Supplementary materials for this article are available at 10.1007/s13253-011-0057-7.

An Efficient Approach to Spatiotemporal Analysis and Modeling of Air Pollution Data

Georgios TSIOTAS and Athanassios A. ARGIRIOU

A statistically efficient approach is adopted for modeling spatial time-series of large data sets. The estimation of the main diagnostic tool such as the likelihood function in Gaussian spatiotemporal models is a cumbersome task when using extended spatial time-series such as air pollution. Here, using the Innovation Algorithm, we manage to compute it for many spatiotemporal specifications. These specifications refer to the spatial periodic-trend, the spatial autoregressive moving average, the spatial autoregressive integrated and fractionally integrated moving average Gaussian models. Our method is applied to daily pollutants over a large metropolitan area like Athens. In the applied part of our paper, we first diagnose temporal and spatial structures of data using non-likelihood based criteria, such as the empirical autocorrelation and covariance functions. Second, we use likelihood and non-likelihood based criteria to select a spatiotemporal model among various specifications. Finally, using kriging we regionalize the resulting parameter estimates of the best-fitted model in space at any unmonitored location in the Athens region. The results show that a specific autoregressive integrated moving average spatiotemporal model can optimally perform in within and out of spatial sample estimation. Supplemental materials for this article are available from the journal website.

Key Words: Gaussian maximum likelihood estimator; Innovation algorithm; Kriging; Spatiotemporal modeling; Urban pollution.

1. INTRODUCTION

Air pollution data are characterized by systematic dependencies in both temporal and spatial dimensions (Thompson et al. 2001). These dependencies are usually modeled using large time-series data of different locations. By using large data sets, the problem of efficiency is very common when it comes to statistical estimation and model selection (Brockwell and Davis 1991). In this paper, we introduce a computational efficient approach for

Georgios Tsiotas (✉) is Lecturer, Department of Economics, University of Crete, Panepistimioupolis, Rethymnon 74100, Greece (E-mail: *tsiotas@ermis.soc.uoc.gr*). Athanassios A. Argiriou is Associate Professor, Department of Physics, Section of Applied Physics, University of Patras, 265 00 Patras, Greece (E-mail: *argiriou@physics.upatras.gr*).

Journal of Agricultural, Biological, and Environmental Statistics, Volume 16, Number 3, Pages 371–388
DOI: 10.1007/s13253-011-0057-7

statistical inference in spatiotemporal models applied to air pollution data of large dimension. This approach is based on the use of the Innovation Algorithm which decreases the time needed for inference by avoiding the inversion of large data matrices (Brockwell and Davis 1991; Yao and Brockwell 2006). The performance of this method is investigated in some benchmark Gaussian spatiotemporal models such as the spatial periodic-trend, the spatial autoregressive moving average, the spatial autoregressive integrated and fractionally integrated moving average models. Initially, we decompose via estimation the various temporal and spatial components of the spatiotemporal models and then we perform model selection using an estimator based on the likelihood function. This inference strategy is then applied to daily ozone (O_3) concentration data collected over a large metropolitan area like Athens.

Thus, the objective of this paper is both methodological and empirical. There are two methodological objectives pursued here: one involves the demonstration of statistical tools for investigating the underlying dynamics and spatial structures of atmospheric spatiotemporal data; the other refers to the creation of a theoretical specification for these data using a computationally efficient algorithm. With regard to our empirical objectives, we introduce apart from within-sample estimation, the spatial forecasting of air pollution for locations that we do not observe. This can potentially reveal extreme air pollution measures giving rise to public health warning issues. Additionally, spatial forecasting estimation can detect areas with measures of high error, which can demonstrate the need for new allocation of monitoring station in order to improve the accuracy of air pollution estimation.

Spatial analysis has long been an issue in statistical literature. Since the pioneering work of Whittle (1954) concerning stationary spatial processes, time-domain methods based on the seminar work of Besag (1974) put forward an ingenious auto-normal specification based on a conditional specification argument. Besag's proposal effectively specifies the inverse of covariance matrix in a Gaussian likelihood function, where parameters bear direct interpretation of the dependence in terms of conditional expectations. However, to avoid the calculation of the determinant of the covariance matrix, approximation methods such as the coding and the pseudo-likelihood were used in place of the true likelihood to derive the estimators (Besag 1975). Attempts to calculate Gaussian likelihood functions directly have also been made, including, among others, by Ord (1975), Mardia and Marshall (1984). See §7.2 in Cressie (1993) for a survey on the estimation for spatial models.

In spatiotemporal analysis of modeling air pollution data there is important but not extensive literature (Thompson et al. 2001). Caroll et al. (1997) use a temporally stationary space–time model that is homogeneous in space. Their purpose is to model ozone concentrations for spatiotemporal prediction. However, in the temporal side of their analysis they do not use purely time-series processes but also their spatial structure is homogeneous. As Cressie (1997), Stein and Fang (1997) criticize the model for lack of spatial trend component. A more thorough approach is adopted later by Kyriakidis and Journel (2001). They present a time-series framework for modeling air pollution data where parametric temporal trend models are established from concentration profiles at monitoring stations. Such parameters, e.g. amplitude of seasonal variation, are then regionalized in space for determining trend models at any unmonitored location. Their methodology is then applied to the sulfur concentration data measured at the European Monitoring and Evaluation Program.

Shaddick and Wakefield (2002) propose a hierarchical dynamic linear model for air pollutants in the London area. Their approach uses a generalized assumption regarding spatial isotropy and temporal stationarity. Romanowicz et al. (2006) model daily nitric oxide concentrations around Paris using a spatiotemporal model consisting of trend and seasonal components assuming anisotropic spatial dependency. However, all these approaches suffer from weaknesses at the model selection stage of inference which in our paper is demonstrated using a computationally efficient time-series algorithm; the Innovation Algorithm. This algorithm, widely used in time-series, is a recursive method for computing best linear predictors. It allows computing the best estimators without having to perform any matrix inversion, which is useful for large data series, as the ones applied here, where direct estimation requires the solution of a large system of linear equations (Brockwell and Davis 1991; Yao and Brockwell 2006).

Considering the empirical part of our paper, this is dedicated to the application of spatiotemporal modeling of O_3 concentration over Athens. Large-scale industrialization, population inflow, complex topographical and meteorological conditions often favor the accumulation of various pollutants in the Athens region. This can induce significant degradation of urban air quality. It is of great interest to evaluate the status of air quality for several pollutants such as O_3, CO, SO_2, NO_2 etc. given their impact on the health of the urban population. With regard to the O_3 measure, this is a very important atmospheric pollutant that depends on the amount of hydrocarbons in the atmosphere, since photochemical processes are responsible for its formation. Additionally, conditions favorable to O_3 production are sunshine, high temperature, and stagnant air (NRC 1991). In the present analysis, all the above meteorological variables are assumed to be incorporated in both the spatial and temporal structure of data. Thus, the choice of the relevant variables depend on the objective of our analysis which is associated with a purely spatiotemporal model.

In this paper, we model temporal dynamics and spatial dependencies simultaneously using O_3 concentration data series in the greater Athens region. Our basic aim is to propose a spatiotemporal model that can best fit our Athens data in space and time. The paper is organized as follows: In Section 2 we use data from the Athens Metropolitan Area (AMA) to investigate the existing spatial and temporal structure. We do so based on spatial and temporal diagnostic tools such as the covariance, the correlation, and the autocorrelation functions. This can help us recommend some models that can possibly characterize our data. In Section 3, we introduce some spatiotemporal models together with tools for statistical inference in a Gaussian framework. To implement this, we introduce three alternative general class of models: the spatial periodic-trend, the spatial autoregressive moving average model, and the non-stationary spatial times-series model. We focus on parametric estimation and model selection by calculating the likelihood function. This is implemented by adopting the computationally efficient 'Innovation Algorithm based on pre-whitening (Brockwell and Davis 1991; Yao and Brockwell 2006). Section 4 is devoted to regional prediction using kriging. Here, the resulting parameter estimates of the best-fitted model are regionalized in space at any unmonitored location. Finally, Section 5 presents a discussion.

2. ATHENS OZONE DATA

The need for air quality control has led to daily monitoring of these pollutants in several locations. For determining the spatiotemporal "distribution" of these pollutants, monitoring stations are usually placed at the center and periphery of urban areas as well as in surrounding rural areas. With regard to the time dimension, long daily data series can be used to determine a robust temporal analysis. Several approaches have been proposed to forecast the pollutant levels in the Athens area (e.g. Ziomas et al. 1995).

The data sets we use in this paper represent measured daily maximum values of O_3 concentrations from eight different locations in the Athens region, covering the period from 01.01.1994 to 31.12.2000. The locations we use are the pollution monitoring stations of *Athinas*, *Geoponiki*, *Liosion*, *Lykovrisi*, *Marousi*, *Patisia*, *Piraeus* and *Nea Smirni*. These values are then logarithmically transformed (by taking the logarithms of the ozone concentrations) in order to achieve a symmetric Gaussian-like frequency behavior.

2.1. ANALYSIS OF TEMPORAL STRUCTURE

In this section we investigate the temporal structure of our data sets. We do this by presenting the data time-series of each station, independent of spatial information. For this purpose we use the autocorrelation function (ACF) since it is a good measure of temporal correlation (Handcock and Wallis 1994). Figure 1 presents the empirical autocorrelation functions from the *Athinas* station. This is derived from daily maximum value observations of the period 1994–2000. What is evident is the high level of autocorrelation plus the cyclical behavior as the correlation dies off at all stations. This high level of autocorrelation together with its slow damping behavior gives rise to a non-stationary analysis as proposed by Box and Jenkins (1970) (hereafter called the BJ). Non-stationary analysis can either adopt the first-difference transformation or set the differencing parameter as random assuming long memory (Brockwell and Davis 1991). In the first case, by introducing the first difference, one can identify the existing time-dependent structure of our data. We also present the empirical ACF of the first-difference data series from the *Athinas* station. Figure 2 shows that the cyclical pattern has been diminished whereas the autoregressive structure does not take statistically significant values after the third lag. Standard identification procedures, introduced by BJ, diagnose patterns of an autoregressive process of order less than four and a possible moving average dependence (see Figures 1 and 2 in the web Appendix). Also, as the ACF of the first-difference series dies off quickly, further differencing is not necessary. Similarly to the above patterns, both original and first-difference data series are observed in the rest of the stations (see Figures 3 and 2 in the web Appendix). Concerning the long-memory effect, Section 3 provides an insight on its spatial specification.

2.2. ANALYSIS OF SPATIAL STRUCTURE

Apart from temporal variability, there is the spatial aspect that needs to be investigated. First, we attempt to derive some measures of data spatial dependency for the whole time

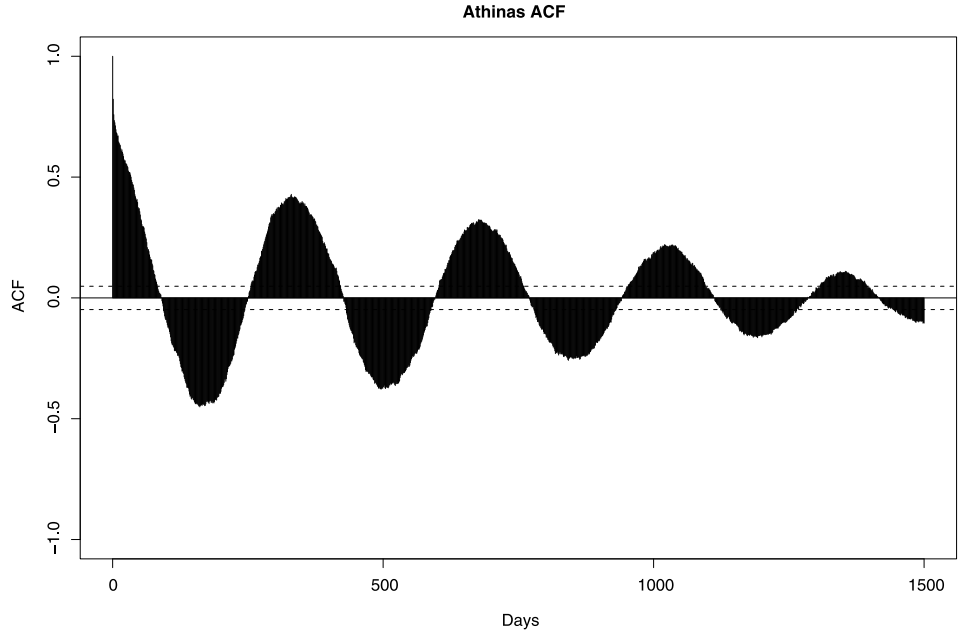

Figure 1. Empirical autocorrelation functions from Athinas station using O_3 daily data from the period 1994–2000.

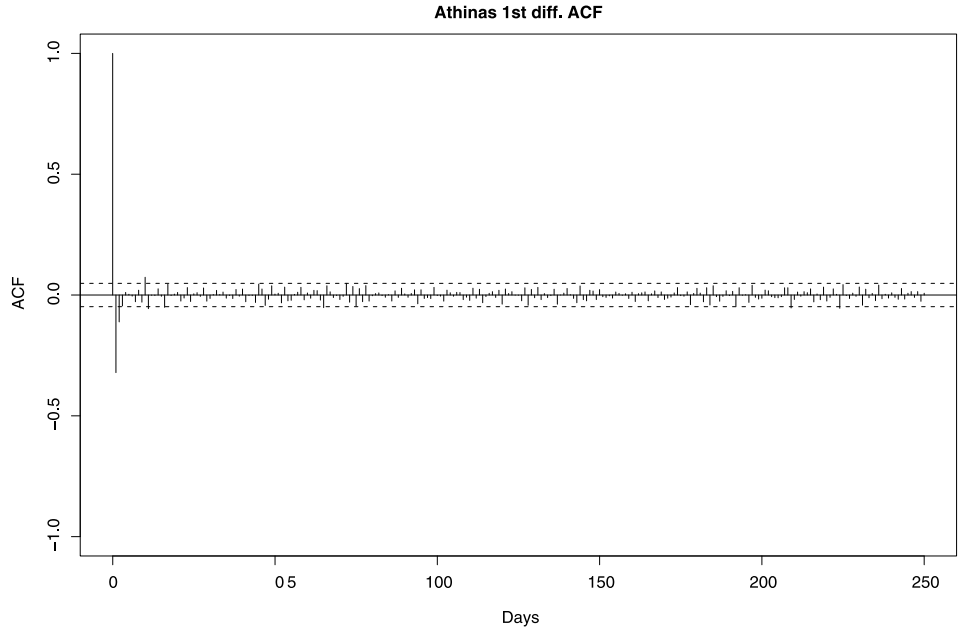

Figure 2. Empirical autocorrelation functions for the first difference series from Athinas station using O_3 daily data from the period 1994–2000.

domain of any two spatial locations. One of these empirical measures of spatial dependencies is the covariance and correlation functions. Second, we investigate whether these empirical spatial dependency functions are time-independent.

Suppose spatial processes $\{Y(\mathbf{s})\}$ and $\{Y(\mathbf{r})\}$ are indexed by a two-dimensional location $\mathbf{s}, \mathbf{r} \in \mathscr{S} \subset \Re^2$. Then their empirical covariance function can be expressed as (Cressie 1993):

$$\operatorname{Cov}\big(Y(\mathbf{s}), Y(\mathbf{r})\big) = \frac{1}{T}\sum_{t=1}^{T}\big(Y_t(\mathbf{s}) - \bar{Y}(\mathbf{s})\big)\big(Y_t(\mathbf{r}) - \bar{Y}(\mathbf{r})\big), \qquad \mathbf{s}, \mathbf{r} \in \mathscr{S},$$

where $Y_t(\mathbf{s})$ and $Y_t(\mathbf{r})$ are observations at time t located in $\mathbf{s}$ and $\mathbf{r}$ respectively, and $\bar{Y}(\mathbf{s}) = 1/T\sum_{t=1}^{T} Y_t(\mathbf{s})$, $\bar{Y}(\mathbf{r}) = 1/T\sum_{t=1}^{T} Y_t(\mathbf{r})$ represent the mean location values. This measure of spatial dependency is usually expressed by various theoretical covariance functions such as the Exponential, the Gaussian, the Wave or the Matérn class (Cressie 1993). These functions will be used in Sections 3 and 4 where various classes of models are assessed for data series of maximum O_3 values. An additional measure of spatial relation is the standardized covariance function known as the spatial correlation function. This takes the form

$$\operatorname{Corr}\big(Y(\mathbf{s}), Y(\mathbf{r})\big) = \operatorname{Cov}\big(Y(\mathbf{s}), Y(\mathbf{r})\big)\big/\big[\operatorname{var}(Y(\mathbf{s}))\big]^{1/2}\big[\operatorname{var}(Y(\mathbf{r}))\big]^{1/2}, \qquad \mathbf{s}, \mathbf{r} \in \mathscr{S}.$$

The above spatial functions assume time-homogeneity. In other words, one supposes that the empirical covariance and correlation functions are the same across time. To test this assumption we calculate the time-dependent empirical spatial covariance and correlation functions. Here, the time-dependent empirical spatial covariance function takes the form

$$\operatorname{Cov}_k\big(Y(\mathbf{s}), Y(\mathbf{r})\big) = \frac{1}{T}\sum_{t=1}^{T}\big(Y_t(\mathbf{s}) - \bar{Y}(\mathbf{s})\big)\big(Y_{t-k}(\mathbf{r}) - \bar{Y}(\mathbf{r})\big), \qquad \mathbf{s}, \mathbf{r} \in \mathscr{S},$$

with $k = 1, 2, 3, \ldots$. It shows the spatial covariance and correlation functions respectively using the time-independence and the time-dependence assumption. In the first case we suppose a time lag $k = 0$, whereas in the second case we assume lags from $k = 1$ to 5. It is well documented for both the time-dependent and the time-independent functions that there is a systematic relationship between locations in the whole time-series. Additionally, the spatial correlation function shows more structure than the covariance one. This is due to the high level of variance of time-series at each location. Finally, the analysis of both covariance and correlation functions shows time-homogeneity since the numerical results show only marginal deviation across time-lagged series (see Figures 5 and 6 in the Web Appendix). This time-homogeneity result influences the way we set the spatial and time dimensions in the spatiotemporal models that follow.

3. SPATIOTEMPORAL MODELS

In this section, we propose three classes of models that can be viewed as candidates for fitting our maximum O_3 data. The first class of models is the periodic-trend which is

selected due to the periodic nature of our data series (see Section 2). The second class of models is the pth and qth order Autoregressive Moving Average ARMA(p, q), which has been selected because of the strong lagged correlation and the sometimes irregular seasonal pattern of our observed data. The third class of models is the non-stationary Autoregressive Moving Average models that assume either first or fractional differencing. In the following, we present some statistical aspects of these three candidate classes of models and give a brief description of the way we implement our inference. Finally, we present results from alternative model specifications.

3.1. PERIODIC MODELS

Here we propose a class of spatiotemporal models to fit with spatial time-series data located at irregular space. More specifically, suppose that $\{Y_t(\mathbf{s}), t = 0, \pm1, \pm2, \ldots\}$ represent a periodic-trend model at location $\mathbf{s} \in \mathscr{S} \subset R^2$ (Kyriakidis and Journel 2001):

$$Y_t(\mathbf{s}) = a_{0,\mathbf{s}} + a_{1,\mathbf{s}} \cos\left(\frac{2\pi}{PER} t\right) + a_{2,\mathbf{s}} \sin\left(\frac{2\pi}{PER} t\right) + \sigma_{\mathbf{s}} \epsilon_t(\mathbf{s}), \tag{3.1}$$

where PER represents the period. Here for each fixed $\mathbf{s} \in \mathscr{S}$, $\{\varepsilon_t(\mathbf{s}), t = 0, \pm1, \pm2, \ldots\}$ is an i.i.d. sequence with mean 0 and variance 1.

Here the $\{a_{i,\mathbf{s}}\}_{i=0,1,2}$ coefficients represent the intercept and the seasonal coefficients. To interpret coefficients, we need to clarify that the $a_{1,\mathbf{s}}$ and $a_{2,\mathbf{s}}$ are linked to the amplitude $\alpha(\mathbf{s})$ and the phase $\phi(\mathbf{s})$ of the periodic component in our temporal specification as:

$$\alpha(\mathbf{s}) = \sqrt{a_{1,\mathbf{s}}^2 + a_{2,\mathbf{s}}^2}, \tag{3.2}$$

$$\phi(\mathbf{s}) = \tan^{-1}\left(\frac{a_{2,\mathbf{s}}}{a_{1,\mathbf{s}}}\right). \tag{3.3}$$

Here the amplitude $\alpha(\mathbf{s})$ is connected with the period component PER at location $\mathbf{s}$ and the phase $\phi(\mathbf{s})$ with the period when the temporal concentration is minimum or maximum.

3.2. ARMA MODELS

The second class of models we propose is the spatial pth and qth order Autoregressive Moving Average ARMA(p, q) models (Besag 1974, 1975; Brockwell and Davis 1991; Yao and Brockwell 2006). Our basic aim is to model both the temporal dynamics and spatial correlations.

Let $\{Y_t(\mathbf{s}), t = 0, \pm1, \pm2, \ldots\}$ be an ARMA process at location $\mathbf{s} \in \mathscr{S} \subset R^2$:

$$Y_t(\mathbf{s}) = b_{1,\mathbf{s}} Y_{t-1}(\mathbf{s}) + \cdots + b_{p,\mathbf{s}} Y_{t-p}(\mathbf{s}) + \varepsilon_t(\mathbf{s}) - c_{1,\mathbf{s}} \varepsilon_{t-1}(\mathbf{s}) - \cdots - c_{q,\mathbf{s}} \varepsilon_{t-q}(\mathbf{s}), \tag{3.4}$$

where $\varepsilon(\mathbf{s}) \sim N(0, \sigma_{\mathbf{s}}^2)$ is a Gaussian spatial process with mean zero and variance $\sigma_{\mathbf{s}}^2$. For the spatiotemporal process $\{Y_t(\mathbf{s})\}$ to be fully specified, we impose the following conditions:

(C1) For each $\mathbf{s} \in \mathscr{S}$, $\sigma_{\mathbf{s}} > 0$, and $b(L) = 1 - \sum_{j=1}^{p} b_{j,\mathbf{s}} z^j \neq 0$ for any $|z| \leq 1$.

(C2) For each $\mathbf{s} \in \mathscr{S}$, $\sigma_{\mathbf{s}} > 0$, and $c(L) = 1 - \sum_{j=1}^{q} c_{j,\mathbf{s}} z^j \neq 0$ for any $|z| \leq 1$.

(C3) For each fixed $\mathbf{s} \in \mathscr{S}$, $\{\varepsilon_t(\mathbf{s}), t = 0, \pm 1, \pm 2, \ldots\}$ is an i.i.d. sequence with mean 0 and variance 1.

(C4) For any fixed t,

$$\operatorname{Cov}\big(\varepsilon_t(\mathbf{s}), \varepsilon_t(\mathbf{r})\big) = \rho(\mathbf{s}, \mathbf{r}; \boldsymbol{\theta}), \qquad \mathbf{s}, \mathbf{r} \in \mathscr{S},$$

where $\rho(\cdot, \cdot; \boldsymbol{\theta})$ is a correlation function on the space.

In the special case of $q = 0$, and using the conditions (C1)–(C3)–(C4), the spatiotemporal process $\{Y_t(\mathbf{s})\}$ is well defined, and the covariance $\operatorname{Cov}(Y_t(\mathbf{s}), Y_k(\mathbf{r}))$ is uniquely determined in terms of $\rho(\cdot, \cdot; \boldsymbol{\theta})$ and $\{b_{j,\mathbf{s}}, \sigma_{\mathbf{s}}\}$, with

$$\operatorname{Cov}\big(Y_t(\mathbf{s}), Y_k(\mathbf{r})\big) = \sigma_{\mathbf{s}} \sigma_{\mathbf{r}} \sum_{j=1}^{\infty} b_{\mathbf{s}}^{j} \, b_{\mathbf{r}}^{k-t+j} \rho(\mathbf{s}, \mathbf{r}; \boldsymbol{\theta}) \qquad (k \geq t).$$

Note that for each fixed $\mathbf{s}$, time-series $\{Y_t(\mathbf{s}), t = 0, \pm 1, \pm 2, \ldots\}$ is strictly stationary. When $b_{j,\mathbf{s}} \equiv b_j$, $\sigma_{\mathbf{s}} \equiv \sigma$ and $\rho(\mathbf{s}, \mathbf{r}; \boldsymbol{\theta}) = \rho(\mathbf{s} - \mathbf{r}; \boldsymbol{\theta})$ for any $\mathbf{s}, \mathbf{r} \in \mathscr{S}$, the spatiotemporal process $\{Y_t(\mathbf{s})\}$ is weakly stationary. We may also let the order $p = p(\mathbf{s})$, which varies over different locations. This will add little complication to our approach. We do not make this explicit in order to avoid congregated notation.

3.3. NON-STATIONARY MODELS

In the case of non-stationary spatiotemporal models, we introduce models of two types. First, we have models with pre-specified differentiation parameter known as ARIMA models, and second, the ones with unknown differentiation parameter known as ARFIMA models.

In the first case, we generalize the class of ARMA models to include differencing to count for a certain type of non-stationarity. Thus, we introduce a class of spatial d-differenced ARIMA(p, d, q) spatial process with pth Autoregressive and qth Moving Average components. Let $\{Y_t(\mathbf{s}), t = 0, \pm 1, \pm 2, \ldots\}$ be an ARIMA process at location $\mathbf{s} \in \mathscr{S} \subset R^2$:

$$b(L) \cdot \Delta^d Y_t(\mathbf{s}) = c(L) \cdot \varepsilon_t(\mathbf{s}), \tag{3.5}$$

where $\varepsilon(\mathbf{s}) \sim N(0, \sigma_{\mathbf{s}}^2)$ is a Gaussian spatial process and $\Delta = (1 - L)$ with $LY_t(\mathbf{s}) = Y_{t-1}(\mathbf{s})$ the backward shift operator. In the case the parameter d is an integer, the above specification becomes a dth differenced ARIMA(p, d, q) spatial process. In the case where the differencing parameter d is a real value, fractional differencing is permitted giving rise to the Autoregressive Fractionally Integrated Moving Average, or ARFIMA(p, d, q) spatial model.

3.4. GAUSSIAN MLE

Here we apply a Gaussian Maximum Likelihood estimation (MLE) method for the existing spatiotemporal models. It is well known, however, that the Gaussian ML (GML)-fitting is very similar to the Generalized Least Squares (GLS)-fitting, whereby the inverse

covariance matrix plays the role of a weight matrix. The additional feature of the GML is the inclusion of the determinant of the variance-covariance matrix, which serves as an overall measure of dispersion (uncertainty). Finally, the GML method can generate ML estimates for the variance-covariance matrix using greed-search suitable for the overall likelihood function estimation.

3.4.1. Estimators

Suppose the simplest version of the (3.4) is a model with $q = 0$. Let $\{Y_t(\mathbf{s}_j),\ t = 1-p, 2-p, \ldots, T;\ j = 1, \ldots, N\}$ be observations from the above model. We introduce some notation first. Let $\mathbf{X}(\mathbf{s})$ be the $T \times p$ matrix with $Y_{j-i}(\mathbf{s})$ as the (i, j)th element, and $\mathbf{X} = \text{diag}\{\mathbf{X}(\mathbf{s}_1), \ldots, \mathbf{X}(\mathbf{s}_N)\}$ a $(T \times N) \times (p \times N)$ matrix for the model (3.4). Let $\mathbf{R}(\boldsymbol{\theta})$ be an $N \times N$ matrix with $\rho(\mathbf{s}_i, \mathbf{s}_j; \boldsymbol{\theta})$ as the (i, j)th element, $\boldsymbol{\Lambda} = \text{diag}(\sigma_{\mathbf{s}_1}, \ldots, \sigma_{\mathbf{s}_N})$ and $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\mathbf{R}(\boldsymbol{\theta})\boldsymbol{\Lambda}$. The actual choice among these covariance functions takes part at the MLE stage. We define, $\mathbf{Y} = (\mathbf{Y}(\mathbf{s}_1) \cdots \mathbf{Y}(\mathbf{s}_N))'$, $\mathbf{Y}(\mathbf{s}) = (Y_1(\mathbf{s}) \cdots Y_T(\mathbf{s}))'$, $\mathbf{b} = (\mathbf{b}(\mathbf{s}_1) \cdots \mathbf{b}(\mathbf{s}_N))'$ and $\mathbf{b}(\mathbf{s}) = (b_{1,\mathbf{s}} \cdots b_{p,\mathbf{s}})'$. Now the conditional Gaussian likelihood function given $\{Y_t(\mathbf{s}_j),\ t \le 0;\ j = 1, \ldots, N\}$ can be written as

$$\frac{|\mathbf{R}(\boldsymbol{\theta})|^{-\frac{T}{2}}}{\{\prod_{j=1}^{N} \sigma_{\mathbf{s}_j}\}^T} \exp\left\{-\frac{1}{2}(\mathbf{Y} - \mathbf{Xb})^{\tau}(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_T)(\mathbf{Y} - \mathbf{Xb})\right\}, \tag{3.6}$$

where $\otimes$ denotes the matrix Kronecker product and $\mathbf{I}_T$ the $T \times T$ identity matrix. By maximizing the above function, we obtain the Gaussian maximum likelihood estimators $\widehat{\mathbf{b}}, \widehat{\boldsymbol{\theta}}$ and $\widehat{\sigma}_{\mathbf{s}_j}$, which should be referred to as *quasi-maximum likelihood estimators*. Here $\widehat{\mathbf{b}}$ and $\widehat{\sigma}_{\mathbf{s}_j}$ are estimated using LS-methods to location by location and $\widehat{\boldsymbol{\theta}}$ by minimizing[1]:

$$\ell(\widehat{\boldsymbol{\theta}}, \widehat{\mathbf{b}}, \widehat{\sigma}_{\mathbf{s}_j}; \mathbf{Y}) = \log\left\{\frac{1}{N}\sum_{j=1}^{N}(Y_j - \widehat{Y}_j)^2/v_{j-1}\right\} + \frac{1}{N}\sum_{j=1}^{N}\log v_{j-1}. \tag{3.7}$$

We refer to this as the *reduced likelihood function*, where Y_j represents the observed process. Here $v_j = \text{Var}(Y_{j+1} - \widehat{Y}_{j+1})$, denotes the estimation error variance and $\widehat{Y}_j$ the estimate in the Innovation Algorithm (see the Appendix) (Brockwell and Davis 1991; Yao and Brockwell 2006).

Although we are able to calculate the likelihood function, we need to solve a minimization problem involving $N \times (p + q + 2) + d$ parameters for model (3.4), so as to obtain the quasi-maximum likelihood estimators. Here d denotes the number of the unknown parameters in the $\boldsymbol{\theta}$ parameter. As an alternative we first estimate $\mathbf{b}(\mathbf{s})$ and $\sigma_{\mathbf{s}}$ based on the observations at the location $\mathbf{s}$ only, and then we estimate $\boldsymbol{\theta}$ from the residuals from all locations in terms of the Innovation Algorithm. The resulting estimators may be less efficient than the pseudo-maximum likelihood estimators, but the minimization problem contains only d unknown parameters.

[1] In the case where $q \neq 0$, MLE is essential for both parameters' vector estimation.

3.5. RESULTS

In this subsection we present some empirical results from the use of the periodic-trend, spatial ARMA, and spatial non-stationary models. By applying the data series described in Section 2 from the eight locations in the Athens region, we estimate the above models using various specification assumptions. We then derive the *reduced likelihood function* for each specification. The purpose of this is to compare the introduced models for the level of this likelihood estimator. The specification that we are going to favor is the one that gets the minimum value. Also, we present the ACF and the covariance function of the fitted residuals as an additional diagnostic tool to the likelihood function.

First, to build the candidate periodic models, we focus on the temporal structure of the data series introduced in Section 2. What we infer from the results in Figure 1, is the autocorrelation structure of our representative series. In all series, we see that the level of periodicity is above 300 due to the nature of our data series which are daily. Thus, the model assumptions we impose are based on these primary data considerations. Table 1 presents the results for the *PER* value between 300 and 380. For $PER = 360$ we get the minimum *reduced likelihood* value, i.e. -278.6. For values far from this range, we observe log-likelihood values further away from its minimum.

Second, in the ARIMA(p, d, q) models we experiment with either stationary ($d = 0$) or non-stationary ($d = 1$) model assumptions. For each stationary and non-stationary case, we select various p and q values that correspond to an informal identification that uses the ACF as key criterion. Table 2 presents these results. For $p = 1$, $d = 1$, and $q = 1$ we get the minimum *reduced likelihood* value, i.e. -309.0. By imposing specification assumptions that deviate from the above order values, we get likelihood estimator much away from the minimum. In the case of ARFIMA models, we see that under various p and q specifications the likelihood criterion values are higher than the minimum recorded in the ARIMA$(1, 1, 1)$ model (see Table 3).

To robustify the above likelihood results, we compare the spatiotemporal fitting of various models using the spatial covariance function and the ACF of the fitted residuals. We use the best likelihood-performed models among the periodic-trend, the ARMA, the ARIMA, and the ARFIMA models to make this comparison. Figure 3 pictures the covariance functions of the fitted residuals from four different model specifications, i.e., the $PER = 360$, the AR(4), the ARIMA$(1, 1, 1)$, and the ARFIMA$(1, d, 1)$. What we see is that the covariance function in the AR(4), the ARIMA$(1, 1, 1)$, and the ARFIMA$(1, d, 1)$ models is nearly eliminated and captures the whole spatial relationship between locations, whereas in the periodic-trend specification, there is still some spatial relationship uncaptured. In Figure 4 we demonstrate the ACF of the fitted residuals from the model specifications used

Table 1. Reduced log-likelihood estimators for various Periodic model specifications using O_3 daily data from the period 1994–2000.

$PER = 300$	$PER = 310$	$PER = 320$	$PER = 330$	$PER = 340$	$PER = 360$	$PER = 370$	$PER = 380$
-4.2	-4.7	-7.2	-17.3	-251.2	-278.6	-22.3	-10.7

Table 2. Reduced log-likelihood estimators for various ARIMA(p, d, q) model specifications using O_3 daily data from the period 1994–2000.

	$p=1$	$p=2$	$p=3$	$p=4$	$p=5$	$p=6$	$p=7$	$p=10$
$d=0, q=0$	−86.3	−93.9	−133.0	−294.5	−214.1	−152.0	−119.2	543.7
$d=0, q=1$	−94.1	−107.2	−141.9	−298.1	−215.5	−154.0	−119.4	561.0
$d=0, q=2$	−35.9	−92.8	−101.9	−118.6	−121.7	−101.6	−69.3	602.9
$d=1, q=0$	−297.1	−139.1	−128.9	−95.1	−92.7	−87.1	−55.8	396.4
$d=1, q=1$	−309.0	−298.3	−215.8	−100.9	−105.8	−99.6	−62.4	245.6
$d=1, q=2$	−289.5	−211.3	−144.9	−98.0	−89.9	−62.0	−57.0	471.1

Table 3. Reduced log-likelihood estimators for various ARFIMA(p, d, q) model specifications using O_3 daily data from the period 1994–2000.

	$p=1$	$p=2$	$p=3$	$p=4$	$p=5$	$p=6$	$p=7$	$p=10$
$q=0$	−109.4	−84.9	−76.4	−22.9	−10.9	109.0	306.4	503.0
$q=1$	−291.2	−192.5	−179.0	−94.9	−90.2	189.3	302.0	483.7
$q=2$	−57.0	−44.0	−23.3	−11.1	−3.8	121.0	401.7	672.0

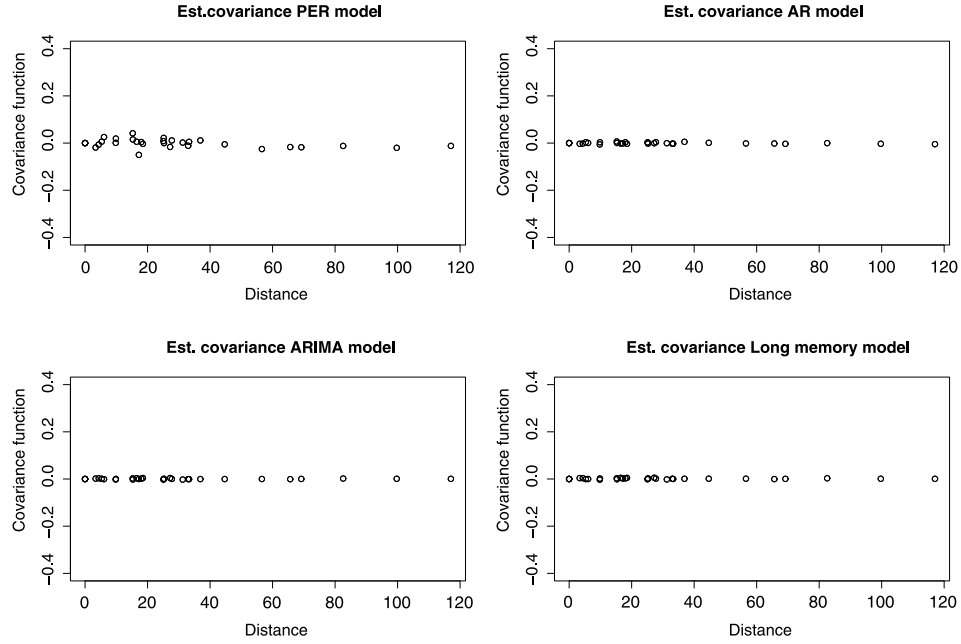

Figure 3. Empirical and estimated errors covariance functions of the $PER = 360$, the AR(4), the ARIMA(1, 1, 1), and the ARFIMA(1, d, 1) models using O_3 daily data from the period 1994–2000.

also in Figure 3. We observe that the ACF for the ARIMA(1, 1, 1) model dies off quicker than in any other spatiotemporal model. Results similar to those obtained for *Athinas* station, which favor the ARIMA(1, 1, 1) specification, can also be seen at the remaining stations.

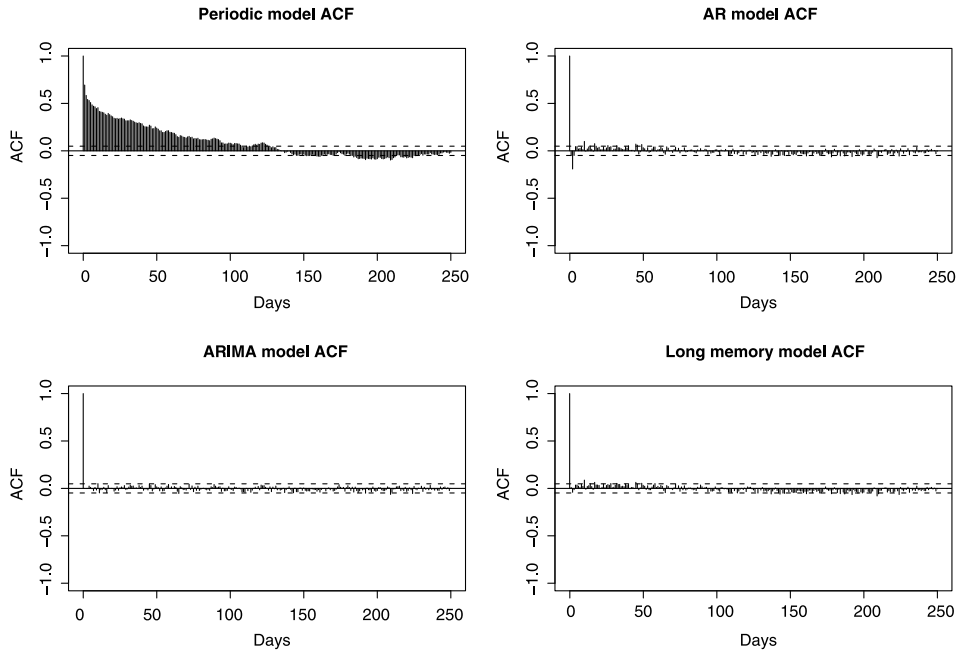

Figure 4. Empirical data and estimated errors ACF of the $PER = 360$, the AR(4), the ARIMA(1, 1, 1), and the ARFIMA(1, d, 1) models using *Athinas* O_3 daily data from the period 1994–2000.

4. SPATIAL PREDICTION

This section is dedicated to spatial estimation at any unmonitored location in the Athens area. This is performed using Ordinary Kriging (OK) (Kyriakidis and Journel 2001). Kriging is an interpolation technique that aims to estimate any unobserved spatial location by using the observed location estimates.

Here, we regionalize the discrete estimation results in the 2D-plane of the Athens area. By using estimation results, such as the autoregressive coefficients, the moving average coefficients, and the error variance of the best fitted ARIMA(1, 1, 1) model, we perform an OK estimation for all the unmonitored locations. In kriging, an assumption has to be made about the spatial behavior of the stochastic term embodied in the semi-variogram. Here kriging is performed using three alternative semi-variogram models such as the Exponential, the Gaussian and the Spherical models (Cressie 1993). With regard to the nugget, sill and range effects, these are specified by an empirical semi-variogram. The choice among competing models is made by the criterion of the minimization of the kriging standard errors. In all the OK estimations, the Spherical models minimize the estimated standard errors.

In Figures 5, 6, 7, we present the maps of the OK-derived parameter estimates in a large area of Athens. In these figures, we demonstrate four representative estimated parameters from the ARIMA(1, 1, 1) model, i.e. $b_{1,\mathbf{s}}$, $c_{1,\mathbf{s}}$ and $\sigma_{\mathbf{s}}$ coefficients. Due to the first-difference transformation we introduce in the ARIMA(1, 1, 1) model, the $b_{1,\mathbf{s}}$ represents the first-order coefficient of the maximum O_3 rate of change. In Figure 5, we demonstrate

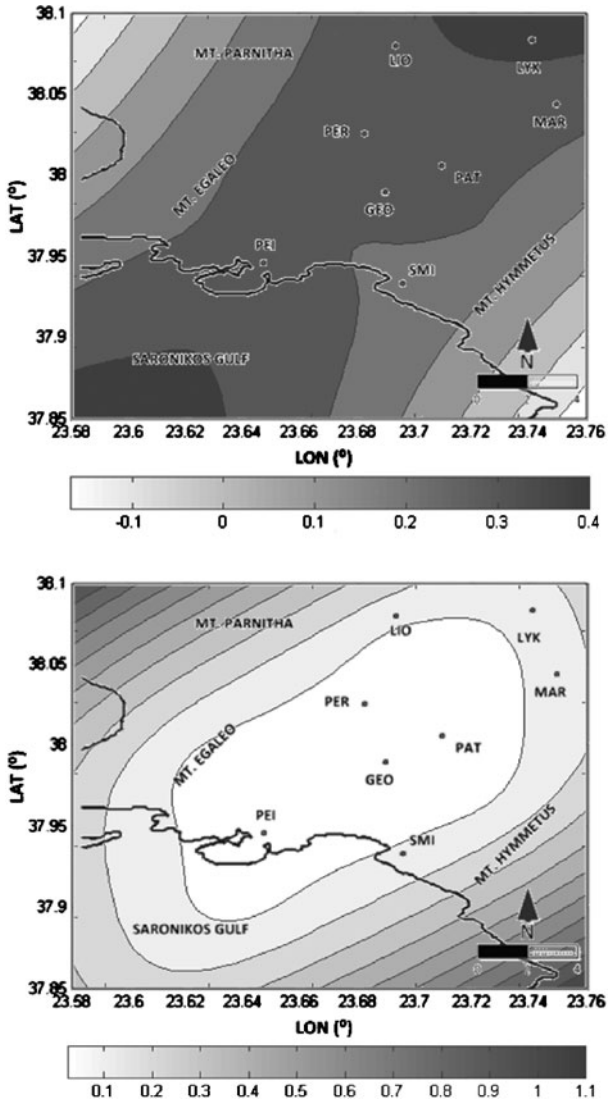

Figure 5. Kriging estimates (top) and standard deviations (bottom) of the AR(1) coefficient using O_3 daily data from the period 1994–2000.

the kriging estimates and standard deviations of the AR(1) parameter values. The AR(1) parameter estimates show homogeneous autoregressive values in the central Athens region. However, *Nea Smirni* and *Lykovrisi* reveal respectively lower and higher than the average values. The kriging standard errors in parametric estimates demonstrate much smoothness with minimum values around the stations.

In Figure 6, we demonstrate the kriging estimates and standard deviations of the MA(1) parameter. The kriging MA(1) parameter estimates show that the central, the southern, and the northern locations have errors with highly negative temporal dependency whereas the

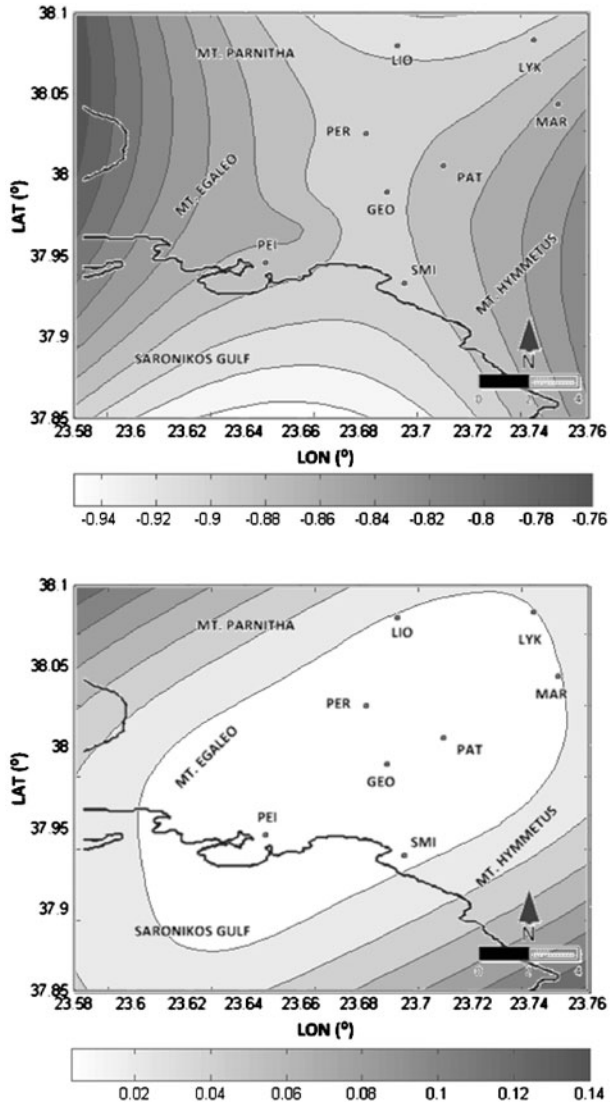

Figure 6. Kriging estimates (top) and standard deviations (bottom) of the MA(1) coefficient using O_3 daily data from the period 1994–2000.

western and the eastern ones demonstrate low negative temporal dependency. The standard deviations estimates of the kriging errors seem very low in all Athens 2-D plane.

In Figure 7, we demonstrate the kriging estimates and standard deviations of the errors' standard deviation parameter. The small standard deviations of the errors reveal that any spatial variation has been eliminated, which is useful for the estimated model. Here, the smallest standard deviations values for the kriging errors are observed in the southwestern and the northeastern regions of the Athens area. Thus, the relatively high values of standard deviations in the northwestern and southeastern parts of the Athens region reveal the need

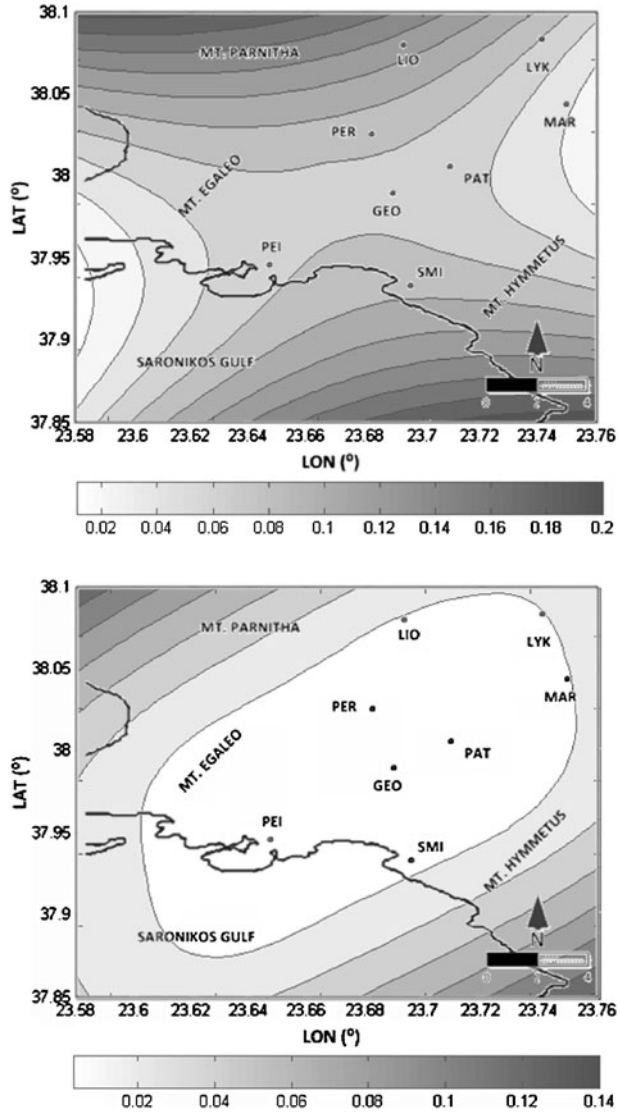

Figure 7. Kriging estimates (top) and standard deviations (bottom) of the error's standard deviation using O_3 daily data from the period 1994–2000.

for a new monitoring station allocation. In terms of the kriging standard errors, the standard deviations demonstrate the lowest values among all the other estimators.

5. DISCUSSION

In this work we have addressed a problem of O_3 concentration modeling over the large metropolitan area of Athens. After investigating the temporal and spatial structure of our data series, we propose alternative spatiotemporal models. Using LS and MLE, we define the parametric estimates as well as estimates of the reduced likelihood function of the alter-

native spatiotemporal models. From these inference results, we derive the fitted residuals spatial covariance functions and ACF. We find that most of the existing spatial and temporal dependencies are best captured by the ARIMA(1, 1, 1) model. By using the minimum reduced log-likelihood criterion we also find that the ARIMA(1, 1, 1) performs better than any other spatiotemporal model. Finally, a spatial prediction estimation is performed using kriging. This is implemented by capturing the ARIMA(1, 1, 1) parametric estimates over a large map over Athens.

Findings show strong spatial and mostly temporal dependencies. The use of these models can further contribute to a better forecasting of various urban air pollutants and also increasing the accuracy of environmental studies in parts of the Athens area not covered currently by monitoring stations.

APPENDIX

A.1. INNOVATION ALGORITHM

Let $\widehat{X}_1 = 0$. For $k \geq 1$, let

$$\widehat{X}_{k+1} \equiv \varphi_{k1} X_k + \cdots + \varphi_{kk} X_1 \tag{A.1}$$

be the best linear predictor for X_{k+1} based on $X_k, \ldots, X_1$ in the sense that

$$E(X_{k+1} - \widehat{X}_{k+1})^2 = \min_{\{\psi_j\}} E\left(X_{k+1} - \sum_{j=1}^{k} \psi_j X_{k-j+1} \right)^2. \tag{A.2}$$

Then the coefficients $\{\varphi_{kj}\}$ are the solutions of the equations

$$\gamma(\mathbf{s}_{k+1}, \mathbf{s}_{k+1-i}; \boldsymbol{\theta}) = \sum_{j=1}^{k} \varphi_{kj} \gamma(\mathbf{s}_{k-j+1}, \mathbf{s}_{k-i+1}; \boldsymbol{\theta}), \qquad i = 1, \ldots, k.$$

Further,

$$v_k \equiv v_k(\boldsymbol{\theta}) \equiv \mathrm{Var}(X_{k+1} - \widehat{X}_{k+1}) = \gamma(\mathbf{s}_{k+1}, \mathbf{s}_{k+1}; \boldsymbol{\theta}) - \sum_{j=1}^{k} \varphi_{kj} \gamma(\mathbf{s}_{k+1}, \mathbf{s}_{k+1-j}; \boldsymbol{\theta}).$$

The least-square property also implies that $\mathrm{Cov}(X_{k+1} - \widehat{X}_{k+1}, X_j) = 0$ for any $1 \leq j \leq k$. Thus $\{X_j - \widehat{X}_j, j = 1, \ldots, N\}$ are uncorrelated. We can see from (A.1) that X_k can be written as a linear combination of $X_k - \widehat{X}_k, \ldots, X_1 - \widehat{X}_1$. Thus

$$\widehat{X}_{k+1} = \sum_{j=1}^{k} \beta_{kj} (X_{k+1-j} - \widehat{X}_{k+1-j}), \qquad k = 1, \ldots, N-1. \tag{A.3}$$

The coefficients $\{\beta_{kj}\}$ and $\{v_j\}$ can be calculated in the recursive manner in terms of the Innovation Algorithm (Proposition 5.2.2 of Brockwell and Davis 1991) described below. Then the MLE $\widehat{\boldsymbol{\theta}}$ will be calculated via the minimization of (3.7) in which $\widehat{X}_k$ is obtained recursively through (A.3) with $\widehat{X}_1 = 0$. Therefore the calculation for $N \times N$ inverse matrix $\boldsymbol{\Gamma}(\boldsymbol{\theta})^{-1}$ is avoided.

Innovation Algorithm: Set $v_0 = \gamma(\mathbf{s}_1, \mathbf{s}_1; \boldsymbol{\theta})$. Based on the cross-recursion equations

$$\beta_{k,k-j} = v_j^{-1}\left\{\gamma(\mathbf{s}_{k+1}, \mathbf{s}_{j+1}; \boldsymbol{\theta}) - \sum_{i=0}^{j-1} \beta_{j,j-i}\beta_{k,k-i}v_i\right\},$$

$$n_k = \gamma(\mathbf{s}_{k+1}, \mathbf{s}_{k+1}; \boldsymbol{\theta}) - \sum_{j=0}^{k-1} \beta_{k,k-j}^2 v_j,$$

compute the values of $\{\beta_{ij}\}$ and $\{v_j\}$ in the order

$$\beta_{11}, v_1,$$

$$\beta_{22}, \beta_{21}, v_2,$$

$$\beta_{33}, \beta_{32}, \beta_{31}, v_3,$$

$$\ldots\ldots$$

$$\beta_{N-1,N-1}, \beta_{N-1,N-2}, \ldots, \beta_{N-1,1}, v_{N-1}.$$

SUPPLEMENTARY MATERIALS

The spatiotemporal model point estimation and model selection is implemented in C language. The computer program is available by the authors upon request.

ACKNOWLEDGEMENTS

The authors are thankful to the Directorate of Atmospheric and Noise Pollution Control, Hellenic Ministry for the Environment, Physical Planning and Public Works, for providing the data used in this study. The assistance of Dr. Ioannis Venetis and Mr. Nikolaos Manalis is greatly acknowledged.

[Accepted February 2011. Published Online March 2011.]

REFERENCES

Besag, J. (1974), "Spatial Interaction and the Statistical Analysis of Lattice Systems" (with discussion). *Journal of the Royal Statistical Society B*, 36, 192–236.

——— (1975), "Statistical Analysis of Non-lattice Data," *The Statistician*, 24, 179–195.

Box, G. E. P., and Jenkins, G. M. (1970), *Time Series Analysis: Forecasting and Control*, San Francisco: Holden-Day.

Brockwell, P. J., and Davis, R. A. (1991), *Time Series: Theory and Methods*, New York: Springer.

Caroll, 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.

Cressie, N. A. K. (1993), *Statistics for Spatial Data*, New York: Wiley.

——— (1997), "Comment," *Journal of the American Statistical Association*, 92, 411–413.

Handcock, M. S., and Wallis, J. R. (1994), "An Approach to Statistical Spatiotemporal Modelling of Meteorological Fields," *Journal of the American Statistical Association*, 89, 368–378.

Kyriakidis, P. C., and Journel, A. G. (2001), "Stochastic Modelling of Atmospheric Pollution: A Spatial Time Series Framework. Part I: Methodology," *Atmospheric Environment*, 35, 2331–2337.

Mardia, K. V., and Marshall, R. J. (1984), "Maximum Likelihood Estimation of models for Residual Covariance in Spatial Regression," *Biometrika*, 71, 135–146.

NRC (National Research Council) (1991), *Rethinking the Ozone Problem in Urban and Regional air Pollution*, Washington: National Academy Press.

Ord, K. (1975), "Estimation Methods for Models of Spatial Interaction," *Journal of the American Statistical Association*, 70, 120–126.

Romanowicz, R., Young, P., Brown, P., and Diggle, P. (2006), "A Recursive Estimation Approach to the Spatiotemporal Analysis and Modelling of Air Quality Data," *Environmental Modelling and Software*, 21, 759–769.

Shaddick, G., and Wakefield, J. (2002), "Modelling Daily Multivariate Pollutant Data at Multiple Sites," *Journal of the Royal Statistical Society C*, 51, 509–511.

Stein, M. L., and Fang, D. (1997), "Comment," *Journal of the American Statistical Association*, 92, 408–411.

Thompson, M. L., Reynold, J., Cox, L. H., Gutton, P., and Sampson, P. D. (2001), "A Review of Statistical Methods for the Meteorological Adjustment of Tropospheric Ozone," *Atmospheric Environment*, 35, 617–630.

Whittle, P. (1954), "On Stationary Processes in the Plane," *Biometrika*, 41, 434–449.

Yao, Q., and Brockwell, P. J. (2006), "Gaussian Maximum Likelihood Estimation for ARMA Models II: Spatial Processes," *Bernouli*, 12, 403–429.

Ziomas, I. C., Melas, D., Zerefos, C. S., Bais, A., and Paliatsos, A. G. (1995), "Forecasting Peak Pollutant Levels From Meteorological Variables," *Atmospheric Environment*, 29, 3703–3711.