Journal of the Korean Statistical Society (2003), **32**: 4, pp 313–335

SPATIAL TRENDS AND SPATIAL EXTREMES IN SOUTH KOREAN OZONE

SEOKHOON YUN[1] AND RICHARD L. SMITH[2]

ABSTRACT

Hourly ozone data are available for 73 stations in South Korea from January, 1988 to August, 1998. We are interested in detecting trends in both the mean levels and the extremes of ozone, and in determining how these trends vary over the country. The latter aspect means that we also have to understand the spatial dependence of ozone. In this connection, therefore, we examine in this paper the following features: determining trends in mean ozone levels at individual stations and combination across stations; determining trends in extreme ozone levels at individual stations and combination across stations; spatial modeling of trends in mean and extreme ozone levels.

AMS 2000 subject classifications. Primary 62H11; Secondary 62P12.
Keywords. AR processes, extreme value theory, hierarchical models, point processes, spatial statistics, threshold methods.

1. INTRODUCTION

As a measure of air pollution, (ground-level) ozone is one of the current important issues of environmental concern. Excessive levels of ozone are generally taken as an indication of high air pollution. The governmental standard is usually specified in terms of percentiles of the distribution of ozone concentrations. For instance, there are two types of the U.S. governmental standards for ozone. One is one-hour standard, which specifies that the number of exceedances by daily maxima of hourly ozone of the level of 120 ppb (parts per billion) should not exceed three in any three-year period. The other is eight-hour standard, which specifies that the average of the annual fourth highest daily maximum eight-hour averages over any three-year period should not exceed the level of 80 ppb. One

Received September 2003; accepted October 2003.
Richard L. Smith was partly funded by NSF grant DMS-0084375.
[1]Department of Applied Statistics, University of Suwon, Suwon 445-743, Korea
[2]Department of Statistics, University of North Carolina, Chapel Hill, NC 27599-3260, U.S.A.

of these two standards is usually violated in major cities such as New York, Chicago, Houston and Los Angeles. South Korean standards for ozone are more stringent, where the levels of 100 ppb and 60 ppb are used instead for one-hour and eight-hour standards, respectively.

There have been a great number of ozone studies over the last decade or so. These include Smith (1989), Feister and Balzer (1991), Cox and Chu (1993), van Ooy and Carroll (1995), Bloomfield *et al.* (1996), Joe *et al.* (1996), Carroll *et al.* (1997), Fiore *et al.* (1998), and Huang and Smith (1999). Most of these studies concern about trend estimation in mean or extreme ozone levels, and support the hypothesis of increasing trends particularly in urban areas.

The purpose of the paper is to search for overall trends in mean and extreme levels in South Korean ozone. Hourly ozone data are available for 73 monitoring stations in South Korea from January, 1988 to August, 1998. The data were supplied by the Ministry of Environment of South Korea. Figure 1.1 displays spatial location of the stations, around half of which are located in the capital area including Seoul city.

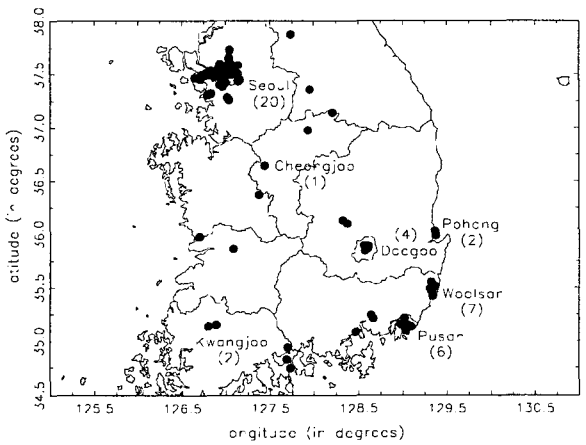

FIGURE 1.1 *Spatial location of the 73 monitoring stations in South Korea (numbers of stations in parentheses)*

First of all we estimate trends in mean and extreme ozone levels at individual stations based on time series models appropriate for the ozone data series of the 73 monitoring stations. The results of these single-station analyses appear to confirm overall positive trends, but are nevertheless hard to interpret because of the huge variability in the trend estimates from station to station. We therefore need to consider ways to combine the results across stations. We propose a

variant of the usual hierarchical model as one way of spatially smoothing the trend estimates and constructing smoothed trend surfaces, by assuming the existence of underlying smooth spatial fields.

Finally, we find that the overall trend in mean ozone levels is an increase of about 8.5 ppb over 1988–1998 and that the overall trend in extreme ozone levels is a rise of about 16.1% over the study period.

In Section 2, we give the maximum likelihood estimates of trends in mean and extreme ozone levels at individual stations. In Section 3, we introduce one way of spatially smoothing the trend estimates and then constructing smoothed trend surfaces by standard kriging. In Section 4, we estimate regional averages of the trends using the smoothed trend surfaces obtained in Section 3. Section 5 contains summary and conclusions.

2. TRENDS IN MEAN AND EXTREME OZONE LEVELS AT INDIVIDUAL STATIONS

The ozone data series typically have great hour-to-hour and season-to-season variability. To determine trends in mean ozone levels, one way is to use the average of hourly ozone concentrations measured from 9 a.m. to 5 p.m. for each day. In the present analysis, however, we focus only on daily maxima of hourly ozone (in ppb) since a lot of missing values for that time period were found in the hourly ozone data. Considering daily ozone maxima also has an effect of removing hour-to-hour variation in the hourly ozone data. Figure 2.1 displays a plot of the daily ozone maxima series at Sangmoondong station which is located in the north-east part of Seoul city (see also Figure 4.2). The plot shows a typical characteristic (high in summer and low in winter) of season-to-season variation of the ozone data series.

All the 73 daily ozone maxima series contain some portions of data missing at random. To provide a guide to the likely consequences of missing data, we computed, for each month of the study period, an overall percentage of missing values from all station $\times$ day combinations within that month, which is plotted in Figure 2.2. Many of the monitoring stations began their operation after 1990 and so the proportion of missing data is very high for each month in 1988–1989. The percentage of missing data drops from around 70% in the early months of 1988 to about 10% by 1993. It is also noted that the data are completely missing for two months (November, 1991 and May, 1998) and 84.45% of the data are missing for January, 1992.

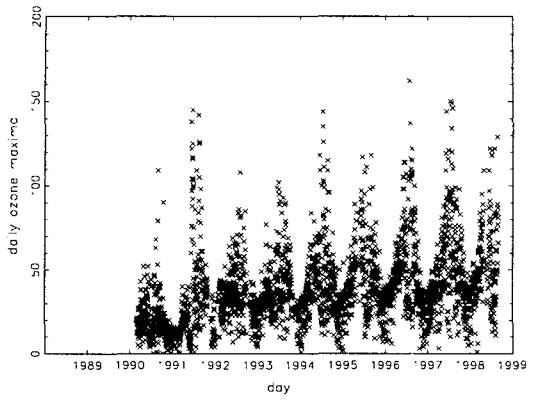

FIGURE 2.1 *Daily ozone maxima at Sangmoondong station*

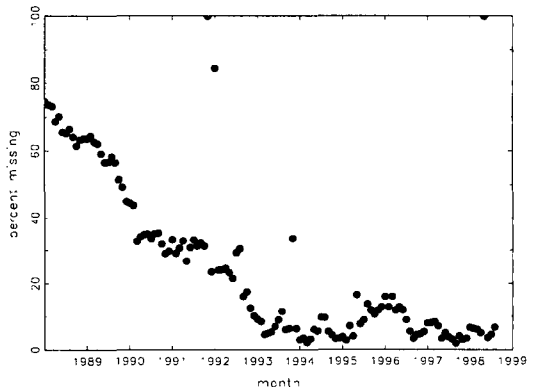

FIGURE 2.2 *Percentage of missing data in the whole network, for each month from 1988 to 1998*

2.1. Trends in mean ozone levels

To investigate a temporal trend in mean ozone levels at a single station, the mean of daily ozone maxima was computed for each month from January, 1988 to August, 1998, for each of the 73 stations. In case that the number of daily ozone maxima in a month is less than 15, the corresponding monthly mean was not computed, being treated as a missing value. We adopt a base time interval of one year, so that $h = 1/12$ is the length of one month. Let Y_t denote the monthly mean ozone for month $t = 1/12, 2/12, \ldots, (10 + 8/12)$ at a single station (*e.g.*, $t = 1/12$ corresponds to January, 1988 and so on). The model considered here is

a general linear regression with seasonal components and AR(1) errors, of form

$$Y_t = \beta_0 + \beta_1 t + \beta_2 \cos(2\pi t) + \beta_3 \sin(2\pi t) + \beta_4 \cos(4\pi t) + \beta_5 \sin(4\pi t) + \epsilon_t,$$
$$\epsilon_t = \rho \epsilon_{t-h} + \delta_t, \tag{2.1}$$

where δ_t's are *iid* $N(0, \sigma^2)$ disturbances and $|\rho| < 1$. The parameter β_1 is a linear trend (in ppb per year) and the seasonal components correspond to one-year and six-month cycles, respectively. In Figure 2.3, we display a plot of the monthly mean ozone series at Sangmoondong station.

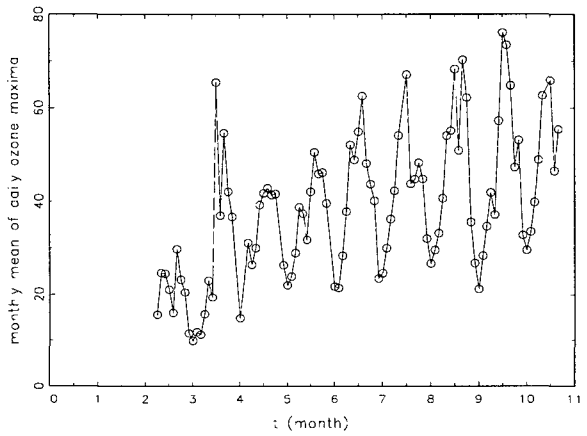

FIGURE 2.3 *Monthly mean ozone at Sangmoondong station*

Using the method of maximum likelihood, model (2.1) was fitted to each of the 73 stations. Monthly means separated by a missing value were considered to be independent. The results show that there is enormous variability in the estimates of the trend parameter β_1 over the stations: from -2.88 to $+3.66$ with a mean of 0.728 and a standard deviation of 1.134. Although the overall mean seems to be slight, when compounded over the $(10 + 2/3)$ years of the data series, it results in an overall increase of about 7.8 ppb ($(10 + 2/3) \times 0.728 = 7.765$) over the period. Our particular attention is given to Seoul city having 20 stations in which the range of the estimates of β_1 was -1.01 to $+3.66$ with a mean of 1.088 (slightly higher than the national mean) and a standard deviation of 1.082. In particular, Sangmoondong station turns out to have the highest trend in the nation, where the estimate of trend was 3.657 with standard error 0.405. The results also show that there is very strong evidence of seasonal effects and of positive correlations among the monthly mean ozone levels over time. The estimates of ρ ranged from 0.145 to 0.813 with a mean of 0.561 and a standard deviation of 0.158.

The t statistic (parameter estimate divided by its estimated standard error) for β_1 was also computed for each station. According to the results, the number of stations for which $t > 2$ is 28 out of the 73 stations, which is substantially greater than the number that would have been expected by chance ($0.025 \times 73 = 1.825$), assuming approximate normality of the parameter estimates. Moreover, the number of stations for which $t > 0$ is 56 out of 73, which is substantially larger than would seem plausible by chance alone if there were no overall trend. Thus we may conclude that there were overall positive trends in mean ozone levels though the estimates show enormous spatial variability.

2.2. Seasonal trends in mean ozone levels

Since the monthly mean ozone levels show great season-to-season variability, a separate analysis was carried out for each of the four seasons (winter, spring, summer and fall). Each season consists of three consecutive months and is considered to continue to the same season of the following year. Thus, winter season, for example, consists of December, January and February in the period 1988–1998. We adopt a base time interval of one season of three consecutive months, so that $h = 1/3$ is the length of one month. For each season, let Y_t denote the monthly mean ozone for month $t = 1/3, 2/3, 1, \ldots$ at a single station (*e.g.*, for winter season, $t = 2/3$ corresponds to January, 1988). The model fitted to each season is a simple linear regression with AR(1) errors, *i.e.*,

$$Y_t = \beta_0 + \beta_1 t + \epsilon_t, \tag{2.2}$$

where ϵ_t's are defined as in (2.1) with $h = 1/3$. Figure 2.4 displays plots of the monthly mean ozone in each season at Sangmoondong station, where the mean ozone levels are clearly increasing in all the four seasons and have the steepest slope in spring season.

In each season, model (2.2) was fitted and the maximum likelihood estimate of the trend parameter β_1 was computed for each of the 73 stations. The range of the estimates and their mean and standard deviation are given in Table 2.1. Results of the t statistics for β_1 at individual stations are also given in Table 2.1. As expected, summer season has the highest overall mean of 1.116, which implies that there was an overall increase of about 12.3 ppb ($11 \times 1.116 = 12.276$) in summer season during the 11 years of the data series. In summer season, the number of stations for which $t > 2$ and for which $t > 0$ are 27 and 60, respectively, out of 73, which are both substantially larger than would have been expected by chance.

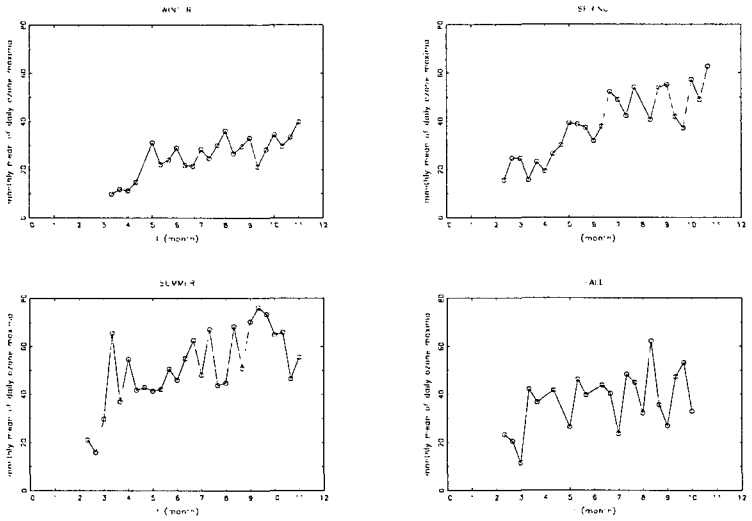

FIGURE 2.4 *Plots of monthly mean ozone in each season at Sangmoondong station*

Among the four seasons, spring season shows the greatest spatial variability in the trend estimates. The results in Table 2.1 also show that there were overall positive trends in all seasons. The overall mean of the trend estimates is higher in winter season than in fall season. The stations with the highest trend in each season were Jangheungdong station in Pohang in winter season, Songjungdong station in Cheongjoo in spring and summer seasons and Shinpyungdong station in Pusan in fall season (see Figure 1.1). The corresponding estimates of trend were 3.528, 5.568, 3.922 and 2.865 with standard errors 1.378, 1.322, 1.697 and 1.016, respectively. In summer season, the station with the second highest trend was Sangmoondong station with estimate 3.872 and standard error 0.895.

TABLE 2.1 *Summary table of estimates and t statistics for* β_1 *in each season*

Season	*Winter*	*Spring*	*Summer*	*Fall*
Range of Estimates	$-3.08 \sim +3.53$	$-3.36 \sim +5.57$	$-2.28 \sim +3.92$	$-3.97 \sim +2.87$
Mean	0.437	0.792	1.116	0.189
Standard Deviation	1.179	1.741	1.299	1.243
$t > 2$	24	21	27	7
$t > 1$	35	31	43	23
$t > 0$	52	48	60	44
$t < 0$	21	25	13	29
$t < -1$	10	10	5	9
$t < -2$	7	6	2	3

2.3. *Trends in extreme ozone levels*

For investigation of a trend in extreme ozone levels at an individual station, the threshold-based method of Smith (1989) was applied to the time series of daily ozone maxima for each of the 73 stations. Threshold methods are based on fitting stochastic models to the exceedances over a fixed high threshold u, say. For a broad discussion of the methodology, the reader is referred to Davison and Smith (1990). Extreme values in a time series typically appear in clusters due to its local dependence. In the present analysis, clusters were defined by the property that two threshold exceedances within three days of each other are considered part of the same cluster.

Following Smith (1989), the two-dimensional point process $\{(T_i, Y_i)\}$, where T_i is the time of the i^{th} cluster maximum and Y_i is the value, may be approximated by a nonhomogeneous Poisson process with intensity measure $\lambda(\cdot)$ defined by

$$\lambda((t_1, t_2] \times (y, \infty)) = \int_{t_1}^{t_2} V(y; \xi_t, \mu_t, \sigma_t)\, dt, \quad 0 \le t_1 < t_2, \ y \ge u, \tag{2.3}$$

where $V(y; \xi, \mu, \sigma) = \{1+\xi(y-\mu)/\sigma\}_+^{-1/\xi}$, $x_+ = \max\{x, 0\}$ and ξ_t, μ_t, σ_t represent respectively a shape parameter, location parameter and scale parameter for time t. Allowing these parameters to be dependent on time t creates the possibility of introducing covariate effects into the analysis. The case $\xi = 0$ is interpreted as a limit when $\xi \to 0$, so that $V(y; 0, \mu, \sigma) = \exp\{-(y-\mu)/\sigma\}$.

Under this model, if we observe the time series on a time interval $(0, T^*]$ and if we observe N cluster maxima $Y_1, \ldots, Y_N$ at time $T_1, \ldots, T_N$, then the likelihood function is given by

$$L = \exp\left(-\int_0^{T^*} V(u; \xi_t, \mu_t, \sigma_t)\, dt\right) \prod_{i=1}^{N} v(Y_i; \xi_{T_i}, \mu_{T_i}, \sigma_{T_i}), \tag{2.4}$$

where $v(y; \xi, \mu, \sigma) = -\partial V(y; \xi, \mu, \sigma)/\partial y$. In practice, the integral in (2.4) is replaced by a sum of form

$$\int_0^{T^*} V(u; \xi_t, \mu_t, \sigma_t)\, dt \approx h \sum_t V(u; \xi_t, \mu_t, \sigma_t),$$

where the sum is over days t and h is the length of one day. We adopt a base time interval of one year, so that $h = 1/365$. If there are missing data, the integral in (2.4) is replaced by an integral over the available period of data.

With the data series of daily ozone maxima for a single station, the model adopted is of form

$$\xi_t = \xi_0,\ \mu_t = \mu_0 e^{\nu_t},\ \sigma_t = \sigma_0 e^{\nu_t}, \tag{2.5}$$

where $\xi_0 \in \mathbb{R}$, $\mu_0 \in \mathbb{R}$ and $\sigma_0 > 0$ are constants and

$$\nu_t = \beta_1 t + \beta_2 \cos(2\pi t) + \beta_3 \sin(2\pi t) + \beta_4 \cos(4\pi t) + \beta_5 \sin(4\pi t). \tag{2.6}$$

For any $q \approx 1$ with $q < 1$, let $y_T(q)$ denote the q-level quantile of the distribution of the annual maximum of daily ozone maxima in a one-year time period $(T, T+1]$. Then, under model (2.3)–(2.6), it can be seen that $y_{T+1}(q) = e^{\beta_1} y_T(q)$, *i.e.*, the linear trend β_1 is interpretable as an "inflation factor" associated with the extreme quantiles of the annual maxima.

Using the method of maximum likelihood, model (2.3)–(2.6) was fitted to each of the 73 stations. In each station, the analysis was repeated, varying the threshold u from the 95^{th} percentile of the empirical distribution of daily ozone maxima to the 98^{th} percentile. For the 95^{th}-percentile threshold, successful model fits were obtained for 65 stations, while for higher-percentile thresholds (the 96^{th}, 97^{th} and 98^{th}) the number of stations with successful model fits decreased (62, 59 and 47, respectively), which might be due to smaller number of exceedances.

Summary statistics for the trend parameter β_1 are given in Table 2.2. For the 95^{th}-percentile threshold, the overall mean of 0.0108 corresponds to a rise of approximately 1.1% ($e^{0.0108} = 1.01086$) per year in the extreme quantiles, which results in an overall increase of about 12.2% ($e^{(10+2/3)\times 0.0108} = 1.1221$) during the $(10+2/3)$ years of the data series. For this threshold, the station with the highest trend was Shinseoldong station in Seoul city (see Figure 4.2), where the estimate of trend was 0.067 with standard error 0.013. For higher-percentile thresholds, the overall means are slightly higher. However, Table 2.2 shows that the analysis results are not sensitive to the selection of the treshold. The t statistics also reveal evidence of overall positive trends for all thresholds considered.

Similar summary statistics for the shape parameter ξ_0 are given in Table 2.3. Though the overall means are negative, the t statistics for higher-percentile thresholds do not seem to support that ξ_0 is negative. In an extreme value context, $\xi_0 = 0$, $\xi_0 < 0$ and $\xi_0 > 0$ correspond to exponential, short and heavy tails, respectively. For the present ozone data set, the results suggest an overall tendency toward short- or exponential-tailed distributions for the daily ozone maxima.

TABLE 2.2 *Summary table of estimates and t statistics for β_1 in extreme value model*

Threshold	95^{th}	96^{th}	97^{th}	98^{th}
Range of Estimates	$-0.065 \sim +0.067$	$-0.065 \sim +0.075$	$-0.063 \sim +0.070$	$-0.051 \sim +0.076$
Mean	0.0108	0.0117	0.0121	0.0118
Standard Deviation	0.0260	0.0276	0.0284	0.0271
$t > 2$	22	20	20	14
$t > 1$	31	31	31	25
$t > 0$	44	43	38	30
$t < 0$	21	19	21	17
$t < -1$	11	12	9	7
$t < -2$	5	5	4	2

TABLE 2.3 *Summary table of estimates and t statistics for ξ_0 in extreme value model*

Threshold	95^{th}	96^{th}	97^{th}	98^{th}
Range of Estimates	$-0.48 \sim +0.38$	$-0.41 \sim +0.37$	$-0.48 \sim +0.41$	$-0.48 \sim +0.68$
Mean	-0.087	-0.065	-0.069	-0.013
Standard Deviation	0.181	0.184	0.219	0.225
$t > 2$	4	5	6	5
$t > 1$	9	11	10	11
$t > 0$	18	21	21	18
$t < 0$	47	41	38	29
$t < -1$	31	28	29	11
$t < -2$	13	13	7	0

3. SPATIAL MODELING OF TRENDS IN MEAN AND EXTREME OZONE LEVELS

The results of Section 2 appear to confirm overall positive trends in mean and extreme ozone levels, but are nevertheless hard to interpret because of the enormous spatial variability in the estimates of the trend parameters. In this section, we explore a variant of the usual hierarchical model as one way of spatially smoothing the trend estimates obtained in Section 2 and then constructing a smoothed trend surface, by assuming the existence of an underlying smooth spatial field. The method used in this section is essentially the same as that of Holland *et al.* (2000).

Let $Z(s)$ denote an unobservable "true" trend of interest which is assumed to vary smoothly as a function of spatial location s lying in some domain $\mathcal{S}$. We assume that for each s_i belonging to a fixed subset of spatial locations

$\{s_1, \ldots, s_n\} \subset \mathcal{S}$, we observe a time series $Y(s_i, t)$, where t is time, whose distribution depends on $Z(s_i)$ as well as possibly other nuisance parameters. The trend parameter β_1 in Section 2 at a specific single station s_i corresponds to $Z(s_i)$. Suppose, for each $i = 1, \ldots, n$, we calculate an estimate of $Z(s_i)$, which we denote by $\widehat{Z}(s_i)$, based just on the time series $Y(s_i, t)$. This may be based on any model appropriate for that time series.

Since most statistical methods lead to approximately normal distributions of estimators in large samples, we may assume

$$\widehat{\mathbf{Z}} = \mathbf{Z} + \boldsymbol{\eta}_0, \tag{3.1}$$

where $\widehat{\mathbf{Z}} = (\widehat{Z}(s_1), \ldots, \widehat{Z}(s_n))^T$, $\mathbf{Z} = (Z(s_1), \ldots, Z(s_n))^T$ and $\boldsymbol{\eta}_0 \sim N(\mathbf{0}, \mathbf{W})$ represents a random error vector. In the present study, we shall assume that the covariance matrix $\mathbf{W}$ is known and diagonal with entries determined by the standard errors of the maximum likelihood analyses in Section 2. Assuming $\mathbf{W}$ to be diagonal contains an implicit assumption that the time series $Y(s_i, t)$, $i = 1, \ldots, n$, are independent.

Concerning the vector field $\mathbf{Z}$, we adopt the standard universal kriging assumption of

$$\mathbf{Z} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\eta}_1, \tag{3.2}$$

where $\mathbf{X}\boldsymbol{\beta}$ represents a spatial regression component and $\boldsymbol{\eta}_1 \sim N(\mathbf{0}, \alpha\mathbf{V}(\boldsymbol{\theta}))$ is a vector of spatially correlated random errors with covariance matrix $\alpha\mathbf{V}(\boldsymbol{\theta})$ specified in terms of a scale parameter $\alpha > 0$ and a finite-dimensional parameter $\boldsymbol{\theta}$. Since $\boldsymbol{\eta}_0$ represents a vector of measurement errors while $\boldsymbol{\eta}_1$ reflects the inherent randomness of the environment and since they represent completely different sources of variation, it is reasonable to assume that they are independent. The model defined by (3.1)–(3.2) is of hierarchical structure, with the "state equation" (3.2) representing the hypothesized true state of nature and the "measurement equation" (3.1) representing the estimate vector $\widehat{\mathbf{Z}}$ as a function of the state of nature.

Under the assumption that $\boldsymbol{\eta}_0$ and $\boldsymbol{\eta}_1$ are independent, the model now becomes

$$\widehat{\mathbf{Z}} \sim N(\mathbf{X}\boldsymbol{\beta}, \alpha\mathbf{V}(\boldsymbol{\theta}) + \mathbf{W}), \tag{3.3}$$

from which the parameters α, $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ may be estimated by the method of maximum likelihood. Once the parameters are estimated, it is then possible to reconstruct smoothed estimates of $Z(s)$, $s \in \mathcal{S}$, by standard kriging.

It remains to specify parametric models for $\mathbf{X}\boldsymbol{\beta}$ and $\alpha\mathbf{V}(\boldsymbol{\theta})$. In the present analysis, the regression component $\mathbf{X}\boldsymbol{\beta}$ was taken as a polynomial of the two-dimensional location vector s_i up to fourth order. For the covariance matrix $\alpha\mathbf{V}(\boldsymbol{\theta})$, the following three isotropic models were considered:

(i) Exponential model:

$$\mathrm{Cov}(Z(s_i), Z(s_j)) = \alpha \exp\left(-\frac{||s_i - s_j||}{\theta}\right), \ \theta > 0.$$

(ii) Gaussian model:

$$\mathrm{Cov}(Z(s_i), Z(s_j)) = \alpha \exp\left(-\frac{||s_i - s_j||^2}{\theta^2}\right), \ \theta > 0.$$

(iii) Matérn model:

$$\mathrm{Cov}(Z(s_i), Z(s_j)) = \frac{\alpha}{2^{\theta_2 - 1}\Gamma(\theta_2)} \left(\frac{2\sqrt{\theta_2}||s_i - s_j||}{\theta_1}\right)^{\theta_2} K_{\theta_2}\left(\frac{2\sqrt{\theta_2}||s_i - s_j||}{\theta_1}\right),$$

$$\boldsymbol{\theta} = (\theta_1, \theta_2), \ \theta_1 > 0, \ \theta_2 > 0.$$

Here, $||\cdot||$ denotes Euclidean distance and $K_{\theta_2}(\cdot)$ is the modified Bessel function of the third kind of order θ_2 (Handcock and Stein (1993) gave a detailed account of the Matérn covariance function). The units of distance were taken to be degrees of latitude or longitude. The Matérn model reduces to the exponential and Gaussian models if $\theta_2 = 1/2$ and $\theta_2 \to \infty$, respectively. Though the measurement error covariance matrix $\mathbf{W}$ has the interpretation of nugget effects, an additional nugget effect in $\alpha\mathbf{V}(\boldsymbol{\theta})$ was considered by substituting $\mathrm{Var}(Z(s_i)) = \alpha + \theta_0$ ($\theta_0 > 0$) for $\mathrm{Var}(Z(s_i)) = \alpha$ in the model. We also considered geometrically anisotropic models (the simplest form of anisotropic models) which can be obtained by replacing $||s_i - s_j||$ in the previous isotropic models by $||\mathbf{A}(s_i - s_j)||$, where

$$\mathbf{A} = \begin{pmatrix} D\cos\phi & D\sin\phi \\ -D^{-1}\sin\phi & D^{-1}\cos\phi \end{pmatrix}, \ D > 0, \ \phi \in [0, 2\pi).$$

By the method of maximum likelihood, model (3.3) with a variety of covariance models was fitted to the trend estimates in mean and extreme ozone levels obtained in Section 2. The fitted negative log likelihood (NLLH) values for the trend estimates in mean and extreme (95^{th}-percentile threshold) ozone

TABLE 3.1 *Comparison of model fits for the trend estimates in mean and extreme (95^{th}-percentile threshold) ozone levels, without a nugget effect*

Degree of Polynomial Trend	*Model*	*Number of Parameters*	*NLLH for Mean-Level Trends*	*NLLH for Extreme-Level (95^{th}) Trends*
0	Exponential	3	37.704	−210.547
0	Gaussian	3	38.799	−209.914
0	Matérn	4	37.702	−210.547
0	Matérn GA	6	36.591	−211.313
1	Exponential	5	37.189	−211.509
1	Gaussian	5	38.255	−211.550
1	Matérn	6	37.189	−211.618
2	Exponential	8	35.219	−213.110
2	Gaussian	8	35.584	−213.127
2	Matérn	9	35.196	−213.217
3	Exponential	12	34.934	−213.514
3	Gaussian	12	35.138	−213.497
3	Matérn	13	34.895	−213.608
4	Exponential	17	34.213	−213.920
4	Gaussian	17	34.256	−213.920
4	Matérn	18	34.175	−214.029

levels are given in Table 3.1. The models were tested for a nugget effect, but none was found. Specifically, when the parameter θ_0 was included in the model, the value of θ_0 quickly converged to 0. The models were also tested for geometric anisotropy, but none was significant. The model indicated as GA in Table 3.1 means the geometrically anisotropic model.

From likelihood ratio tests based on the fitted NLLH values in Table 3.1, there is no evidence of any spatial polynomial trends of higher order than 0 in each model. For example, in testing the quadratic trend (degree 2) against no trend (degree 0) under the Matérn model fit for the mean-level trends, the likelihood ratio test statistic is $2 \times (37.702 - 35.196) = 5.012$ with $9 - 4 = 5$ degrees of freedom, yielding p-value 0.4144 based on the asymptotic $\chi^2(5)$ distribution, the null distribution of the test statistic, which is not statistically significant. It was also found that there is no big difference among the three covariance models.

The Matérn model with a constant regression component now resulted in parameter estimates $\widehat{\alpha} = 0.847$, $\widehat{\theta}_1 = 0.144$, $\widehat{\theta}_2 = 0.469$ for the mean-level trends and $\widehat{\alpha} = 0.00046$, $\widehat{\theta}_1 = 0.089$, $\widehat{\theta}_2 = 0.492$ for the extreme-level (95^{th}-percentile threshold) trends. The smoothed estimates of trends in mean and extreme ozone

TABLE 3.2 *Summary table of estimates and t statistics for the trends in mean and extreme (95^{th}-percentile threshold) ozone levels before and after smoothing*

	Mean-Level Trends		*Extreme-Level (95^{th}) Trends*	
	Before Smoothing	*After Smoothing*	*Before Smoothing*	*After Smoothing*
Range of Estimates	$-2.88 \sim +3.66$	$-1.81 \sim +3.11$	$-0.065 \sim +0.067$	$-0.036 \sim +0.048$
Mean	0.728	0.730	0.0108	0.0113
Standard Deviation	1.134	0.762	0.0260	0.0188
$t > 2$	28	31	22	23
$t > 1$	43	47	31	34
$t > 0$	56	62	44	44
$t < 0$	17	11	21	21
$t < -1$	7	2	11	8
$t < -2$	2	2	5	4

levels at individual stations were computed by standard kriging. They were much smoother than the original estimates. For example, at Sangmoondong station the original estimate of the mean-level trend was 3.657 with standard error 0.405. After smoothing, the estimate was 3.111 with prediction standard error 0.355. At Shinseoldong station, the original estimate of the extreme-level (95^{th}) trend was 0.067 with standard error 0.013, but the estimate was 0.047 with prediction standard error 0.010 after smoothing.

Summary statistics for both the before and after smoothing estimates are given in Table 3.2, the "after smoothing" estimates indicating the estimates obtained by kriging. The decrease of standard deviations of the estimates from 1.134 to 0.762 for the mean-level trends and from 0.0260 to 0.0188 for the extreme-level trends is a result of the smoothing. According to the t statistics in Table 3.2, the smoothing results in a stronger evidence of an overall positive trend in mean ozone levels, while the smoothing effect is minor for the extreme-level trends.

In Figure 3.1, we display the resulting smoothed mean-level trend surface, represented both as a contour plot and as a perspective plot. Figure 3.2 shows the contour and perspective plots of the smoothed extreme-level (95^{th}-percentile threshold) trend surface. For the mean-level trends, the north-west area tends to have relatively higher trends than the south-east area. On the other hand, for the extreme-level trends there is very little spatial structure with just local random variation.

A separate spatial analysis was also carried out for the mean-level trend estimates in each season which were obtained in Section 2. In each season, model (3.3)

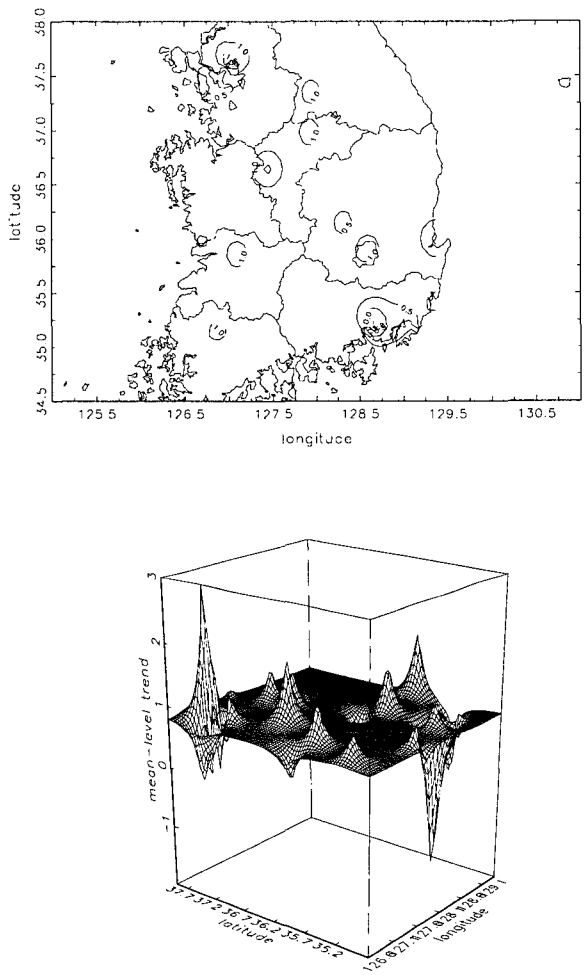

FIGURE 3.1 *Contour and perspective plots of the smoothed mean-level trend surface*

was fitted and the maximum likelihood estimates of the parameters were computed. As before, there is no evidence of a nugget effect, of geometric anisotropy and of any spatial polynomial trends. Based on the Matérn model with a constant regression component, the smoothed estimates of the mean-level trends in each season were computed by standard kriging. The perspective plots of the resulting smoothed trend surfaces are given in Figure 3.3. While summer season shows the

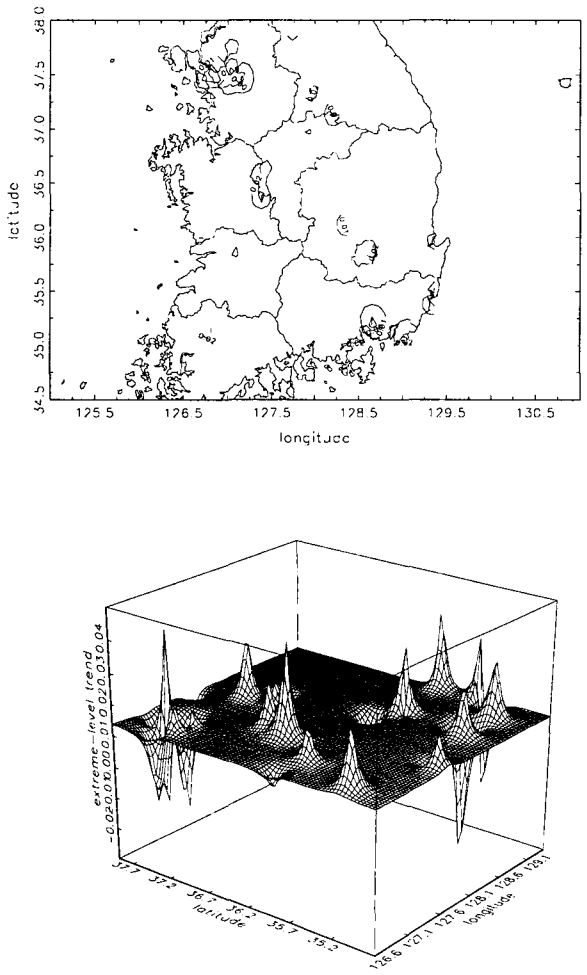

FIGURE 3.2 *Contour and perspective plots of the smoothed extreme-level (95^{th}-percentile threshold) trend surface*

highest overall mean of the smoothed trend estimates, the greatest variability in the estimates is in spring season, which is coincident with the results in Table 2.1 for the before smoothing estimates. For the extreme-level trend estimates in the higher-percentile thresholds (96^{th}, 97^{th} and 98^{th}), the results of spatial analysis were very similar to that of the 95^{th}-percentile trend analysis and are omitted here.

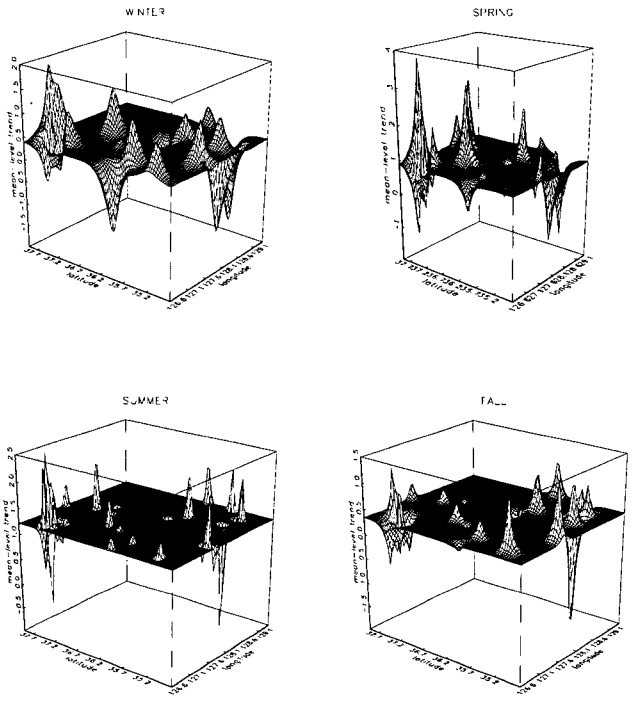

FIGURE 3.3 *Perspective plots of the smoothed mean-level trend surface in each season*

4. REGIONAL AVERAGES

An application of the smoothed trend surfaces obtained in Section 3 from the spatial analysis is to estimate regional averages of the trends. Suppose we are interested in estimating the integral

$$I(R) = \int_R Z(s)\, ds,$$

where $Z(s)$ is the trend of interest at spatial location s and $R \subset \mathcal{S}$ is some region of space. If we denote by $\widetilde{Z}(s)$ the smoothed spatial estimate of $Z(s)$ at s, then a natural estimate of $I(R)$ is

$$\widetilde{I}(R) = \int_R \widetilde{Z}(s)\, ds. \tag{4.1}$$

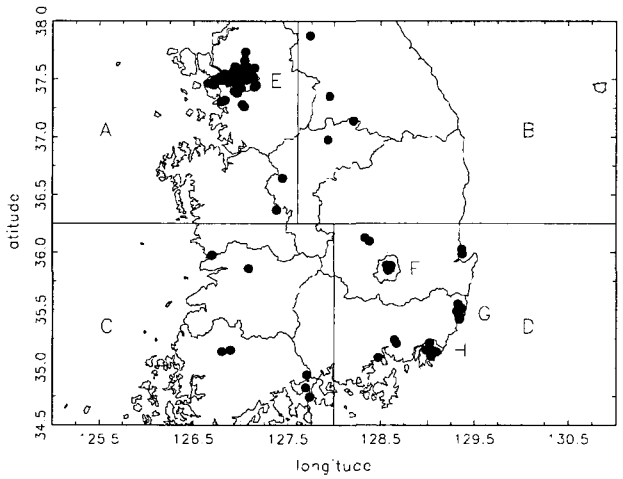

FIGURE 4.1 *Regions used for regional trend calculations. Regions A, B, C, D refer respectively to the north-west, north-east, south-west and south-east quadrants, bounded by the solid lines. Regions E, F, G, H are respectively the capital area, Daegoo, Woolsan and Pusan, bounded by the dotted lines.*

Moreover, from kriging theory, we have

$$\mathrm{Var}\left\{\widetilde{I}(R) - I(R)\right\} = \int_R \int_R \mathrm{Cov}\left(\widetilde{Z}(s_1) - Z(s_1), \widetilde{Z}(s_2) - Z(s_2)\right) ds_1 ds_2. \quad (4.2)$$

There is also a standard formula from kriging theory for the covariance in (4.2) (*cf.* Cressie, 1993, pp. 154–155), which may be substituted into (4.2) to obtain an approximate standard error for $\widetilde{I}(R)$. In practice, the integrals in (4.1) and (4.2) are replaced by sums over a suitable dense grid.

To apply the method, South Korea was divided into several subregions (see Figure 4.1), with A, B, C and D representing respectively the north-west, north-east, south-west and south-east quarters of the country. We also considered smaller regions E, F, G and H which represent respectively the capital area, Daegoo, Woolsan and Pusan.

Applying the formulae (4.1)-(4.2) to the smoothed trend estimates in mean and extreme (95^{th}-percentile threshold) ozone levels from Section 3, we computed estimates of the average trends and standard errors for the regions A–H and for the country overall, which are given in Table 4.1. All the regions considered clearly have strong positive trends in both the mean and extreme levels. The

TABLE 4.1 *Estimates and standard errors of regional averages of the trends in mean and extreme (95^{th}-percentile threshold) ozone levels for eight regions of South Korea, and overall*

	Mean-Level Trends		*Extreme-Level (95^{th}) Trends*	
Region	*Average*	*S.E.*	*Average*	*S.E.*
A	0.831	0.011	0.0135	0.0003
B	0.810	0.012	0.0140	0.0003
C	0.818	0.011	0.0145	0.0003
D	0.714	0.011	0.0139	0.0003
E	0.772	0.010	0.0099	0.0003
F	1.058	0.011	0.0175	0.0003
G	0.551	0.010	0.0122	0.0003
H	0.408	0.011	0.0164	0.0003
All	0.792	0.011	0.0140	0.0003

region with the highest trends in both levels is region F, which is not surprising since region F has a particular geograpical structure of a basin surrounded by several mountains.

For the mean-level trends, regions G and H located near the coast appear to have relatively lower trends though region G is an industrial area and region H has the second highest trend in extreme level. For the extreme-level trends, region E has the lowest trend though Seoul city is included in that region. This is probably because there is still great variability of the smoothed trend estimates within region E (*cf.* Figure 3.2). Concerning the regions A, B, C and D, the regionally averaged mean-level trend estimates are fairly stable except for region D, with an overall average of 0.792, while the regionally averaged extreme-level trend estimates are very consistent, being about 0.014 for all the regions.

Region E, the capital area, contains 37 monitoring stations which are more than half of the total stations considered in South Korea. Also, around one third of the whole population of over 47 millions belongs to region E. Since it is suspected that Seoul city is highly contaminated, region E was redivided into four subregions (see Figure 4.2), with E1, E2, E3 and E4 representing respectively the north-west, north-east, south-west and south-east quarters.

Figures 4.3 and 4.4 show the smoothed trend surfaces in mean and extreme (95^{th}-percentile threshold) ozone levels respectively over region E, recomputed with a more dense grid than in Figures 3.1 and 3.2. The peaks of the smoothed surfaces are at Sangmoondong station and at Shinseoldong station for the mean-level and extreme-level trends, respectively.

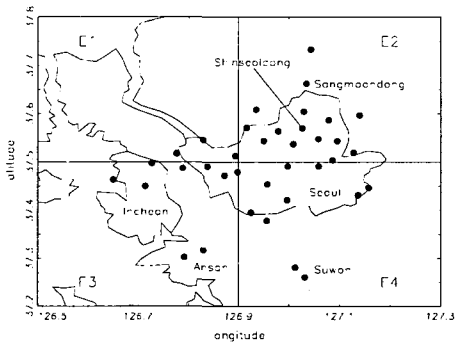

FIGURE 4.2 *Spatial location of the 37 stations in region E, the capital area, and redivision of E into four subregions E1, E2, E3, E4 representing respectively the north-west, north-east, south-west and south-east quadrants*

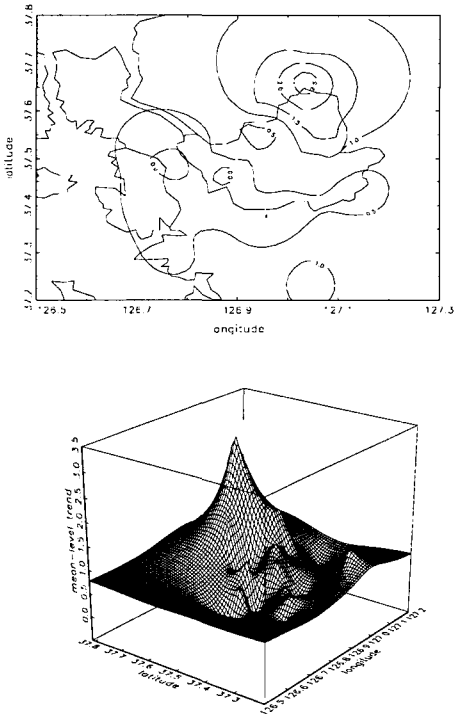

FIGURE 4.3 *Contour and perspective plots of the smoothed mean-level trend surface over region E, the capital area*

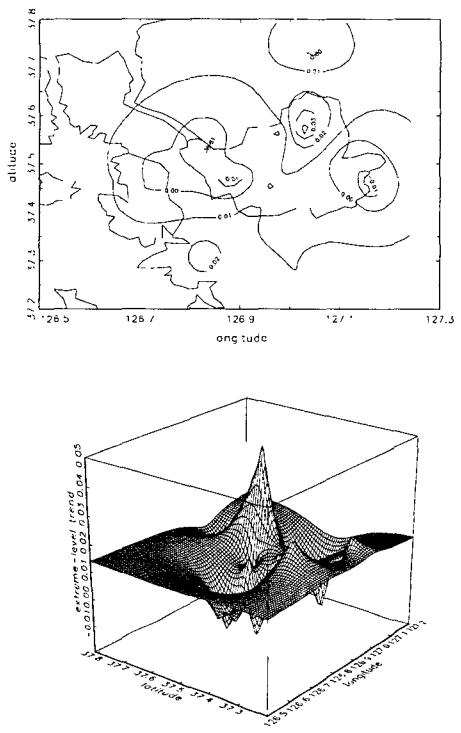

FIGURE 4.4 *Contour and perspective plots of the smoothed extreme-level (95^{th}-percentile threshold) trend surface over region E, the capital area*

Using the formulae (4.1)–(4.2), we again computed estimates of the average mean- and extreme-level trends and standard errors for the subregions E1–E4 of region E. The results are given in Table 4.2. For the mean-level trends, region E2 has the highest trend which is even higher than in region F in Table 4.1, whereas region E3 connected to the coast appears to have the lowest trend. For the extreme-level trends, the regions with the highest and lowest trends are regions E3 and E4, respectively. The results are understandable if we note that the north-east part of Seoul city contained in region E2 has many old streets and is surrounded by high mountains to the north, while the south-east part of Seoul city contained in region E4 has been newly constructed and is composed of relatively plain areas.

TABLE 4.2 *Estimates and standard errors of regional averages of the trends in mean and extreme (95^{th}-percentile threshold) ozone levels for four subregions of region E, the capital area*

	Mean-Level Trends		*Extreme-Level (95^{th}) Trends*	
Region	*Average*	*S.E.*	*Average*	*S.E.*
E1	0.685	0.011	0.0091	0.0003
E2	1.262	0.010	0.0104	0.0002
E3	0.487	0.010	0.0111	0.0003
E4	0.660	0.010	0.0085	0.0002

5. CONCLUSIONS

Based on monthly means of daily ozone maxima and all exceedances over the 95^{th}-percentile threshold at each station, the final results of our spatial analysis for South Korean ozone are as follows.

For mean ozone levels, the analysis results show an overall positive trend of about 0.792 ppb per year, which results in an overall increase of about 8.5 ppb ($(10 + 2/3) \times 0.792 = 8.448$) over the study period. For extreme ozone levels, the overall average of trends was 0.014, which corresponds to a rise of about 1.4% ($e^{0.014} = 1.0141$) per year. This results in an overall increase of about 16.1% ($e^{(10+2/3)\times 0.014} = 1.16106$) in extreme ozone levels over the study period.

For the mean-level trends of ozone, the north-west region of the country including the capital area appears to have higher trends than the south-east region. In contrast, for the extreme-level trends, there seems to be no big difference of trends among different regions. These are reflected in the smoothed trend surfaces as well as in the regionally averaged trends.

Within the capital area, the north-east part shows an exceptionally high regional trend in mean ozone levels (about 159% of the nationally averaged trend), while for the trends in extreme ozone levels the south-west part appears to have the highest regional trend.

The results of our spatial analysis also indicate that the ozone levels in South Korea are largely affected by geographical structures of the cities considered. For example, the mean-level trends tend to be very high in Daegoo and the north-east part of Seoul, both areas being surrounded by several mountains. On the other hand, Pusan, Woolsan and Incheon located near the coast show low trends in mean ozone levels.

REFERENCES

BLOOMFIELD, P. J., ROYLE, J. A., STEINBER, L. J. AND YANG, Q. (1996). "Accounting for meteorological effects in measuring urban ozone levels and trends", *Atmospheric Environment*, **30**, 3067–3077.

CARROLL, R. J., CHEN, R., GEORGE, E. I., LI, T. H., NEWTON, H. J., SCHMIEDICHE, H. AND WANG, N. (1997). "Ozone exposure and population density in Harris County, Texas", *Journal of the American Statistical Association*, **92**, 392–404.

COX, W. AND CHU, S. (1993). "Meteorologically adjusted ozone trends in urban areas : A probabilistic approach", *Atmospheric Environment*, **27**, 425–434.

CRESSIE, N. A. C. (1993). *Statistics for Spatial Data*, 2nd ed., John Wiley, New York.

DAVISON, A. C. AND SMITH, R. L. (1990). "Models for exceedances over high thresholds (with discussion)", *Journal of the Royal Statistical Society*, **B52**, 393–442.

FEISTER, U. AND BALZER, K. (1991). "Surface ozone and meteorological predictors on a subregional scale", *Atmospheric Environment*, **25**, 1781–1790.

FIORE, A. M., JACOB, D. J., LOGAN, J. A. AND YIN, J. H. (1998). "Long-term trends in ground level ozone over the contiguous United States, 1980–1995", *Journal of Geophysical Research*, **103**, 1471–1480.

HANDCOCK, M. S. AND STEIN, M. L. (1993). "A Bayesian analysis of kriging", *Technometrics*, **35**, 403–410.

HOLLAND, D. M., DE OLIVEIRA, V., COX, L. H. AND SMITH, R. L. (2000). "Estimation of regional trends in sulfur dioxide over the eastern United States", *Environmetrics*, **11**, 373–393.

HUANG, L. S. AND SMITH, R. L. (1999). "Meteorologically-dependent trends in urban ozone", *Environmetrics*, **10**, 103–118.

JOE, H., STEYN, D. G. AND SUSKO, E. (1996). "Analysis of trends in tropospheric ozone in the lower Fraser Valley, British Columbia", *Atmospheric Environment*, **30**, 3413–3421.

SMITH, R. L. (1989). "Extreme value analysis of environmental time series : An application to trend detection in ground-level ozone (with discussion)", *Statistical Science*, **4**, 367–393.

VAN OOY, D. J. AND CARROLL, J. J. (1995). "The spatial variation of ozone climatology on the western slope of the Sierra Nevada", *Atmospheric Environment*, **29**, 1319–1330.