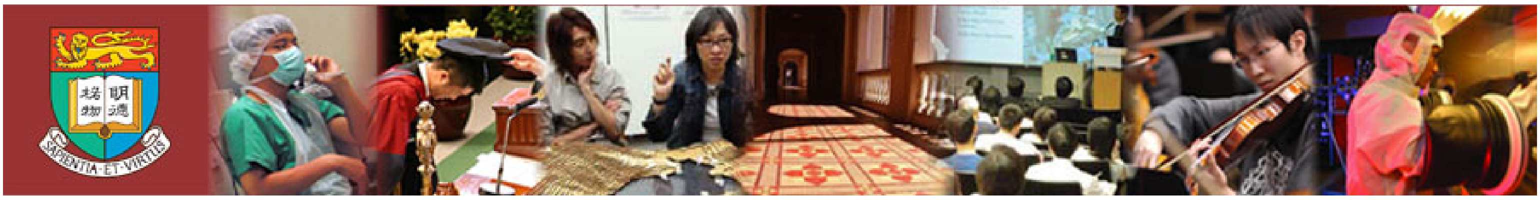
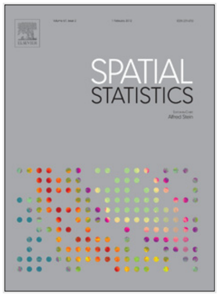


Accepted Manuscript



Time varying spatio-temporal covariance models

Ryan H.L. Ip, W.K. Li

PII: S2211-6753(15)00054-8
DOI: http://dx.doi.org/10.1016/j.spasta.2015.06.006
Reference: SPASTA 113

To appear in: *Spatial Statistics*

Received date: 12 February 2015
Accepted date: 22 June 2015

Please cite this article as: Ip, R.H.L., Li, W.K., Time varying spatio-temporal covariance models. *Spatial Statistics* (2015), http://dx.doi.org/10.1016/j.spasta.2015.06.006

Time Varying Spatio-Temporal Covariance Models

Ryan H.L. IP[1,*], W.K. LI

Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam, Hong Kong

Abstract

In this paper, we introduce valid parametric covariance models for univariate and multivariate spatio-temporal random fields. In contrast to the traditional models, we allow the model parameters to vary over time. Since variables in applications usually exhibit seasonality or changes in dependency structures, the allowance of varying parameters would be beneficial in terms of improving model flexibility. Conditions in constructing valid covariance models and discussions on practical implementation will be provided. As an application, a set of air pollution data observed from a monitoring network will be modeled. It is found that the time varying model performs better in prediction compared with the traditional models.

Keywords: Valid covariance models, Multivariate processes, Prediction, Monitoring datasets

[*]Correspondence for: R.H.L. Ip, Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam, Hong Kong.

[1]Email: ryanhlip@hku.hk

1. Introduction

Spatial or spatio-temporal models have been proven to be useful in a variety of fields including environmetrics, hydrology, economics, among many others. One of the most important parts in spatial and spatio-temporal analysis is the modeling of the covariance function. A function is said to be a covariance function if the matrix defined by the covariance function is valid for all finite sets of locations and times. A covariance matrix is said to be valid if and only if it is positive semi-definite (p.s.d.). Recall that a matrix $\boldsymbol{A}$ is said to be p.s.d. if and only if $\boldsymbol{a}^\top \boldsymbol{A}\boldsymbol{a} \geq 0$ for all vectors $\boldsymbol{a}$ of comfortable dimensions. One way to guarantee the positive semi-definiteness is to define the covariance matrix based on some positive definite functions and make use of the celebrated Bochner's theorem (Bochner, 1955), see also Chilès and Delfiner (1999, Ch. 2). Over the past few decades, many authors introduced different kinds of valid parametric covariance models. For spatial models, the covariance is usually modeled in the form of $\mathrm{Cov}(X(\boldsymbol{s}_1), X(\boldsymbol{s}_2))$ where $\boldsymbol{s}$ is the location where X is observed. Interested readers may consult Cressie (1993) and Finkenstädt et al. (2007) for further details. For spatio-temporal models, the situation is more challenging. Traditionally, scholars built valid spatio-temporal covariance models based on the assumption that the spatial and time components are separable. Recall that a spatio-temporal covariance model is called separable if $\mathrm{Cov}(X(\boldsymbol{s}_1, t_1), X(\boldsymbol{s}_2, t_2))$ can be written as $C_S(\boldsymbol{s}_1, \boldsymbol{s}_2)C_T(t_1, t_2)$, a product of a purely spatial covariance function C_S and a purely temporal covariance function C_T. A review regarding separable models can be found in Kyriakidis and Journel (1999) and an application of separable model can be found in Rodríguez-Iturbe and Mejía (1974). The main drawback of separable models is the disallowance of the space-time interaction which leads to undesirable properties in some occasion. Hence, literature concerning non-separable covariance models appeared. Cressie and Huang (1999) introduced some classes of valid non-separable spatio-temporal covariance models. Based on the results of Cressie and Huang (1999), Gneiting (2002) introduced other classes of valid models based on completely monotonic functions. Other works include De Iaco et al. (2002), Stein (2005) and Fuentes et al. (2008), among many others. For non-separable anisotropic (depending on the directions) spatio-temporal models, see Porcu et al. (2006). We note that under our approach, the resulting spatio-temporal covariance is non-separable in general.

In the above works, in order to achieve validity, covariance parameters were assumed to be fixed both spatially and temporally. But such an assumption is clearly unnecessarily restrictive. Relaxation of the constant parameter assumption will surely be beneficial since it enhances the model flexibility. Under the purely spatial setting, Gelfand et al. (2003) attempted to include spatially varying coefficients in their models under the Bayesian framework. For the multivariate spatial settings, some details can be found in Gelfand et al. (2003), Gelfand et al. (2004) and Kleiber and Genton (2013). Our work is closely related to Kleiber and Genton (2013). In Kleiber and Genton (2013), they introduced the spatial covariance models for multivariate spatial processes which are spatial varying. Analogously, one could consider spatio-temporal processes as multivariate spatial processes, with each time point regarded as a component from the multivariate process. The temporal correlation in our work can be analogous to the cross-covariance correlation in their work. Nevertheless, we must emphasize the difference between our work and Kleiber and Genton (2013). First, forecasting in time, which cannot be done in their models, can be easily done under our proposed models. Second, in Kleiber and Genton (2013), estimation of parameters were done under non-parametric methods. In our work, full parametric methods will be employed. Under full parametric methods, predictions can be done using classical methods. In later parts, we will compare the time varying models with an ordinary separable model in terms of the predictive powers.

The rest of this paper is organized as follows. In Section 2, details for the univariate case are provided while the results for the multivariate case are given in Section 3. In Section 4, the empirical coverage rates of confidence intervals are assessed via a simulation study. In Section 5, we applied the models to a set of trivariate air pollution data recorded in California. Conclusions and discussions are provided in the last section.

2. Univariate Time Varying Spatio-Temporal Covariance Models

2.1. Main Results

Consider the spatio-temporal random process

$$\boldsymbol{X} = \left(X\left(\boldsymbol{s}_1, t_1\right), X\left(\boldsymbol{s}_2, t_1\right), \ldots, X\left(\boldsymbol{s}_m, t_1\right), \ldots, X\left(\boldsymbol{s}_1, t_T\right), \ldots, X\left(\boldsymbol{s}_m, t_T\right)\right)^\top$$

containing time series of length T at each of the m locations. In practice the sites $\boldsymbol{s}_i \in \mathbb{R}^d$ for $d \leq 3$. We focus on the modeling of the covariance of $\boldsymbol{X}$ and therefore, throughout the whole work, it is assumed that the mean of $\boldsymbol{X}$ is $\mathbf{0}$. Note that the assumption is not restrictive since in practice one can always subtract the original data by the sample mean to remove the mean component. The main objective here is to introduce valid temporally varying spatio-temporal covariance models which are computationally estimable. Let

$$\mathrm{Var}\,(\boldsymbol{X}) = \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{C}_{11} & \boldsymbol{C}_{12} & \cdots & \boldsymbol{C}_{1T} \\ & \boldsymbol{C}_{22} & & \vdots \\ & & \ddots & \vdots \\ & & & \boldsymbol{C}_{TT} \end{pmatrix}. \tag{1}$$

In (1), $m \times m$ block matrices $\boldsymbol{C}_{k\ell}$ constitute the large matrix $\boldsymbol{\Sigma}$ of dimension $mT \times mT$. When $k = \ell$ (i.e., the diagonal blocks), each $\boldsymbol{C}_{kk}$ is a (possibly different) spatial covariance matrix governing the spatial dependency structure of the random process at time k. When $k \neq \ell$ (i.e., the off-diagonal blocks), $\boldsymbol{C}_{kl}$ are covariance matrices capturing the temporal association of the process between different time points k and ℓ. Under classical approaches, $\boldsymbol{\Sigma}$ is modeled using a parametric spatio-temporal covariance model with the assumption that $\boldsymbol{C}_{11} = \boldsymbol{C}_{22} = \cdots = \boldsymbol{C}_{TT}$ and all the $\boldsymbol{C}_{k\ell}$ are the same when the differences between k and ℓ are equal. However, empirical evidence (see Section 5) reveals that it maybe sometimes restrictive to assume constant parameters over time, especially when the variables on hand, such as environmental quantities, vary with time.

Under the new proposed approach, the only requirement posed on the diagonal blocks $\boldsymbol{C}_{kk}$ is the p.s.d. requirement. However, it should be noted that, in general, for $k \neq \ell$, the parametric forms of $\boldsymbol{C}_{kk}$ and $\boldsymbol{C}_{\ell\ell}$ can be different from each other. For example, $\boldsymbol{C}_{kk}$ can be in the Matérn class while $\boldsymbol{C}_{\ell\ell}$ can be in the Cauchy class. In addition, even they are in the same class, the parameters can be different from each other. For instance, the decaying parameter can be time varying. It is easy to see that if one assumes zero correlations across time, i.e., $\boldsymbol{C}_{k\ell} = \boldsymbol{0}$ when $k \neq \ell$, then $\boldsymbol{\Sigma}$ is always valid as long as all the diagonal blocks are p.s.d. (Horn and Johnson, 1990, Ch. 7). Hence, it remains to find the conditions for the off-diagonal blocks such that $\boldsymbol{\Sigma}$ is valid. In the following, for any square matrix $\boldsymbol{M}$, define $\boldsymbol{M}^{1/2}$ to be a square root matrix such that $\boldsymbol{M}^{1/2}\boldsymbol{M}^{1/2} = \boldsymbol{M}$. The following Lemma will be useful for further development of our proposed model.

Lemma 1. *Let $\boldsymbol{I}$ be the $m \times m$ identity matrix, for diagonal matrices $\boldsymbol{D}_{ij} = diag(d_{ij1}, \ldots, d_{ijm})$, $i, j = 1, \ldots, T$, the matrix*

$$\begin{pmatrix} \boldsymbol{I} & \boldsymbol{D}_{12} & \cdots & \boldsymbol{D}_{1T} \\ & \boldsymbol{I} & & \vdots \\ & & \ddots & \vdots \\ & & & \boldsymbol{I} \end{pmatrix} \tag{2}$$

is p.s.d. if all matrices of the form

$$\boldsymbol{D}_k = \begin{pmatrix} 1 & d_{12k} & \cdots & d_{1Tk} \\ & 1 & & \vdots \\ & & \ddots & \vdots \\ & & & 1 \end{pmatrix} \tag{3}$$

for $k = 1, \ldots, m$ are p.s.d..

Proof. By rearranging the rows and columns, (2) can be written in the block diagonal form as

$$\begin{pmatrix} \boldsymbol{D}_1 & 0 & \cdots & 0 \\ 0 & \boldsymbol{D}_2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{D}_m \end{pmatrix}.$$

Therefore, if all $\boldsymbol{D}_k$, $k = 1, \ldots, m$, are p.s.d., (2) is p.s.d.. $\square$

The time varying spatio-temporal covariance models can be developed as specified in the theorem below.

Theorem 1. *Let $\boldsymbol{K}$ be an $mT \times mT$ block matrix with $T \times T$ blocks $\boldsymbol{K}_{k\ell}$, $k, \ell = 1, \dots, T$, such that $\boldsymbol{K}_{kk} = \boldsymbol{I}$ and $\boldsymbol{K}_{k\ell} = \boldsymbol{C}_{kk}^{-1/2} \boldsymbol{C}_{k\ell} \boldsymbol{C}_{\ell\ell}^{-1/2}$ for $k \neq \ell$. Then $\boldsymbol{\Sigma}$ is p.s.d. if and only if $\boldsymbol{K}$ is p.s.d..*

Proof. Denoted matrix congruence by $\sim$, using proposition 1.3.2 of Bhatia (2007), it can be shown that

$$
\begin{aligned}
\boldsymbol{\Sigma} &= \begin{pmatrix} \boldsymbol{C}_{11} & \boldsymbol{C}_{12} & \cdots & \boldsymbol{C}_{1T} \\ & \boldsymbol{C}_{22} & & \vdots \\ & & \ddots & \vdots \\ & & & \boldsymbol{C}_{TT} \end{pmatrix} \\
&\sim \begin{pmatrix} \boldsymbol{I} & \boldsymbol{C}_{11}^{-1/2} \boldsymbol{C}_{12} \boldsymbol{C}_{22}^{-1/2} & \cdots & \boldsymbol{C}_{11}^{-1/2} \boldsymbol{C}_{11} \boldsymbol{C}_{TT}^{-1/2} \\ & \boldsymbol{I} & & \vdots \\ & & \ddots & \vdots \\ & & & \boldsymbol{I} \end{pmatrix} = \boldsymbol{K}.
\end{aligned} \tag{4}
$$

Hence, $\boldsymbol{\Sigma}$ is p.s.d. if and only if $\boldsymbol{K}$ is p.s.d.. $\square$

Recall that two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are said to be congruent if there exists another matrix $\boldsymbol{C}$ such that $\boldsymbol{B} = \boldsymbol{C}^\top \boldsymbol{A} \boldsymbol{C}$, more details can be found in Bhatia (2007). Congruent matrices are equivalent in certain aspects. In particular, if a matrix is p.s.d., its congruent matrices are p.s.d. as well. Using Theorem 1, the following corollaries hold.

Corollary 1. *For $k \neq \ell$, let $\boldsymbol{C}_{k\ell} = \boldsymbol{C}_{kk}^{1/2} \boldsymbol{D}_{k\ell} \boldsymbol{C}_{\ell\ell}^{1/2}$ where $\boldsymbol{D}_{k\ell}$ are diagonal* *matrices with positive entries,* $\boldsymbol{\Sigma}$ *is p.s.d. if*

$$\begin{pmatrix} \boldsymbol{I} & \boldsymbol{D}_{12} & \cdots & \boldsymbol{D}_{1T} \\ & \boldsymbol{I} & & \vdots \\ & & \ddots & \vdots \\ & & & \boldsymbol{I} \end{pmatrix}$$

is p.s.d..

Proof. Corollary 1 follows directly from Corollary 2 of Kleiber and Genton (2013). □

Under the construction of Corollary 1, the ij-th element of $\boldsymbol{C}_{k\ell}$, denoted by $[\boldsymbol{C}_{k\ell}]_{ij}$, can be expressed as

$$[\boldsymbol{C}_{k\ell}]_{ij} = \sum_{r=1}^{m} c_{kk}^{ir} d_{k\ell r} c_{\ell\ell}^{rj}$$

where c_{st}^{uv} denotes the uv-th element of $\boldsymbol{C}_{st}^{1/2}$ and $d_{k\ell r}$ denotes the r-th diagonal element of $\boldsymbol{D}_{k\ell}$. It is essentially a weighted sum of the products of the elements from the square root matrices. Under Corollary 2 below, when $d_{k\ell r}$ represents temporal correlations, $[\boldsymbol{C}_{k\ell}]_{ij}$ is a mixture of temporal correlation and spatial covariances at different times.

Corollary 2. *For* $k \neq \ell$*, let* $\boldsymbol{D}_{k\ell}$ *as defined in Corollary 1 and let* $d_{k\ell n}$ *be the* n*-th diagonal element of* $\boldsymbol{D}_{k\ell}$ *for* $n = 1, \ldots, m$*. If* $d_{k\ell n} = g(k, \ell)$ *where* $g(\cdot, \cdot)$ *is a valid purely temporal correlation function, then* $\boldsymbol{\Sigma}$ *is p.s.d..*

Proof. When $d_{k\ell n} = g(k, \ell)$, $\boldsymbol{D}_{k\ell} = \mathrm{diag}(g(k, \ell), \ldots, g(k, \ell))$. Hence, from (3), $\boldsymbol{D}_1 = \cdots = \boldsymbol{D}_m$. Since g is a valid correlation function, it implies that all $\boldsymbol{D}_i, i = 1, \ldots, m$ are p.s.d.. Therefore, Corollary 2 follows from Lemma 1 and Corollary 1. □

Remark 1. The purely temporal function g in Corollary 2 need not be isotropic or stationary. Indeed, any valid positive definite functions can be used. In particular, a valid stationary purely temporal positive definite function satisfies $f_\omega = \int_{\mathbb{R}} e^{-iu\omega} g\left(u\right) \mathrm{d}u \geq 0$ for all values of $u = |k - \ell| \in \mathbb{R}$, see Bochner (1955).

2.2. Estimation

Before proceeding to parameter estimation, one has to choose the spatial as well as the temporal covariance/ correlation models. For instance, one can choose the Matérn covariance model (Matérn, 1986) as the spatial covariance model, i.e., the ij-th element of the matrix $\boldsymbol{C}_{kk}$, $i, j = 1, \ldots, m$, $k = 1, \ldots, T$ is given by

$$\sigma_k^2 \frac{\left(\alpha_k \left\|\boldsymbol{s}_i - \boldsymbol{s}_j\right\|\right)^{\nu_k}}{\Gamma\left(\nu_k\right) 2^{\nu_k - 1}} \mathcal{K}_{\nu_k}\left(\alpha_k \left\|\boldsymbol{s}_i - \boldsymbol{s}_j\right\|\right), \quad \nu_k, \alpha_k, \sigma_k^2 > 0 \tag{5}$$

where $\mathcal{K}_{\nu_k}$ is the modified Bessel function of the second kind, see Stein (1999) and Abramowitz and Stegun (1972) for further details about the Matérn model. The Matérn model has been used and discussed extensively in spatial and spatio-temporal literature, examples include Fernández-Casal et al. (2003); Christakos (2000); Cressie and Wikle (2011); Matheron (1962). Note that under the current setting, the parameters are allowed to be varying in time. For the temporal correlation models, one can, for example, choose the third entry in Table 1 of Gneiting (2002), i.e., the diagonal elements of $\boldsymbol{D}_{k\ell}$ are given by

$$\left(1 + a|k - \ell|^{\gamma}\right)^{-b}, \quad a, b > 0, \quad 0 < \gamma \leq 1. \tag{6}$$

If the time varying spatio-temporal covariance model was built using (5) and (6), then the spatial parameters are $\boldsymbol{\theta}_S = (\alpha_1, \ldots, \alpha_T, \sigma_1^2, \ldots, \sigma_T^2)$ and the temporal parameters are $\boldsymbol{\theta}_T = (a, b)$.

In general, following the results given in the last section, any valid spatial covariance functions together with valid temporal correlation functions can be used to construct valid time varying spatio-temporal covariance functions. More examples can be found in Cressie and Huang (1999), Gneiting (2002) and Sherman (2011), among many others.

To estimate the model parameters, when distributional assumptions are imposed, likelihood methods are suggested. In Section 5 of this work, we have assumed that the data follow the normal distribution. If one does not impose any distributional assumption, least squares methods can be used. It is noted that estimation using full likelihood methods can lead to heavy computational burden when the dimensionality is high. Concerning the computational burden, one may use approximate likelihood methods as provided in Varin and Vidoni (2005) and Bevilacqua et al. (2012). It is suggested to estimate the spatial parameters at each time point first. Hence, fixing the spatial parameters, one can estimate the temporal parameters.

2.3. Prediction

Suppose one wishes to do interpolation, that is, predicting the value of X at an unobserved location $\boldsymbol{s}_0$ and time t_0 where $t_0 \in \{1, \ldots, T\}$, the interpolation can be done following the steps below:

1. Insert $\{X(\boldsymbol{s}_0, t_j)\}_{j=1,\ldots,T}$ into $\boldsymbol{X}$, call the new augmented observation vector $\boldsymbol{X}_*$, so that

$$\begin{aligned}\boldsymbol{X}_* \;=\; & \left(X\left(\boldsymbol{s}_1, t_1\right), \ldots, X\left(\boldsymbol{s}_m, t_1\right), X\left(\boldsymbol{s}_0, t_1\right), \ldots,\right. \\ & \left. X\left(\boldsymbol{s}_1, t_T\right), \ldots, X\left(\boldsymbol{s}_m, t_T\right), X\left(\boldsymbol{s}_0, t_T\right)\right)^\top .\end{aligned}$$

2. Denote by $\hat{\boldsymbol{\theta}}_S$ and $\hat{\boldsymbol{\theta}}_T$ the estimated spatial and temporal parameters respectively, compute $\hat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}\left(\hat{\boldsymbol{\theta}}_S, \hat{\boldsymbol{\theta}}_T\right)$ as in (1). Similarly, compute $\hat{\boldsymbol{\Sigma}}_* = \hat{\mathrm{Var}}\left(\boldsymbol{X}_*\right)$.

3. Denote by k the position of $X\left(\boldsymbol{s}_0, t_0\right)$ in $\boldsymbol{X}_*$ and define $\boldsymbol{c}_*$ to be the mT vector with elements $\hat{\mathrm{Cov}}\left(X\left(\boldsymbol{s}_0, t_0\right), X\left(\boldsymbol{s}_i, t_j\right)\right)$, $i = 1, \ldots, m$ and $j = 1, \ldots, T$, extracted from the k-th row of $\hat{\boldsymbol{\Sigma}}_*$, then the predicted value of $X\left(\boldsymbol{s}_0, t_0\right)$ is given by $\hat{X}\left(\boldsymbol{s}_0, t_0\right) = \boldsymbol{\lambda}^\top \boldsymbol{X}$ where

$$\boldsymbol{\lambda}^\top = \left(\boldsymbol{c}_* + \mathbf{1}\frac{1 - \mathbf{1}^\top \hat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{c}_*}{\mathbf{1}^\top \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{1}}\right)^\top \hat{\boldsymbol{\Sigma}}^{-1} \tag{7}$$

and $\mathbf{1}$ is the column vector of 1's (Cressie and Wikle, 2011, p. 324).

Following Cressie and Wikle (2011), the squared prediction error can be computed as

$$e_p^2(s_0, t_0) = \hat{\mathrm{Cov}}\left(X\left(\boldsymbol{s}_0, t_0\right), X\left(\boldsymbol{s}_0, t_0\right)\right) - \boldsymbol{\lambda}^\top \boldsymbol{c}_* + \frac{1 - \mathbf{1}\hat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{c}_*}{\mathbf{1}\hat{\boldsymbol{\Sigma}}^{-1} \mathbf{1}}. \tag{8}$$

If Gaussianity is assumed, the 95% prediction interval can be constructed as

$$\left(\hat{X}\left(\boldsymbol{s}_0, t_0\right) - 1.96 e_p(s_0, t_0), \hat{X}\left(\boldsymbol{s}_0, t_0\right) + 1.96 e_p(s_0, t_0)\right). \tag{9}$$

Suppose instead of interpolation, one wishes to do forecasting at station $\boldsymbol{s}_0$ and time $t_0 = T + q$ for some $q > 0$. Without loss of generality, fix $q = 1$, the predicted value of $X\left(\boldsymbol{s}_0, t_0\right)$ can be obtained by some minor modification of the above steps:

1(a). Define a augmented observation vector $\boldsymbol{X}_{**}$, so that

$$\boldsymbol{X}_{**} = \left(X\left(\boldsymbol{s}_1, t_1\right), \ldots, X\left(\boldsymbol{s}_0, t_1\right), \ldots, X\left(\boldsymbol{s}_1, t_{T+1}\right), \ldots, X\left(\boldsymbol{s}_0, t_{T+1}\right)\right)^\top.$$

2(a). Predict the spatial parameters at time $T+1$ (see Remark 2, compute $\hat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}\left(\hat{\boldsymbol{\theta}}_S, \hat{\boldsymbol{\theta}}_T\right)$ and $\hat{\boldsymbol{\Sigma}}_{**} = \hat{\text{Var}}\left(\boldsymbol{X}_{**}\right)$.

3(a). Extract $\boldsymbol{c}_{**}$ from the last row of $\hat{\boldsymbol{\Sigma}}_{**}$ and $\hat{X}\left(\boldsymbol{s}_0, t_0\right)$ is given by (7) with $\boldsymbol{c}_{*}$ replaced by $\boldsymbol{c}_{**}$.

Remark 2. In Step 2(a), researchers can treat the estimated spatial parameters at different time points, $\hat{\boldsymbol{\theta}}_{S,k}$, $k = 1, \ldots, T$, as a time series and use various time series models such as autoregressive moving-average (ARMA) models (see, for example, Wei (2006, Ch. 5) and Brockwell and Davis (2009, Ch. 9)) to do the prediction.

3. Multivariate Time Varying Spatio-Temporal Covariance Models

In practice, researchers often consider more than one variable at a time. In this section, we aimed to construct the multivariate time varying spatio-temporal covariance models using similar techniques as given in section 2.

3.1. Main Results

Assume there are p variables on hand such that the multivariate spatio-temporal data set is $\underset{\sim}{\boldsymbol{X}} = \left(\boldsymbol{X}_1^\top, \boldsymbol{X}_2^\top, \ldots, \boldsymbol{X}_p^\top\right)^\top$ where

$$\boldsymbol{X}_j = \left(X_j\left(s_1, t_1\right), X_j\left(s_2, t_1\right), \ldots, X_j\left(s_m, t_1\right), \ldots, X_j\left(s_m, t_T\right)\right)^\top$$

is an $m \times T$ vector representing the j-th variable of interest. The variance covariance matrix of $\underset{\sim}{\boldsymbol{X}}$, $\text{Var}\left(\underset{\sim}{\boldsymbol{X}}\right)$, consists of p^2 large block matrices. For each block matrix, the dimension is $mT \times mT$, as given in (1). In matrix form, write

$$\mathrm{Var}\left(\underset{\sim}{\boldsymbol{X}}\right) = \underset{\sim}{\boldsymbol{\Sigma}} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \cdots & \boldsymbol{\Sigma}_{1p} \\ & \boldsymbol{\Sigma}_{22} & & \vdots \\ & & \ddots & \vdots \\ & & & \boldsymbol{\Sigma}_{pp} \end{pmatrix}$$

where the diagonal blocks $\boldsymbol{\Sigma}_{jj}$ denotes the spatio-temporal covariance matrix of the j-th variable, i.e., $\boldsymbol{\Sigma}_{jj} = \mathrm{Var}\,(\boldsymbol{X}_j)$, as defined in the (1). The off-diagonal block matrices $\boldsymbol{\Sigma}_{ij}$ are the cross-covariance matrices between $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$ for $i \neq j$. Using similar techniques as in the previous section, valid multivariate time varying spatio-temporal covariance models can be constructed.

Theorem 2. *Let $\underset{\sim}{\boldsymbol{K}}$ be a $pmT \times pmT$ block matrix with p^2 blocks, $\left\{\underset{\sim}{\boldsymbol{K}}_{ij}\right\}_{ij=1}^{p}$, such that $\underset{\sim}{\boldsymbol{K}}_{ii} = \boldsymbol{I}$ where I denotes the identity matrix and $\underset{\sim}{\boldsymbol{K}}_{ij} = \boldsymbol{\Sigma}_{ii}^{-1/2}\boldsymbol{\Sigma}_{ij}\boldsymbol{\Sigma}_{jj}^{-1/2}$ for $i \neq j$. Then $\underset{\sim}{\boldsymbol{\Sigma}}$ is p.s.d. if and only if $\underset{\sim}{\boldsymbol{K}}$ is p.s.d..*

Proof. Similar to Theorem 1, by matrix congruence, denoted by $\sim$, and proposition 1.3.2 of Bhatia (2007), it can be shown that

$$\begin{aligned} \underset{\sim}{\boldsymbol{\Sigma}} &= \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \cdots & \boldsymbol{\Sigma}_{1p} \\ & \boldsymbol{\Sigma}_{22} & & \vdots \\ & & \ddots & \vdots \\ & & & \boldsymbol{\Sigma}_{pp} \end{pmatrix} \\ &\sim \begin{pmatrix} \boldsymbol{I} & \boldsymbol{\Sigma}_{11}^{-1/2}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1/2} & \cdots & \boldsymbol{\Sigma}_{11}^{-1/2}\boldsymbol{\Sigma}_{1p}\boldsymbol{\Sigma}_{pp}^{-1/2} \\ & I & & \vdots \\ & & \ddots & \vdots \\ & & & \boldsymbol{I} \end{pmatrix} = \underset{\sim}{\boldsymbol{K}}. \end{aligned}$$

Hence, $\underset{\sim}{\boldsymbol{\Sigma}}$ is p.s.d. if and only if $\underset{\sim}{\boldsymbol{K}}$ is p.s.d.. $\square$

Remark 3. In fact, the multivariate covariance matrix $\underset{\sim}{\boldsymbol{K}}$ defined in Theorem 2 is very flexible in the sense that no particular form is imposed on each marginal covariance matrix $\boldsymbol{\Sigma}_{ii}$. In particular, one can construct $\boldsymbol{\Sigma}_{ii}$ using Theorem 1 and then construct $\underset{\sim}{\boldsymbol{\Sigma}}$ using Theorem 2. Under this construction, the resulting multivariate covariance matrix is also time varying. In addition, similar to the univariate case, we can construct $\boldsymbol{\Sigma}_{ij}$ using some specially designed diagonal matrices as follows.

Corollary 3. *For $i \neq j$, let $\boldsymbol{\Sigma}_{ij} = \boldsymbol{\Sigma}_{ii}^{1/2}\boldsymbol{S}_{ij}\boldsymbol{\Sigma}_{jj}^{1/2}$ where $\boldsymbol{S}_{ij}$ are diagonal matrices for $i, j = 1, \ldots, p$, $\underset{\sim}{\boldsymbol{\Sigma}}$ is p.s.d. if*

$$\begin{pmatrix} \boldsymbol{I} & \boldsymbol{S}_{12} & \cdots & \boldsymbol{S}_{1p} \\ & \boldsymbol{I} & & \vdots \\ & & \ddots & \vdots \\ & & & \boldsymbol{I} \end{pmatrix}$$

is p.s.d..

Proof. Similar to the proof of Corollary 1. □

Corollary 4. *Let $\boldsymbol{S}_{ij}$ as defined in Corollary 3 and S_{ijn} be the n-th diagonal element of $\boldsymbol{S}_{ij}$, $i, j = 1, \ldots, p$, $n = 1, \ldots, mT$, where $|S_{ijn}| \leq 1$ such that all matrices of the form*

$$\boldsymbol{S}_n = \begin{pmatrix} 1 & S_{12n} & \cdots & S_{1pn} \\ & 1 & & \vdots \\ & & \ddots & \vdots \\ & & & 1 \end{pmatrix}$$

are p.s.d., then $\underset{\sim}{\boldsymbol{\Sigma}}$ is p.s.d..

Proof. Similar to the proof of Corollary 2. □

Remark 4. For $p = 2$, it requires all $|S_{12n}| \leq 1$, $n = 1, \ldots, mT$. For $p = 3$, it requires all matrices of the form

$$\begin{pmatrix} 1 & S_{12n} & S_{13n} \\ & 1 & S_{23n} \\ & & 1 \end{pmatrix}$$

are p.s.d. which may be difficult to check in practice, especially when mT is large. However, if we assume $S_{ijn} = \rho_{ij}$ where $|\rho_{ij}| \leq 1$ represents the cross-correlation coefficient, then the validity of $\underset{\sim}{\boldsymbol{\Sigma}}$ can be guaranteed. Yet, under such a specification, the model would be simplified to a proportional coregionalization model (Wackernagel, 2003, Ch. 26).

3.2. Estimation

To estimate the multivariate model, it is suggested to first estimate the marginal covariance models $\boldsymbol{\Sigma}_{ii}$, $i = 1, \ldots p$, using the procedures as described in Section 2.2. The cross-variable parameters can be estimated using likelihood or least squares methods similarly as in the univariate case. In the next section, we estimated the multivariate model using the method as described in Remark 4, i.e., we set $S_{ijn} = \rho_{ij}$ where $|\rho_{ij}| \leq 1$.

4. Simulation Study

In this section, the empirical coverage of the confidence intervals, arising from Equations (8) and (9), are assessed. In (9), although the nominal coverage is 95%, the empirical coverage could be quite different when the true

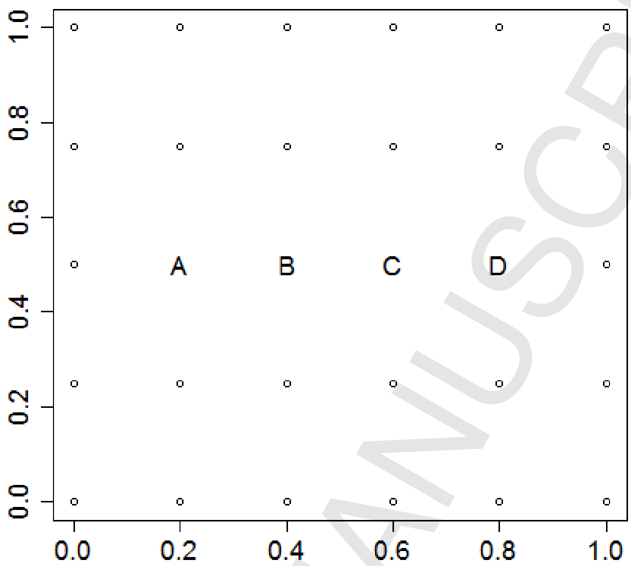


Figure 1: Sampling locations used in the simulation study (circles). Alphabetical letters represent locations used for interpolation.

parameters are replaced by the estimated ones, as suggested in Zimmerman and Cressie (1992).

In this simulation study, the spatial domain is set to be a two-dimensional regular grid in $[0, 1]^2$, as shown in Figure 1. Four sites (named as A, B, C and D) are reserved for the purpose of interpolation. The number of sites used in the modelling stage is therefore 26. The number of time points is 30.

Throughout the simulation study, the Matérn model (5) is used as the spatial covariance function $\boldsymbol{C}_{kk}$ and the temporal correlation model (6) is employed to construct the off-diagonal blocks $\boldsymbol{C}_{k\ell}$ following Corollary 2. The parameters are fixed as $\nu = 1.5$, $a = 0.3$ and $\gamma = 1$. Both α and σ are allowed to vary with time. The time series models are respectively:

$$\begin{aligned} \alpha_k &= 2 + 0.5\alpha_{k-1} + \varepsilon_k \\ \sigma_k^2 &= 4 + 0.8\sigma_{k-1}^2 - 0.5\sigma_{k-2}^2 + \varepsilon_k \end{aligned}$$

where ε_k are independent standard normal random variables representing the white noise series. In order to compare the effect of temporal correlation on the empirical coverage of confidence intervals, b takes the values 50 and 200. The temporal correlation is weaker when b is larger. For each set of parameters, 500 independent realizations are simulated using the MASS package in R (Venables and Ripley, 2002). The realizations are assumed to follow a mean zero Gaussian process with covariogram defined using Corollary 2.

For each independent replicate, Steps 1 to 3 listed in Section 2.3 are implemented to perform interpolation from time 1 to 30 at the four reserved sites. Table 1 summarizes the empirical coverages of the confidence intervals and the mean squared prediction errors. It can be seen that the empirical coverage rates are close to the nominal level 95%. Although the empirical coverage seemed to be closer to 95% when the temporal correlation is weaker, the differences are slight. Regarding the prediction errors, comparing the MSPE under the two values of b, the strengths of temporal correlations seemed to produce negligible effects.

5. Application

5.1. Data

As an application, we made use of an air pollution dataset recorded by the California Air Resources Board which is available online. The dataset

Table 1: Empirical coverage rates (ECR) of nominal 95% confidence intervals, expressed as percentages, and the mean squared prediction errors (MSPE).

	$b = 50$		$b = 200$	
Site	ECR	MSPE	ECR	MSPE
A	93.93	0.1728	94.05	0.1694
B	94.03	0.2184	94.15	0.2175
C	93.85	0.2216	93.71	0.2237
D	93.88	0.1711	94.13	0.1697

consists of daily averages of nitrogen dioxide (NO2), nitrogen oxide (NO) and carbon monoxide (CO) observed from September and October 2010 (61 days). We have removed the stations which contain missing values of at least one variable. Finally, 31 sites were found to have full record during the period and were retained in the dataset. In other words, the spatio-temporal dataset consists of 1891 observations for each variable. Apanasovich and Genton (2010) and Schmidt and Gelfand (2003) studied a similar air pollution dataset in the context of multivariate spatial modeling. To investigate the predictive power of the proposed time varying models, nine extra sites that contain no missing value over the period of interpolation and forecast are chosen to compare the performance of prediction under different models. The 31 sampling sites and the nine extra stations are shown in Figure 2.

In order to achieve approximate normality, logarithm transformation was made, as suggested by Schmidt and Gelfand (2003). Micro-scale effects were removed using ANOVA considering each site as a factor. After that, we standardized each time series at each site with the respective empirical mean and standard deviation for numerical computational stability. Finally, we divided the distance between stations and the time lags by their corresponding maximum values so that the sampling region becomes $[0,1]^2 \times [0,1]$. The estimation procedure can be summarized as follows.

Before proceeding to do estimation for the multivariate process, we estimated the marginal parameters first. For each variable, we assumed that the observations came from a mean-zero multivariate normal distribution with covariance matrix $\mathbf{\Sigma}$ as given in (1). For simplicity, we assumed the spatial covariance functions at each time point are in the same class but the parameters could be different. We used the Matérn covariance function (5) to model the data set on hand owing to its flexibility. For each time $k = 1, \ldots, T$, we employed the profile likelihood approach as given in Zhang (2004) to obtain $\hat{\alpha}_k$ and $\hat{\sigma}_k^2$. Hence, we set $\boldsymbol{C}_{k\ell} = \boldsymbol{C}_{kk}^{1/2} \boldsymbol{D}_{k\ell} \boldsymbol{C}_{\ell\ell}^{1/2}$, $k, \ell = 1, \ldots, T$, where $\boldsymbol{D}_{k\ell}$ are diagonal matrices defined by the temporal correlation function (6) with $\gamma = 1$. The temporal parameters were estimated through a likelihood approach with the spatial parameters fixed as described in Section 2.2. With the estimates of the marginal parameters, we proceeded to do estimation for the multivariate process. As described in Section 3.2, we fixed the marginal parameters and estimate ρ_{ij}, $i, j \in \{\mathrm{NO2}, \mathrm{NO}, \mathrm{CO}\}$.

5.2. Estimation Results

For the spatial covariance functions, we assume the values of ν are fixed. For better comparison, we simply use the estimated values of ν under the separable model as provided in Ip (2015, Ch. 2). The estimates of the spatial parameters for the three variables are provided in Figure 3. The fluctuations of both $\hat{\alpha}$ and $\hat{\sigma}^2$ suggest that constant values of $\hat{\alpha}$ and $\hat{\sigma}^2$ may not

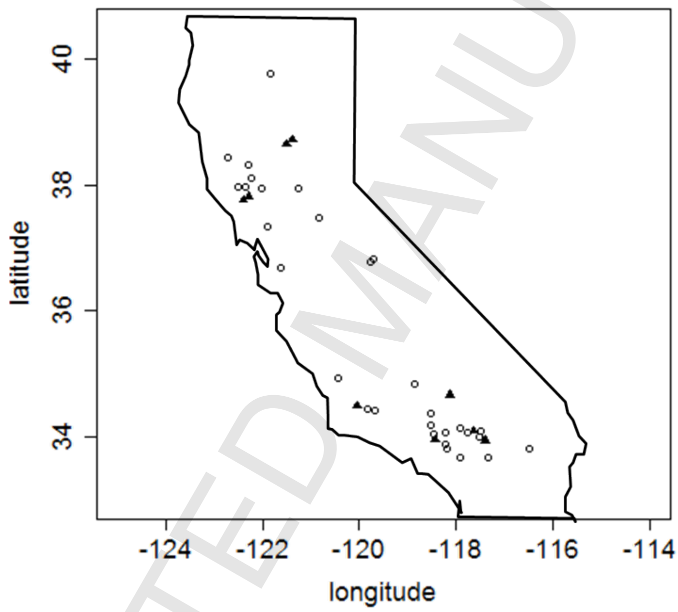


Figure 2: Map of sampling stations (circles) and the nine extra stations (triangles).

Table 2: Estimated parameters of the temporal correlation models for NO2, NO and CO.

	$\hat{a}$	$\hat{b}$
NO2	0.1591	300.00
NO	0.3198	192.00
CO	0.1523	295.11

provide accurate inferences, especially during times when jumps occurred. The estimated parameters $\hat{a}$ and $\hat{b}$ of the temporal correlation model are given in Table 2. For the multivariate models, the estimated parameters are $\hat{\rho}_{\mathrm{NO2,NO}} = 0.6618$, $\hat{\rho}_{\mathrm{NO2,CO}} = 0.6253$ and $\hat{\rho}_{\mathrm{NO,CO}} = 0.5965$. It can be easily check that the correlation matrix formed by $\hat{\rho}_{\mathrm{NO2,NO}}, \hat{\rho}_{\mathrm{NO2,CO}}$ and $\hat{\rho}_{\mathrm{NO,CO}}$ is positive definite, and therefore the multivariate time varying spatio-temporal covariance model is valid following Corollary 4. Throughout this section, we have demonstrated that the proposed time varying spatio-temporal covariance model can be applied to real life data. Next, we examine the predictive performance of the model.

5.3. Performance of Interpolation

To compare the proposed time varying model with ordinary models that assume fixed parameters, we compare the predictive performances of our proposed model with the classical approach in which the parameters are assumed to be constants. Under the classical approach, the variance-covariance matrix $\mathbf{\Sigma}$ is built from the separable spatio-temporal covariance function as given in Fuentes et al. (2008) and further discussed in Ip (2015, Ch. 2).

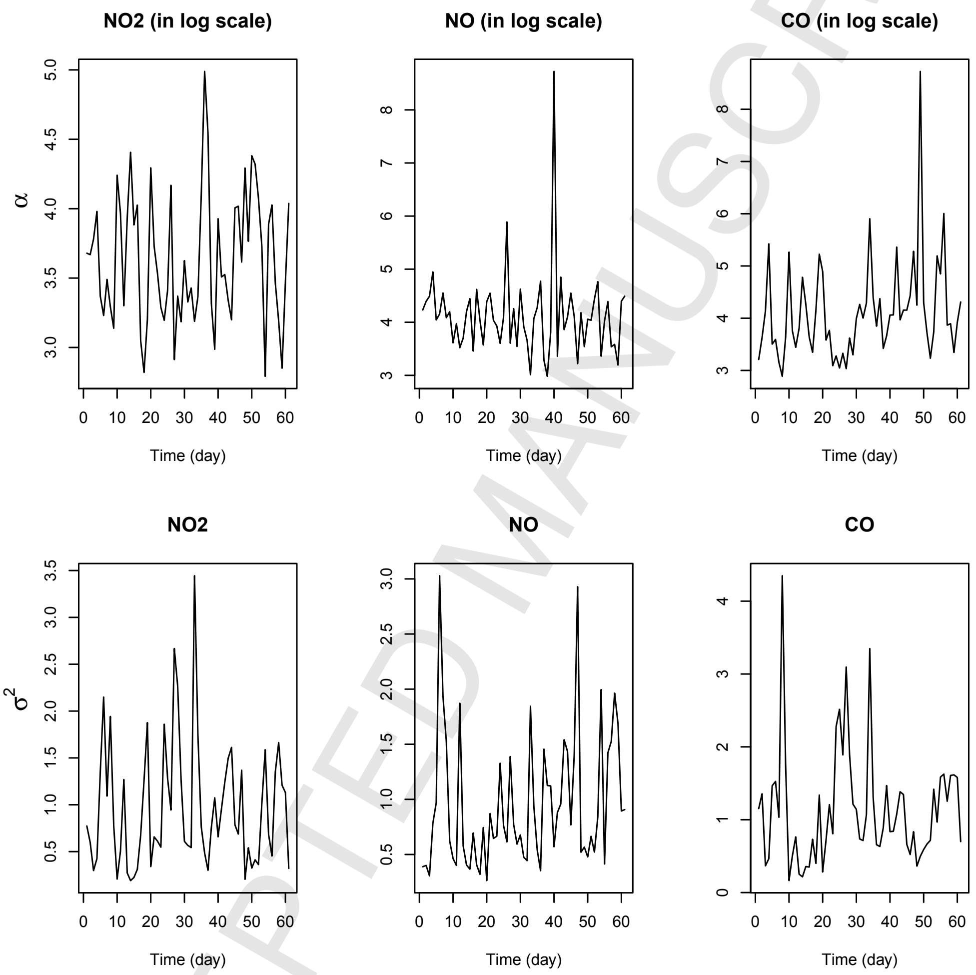


Figure 3: Estimates of α (in log scale, upper row) and σ^2 (lower row) from time 1 to time 61 for NO2 (left column), NO (middle column) and CO (right column).

Specifically, the covariogram is given as

$$\begin{aligned}&\operatorname{Cov}\left(X\left(\boldsymbol{s}_1, t_1\right), X\left(\boldsymbol{s}_2, t_2\right)\right) \\ =\ &\frac{\sigma^2 2^{2-2\nu+\frac{d+1}{2}}}{\Gamma\left(\nu-\frac{d}{2}\right)\Gamma\left(\nu-\frac{1}{2}\right)}\left(\alpha h\right)^{\nu-\frac{d}{2}}\left(\beta u\right)^{\nu-\frac{1}{2}}\mathcal{K}_{\nu-\frac{d}{2}}\left(\alpha h\right)\mathcal{K}_{\nu-\frac{1}{2}}\left(\beta u\right) \quad (10)\end{aligned}$$

where $h = \|\boldsymbol{s}_1 - \boldsymbol{s}_2\|$, $u = |t_1 - t_2|$, $\nu > d/2$ and $\alpha, \beta, \sigma^2 > 0$. The estimated parameters of the covariance function (10) can be found in Ip (2015, Ch. 2). Here, we discuss the performance of interpolation in this section and that of forecasting in the next section. In particular, interpolation was done using Steps 1 to 3 as described in Section 2.3 for the extra stations from day 24 to 33 (i.e., from 24 September to 3 October). The period was picked to minimize the number of missing values. Over the period of interpolation (and the period of forecast described in the next section), there is no missing value observed for NO2 in these nine stations. Unfortunately, for variables NO and CO, interpolation was performed only in three among these nine stations, owing to the presence of missing values in the remaining sites. Figures 4 and 5 show the plots of observed and predicted values of NO2 under the separable model (10) and time varying models respectively. In these figures, the solid lines represent the observed values while the dotted lines represent the interpolated values (from day 24 to 33) and the forecast values (from day 62 to 66). The prediction intervals (grey area) were calculated using (8) and (9). Figures 6 and 7 show the corresponding graphs for NO and CO respectively.

For NO2, as shown in Figure 5, the interpolation matches well with the observed data when time varying model is used. Compared with Figure 4, under the time varying models, the predictive intervals are usually narrower but are still able to cover most of the observed values. For NO and CO, it is hard to make judgments due to the limited number of stations. As a by-product of the time varying models, the predictive intervals are also temporally varying since the prediction errors (8) vary with time. Under the separable model (10), the widths of the predictive intervals are always fixed over time, which may not be sensible in practice.

The left panel of Table 3 summarizes the mean squared prediction error (MSPE) between the predicted values and the observed values at the extra stations. The MSPE at station $\boldsymbol{s}_{0i}$ is defined as

$$\frac{1}{10}\sum_{j=24}^{33}\left[X_p\left(\boldsymbol{s}_{0i},t_j\right)-\hat{X}_p\left(\boldsymbol{s}_{0i},t_j\right)\right]^2$$

where $\hat{X}_p\left(\boldsymbol{s}_{0i},t_j\right)$ denotes the predicted value of variable p at station $\boldsymbol{s}_{0i}$ and time t_j using either the separable model (10) or the proposed time varying model.

From the left panel of Table 3, it can be observed that the proposed time varying model performs better in 6 out of 9 stations in terms of smaller values of MSPE for NO2. The time varying models perform better in one and two stations for NO and CO respectively. Across all the variables and sites, the maximum reduction in MSPE is 79% and the median reduction is 17%.

5.4. Performance of Forecasting

As previously discussed in Section 2.3, it is possible to forecast the values of X at some unobserved locations. Take NO2 as an example, ARMA models were first fitted to each of the estimated spatial covariance parameter series $\hat{\alpha}_k$ and $\hat{\sigma}_k^2$, $k=1,\ldots,T$. The fitted time series models for α_k and σ_k^2 are

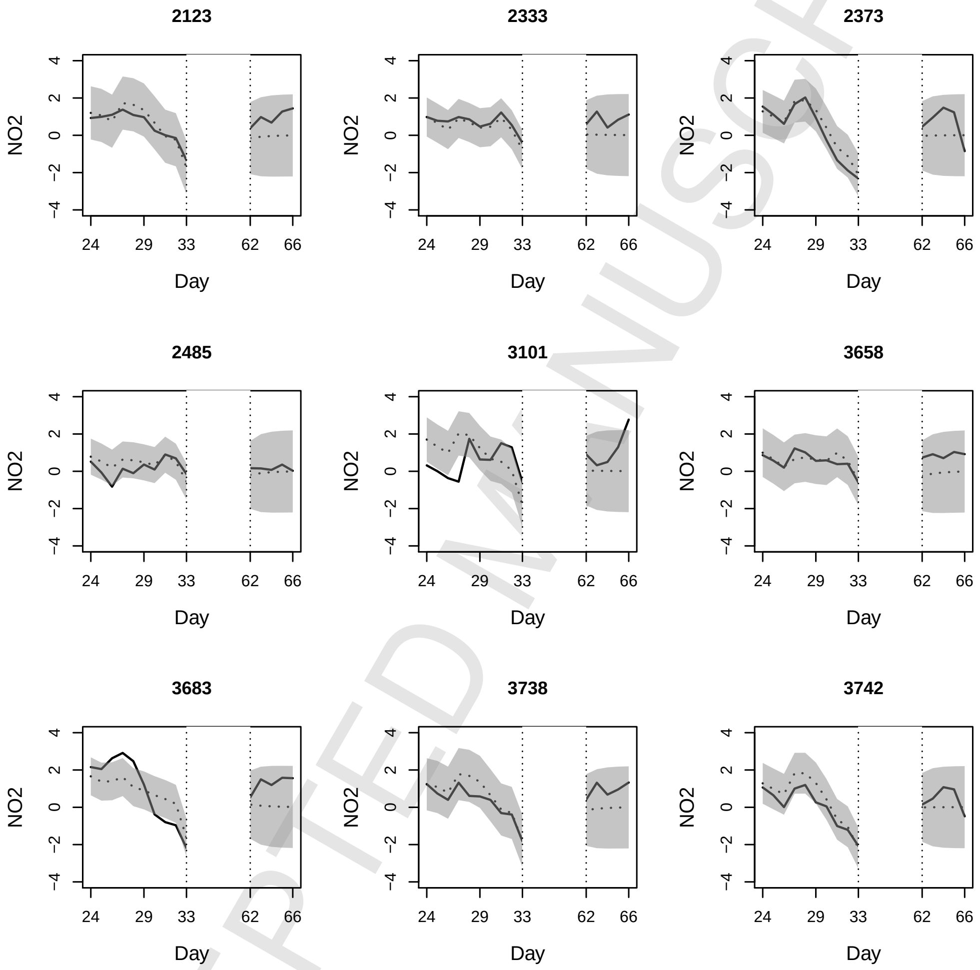


Figure 4: Plots of observed, interpolated and forecast values of NO2 under the separable model (10) at different sites. Refer to the text for details.

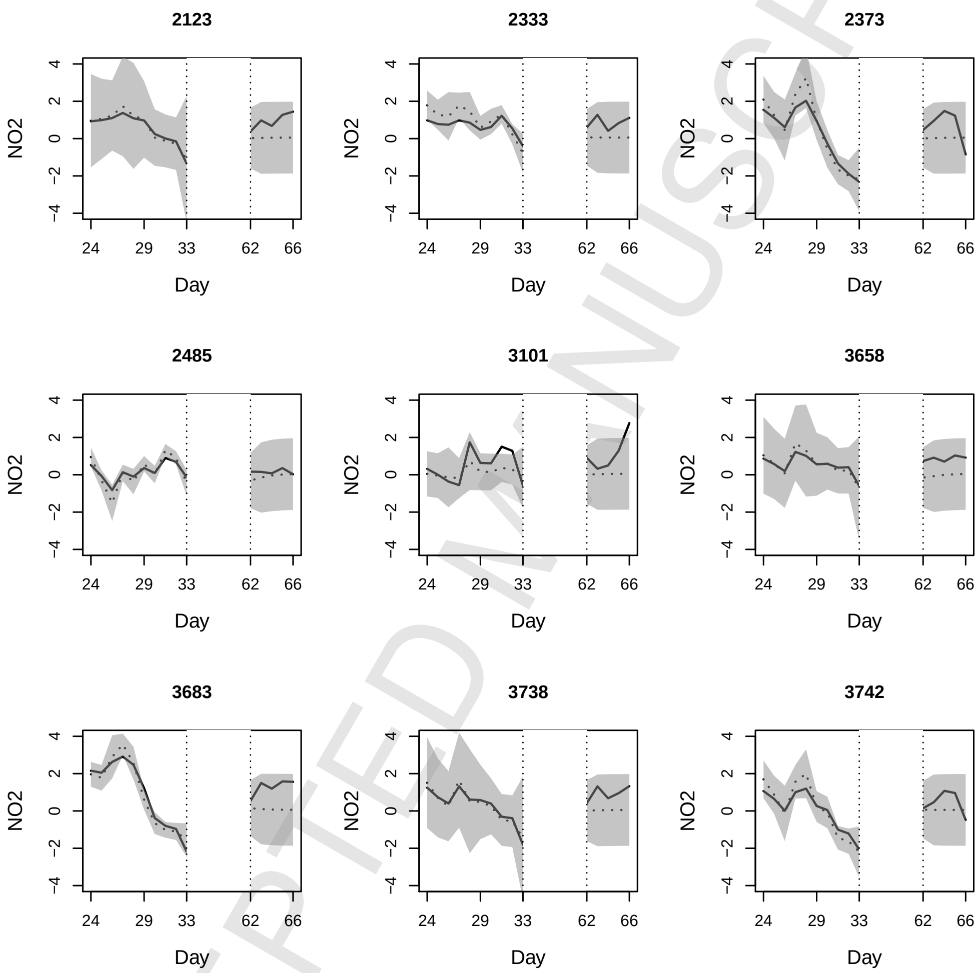


Figure 5: Plots of observed, interpolated and forecast values of NO2 under the time varying model at different sites. Refer to the text for details.

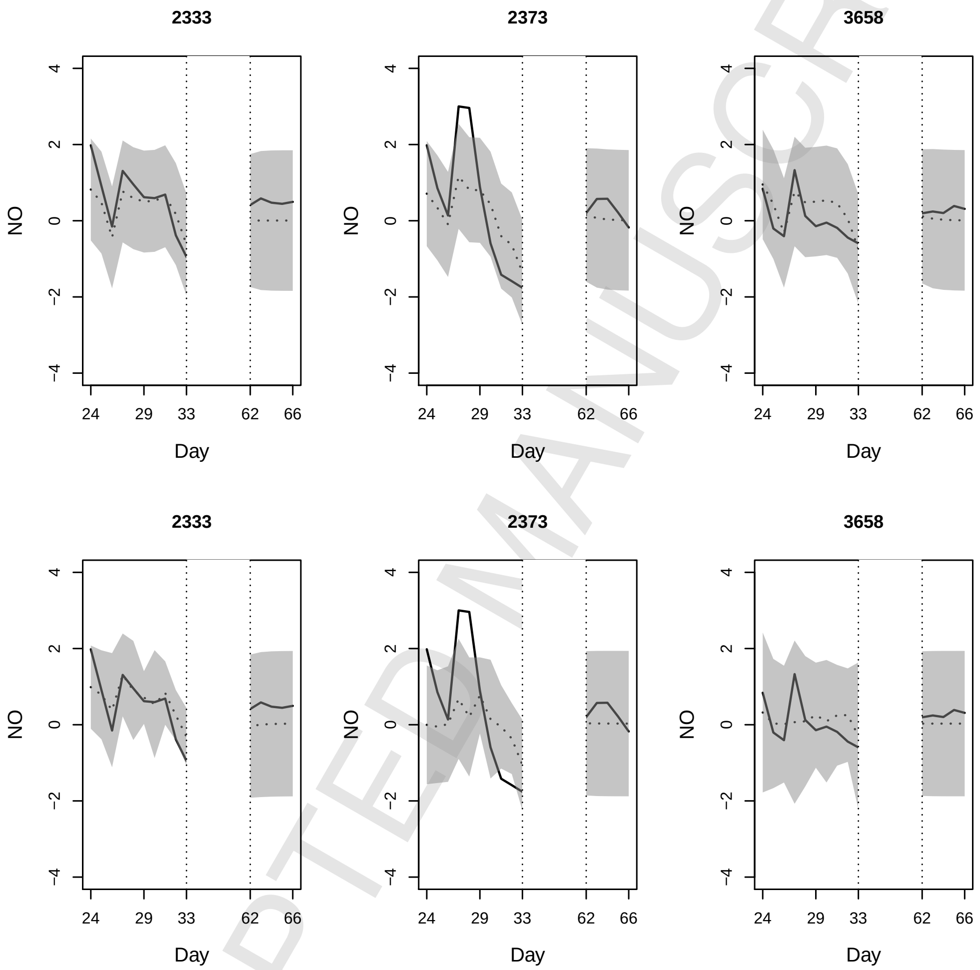


Figure 6: Plots of observed, interpolated and forecast values of NO at different sites. Results from the separable model (10) and time varying models are shown in the top and bottom panels respectively. Refer to the text for details.

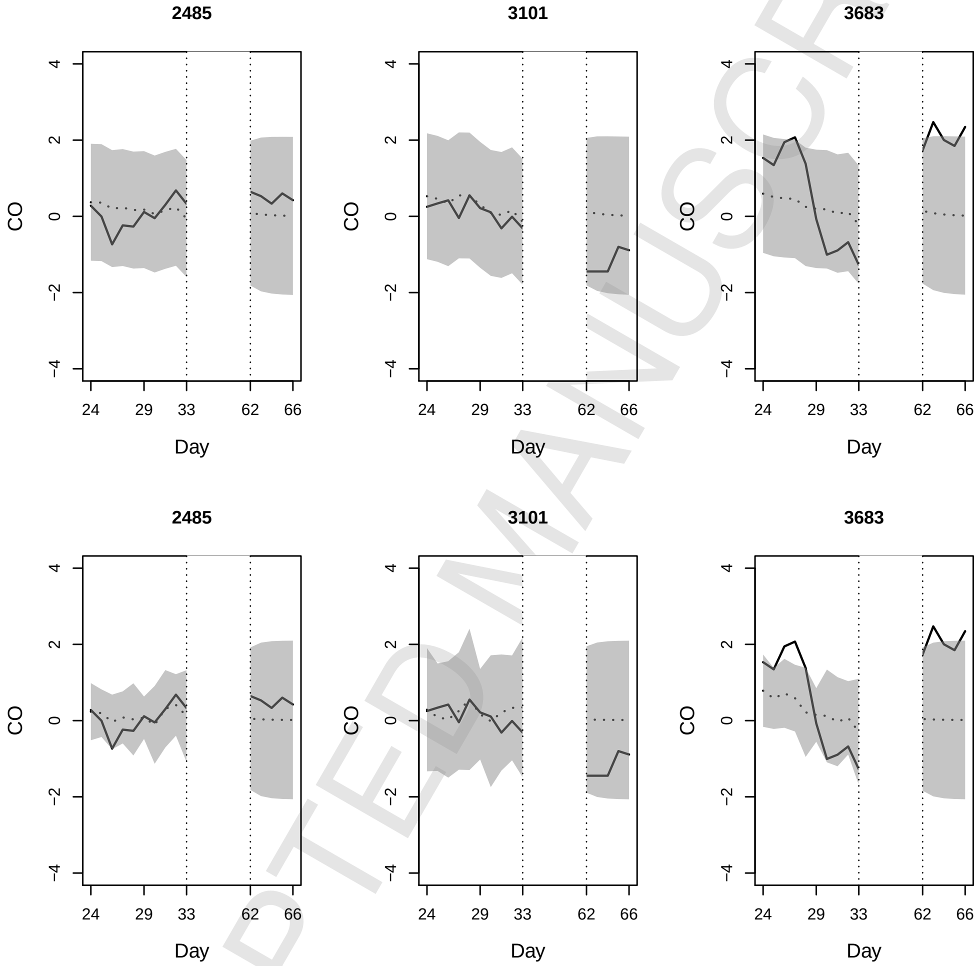


Figure 7: Plots of observed, interpolated and forecast values of CO at different sites. Results from the separable model (10) and time varying models are shown in the top and bottom panels respectively. Refer to the text for details.

Table 3: Mean squared prediction errors (MSPE) for interpolation from day 24 to 33 and forecasting from day 62 to 66 under the separable model (10) and time varying models for NO2, NO and CO at different sites. Bold entries indicate smaller MSPE.

		Interpolation		Forecasting	
Variable	Station	Separable	Time Varying	Separable	Time Varying
NO2	2123	**0.1106**	0.1583	1.1335	**0.9660**
	2333	0.0553	**0.0223**	0.7765	**0.7208**
	2373	**0.1843**	0.3788	1.1234	**1.0696**
	2485	0.2440	**0.1251**	**0.0706**	0.0897
	3101	1.6593	**1.2997**	2.0676	**2.0072**
	3658	**0.0828**	0.6461	0.9241	**0.8085**
	3683	1.0170	**0.4880**	1.6479	**1.5811**
	3738	0.2356	**0.1077**	1.0972	**0.9314**
	3742	0.3102	**0.0658**	0.5088	**0.4665**
NO	2333	0.2472	**0.2128**	0.2315	**0.2287**
	2373	**1.3075**	2.2165	**0.1246**	0.1383
	3658	**0.2317**	0.3034	**0.0606**	0.0621
CO	2485	0.1808	**0.0848**	**0.2293**	0.2401
	3101	**0.0639**	0.1156	1.6949	**1.5925**
	3683	1.1813	**0.9803**	**4.1560**	4.3041

both MA(1). Then, the q-step ahead forecast can be done by first predicting the values of spatial covariance parameters at time $T+q$, $q \in \mathbb{Z}^{+}$ and hence, $\hat{X}_p(\boldsymbol{s}_{0i}, T+q)$ can be computed using steps 1(a) to 3(a) as given in section 2.3. The same procedures were applied to NO and CO. As an illustration, we performed forecasting from day 62 to 66, i.e., five-step ahead forecast from 1 to 5 November, 2010. The right panel of Table 3 shows the mean squared errors of the five-step ahead forecast using different models at the extra stations. The plots of observed versus forecast values at different sites can be found in Figures 4 to 7.

As naturally expected, the performances of forecasting for all models and all variables are poorer than those of interpolation. As shown in Table 3, the MSPE for forecasting are usually larger than that of interpolation. Nevertheless, comparing the performance of forecasting, the time varying model produces smaller values of MSPE in general. Overall speaking, across all variables and sites, the maximum reduction in MSPE is 15.1% and the median reduction is 4.05%.

From Figures 4 to 7, it can be observed that the forecast values are usually flat and are often biased from the observed values under the time varying models. Yet, the situation is also observed under the separable model (10).

Combining both results of interpolation and forecasting, we have shown that, at least in some cases, it is beneficial to allow the spatio-temporal models to change with time and therefore the proposed time varying spatio-temporal covariance model would be useful in practice. Although the time varying model does not always provide a better fit, it is at least comparable to the separable model (10). Therefore, the time varying model is worthwhile to be considered in practice.

6. Conclusion and Discussions

To summarize, we have introduced univariate and multivariate spatio-temporal models that the spatial dependency structures are allowed to vary over time. The models allow the covariance models and parameters to vary with time, which relax the constant parameter assumption imposed on ordinary spatio-temporal covariance models. The models were demonstrated to be useful in practice through applying them to a set of air pollution data. The proposed time varying models often show better performance in terms of interpolation and forecast. Although the proposed method is not always a superior one, it is worthy to be considered in practice as an alternative model.

In Corollaries 1 and 4, the matrices $\boldsymbol{S}$ and $\boldsymbol{D}$ were restricted to be diagonal. It is noted that such a restriction reduces the flexibility of the models. In light of Kleiber and Genton (2013), it is noted that $\boldsymbol{S}$ and $\boldsymbol{D}$ can be replaced by $\boldsymbol{U}^\top \boldsymbol{S}\boldsymbol{U}$ and $\boldsymbol{U}^\top \boldsymbol{D}\boldsymbol{U}$ respectively for some unitary matrices $\boldsymbol{U}$. However, the inclusion of $\boldsymbol{U}$ may lead to a huge increase in the number of parameters which is not desirable in most of the cases. Nevertheless, there may appear some unitary matrices $\boldsymbol{U}$ with nice structures such that one can strike a balance between model complexity and flexibility. Similarly, more flexible multivariate structures are yet to be investigated. We leave these problems for future research.

Acknowledgment

The authors thank an associate editor and an anonymous reviewer for their helpful and critical suggestions.

Reference

Abramowitz, M., Stegun, I. A., 1972. Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. National Bureau of Standards Applied Mathematics Series 55. Tenth printing. Dover.

Apanasovich, T. V., Genton, M. G., 2010. Cross-covariance functions for multivariate random fields based on latent dimensions. Biometrika 97 (1), 15–30.

Bevilacqua, M., Gaetan, C., Mateu, J., Porcu, E., 2012. Estimating space and space–time covariance functions for large data sets: a weighted composite likelihood approach. J. Am. Stat. Assoc. 107 (497), 268–280.

Bhatia, R., 2007. Positive Definite Matrices. Princeton University Press, Princeton, NJ.

Bochner, S., 1955. Harmonic Analysis and the Theory of Probability. Berkeley and Los Angeles: University of Califronia Press.

Brockwell, P. J., Davis, R. A., 2009. Time series: theory and methods. Springer.

Chilès, J. P., Delfiner, P., 1999. Geostatistics: Modeling Spatial Uncertainty. John Wiley & Sons, Inc.

Christakos, G., 2000. Modern Spatiotemporal Geostatistics. Oxford University Press, New York.

Cressie, N., 1993. Statistics for Spatial Data. John Wiley & Sons, Inc.

Cressie, N., Huang, H.-C., 1999. Classes of nonseparable, spatio–temporal stationary covariance functions. J. Am. Stat. Assoc. 94 (448), 1330–1339.

Cressie, N., Wikle, C. K., 2011. Statistics for Spatio–Temporal Statistics. John Wiley & Sons, Inc.

De Iaco, S., Myers, D. E., Posa, D., 2002. Nonseparable space-time covariance models: some parametric families. Math. Geol. 34 (1), 23–42.

Fernández-Casal, R., González-Manteiga, W., Febrero-Bande, M., 2003. Space-time dependency modeling using general classes of flexible stationary variogram models. J. Geophys Res: Atmos 108 (D24).

Finkenstädt, B., Held, L., Isham, V., 2007. Statistical Methods for Spatio–Temporal Systems. Chapman & Hall/CRC.

Fuentes, M., Chen, L., Davis, J. M., 2008. A class of nonseparable and nonstationary spatial temporal covariance functions. Environmetrics 19 (5), 487–507.

Gelfand, A. E., Kim, H. J., Sirmans, C. F., Banerjee, S., 2003. Spatial modeling with spatially varying coefficient processes. J. Am. Stat. Assoc. 98 (462), 387–396.

Gelfand, A. E., Schmidt, A. M., Banerjee, S., Sirmans, C., 2004. Nonstationary multivariate process modeling through spatially varying coregionalization. Test 13 (2), 263–312.

Gneiting, T., 2002. Nonseparable, stationary covariance functions for space–time data. J. Am. Stat. Assoc. 97 (458), 590–600.

Horn, R. A., Johnson, C. R., 1990. Matrix Analysis. Cambridge University Press.

Ip, H. L., 2015. On some topics in spatio-temporal modelling. Ph.D. thesis, The University of Hong Kong, unpublished.

Kleiber, W., Genton, M. G., 2013. Spatially varying cross–correlation coefficients in the presence of nugget effects. Biometrika 100 (1), 213–220.

Kyriakidis, P. C., Journel, A. G., 1999. Geostatistical space–time models: a review. Math. Geol. 31 (6), 651–684.

Matérn, B., 1986. Spatial Variation, 2nd Edition. Springer, New York.

Matheron, G., 1962. Traité de Géostatistique Appliquée. Paris: Editions Technip.

Porcu, E., Gregori, P., Mateu, J., 2006. Nonseparable stationary anisotropic space–time covariance functions. Stoch. Env. Res. Risk A. 21 (2), 113–122.

Rodríguez-Iturbe, I., Mejía, J. M., 1974. The design of rainfall networks in time and space. Water Resour. Res. 10 (4), 713–729.

Schmidt, A. M., Gelfand, A. E., 2003. A Bayesian coregionalization approach for multivariate pollutant data. J. Geophys. Res.–Atmos. 108 (D24), STS10–1–8.

Sherman, M., 2011. Spatial Statistics and Spatio–Temporal Data: Covariance Functions and Directional Properties. John Wiley & Sons Ltd.

Stein, M. L., 1999. Interpolation of Spatial Data: Some Theory for Kriging. Springer–Verlag, New York.

Stein, M. L., 2005. Space–time covariance functions. J. Am. Stat. Assoc. 100 (469), 310–321.

Varin, C., Vidoni, P., 2005. A note on composite likelihood inference and model selection. Biometrika 92 (3), 519–528.

Venables, W. N., Ripley, B. D., 2002. Modern Applied Statistics with S, 4th Edition. Springer, New York.

Wackernagel, H., 2003. Multivariate Geostatistics: An Introduction with Application, 3rd Edition. Springer–Verlag, Berlin.

Wei, W. W. S., 2006. Time Series Analysis: Univariate and Multivariate Methods, 2nd Edition. Pearson Education, Inc.

Zhang, H., 2004. Inconsistent estimation and asymptotically equal interpolations in model–based geostatistics. J. Am. Stat. Assoc. 99 (465), 250–261.

Zimmerman, D. L., Cressie, N., 1992. Mean squared prediction error in the spatial linear model with estimated covariance parameters. Ann. Inst. Stat. Math. 44 (1), 27–43.