m^3)	48	1.0	1.1	1.2	1.2	1.0	1.0	1.1	1.2

Table 2. Various percentiles of observed and augmented BaP concentrations (ng/m^3) for four sites.

Percentile	Observed				Semiparametic				Parametic			
	80th	90th	95th	99th	80th	90th	95th	99th	80th	90th	95th	99th
Site A	1.5	2.4	3.2	4.3	1.7	2.2	2.5	3.7	1.8	2.4	3.1	3.8
Site B	1.9	2.9	3.3	4.8	1.8	2.4	2.8	3.9	1.9	2.4	3.2	4.3
Site C	1.4	2.5	3.4	3.7	1.6	2.2	2.5	3.4	1.6	2.3	2.6	3.6
Site D	1.7	2.9	3.7	4.4	2.2	2.7	3.4	3.8	2.1	2.6	3.1	4.0

air quality monitoring stations operated by the Korean Ministry of Environment (see Fig. 1). Hourly data on other general air quality parameters such as SO_2, CO, NO_2, PM_{10}, and O_3, were obtained from the national air quality database. In addition, hourly data on ambient temperature and wind speed were obtained from the National AWS (automatic weather station), which is located approximately 5 km from the PAH sampling sites. Data for relative humidity and solar radiation data were not available from the AWS. The daily average air quality and meteorological data was recalculated so as to match the PAH sampling time. It has been reported that atmospheric concentrations of PAHs are well related to CO and NO_2 as byproducts of combustion from many sources (Saud *et al*., 2011; Ravindra *et al*., 2008; Park *et al*., 2002). In addition, ambient levels of PAHs are higher in winter and lower in summer, and thus, indicate a negative correlation between PAHs and ambient temperature (Lee *et al*., 2011; Park *et al*., 2002).

3.2 Prediction of BaP Concentrations

Through the semi-parametric approach, we considered the weight function to be proportional to the exponentiated distance, as the augmented BaP concentrations based on the weight function proportional to the inverse of distance fail to produce enough variability of BaP concentrations. Various percentiles of daily BaP concentrations are compared in Table 2. Based on the parametrically augmented BaPs, BaP concentrations show similar variability to the observed BaP concentrations. However, semiparametrically augmented BaPs tends to be underestimated, especially during their peak season (December). The proposed approaches can provide a useful tool for exploring unobserved periods. Because of the small number of disconnected observations, the effects of autocorrelation terms in the long-term trend model are limited. However, the autocorrelation of air quality and meteorological variables still indirectly takes into account the effect of the autocorrelation term on BaP concentrations. In addition, retaining the autocorrelation term as a covariate makes the interpretation of the model more difficult and complicates the relationships between air quality and meteorological variables.

Figs. 2 and 3 show the performance of the proposed approaches in reproducing BaP concentrations for four sites. Time series of daily BaP concentrations are simulated over a one-year period. Augmented BaP concentrations show much less variability during summer months than winter months, and the observation period may miss the peak season (December) of BaP concentrations. Estimated cumulative density functions of seasonally observed BaP concentrations were compared with those of semiparametrically and parametrically augmented BaP concentrations (see Fig. 4). Figs. 5 and 6 clearly show the variability of observed and augmented BaP concentrations, both regionally and

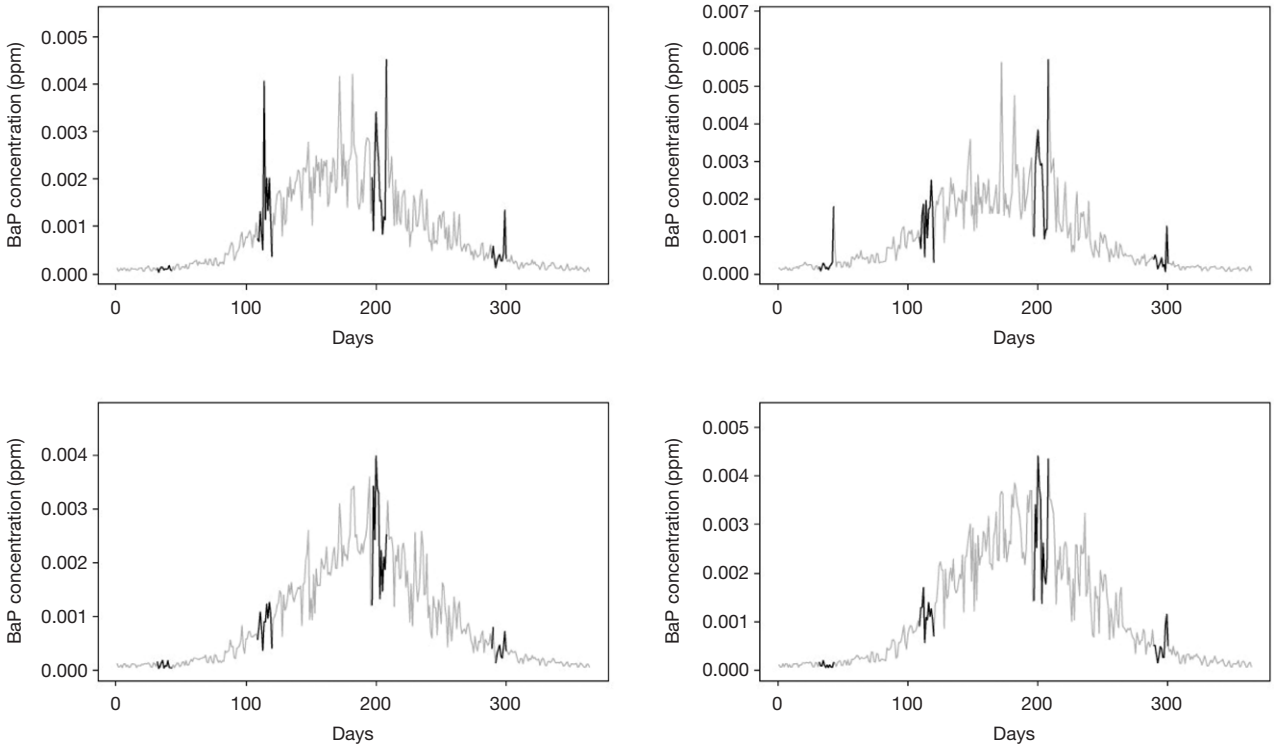

Fig. 2. Observed BaP concentrations (black) for the period 2005-2006 and (semi-parametrically) augmented BaP concentrations (grey) for four sites (A, B, D, and C; clockwise from top left).

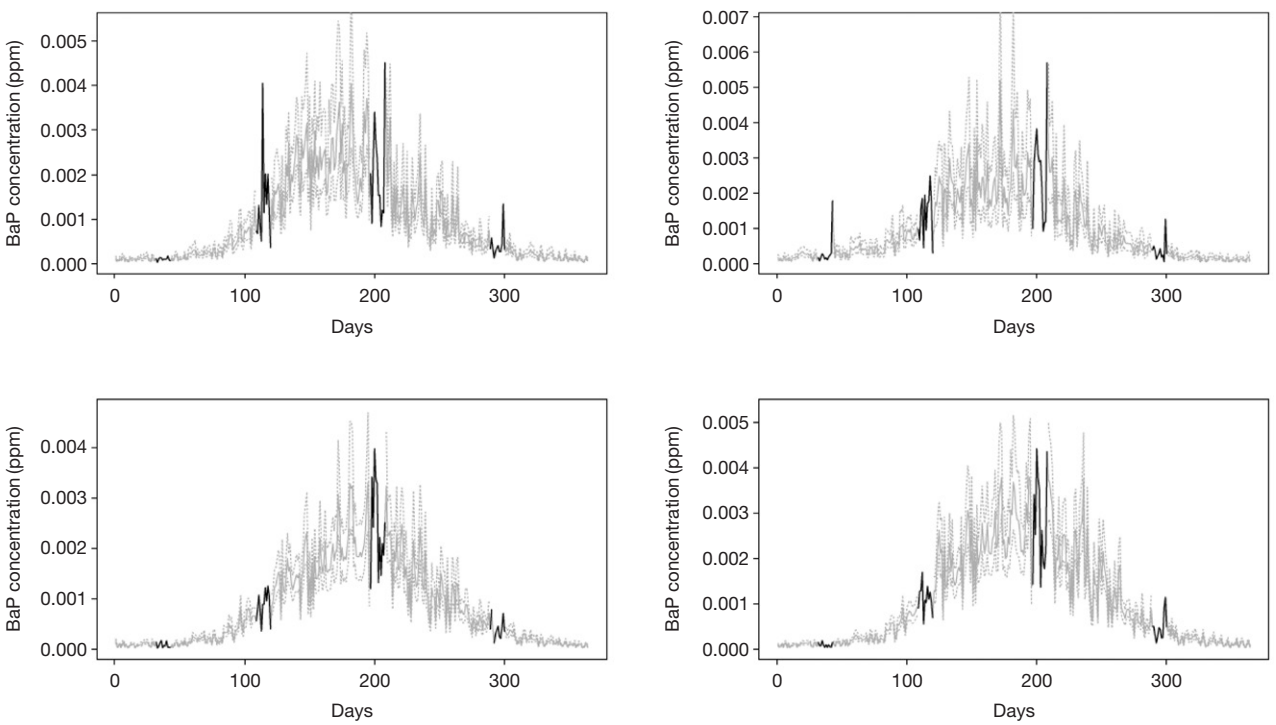

Fig. 3. Observed BaP concentrations (black) for the period 2005-2006 and (parametrically) augmented BaP concentrations (grey) with uncertainty (dotted) for four sites (A, B, D, and C; clockwise from top left).

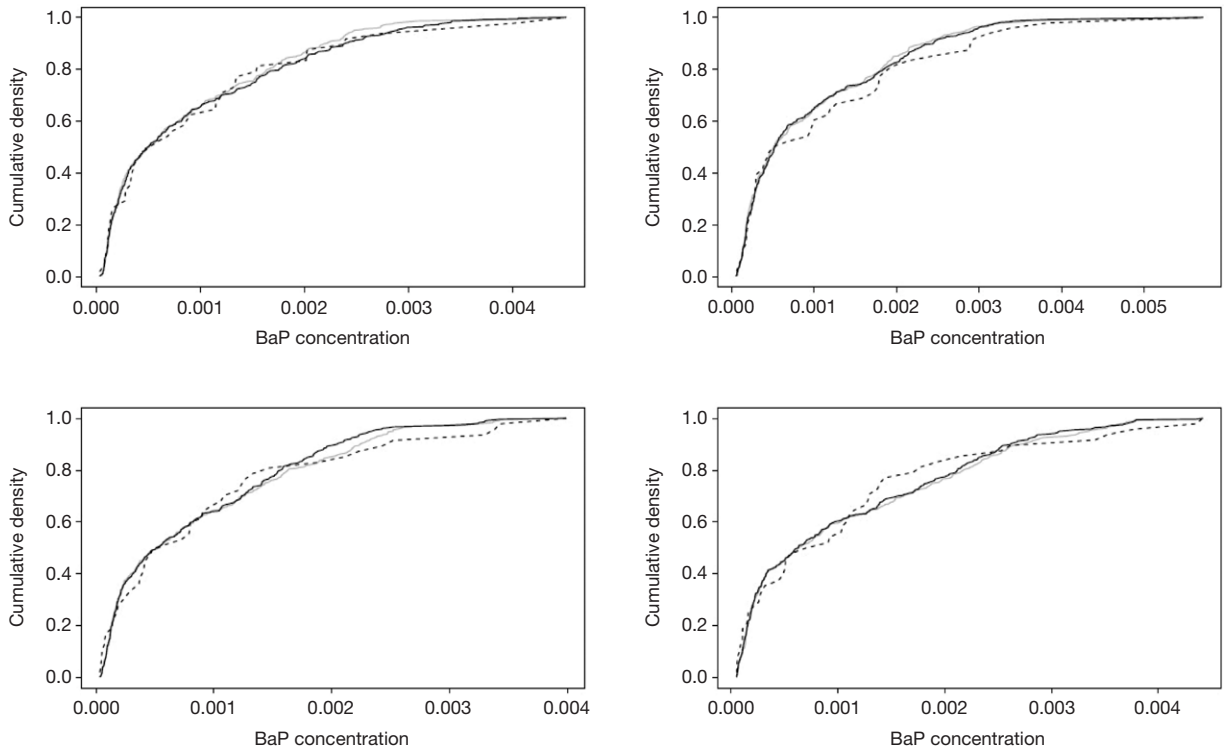

Fig. 4. Esitmated cumulative density functions based on semi-parametrically augmented BaP concentrations (grey) and parametrically augmented BaP concentrations (black) for four sites (A, B, D, and C; clockwise from top left; the dashed line corresponds to observed BaP concentrations).

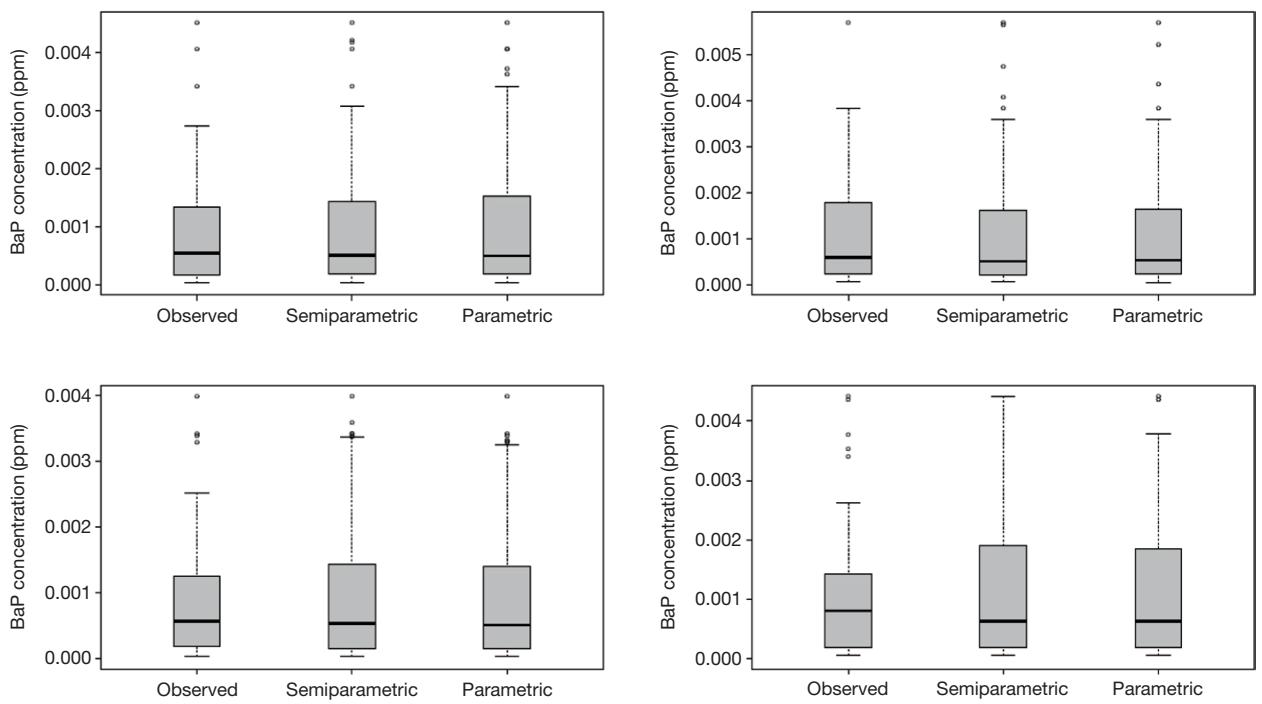

Fig. 5. Boxplots of observed and modeled BaP concentrations for four sites (A, B, D, and C; clockwise from top left).

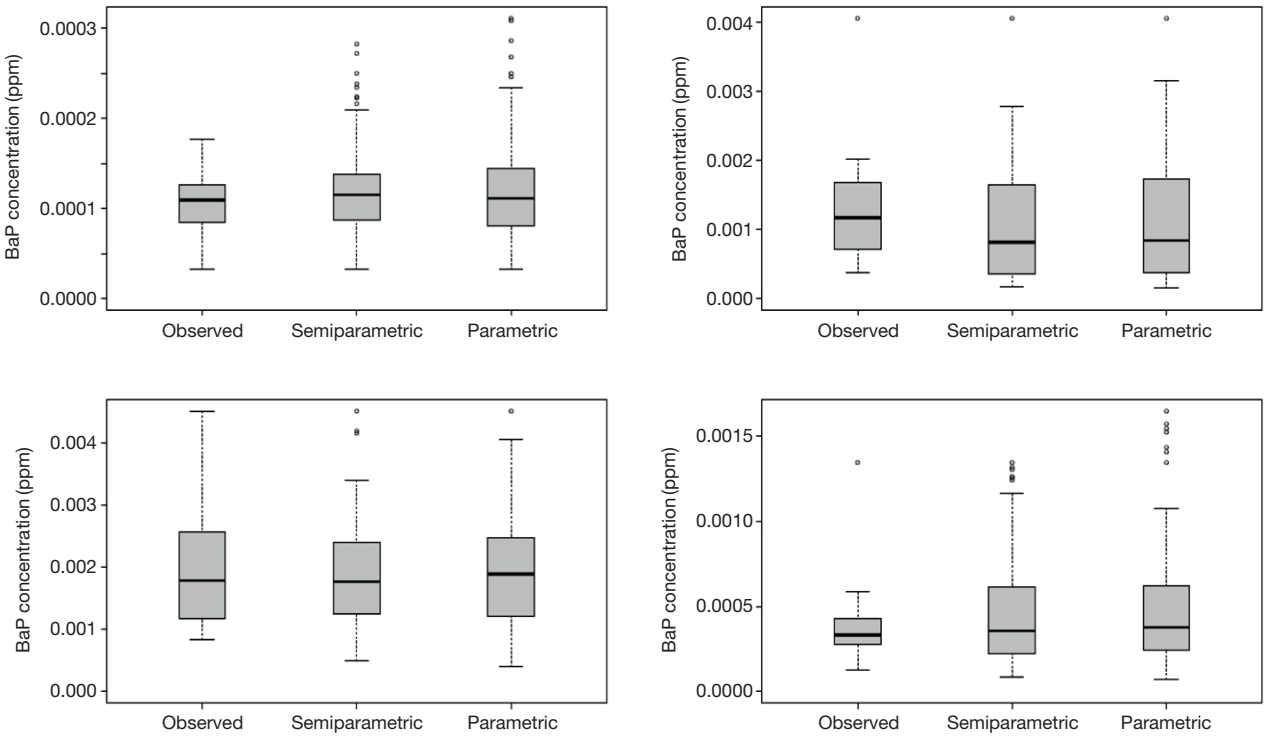

Fig. 6. Boxplots of observed and modeled BaP concentrations for site A for spring (top left), summer (top right), autumn (bottom left), and winter (bottom right).

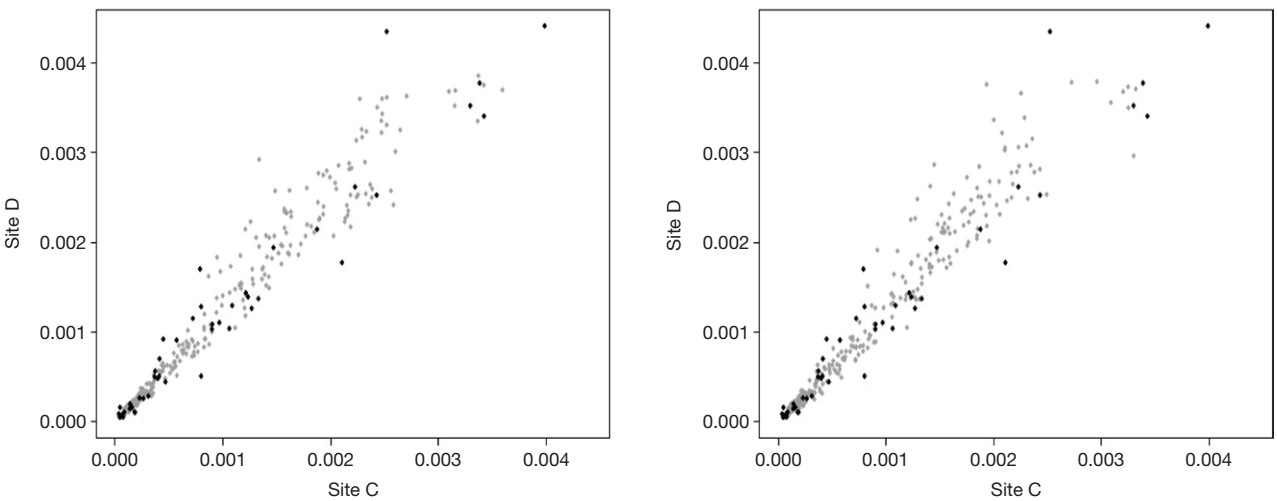

Fig. 7. Plots of observed BaP concentrations (black), semi-parametrically (left) and parametrically (right) augmented BaP concentrations 9 gray) near sites.

seasonally, particularly for high BaP concentrations. Their inconsistency is partly due to the lack of alignment of observation times, and is potentially due to adverse meteorological conditions (e.g., severe episodes). Fig. 7 shows that the proposed model well preserves the relationship between BaP concentrations near sites. The forecasting behavior of the proposed model and the approach taken in the previous section is assessed through a standard "hold-out" experiment (see West and Harrison, 1999). More specifically, the

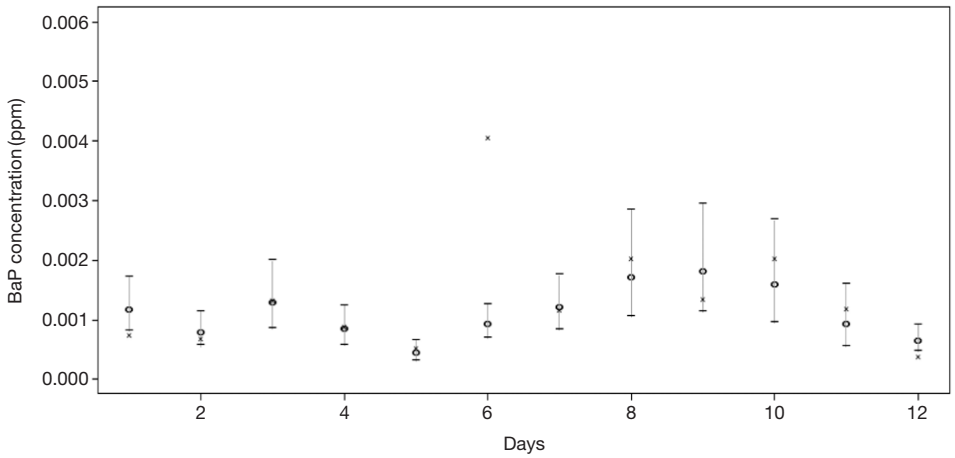

Fig. 8. Observed BaP concentrations (gray x) for the winter 2005-2006 period and corresponding predicted intervals of BaP concentrations (gray) for site A.

model parameters are estimated using observed BaP concentrations after the removal of a subset (called the validation set) of data from the 2005-2006 season, and forecasting of the removed BaP concentrations. The forecast is then compared with actual BaP concentrations during the validation period. During the forecast period, variables for the observed air quality data and meteorological conditions are used for each region. Fig. 8 shows the actual BaP concentrations for the 2005-2006 season, as well as their predicted intervals for the Shiwha residential area (site A).

3.3 Effects on Risk Assessment

For the assessment of our approaches, as well as to determine the distribution of this variable over the population residing in the risk area, cancer risk was examined for different HAP sources. First, an exposure model was first constructed based on the lifetime average daily dose (*LADD*) involving the active surface area concentration of the HAP (*C*, μg/m^3), the inhalation rate (*IR*, m^3/day), exposure duration (*ED*, years), exposure frequency (*EF*, days/year), absorption efficiency (*AB*, %), body weight (*BW*, kg), and the averaging time (*AT*, days):

$$LADD = \frac{C \times IR \times ED \times EF \times AB}{BW \times AT}. \quad (12)$$

Finally, the cancer risk for a specific HAP can be obtained using inhalation slope factor (*SF*, (μg/kg/day)$^{-1}$) as follows:

$$Cancer\ Risk = LADD \times SF, \quad (13)$$

where $SF = \frac{UR \times BW}{IR}$ and *UR* is inhalation unit risk.

Table 3. Summary of terms associated with the cancer risk.

Terms	Unit	Variability
Unit risk (UR)	per μg/m^3	7.8e−06
Inhalation rate (IR)	m^3/day	$N(13, 0.9^2)$
Exposure duration (ED)	years	25
Frequency of exposure (EF)	days/year	Triangle distribution (180,345,365)
Absorption efficiency (AB)	%	100
Body weight (BW)	kg	$N(62, 8.8^2)$
Averaging time (AT)	days	70 years = 25,500 days

The level of chemical pollution was modeled as a function of other variables, some of which are random variables, and the distribution of the variable of interest was generated through a computer simulation. This then allowed for the determination of whether the probability of the variable of interest exceeds acceptable levels.

The Monte Carlo simulation has been used for risk assessment purposes because of its increased computing power. Here, it includes *LADD* and *Cancer Risk* as functions of the other variables. Each recalculation produces new random values for *IR*, *EF*, and *BW*, and consequently, for *LADD* and *Cancer Risk* to simulate the situation for a random individual from the population at risk. Fig. 9 compares the distribution of the cancer risk for the observed BaP concentrations with that for augmented BaP concentrations, for both industrial and residential areas in Korea. Note that augmented BaP concentrations provide more variations in the cancer risk (ng/m^3) of residential area but there is not "worst-case" scenario with extremely high cancer risk values.

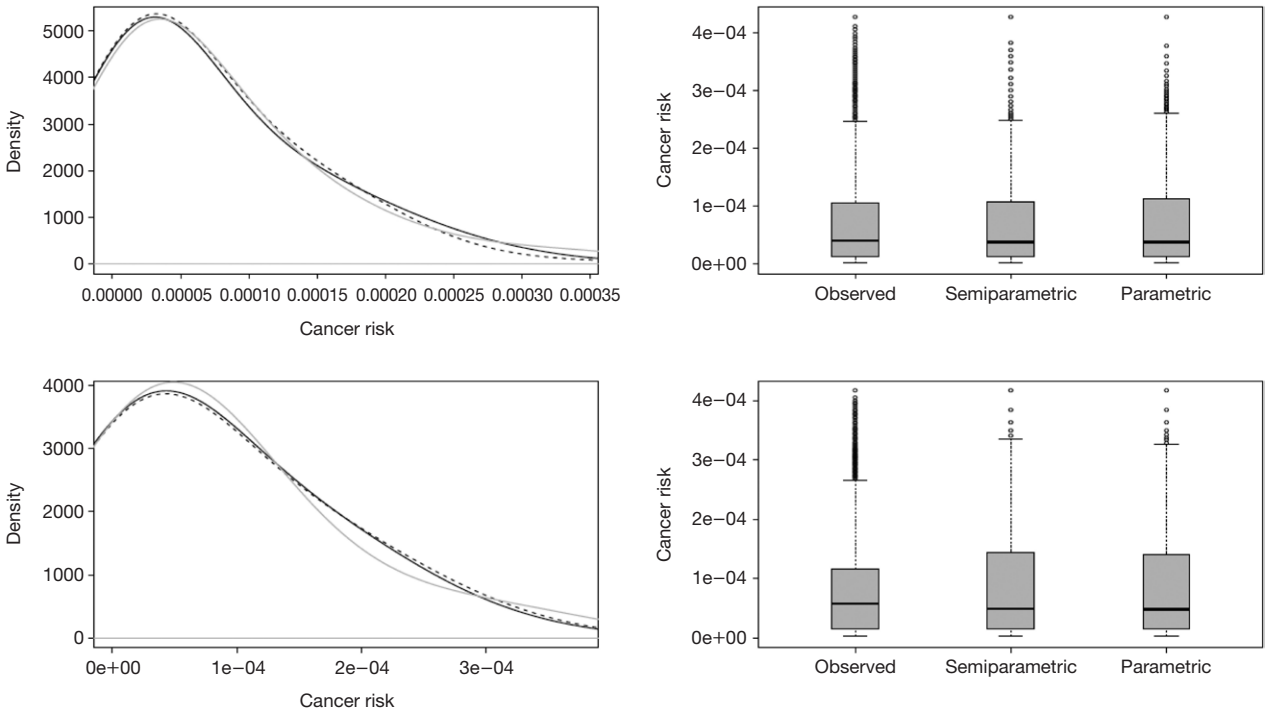

Fig. 9. Esitimated density (left) and boxplots (right) of cancer risks based on observed BaP concentrations (gray), semi-parametrically augmented BaP concentrations (black; dash) and parametrically augmented BaP concentrations (black; line) for site A (top) and site D (bottom).

4. CONCLUDING REMARKS

For various reasons, BaP concentrations are observed only for 48 days in a full year at the site in question (12 days in each season). Thus, it is very difficult to estimate unobserved HAP concentrations. It has been shown that the proposed approaches can be extended to effectively estimate BaP concentrations. The proposed nearest-neighboring structure in terms of observed meteorological conditions, including temperature, precipitation, wind speed, and air quality. Based on augmented BaP concentrations, it shows much less variability during summer months, and unobserved BaP concentrations can be estimated through more realistic low-frequency statistical properties and without any apparent deterioration in high-frequency characteristics. With this improvement, the proposed approach should provide more realistic risk assessments based on Monte Carlo simulations.

Studies of the behavior of BaPs should rely heavily on both observations and physically based models. However, our approach might fail to give a full physical explanation of the heterogeneity of BaP dynamics. This is different from traditional physical modeling, perhaps with data-based parameter estimates, and traditional statistical modeling, perhaps relying on vague, qualitative physical reasoning. That is, analysis was mainly focused on parameters and stochastic components of uncertainty. The structural uncertainty is satisfied by performing validation of the study, in which the goodness of fit of the model is assessed.

Implanting spatial dependence, which was not considered here, into our approaches is another challenge. In principle, it can be incorporated by a conditional autoregressive (CAR) model based on proximity matrix or by modeling covariance matrix Σ, which is parameterized by variance term and correlation function. That is, $\sigma(s,s') = \sigma^2 R_{\rho,\nu}(s,s')$, where $R_{\rho,\nu}(s,s')$ is a positive definite correlation function at two sites s and s' in $\mathbb{R}^2$. One important limitation of the proposed approaches is the unrealistic estimation of extreme values. That is, it is hard to accurately estimate the maximum BaP concentrations which occur.

ACKNOWLEDGEMENT

This study was conducted as a part of projects “Monitoring of Hazardous Air Pollutants in Sihwa-Banwol Industrial Complex in Korea, 2005 to 2007”, which were financially supported by National Institute of Environmental Research in Korea. The research of

Yongku Kim was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. 2015R1D1A1A01057847).

REFERENCES

Bae, S.Y., Yi, S.M., Kim, Y.P. (2002) Temporal and spatial variations of the particle size distribution of PAHs and their dry deposition fluxes in Korea. Atmospheric Environment 36, 5491-5500.

Berliner, L.M. (2003) Physical-statistical modeling in geophysics. Journal of Geophysical Research 108(D24), STS3 1-10.

Callén, M.S., López, J.M., Mastral, A.M. (2010) Seasonal variation of benzo(a)pyrene in the Spanish airborne PM_{10}. Multivariate linear regression model applied to estimate BaP concentrations. Journal of Hazardous Materials 180, 648-655.

Fukunaga, K. (1990) Introduction to statistical pattern recognition, Academic, San Diego, California.

ISO (2000) Ambient air - Determination of total (gas and particle-phase) poly-cyclic aromatic hydrocarbons - Collection on sorbent backed filters with gas chromatographic/mass spectrometric analyses. International standard, ISO 12884. 25 p.

Kim, J.Y., Lee, J.Y., Choi, S.D., Kim, Y.P., Ghim, Y.S. (2012) Gaseous and particulate polycyclic aromatic hydrocarbons at the Gosan background site in East Asia. Atmospheric Environment 49, 311-319.

Kim, Y., Seo, Y.K, Baek, S.O. (2013) A statistical inference for concentrations of benzo[a]pyrene partially measured in the ambient air of an industrial city in Korea. Atmospheric Environment 81, 92-101.

Lall, U., Sharma, A. (1996) A nearest neighbor bootstrap for resampling hydrologic time series. Water Resources Research 32, 679-693.

Lee, J.Y., Kim, Y.P., Kang, C.H. (2011) Characteristics of the ambient particulate PAHs at Seoul, a mega city of Northeast Asia in comparison with the characteristics of a background site. Atmospheric Environment 99, 50-56.

Mastral, A.M., López, J.M., Callén, M.S., García, T., Murillo, R., Navarro, M.V. (2003) Spatial and temporal PAH concentrations in Zaragoza, Spain. Science of the Total Environment 307, 111-124.

Matérn, B. (1986) Spatial Variation, Springer-Verlag, second edition.

McCullagh, P., Nelder, J.A. (1989) Generalized Linear Models, London: Chapman & Hall/CRC.

NRC (1983) Risk assessment in the federal government: managing the process, National Research Council. National Academy Press, Washington, DC. 191 p.

Park, S.S., Kim, Y.J., Kang, C.H. (2002) Atmospheric polycyclic aromatic hydrocarbons in Seoul, Korea. Atmospheric Environment 36, 2917-2924.

Ravindra, K., Sokhi, R., Grieken, R.V. (2008) Atmospheric polycyclic aromatic hydrocarbons: Source attribution, emission factors and regulation. Atmospheric Environment 42, 2895-2921.

Saud, T., Mandal, T.K., Gadi, R., Singh, D.P., Sharma, S.K., Saxena, M., Mukherjee, A. (2011) Emission estimates of particulate matter (PM) and trace gases (SO_2, NO, and NO_2) from biomass fuels used in rural sector of Indo-Gangetic Plain, India. Atmospheric Environment 45, 5913-5923.

Seo, Y.K. (2010) Occurrence and behaviour of hazardous air pollutants in a large industrial area, Ph.D. thesis, Yeungnam University.

Suvarapu, L.N., Seo, Y.K., Lee, B.S., Baek, S.O. (2012a) A review on the atmospheric concentrations of polycyclic aromatic hydrocarbons (PAHs) in Asia since 2000 - part I: data from developed countries. Asian Journal of Atmospheric Environment 6, 147-168.

Suvarapu, L.N., Seo, Y.K., Cha, Y.C., Baek, S.O. (2012b) A review on the atmospheric concentrations of polycyclic aromatic hydrocarbons (PAHs) in Asia since 2000 - part II: data from developing countries. Asian Journal of Atmospheric Environment 6, 169-191.

US EPA (1999) Compendium method TO-13A determination of polycyclic aromatic hydrocarbons (PAHs) in ambient air using gas chromatography/mass spectrometry (GC/MS). U.S. environmental protection agency, Cincinnati, 78 p.

West, M., Harrison, J. (1999) Bayesian Forecasting and Dynamic Models, Springer, New York.

WHO (1983) IARC Monographs on the evaluation of carcinogenic risks to humans; Polynuclear aromatic compounds, part 1, chemical, environmental and experimental data, 32, Lyon, France. 55 p.

WHO (2000) Air quality guidelines for Europe second edition, WHO regional publication, European Series No. 91. 273 p.

WHO (2010) IARC Monographs on the evaluation of carcinogenic risks to humans, Some non-heterocyclic polycyclic aromatic hydrocarbons and some related exposures, 92, Lyon, France. 853 p.

(Received 16 August 2016, revised 26 September 2016, accepted 26 September 2016)