Spatio-temporal models for generating a map of high resolution NO2 level†

Sanghoo Yoon1 · Mingyu Kim2

1Department of Computer Science and Statistics, Daegu University
2WISE institute, Hankuk University of Foreign Studies
Received 4 March 2016, revised 11 May 2016, accepted 17 May 2016

Abstract

Recent times have seen an exponential increase in the amount of spatial data, which is in many cases associated with temporal data. Recent advances in computer technology and computation of hierarchical Bayesian models have enabled to analyze complex spatio-temporal data. Our work aims at modeling data of daily average nitrogen dioxide (NO2) levels obtained from 25 air monitoring sites in Seoul between 2003 and 2010. We considered an independent Gaussian process model and an auto-regressive model and carried out estimation within a hierarchical Bayesian framework with Markov chain Monte Carlo techniques. A Gaussian predictive process approximation has shown the better prediction performance rather than a Hierarchical auto-regressive model for the illustrative NO2 concentration levels at any unmonitored location.

Keywords: Auto-regressive model, bayesian inference, gaussian predictive process, nitrogen dioxide, space-time modelling.

1. Introduction

Air pollutants are well known to have a negative impact on human health and they have therefore been studied by many researchers. In Korea, five air pollutants are monitored and managed by the government, which are particulate matter10 (PM10), ozone (O3), nitrogen dioxide (NO2), sulfur dioxide (SO2), and carbon monoxide (CO). This study concerns the spatio-temporal characteristics of NO2, which affects respiratory diseases (Choi et al., 2000) and leads to the formation of photochemical smog.

The main source of NO2 in Korea is the burning of fossil fuels, such as coal, oil, and gas. In urban areas, most of the NO2 is generated by motor vehicle exhaust emissions (around 80%). Other sources of NO2 are petrol combustion and the vaporization of metal particles, coal-fired power stations for electricity generation, other manufacturing industries, and the heating of boilers in residential environments. Hence, in the winter season, NO2 levels are high because of the necessity to produce heat.

† This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education (NRF-2014R1A1A2009060).
1 Corresponding author: Assistant professor, Department of Computer Science and Statistics, Daegu University, Gyeongbuk 38453, Korea. E-mail: statstar@daegu.ac.kr
2 Researcher, WISE institute, Hankuk University of Foreign Studies, Gyeonggi 17035, Korea.
Despite the importance of monitoring and assessing the effects of NO2, the current understanding based on studies performed in Korea is inadequate. Examples of Korean studies are those of Choi et al. (2000), who studied the interrelationship of air pollution and respiratory diseases, Lee (2005, 2006, 2009), who established a statistical model for ozone concentration based on time series analysis, and Kwon and Cho (1999), who examined the relationship between air pollution and daily mortality using a generalized additive model. However, no study has previously attempted to investigate the spatio-temporal characteristics of NO2 in Korea, although many studies are underway outside of Korea. For example, Lu and Tian (2007) employed the kriging technique to interpolate data and showed the existence of a spatial pattern in the NO2 concentration, and Sitnov (2009), Prasad et al. (2012), and Jacob et al. (1986) investigated and proposed spatial and temporal models for Moscow, India, and San Francisco, respectively. However, the statistical models resulting from the above studies are insufficient in that they do not take into account both time and space. Some related research about space and time are disease mapping (Ahn et al., 2015) and apartment prices (Lee and Park, 2015).

Several authors have developed generic models for analyzing spatio-temporal data (Cressie, 1994; Goodall and Mardia, 1994). More recent models were those developed by Kyriakidis and Journel (1999), Stroud et al. (2001), Wikle and Cressie (1999), Wikle (2003), and Gelfand et al. (2005). Bakar and Sahu (2015) developed the package spTimer for the hierarchical Bayesian modeling of stylized environmental space-time monitoring data in R language. The package is able to fit and spatially and temporally predict large amounts of space-time data.

The work we present in this paper investigated the use of two hierarchical Bayesian spatial-temporal models, namely an independent Gaussian process model and a hierarchical autoregressive model. The work led to the establishment of models to predict patterns based on the daily average NO2 level between 2003 and 2010 in Seoul.

2. Methodology

A Bayesian spatial model was used to predict the three important meteorological variables at air emission monitoring sites based on observed meteorological data. The formula of a Bayesian spatial model is the same as that of a spatial linear model. Separate inferential statements about parameters require a marginal posterior distribution. In this regard, a Markov chain Monte Carlo (MCMC) technique was selected to estimate the parameters, because there is no closed form for the integrations of marginal posterior distribution. We ran a total of 3,000 iterations, and, although it was not possible to simultaneously perform formal tests of convergence of the sample chain. To provide the necessary background information, two hierarchical Bayesian spatial-temporal models and some validation methods are briefly described in this section.

2.1. Bayesian spatio-temporal models

According to Gelfand (2005), the Bayesian spatio-temporal models can be described in a hierarchical structure. Data, process and parameters in three stages are specify by:

First \[ \text{data} | \text{process, parameter} \]
Second \[ \text{process} | \text{parameter} \]
Third \[ \text{parameter} \]
The first level represents true underlying process and the second level of the hierarchy explains the spatio-temporal random effect.

The models are described for time series data that are segmented using two different units of time. Let \( l \) and \( t \) denote the two units of time where \( l \) denotes the longer unit (year, \( l = 1, \ldots, r \)) and \( t \) denotes the shorter unit (day, \( t = 1, \ldots, T_l \)) where \( r \) and \( T_l \) indicate the total number of two time units. Let \( Z_l(t_i, t) \) at site \( s_i, i, \ldots, n \) at time denoted by two indices \( l \) and \( t \). Let \( Z_l = (Z_l(s_1, t_1), \ldots, Z_l(s_n, t))' \), \( O_l = (O_l(s_1, t_1), \ldots, O_l(s_n, t))' \) and \( N = n \sum_{l=1}^{r} T_l \) which is the total number of observations. The nugget effect or the pure error term is \( \epsilon_l = (\epsilon_l(s_1, t_1), \ldots, \epsilon_l(s_n, t))' \) that follow independently normally distribution \( N(0, \sigma^2_{\epsilon_l}I_n) \), where \( \sigma^2_{\epsilon_l} \) is the unknown pure error variance and \( I_n \) is the identity matrix of order \( n \). \( \eta_l = (\eta_l(s_1, t_1), \ldots, \eta_l(s_n, t))' \) is the spatio-temporal random effects that are assumed to follow \( N(0, \sum_\eta) \) independently in time. Variance covariance matrix \( (\sum_\eta) \) is \( \sigma^2_\eta S_\eta \), where \( \sigma^2_\eta \) is the site invariant spatial variance and \( S_\eta \) is the spatial correlation matrix obtained from exponential correlation function defined as follows:

\[ \kappa(s_i, s_j, \phi) = \exp(-\phi|s_i - s_j|), \phi > 0. \]  

(2.1)

The parameter \( \phi \) controls the rate of spatial decay of the correlation according to the distance \( |s_i - s_j| \) (Cressie, 1994).

### 2.2. Independent Gaussian process model (GP model)

Let \( Z_{lt} \) denote true underlying process and \( O_{lt} \) denote the spatio-temporal random effect. The independent Gaussian process model is defined by:

\[ Z_{lt} = O_{lt} + \epsilon_{lt}, \]  

(2.2)

\[ O_{lt} = X_{lt}\beta + \eta_{lt}, \]  

(2.3)

for each \( l = 1, \ldots, r \) and \( t = 1, \ldots, T_l \), where we assume that \( \epsilon_{lt} \) and \( \eta_{lt} \) are independent and each shows a normal distribution in terms of its respective parameters. We assume that there are \( p \) covariates, including the intercept, denoted by the \( n \times p \) matrix \( X_{lt} \), \( \beta = (\beta_1, \ldots, \beta_p) \) denote the \( p \times 1 \) vector of regression coefficients. Let \( O \) denote all the random effects, \( O_{lt} \), and \( \theta = (\beta, \sigma^2_{\epsilon}, \sigma^2_\eta, \phi) \) denote all the parameters of this model. \( \pi(\theta) \) denote the prior distribution. The logarithm of the joint posterior distribution of the parameters and the missing data for this GP model is defined by:

\[
\log \pi(\theta, O, z^* | z) \propto -\frac{N}{2} \log \sigma^2_{\epsilon} - \frac{1}{2\sigma^2_{\epsilon}} \sum_{l=1}^{r} \sum_{t=1}^{T_l} (Z_{lt} - O_{lt})'(Z_{lt} - O_{lt}) - \frac{\sum_{l=1}^{r} T_l}{2} \log |\sigma^2_{\eta} S_\eta | \\
- \frac{1}{2\sigma^2_\eta} \sum_{l=1}^{r} \sum_{t=1}^{T_l} (O_{lt} - X_{lt}\beta)' S^{-1}_\eta (O_{lt} - X_{lt}\beta) + \log \pi(\theta),
\]  

(2.4)

where \( z \) denote the observed data and \( z^* \) denote the missing data. Gibbs sampling (Gelfand and Smith, 1990) is required for the full conditional distributions of parameters. Missing data values are sampled from their conditional distribution at each iteration of the Gibbs sampler.

The full conditional distributions of the parameters are provided by Bakar and Sahu (2015). That are, the full conditional distribution of \( \beta \) can be obtained by: \( \pi(\beta, ..., z) \sim \)
conditional distributions for $\sigma_i^2$ and $\sigma_\eta^2$ is as follows:

$$
\pi(1/\sigma_i^2, \ldots, z) \sim G \left( \frac{N}{2} + a, b + \frac{1}{2} \sum_{t=1}^{r} \sum_{l=1}^{T_t} (Z_{lt} - O_{lt})'(Z_{lt} - O_{lt}) \right),
$$

$$
\pi(1/\sigma_\eta^2, \ldots, z) \sim G \left( \frac{N}{2} + a, b + \frac{1}{2} \sum_{t=1}^{r} \sum_{l=1}^{T_t} (O_{lt} - X_{lt}\beta)'S_\eta^{-1}(O_{lt} - X_{lt}\beta) \right).
$$

The full conditional distribution $\phi$ is given by:

$$
\pi(\phi, \ldots, z) \propto \pi(\phi) \times |S_\eta|^{-\frac{1}{2}} \times \exp \left[ -\frac{1}{2\sigma_\eta^2} \sum_{l=1}^{r} \sum_{t=1}^{T_t} (O_{lt} - X_{lt}\beta)'S_\eta^{-1}(O_{lt} - X_{lt}\beta) \right].
$$

### 2.3. Hierarchical auto-regressive model (AR model)

The hierarchical auto-regressive model was proposed by Sahu and Mardia (2005) and is given hierarchically by:

$$
Z_{lt} = O_{lt} + \varepsilon_{lt}, \quad (2.5)
$$

$$
O_{lt} = \rho O_{l,t-1} + X_{lt}\beta + \eta_{lt}, \quad (2.6)
$$

where $\rho$ denotes the unknown temporal correlation parameter within the interval (-1,1). These model reduce to the GP model when $\rho = 0$ and it is known as autoregressive stated models. The AR model requires specification of the initial term $O_{l0}$ for each $l = 1, \ldots, r$. An independent spatial model for each $O_{l0}$ with mean $\mu_l$ and the covariance matrix $\sigma_l^2 S_0$ where the correlation matrix $S_0$ is obtained using exponential correlation function with the same set of correlation parameters $\phi$ for $\eta_l$ is used.

Let $\theta$ denote all the parameters, $\theta = (\beta, \rho, \sigma_z^2, \sigma_\eta^2, \phi, \mu, \sigma_\phi^2), l = 1, \ldots, r$ and $O$ contains all the random effect $O_{lt}$. The logarithm of the joint posterior distribution of the parameters and the missing data is given by:

$$
\log \pi(\theta, O, z^* | z) \propto -\frac{N}{2} \log \sigma_z^2 - \frac{1}{2 \sigma_\eta^2} \sum_{l=1}^{r} \sum_{t=1}^{T_t} (Z_{lt} - O_{lt})'(Z_{lt} - O_{lt}) - \frac{1}{2} \sum_{l=1}^{r} T_l \log |S_\eta|
$$

$$
- \frac{1}{2 \sigma_\eta^2} \sum_{l=1}^{r} \sum_{t=1}^{T_t} (O_{lt} - \rho O_{l,t-1} - X_{lt}\beta)'S_\eta^{-1}(O_{lt} - \rho O_{l,t-1} - X_{lt}\beta)
$$

$$
- \frac{1}{2} \sum_{l=1}^{r} \log |S_\eta| - \frac{1}{2} \sum_{l=1}^{r} \frac{1}{\sigma_\phi^2} (O_{l0} - \mu_l)'S_\phi^{-1}(O_{l0} - \mu_l) + \log \pi(\theta).
$$

The full conditional distribution of $\beta$ is obtained as $\pi(\beta | \ldots, z) \sim N(\Delta \chi, \Delta)$, where $\Delta^{-1} = \sum_{l=1}^{r} \sum_{t=1}^{T_t} X_{lt}' \sum_{l=1}^{T_t} X_{lt} + I_p/\delta_\phi^2$ and $\chi = \sum_{l=1}^{r} \sum_{t=1}^{T_t} X_{lt}' \sum_{l=1}^{T_t} (O_{lt} - \rho O_{lt})$. The full conditional distribution of $\rho$ is calculated as $\pi(\rho | \ldots, z) \sim N(\nabla \chi, \nabla)$, where $\nabla^{-1} = \sum_{l=1}^{r} \sum_{t=1}^{T_t} O_{lt}' \sum_{l=1}^{T_t} O_{lt} + I_p/\delta_\rho^2$ and $\chi = \sum_{l=1}^{r} \sum_{t=1}^{T_t} O_{lt}' \sum_{l=1}^{T_t} (O_{lt} - X_{lt}\beta)$. 

The conditional distributions for $\sigma^2_a$ and $\sigma^2_b$ is as following as follows:

$$\pi(1/\sigma^2_a|... , z) \sim G\left(\frac{N}{2} + a, b + \frac{1}{2} \sum_{l=1}^{r} \sum_{t=1}^{T_l} (Z_{lt} - O_{lt})' (Z_{lt} - O_{lt}) \right),$$

$$\pi(1/\sigma^2_b|... , z) \sim G\left(\frac{N}{2} + a, b + \frac{1}{2} \sum_{l=1}^{r} \sum_{t=1}^{T_l} (O_{lt} - \rho O_{lt-1} - X_{lt}\beta)' S^{-1}_{\eta}(O_{lt} - \rho O_{lt-1} - X_{lt}\beta) \right).$$

The full conditional distribution $\phi$ is given by:

$$\pi(\phi|... , z) \propto \pi(\phi) \times |S|^{-\tau/2} \times \exp\left[-\frac{1}{2}\sum_{l=1}^{r} \left(\sum_{t=1}^{T_l} (O_{lt} - \rho O_{lt-1} - X_{lt}\beta)' S^{-1}_{\eta}(O_{lt} - \rho O_{lt-1} - X_{lt}\beta) \right) \right].$$

The full conditional distribution for $\mu_t$ as $N(\Delta_{lt}\chi_t, \Delta_{lt})$, where $\Delta_{lt}^{-1} = \frac{1}{\sigma^2_{\beta}} S_{0}^{-1}\mathbf{1}_n + \frac{1}{\sigma^2_{\rho}} \mathbf{1}_n$ and $\chi_t = \frac{1}{\sigma^2_{\beta}} S_{0}^{-1} O_{lt}$, and for $\sigma^2_t$ following as:

$$\pi(1/\sigma^2_t|... , z) \sim G\left(\frac{N}{2} + a, b + \frac{1}{2}(O_{lt0} - \mu_t)' S^{-1}_{0}(O_{lt0} - \mu_t) \right).$$

### 2.4. Prior distribution

The spatio-temporal models based on Bayesian framework are completed by assuming suitable prior distributions for the underlying parameters. Three different types of parameters (the mean, the variance or the correlation) are grouped for simplicity and convenience. We have taken all the means, e.g., $\beta, \rho$, are given independent normal prior distribution with mean 0 and $\sigma^2_{\mu}$, assumed to be $10^{10}$. That correspond to our assumption of flat prior distributions for any parameter is missing. The prior distribution for the inverse of variance is specified through a gamma distribution with mean $a/b$ and variance $a/b^2$. We have chosen $a = 2$ and $b = 1$ because a proper prior distribution for any variance component that will satisfy a proper posterior distribution (Gelman, et al., 2004). We chosen a gamma prior distribution and uniform prior distribution over an interval for the decay parameter $\rho$. However, the trace of estimated value of the decay parameter $\phi$ was unstable. The fixed $\rho$ at a particular value is used for modelling.

### 2.5. Predicting NO2 level at a new location

Spatial prediction at location $s'$ and time $t'$ proceeds on the basis of the posterior predictive distribution for $Z_l(s', t')$. The posterior predictive distribution for $Z_l(s_0, t')$ is obtained by integrating over the parameters with respect to the joint posterior distribution as:

$$\pi(Z_l(s', t')|z) = \int \pi(Z_l(s', t')|O_l(s_0, t'), \sigma^2_z)\pi(O_l(s', t')|\theta, z^*)\pi(\theta|z)d\theta dz^*$$

Predictions are calculated by performing composition sampling. First, a random sample $\theta^{(j)}$ is drawn from the posterior distribution $\pi(\theta, z^*|z)$. Then, Bayesian kriging is applied to
draw a sample \( O_l^{(j)}(s', t') \) from the conditional distribution of \( O_l(s_1, t') \), \( O_l(s_2, t') \), \( O_l(s_n, t') \). Finally a sample \( Z_l^{(j)}(s', t') \) is drawn from the top-level model.

At the end of MCMC run, the samples \( Z_l^{(j)}(s', t') \), \( j = 1, \ldots, J \) are summarized to enable us to make a predictive inference. We use the median of the MCMC samples and the lengths of the 95% intervals to summarize the predictions. Further details regarding these predictions methods can be found in the articles by Bakar and Sahu (2015) for the GP model and Sahu and Mardia (2005) for the AR model respectively.

### 2.6. Validation criteria

The following validation criteria are calculated: root mean squared error (RMSE), mean absolute error (MAE), relative bias (rBias), and relative mean separation (rMSEP)

\[
MAE = \frac{1}{m} \sum_{i=1}^{m} |\hat{z}_i - z_i|, \quad RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^{m} (\hat{z}_i - z_i)^2}
\]

\[
rBIAS = \frac{1}{m} \sum_{i=1}^{m} (\hat{z}_i - z_i), \quad rMSEP = \frac{\sum_{i=1}^{m} (\hat{z}_i - z_i)^2}{\sum_{i=1}^{m} (\bar{z}_p - z_i)^2},
\]

where \( m \) is the total number of observations we want to validate, \( z_i \) is the data indexed by \( I \), \( \hat{z}_i \) is the prediction value, \( \bar{z} \) and \( \bar{z}_p \) are the arithmetic mean values of the observations and predictions.

### 3. Data and study area

The data we used in our analysis contained the average daily NO2 levels and were obtained from \( n = 25 \) sites in Seoul. We additionally obtained meteorological data from 25 automated weather stations (AWS) in Seoul. The locations of all 50 sites are plotted in Figure 3.1, which shows that the air emission monitoring sites and meteorological observatories are evenly distributed because they were installed according to administrative district. However, none of the sites overlap; thus, unobserved meteorological variables at air emission monitoring sites were interpolated by using a Bayesian spatial model for further analysis. Data from five of the sites (in addition to the 20 modelling sites) have been set aside for validation purposes. These sites are also indicated in Figure 3.1 and they are randomly selected.

![Figure 3.1 Air emission monitoring sites and meteorological sites in Seoul](image-url)
We consider data for \( l = 8 \) years from 2003 to 2010. In each year we have data for \( r = 365 \) days, and data collected on the additional day in leap years were omitted for convenience of analysis. The total number of datasets \( N \) is 73,000 \((n \times l \times r)\) of which 2,144 (2.93%) are missing. The boxplots of NO2 levels by year and month are shown in Figure 3.2. Overall, the average daily NO2 levels by year can be considered constant although slight variation exists. Moreover, in winter, NO2 levels are higher than in summer because of the use of heating systems.

Three important meteorological variables, i.e., maximum daily temperature in degrees Celsius, wind speed, and daily precipitation, are considered. Relative humidity was also considered for inclusion in the model, but was eventually excluded because the observations collected from weather stations failed to provide satisfactory results. Therefore, precipitation was screened instead. The results of the standard multiple regression model indicated that the three meteorological variables were significant.

Histograms and normal quantile-quantile (Q-Q) plots are provided in Figure 3.3 on three measurement scales of original, logarithmic, and square root. The data on the original scale introduces positive skewness, whereas the log scale shows negative skewness. The square root scale seems most attractive in terms of both symmetry and stabilizing the variance. This is in accordance with the results of other work involving the modeling of air pollutants (Sahu and Mardia, 2005).
4. Result and discussion

The results of the validation and parameter estimation in each Bayesian spatial-temporal model are listed in Tables 4.1 and 4.2. The parameter $\phi$ represents spatial decay was estimated by the MCMC technique. The trace of estimated value of the parameter $\phi$ was unstable. It was kept constant at 0.152 and 0.055 for the GP and the AR models, respectively, for stabilizing purposes based on the density. The GP model is reflected spatially than AR model because the higher value of the parameter $\phi$. The parameter $\rho$, represents the temporal effect, showed a significant effect in AR model. Although the parameter $\rho$ was significant, the results of the validation provided that GP model explained the spatio-temporal effects well than AR model for daily NO2 levels. The spatial variance component $\sigma^2_\eta$ is greater than the pure error component $\sigma^2_\epsilon$ in both models, thereby indicating that more variation is explained by the spatio-temporal effects than by the pure error process. We also found strong dependence among NO2 concentration levels (estimate of $\rho = 0.736$). Among the meteorological variables, the minimum temperature (tmin) and wind speed (wd) show a significant effect. However, the effect of precipitation (prec) was not significant. This should be taken to be a consequence of including the daily precipitation without considering the relative humidity, which is a point of constraint of this research. The estimation of $\mu_1, \ldots, \mu_8$ in Table 4.3 lists the initial parameters for the AR model. Figures 4.1 and 4.2 show the trace plots of the iteration and show that there is not much difference between the GP and AR models in terms of performance.

### Table 4.1 The results of validation

<table>
<thead>
<tr>
<th></th>
<th>RMSE</th>
<th>MAE</th>
<th>rBIAS</th>
<th>rMSEP</th>
</tr>
</thead>
<tbody>
<tr>
<td>GP model</td>
<td>8.096</td>
<td>6.027</td>
<td>-0.038</td>
<td>0.288</td>
</tr>
<tr>
<td>AR model</td>
<td>8.485</td>
<td>6.420</td>
<td>-0.036</td>
<td>0.317</td>
</tr>
</tbody>
</table>

### Table 4.2 The result of parameters estimation

<table>
<thead>
<tr>
<th></th>
<th>GP model</th>
<th>AR model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Intercept</td>
<td>7.486</td>
<td>0.026</td>
</tr>
<tr>
<td>tmin</td>
<td>-0.014</td>
<td>0.001</td>
</tr>
<tr>
<td>prec</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>wd</td>
<td>-0.791</td>
<td>0.013</td>
</tr>
<tr>
<td>sig2eps</td>
<td>0.027</td>
<td>0.001</td>
</tr>
<tr>
<td>sig2eta</td>
<td>0.818</td>
<td>0.005</td>
</tr>
<tr>
<td>rho</td>
<td>0.736</td>
<td>0.003</td>
</tr>
</tbody>
</table>

### Table 4.3 Estimation of $\mu_l$ and $\sigma^2_l$ for AR model, $l = 1, \ldots, 8$

<table>
<thead>
<tr>
<th>Year</th>
<th>Mean</th>
<th>SD</th>
<th>95% Credible interval</th>
<th>Mean</th>
<th>SD</th>
<th>95% Credible interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>2004</td>
<td>5.359</td>
<td>0.380</td>
<td>4.620</td>
<td>6.104</td>
<td>0.125</td>
<td>0.045</td>
</tr>
<tr>
<td>2005</td>
<td>6.635</td>
<td>0.773</td>
<td>5.085</td>
<td>8.160</td>
<td>1.771</td>
<td>0.717</td>
</tr>
<tr>
<td>2006</td>
<td>5.006</td>
<td>0.387</td>
<td>4.254</td>
<td>5.773</td>
<td>0.120</td>
<td>0.043</td>
</tr>
<tr>
<td>2007</td>
<td>6.426</td>
<td>0.387</td>
<td>5.677</td>
<td>7.193</td>
<td>0.135</td>
<td>0.050</td>
</tr>
<tr>
<td>2008</td>
<td>6.943</td>
<td>0.379</td>
<td>6.204</td>
<td>7.693</td>
<td>0.121</td>
<td>0.044</td>
</tr>
<tr>
<td>2009</td>
<td>3.807</td>
<td>0.368</td>
<td>3.077</td>
<td>4.501</td>
<td>0.113</td>
<td>0.040</td>
</tr>
<tr>
<td>2010</td>
<td>4.027</td>
<td>0.383</td>
<td>3.256</td>
<td>4.789</td>
<td>0.102</td>
<td>0.035</td>
</tr>
<tr>
<td>2011</td>
<td>5.240</td>
<td>0.404</td>
<td>4.442</td>
<td>6.054</td>
<td>0.101</td>
<td>0.033</td>
</tr>
</tbody>
</table>
Spatio-temporal models for generating a map of high resolution NO2 level

Figure 4.1 The trace plot and density plot in GP model

Figure 4.2 The trace plot and density plot in AR model
Figure 4.3 The map of annual daily average levels of NO2
In addition, we performed predictions at 300 locations on the regular grid, at an approximated resolution of 1km × 1km resolution, using the GP model. At each of these sites we spatially interpolated the daily average NO2 level on each of the 365 days in every year. These daily levels were then aggregated up to the annual levels (Figure 4.3). Although the NO2 levels were high in the south-eastern area in 2003, the overall NO2 levels were found to be high in the north-western area and low in the south-eastern area. Although it is beyond the scope of this study, it is necessary to investigate the main factors responsible for the increase in the level of NO2. We trust that the results of this research would be useful for assessing the impact of NO2 on the environment and on human beings.

The other spatio-temporal model using the grid unit, called the knots, was suggested by Sahu and Bakar (2012). The main idea is to define the random spatio-temporal effects at the knots. An AR model is only affected for the random effects at the knot locations and then kriging is used to predict the random effects at the observation locations and prediction locations.

In addition to the above-mentioned data, a composite air quality index is provided by a community unit in Seoul. However, it does not contain sufficient information to ensure an improvement in the quality of life of the citizens. The WISE project aims to provide high-resolution weather information for metropolitan citizens. In this regard, the hierarchical modeling methods described in this paper allow for the accurate modeling of localized spatial variation that may change over time based on the high-resolution output generated by the numerical weather model.

References


