Modeling Transport Effects on Ground-Level Ozone Using a Non-Stationary Space-Time Model

Hsin-Cheng Huang
Institute of Statistical Science
Academia Sinica
Taipei 115, Taiwan
hchuang@stat.sinica.edu.tw
Tel: 886-2-27835611 ext.416
Fax: 886-2-27831523

Nan-Jung Hsu
Institute of Statistics
National Tsing Hua University
Hsin-Chu 300, Taiwan
njhsu@stat.nthu.edu.tw

August 17, 2003

Short Title: Modeling Transport Effects on Ground-Level Ozone
Summary

This article presents a novel autoregressive space-time model for ground-level ozone data, which models not only spatio-temporal dynamics of hourly ozone concentrations, but also relationships between ozone concentrations and meteorological variables. The proposed model has a nonseparable spatio-temporal covariance function that depends on wind speed and wind direction, and hence is nonstationary in both time and space. Ozone concentration for a given location and time is assumed to be directly influenced by ozone concentrations at neighboring locations at the previous time, via a weight function of space-time dynamics caused by wind speed and wind direction.

To our knowledge, the proposed method is the first one to incorporate transport effect of ozone into the spatio-temporal covariance structure. Moreover, it uses a computationaly efficient space-time Kalman filter and can compute optimal spatio-temporal prediction at any location and time very fast for given meteorological conditions. Ozone data from Taipei are used for illustration, in which the model parameters are estimated by maximum likelihood.

Key words: Empirical orthogonal function, Kalman filter, kriging, non-separable spatio-temporal covariance function, wind direction, wind speed.
1 Introduction

Ambient ozone pollution has become a major health hazard in cities worldwide, because high-level ozone may damage human respiratory systems. It is therefore essential to assess ozone trends, predict ozone concentrations, and understand the transform patterns of ozone. Since topography, meteorology, and emission patterns of ozone precursors can heavily affect ozone concentrations, an appropriate spatio-temporal covariance function would be heterogeneous across space and nonstationary over time. Additionally, the spatial correlation of ozone concentrations between two sites lying in the direction of wind tends to be stronger than those lying in other directions (i.e., anisotropy), and the cross-correlation function of ozone between two sites lying in the direction of wind is typically asymmetric about zero due to the transport effect. This asymmetric feature can not be produced from a commonly used separable spatio-temporal covariance function, which is a purely spatial covariance function multiplied by a purely temporal covariance function. Therefore, modeling the spatio-temporal covariance structure of ground-level ozone by a separable spatio-temporal covariance function would be inappropriate.

Although statistical modeling of ground-level ozone has received considerable interest (see Thompson et al., 1999, for a review), statistical analysis that accounts for both spatial and temporal dependence has seldom been addressed. Notable exceptions are Guttorp et al. (1994), Carroll et al. (1997), and Guttorp et al. (1997). However, these investigations did not incorporate meteorological variables, such as wind speed and wind direction, into the spatio-temporal covariance function.

Modeling ground-level ozone poses several challenges. First, conventional space-time models mostly have separable spatio-temporal covariance functions, which is inappropriate when modeling the transport patterns of ozone concentrations. Recent
investigations have proposed nonseparable space-time models (Cressie and Huang, 1999; Brown et al., 2000; Gneiting, 2002). Although these models may capture some nonseparable feature caused by the transport effect, their spatio-temporal covariance functions are stationary and symmetric in both space and time, making it difficult to provide further insight into the asymmetric feature caused by wind. Second, since the amount of ozone monitoring data is typically very large, the model must be computationally feasible.

To obtain a more realistic space-time model for ground-level ozone, we incorporate the transport pattern of ozone into the spatio-temporal covariance structure. We propose an autoregressive space-time model for ground-level ozone data, which models the spatio-temporal variations of hourly ozone concentrations, as well as relationships between ozone concentrations and meteorological variables. With a spatio-temporal covariance structure depending on wind speed and wind direction, the proposed model has a nonseparable spatio-temporal covariance function that is nonstationary in both time and space. Specifically, the ozone concentration at a location and time is assumed to be directly influenced by ozone concentrations at neighboring locations at previous time via a weight function. This weight function is further specified parametrically to have space-time dynamics caused by wind conditions.

The proposed model is extended and modified from the space-time model derived by Wikle and Cressie (1999, 2000) for modeling atmospheric processes, where no covariate is incorporated in the spatio-temporal covariance structure, and temporal stationarity is assumed. The proposed method allows us to model the patterns of transport of ozone. Moreover, the optimal spatio-temporal predictor and the corresponding conditional mean-squared prediction error can be obtained at any location and time using a space-time Kalman filter (Huang and Cressie, 1996; Mardia et al., 1998; Wikle and Cressie, 1999) under given meteorological conditions.
The rest of this paper is organized as follows. Section 2 introduces the proposed method, for which a state-space representation is derived. Section 3 describes how the model parameters are estimated using maximum likelihood (ML). Section 4 applies the proposed method to the ground-level ozone data from Taiwan Air Quality Monitoring Network. A brief discussion is drawn in Section 5. The Appendix section contains a technical proof.

2 The Proposed Model

Consider a spatio-temporal process \( \{S(s, t) : s \in D, t = 1, 2, \ldots\} \) defined over a spatial region \( D \), with \( |D| > 0 \), and time points \( t \in \mathbb{N} \equiv \{1, 2, \ldots\} \), where \( S(s, t) \) is used to represent the ozone concentration at location \( s \) and time \( t \). Assume that the process \( S(\cdot, \cdot) \) can be decomposed into the following:

\[
S(s, t) = \mu(s, t) + Y(s, t) + \nu(s, t); \quad s \in D, \ t \in \mathbb{N},
\]

where \( \mu(\cdot, \cdot) \) is a deterministic mean process, \( Y(\cdot, \cdot) \) is a zero-mean spatio-temporal process, \( \nu(\cdot, \cdot) \) is a spatially stationary and temporally uncorrelated process that represents the small-scale spatio-temporal variation, and \( Y(\cdot, \cdot) \) and \( \nu(\cdot, \cdot) \) are uncorrelated. The process \( Y(\cdot, \cdot) \) is assumed to evolve according to the state equation:

\[
Y(s, t) = \int_D w(s, u, x(s, t - 1))Y(u, t - 1)du + \eta(s, t); \quad s \in D, \ t \in \mathbb{N},
\]

where \( \eta(\cdot, \cdot) \) is a temporally uncorrelated and identically distributed spatial error process with \( \text{cov}(\eta(s, t), Y(s', t - 1)) = 0; s, s' \in D, t \in \mathbb{N} \), and \( w(s, u, x(s, t - 1)) \) is a weight function that depends on some covariates \( x(s, t - 1) \). Herein, \( x(s, t - 1) \equiv (x_1(s, t - 1), x_2(s, t - 1))^\prime \) is assumed to be the vector of wind speeds in the east-west and the north-south directions. The weight function \( w(s, u, x(s, t - 1)) \) represents the relationship between the ozone concentration at \( (s, t) \) and that at \( (u, t - 1) \) under the
wind-speed field \( x(\cdot, t - 1) \). Since the weight function changes with the wind-speed field, the process \( Y(\cdot, \cdot) \) is in general temporally nonstationary.

We assume that the data \( \{Z(s_i, t) : i = 1, \ldots, n, \ t = 1, \ldots, T\} \) are observed (perhaps incompletely) at \( n \) monitoring sites and \( T \) time points according to the following measurement equation:

\[
Z(s, t) = S(s, t) + \varepsilon(s, t) = \mu(s, t) + Y(s, t) + \nu(s, t) + \varepsilon(s, t); \quad s \in D, \ t = 1, \ldots, T, \quad (3)
\]

where \( \varepsilon(\cdot, \cdot) \) is a white-noise process with variance \( \sigma^2_{\varepsilon} \), representing the measurement error. The model is an extension of the space-time model proposed by Wikle and Cressie (1999, 2000), where the weight function does not depend on any covariate, and is time invariant. This article focuses on modeling the transport patterns of ozone through the weight function. This weight function is further specified with some parametric structure corresponding to diffusion and transport effects.

## 2.1 Reduced State-Space Representation

In Wikle and Cressie (1999, 2000), the process \( Y(s, t) \) and the time-invariant weight function are modeled in terms of a small number of basis functions \( \{\phi_1(\cdot), \ldots, \phi_K(\cdot)\} \), where \( K \) is small relative to \( n \). In our study, a different approach is adapted by first considering a large number of orthonormal basis functions \( \{e_1(\cdot), \ldots, e_N(\cdot)\} \), which are flexible enough to model the weight function and \( Y(s, t) \) at a fine resolution.

For convenience, we use a very simple class of orthonormal basis functions for the application in Section 4. Specifically, we assume that \( D = \bigcup_{i=1}^N D_i \) is a rectangular region, where \( \{D_1, \ldots, D_N\} \) are disjoint regular grid cells with unit areas, and let \( e_i(s) = I(s \in D_i); \ i = 1, \ldots, N, \) resulting in a space-time model whose spatial indices can be regarded as defined on a lattice. The reader should keep in mind that the
model developed in this paper applies to more general (e.g., smooth) orthonormal basis functions \(\{e_1(\cdot), \ldots, e_N(\cdot)\}\).

Additionally, we assume that \(Y(s,t), w(s,u,x(s,t))\) and \(\eta(s,t)\) can be decomposed in terms of \(\{e_i(\cdot)\}\) into

\[
Y(s,t) = \sum_{k=1}^{N} a_k(t)e_k(s); \quad s \in D, \ t \in \mathbb{N}, \tag{4}
\]

\[
w(s,u,x(s,t-1)) = \sum_{k=1}^{N} \sum_{l=1}^{N} c_{k,l}(t-1)e_l(u)e_k(s); \quad s,u \in D, \ t \in \mathbb{N}, \tag{5}
\]

\[
\eta(s,t) = \sum_{k=1}^{N} \xi_k(t)e_k(s); \quad s \in D, \ t \in \mathbb{N}, \tag{6}
\]

where \(\{a_k(t) : k = 1, \ldots, N, \ t = 0,1,\ldots\}\) and \(\{\xi_k(t) : k = 1, \ldots, N, \ t \in \mathbb{N}\}\) are zero-mean random variables. By (5), the coefficients \(\{c_{k,l}(t-1)\}\) satisfy

\[
c_{k,l}(t) = \int_D \int_D w(s,u,x(s,t))e_k(s)e_l(u)ds \, du; \quad k,l = 1,\ldots,N, \ t \in \{0,1,\ldots\},
\]

in which \(x(\cdot,t-1)\) is omitted in the notation of \(\{c_{k,l}(t-1)\}\) for simplicity. Note that the decompositions given in (4)–(6) are quite general. In fact, any \(L^2\)-continuous spatio-temporal processes \(Y(s,t)\) and \(\eta(s,t)\), and any continuous weight function \(w(s,u,x(s,t-1))\) can be well approximated by these representations if the function space generated by these orthonormal basis functions \(\{e_1(\cdot), \ldots, e_N(\cdot)\}\) is sufficiently large. Note that, for \(e_k(s) = I(s \in D_k)\), we have \(Y(s,t) = a_k(t)\) for \(s \in D_k\).

Based on (4)–(6), the following proposition is obtained:

**Proposition 1** Under the assumptions (4)–(6), the state equation (2) is equivalent to the following multivariate representation:

\[
a(t) = C_{t-1} a(t-1) + \xi(t); \quad t \in \mathbb{N}, \tag{7}
\]

where \(a(t) = (a_1(t), \ldots, a_N(t))'\), \(\xi(t) = (\xi_1(t), \ldots, \xi_N(t))'\) and \(C_{t-1}\) is the \(N \times N\) matrix with the \((k,l)\)-th element \(c_{k,l}(t-1)\).
In (7), the transition matrix $C_{t-1}$ changes over time since it is a function of $x(\cdot, t-1)$. Consequently, $\{a(t)\}$ is a nonstationary $N$-dimensional AR(1) process.

Instead of handling the high-dimensional time series $\{a(t)\}$ directly, we look for a linear transformation $U$ of $a(t)$ such that (7) can be well approximated by a lower dimensional representation. Let $a^*(t) \equiv (a_1^*(t), \ldots, a_N^*(t))^\prime \equiv Ua(t)$. Then (7) can be represented as:

$$a^*(t) = (UC_{t-1}U^{-1})a^*(t-1) + U\xi(t); \quad t \in \mathbb{N}.$$  

Suppose there exists $U$ satisfying $\sum_{k=L+1}^{N} \text{var}(a_k^*(t)) \approx 0$ for all $t \in \{0, 1, \ldots\}$, then (7) can be approximately reduced to

$$a_L^*(t) = C_{t-1}^*a_L^*(t-1) + \xi_L^*(t); \quad t \in \mathbb{N}, \quad (8)$$

where $a_L^*(t)$ and $\xi_L^*(t)$ are the vectors consisting of the first $L$ components of $a^*(t)$ and $\xi^*(t)$, respectively, $C_{t-1}^* = (U_LC_{t-1}\Sigma_tU_L^{-1})(U_L\Sigma_tU_L^{-1})^{-1}$, and $U_L$ is the $L \times N$ matrix consisting of the first $L$ rows of $U$. In particular, if the row vectors $u_1, \ldots, u_N$ of $U$ are eigenvectors (principal components) of $\Sigma_{t_0} = \text{var}(a(t_0))$ for some $t_0 \in \mathbb{N}$, the basis functions $\{\phi_k(\cdot) \equiv u_k^\prime \epsilon(\cdot); \; k = 1, \ldots, N\}$ are called the empirical orthogonal functions (Cohen and Jones, 1969; Buell, 1972) of the spatial process $Y(\cdot, t_0)$, where $\epsilon(s) \equiv (\epsilon_1(s), \ldots, \epsilon_N(s))^\prime$.

Incorporating (8) into (4), we have

$$Y(s, t) = \sum_{k=1}^{N} a_k(t)e_k(s) = \sum_{k=1}^{N} a_k^*(t)\phi_k(s) \approx \sum_{k=1}^{L} a_k^*(t)\phi_k(s); \quad s \in D, \; t \in \{0, 1, \ldots\}.$$  

Therefore, the measurement equation (3) satisfies

$$Z(t) = \mu(t) + \Phi_La_L^*(t) + \nu(t) + \epsilon(t); \quad t \in \mathbb{N}, \quad (9)$$

where $Z(t) \equiv (Z(s_1, t), \ldots, Z(s_n, t))^\prime$, $\mu(t) \equiv (\mu(s_1, t), \ldots, \mu(s_n, t))^\prime$, $\nu(t) \equiv (\nu(s_1, t), \ldots, \nu(s_n, t))^\prime$, $\epsilon(t) \equiv (\epsilon(s_1, t), \ldots, \epsilon(s_n, t))^\prime$, and $\Phi_L$ is the $n \times L$ matrix with the $(i, k)$-th element
\( \phi_k(s_i) \). Consequently, (8) and (9) form a reduced state-space representation under (1), (2) and (4)–(6).

### 2.2 Selection of Linear Transformation \( U \)

We now describe how to select the linear transformation \( U \). According to the concept of principal component analysis, \( \{u_1, \ldots, u_N\} \) are selected as the eigenvectors of an estimated covariance matrix of \( a(t) = (a_1(t), \ldots, a_N(t))^\prime \) for some \( t \in N \), with the corresponding eigenvalues in non-descending order. Since the data are only observed at \( n \) out of \( N \) sites, one common method for estimating \( \text{var}(a(t)) \) is first to predict \( a(t) \) by a stochastic or nonstochastic method, and then compute the sample covariance matrix of the predicted \( a(t) \) as an estimate (e.g., Karl et al., 1982; Wikle and Cressie, 1999). However, the underlying spatial covariance matrix is usually underestimated by this procedure. The reason is explained by the following example. Suppose that \( a(t) \) is predicted by a kriging method based on \( Z(t) \). Under the Gaussian assumption, the optimal predictor and the corresponding prediction-error variance are given by \( E(a(t)|Z(t)) \) and \( \text{var}(a(t)|Z(t)) \), respectively. Using the following equality

\[
\text{var}(a(t)) = \text{var}[E(a(t)|Z(t))] + E[\text{var}(a(t)|Z(t))],
\]

it is clear that the sample covariance matrix underestimates \( \text{var}(a(t)) \), since it only estimates the first part \( \text{var}[E(a(t)|Z(t))] \) but ignores the second part \( E[\text{var}(a(t)|Z(t))] \) in (10). Note that \( \text{var}(a(t)|Z(t)) = E[\text{var}(a(t)|Z(t))] \) under the Gaussian assumption.

To adjust for underestimation, the second term in (10) is estimated using the prediction-error covariance matrix obtained from a kriging method, which is then added to the sample covariance matrix to obtain the final estimate of \( \text{var}(a(t)) \). We select \( \{u_1, \ldots, u_N\} \) as the eigenvectors (eigenimages) of this new estimate with
the corresponding eigenvalues in non-ascending order. Note that only the first $L$ eigenvectors are used in the reduced model (8) and (9).

An optimal choice of $L$ can be determined via some model-selection criterion, such as AIC. In practice, inference and prediction are believed to be insensitive to the choice of $L$ in the presence of the stationary process $\nu(\cdot, \cdot)$. The reason is that if spatial dependence in $Y(\cdot, \cdot)$ is not captured by the nonstationary mechanism formed by $\Phi_L a^*_L(t)$, it can be accounted for by $\nu(\cdot, \cdot)$, and vice versa. Therefore, a suboptimal choice of $L$ would be sufficient. The decomposition of a spatial process into a stationary process and a linear combination of empirical orthogonal functions has also been used in Nychka and Saltzman (1998), Holland et al. (1999), and Wikle and Cressie (1999).

### 2.3 Specification of Weight Function

In our approach, the transport pattern of ozone is characterized through a weight function. First, the following baseline function is specified

$$w^*(s, u, x; \theta) = \begin{cases} \frac{\theta_0}{c(\theta)} \left[ 1 - \left( \frac{(u_1^*)^2 + (u_2^*)^2}{\theta_1^2} \right)^{1/2} \right]; & \text{if } (u_1^*, u_2^*) \in D^*, \; u_1^* \geq 0, \\ \frac{\theta_0}{c(\theta)} \left[ 1 - \left( \frac{(u_1^*)^2 + (u_2^*)^2}{\theta_1^2} \right)^{1/2} \right]; & \text{if } (u_1^*, u_2^*) \in D^*, \; u_1^* < 0, \\ 0; & \text{otherwise,} \end{cases}$$

(11)

where $\theta \equiv (\theta_0, \theta_1, \theta_2, \theta_3)'$ presents the parameter vector, $v$ and $\tau$ are the polar coordinates of $x$ representing the magnitude and the direction of wind speed, $c(\theta) \equiv \pi[\theta_1 + \theta_2 v + \theta_1 \exp(-\theta_3 v)] \theta_1/6$ is the normalizing constant such that $\int w^*(s, u, x; \theta) du = \theta_0$, and

$$D^* \equiv \left\{ (u_1^*, u_2^*) : \frac{(u_1^*)^2}{(\theta_1 + \theta_2 v)^2} + \frac{(u_2^*)^2}{\theta_1^2} \leq 1 \right\},$$

$$\begin{pmatrix} u_1^* \\ u_2^* \end{pmatrix} \equiv \begin{pmatrix} \cos(\tau) & \sin(\tau) \\ -\sin(\tau) & \cos(\tau) \end{pmatrix} (u - s).$$
Figure 1: about here

This function is characterized by four parameters: \( \theta_0 \) is an autoregressive parameter satisfying \(|\theta_0| < 1\) to ensure the stability of the process \( Y(\cdot, \cdot) \), \( \theta_1 > 0 \) represents the transport range of ozone from neighboring locations when there is no wind, \( \theta_2 > 0 \) and \( \theta_3 > 0 \) are range parameters corresponding to the downwind and upwind directions, respectively. Note that, for \( s \) close to the boundary of \( D \), the function \( w^*(s, \cdot, x; \theta) \) is reweighted so that \( \int_{u \in D} w^*(s, u, x; \theta) du = \theta_0 \).

Since the orthonormal basis functions \( \{e_k(\cdot)\} \) determine the finest resolution for all of our representations, the weight function \( w(s, u, x; \theta) \) is defined by projecting the baseline function \( w^*(s, u, x; \theta) \) to the function space generated by \( \{e_k(s)e_l(u) : k, l = 1, \ldots, N\} \) so that (5) is satisfied. Figure 1 shows the weight function \( w(s, u, x; \theta) \) for various wind speeds \( (v = 0, 3, 6, 9) \) at \( s = 0 \) under northeasterly wind (i.e., \( \tau = \pi/4 \)), with \( \theta_0 = 0.93, \theta_1 = 1.42, \theta_2 = 0.39 \) and \( \theta_3 = 0.40 \). The parameter values in these illustrations are actual estimates from the real data described in Section 4. Note that the weight function is a cone function of \( u \) centered at \( s \) when there is no wind at time \( t \) (i.e., \( x = 0 \)).

### 2.4 Optimal Prediction

Under the Gaussian assumption, if all the parameters are known, the optimal mean-squared-error predictor \( \hat{a}^*(t) \) of \( a^*(t) \) based on observed data \( \{Z(1), \ldots, Z(t)\} \) and the corresponding conditional mean-squared prediction error (CMSPE),

\[
\hat{\Sigma}_t \equiv E((a^*(t) - \hat{a}^*(t))(a^*(t) - \hat{a}^*(t))' \mid Z(1), \ldots, Z(t))
= \text{var}(a^*(t) \mid Z(1), \ldots, Z(t)); \quad t \in \mathbb{N},
\]

can be obtained recursively using the Kalman-filter algorithm (Harvey, 1989). The initial values for implementing the Kalman-filter algorithm can be chosen as \( \hat{a}^*(0) = \)
$\Sigma_0 = \Lambda_0^*$, where $\Lambda_0^*$ is the limiting stationary covariance matrix of $a^*(t)$ corresponding to $x(\cdot, \cdot) \equiv 0$.

The optimal spatio-temporal predictor of $S(s, t)$ based on $Z(1), \ldots, Z(t)$ is given by

$$
\hat{S}(s, t) = E[S(s, t)|Z(1), \ldots, Z(t)] = \mu(s, t) + \hat{Y}(s, t) + \hat{\nu}(s, t); \ s \in D, \ t \in \mathbb{N},
$$

where

$$
\hat{Y}(s, t) = (\phi(s))' \hat{\alpha}_L^*(t),
$$

$$
\hat{\nu}(s, t) = (\gamma_L(s))'[\text{var}(Z(t))]^{-1}(Z(t) - \mu(t)),
$$

in which $\phi(s) \equiv (\phi_1(s), \ldots, \phi_L(s))'$ and

$$
\gamma_L(s) \equiv (\text{cov}(\nu(s, t), \nu(s_1, t)), \ldots, \text{cov}(\nu(s, t), \nu(s_n, t)))'.
$$

The CMSPE of $\hat{S}(s, t)$ is given by

$$
E[(\hat{S}(s, t) - S(s, t))^2|Z(1), \ldots, Z(t)]
= (\phi(s))' \Sigma_L \phi(s) + \text{var}[\nu(s, t)] - (\gamma_L(s))'[\text{var}(Z(t))]^{-1} \gamma_L(s)
$$

$$
-2(\phi(s))' \text{var}[a^*(t)] \Phi_L [\text{var}(Z(t))]^{-1} \gamma_L(s); \ s \in D, \ t \in \mathbb{N}.
$$

### 3 Parameters Estimation

In this section, the data and the processes are assumed to be appropriately detrended. Moreover, the linear transformation $U$ and the number $L$ used in determining the reduced representation of $Y(\cdot, \cdot)$ are assumed to be known. In (9), assume that the covariance structure of $\nu(\cdot, \cdot)$ is characterized by the parameter vector $\psi$. Let

$Q \equiv \text{var} (\xi_L^*(t)) = BB'$ be the Cholesky decomposition where $B = (b_{k,l})$ is a lower triangular matrix with positive diagonal elements.
For the reduced state-space representations (8) and (9), the model parameters consist of $\mathbf{b} \equiv \{b_{k,l} : 1 \leq l \leq k \leq L\}$ characterizing the variations of the variables $\{\xi^*(t)\}$, $\theta$ characterizing the weight function, $\psi$ characterizing the small-scale spatial variations, and $\sigma_{e}^2$ representing the measurement-error variance. The ML method is used for parameter estimation by maximizing the likelihood function:

$$L(\mathbf{b}, \theta, \psi, \sigma_{e}^2) = p(Z(1); \mathbf{b}, \theta, \psi, \sigma_{e}^2) \left\{ \prod_{t=2}^{T} p(Z(t)|Z(1), \ldots, Z(t-1); \mathbf{b}, \theta, \psi, \sigma_{e}^2) \right\},$$

where the one-step conditional distribution satisfies

$$Z(t) | Z(1), \ldots, Z(t-1) \sim N \left( \mu(t) + \Phi_L C^*_{t-1} \hat{a}^*_L(t-1), \Phi_L \hat{\Sigma}_{t-1} \Phi_L' + R + \sigma_e^2 I \right),$$

in which $R \equiv \text{var}(\nu(t))$, and $\hat{\Sigma}_{t-1} \equiv \text{var}(\mathbf{a}^*(t)|Z(1), \ldots, Z(t-1))$ is the one-step CMSPE of $\mathbf{a}^*(t)$. Note that the transition matrix $C^*_t$ also depends on the covariates $x(\cdot,t)$ and the unknown parameters $\theta$.

For implementation, the likelihood at given parameter values can be evaluated efficiently via the Kalman-filter algorithm in which $\hat{\mathbf{a}}^*(t-1)$ and $\hat{\Sigma}_{t-1}$ can be calculated recursively. Alternatively, a less efficient method can be derived such as the method of moments as used in Wikle and Cressie (1999, 2000). Although the method of moments is computationally more efficient than the ML method, care has to be taken to ensure positive definiteness of various covariance matrices.

### 4 Illustrative Example

For illustration, this section describes the procedure for modeling ozone data from Taiwan Air Quality Monitoring Network based on our space-time model. The data consist of 8760 hourly ozone measurements (in parts per billion) at twelve stations in Taipei (Figure 2) in year 1999. Meteorological variables such as wind speed, wind direction, and temperature were also collected hourly at each station. As Taiwan is
located in the monsoon area, the prevailing wind direction in this area is northeasterly during the winter and southwesterly during the summer.

4.1 Exploratory Data Analysis

Similar to Carroll et al. (1997) in analyzing ozone exposure, this study uses the square-root transformation of the ozone data. Figure 3 and Figure 4 reveal that the square-root ozone data \( \{Z(s_i, t) : i = 1, \ldots, 12, t = 1, \ldots, 8760\} \) have strong diurnal and monthly patterns, respectively. Also, the ozone concentrations tend to be higher for higher temperatures (Figure 5). Therefore, the trend component \( \mu(s, t) \) is modeled as

\[
\mu(s, t) = \alpha_{\text{hour}} + \sum_{j=1}^{12} \beta_j f_j(t) + \sum_{k=1}^{9} \gamma_k g_k(\text{temp}(s, t)); \quad s \in D, \quad t \in \mathbb{N},
\]

where \( \alpha_{\text{hour}} \) represents the hour effects, \( \text{temp}(s, t) \) is the temperature (in Celsius) at location \( s \) and time \( t \), \( \{f_1, \ldots, f_{12}\} \) are natural cubic periodic spline basis functions with equally spaced knots at hours \( \{0, 730, \ldots, 8760\} \), and \( \{g_1, \ldots, g_9\} \) are natural cubic spline basis functions with knots at \( \{5, 10, 15, 20, 25, 30, 35, 40\} \). The least-squares (LS) method is used for estimating parameters \( \{\alpha_{\text{hour}}, \beta_j, \gamma_k\} \), where the multiple correlation coefficient of the regression is 0.649. Although statistically more efficient methods such as generalized-least-squares (GLS) method or ML method can be applied, the performance of LS estimates is believed to be nearly equivalent to those estimates due to the large amount of data. Other methods in modeling the trend for ground-level ozone can be found in Huerta et al. (2000). After adjusting for the mean function, the residuals \( \{r(s_i, t) = Z(s_i, t) - \hat{\mu}(s_i, t)\} \) are used for fitting the zero-mean stochastic process \( Z(\cdot, \cdot) - \mu(\cdot, \cdot) = Y(\cdot, \cdot) + \nu(\cdot, \cdot) + \varepsilon(\cdot, \cdot) \). Note that the temperature observations can be seen to vary slowly in space at each time \( t \). Therefore,
for a location where no temperature measurements are taken, it is appropriate to use the temperature value taken at the nearest monitoring site for each time.

Before fitting the model, the cross-correlation function (CCF) is used for exploring the dynamic structure of \( \{Y(s, t)\} \) in time among different locations, defined as follows

\[
\rho(s, s', u) = \text{cor}(Y(s, t), Y(s', t + u)).
\]

These CCFs can be estimated by the sample CCFs of ozone residuals \( \{r(s, t)\} \). Figures 6(a) and 6(b) show the sample CCFs of ozone between two monitoring stations \( s \) (Shilin) and \( s' \) (Tamsui) under the wind speed greater than 2 (km/hour) and the wind direction within 20 degrees from \( s' \) to \( s \) and from \( s \) to \( s' \), respectively. Figure 6 shows that the sample CCF skews to the left when wind blows from \( s' \) to \( s \), and it skews to the right when wind blows from the opposite direction. A similar behavior can be found for most pairs of monitoring stations. This empirical evidence supports the importance of incorporating the transport effect into the space-time model for ozone concentrations.

As mentioned earlier, the transport patterns of ozone, characterized by the weight function, change according to the wind-speed field. Herein, the wind-speed vectors \( \{x(s, t)\} \) are only measured at ten monitoring stations. For locations with no wind speed observations, they are predicted from the observed data. Owing to that the observed wind-speed vectors are nearly the same for all available stations at each time \( t \), we estimate \( x(s, t) \) for all \( s \in D \) using the average of the observed wind-speed
vectors at $t$. Note that when spatial homogeneity of the wind-speed field is violated, a kriging method can be applied to predict the wind-speed field. For instance, east-westerly wind speeds $\{x_1(s,t)\}$ and north-southerly wind speeds $\{x_2(s,t)\}$ can be predicted separately using an autoregressive space-time model (Huang and Cressie, 1996), in which the optimal predictors for both $\{x_1(s,t)\}$ and $\{x_2(s,t)\}$ at any $(s,t)$ can be obtained by a fast space-time Kalman filter.

### 4.2 Estimation and Prediction

The space-time model given by (8) and (9) is now applied to the residual ozone data $\{r(s_i,t)\}$. We consider a square region with a size of about 32 kilometers, where the whole Taipei city and the twelve monitoring stations are included (see the square area in Figure 2). This region is divided into $32 \times 32$ longitude-latitude grid cells $\{D_k : k = 1, \ldots, 1024\}$ as the finest resolution so that each grid cell is about 1 kilometer in both east-westerly and north-southerly directions. Based on $\{D_k\}$, the orthonormal basis functions are defined as

$$e_k(s) \equiv I(s \in D_k), \quad k = 1, \ldots, 1024.$$

Next, $\{u_1, \ldots, u_N\}$ are selected based on the procedure described in Section 2.2, where $a(t)$ are first predicted at each time $t \in \{1, \ldots, 8760\}$ using ordinary kriging based on the residuals $r(t) \equiv Z(t) - \mu(t)$. The weighted-least-squares method (Cressie, 1985) is applied to fit an exponential variogram model based on the residuals $\{r(t) : t = 1, \ldots, 8760\}$. The empirical and the fitted variograms are shown in Figure 7. Figure 8 displays the first six eigenimages $u_1, \ldots, u_6$, obtained from the method described in Section 2.2. The first eigenimage corresponds to the overall mean level of ozone concentration, which is also related to the topographic effect of
the Taipei basin. Roughly speaking, the second eigenimage and the third eigenimage correspond to a northwest-southeasterly linear trend and a northeast-southwesterly linear trend, respectively. Note that any linear trend can be approximately produced by a linear combination of the first three eigenimages. The first three eigenimages account for 90.3% of the total spatial variation of the estimated spatial covariance matrix among \( N \) sites. Therefore, \( L = 3 \) is chosen in this study.

Finally, the exponential covariance family is specified for modeling the small-scale spatial process \( \nu(\cdot, \cdot) \), which satisfies

\[
\text{cov}(\nu(s, t), \nu(s', t)) = \sigma^2_\nu \exp\left(-\|s - s'||/\gamma_\nu\right); \quad t \in \mathbb{N}.
\]

The ML estimates of parameters are solved numerically in Splus using the “nlminb” function, in which the starting values are obtained from a two-step procedure described in the following. In the first step, the parameters \( b, \theta_0, \theta_1, \psi, \) and \( \sigma^2_\xi \) are estimated using the ML method based on a selected subset of ozone data \( \{Z(t_1), \ldots, Z(t_m)\} \), where \( x(\cdot, t_i) = x(\cdot, t_i - 1) = x(\cdot, t_i - 2) = 0; \ i = 1, \ldots, m \). This subset corresponds to the data under no wind condition and it is reasonable to assume that \( \text{var}(a_L^*(t_i)) \) are equal for \( i = 1, 2, \ldots, m \). Moreover, \( Z(t_1), \ldots, Z(t_m) \) can be regarded as independent since the resulting \( \{t_i\} \) are far apart from each other. In the second step, \( \theta \) is estimated using the ML method after plugging the estimated values of \( b, \psi, \) and \( \sigma^2_\xi \) into the likelihood function. The final ML estimates are given in the following:

\[
\hat{Q} = \begin{pmatrix}
154.33 & -8.88 & -0.45 \\
-8.88 & 25.47 & -2.13 \\
-0.45 & -2.13 & 13.36
\end{pmatrix},
\]

\[
\hat{\sigma}^2_\nu = 0.66, \ \hat{\gamma}_\nu = 1.91 \ (\text{km}), \ \hat{\sigma}^2_\varepsilon = 0.00, \ \text{and} \ \hat{\theta} = (0.93, 1.42, 0.39, 0.40)', \quad \text{The autore-}
\]
gressive parameter is estimated by $\hat{\theta}_0 = 0.93$, indicating strong temporal dependence. The practical range, namely 95% of the sill, of the $\nu(\cdot)$ process is about $3\hat{\gamma}_\nu = 5.73$ (km). The estimated weight functions for various wind speeds (0, 3, 6, 9 km/hour) under northeasterly wind are shown in Figure 1.

Exactly how the autoregressive matrices $\{C_t^*\}$ change with respect to different wind conditions after plugging in the estimated parameters is of interest. Modeling results indicate that when there is no wind, $\{C_t^*\}$ is almost a diagonal matrix. With either a northwesterly or southeasterly wind, the second diagonal element of $C_t^*$ becomes smaller. Correspondingly, with either a northeasterly or southwesterly wind, the third diagonal element of $C_t^*$ becomes smaller. As expected, the second eigenimage and the third eigenimage represent the northwest-southeasterly linear trend and northeast-southwesterly linear trend, which would be destroyed by northwest-southeasterly wind and northeast-southwesterly wind, respectively.

The optimal spatio-temporal predictor $\mu(s, t) + \hat{Y}(s, t) + \hat{\nu}(s, t)$ of ozone under the square-root-transformed scale for any location $s$ and time $t$ based on $\{Z(1), \ldots, Z(t)\}$ can be obtained from (12). The corresponding CMSPE, $V(s, t)$, is given in (13). Moreover, the optimal spatio-temporal predictor of ozone concentration $o\hat{z}(s, t)$ on the original scale is

$$o\hat{z}(s, t) \equiv E[(\mu(s, t) + Y(s, t) + \nu(s, t))^2|Z(1), \ldots, Z(t)]$$

$$= [\mu(s, t) + \hat{Y}(s, t) + \hat{\nu}(s, t)]^2 + V(s, t),$$

for any location $s$ and time $t$. The corresponding CMSPE is

$$E[(o\hat{z}(s, t) - o\hat{z}(s, t))^2|Z(1), \ldots, Z(t)]$$

$$= 4 [\mu(s, t) + \hat{Y}(s, t) + \hat{\nu}(s, t)]^2 V(s, t) + 2[V(s, t)]^2.$$

Carroll et al. (1997) provide details of these formulations.
4.3 Diagnostics

As a simple method for model diagnostics, the fitted CCFs are compared with their corresponding sample CCFs. Figures 9(a1-a3) and 9(b1-b3) show the fitted CCFs of ozone between \( s \) (Shilin) and \( s' \) (Tamsui) when wind directions are from \( s' \) to \( s \) and from \( s \) to \( s' \), respectively, under various wind speeds. Note that the average wind speed in Taipei during year 1999 is 6.02 km/hour. Comparing the fitted CCFs with the sample CCFs (Figure 6) reveals that although the fitted CCFs do not fit perfectly well with the sample ones, they do capture the asymmetry feature of cross correlation due to the transport effect.

Goodness of fit of our space-time model can be assessed by comparing the variation of the residual ozone series \( r(t) = Z(t) - \mu(t); \ t = 1, \ldots, 8760, \) with that of the one-step prediction-error series

\[
\delta(t) \equiv (\delta(s_1, t), \ldots, \delta(s_n, t))' \equiv r(t) - E(r(t) \mid r(1), \ldots, r(t - 1)); \ t = 1, \ldots, 8760,
\]

under the fitted model. First, the proportions of variation reduction given by

\[
\frac{\text{var}(r(s_i, t)) - \text{var}(\delta(s_i, t))}{\text{var}(r(s_i, t))}; \ i = 1, \ldots, 12,
\]

are examined. These values range from 45.9% to 72.6% with the average of 64.8% for the twelve monitoring sites. That is, about 64.8% of the variation of \( r(t) \) can be accounted for by the past information of \( r(1), \ldots, r(t - 1) \) under the fitted model. Next, the sample correlation matrix of \( r(t) \) is compared with that of \( \delta(t) \). Figure 10 shows the sample spatial correlations of \( r(t) \) (upper-left triangle) and those of \( \delta(t) \) (lower-right triangle) among the twelve monitoring sites. It is interesting to see that the sample spatial correlations of \( \delta(t) \) are much smaller than those of \( r(t) \).
with the averaged absolute correlation values for $r(t)$ and $\delta(t)$ being 0.68 and 0.23, respectively. The result indicates that large spatial correlation values of $r(t)$ can be mostly explained by the diffusion and the transport effects through the proposed space-time model.

Finally, whether our model is preferable over a simpler model with $\theta_2 = \theta_3 = 0$ (i.e., no wind effects) is examined. The AIC values for our model and the reduced model with $\theta_2 = \theta_3 = 0$ are 74174.6 and 74235.4, respectively, indicating the effectiveness of incorporating the transport effect into the model.

5 Discussion

This work develops a new space-time model for ground-level ozone, which accounts for the transport effect. The proposed model has a spatio-temporal covariance structure that depends on wind speed and wind direction. Moreover, it is computationally efficient and can handle large amounts of data. An illustrated example shows that the proposed model is more realistic than other space-time models in analyzing ground-level ozone data owing to its ability in capturing the transport feature previously neglected. We believe that the proposed method can be adopted in analyzing other environmental processes that are affected by wind.

Appendix

Proof of Proposition 1. It is clear that (2) can be written as:

$$
\sum_{k=1}^{N} a_k(t) e_k(s) = \sum_{k=1}^{N} \sum_{l=1}^{N} c_{k,l}(t-1)a_l(t-1)e_k(s) + \sum_{k=1}^{N} \xi_k(t)e_k(s); \quad s \in D, \ t \in \mathbb{N},
$$
which is equivalent to

$$a_k(t) = \sum_{l=1}^{N} c_{k,l}(t-1)a_l(t-1) + \xi_k(t); \quad k = 1, \ldots, N, \; t \in \mathbb{N},$$

since \( \{e_1(\cdot), \ldots, e_N(\cdot)\} \) are linearly independent. This completes the proof. \qed

References


Captions for Figures

Figure 1. Weight functions $w(s, u, x(s, t); \theta)$ under different values of wind speed $v(s, t)$ at $s = 0$. (a) $v(0, t) = 0$, (b) $v(0, t) = 3$, (c) $v(0, t) = 6$, and (d) $v(0, t) = 9$, where $\tau(0, t) = \pi/4$, $\theta_0 = 0.93$, $\theta_1 = 1.42$, $\theta_2 = 0.39$, and $\theta_3 = 0.40$.

Figure 2. Locations of twelve monitoring stations in Taipei.

Figure 3. Average of the square root of the hourly ozone concentrations over days for each hour and each station.

Figure 4. Average of the square root of the hourly ozone concentrations over days for each month and each station.

Figure 5. Boxplots of the square root of the hourly ozone concentrations with respect to temperature.

Figure 6. Sample cross correlation functions of ozone between Shilin and Tamsui under the wind direction from (a) Tamsui to Shilin; (b) Shilin to Tamsui.

Figure 7. Empirical variogram and fitted variogram based on the residual ozone values at twelve sites and 8760 time points.

Figure 8. First six eigenimages of the average spatial covariance matrix.

Figure 9. Estimated cross correlation functions of ozone between Shilin and Tamsui with wind speed = 0, 6, 12 (km/hour), respectively, under the wind direction from (a1-a3) Tamsui to Shilin; (b1-b3) Shilin to Tamsui.

Figure 10. Sample spatial correlations of the residual ozone series (shown in the upper-left triangle) and those of the one-step prediction errors (shown in the lower-right triangle) among the twelve monitoring stations.
Hsin-Cheng Huang, Figure 1
Hsin-Cheng Huang, Figure 2
Hsin-Cheng Huang, Figure 3
Hsin-Cheng Huang, Figure 4
Hsin-Cheng Huang, Figure 5
Hsin-Cheng Huang, Figure 6
Hsin-Cheng Huang, Figure 7
Hsin-Cheng Huang, Figure 8
Hsin-Cheng Huang, Figure 9
Hsin-Cheng Huang, Figure 10