Statistica Sinica **29** (2019), 1181-1207
doi:https://doi.org/10.5705/ss.202017.0561

SPATIO-TEMPORAL MODELS WITH SPACE-TIME INTERACTION AND THEIR APPLICATIONS TO AIR POLLUTION DATA

Soudeep Deb and Ruey S. Tsay

University of Chicago

Abstract: It is important to have a clear understanding of the status of air pollution and to provide forecasts and insights related to air quality to both the public and environmental researchers. Previous studies have shown that even a short-term exposure to high concentrations of atmospheric fine particulate matter can be hazardous to people's health. In this study, we develop a spatio-temporal model with space-time interaction for air pollution data ($PM_{2.5}$). Along with the spatial and temporal components, the proposed model uses a parametric space-time interaction component in the mean structure, as well as a random-effects component specified in the form of zero-mean spatio-temporal processes. To apply the model, we analyze air pollution data ($PM_{2.5}$) from 66 monitoring stations across Taiwan.

Key words and phrases: Dynamical dependence, fine particulate matter, Lagrange multiplier test, spatial dependence.

1. Introduction

The effects of air pollution on public health and vegetation, as well as on human society and various ecosystems in general, have become critical issues. Several epidemiological studies have established that particulate matter (PM) is linked to a range of serious cardiovascular, respiratory, and visibility problems. Detailed discussions can be found in Pope III et al. (1995), Jerrett et al. (2013), Blangiardo, Finazzi and Cameletti (2016), and Thurston et al. (2016). In 1997, taking the severe effects of PM into account, the US Environmental Protection Agency (EPA) provided new regulations that established National Ambient Air Quality Standards (NAAQS) for PM with an aerodynamic diameter of less than 2.5 microns. These standards are usually measured in units of micrograms per cubic meter (μgm^{-3}) (henceforth, denoted as $PM_{2.5}$). According to the NAAQS, the hourly average $PM_{2.5}$ concentration should not be higher than 35 μgm^{-3}. However, in practice, levels often exceed this level. Our study data are obtained from 66 monitoring stations in Taiwan, covering a period of 10 years (from 2006

to 2015), and have median $PM_{2.5}$ values slightly above 37 μgm^{-3}. For additional information, refer to Mayer (1999), who discussed how the air quality is deteriorating in various cities worldwide. Overall, there is a growing demand to identify the main factors that contribute to air pollution.

A brief discussion on PM is in order. In general, $PM_{2.5}$ contains particles either emitted directly or formed in the atmosphere from gaseous emissions. Examples include sulfates formed from sulfur dioxide (SO_2) emissions, nitrates formed from NOx emissions, and carbon formed from organic gas emissions. The rates of conversion of gases to particles are often reliant on regional and temporal factors, including the topography, land cover, and seasonal climatic variables. As a result, the $PM_{2.5}$ concentrations are also affected by these variables. This increases the need for a spatio-temporal model that can be used to assess air quality. A good model can provide better predictions, which will help to determine an efficient strategy to combat air pollution.

The spatial and spatio-temporal modeling of air pollutants began in the previous century. Elsom (1978) studied the spatial correlation fields for air pollution in an urban area. Furthermore, various geostatistical space-time models have been applied to examine the trends in deposits of atmospheric pollutants by Eynon and Switzer (1983), Bilonick (1985), Rouhani and Hall (1989), and Vyas and Christakos (1997), among others. Other earlier notable works in this regard include those of Guttorp, Meiring and Sampson (1994), Haas (1995), and Carroll et al. (1997). Most of these approaches rely on a spatio-temporal random field, where the spatial or temporal dependencies are incorporated in either the mean function or in the error process, and the parameters are estimated using frequentist procedures. For a good discussion on geostatistical space-time models, see Kyriakidis and Journel (1999).

In comparison, many 21st century studies on related problems use hierarchical Bayesian approaches for spatial predictions of air pollution. For example, Sun et al. (2000) analyzed the PM_{10} (PM with a diameter less than 10 μgm^{-3}) concentrations in Vancouver and developed posterior predictive distributions using Bayesian techniques. Kibria et al. (2002), on the other hand, used a multivariate setup to analyze $PM_{2.5}$ concentrations in Philadelphia, developing a spatial prediction methodology in a Bayesian context in order to do so. For a related problem, in order to predict PM_{10} concentrations in London, Shaddick and Wakefield (2002) proposed a short-term space-time modeling technique.

In a hierarchical Bayesian setting, Sahu and Mardia (2005) modeled the spatial structure using principal kriging functions and the time component using

a random walk process in order to present a short-term forecasting analysis of $PM_{2.5}$ data on New York City. There are two other notable works. Sahu, Gelfand and Holland (2006) used indicators for urban or rural sites to employ different spatio-temporal processes in the error structure when modeling the $PM_{2.5}$ series of several states in the US Midwest. In a later work, Sahu, Yip and Holland (2009) developed a space-time model that includes a spatially varying regression term and an autoregressive term in the mean structure to analyze the ozone concentrations in eastern states of the United States. Berrocal, Gelfand and Holland (2010) extended their earlier work to introduce a bivariate downscaler and, thus, provide a flexible class of space-time assimilation models. In addition, Cameletti, Ignaccolo and Bande (2011) compare available space-time models using data from Piemonte, Italy.

In this study, we develop a new spatio-temporal model in order to determine whether there is any space-time interaction in the behavior of $PM_{2.5}$ concentrations. Here, an interaction means that the temporal trends of pollution are more similar for sites that are closer together on a spatial scale. An early study in this regard is that of Wikle, Berliner and Cressie (1998), who developed a hierarchical Bayesian model that allows for an interaction in space and time. However, with the exception of their study and a few other related works, there has been little effort to quantify the space-time interactions in air pollution data, although numerous studies have examined spatio-temporal modeling on the same topic, as already discussed.

Note that the problem of identifying a space-time interaction is not specific to air pollution data, and has been studied in several other fields, including seismology, epidemiology, criminology, and transportation research. Meyer et al. (2016) offer a good discussion on tests for space-time interactions in problems related to medical studies. Here, the most popular techniques are the Knox test, Mantel test, and space-time K-function analysis. These tests have test statistics of the form $T = \sum_{j \neq i} a_{ij}^s a_{ij}^t$, where a_{ij}^s and a_{ij}^t are measures of the spatial and temporal adjacency, respectively, of events i and j. For further information on space-time interactions and related problems from other fields, refer to Kulldorff and Hjalmars (1999), Legendre, Cáceres and Borcard (2010), and Vanem, Huseby and Natvig (2014), and the references therein.

Note that most of the studies described here focus on pollution issues in different parts of Canada and the United States, despite this being a global problem. As a result, few studies have analyzed air pollution data from countries

in Asia or Africa. A noteworthy exception is the work of Al-Awadhi and Al-Awadhi (2006), who used hierarchical Bayesian approaches to develop dynamic linear models for air pollutants in Kuwait. They dealt with the temporal and spatial effects independently, defining separate structures for the two processes. In general, there is a dearth of research on pollution in Asia and Africa. Thus, by focusing on Taiwan, we contribute to the literature in this regard.

The rest of the paper is organized as follows. Section 2 provides an exploratory analysis of the data under study. The proposed model, its properties, and related results are described in Section 3. Section 4 discusses the results of our data analysis. Section 5 provides concluding remarks and discusses possible future work.

2. Preliminary Analysis

2.1. Data

Our $PM_{2.5}$ data were collected from 71 official monitoring stations across Taiwan. However, for five of those stations, we were not able to obtain the information necessary for the covariates considered here; thus, we excluded these stations from the study. The remaining 66 monitoring stations are irregularly located in space, spread across the country with some concentrations around big cities or industrial areas. The minimum and maximum pairwise distances of these stations are 0.58 km and 366.7 km, respectively; the arithmatic mean is 140.7 km. A major feature of our study is that we consider a space-time interaction as a property of a region, rather than that of an individual station. Thus, we divide the data into several clusters, based on the latitude and longitude of the stations. Figure 1 shows the locations of the 66 stations. The concentration of the stations on the west coast is understandable because it is the most populated area of Taiwan. The colors on the map indicate different regions (clusters) in our study.

Our temporal data are obtained on an hourly basis for a period of 10 years, from January 1, 2006, to December 31, 2015. However, the sampling frequencies vary across stations and sites, and some values are missing. Moreover, because we are mainly interested in identifying space-time interactions in the air pollution data, a lower temporal resolution is both desirable and simpler. Thus, in our analysis, we aggregate the hourly data into weekly averages based on all available measurements within a week. Note that this is a common practice when working with monitoring data, as discussed in Smith, Kolenikov and Cox (2003). In order

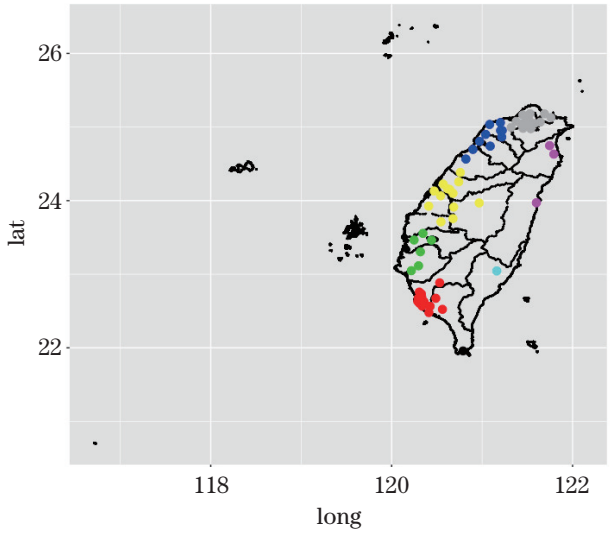

Figure 1. The locations and clustering of the 66 monitoring stations in Taiwan.

to maintain continuity in the time scale, we consider the whole set of 3,652 days (from 2006 to 2015), which we divide into 522 weeks. Hence, the total number of data points considered in this study is $66 \times 522 = 34{,}452$.

To begin, we present an exploratory analysis on the data. As in several related studies, we convert the $PM_{2.5}$ values to a square root scale and base our study on the transformed data. This type of transformation is common when working with air pollution data, as discussed in Smith, Kolenikov and Cox (2003). Time series plots of the transformed $PM_{2.5}$ values for three stations are displayed in Figure 2.

Next, we decompose the time series data for Xinying into three components: seasonal, trend, and the residuals. It is evident (see Figure 3) that there is a decreasing trend, which appears to be a linear component with a small slope.

The graph in Figure 4 shows the overall means and variances of the transformed $PM_{2.5}$ observations for various stations and months. The top, left panel describes the variances against the means of the stations. The top, right panel shows the same, but for different months. The bottom two plots show the behavior of the means and variances, respectively, corresponding to different months.

From the top two plots, it is clear that the variance increases in a nonlinear manner as the mean increases. On the other hand, the bottom, left plot establishes that there is seasonality in the data, which is expected for most related time series problems. Moreover, interestingly, the bottom, right plot of the variance

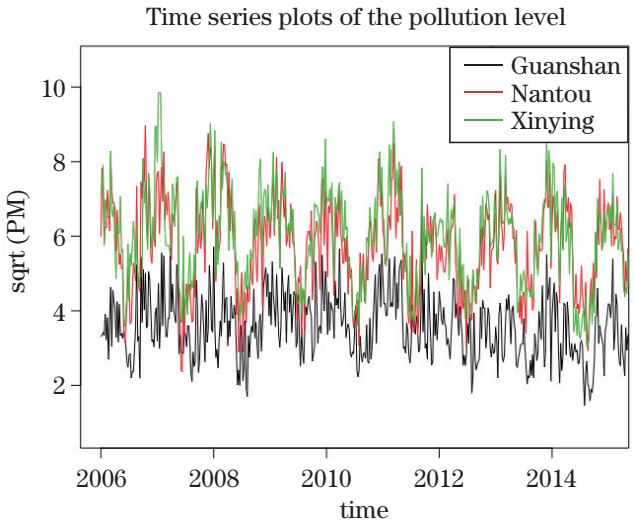

Figure 2. Time series plots of $\sqrt{\text{PM}_{2.5}}$ observations for three stations: Guanshan, Nantou, and Xinying.

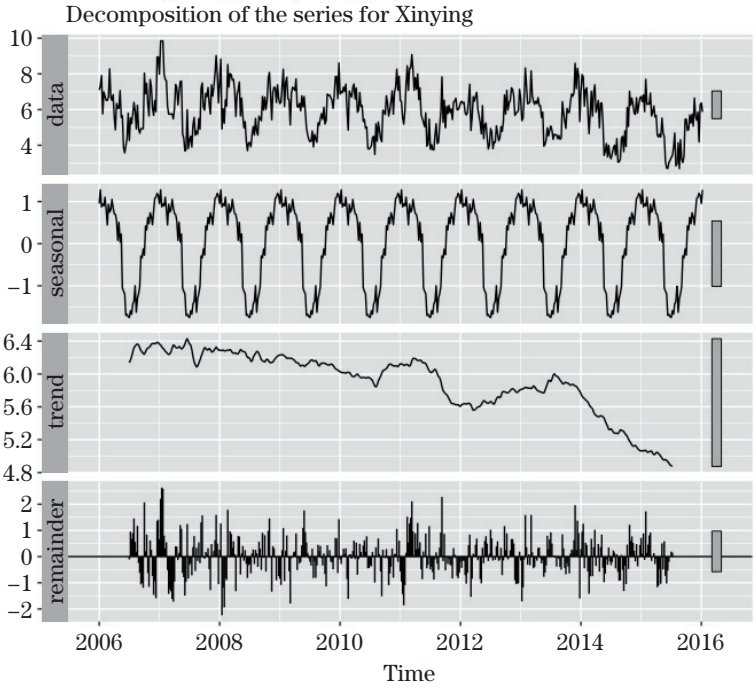

Figure 3. Decomposition of the $\sqrt{\text{PM}_{2.5}}$ observations for Xinying. The four panels show the original series, seasonality, trend, and residuals, respectively.

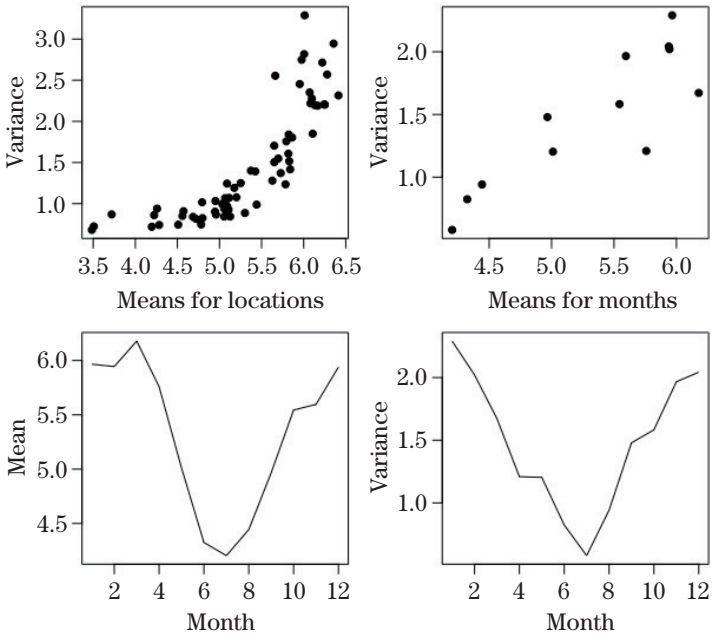

Figure 4. (Top left) Sample variances versus means of the weekly $\sqrt{\text{PM}_{2.5}}$ observations for 66 stations; (Top right) means versus variances of the $\sqrt{\text{PM}_{2.5}}$ observations for 12 months; (Bottom left) means of the $\sqrt{\text{PM}_{2.5}}$ observations corresponding to different months; (Bottom right) variances of the $\sqrt{\text{PM}_{2.5}}$ observations corresponding to different months.

corresponding to the months suggests that the variance is not homoskedastic. In fact, it varies seasonally, motivating us to consider the heteroskedastic nature of the error variance in the Gaussian process of the proposed model.

We also show a heat map (Figure 5) of the weekly average of the $\text{PM}_{2.5}$ observations across all locations. In the map, we show the averages per season. It is evident that the spatial pattern of the weekly averages changes according to the season, motivating us to consider a space-time interaction coefficient in the model, as described in Section 3.

2.2. Covariates used in the study

For each station, along with the observations of $\text{PM}_{2.5}$, data were collected on temperature, relative humidity, and wind speed and direction. Similarly to the air pollution observations, these data were collected on an hourly basis for the period January 1, 2006, to December 31, 2015. However, we aggregate the values per week and use these values as covariates in our study.

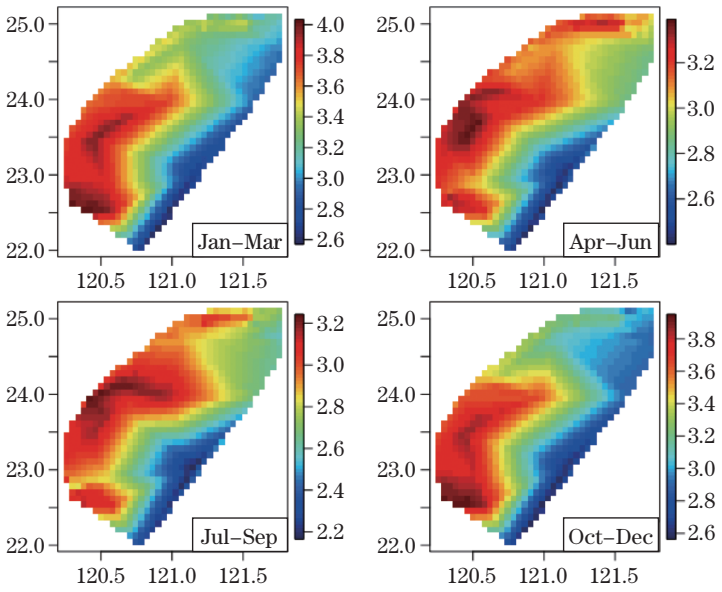

Figure 5. Seasonal heat map of the $PM_{2.5}$ levels in Taiwan.

First, we examine the relationships between the pollution data and relative humidity and temperature. Our exploratory analysis suggests that air pollution is not significantly affected by temperature, but is more dependent on humidity, where there is a slight decrease in pollution with an increase in relative humidity. The Spearman correlation coefficients between the transformed $PM_{2.5}$ and humidity and temperature are -0.326 and -0.254, respectively. Relevant plots are provided in the online Supplementary Material.

Wind speed is another variable that may affect air pollution. The data on wind speed and wind direction were available on an hourly basis. The behavior of the wind speed and direction varies widely by location as shown in the box plots in Figure 6. These plots show clearly that the wind speed range varies markedly by location, whereas there is little variation in the mean. Naturally, in order to account for the effect of the wind, unlike the previous covariates, we cannot employ the weekly average. Instead, we use the maximum daily wind speed for each week. Furthermore, for most weeks, the maximum wind speed occurred on the same day in all locations, justifying our choice of the maximum rather than the average as a measure. The sample correlation coefficient between the transformed $PM_{2.5}$ and the maximum wind speed is -0.146.

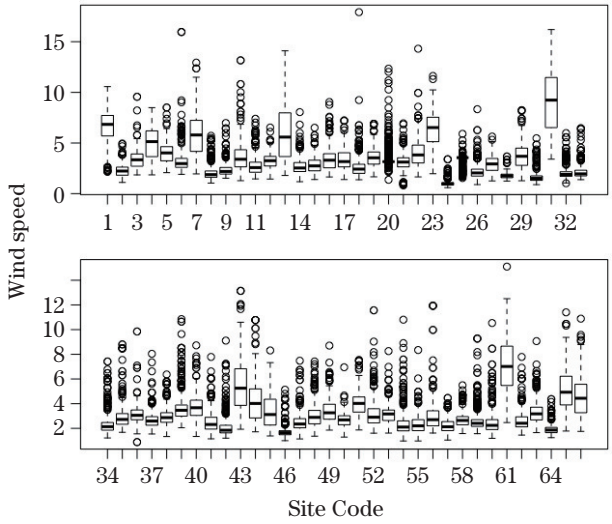

Figure 6. Box plots of the wind speed for 66 stations.

3. Methods

3.1. Covariates used in the study

The proposed model is inspired by that of Sahu, Gelfand and Holland (2006). The key feature of our model is that it accounts for possible space-time interactions in air pollution data. We describe the model in a general setup, and then discuss the Taiwan data in Section 4.2.

Suppose the data are collected for n locations and over T consecutive time points. Let us denote the square root of the $PM_{2.5}$ at time t_j and location s_i by $Z(s_i, t_j)$, for $i = 1, \ldots, n$ and $j = 1, \ldots, T$. Then, we assume that the overall mean value varies by location, and we subtract the location-wise means from the actual transformed $PM_{2.5}$ values to obtain the mean-adjusted values, which we denote by $Y(s_i, t_j)$. For convenience, we drop the subscripts whenever they are not needed. In the proposed model, we consider the following hierarchical structure:

$$Y(s,t) := U(s,t) + \epsilon(s,t), \tag{3.1}$$

where $U(s,t)$ describes a spatio-temporal process, and $\epsilon(s,t)$ denotes a white noise process that accounts for measurement errors and other variability not addressed by $U(s,t)$. We assume the white noise process follows a heteroskedastic $N(0, \sigma_t^2)$ distribution independently, where σ_t^2 is chosen to reflect a season. On the other hand, $U(s,t)$ assumes the following structure:

$$U(s,t) := \mu(s,t) + v(s,t), \tag{3.2}$$

where $\mu(s,t)$ is the mean of the $U(s,t)$ process, and $v(s,t)$ denotes a zero-mean spatio-temporal process.

The aforementioned mean function is an additive combination of the effects of the covariates, the seasonal effects, and a spatio-temporal interaction effect. To capture the effects of the available covariate information, we consider a term of the form $B\alpha$, where B is the design matrix for p covariates. The population density, temperature, humidity, number of factories, number of cars, and so on, can be taken as covariates for the analysis. The seasonal variation is captured by introducing indicators for different seasons. Throughout this study, we consider monthly indicators, but the method can be applied to other cases too. In general, let J be the number of seasons in a year.

The spatio-temporal effect used in the mean structure allows us to test for space-time interaction, if any. In principle, the effect is described as a linear function of the form $\gamma_0 + \gamma_s t$, thereby allowing us to test whether the γ_s values are the same across locations. However, this would increase the complexity and computational burden significantly when there are many sites. To address this issue, we make a logical assumption that the coefficients γ_s for locations that are close to each other are the same, and that it should be a characteristic of a region rather than a station. Thus, based on their coordinates, we divide the stations into clusters, and employ γ_k for all stations in the kth cluster. The number of clusters used is based on the number of locations we have. If n, the number of locations, is five or less, we do not use clustering. For $n > 5$, we use $[\sqrt{n}]$ clusters ($[n]$ denotes the integer part of n) to divide the locations into regions. In practice, we use the k-means clustering method, based on the latitude and longitude, for this purpose.

Furthermore, we use an intercept-free model; thus, along with the response variable, we subtract the overall means of the covariates from the respective observations. Therefore, the mean structure of the process can be described as:

$$\mu(s,t) = c(s,t)'\alpha + \sum_{j=2}^{J} \beta_j \; m(t,j) + \sum_{k=1}^{K} \gamma_k \; r(s,k)t. \tag{3.3}$$

In the above, $c(s,t)$ is a column vector containing the values of the covariates, and $m(t,j)$ and $r(s,k)$ are indicators for the season and location region, respectively, where J and K are the seasonality and the number of regions, respectively. That is, $m(t,j) = 1$ if time t is in the jth season, and zero otherwise. Similarly, $r(s,k) = 1$ if location s is in the kth region, and zero otherwise. For identifiability

purposes, we take $\beta_1 = 0$. Note that because we use mean-adjusted values for the response and the covariates, we do not have an intercept term in the model. On the other hand, we scale down the time points (t) to equi-spaced points in the interval $[0, 1]$.

The term $v(s,t)$ in Equation (3.2) can be treated as a spatially varying temporal trend. By averaging over different sites in a region, we obtain the adjustment to the regional trends; averaging over time helps us to obtain the adjustment at the temporal scale. For convenience, we consider a separable structure for the covariance of this process. Moreover, we assume that the locations in different clusters are independent of each other. In particular, the covariance between $v(s_1,t_1)$ and $v(s_2,t_2)$ when s_1, s_2 are in the same cluster is taken as the product of the spatial dependence and the temporal dependence, which we express as

$$\mathrm{Cov}(v(s_1,t_1), v(s_2,t_2)) = \sigma_v^2 \; \rho(\|s_1 - s_2\|, \phi_s) \cdot \rho(\|t_1 - t_2\|, \phi_t) \cdot \mathbb{I}\{s_1, s_2 \in \mathcal{C}_k\}, \tag{3.4}$$

where $\rho(x,d)$ denotes the exponential covariance function e^{-dx}, and $\mathcal{C}_k$ denotes a particular cluster. The distance functions $\|s_1 - s_2\|$ or $\|t_1 - t_2\|$ are taken as the Euclidean distance between two points.

Throughout the paper, Y denotes the vector of $N = nT$ data points, arranged first by cluster, then by time points and sites. Thus, if $\{s_1, s_2\}$ form a cluster, then the first few observations will be $Y(s_1,t_1), Y(s_2,t_1), Y(s_1,t_2), Y(s_2,t_2)$, and so on. The vector $v = (v(s_i,t_j))$ is formed in a similar way. We denote the full covariance matrix of v by Σ_v. Note that Σ_v is a block diagonal matrix, where each block corresponds to the covariance matrix of a cluster of locations, taking the form $\sigma_v^2(\Sigma_t \otimes \Sigma_s)$, such that $\Sigma_t(i,j) = \rho(\|t_i - t_j\|, \phi_t)$ and $\Sigma_s(i,j) = \rho(\|s_i - s_j\|, \phi_s)$. Furthermore, note that when the sampling design considers equally spaced time points, the common spacing being d_t, we can write $\Sigma_t(i,j) = \psi^{|i-j|}$, where $\psi = e^{-\phi_t d_t}$.

On the other hand, as mentioned before, we entertain a heteroskedastic error function for $\epsilon(s,t)$. The exploratory analysis suggests that the variances differ by season; thus, we assume $\epsilon(s,t) \sim N(0, \sigma^2_{m(t)})$, where $m(t)$ denotes the season that applies to time t. If we use ϵ to denote the vector of $\epsilon(s,t)$, arranged similarly to Y and v, then the above discussion implies that $\epsilon \sim N(0, \sigma^2 D)$, where D is a diagonal matrix, such that the diagonal element corresponding to $\epsilon(s,t)$ is $\sigma^2_{m(t)}/\sigma^2$. We denote these parameters by $\tau_1^2, \ldots, \tau_J^2$, where τ_i^2 is the variance parameter associated with the ith season. In addition, we set $\tau_1^2 = 1$ to avoid any potential identifiability problem.

Now, to write the full model in vector-matrix notation, recall that Y is the observed vector of the dependent variable, and v and ϵ denote the corresponding vectors of the zero-mean spatio-temporal process and the Gaussian error process, respectively. We can write the mean function as the sum of $c(s,t)'\alpha$, $m_t'(\beta_j)$, and $r_{st}'(\gamma_k)$, where m_t and r_{st} are the column vectors corresponding to the parameter vectors $(\beta_j)_{2\leq j\leq J}$ and $(\gamma_k)_{1\leq k\leq K}$, respectively. Then, denoting the vector of all the parameters by θ, and letting X be a design matrix such that each row is of the form $X(s,t)' = (c(s,t)', m_t', r_{st}')$, the model can be written as:

$$Y = X\theta + v + \epsilon, \tag{3.5}$$

where $v \sim N(0, \sigma_v^2\Sigma_v)$ and $\epsilon \sim N(0, \sigma^2 D)$. It is evident that there are $(p+J+K)$ components in the parameter vector θ if we consider $p+1$ covariates and K regions for the locations. On the other hand, we write the two variance components σ^2 and σ_v^2 as being equal. Here, the estimate of σ^2 suggests the variance explained by the spatio-temporal process, and the estimates of the diagonal elements of D tell us how much is explained by the pure error process.

Finally, the best estimates for ϕ_s and ϕ_t are obtained using a cross-validation scheme. The validation scheme considers predictions for the sites at some time points, and obtains the mean squared error for those predictions. In this study, for every site, we take the first 90% of the time points under consideration (e.g., in case of weekly data for 10 years, we consider the first 9 years or 470 weeks) for model fitting, and then make predictions for the last 10% (52 weeks for the aforementioned data) to determine which combination of (ϕ_s, ϕ_t) works best. The possible choices for ϕ_s in the study were $(0.001, 0.005, 0.01, 0.025, 0.05, 0.075, 0.1, 0.25, 0.5, 0.75)$, and the choices for ϕ_t were $(0.25, 0.5, 0.75, 1, 1.25, 1.5)$. To determine the optimal choice, we identify the combination with the smallest mean squared error of the predictions, which is calculated as follows. For each site s_i, denote the validation time points by $t_1, \ldots, t_b$ and the predictions by $\hat{Y}(s_i, t_j)$, for $i = 1, \ldots, n$; $j = 1, \ldots, b$. Then, for each of the $5 \times 4 = 20$ combinations, the prediction mean squared error is computed as:

$$\text{MSE} = \frac{1}{nb}\sum_{i=1}^{n}\sum_{j=1}^{b}\{Y(s_i,t_j) - \hat{Y}(s_i,t_j)\}^2. \tag{3.6}$$

The prediction procedure is discussed in Section 3.3.

3.2. Testing for interaction and parameter estimation

A two-stage procedure is adopted in our method to ensure better results. First, we use an appropriate hypothesis-testing technique to detect whether there

is a space-time interaction present in the data. Following that, an iterative estimation method is executed to identify the effect of the factors that lead to a higher pollution level.

Test for interaction:

The model described in Equation (3.5) allows us to test for a space-time interaction. The model includes no interaction term if $\gamma_1 = \cdots = \gamma_K = 0$. We employ the Lagrange multiplier (LM) test for this purpose. If $\hat{\theta}_0$ denotes the maximum likelihood estimate of θ under the null hypothesis (the above condition), then the LM test statistic is defined by $U'(\hat{\theta}_0)I^{-1}(\hat{\theta}_0)U(\hat{\theta}_0)$, where $U(\cdot)$ is the first derivative of the log-likelihood and $I(\cdot)$ is the information matrix.

Recall that the main advantage of the LM test or the score test is that, unlike the Wald test or the likelihood ratio test, it does not require an estimate of the information under the alternative hypothesis or an unconstrained maximum likelihood. The LM test uses only the assumptions in the null hypothesis to obtain the maximum likelihood estimates. Then, it calculates the value of the test statistic (which follows a chi-squared distribution with appropriate degrees of freedom) to make a decision.

Once we perform the LM test, we can decide which model to use. Here, we use the full model (3.5) if the decision is to reject the null hypothesis that there is no space-time interaction.

Parameter estimation:

Next, to estimate the parameters, we employ the generalized least squares techniques, with some modifications. Observe that the proposed model can be expressed as $Y = X\theta + \varepsilon$, where $\varepsilon \sim N(0, \sigma^2\Omega)$, such that $\Omega = \Sigma_v + D$. This is in the setup of a generalized least squares problem. Furthermore, note that the number of unknown parameters in the error covariance matrix Ω is $J+2$, namely $\sigma^2, \tau_2^2, \ldots, \tau_J^2; \phi_s, \phi_t$. Because we obtain the optimal choices for the last two parameters using a cross-validation procedure, we need to estimate J additional variance parameters from the model.

From a practical point of view, it is more important to identify those cases where air pollution is hazardous to people's health and to the environment, rather than those that are less harmful. According to the standards set by the EPA, the average for $PM_{2.5}$ should not exceed 35 μgm^{-3}. Thus, while minimizing the sum of the squared residuals, we place more weight on those where the actual $PM_{2.5}$ values are greater than 35. Thus, the loss function we need to minimize

is of the form

$$L = \sum_{i=1}^{n} \sum_{j=1}^{T} w(s_i, t_j) \hat{e}(s_i, t_j)^2. \tag{3.7}$$

Here, $w(s_i, t_j)$ should be higher for $Z(s_i, t_j) \geq \sqrt{35}$. We take it to be $(1 + 2/\log N)$ for $Z(s_i, t_j) \geq \sqrt{35}$, and $(1 - 2/\log N)$ otherwise, where $N = nT$. An attractive feature of this method is that the weights approach one as $N \to \infty$; thus, the importance will be approximately equal on all observations for a very large sample size. On the other hand, $\hat{e}(s_i, t_j)$ is the standardized residual for location s_i and time t_j. Note that it can be written as $\hat{\Omega}^{-1/2}\hat{\varepsilon}/\hat{\sigma} = \hat{\Omega}^{-1/2}(Y - X\hat{\theta})/\hat{\sigma}$. Thus, assuming that $\hat{\Omega}$ and $\hat{\sigma}$ are known, we can say that minimizing L is equivalent to minimizing $(Y - X\theta)'\hat{\Omega}^{-1/2}W\hat{\Omega}^{-1/2}(Y - X\theta)$ with respect to θ. Here, W is a diagonal matrix, with the diagonal elements being the same as $w(s_i, t_j)$. It is evident that the minimizer is $\hat{\theta} = (X^T V X)^{-1}(X^T V Y)$, where $V = \hat{\Omega}^{-1/2}W\hat{\Omega}^{-1/2}$. In light of the above discussion, we propose the following procedure to estimate the parameters for the model:

1. Set $\hat{\tau}_j^2 = 1$ for $j = 1, \ldots, J$.
2. Evaluate $\hat{\Omega}$ and $V = \hat{\Omega}^{-1/2}W\hat{\Omega}^{-1/2}$.
3. Compute $\hat{\theta} = (X^T V X)^{-1}(X^T V Y)$ and set $\hat{\varepsilon} = Y - X\hat{\theta}$.
4. Compute $\hat{\sigma}^2 = \hat{\varepsilon}^T\hat{\Omega}^{-1}\hat{\varepsilon}/N$, where $N = nT$ is the total number of observations.
5. Let $\hat{\varepsilon}_j$ be the error corresponding to the jth season, for $j = 1, \ldots, J$.
6. For $j = 2, \ldots, J$, note that $\varepsilon_j \sim N(0, \sigma^2(\Sigma_v^{(j)} + \tau_j^2 I))$, where $\Sigma_v^{(j)}$ is the submatrix of Σ_v corresponding to the jth month.
7. Use optimization methods to compute the MLE $\hat{\tau}_j^2$ using the above.
8. Repeat steps 2 to 7 until convergence.

In the above procedure, in order to reduce the computational burden, we exploit the block-diagonal structure of the matrix Ω. Note that if each block of Ω is denoted by B_i, then $\Omega^{-1/2}$ can be written as a block diagonal matrix, where the blocks are $B_i^{-1/2}$. Using this, and writing W and X following the above steps, we substantially reduce the computational burden in the estimation.

The following theorem describes the asymptotic properties of the above estimators. The proof of the theorem is provided in the online Supplementary Material.

Theorem 1. *The estimate $\hat{\theta}$ obtained from the above procedure is consistent, in the sense that, as the number of locations (n) and the number of time points (T) approach infinity, $\hat{\theta} \to \theta$, in probability. Furthermore, $\sqrt{nT}(\hat{\theta} - \theta) \to N(0, \sigma^2 Q^{-1})$, where Q is the limit of $(X'\Omega^{-1}X)/nT$ as $n, T \to \infty$.*

3.3. Prediction

To make a new prediction for site s' at time t', we use the parameter estimates obtained from the method mentioned in the previous section. Let $X(s', t')$ denote the new set of covariate vectors, similarly to Section 3.1. Then, $\hat{Y}(s', t') = X(s', t')'\theta + \varepsilon(s', t')$. Thus far, to assign the clusters, we use techniques similar to the one-nearest-neighbor technique in k-means clustering; that is, the location s' is assigned to the cluster it is closest to in terms of the distance from the center.

Next, in view of the fact that Ω is not a constant multiple of the identity matrix, we cannot simply assume that the prediction error is going to be independent of the sample disturbances. The prediction procedure has to take the dependence into account.

If ε is the error vector corresponding to the original data, let us denote $\mathrm{Cov}[\varepsilon(s', t'), \varepsilon]$ by η, which is an nT-dimensional column vector. It is evident that η and Ω depend on the values of ϕ_s, ϕ_t, σ^2, and τ_j^2, for $j = 1, \ldots, J$. We use the estimates of these parameters to obtain $\hat{\eta}$ and $\hat{\Omega}$. Then, following Goldberger (1962), we can say that the best linear unbiased predictor is

$$\hat{Y}(s', t') = X(s', t')'\hat{\theta} + \frac{1}{\hat{\sigma}^2} \cdot \hat{\eta}'\hat{\Omega}^{-1}(Y - X\hat{\theta}). \tag{3.8}$$

Following the above, we add the mean adjustment (recall that we are using an intercept-free approach) using the historical overall mean of the station. In the case of a new location, we use the overall mean of the sites from the cluster to which it is assigned. Finally, we transform the predictions to obtain the estimate of the actual air pollution measurement.

To obtain a prediction interval, we use the asymptotic results for the distribution of $\hat{\theta}$ (cf. Theorem 1). Assuming that the variance components are known, note that the asymptotic variance of $\hat{Y}(s', t') - Y(s', t')$ is of the form $(nT)^{-1}\sigma^2\zeta' Q^{-1}\zeta$, where Q is the limit of $(X'\Omega^{-1}X)/nT$ and $\zeta = X(s', t') - X'(\sigma^2\Omega)^{-1}\eta$. Then, using the estimates obtained from our method, the squared standard error of the prediction can be estimated as $\hat{\mathrm{SE}} = \hat{\sigma}^2\hat{\zeta}'(X'\hat{\Omega}^{-1}X)^{-1}\hat{\zeta} + \hat{\sigma}^2/m$. The second term ($m$ is an appropriate constant) is added because of the mean adjustment, which is done using the overall mean estimate for the existing sites in the data. Hence, if $\hat{Z}$ denotes the mean-adjusted predictions, we can

evaluate the $100(1-\alpha)\%$ prediction interval as $(\hat{Z}(s',t') \pm z_{\alpha/2} \times \hat{\text{SE}})^2$. Here, z_α denotes the $100\alpha\%$ upper quantile for the standard normal distribution.

4. Detailed Analysis

4.1. Simulation studies

To begin, we present simulation studies to show that our methods are capable of capturing a space-time interaction that might be present in our empirical application of interest.

We perform the simulation experiment in three stages, first with a linear space-time interaction term, and then with two nonlinear types of interactions. Throughout this study, we consider the weekly average of the air pollution data. Then, for various time intervals, we evaluated the type-I error and the power for different values of n. In the simulation study, we considered three different values of T: 52 (1 year), 104 (2 years), and 261 (5 years). We worked with different numbers of locations ($n = 10, 20, 50$) to understand how the proposed method performs as the amount of data increases. For n locations, the xy-coordinates were generated randomly from a $(0, n^2) \times (0, n^2)$ grid. The values of the variance parameters σ^2, τ_j^2 (see Section 3.1) were generated from an inverse-gamma distribution with parameters $(4, 3)$. All coefficients in the model were simulated from independent normal distributions with mean zero and standard deviation one. Finally, we used $\phi_s = 0.1, \phi_t = 0.75$ for the covariance matrix of the spatio-temporal process $w(s,t)$.

For the case of a linear space-time interaction, the observations (square root of the $PM_{2.5}$ data) were obtained from model (3.5). However, for the first nonlinear case, we replaced the linear term with the quadratic term $(t/T)^2$, and assumed that the space-time interaction coefficients are negative. This type of interaction is common, and usually the coefficients are not big in magnitude. We wanted to observe whether a linear approximation is capable of capturing this type of interaction as well. Then, we consider another nonlinear case, where the interaction term is $\sin(2\pi t/T)$. This example addresses cases where the interaction can be an oscillatory function of time. The results for the three scenarios are displayed in Table 1, based on 500 iterations of the simulation.

While the type-I error remains under control across all scenarios, the power improves significantly for 20 or more locations. In addition, this is true for both the linear dependence and the nonlinear dependence. The power also increases with the sample size in all cases. The results confirm that the proposed testing

Table 1. Results when the time-dependence is linear, quadratic, or oscillatory. 5% critical values and 500 iterations are used.

		Type I error			Power		
Interaction	Number of locations	10	20	50	10	20	50
Linear	Weekly (1 year)	0.052	0.050	0.048	0.224	0.512	0.786
	Weekly (2 years)	0.056	0.058	0.042	0.312	0.578	0.820
	Weekly (5 years)	0.054	0.044	0.046	0.428	0.702	0.868
Quadratic	Weekly (1 year)	0.058	0.046	0.046	0.232	0.394	0.678
	Weekly (2 years)	0.048	0.042	0.062	0.268	0.478	0.714
	Weekly (5 years)	0.058	0.050	0.066	0.358	0.500	0.794
Oscillatory	Weekly (1 year)	0.062	0.028	0.032	0.440	0.686	0.900
	Weekly (2 years)	0.032	0.056	0.050	0.586	0.820	0.902
	Weekly (5 years)	0.038	0.040	0.040	0.542	0.708	0.960

procedure does not have large size distortions and, as expected, fares well for larger sample sizes.

We further extended the simulation studies to see how the proposed method behaves for larger data. To this end, we concentrated on weekly data from three years (157 observations for each site) and used different values for n, ranging from 10 to 50. The results of 500 iterations for various cases are shown in Figure 7. It is evident that, when the dependence on time is linear, the proposed testing procedure fares well for a large number of sites. The procedure also works well for the nonlinear case as its power increases, reaching more than 0.8 when the number of locations is 50.

Next, we show that the parameter estimation process works well in terms of identifying the effect of the factors. In this example, we used 20 different locations, two covariates (humidity and temperature), and simulated data from our model to obtain weekly averages for five years (divided into 12 seasons). Thus, the number of observations in the data were $20 \times 261 = 5{,}220$. We then estimated the parameters using the procedure described in Section 3.2. In Figure 8, the true values and the estimates of all the parameters in the model are plotted. We can see that most points lie along the line $y = x$ (displayed in the figure), showing that the estimates are not too different from the true values. This confirms that the proposed iterative estimation procedure is reliable.

Finally, to illustrate that our method works well in predicting the pollution level, we simulate data for 20 stations for six years. Then, we use data for five years from 19 stations to train our model. Then, we use the model to make predictions for the sixth year of the 20th station. This is displayed in

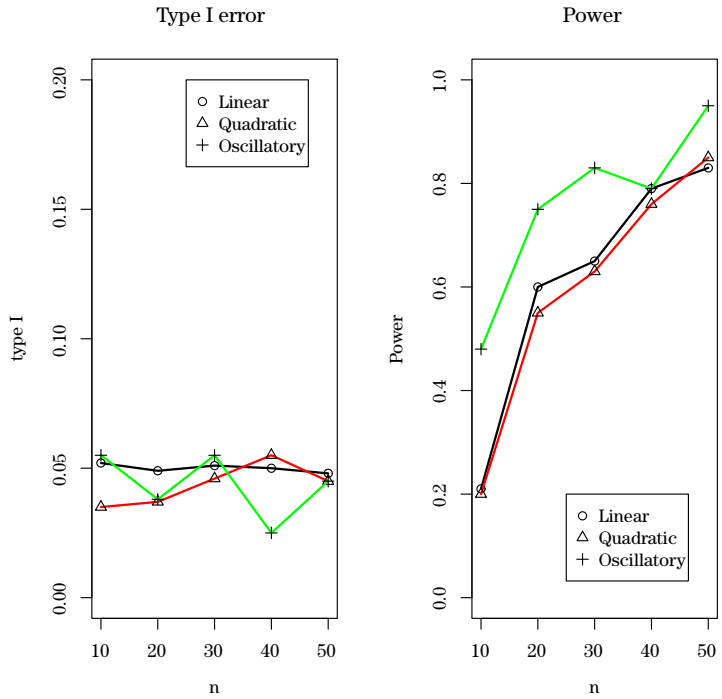

Figure 7. Type-I error (left) and power (right) for a linear (black), quadratic (red), and sinusoidal (green) interaction term, corresponding to different number of locations. $T = 157$ (three years) and 500 iterations are used for all cases.

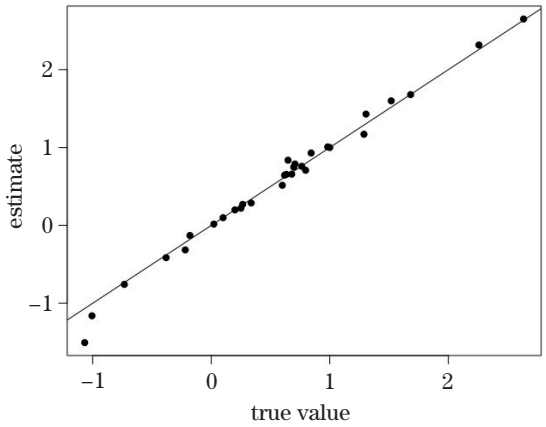

Figure 8. True and estimated values for the parameters of the model, where data are generated for 20 locations and five years.

Figure 9. In general, the forecasts are close to the original values. However, it is more interesting to note that the predictions are usually higher than the original

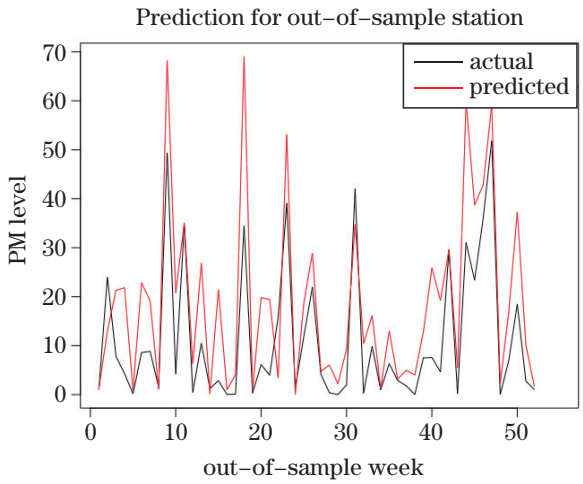

Figure 9. True value and predicted $PM_{2.5}$ observations for one station and one year. The model was trained using data generated for 19 locations and five years.

values whenever the actual values are above 35 μgm^{-3}. Recall from our earlier discussion that the method is aimed at finding significant factors and using their contribution to predict higher pollution levels. We can see that, even with 20 stations and five years, the model performs well.

4.2. Model selection

The first task when implementing the proposed method is to choose proper decay parameters and to identify whether any space-time interaction exists. For the decay parameters, we searched a two-dimensional array to determine which combination gives the least mean-squared error (refer to Equation (3.6)). The choices for ϕ_s were $(0.001, 0.005, 0.01, 0.025, 0.05, 0.075, 0.1, 0.25, 0.5, 0.75)$, and those for ϕ_t were $(0.25, 0.5, 0.75, 1, 1.25, 1.5)$. We used nine years and 64 stations to train the model; the rest were used for validation purposes.

We found that the combination of $\phi_s = 0.05$ and $\phi_t = 0.75$ yielded the best results for our data. To put this into perspective for the actual spatial and temporal scales, we can say that these choices correspond to a significant correlation in an approximate range of 60 km and a time span of four weeks. On the other hand, the LM test returned a p-value less than the level of significance (0.05), establishing that a space-time interaction effect is indeed present in the data. Thus, the model we decide to fit for our empirical analysis is the same as that given in Equation (3.5) shown below for ease of reference:

Table 2. Estimates of the overall location mean parameters.

Parameter	Estimate	Standard error	Confidence interval
α_1 (Humidity)	−0.0370	0.0008	(−0.0387, −0.0353)
α_2 (Temperature)	0.0226	0.0024	(0.0174, 0.0278)
α_3 (Wind speed)	−0.1022	0.0038	(−0.1105, −0.0939)
β_2 (February)	0.0186	0.0259	(−0.0321, 0.0692)
β_3 (March)	0.0593	0.0269	(0.0066, 0.1121)
β_4 (April)	−0.3770	0.0302	(−0.4362, −0.3178)
β_5 (May)	−1.1663	0.0355	(−1.2358, −1.0968)
β_6 (June)	−1.8376	0.0396	(−1.9151, −1.7600)
β_7 (July)	−1.9948	0.0416	(−2.0764, −1.9133)
β_8 (August)	−1.6383	0.0409	(−1.7186, −1.5581)
β_9 (September)	−1.0437	0.0393	(−1.1206, −0.9668)
β_{10} (October)	−0.4823	0.0341	(−0.5491, −0.4155)
β_{11} (November)	−0.3396	0.0299	(−0.3982, −0.2809)
β_{12} (December)	0.0379	0.0258	(−0.0127, 0.0885)

$$Y(s,t) = c(s,t)^T\alpha + \sum_{j=2}^{12} m(t,j)\beta_j + \sum_{k=1}^{K} r(s,k)\ \gamma_k t + v(s,t) + \epsilon(s,t),$$

for $s = 1, \ldots, n;\ t = 1, \ldots, T$. Recall that $c(s,t)$ describes the covariates in the study, and we use relative humidity, temperature, and wind speed in the analysis. Furthermore, $T = 522$ and $n = 66$ and, thus, the number of clusters (K) is taken to be eight.

4.3. Parameter estimates

Next, we fitted the above model to our Taiwan data to obtain the parameter estimates. In this application, we have 22 parameters in the mean structure and 12 more in the variance structure. In what follows, we provide a step-by-step discussion for these estimates. Table 2 shows the coefficient estimates corresponding to the covariates and the estimates of the seasonal effects.

From the table, almost all of the estimates show a significant effect. The $PM_{2.5}$ is higher whenever the humidity is less and the temperature is higher. In addition, as expected, a higher wind speed reduces the amount of pollution significantly. On the other hand, there exists a strong seasonal pattern. The winter months (from December to February) do not show a significant change in pollution. March shows a significant increase, whereas the $PM_{2.5}$ decreases over the summer months, especially between June and August. This is understandable, given the geographical location and climate pattern of Taiwan. It is more

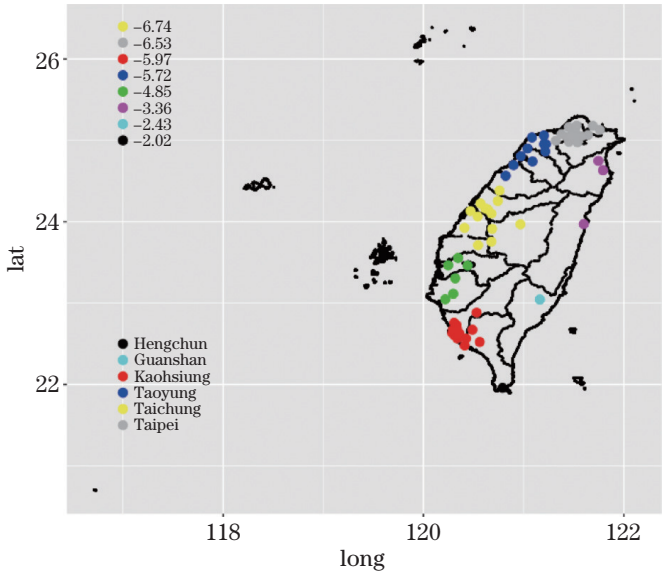

Figure 10. Estimates of the space-time interaction terms for different regions.

likely to rain in summer in Taiwan, and the wind tends to be strong and from the south-east with possible typhoons. On the other hand, the winter months tend to be dry, with wind from the north-west.

We then plotted the space-time interaction coefficients to see how they are spread across the country. Figure 10 shows these estimates on a spatial scale. We found that all regions showed significantly negative estimates for the space-time interaction coefficients, with the extent of the interaction varying by region. In an absolute sense, it is lowest for stations Hengchun (denoted by black) and Guanshan (denoted by light blue), and is highest for stations around Taipei (gray), Taichung (yellow), Taoyung (blue), and Kaohsiung (red). This is understandable, because Hengchun is at the southern tip of Taiwan, and Guanshan is less populated without heavy industry. On the other hand, the most significant stations are around the four largest cities in Taiwan.

The estimates of the variance parameters by month are shown in Figure 11. The estimates show an oscillatory behavior across months, and the magnitudes of the variances range between 0.4 and 1. These estimates explain the extent of the heteroskedasticity in the $PM_{2.5}$ process in our model. For the spatio-temporal process, the variance σ^2 was estimated at 0.7974, which is more than most variance terms of the white noise. This indicates that the white noise process explains less variability in the data, and that the spatio-temporal process

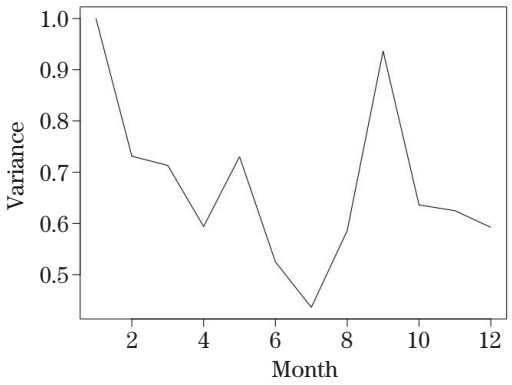

Figure 11. Estimates of the variance term by month.

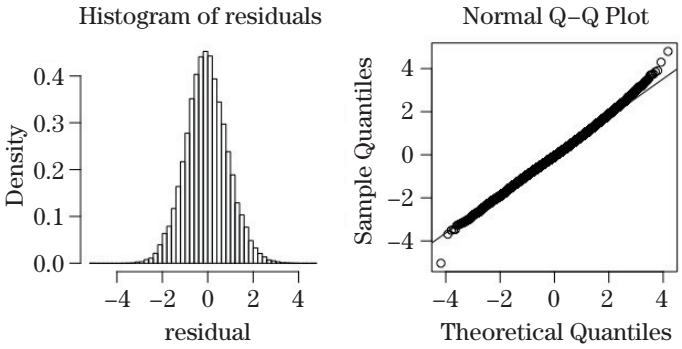

Figure 12. (Left) Histogram and (Right) QQ plot of the standardized residuals.

is more significant in this regard.

4.4. Model diagnostics

In this section, we discuss the behavior of the residuals to show that the proposed model fares well in capturing the effects of Taiwan $PM_{2.5}$ data. Standard statistical procedures reveal that there is no particular pattern in the residuals, establishing that the residuals can be assumed to be uncorrelated. Furthermore, we found that the standardized residuals are more or less evenly distributed across months, showing that we have resolved the issue of heteroskedasticity. These plots are available in the Supplementary Material.

Next, the left panel of Figure 12 shows that the histogram follows an approximately normal shape, which is corroborated by the QQ plot presented in the right panel of the same figure. Consequently, the introduction of heteroskedasticity and the spatio-temporal process works well for the data under study.

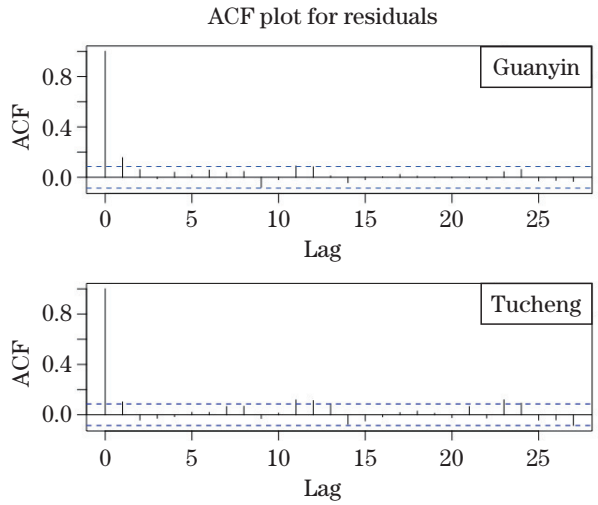

Figure 13. ACF plots of the standardized residuals, corresponding to two sample stations.

We also provide the autocorrelation functions (Figure 13) for the residuals corresponding to two particular stations: Guanyin and Tucheng. They show that the time dependence has been resolved, and that there is not a lot of time-dependency left in the residuals.

4.5. Prediction

Finally, we check the predictive ability of the proposed model using cross-validation techniques. For this, we use 90% of the available data (from 2006 to 2014, for 64 stations) and predict the $PM_{2.5}$ levels for the year 2015 for all stations. This provides us with a detailed view of the out-of-sample predictive performance, in both the temporal and the spatial sense.

In order to evaluate how good the predictions are, we calculated the root mean squared error for the observations, yielding a value of approximately 6.785. In order to understand the effect of the space-time interaction term in our model, we also measured the predictive ability of the same model, but without the interaction component. In that case, the root mean squared error was approximately 37% more than that of the proposed model.

In Figure 14, we show the forecast, along with the prediction interval for two stations. It is clear that the performance of the proposed model is satisfactory. In general, our method tends to overestimate the pollution level, owing to the weighted scheme of the method. However, the forecasts are usually inside the prediction interval, showing that the method works well for prediction purposes.

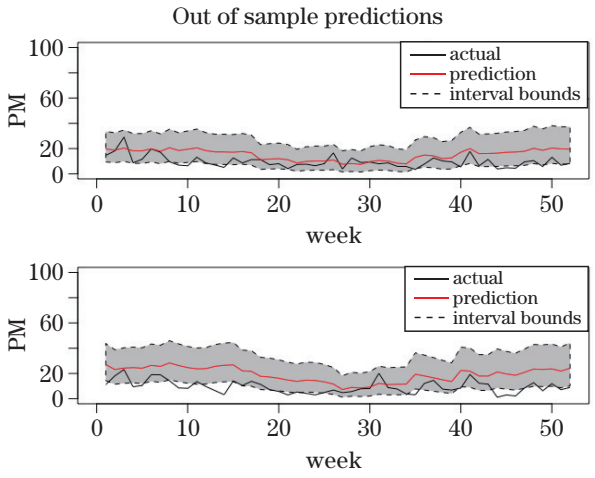

Figure 14. True value and predicted $PM_{2.5}$ observations, along with the prediction interval, for two stations and one year.

5. Conclusion

We have developed a new spatio-temporal modeling technique for identifying a space-time interaction. Our simulation studies and data analysis confirm that the method performs well. In particular, we have shown that the estimates obtained using the proposed method are consistent. Furthermore, we used standard diagnostic techniques to establish that the model assumptions are reasonable. This modeling technique can successfully detect and estimate a space-time interaction in air pollution data. Furthermore, because of the weighting scheme we use in the method, it has the potential to predict a higher level of pollution with more precision. This is useful from a practical point of view.

We finish with some comments on possible future studies. An important potential future direction of this work is to consider a more generalized framework in the spatio-temporal process. In particular, we consider a separable structure for the spatial and temporal dependence, a condition that can be relaxed to address a more general setup. Moreover, in the aforementioned data analysis example, it was found that the maximum wind speed at each location plays an important role in the level of air pollution, while there is a significant space-time interaction effect as well. Taken together, these points raise an important question: how much of an effect does the wind flux, calculated from the wind speed, wind direction, and the coordinates of a location, have on pollution? Answering this question requires knowing about the physical behavior of the

wind. This process can be incorporated into the model to develop more efficient techniques.

Another important aspect of this topic is to identify the effect of air pollution on human life. Pollution is responsible for many respiratory and cardiovascular diseases; thus, it is crucial that we use appropriate modeling techniques for such problems. This will potentially have a significant impact on the health sciences.

Supplementary Materials

The proof of Theorem 1 and additional empirical data analyses are available in the online Supplementary Material.

References

Al-Awadhi, F. A. and Al-Awadhi, S. A. (2006). Spatial-temporal model for ambient air pollutants in the state of kuwait. *Environmetrics* **17**, 739–752.

Berrocal, V. J., Gelfand, A. E. and Holland, D. M. (2010). A bivariate space-time downscaler under space and time misalignment. *The Annals of Applied Statistics* **4**, 1942.

Bilonick, R. A. (1985). The space-time distribution of sulfate deposition in the northeastern united states. *Atmospheric Environment (1967)* **19**, 1829–1845.

Blangiardo, M., Finazzi, F. and Cameletti, M. (2016). Two-stage bayesian model to evaluate the effect of air pollution on chronic respiratory diseases using drug prescriptions. *Spatial and Spatio-temporal Epidemiology* **18**, 1–12.

Cameletti, M., Ignaccolo, R. and Bande, S. (2011). Comparing spatio-temporal models for particulate matter in piemonte. *Environmetrics* **22**, 985–996.

Carroll, R., Chen, R., George, E., Li, T., Newton, H., Schmiediche, H. and Wang, N. (1997). Ozone exposure and population density in harris county, texas. *Journal of the American Statistical Association* **92**, 392–404.

Elsom, D. (1978). Spatial correlation analysis of air pollution data in an urban area. *Atmospheric Environment (1967)* **12**, 1103–1107.

Eynon, B. P. and Switzer, P. (1983). The variability of rainfall acidity. *Canadian Journal of Statistics* **11**, 11–23.

Goldberger, A. S. (1962). Best linear unbiased prediction in the generalized linear regression model. *Journal of the American Statistical Association* **57**, 369–375.

Guttorp, P., Meiring, W. and Sampson, P. D. (1994). A space-time analysis of ground-level ozone data. *Environmetrics* **5**, 241–254.

Haas, T. C. (1995). Local prediction of a spatio-temporal process with an application to wet sulfate deposition. *Journal of the American Statistical Association* **90**, 1189–1199.

Jerrett, M., Burnett, R. T., Beckerman, B. S., Turner, M. C., Krewski, D., Thurston, G., Martin, R. V., van Donkelaar, A., Hughes, E., Shi, Y. and et al. (2013). Spatial analysis of air pollution and mortality in california. *American Journal of Respiratory and Critical Care Medicine* **188**, 593–599.

Kibria, B. G., Sun, L., Zidek, J. V. and Le, N. D. (2002). Bayesian spatial prediction of random

space-time fields with application to mapping pm2. 5 exposure. *Journal of the American Statistical Association* **97**, 112–124.

Kulldorff, M. and Hjalmars, U. (1999). The knox method and other tests for space-time interaction. *Biometrics* **55**, 544–552.

Kyriakidis, P. C. and Journel, A. G. (1999). Geostatistical space–time models: a review. *Mathematical Geology* **31**, 651–684.

Legendre, P., Cáceres, M. D. and Borcard, D. (2010). Community surveys through space and time: testing the space–time interaction in the absence of replication. *Ecology* **91**, 262–272.

Mayer, H. (1999). Air pollution in cities. *Atmospheric Environment* **33**, 4029–4037.

Meyer, S., Warnke, I., Rössler, W. and Held, L. (2016). Model-based testing for space–time interaction using point processes: An application to psychiatric hospital admissions in an urban area. *Spatial and Spatio-temporal Epidemiology* **17**, 15–25.

Pope III, C. A., Thun, M. J., Namboodiri, M. M., Dockery, D. W., Evans, J. S., Speizer, F. E. and Heath Jr, C. W. (1995). Particulate air pollution as a predictor of mortality in a prospective study of us adults. *American Journal of Respiratory and Critical Care Medicine* **151**, 669–674.

Rouhani, S. and Hall, T. J. (1989). Space-time kriging of groundwater data. In: *Geostatistics*. 639–650, Springer.

Sahu, S. K., Gelfand, A. E. and Holland, D. M. (2006). Spatio-temporal modeling of fine particulate matter. *Journal of Agricultural, Biological, and Environmental Statistics* **11**, 61–86.

Sahu, S. K. and Mardia, K. V. (2005). A bayesian kriged kalman model for short-term forecasting of air pollution levels. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **54**, 223–244.

Sahu, S. K., Yip, S. and Holland, D. M. (2009). Improved space–time forecasting of next day ozone concentrations in the eastern US. *Atmospheric Environment* **43**, 494–501.

Shaddick, G. and Wakefield, J. (2002). Modelling daily multivariate pollutant data at multiple sites. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **51**, 351–372.

Smith, R. L., Kolenikov, S. and Cox, L. H. (2003). Spatiotemporal modeling of pm2. 5 data with missing values. *Journal of Geophysical Research: Atmospheres* **108**.

Sun, L., Zidek, J. V., Le, N. D. and Özkaynak, H. (2000). Interpolating vancouver's daily ambient pm10 field. *Environmetrics* **11**, 651–663.

Thurston, G. D., Burnett, R. T., Turner, M. C., Shi, Y., Krewski, D., Lall, R., Ito, K., Jerrett, M., Gapstur, S. M., Diver, W. R. and et al. (2016). Ischemic heart disease mortality and long-term exposure to source-related components of us fine particle air pollution. *Environmental Health Perspectives (Online)* **124**, 785.

Vanem, E., Huseby, A. B. and Natvig, B. (2014). Bayesian hierarchical spatio-temporal modelling of trends and future projections in the ocean wave climate with a co_2 component. *Environmental and Ecological Statistics* **21**, 189–220.

Vyas, V. M. and Christakos, G. (1997). Spatiotemporal analysis and mapping of sulfate deposition data over eastern USA. *Atmospheric Environment* **31**, 3623–3633.

Wikle, C. K., Berliner, L. M. and Cressie, N. (1998). Hierarchical bayesian space-time models. *Environmental and Ecological Statistics* **5**, 117–154.

5747 South Ellis Avenue, Chicago, IL 60637, USA.
E-mail: sdeb@uchicago.edu
5807 South Woodlawn Avenue, Chicago, IL 60637. USA.
E-mail: ruey.Tsay@chicagobooth.edu

(Received January 2018; accepted January 2019)