Statistica Sinica Preprint No: SS-2017-0561

Title	Spatio-Temporal Models with Space-Time Interaction and Their Applications to Air Pollution Data
Manuscript ID	SS-2017-0561
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202017.0561
Complete List of Authors	Soudeep Deb and Ruey S. Tsay
Corresponding Author	Ruey S. Tsay
E-mail	Ruey.Tsay@chicagobooth.edu

Statistica Sinica

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 of utmost importance to have a clear understanding of the status of air pollution and to provide forecasts and insights about the air quality to the general public and researchers in environmental studies. Previous studies showed that even a short-term exposure to high concentrations of atmospheric fine particulate matters can be hazardous to the health of ordinary people. In this study, we develop a spatio-temporal model with space-time interaction for air pollution data ($PM_{2.5}$). The proposed model uses a parametric space-time interaction component along with the spatial and temporal components in the mean structure, and introduces a random-effects component specified in the form of zero-mean spatio-temporal processes. For application, we analyze the air pollution data ($PM_{2.5}$) from 66 monitoring stations across Taiwan.

Key words and phrases: Fine particulate matter; Dynamical dependence; Spatial dependence; Lagrange multiplier test

1. Introduction

The effect of air pollution on public health, vegetation, and more generally, on the human society and the ecosystem has been a burning issue in recent years. Several epidemiological studies have established that the particulate matters (PM) are 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 et al. [2016], and Thurston et al. [2016]. Taking the severe effect of the PM into account, the Environmental Protection Agency (EPA) of the United States of America (USA) provided in 1997 new regulations that established National Ambient Air Quality Standards (NAAQS) for PM with aerodynamic diameters less than 2.5 micron. This is usually measured in units of micrograms per cubic meter (μgm^{-3}), and we will denote it as PM$_{2.5}$ henceforth. According to NAAQS, the hourly average of the PM$_{2.5}$ concentration should not be over 35 μgm^{-3}. However, the real situation often goes beyond the standard. A quick example is the data analyzed in this study. They are obtained from 66 different monitoring stations in Taiwan, for a span of 10 years (from 2006 to 2015) and the median of the PM$_{2.5}$ values is slightly above 37 μgm^{-3}. The reader is also referred to the earlier study by Mayer [1999] who discussed how the air quality is deteriorating in different cities across the world. All in all, there is a growing demand to identify the main factors that contribute to the air pollution.

A brief discussion on PM is in order. Generally speaking, PM$_{2.5}$ contains different particles either emitted directly or formed in the atmosphere from gaseous emissions. The examples include sulfates formed from sulfur dioxide (SO_2) emissions, nitrates formed from NOx emissions, and carbon formed from organic gas emissions. Now, the rates of conversion of gases to particles are often reliant on different regional and temporal factors, including the topography, the land cover and several seasonal climatic variables, and as a consequence, the PM$_{2.5}$ concentrations are also affected by these variables. This immediately increases the need for a spatio-temporal model to assess the air quality. A good model can provide better predictions and that would in turn help in determining an efficient strategy to control the air pollution.

The spatial and spatio-temporal modeling of air pollutants started early in the past century. Elsom [1978] studied the spatial correlation fields for air pollution in an urban area while different geostatistical space-time models have been applied to examine the trend in deposition of atmospheric pollutants in Eynon and Switzer [1983], Bilonick [1985], Rouhani and Hall [1989], and Vyas and Christakos [1997]. Other earlier notable works in this regard include Guttorp et al. [1994], Haas [1995], and Carroll et al. [1997]. Most of these approaches rely on a spatio-temporal random field where the spatial or temporal dependences are incorporated in either the mean function or in the error process, and the parameters are estimated using frequentist procedures. A nice discussion on the geostatistical space-time models can be found in Kyriakidis and Journel [1999].

In comparison, many 21st century studies in the related problem have made use of hierarchical Bayesian approaches for spatial prediction of air pollution. For example, Sun et al. [2000] analyzed the PM_{10} (particulate matters with diameter less than 10 $\mu\mathrm{gm}^{-3}$) 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 $\mathrm{PM}_{2.5}$ concentrations in Philadelphia and developed spatial prediction methodology in a Bayesian context for this purpose. 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 by a random walk process to present a short-term forecasting analysis of $\mathrm{PM}_{2.5}$ data in New York City. There are two other notable works by the same authors. Sahu et al. [2006] used indicators for urban or rural sites to employ different spatio-temporal processes in the error structure for modeling the $PM_{2.5}$ series of several states in the Midwest of the United States (US). In a later paper, Sahu et al. [2009] developed a space-time model, that includes a spatially varying regression term along with an auto-regressive term in the mean structure, to analyze the ozone concentrations in the eastern states of the US. Berrocal et al. [2010] extended one of their earlier works to introduce a bivariate downscaler and provided a flexible class of space-time assimilation models. On the other hand, Cameletti et al. [2011] provided a nice discussion on the comparison of available space-time models using data from Piemonte, Italy.

In this study, we consider the problem of developing a new spatio-temporal model, with the main focus on identifying if there is any space-time interaction in the behavior of the $PM_{2.5}$ concentrations. An interaction means that the temporal trend of the pollution is more similar for sites closer in a spatial scale. An early study in this regard is by Wikle et al. [1998]. The authors developed a hierarchical Bayesian model that allowed an interaction in space and time. However, except for that and a few related works, there have been very few efforts to quantify the space-time interaction for air pollution data, albeit there is a vast literature on the spatio-temporal modeling for the same, as we have already discussed.

Note that the problem of identifying the space-time interaction is not at all specific to the air pollution data. Rather, it has been studied in several other fields, ranging from seismology, epidemiology, criminology to transportation research. Meyer et al. [2016] provided a nice discussion on the tests for space-time interaction in problems related to medical studies. The most popular techniques in this regard are Knox test, Mantel test, and space-time K-function analysis. These tests mainly revolve around test statistics of the form $T = \sum_{j \neq i} a^s_{ij} a^t_{ij}$, where a^s_{ij} and a^t_{ij} are measures of the spatial and temporal adjacency of the events i and j. Further reading on space-time interaction and related problems from other fields can be found in many articles in the literature; Kulldorff and Hjalmars [1999], Legendre et al. [2010], Vanem et al. [2014] and the references therein being a few examples.

We conclude the section by noting that most of the air pollution studies mentioned earlier in this section concentrated on the pollution issues in different parts of Canada and the USA, while the problem is certainly not limited to this continent only. In fact, there have been only a few good studies that analyze air pollution data from countries in Asia or Africa. A noteworthy paper in this regard is Al-Awadhi and Al-Awadhi [2006], who took hierarchical Bayesian approaches to develop a 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 serious dearth of related papers that focus on the pollution situations in Asia and Africa. And that is another contribution of this work, as we develop analyze the data for Taiwan.

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 shows the results of detailed data analysis while Section 5 provides some concluding remarks and the scopes of future work.

2. Preliminary Analysis

2.1 Data

The $PM_{2.5}$ data we analyze were collected from 71 official monitoring stations across Taiwan. However, for five of those stations, we do not have the necessary information on the covariates considered and so, we decide to exclude them from this study. The remaining 66 monitoring stations are irregularly located in space, spread over the whole Island with some concentrations around big cities or industrial areas. The minimum and maximum of the pairwise distances of these stations are 0.58 kilometers and 366.7 kilometers respectively while the arithmatic mean of the same is 140.7 kilometers. One major talking point of our study is that we consider the issue of space-time interaction as a property of a region, rather than that of every individual station. For that, we divide the data into several clusters, based on the latitude and longitude of the stations. More on this is discussed in the following sections. Figure 1 shows the exact locations of the 66 stations. The concentration of the stations on the west coast is understandable because it is the most populated area of the Island. Different colors of the map indicate different regions (clusters) in our study.

Turn to the temporal scale of our study. The data are obtained on an hourly basis for ten years, starting from January 1, 2006 to December 31, 2015. However, the sampling frequencies vary from station to station and for some sites, there are missing values. Moreover, as we are mainly interested in identifying space-time interaction for the air pollution data, lower temporal resolution is desired and simpler. Thus, in this analysis, we decide to aggregate the hourly data into weekly averages based on all available measurements within a week. It is worth mentioning that this is a

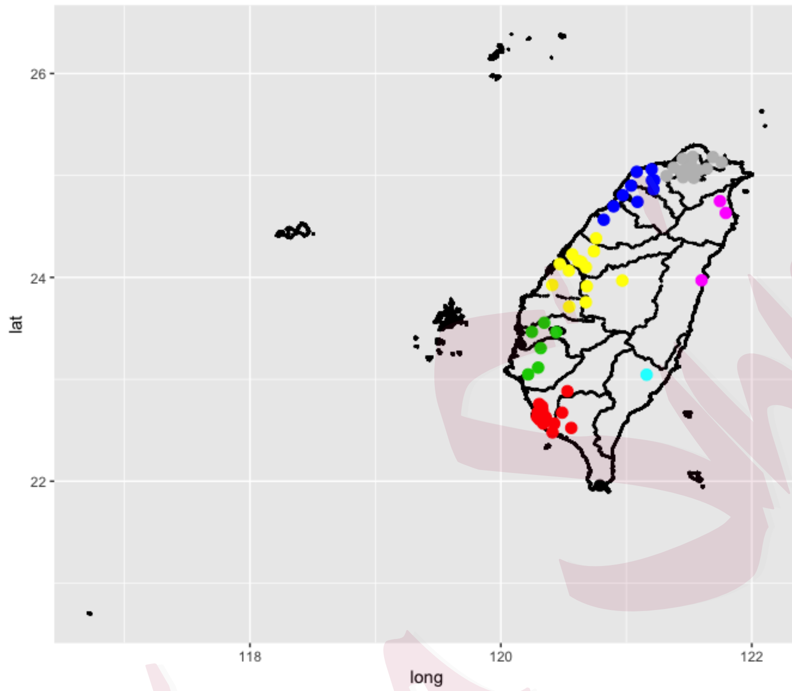

Figure 1: Actual map of the locations and clustering for the 66 monitoring stations on Taiwan.

common practice while dealing with monitoring data, as discussed in Smith et al. [2003]. In order to maintain continuity in the time scale, we consider the whole set of 3652 days (from 2006 to 2015) together and divide it into 522 weeks. Hence, the total number of data points considered in this study is $66 \times 522 = 34452$.

To begin, we present some exploratory analysis on the data. As has been done in several related studies, we convert the $PM_{2.5}$ values to the square root scale and the rest of the study is performed with the transformed data. This type of transformation is a common practice for air pollution data, as discussed in Smith et al. [2003]. First, the time series plots of the transformed $PM_{2.5}$ for three stations are displayed in Figure 2.

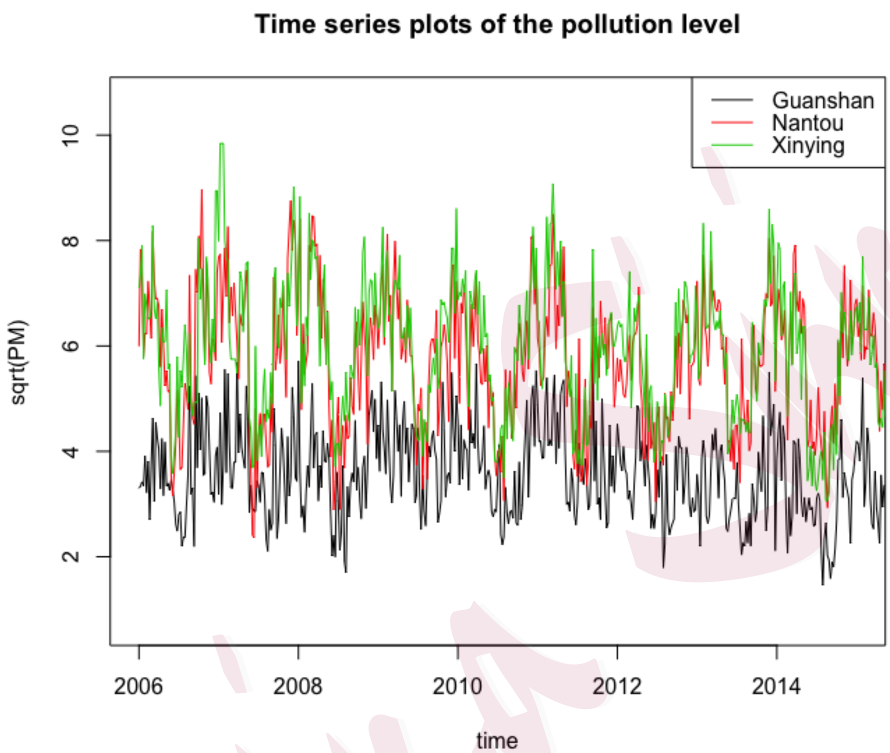

Figure 2: Time series plots of $\sqrt{\mathrm{PM}_{2.5}}$ observations for three particular stations - Guanshan, Nantou and Xinying.

Next, we take one of those series (corresponding to Xinying) and decompose the time series data into three components - seasonal, trend and the residuals. It is evident (see Figure 3) that there is a decreasing trend, which seems like a linear component with a small slope.

The graph in Figure 4 shows the overall means and the variances of the transformed $\mathrm{PM}_{2.5}$ observations for different stations and different 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 in

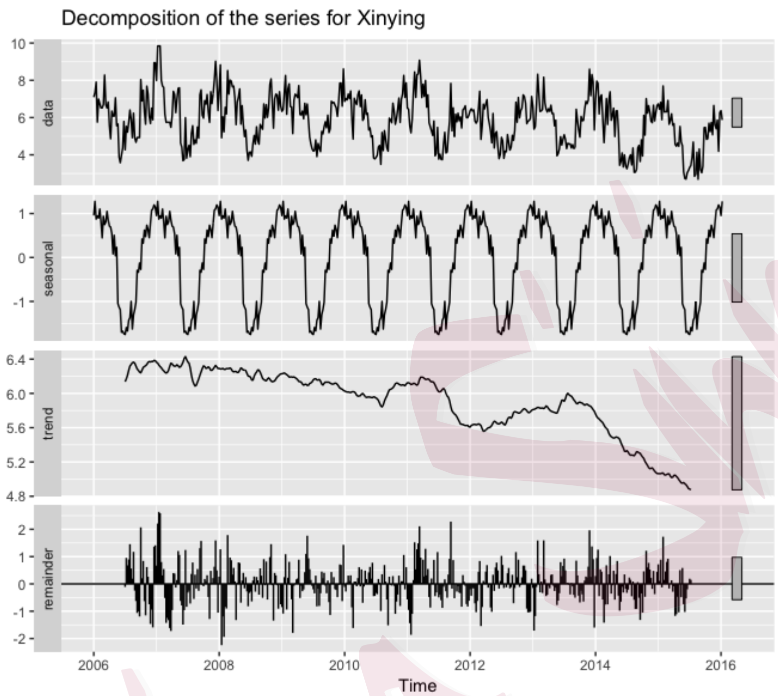

Figure 3: Decomposition of the $\sqrt{\mathrm{PM}_{2.5}}$ observations for Xinying. The four panels show the original series, seasonality, trend and residuals, respectively.

the study. The bottom two plots show the behavior of the means and variances 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 corresponding to the months suggests that the variance is not homoskedastic. In fact, it varies seasonally, and that motivates us to consider a heteroskedastic nature for the error variance in the Gaussian process of the proposed model.

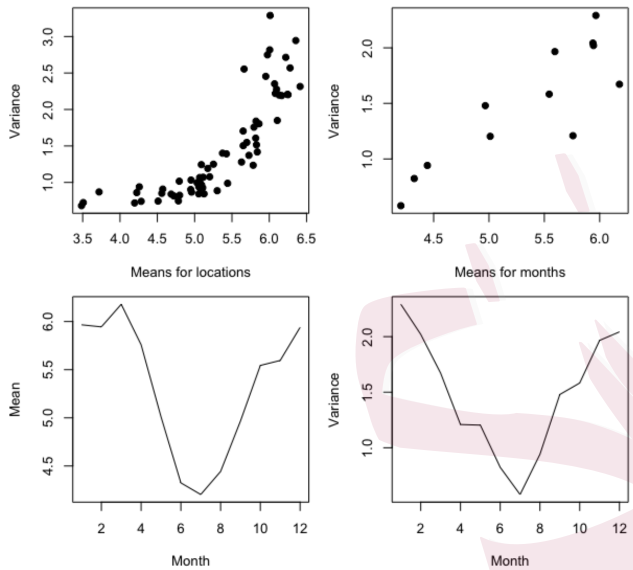

Figure 4: (Top left) Sample variances versus means of the weekly $\sqrt{\mathrm{PM}_{2.5}}$ observations for 66 stations, (Top right) Means versus variances of the $\sqrt{\mathrm{PM}_{2.5}}$ observations for 12 months, (Bottom left) Means of the $\sqrt{\mathrm{PM}_{2.5}}$ observations corresponding to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

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

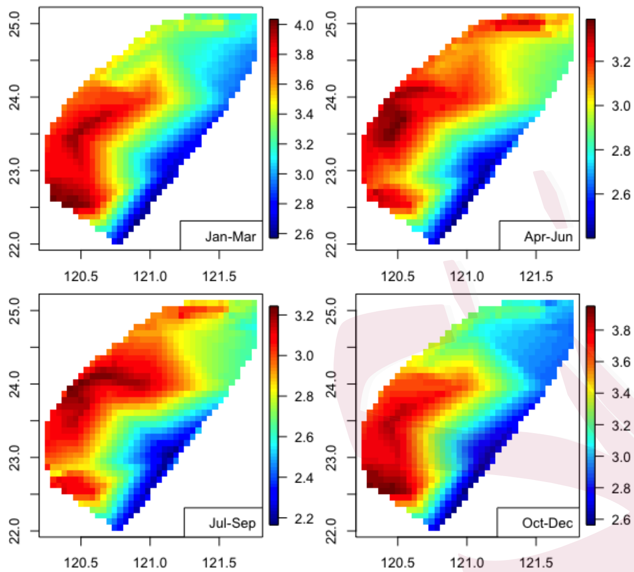

Figure 5: Seasonal heat map of the $PM_{2.5}$ levels on Taiwan.

2.2 Covariates used in the study

For each station, along with the observations of $PM_{2.5}$, data were collected on temperature, relative humidity, and wind speed and direction. Similar to the air pollution observations, these data were also collected on an hourly basis from January 1, 2006 to December 31, 2015. But, we aggregate them per week and use the representative values as covariates in our study.

At first, we examine the relationship of the pollution data with the relative humidity and the temperature. Exploratory analysis suggests that the air pollution is not so significantly affected by the temperature, but is more dependent on the humidity. It shows a slight decrease in pollution with the increase in the relative humidity. The Spearman correlation coefficient between transformed $PM_{2.5}$ and humidity was -0.326, while the same for the temperature was -0.254. Relevant plots on this are provided in the online supplement.

The wind speed is another variable that may have some effects on the air pollution. The data on wind speed and wind direction were available on an hourly basis. It was noted that the behavior of the wind speed and direction varied widely for different locations as shown in the boxplots in Figure 6. These plots shows clearly that the ranges of the wind speed vary markedly for different locations, while the means are not so much. Naturally, in order to account for the effect of the wind, unlike the previous covariates, it will be wise not to take the weekly average. Instead, we decided to use the maximum daily wind speed for every week. It is also worth mentioning that for most weeks, the maximum wind speed occurred on same day in all the locations, which again justifies why the maximum in lieu of the average is a good measure in this regard. The sample correlation coefficient between transformed $\mathrm{PM}_{2.5}$ and the maximum wind speeds is -0.146.

3. Methods

3.1 The proposed model

The proposed model is in some sense inspired by the one studied by Sahu et al. [2006]. The key feature of our model is that it accounts for possible space-time interactions in the air-pollution data. We describe the model in a general setup, and discuss the particular case of 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 $\mathrm{PM}_{2.5}$ at time t_j and location s_i by $Z(s_i, t_j)$, for $i = 1, \ldots, n$ and

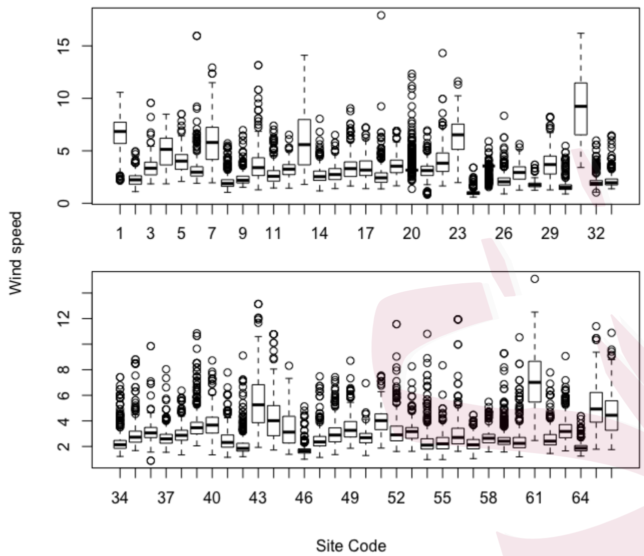

Figure 6: Boxplots of the wind speed for 66 stations

$j = 1, \ldots, T$. Then, we assume that the overall mean values for different locations are different, and we subtract the location-wise overall means from the actual values of the transformed $PM_{2.5}$ values to obtain the mean-adjusted numbers. These mean-adjusted values are denoted by $Y(s_i, t_j)$. For convenience, we drop the subscripts whenever 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, to account for measurement errors and other variability not addressed by $U(s,t)$. We assume the white noise process to follow heteroskedastic $N(0, \sigma_t^2)$ distributions independently, where σ_t^2 is chosen according to different seasons. $U(s,t)$, on the other hand, assumes the following structure:

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

where $\mu(s,t)$ stands for 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 available covariate information, we consider a term of the form $B\alpha$, where B is the design matrix for p covariates. The population densities, temperature, humidity, number of factories, number of cars, etc can be taken as the 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 used for 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 if the γ_s values are same across different locations. However, this would increase the complexity and computational burden dramatically, if we have a lot of sites. To address this issue, we make a logical assumption that the coefficients γ_s for locations which are close to each other are the same, and that it should be a characteristic of a region rather a station. So, based on the coordinates of the stations, we divide the stations into clusters first 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 5 or less, we do not use clustering. But for $n > 5$, we use $[\sqrt{n}]$

clusters ($[n]$ denotes the integer part of n) to divide the locations into different regions. In practice, we use the k-means clustering method, based on the latitude and longitude, for this purpose.

Further, we propose to use an intercept-free model, and so, along with the response variable, we also subtract the overall means of the covariates from the respective observations. Thus, 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 with the values of the covariates and $m(t,j)$ and $r(s,k)$ are the indicators for the season and location region, respectively, where J and K are the seasonality and the number of regions. That is, $m(t,j) = 1$ if time t is in the jth season and 0 otherwise. Similarly, $r(s,k) = 1$ if location s is in kth region and 0 otherwise. For identifiability purposes, we would take $\beta_1 = 0$. Note that because of using mean-adjusted values for the response and the covariates, we do not have any intercept term in the model. On the other hand, we would 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. Averaging over different sites in a region, we can get the adjustment to the regional trends and averaging over time can help us 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 cluster 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 a product of the spatial dependence and temporal dependence and we write it 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 of the two points.

Throughout the paper, Y denotes the vector of $N = nT$ data points, arranged according to clusters at first, time points next, and sites then. Thus, if $\{s_1,s_2\}$ form a cluster, then the first few observations will be $Y(s_1,t_1), Y(s_2,t_1), Y(s_1,t_2), Y(s_2,t_2)$, etc. 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, and is of the form $\sigma_v^2(\Sigma_t\otimes\Sigma_s)$ such that $\Sigma_t(i,j)=\rho(\|t_i-t_j\|,\phi_t)$ and $\Sigma_s(i,j)=\rho(\|s_i-s_j\|,\phi_s)$. Further, note that when the sampling design considers equally-spaced time points, the common spacing being d_t, one 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 are different for different seasons, and so, we assume $\epsilon(s,t)\sim N(0,\sigma^2_{m(t)})$, where $m(t)$ denotes the season the time t is in. If we use ϵ to denote the vector of $\epsilon(s,t)$, arranged similarly as 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 parameters associated with the ith season. Also, we set $\tau_1^2=1$ to avoid any potential identifiability problem.

Now, to write the full model in a vector-matrix notation, recall that Y is the observed vector of dependent varaible, 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, σ_v^2 to be equal. Here, the estimate of σ^2 would give us an idea about the variance explained by the spatio-temporal process while 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 prediction 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 would 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 see which combination of (ϕ_s, ϕ_t) would work the best. The possible choices for ϕ_s used in the study were $(0.001, 0.005, 0.01, 0.025, 0.05, 0.075, 0.1, 0.25, 0.5, 0.75)$ while the choices for ϕ_t were $(0.25, 0.5, 0.75, 1, 1.25, 1.5)$. To find out the optimal choice, we find the combination with the smallest mean squared error of the predictions, which is calculated as follows.

For each site s_i, let us 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 if there is any space-time interaction present in the data. Following that, an iterative estimation method is executed. This method is focused on identifying the effect of the factors that lead to higher pollution level.

Test for interaction:

The model described in Equation (3.5) allows us to test for space-time interaction. The model includes no interaction term if $\gamma_1 = \ldots = \gamma_K = 0$. 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 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 it, unlike Wald test or likelihood ratio test, does not require an estimate of the information under the alternative hypothesis or unconstrained maximum likelihood. The LM test uses only the assumptions in the null hypothesis to get the maximum likelihood estimates and then 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 make a decision about the model to use. 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 thought of in the form $Y = X\theta + \varepsilon$, 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. Further, note that the number of unknown parameters in the error covariance matrix Ω is $J + 2$, namely $\sigma^2, \tau_2^2, \ldots, \tau_J^2; \phi_s, \phi_t$. As we decided to get the optimal choices for the last two parameters by a cross-validation procedure, that leaves us the task of estimating J additional variance parameters from the model.

From a practical point of view, it is more important to identify the cases where the air pollution is hazardous for health and environment, rather than the ones when the situation is less harmful. According to the standards set by the EPA, the average for the fine particulate matters should not exceed 35 μgm^{-3}. So, while minimizing the squared sum of residuals, we put more weight on the cases where the actual PM$_{2.5}$ values are more than 35. Thus, the loss function that we want 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 taken 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 is that the weights will approach 1 as $N \to \infty$, thereby establishing that the importance will be approximately equal on all observations when the sample size is huge. $\hat{e}(s_i, t_j)$, on the other hand, 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 to us, 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 $\hat{\theta} = (X^TVX)^{-1}(X^TVY)$, where $V = \hat{\Omega}^{-1/2}W\hat{\Omega}^{-1/2}$. In light of the above discussion, we propose to use the following procedure to estimate the parameters in 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 $\hat{\theta} = (X^TVX)^{-1}(X^TVY)$ 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 accordingly in the above steps, we substantially reduce the computational burden in the estimation.

The following theorem describes the asymptotic properties of the above estimators. Proof of the theorem is provided in the online supplementary file.

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, similar to what we did in Section 3.1. Then, it can be said that $\hat{Y}(s', t') = X(s', t')'\theta + \varepsilon(s', t')$. So far as the assignment of the cluster goes, we use techniques similar to 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 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 going to be a 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 get $\hat{\eta}$ 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) by using the historical overall mean of the station. In case of a new location, we use the overall mean of the sites from the cluster it is assigned to. Finally, we transform the predictions to get the estimate of the actual air pollution measurement.

To obtain prediction interval, we make use of 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 $\hat{\text{SE}} = \hat{\sigma}^2\hat{\zeta}'(X'\hat{\Omega}^{-1}X)^{-1}\hat{\zeta} + \hat{\sigma}^2/m$. 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 some simulation studies to show that our methods are capable of capturing the space-time interaction that might be present in the empirical application we are interested in.

We perform the simulation experiment in three stages - first with a linear space-time interaction term and then with two different non-linear types of interactions. Throughout this study, we consider weekly average of the air pollution data. And then, for different time intervals, we evaluated the type-I error and the power for different values of n. In the simulation study, we considered four 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 we have more data. 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 the coefficients in the model were simulated from independent normal distributions with mean 0 and standard deviation 1. 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 linear space-time interaction, the observations (square root of the $\text{PM}_{2.5}$ data) were obtained from the 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 non-linear case where the interaction term is $\sin(2\pi t/T)$. This example would address the cases where the interaction can be an oscillatory function of time. The results for different cases under the three different scenarios are displayed in Table 1. The results reported are based on 500 iterations of simulation.

While the type I error remains under control across all scenarios, the power improves significantly for 20 or more locations and this is true both for the linear dependence and for the nonlinear dependence. The power also increases with sample size for all cases. The results confirm that the proposed testing 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 3 years (157 observations for each site) and took different values for n, ranging from 10 to 50. The results of 500 iterations for different cases are shown in Figure 7. It is evident that, when the dependence on time is linear, the proposed testing procedure fares nicely for large number of sites. The procedure also works well for the nonlinear case as its power also increases, reaching more than 0.8 when the number of locations is 50.

Next, we present a particular case to show that the parameter estimation process indeed works well to identify 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 get weekly averages for 5 years (divided in 12 seasons). Thus, the number of observations in the data were $20 \times 261 = 5220$. 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

Table 1: Results for Different Cases, 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

points lie along the line $y = x$ (displayed in the figure), thereby showing that the estimates are not too different from the true values. This study 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 6 years. Then, use five years' data from 19 stations to train our model and use that to make predictions for the sixth year of the 20th station. This is displayed in Figure 9. In

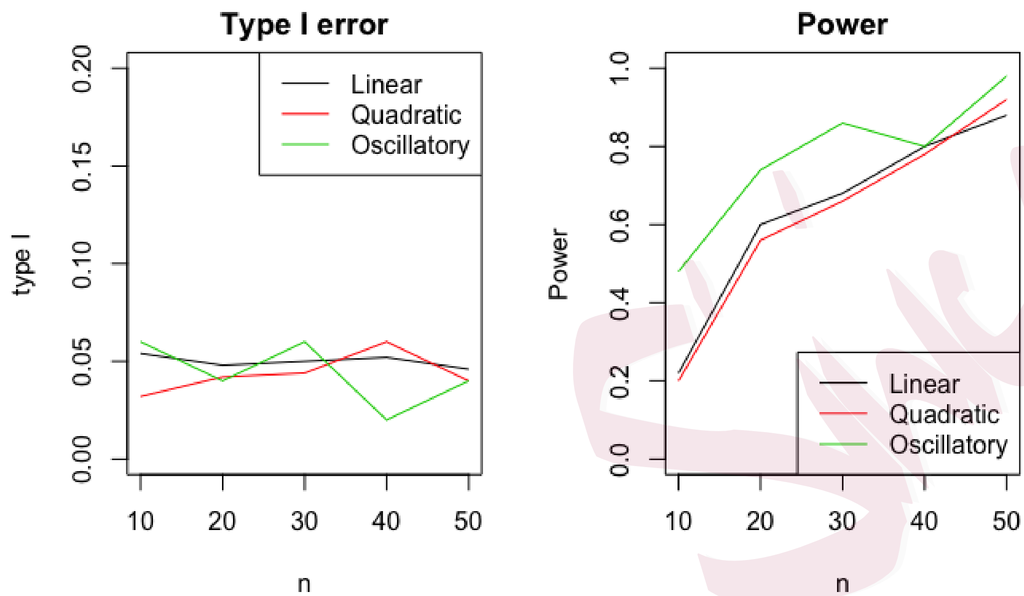

Figure 7: Type-I error (on left) and power (on right) for linear (black), quadratic (red) and sinusoidal (green) interaction term, corresponding to different number of locations. $T = 157$ (3 years) and 500 iterations are used for all cases.

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 whenever the actual values are above $35\mu\text{gm}^{-3}$. This brings into attention 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 5 years, the performance is good in that regard.

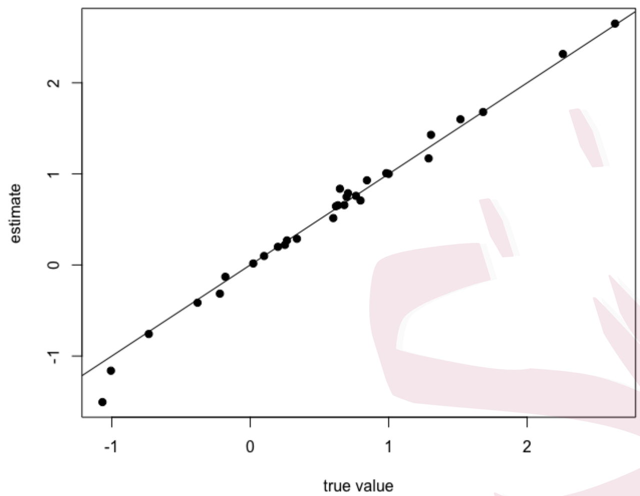

Figure 8: True value and estimated value for the parameters in the model, where data is generated for 20 locations and 5 years

4.2 Model selection

The first task to implement the proposed method is to choose proper decay parameters and to identify if there exists any space-time interaction. For the decay parameters, we searched in a two-dimensional array to find out which combination gives the least mean-squared error (refer to Equation (3.6)). The choices for ϕ_s were $(0.001, 0.005, 0.01, 0.025, 0.05, 0.075, 0.1, 0.25, 0.5, 0.75)$ while the same for ϕ_t were $(0.25, 0.5, 0.75, 1, 1.25, 1.5)$. We used 9 years and 64 stations to train the model and the rest were used for validation purposes in this regard.

We found that the combination of $\phi_s = 0.05$ and $\phi_t = 0.75$ was the best one for the data at hand. To put it into the perspective of the actual spatial and temporal scale, we can say that these

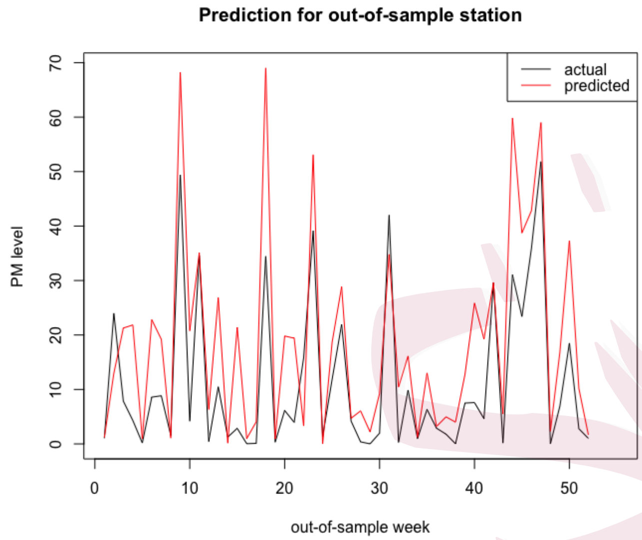

Figure 9: True value and predicted $PM_{2.5}$ observations for one station and for one year. Model was trained using a data generated for 19 locations and 5 years.

choices correspond to a significant correlation in an approximate range of 60 kilometers and in a time span of 4 weeks, respectively. 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 Equation (3.5), which is shown below for ease in reference.

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

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, $T = 522$, $n = 66$ and thus, the number of clusters (K) is taken to be 8.

4.3 Parameter estimates

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

From the table, it is seen that 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. Also, 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) does not show a significant change in the 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 likely to rain over the summer in Taiwan and the wind tends to be from the south-east and the speed can be high with possible strong typhoons. On the other hand, winter months tend to be dry with wind from north-west in Taiwan.

We then plotted the space-time interaction coefficients to see how they spread across the whole Island. Figure 10 shows these estimates on a spatial scale. It was found that all of the regions showed significantly negative estimates for the space-time interaction coefficients, the extent of interaction being different from region to region. In absolute sense, it is the least for stations Hengchun (denoted by black) and Guanshan (denoted by light blue) while it is the most for stations around Taipei (gray), Taichung (yellow), Taoyung (blue), and Kaohsiung (red) of Taiwan. This is also

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	$(-0.0127, 0.0885)$

understandable as Hengchun is at the southern tip of Taiwan and Guanshan is less populated and without any heavy industry. On the other hand, the most significant stations are around the four largest cities in Taiwan.

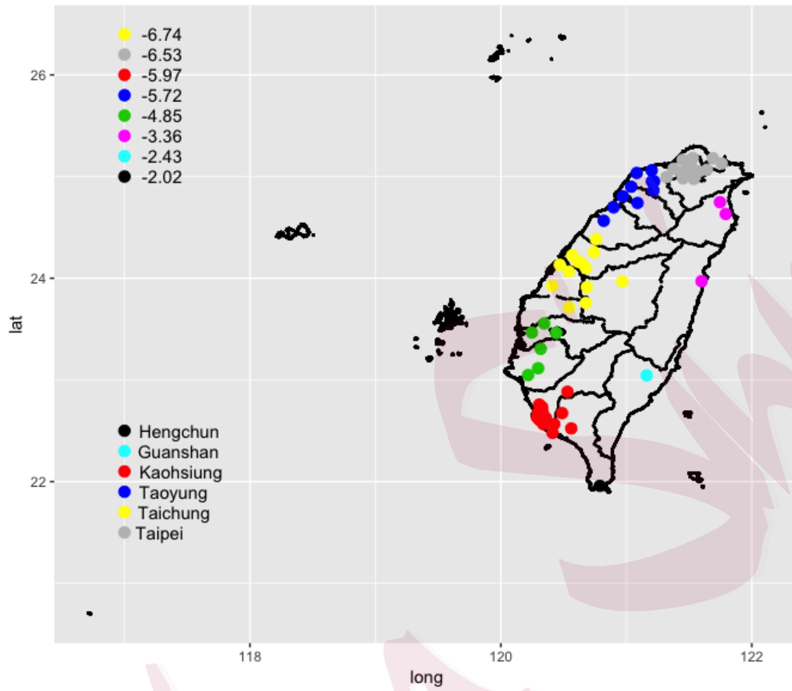

Figure 10: Estimates of the space-time interaction terms for different regions

The estimates of the variance parameters for different months are shown in Figure 11. We can see that the estimates show an oscillatory behavior across different months, and the magnitudes of the variances range between 0.4 and 1. These estimates can explain the extent of the heteroskedasticity in the $PM_{2.5}$ process in our model. So far as the spatio-temporal process is concerned, its 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 can explain less variability in the data, and that the spatio-temporal process is more significant in this aspect.

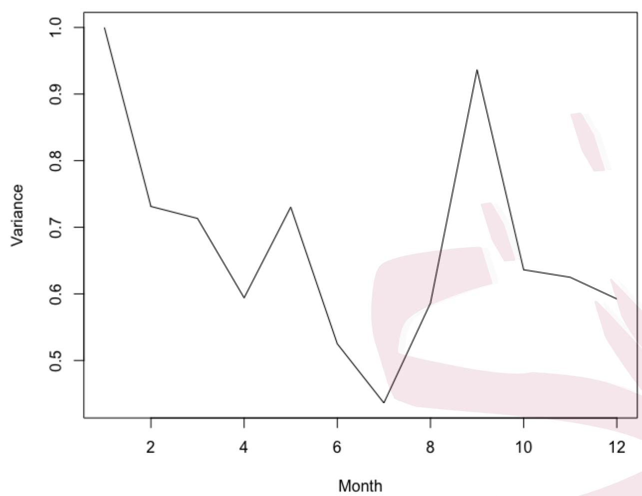

Figure 11: Estimates of the variance term for different months

4.4 Model diagnostics

In this section, we discuss the behavior of the residuals to show that the proposed model fares well to capture the effects of Taiwan $PM_{2.5}$ data. Standard statistical procedures reveal that there is no particular pattern in the residuals, thereby establishing that the residuals can be assumed to be uncorrelated. Further, for different months, we found that the standardized residuals are more or less evenly distributed across the months, thereby showing that the issue of heteroskedasticity has been taken care of. These plots are available in the online supplement.

Next, the left panel of Figure 12 shows that the histogram is approximately normally shaped, while the QQ plot presented in the right panel of the same figure corroborates that as well. Consequently, the introduction of heteroskedasticity and the spatio-temporal process works well for the data under study.

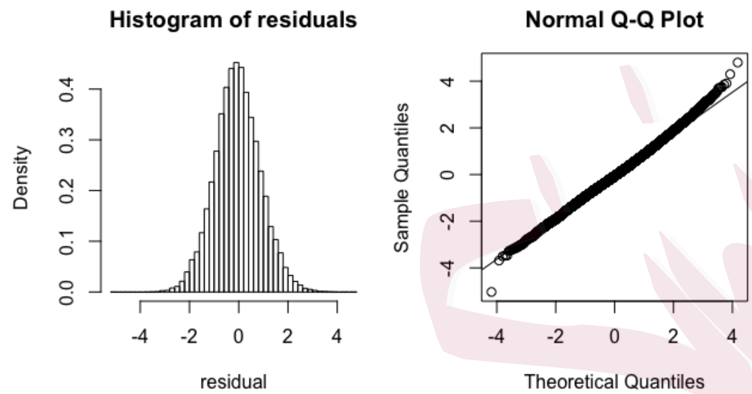

Figure 12: (Left) Histogram and (Right) QQ plot of the standardized residuals

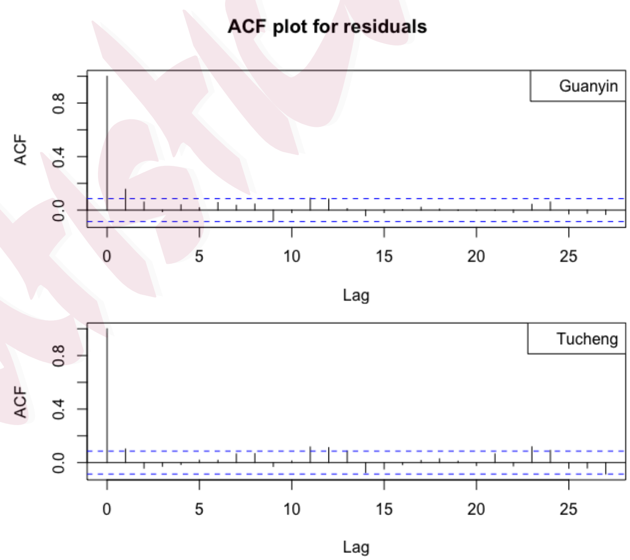

Figure 13: ACF plots of the standardized residuals, corresponding to two sample stations.

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

4.5 Prediction

As a final piece, we check the prediction abilities 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 the stations. This way, we get detailed idea about out-of-sample prediction performances, in both temporal and spatial sense.

In order to evaluate how good the predictions are, we have calculated the root mean squared error for the observations, and it is approximately 6.785. In order to understand the effect of the space-time interaction term in our model, we also measured the prediction abilities of the same model, but without the interaction component. In that case, the root mean squared error was found out to be approximately 37% more than the root mean squared error for 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, and that is because of our weighted scheme in the method. However, the forecasts are usually inside the prediction interval and it shows that the method works well for prediction purposes.

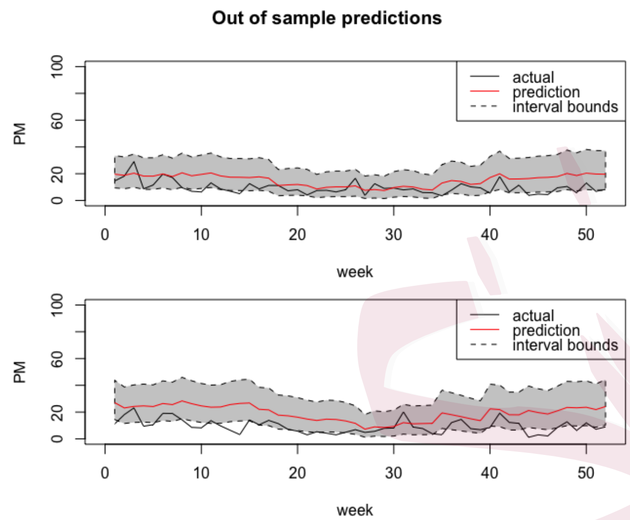

Figure 14: True value and predicted $PM_{2.5}$ observations, along with prediction interval, for two stations and for one year.

5. Conclusion

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

We finish with some notes on 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, and that condition can be relaxed to address a more general setup. Moreover, in the aforementioned data analysis example, it was found that maximum wind speed at every location plays an important role in the air pollution, while there is significant space-time interaction effect as well. Combined, they raise an important question - how much effect does the wind flux, calculated from the wind speed, wind direction and the coordinates of different locations, have on the pollution? In order to address that, it will be necessary to know about the physical behavior of the wind, and that process can be incorporated in 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, and it is immensely important to use appropriate modeling techniques for such problems. It has the potential to impact health sciences significantly.

References

Fahimah A Al-Awadhi and Shafiqah A Al-Awadhi. Spatial-temporal model for ambient air pollutants in the state of kuwait. *Environmetrics*, 17(7):739–752, 2006.

Veronica J Berrocal, Alan E Gelfand, and David M Holland. A bivariate space-time downscaler under space and time misalignment. *The annals of applied statistics*, 4(4):1942, 2010.

Richard A Bilonick. The space-time distribution of sulfate deposition in the northeastern united states. *Atmospheric Environment (1967)*, 19(11):1829–1845, 1985.

Marta Blangiardo, Francesco Finazzi, and Michela Cameletti. 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, 2016.

Michela Cameletti, Rosaria Ignaccolo, and Stefano Bande. Comparing spatio-temporal models for particulate matter in piemonte. *Environmetrics*, 22(8):985–996, 2011.

RJ Carroll, R Chen, EI George, TH Li, HJ Newton, H Schmiediche, and N Wang. Ozone exposure and population density in harris county, texas. *Journal of the American Statistical Association*, 92(438):392–404, 1997.

DM Elsom. Spatial correlation analysis of air pollution data in an urban area. *Atmospheric Environment (1967)*, 12(5):1103–1107, 1978.

Barrett P Eynon and Paul Switzer. The variability of rainfall acidity. *Canadian Journal of Statistics*, 11(1):11–23, 1983.

Arthur S Goldberger. Best linear unbiased prediction in the generalized linear regression model. *Journal of the American Statistical Association*, 57(298):369–375, 1962.

Peter Guttorp, Wendy Meiring, and Paul D Sampson. A space-time analysis of ground-level ozone data. *Environmetrics*, 5(3):241–254, 1994.

Timothy C Haas. Local prediction of a spatio-temporal process with an application to wet sulfate deposition. *Journal of the American Statistical Association*, 90(432):1189–1199, 1995.

Michael Jerrett, Richard T Burnett, Bernardo S Beckerman, Michelle C Turner, Daniel Krewski, George Thurston, Randall V Martin, Aaron van Donkelaar, Edward Hughes, Yuanli Shi, et al. Spatial analysis of air pollution and mortality in california. *American journal of respiratory and critical care medicine*, 188(5):593–599, 2013.

BM Golam Kibria, Li Sun, James V Zidek, and Nhu D Le. Bayesian spatial prediction of random space-time fields with application to mapping pm2. 5 exposure. *Journal of the American Statistical Association*, 97(457):112–124, 2002.

Martin Kulldorff and Ulf Hjalmars. The knox method and other tests for space-time interaction. *Biometrics*, 55(2):544–552, 1999.

Phaedon C Kyriakidis and André G Journel. Geostatistical space–time models: a review. *Mathematical geology*, 31(6):651–684, 1999.

Pierre Legendre, Miquel De Cáceres, and Daniel Borcard. Community surveys through space and time: testing the space–time interaction in the absence of replication. *Ecology*, 91(1):262–272, 2010.

Helmut Mayer. Air pollution in cities. *Atmospheric environment*, 33(24):4029–4037, 1999.

Sebastian Meyer, Ingeborg Warnke, Wulf Rössler, and Leonhard Held. 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, 2016.

C Arden Pope III, Michael J Thun, Mohan M Namboodiri, Douglas W Dockery, John S Evans, Frank E Speizer, and Clark W Heath Jr. Particulate air pollution as a predictor of mortality in a prospective study of us adults. *American journal of respiratory and critical care medicine*, 151 (3_pt_1):669–674, 1995.

Shahrokh Rouhani and Timothy J Hall. Space-time kriging of groundwater data. In *Geostatistics*, pages 639–650. Springer, 1989.

Sujit K Sahu and Kanti V Mardia. A bayesian kriged kalman model for short-term forecasting of air pollution levels. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(1): 223–244, 2005.

Sujit K Sahu, Alan E Gelfand, and David M Holland. Spatio-temporal modeling of fine particulate matter. *Journal of Agricultural, Biological, and Environmental Statistics*, 11(1):61–86, 2006.

Sujit K Sahu, Stan Yip, and David M Holland. Improved space–time forecasting of next day ozone concentrations in the eastern us. *Atmospheric Environment*, 43(3):494–501, 2009.

Gavin Shaddick and Jon Wakefield. Modelling daily multivariate pollutant data at multiple sites. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 51(3):351–372, 2002.

Richard L Smith, Stanislav Kolenikov, and Lawrence H Cox. Spatiotemporal modeling of pm2. 5 data with missing values. *Journal of Geophysical Research: Atmospheres*, 108(D24), 2003.

Li Sun, James V Zidek, Nhu D Le, and Halûk Özkaynak. Interpolating vancouver's daily ambient pm10 field. *Environmetrics*, 11(6):651–663, 2000.

George D Thurston, Richard T Burnett, Michelle C Turner, Yuanli Shi, Daniel Krewski, Ramona Lall, Kazuhiko Ito, Michael Jerrett, Susan M Gapstur, W Ryan Diver, et al. Ischemic heart disease mortality and long-term exposure to source-related components of us fine particle air pollution. *Environmental Health Perspectives (Online)*, 124(6):785, 2016.

Erik Vanem, Arne Bang Huseby, and Bent Natvig. 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(2):189–220, 2014.

Vikram M Vyas and George Christakos. Spatiotemporal analysis and mapping of sulfate deposition data over eastern usa. *Atmospheric Environment*, 31(21):3623–3633, 1997.

Christopher K Wikle, L Mark Berliner, and Noel Cressie. Hierarchical bayesian space-time models. *Environmental and Ecological Statistics*, 5(2):117–154, 1998.

Department of Statistics, University of Chicago

E-mail: sdeb@uchicago.edu

University of Chicago Booth School of Business

E-mail: ruey.Tsay@chicagobooth.edu