Statistica Sinica Preprint No: SS-2022-0064	
Title	Modeling and Predicting Spatio-Temporal Dynamics of PM$_{2.5}$ Concentrations Through Time-Evolving Covariance Models
Manuscript ID	SS-2022-0064
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202022.0064
Complete List of Authors	Ghulam A. Qadir and Ying Sun
Corresponding Authors	Ghulam A. Qadir
E-mails	ghulam.qadir@h-its.org
Notice: Accepted version subject to English editing.	

Statistica Sinica

MODELING AND PREDICTING SPATIO-TEMPORAL DYNAMICS OF $PM_{2.5}$ CONCENTRATIONS THROUGH TIME-EVOLVING COVARIANCE MODELS

Ghulam A. Qadir[1] and Ying Sun[2]

[1]*Heidelberg Institute for Theoretical Studies*

[2]*King Abdullah University of Science and Technology*

Abstract: Fine particulate matter ($PM_{2.5}$) has become a great concern worldwide due to its adverse health effects. $PM_{2.5}$ concentrations typically exhibit complex spatio-temporal variations. Both the mean and the spatio-temporal dependence evolve with time due to seasonality, which makes the statistical analysis of $PM_{2.5}$ challenging. In geostatistics, Gaussian process is a powerful tool for characterizing and predicting such spatio-temporal dynamics, for which the specification of a spatio-temporal covariance function is the key. While the extant literature offers a wide range of choices for flexible stationary spatio-temporal covariance models, the temporally evolving spatio-temporal dependence has received scant attention only. To this end, we propose a time-varying spatio-temporal covariance model for describing the time-evolving spatio-temporal dependence in $PM_{2.5}$ concentrations. The proposed model is shown to outperform traditionally used models through simulation studies in terms of predictions. We apply our model

to analyze the $PM_{2.5}$ data in the state of Oregon, US. Therein, we show that the spatial scale and smoothness exhibit noticeable temporal variation. The proposed model is also shown to be beneficial over traditionally used models on this dataset for predictions.

Key words and phrases: Bernstein functions, Nonstationarity, Matérn covariance function, Spatio-temporal covariance.

1. Introduction

In the context of air quality control, particulate matter concentrate with diameter $\leq 2.5\mu m$ ($PM_{2.5}$) is a crucial pollutant of concern because of its deleterious effects on human health (Dominici et al., 2006; Pope III and Dockery, 2006). In consequence, $PM_{2.5}$ has been a focal topic in numerous air quality control oriented research where it has been studied for its chemical composition (Zhang et al., 2020), risk assessment (de Oliveira et al., 2012; Amoatey, Omidvarborna, and Baawain, 2018), statistical modeling and prediction (Qadir and Sun, 2020; Qadir, Euán, and Sun, 2020), etc. $PM_{2.5}$ is closely connected to the meteorology (Dawson, Adams, and Pandis, 2007), which causes the seasonality or other time-varying factors to have a strong influence on it. This strong seasonality effect has been noted in numerous case studies pertaining to the spatio-temporal variations of $PM_{2.5}$ (Bell et al., 2007; Zhao et al., 2019). Much of the literature con-

cerning the spatio-temporal modeling of $PM_{2.5}$, such as Li et al. (2017); Xiao, Lang, and Christakos (2018), account for such time-varying effects in the mean but ignore those effects in the spatio-temporal covariance of $PM_{2.5}$. In order to achieve a more comprehensive spatio-temporal modeling of $PM_{2.5}$, both the mean and spatio-temporal dependence should be allowed to evolve temporally for the seasonality effect. This necessitates the development of flexible time-varying model for a beneficial statistical modeling and prediction of $PM_{2.5}$.

Statistical modeling of $PM_{2.5}$ as a spatio-temporal stochastic process allows us to delve into the associated spatio-temporal uncertainties and perform predictions at unobserved space-time points, which can be beneficial in planning strategies for air quality control and formulating health care policies. The Gaussian process models are the typical choice of stochastic process models in spatio-temporal modeling where the joint distribution of random variables continuously indexed with space and time is multivariate normal. In particular, let $X(\mathbf{s}, t)$, $(\mathbf{s}, t) \in \mathbb{R}^d \times \mathbb{R}$, be the Gaussian process, indexed by space-time coordinates $(\mathbf{s}, t)$, then for any finite set of space-time pairs $\{(\mathbf{s}_1, t_1), \ldots, (\mathbf{s}_n, t_n)\}$, $n \geq 1$, the random vector $\{X(\mathbf{s}_1, t_1), \ldots, X(\mathbf{s}_n, t_n)\}^{\mathrm{T}} \sim \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = [\mathrm{Cov}\{X(\mathbf{s}_i, t_i), X(\mathbf{s}_j, t_j)\}]_{i,j=1}^n$ is the covariance matrix and $\boldsymbol{\mu} = [\mathbb{E}\{X(\mathbf{s}_1, t_1)\}, \ldots, \mathbb{E}\{X(\mathbf{s}_n, t_n)\}]^{\mathrm{T}}$

is the mean vector of the multivariate normal distribution. The entries of $\mathbf{\Sigma}$ are usually defined through some nonnegative definite parametric function $\mathrm{K}(\mathbf{s}_i, \mathbf{s}_j, t_i, t_j)$. The optimal prediction of an unobserved part of $X(\mathbf{s}, t)$ is given by kriging predictor (Cressie, 1993), which is a weighted linear combination of the observed part of $X(\mathbf{s}, t)$. These weights are affected by the covariance structure of the process, and therefore, the function $\mathrm{K}(\mathbf{s}_i, \mathbf{s}_j, t_i, t_j)$ must be specified diligently.

For practical convenience, it is often assumed that the covariance function $\mathrm{K}(\mathbf{s}_i, \mathbf{s}_i + \mathbf{h}, t_i, t_i + u)$ is stationary in space and time, i.e., $\mathrm{K}(\mathbf{s}_i, \mathbf{s}_i + \mathbf{h}, t_i, t_i + u) = \mathrm{C}(\mathbf{h}, u)$ depends only on the spatial lag $\mathbf{h}$ and temporal lag u. The particular restrictions $\mathrm{C}(\mathbf{h}, 0)$ and $\mathrm{C}(\mathbf{0}, u)$ represent purely spatial and purely temporal covariance functions, respectively. The existing literature on stationary spatio-temporal models provides practitioners with numerous alternatives, and those are comprehensively summarized in review papers by Gneiting, Genton, and Guttorp (2006) and Chen, Genton, and Sun (2021). A rudimentary approach to build a valid spatio-temporal covariance function is to impose separabilty, in which $\mathrm{C}(\mathbf{h}, u)$ can be decomposed into purely spatial and purely temporal covariance function. This decomposition can be in the form of a product: $\mathrm{C}(\mathbf{h}, u) = \mathrm{C}_s(\mathbf{h})\mathrm{C}_t(u)$, or in the form of a sum: $\mathrm{C}(\mathbf{h}, u) = \mathrm{C}_s(\mathbf{h}) + \mathrm{C}_t(u)$. The sum-based model suffers the rendering

of singular covariance matrices for some configurations of spatio-temporal data (Myers and Journel, 1990). Besides, a major shortcoming of separable model is its inability to allow for any space-time interactions which are often present in real data. For space-time interactions, Cressie and Huang (1999) introduced some classes of stationary nonseparable spatio-temporal covariance functions based on Fourier transform pairs in $\mathbb{R}^d$. Their approach led Gneiting (2002) to develop general classes of stationary nonseparable spatio-temporal covariance functions by using completely monotone functions and positive functions with completely monotone derivatives. Some further developments of nonseparable models include Stein (2005); De Iaco, Myers, and Posa (2002)

Among the existing stationary nonseparable spatio-temporal covariance functions, Gneiting (2002)'s classes have been notably popular and are explored for further generalizations (Porcu, Gregori, and Mateu, 2006; Bourotte, Allard, and Porcu, 2016). Specifically, Gneiting's class is defined as:

$$\mathrm{C}(\mathbf{h}, u) = \frac{\sigma^2}{\psi(|u|^2)^{d/2}} \varphi\left(\frac{\|\mathbf{h}\|^2}{\psi(|u|^2)}\right), (\mathbf{h}, u) \in \mathbb{R}^d \times \mathbb{R}, \tag{1.1}$$

where $\sigma > 0$ is the standard deviation of the process, $\varphi(w)$, $w \geq 0$, be any completely monotone function and $\psi(w)$, $w \geq 0$, be any positive function with a completely monotone derivative which is commonly termed as Bern-

stein function (Bhatia and Jain, 2015). Table 1 and Table 2 of Gneiting (2002) provide different choices of $\varphi(\cdot)$ and standardized $\psi(\cdot), \psi(0) = 1$, respectively. For a particular choice $\varphi(w) = (\alpha w^{1/2})^{\nu} K_{\nu}(\alpha w^{1/2})/\{2^{\nu-1}\Gamma(\nu)\}$, $\alpha > 0, \nu > 0$, where $K_{\nu}(\cdot)$ is a modified Bessel function of the second kind of the order ν (Abramowitz and Stegun, 1965), (1.1) reduces to:

$$\mathrm{C}(\mathbf{h}, u) = \frac{1}{\psi(|u|^2)^{\frac{d}{2}}} \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} \left(\frac{\alpha\|\mathbf{h}\|}{\psi(|u|^2)^{\frac{1}{2}}} \right)^{\nu} K_{\nu}\left(\frac{\alpha\|\mathbf{h}\|}{\psi(|u|^2)^{\frac{1}{2}}} \right), (\mathbf{h}, u) \in \mathbb{R}^d \times \mathbb{R}. \tag{1.2}$$

The purely spatial covariance function in (1.2) reduces to: $\mathrm{C}(\mathbf{h}, 0) = \sigma^2(\alpha\|\mathbf{h}\|)^{\nu} K_{\nu}(\alpha\|\mathbf{h}\|) 2^{1-\nu}/\Gamma(\nu)$, which belongs to the Matérn class (Matérn, 1986), henceforth denoted as $\sigma^2 \mathrm{M}(\mathbf{h} \mid \alpha, \nu)$, where $\alpha > 0$ and $\nu > 0$ represent spatial scale and smoothness parameters, respectively. The Matérn class has become an extremely preferred and important class of isotropic covariance functions for modeling spatial data (Stein, 1999; Gneiting, Kleiber, and Schlather, 2010), and therefore, (1.2), which we hereafter refer as "Gneiting-Matérn" class is also particularly important.

The class of stationary spatio-temporal models which does not allow covariance to evolve either in space or time, can be restrictive for many real applications. Therefore, while this class of models is essential, its relevance to the considered data must be assessed individually. Moreover, the spatio-temporal data which is likely to demonstrate heterogeneity of dependence

(nonstationarity) in space and/or time must be served with flexible space and/or time-varying models for satisfactory inference and prediction. Consequently, numerous nonstationary spatio-temporal models with varied fundamental constructions have been proposed in the last two decades. Stroud, Müller, and Sansó (2001) proposed a state-space model in which the nonstationarity operates through locally-weighted mixture of regression surfaces with time-varying regression coefficients. Ma (2002) proposed nonstationary spatio-temporal covariance model constructions through scale and positive power mixtures of stationary covariance functions. Set in the spectral domain, Fuentes et al. (2008) derived the nonstationary spatio-temporal covariance model via mixture of locally stationary spatio-temporal spectral densities. Shand and Li (2017) extended the idea of dimension expansion by Bornn et al. (2012) to the spatio-temporal case, resulting in nonstationary spatio-temporal covariances. Some other important works in the nonstationary spatio-temporal modeling include Huang and Hsu (2004); Sigrist, Künsch, and Stahel (2012); Xu and Gardoni (2018). The nonstationary extension of the Gneiting-Matérn class is also of particular interest and have been explored by Porcu et al. (2006) and Porcu, Mateu, and Bevilacqua (2007) for spatial anisotropy and spatial nonstationarity, respectively, but the resulting models in those approaches are stationary in time.

Alternatively, one can impart space-time nonstationarity to the Gneiting-Matérn class through the process convolution-based spatio-temporal covariance model of Garg, Singh, and Ramos (2012), however, their model does not allow for evolving smoothness.

We propose a time-varying spatio-temporal covariance model by generalizing the Gneiting-Matérn class (1.2) to include temporally varying spatial scale and smoothness. The foundational idea of the proposed model is analogous to the work of Ip and Li (2015), however, our model construction significantly differs from that of Ip and Li (2015). The time-varying model of Ip and Li (2015) requires computing the square root of purely spatial covariance matrices for the evaluation of the full space-time covariance matrix, which can be a computationally expensive evaluation for a large number of spatial locations. In contrast, the proposed model avoids computing the square root of matrices and provides a simple parametric functional form for evaluation of the spatio-temporal covariance. Additionally, since the proposed model is a generalization of the Gneiting-Matérn class, it inherits all the desirable properties of the original class and beyond.

The rest of the paper is organized as follows: we introduce the considered $PM_{2.5}$ data and perform a preliminary data analysis in Section 2 to motivate time-varying model construction. In Section 3, we describe the

proposed time-varying model and its properties. We conduct a simulation study to compare the performance of the proposed model with Gneiting-Matérn class and separable model in Section 4. The proposed time-varying model is then applied to analyze the $PM_{2.5}$ data in Section 5. We conclude with discussion and potential future extensions in Section 6.

2. The $PM_{2.5}$ Data and the Preliminary Analysis

The $PM_{2.5}$ data in consideration is sourced from the Environmental Protection Agency (EPA), which provides daily average measurements of $PM_{2.5}$ across the United States. To generate the $PM_{2.5}$ measurements, the EPA integrates the monitoring data from National Air Monitoring Stations/State and Local Air Monitoring Stations (NAMS/SLAMS) with the Community Multiscale Air Quality (CMAQ) modeling system (`https://www.epa.gov/cmaq`), which generates 12 km gridded output. For our analysis, we focus on the weekly average data for the year 2017. Prior to calculating the weekly averages, we apply a log transformation to the $PM_{2.5}$ values. Through the log transformation, we aim to normalize the data, which in turn makes it suitable for Gaussian process modeling. For the rest of our study, we use the term $PM_{2.5}$ synonymously with the weekly averaged and log transformed $PM_{2.5}$, for simplicity.

While the raw dataset covers the entire United States, our analysis focuses specifically on the state of Oregon for the year 2017. We choose the state of Oregon due to its susceptibility to wildfires, making the study of $PM_{2.5}$ concentration in this region particularly relevant. The dataset consists of 41,860 observations, representing weekly time series data (52 weeks) at 805 spatial locations. These spatial locations are depicted in Figure 1 and are denoted by the set $\mathcal{D}_s \subset \mathbb{R}^2$. Additionally, we define a set of 52 equally spaced points in the interval $[0, 1]$, denoted by $\mathcal{D}_t \subset \mathbb{R}$. Figure 1 visualizes the spatial fields of the considered $PM_{2.5}$ data. Clearly, the 805 observed spatial locations shown in Figure 1 are not uniformly distributed across the state of Oregon. These observed locations are concentrated in the western, northwestern, and southwestern regions of Oregon, which have higher population density. Conversely, the eastern region has limited or no observations available. Therefore, the primary objective of our analysis is to effectively model the spatio-temporal dependence of the $PM_{2.5}$ data and accordingly develop an accurate predictive model that can predict PM2.5 levels in the space-time continuum of the unobserved regions of the study area.

The accuracy of any predictive model must be assessed through cross-validation, and as such, we conduct multiple rounds of cross-validation to

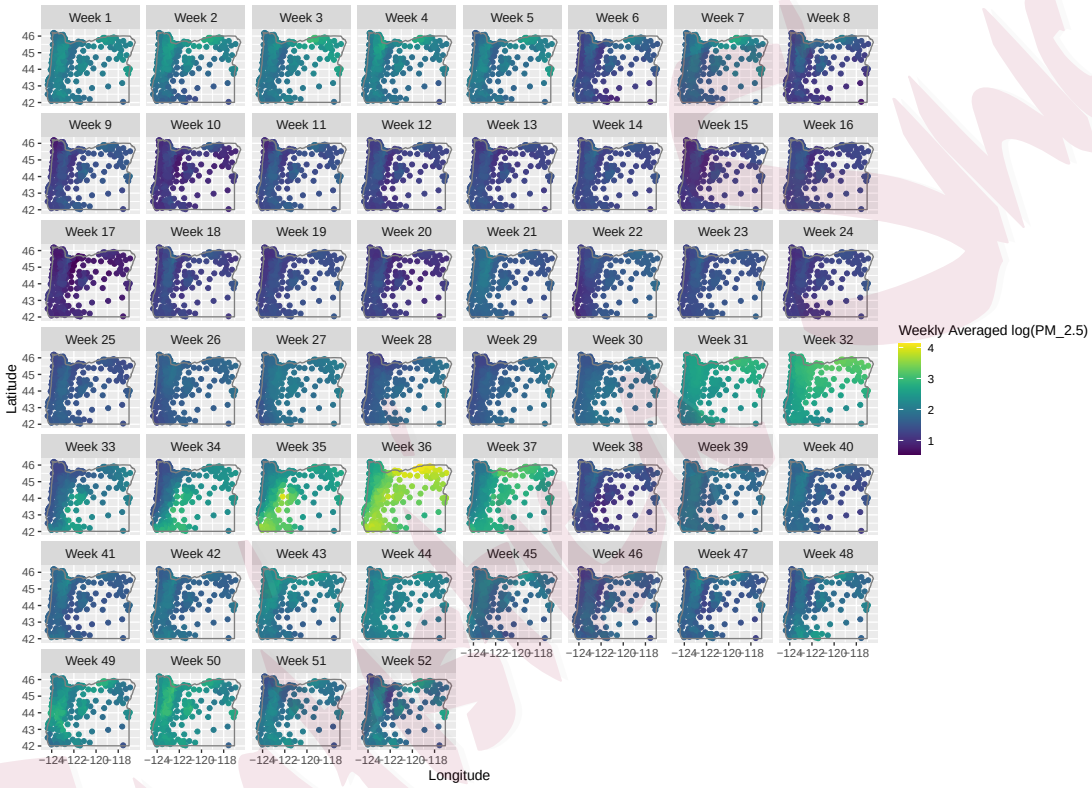

Figure 1: Observed $PM_{2.5}$ data over the state of Oregon, US.

ensure a robust analysis. To achieve this, we create 100 random spatial splits of the complete dataset, which consists of 805 spatial locations, into training sets comprising 322 spatial locations and validation sets comprising 483 locations. Specifically, we randomly generate 100 distinct combinations of training and validation sets based on the training locations $\mathcal{T}_s^i \subset \mathcal{D}_s$ and validation locations $\mathcal{V}_s^i \subset \mathcal{D}_s$, respectively, such that $\mathcal{T}_s^i \cup \mathcal{V}_s^i = \mathcal{D}_s$, $i = 1, \ldots, 100$.

Let $Y(\mathbf{s}, t)$, $(\mathbf{s}, t) \in \mathcal{D}_s \times \mathcal{D}_t$, denote the $\text{PM}_{2.5}$ indexed at spatial location $\mathbf{s} = (\text{Longitude, Latitude}) \in \mathcal{D}_s$ and the time $t = (\text{week of the year 2017} - 1)/(52 - 1) \in \mathcal{D}_t$, where time t is scaled such that $t = 0$ and $t = 1$ represent the first and the last week of the year 2017, respectively. We aim to model $Y(\mathbf{s}, t)$, as a spatio-temporal Gaussian process such that $\mathbb{E}\{Y(\mathbf{s}, t)\} = \mu(\mathbf{s}, t)$, and $\text{Cov}\{Y(\mathbf{s}_1, t_1), Y(\mathbf{s}_2, t_2)\} = K(\mathbf{s}_1, \mathbf{s}_2, t_1, t_2)$. To capture the spatio-temporal trends and seasonality in the mean, we model the mean function $\mu(\mathbf{s}, t)$ independently across the 100 training sets, as the following linear combination of spatial trend and temporal harmonic terms:

$$\mu(\mathbf{s}, t) = \beta_0 + \boldsymbol{\beta}_s^{\mathrm{T}} \mathbf{s} + \beta_1 \cos(3\pi t) + \beta_2 \sin(3\pi t), \ \beta_0, \beta_1, \beta_2 \in \mathbb{R}, \ \boldsymbol{\beta}_s \in \mathbb{R}^2.$$

We fit the specified linear model for $\mu(\mathbf{s}, t)$ to detrend the process $Y(\mathbf{s}, t)$ with the fitted $\hat{\mu}(\mathbf{s}, t)$ and obtain the residual process $\epsilon(\mathbf{s}, t) = Y(\mathbf{s}, t) - \hat{\mu}(\mathbf{s}, t)$, for each of the 100 training sets. The residual process is then stan-

dardized and further investigated to explore the properties of $K(\mathbf{s}_1, \mathbf{s}_2, t_1, t_2)$. For simplicity, we assume that the process is stationary in space, and therefore, the covariance function $K(\mathbf{s}_1, \mathbf{s}_2, t_1, t_2) = \mathrm{C}(\mathbf{s}_2 - \mathbf{s}_1, t_1, t_2)$, depends only on spatial lag $\mathbf{s}_2 - \mathbf{s}_1$ and time-points t_1, t_2. Furthermore, we also assume that the purely spatial covariance function for any arbitrary time t: $\mathrm{C}(\mathbf{s}_2 - \mathbf{s}_1, t, t)$ is of the Matérn class. Next, we want to explore the time-varying properties of $\mathrm{C}(\mathbf{s}_2 - \mathbf{s}_1, t, t)$; and therefore, we fit a purely spatial Matérn covariance function, $\sigma^2\mathrm{M}(\mathbf{h} \mid \alpha, \nu)$, independently for each $t \in \mathcal{D}_t$, on the residual processes from the 100 training sets, using the maximum likelihood estimation (MLE) method. Figure 2 presents the boxplots illustrating the weekly estimates of the spatial scale parameter α and the smoothness parameter ν obtained from fitting the spatial Matérn covariance function across 100 training sets. Notably, both parameters, α and ν, display a clear temporal evolution. During the middle of the year, there is a noticeable dip in the values of both α and ν. This temporal pattern suggests a change in the spatial structure and smoothness of the underlying process during that period. Hence, it is crucial to consider the time-varying nature of the spatial dependence when defining the spatio-temporal covariance function $K(\mathbf{s}_1, \mathbf{s}_2, t_1, t_2)$. Neglecting this aspect can result in the misrepresentation of the underlying process, potentially leading to subopti-

mal inference and prediction. This serves as the motivation for developing our time-varying class of spatio-temporal covariance functions, which are discussed in detail in Section 3. The analysis of the data, as presented here, is continued in Section 5.

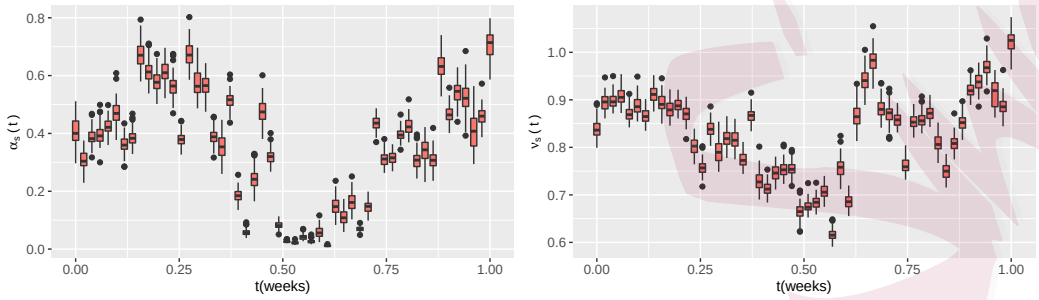

Figure 2: Boxplots illustrating the Maximum Likelihood estimates for (left) the spatial scale parameter α and (right) the smoothness parameter ν of the covariance function $\sigma^2\mathrm{M}(\mathbf{h} \mid \alpha, \nu)$. The estimates are obtained by fitting the covariance function independently over the standardized residuals for each week on the 100 training sets.

3. Time-varying Spatio-temporal Covariance Model

In this section, we introduce our proposed class of time-varying spatio-temporal covariance functions with discussion on its properties and validity conditions. We consider the Gneiting-Matérn class (1.2), and provide its time-varying generalization in the following theorem:

Theorem 1. *Let $\alpha_s(t) > 0, t \in \mathbb{R}$ and $\nu_s(t) > 0, t \in \mathbb{R}$, be any positive real valued functions, then the following time-varying spatio-temporal covariance function in* (3.3)*:*

$$Cov\{X(\boldsymbol{s}, t_i), X(\boldsymbol{s}+\boldsymbol{h}, t_j)\} = C(\boldsymbol{h}, t_i, t_j) = \sigma^2 \frac{\Gamma\{\frac{\nu_s(t_i)+\nu_s(t_j)}{2}\}}{\sqrt{\Gamma\{\nu_s(t_i)\}\Gamma\{\nu_s(t_j)\}}} \times \tag{3.3}$$
$$\frac{1}{\{\alpha_s^2(t_i)\}^{d/4}\{\alpha_s^2(t_j)\}^{d/4}\{\frac{\psi(|t_i-t_j|^2)}{\overline{\alpha}_s^2} + \frac{1/\alpha_s^2(t_i)+1/\alpha_s^2(t_j)}{2} - \frac{\psi(0)}{\overline{\alpha}_s^2}\}^{d/2}} \times$$
$$M[\boldsymbol{h} \mid \frac{1}{\{\frac{\psi(|t_i-t_j|^2)}{\overline{\alpha}_s^2} + \frac{1/\alpha_s^2(t_i)+1/\alpha_s^2(t_j)}{2} - \frac{\psi(0)}{\overline{\alpha}_s^2}\}^{1/2}}, \frac{\nu_s(t_i)+\nu_s(t_j)}{2}],$$

is valid for any Bernstein function $\psi(w) > 0,\ w \geq 0$, and $\overline{\alpha}_s > 0$.

While the spatio-temporal covariance function in (3.3) is valid for any positive value of the parameter $\overline{\alpha}_s$, we now onwards choose to constrain it as $\overline{\alpha}_s = \sum_{t_i \in \mathrm{T}} \alpha_s(t_i)/\mathrm{T}$, where T represents the set of all the training time-points. This constraint is beneficial for two reasons: (i) it renders a simpler model with one less parameter to be estimated, and (ii) include (1.2) as a special case when $\alpha_s(t_i)$ and $\nu_s(t_i)$ are constant over time. Specifically, with any standardized Bernstein function $\psi(w), w \geq 0, \psi(0) = 1$, let $\alpha_s(t) = \alpha > 0,\ \nu_s(t) = \nu > 0,\ t \in \mathbb{R}$, in (3.3), then $\overline{\alpha}_s = \alpha$ and (3.3) reduces to:

$$\mathrm{C}(\mathbf{h}, t_i, t_j) = \frac{\sigma^2}{\{\psi(|t_i - t_j|^2)\}^{d/2}} \left(\frac{\alpha \|\mathbf{h}\|}{\psi(|t_i - t_j|^2)^{1/2}} \right)^{\nu} K_\nu \left(\frac{\alpha \|\mathbf{h}\|}{\psi(|t_i - t_j|^2)^{1/2}} \right),$$

which is a Gneiting-Matérn class (1.2), and on that account, (3.3) is a time-varying generalization of (1.2). The time-varying properties of the

spatio-temporal covariance model in (3.3) become intelligible in its purely spatial and purely temporal restrictions. In particular, let $t_i = t_j = t$ in (3.3) to evaluate the purely spatial covariance at any arbitrary time point t, we get: $\mathrm{C}(\mathbf{h}, t, t) = \sigma^2 \mathrm{M}\{\mathbf{h} \mid \alpha_s(t), \nu_s(t)\}$, which is a spatial Matérn covariance function with spatial scale $\alpha_s(t)$ and smoothness $\nu_s(t)$ as a function of time t, thus allowing the temporal evolution of spatial dependence. As an immediate consequence of functions $\alpha_s(t)$ and $\nu_s(t)$, the following purely temporal restriction of (3.3) becomes nonstationary in time:

$$\mathrm{C}(\mathbf{0}, t_i, t_j) = \frac{\sigma^2 \Gamma\{\frac{\nu_s(t_i)+\nu_s(t_j)}{2}\}\{\alpha_s^2(t_i)\}^{-d/4}\{\alpha_s^2(t_j)\}^{-d/4}}{\sqrt{\Gamma\{\nu_s(t_i)\}\Gamma\{\nu_s(t_j)\}}\{\frac{\psi(|t_i - t_j|^2)}{\bar{\alpha}_s^2} + \frac{1/\alpha_s^2(t_i)+1/\alpha_s^2(t_j)}{2} - \frac{\psi(0)}{\bar{\alpha}_s^2}\}^{d/2}}. \tag{3.4}$$

While the nonstationary behavior of the purely temporal covariance in (3.4) is entirely controlled by nontrivial interactions of functions $\alpha_s(t)$ and $\nu_s(t)$, the individual interpretation of those functions in the context of temporal nonstationarity becomes clear when we vary them singly in (3.4). The detailed discussion on the nonstationarity effects of $\alpha_s(t)$ and $\nu_s(t)$ on (3.4) is provided in Supplementary Material S3. The proper specification of functional forms for $\alpha_s(t)$ and $\nu_s(t)$ in (3.3), which can flexibly capture the time-varying dependence of any considered spatio-temporal data, is consequential to achieve an advantageous modeling and inference. Thus, the definition of the functions $\alpha_s(t)$ and $\nu_s(t)$ should ideally be based on empiri-

cal evidence from some exploratory data analysis. One such alternative is to follow the analysis of Section 2 and utilize the boxplots or estimates analogous to Figure 2 for defining the functions $\alpha_s(t)$ and $\nu_s(t)$ as those estimated boxplots are indeed an empirical counterpart of the corresponding functions $\alpha_s(t)$ and $\nu_s(t)$. While there can be innumerable possible constructions of positive real-valued functions for $\alpha_s(t)$ and $\nu_s(t)$, we consider the following definition of $\alpha_s(t)$ and $\nu_s(t)$ for the time-varying model estimation in our work: $\alpha_s(t) = \exp(p^{\alpha}_{n.\alpha}(t)), n.\alpha \geq 0$, and $\nu_s(t) = \exp(p^{\nu}_{n.\nu}(t)), n.\nu \geq 0$, where $p^{\alpha}_k(t)$ and $p^{\nu}_k(t)$, both are $k-$order polynomial of t. The polynomial based specification requires an informed choices for the polynomial orders $n.\alpha \geq 0$, and $n.\nu \geq 0$, which should be guided by an objective criteria. To this end, we provide a stepwise polynomial regression based exploratory analysis in Supplementary Material Section S5, where we also give the detailed explanations on the specific polynomial order choices considered in our simulation study and the data application. For the choice of $\psi(w) > 0, w \geq 0$ in (3.3), there are several options available from the list of Bernstein functions given in Van Den Berg and Forst (2012) and Gneiting (2002). In this work, we only consider the following choice of Bernstein function: $\psi(w) = (aw^{\gamma} + 1)^{\beta},\ a > 0,\ 0 < \gamma \leq 1,\ 0 \leq \beta \leq 1$. Moreover, to impart additional flexibility in the temporal part similar to Example 2 of

Gneiting (2002), we multiply (3.3) with a purely temporal covariance function: $(a|t_i-t_j|^{2\gamma}+1)^{-\delta}, \delta \geq 0$, where a and γ are the parameters common to the chosen function $\psi(w)$. Consequently, the time-varying spatio-temporal model in (3.3) reduces to:

$$\mathrm{C}(\mathbf{h}, t_i, t_j) = \sigma^2 \frac{\Gamma\{\frac{\nu_s(t_i)+\nu_s(t_j)}{2}\}}{\sqrt{\Gamma\{\nu_s(t_i)\}\Gamma\{\nu_s(t_j)\}}} \times \frac{\{\alpha_s^2(t_i)\}^{-d/4}\{\alpha_s^2(t_j)\}^{-d/4}}{\{\frac{(a|t_i-t_j|^{2\gamma}+1)^{\beta}}{\overline{\alpha}_s^2} + \frac{1/\alpha_s^2(t_i)+1/\alpha_s^2(t_j)}{2} - \frac{1}{\overline{\alpha}_s^2}\}^{d/2}(a|t_i-t_j|^{2\gamma}+1)^{\delta}} \times \mathrm{M}[\mathbf{h} \mid \frac{1}{\{\frac{(a|t_i-t_j|^{2\gamma}+1)^{\beta}}{\overline{\alpha}_s^2} + \frac{1/\alpha_s^2(t_i)+1/\alpha_s^2(t_j)}{2} - \frac{1}{\overline{\alpha}_s^2}\}^{1/2}}, \frac{\nu_s(t_i)+\nu_s(t_j)}{2}], \tag{3.5}$$

where $\sigma > 0,\ \delta \geq 0,\ 0 \leq \beta \leq 1,\ a > 0,\ 0 < \gamma \leq 1,\ \alpha_s(t) > 0, \nu_s(t) > 0, t \in \mathbb{R}$, and $\overline{\alpha}_s = \sum_{t_i \in \mathrm{T}} \alpha_s(t_i)$. Now, if we set $\alpha_s(t) = \alpha, t \in \mathbb{R}$ and $\nu_s(t) = \nu, t \in \mathbb{R}$, then (3.5) reduces to the following Gneiting-Matérn class:

$$\mathrm{C}(\mathbf{h}, t_i, t_j) = \sigma^2 \frac{1}{(a|t_i-t_j|^{2\gamma}+1)^{\beta d/2+\delta}} \mathrm{M}\{\mathbf{h} \mid \frac{\alpha}{(a|t_i-t_j|^{2\gamma}+1)^{\beta/2}}, \nu\}, \tag{3.6}$$

where β represents the parameter to control the degree of nonseparability such that $\beta = 1$ corresponds to fully nonseparable model and $\beta = 0$ leads to the following separable model:

$$\mathrm{C}(\mathbf{h}, t_i, t_j) = \sigma^2 \frac{1}{(a|t_i-t_j|^{2\gamma}+1)^{\delta}} \mathrm{M}(\mathbf{h} \mid \alpha, \nu). \tag{3.7}$$

The three spatio-temporal covariance models in (3.5), (3.6) and (3.7) are considered as the candidate models in the simulation study and data

application presented in Section 4 and Section 5, respectively. The model in (3.5) can be made arbitrarily flexible through parametric functions for $\alpha_s(t)$ and $\nu_s(t)$, however, the estimation of model parameters through Gaussian MLE also then becomes increasingly challenging. Higher values of $n.\alpha$ and $n.\nu$ in our polynomial-based specification lead to increased flexibility in the proposed time-varying model. In principle, large volume of data is preferred to fit a highly parameterized complex model such as (3.5) to avoid over-fitting, but at the same time the Gaussian MLE becomes time-prohibitive and computationally infeasible with a high volume of spatio-temporal data. The main issue lies in storing and performing the Cholesky factorization of the large covariance matrix. To overcome this, we implement a random composite likelihood (RCL)-based estimation procedure, which is described in Supplementary Material Section S2.

4. Simulation Study

In this section, we conduct a simulation study to empirically evaluate the advantage of using the proposed time-varying class of spatio-temporal covariance models against the commonly used Gneiting-Matérn class and the separable class of spatio-temporal covariance models. For this simulation study, we particularly consider the three nested models (3.5), (3.6)

and (3.7), which now onwards, are referred as "Tvar.M", "Gneit.M" and "Sep.M", respectively, for brevity. These models are compared on the basis of interpolation and forecasting performance under four different cases.

The spatial domain of interest, $\mathcal{D}_s$, is set to be 25×25 equally spaced grid points on a unit square, i.e., $[0,1]^2$ and temporal domain of interest, $\mathcal{D}_t$, is set as 21 equally spaced points in $[0,1]$. We simulate 100 realizations of a zero mean spatio-temporal Gaussian process $Z(\mathbf{s},t),\ \mathbf{s} \in \mathcal{D}_s \subset \mathbb{R}^2,\ t \in \mathcal{D}_t \subset \mathbb{R}$, with Tvar.M covariance model, under four different parameter settings listed as Cases 1–4 in Table 1. Observe that for Case 1, 2 and 3, the true functions $\alpha_s(t)$ and $\nu_s(t)$ are time-varying, whereas, for Case 4, $\alpha_s(t)$ and $\nu_s(t)$ are constant; therefore, true data generating model for Case 4 is in fact Gneit.M. The true functions $\alpha_s(t)$ and $\nu_s(t)$, for $t \in \mathcal{D}_t$, are also shown in Figure 3. As a consequence of specified $\alpha_s(t)$ and $\nu_s(t)$, the purely spatial dependence of Z varies periodically over time in Case 1, linearly over time in Case 2, nonlinearly over time in Case 3, and stays constant over time in Case 4. The resulting temporal nonstationarity is discussed in the Supplementary Material Section S4.

For comparison of interpolation and forecasting performance, we perform cross-validation, and accordingly, we split the data into a training set, a validation set for interpolation $\mathcal{V}_i$ and a validation set for forecast-

Table 1: Parameter settings for the four simulation cases. For all the four cases, the true values of the parameters $\{\sigma, \gamma, \beta, a, \delta\} = \{1, 0.60, 0.80, 10, 0.10\}$.

Functional	Case 1	Case 2	Case 3	Case 4
$\alpha_s(t)$	$20 + 15\sin(\frac{\pi t}{20})$	$25 - 10t$	$20 - 10\frac{\exp(10t-5)}{1+\exp(10t-5)}$	20
$\nu_s(t)$	$0.5 + \sin(\frac{\pi t}{20})$	$0.5 + t$	$0.5 + \frac{\exp(10t-5)}{1+\exp(10t-5)}$	1

ing $\mathcal{V}_f$. The entire spatial field at the last two time points of $\mathcal{D}_t$, i.e., at $t = \{0.95, 1.00\}$, constitute $\mathcal{V}_f$. We randomly select 125 spatial locations as our validation locations for interpolation and the data at those locations for the remaining 19 time-points of $\mathcal{D}_t$, i.e., $t = \{0.00, 0.05, \ldots, 0.90\}$, form $\mathcal{V}_i$. All the data that remains after removing the two validation sets make our training data. For each of the four simulation cases, we fit three candidate models Tvar.M, Gneit.M and Sep.M by using the RCL estimation, on the training data in each of the 100 simulation runs. During estimation, the functions $\alpha_s(t)$ and $\nu_s(t)$ in the candidate model Tvar.M are specified as: $\alpha_s(t) = \exp(p_2^{\alpha}(t)),\ \nu_s(t) = \exp(p_2^{\nu}(t)), t \in \mathcal{D}_t$ for Case 1, Case 2 and Case 4, whereas for Case 3, $\alpha_s(t) = \exp(p_3^{\alpha}(t)),\ \nu_s(t) = \exp(p_3^{\nu}(t)), t \in \mathcal{D}_t$. Note that the specification of functions $\alpha_s(t)$ and $\nu_s(t)$ in the candidate model Tvar.M are different from that data generating model Tvar.M. Additionally, we fix a to its true value, i.e., $a = 10$, in all the three candidate models during the estimation to slightly reduce the optimization burden.

The RCL parameter estimates and their description can be found in the Supplementary Material Section S4. The estimated $\alpha_s(t)$ and $\nu_s(t)$ from the candidate Tvar.M shown in Figure 3 display conspicious comparability with the corresponding true functions in all the four cases. Although, it is worth noting that, in Case 3, the estimate for $\nu_s(t)$ displays increasing offset from the true values for the time period t outside the training data, i.e. $t > 0.90$. This points out to the fact that, outside the training temporal domain, the estimated functions $\alpha_s(t)$ and $\nu_s(t)$ should be interpreted with caution. Note that the candidate Tvar.M model is also slightly misspecified in Cases 1–3 due to its functional specification of $\alpha_s(t)$ and $\nu_s(t)$, which is different from that of the true model; however, the estimated functions $\alpha_s(t)$ and $\nu_s(t)$, in general, still recover the corresponding true functions because the specification in the candidate Tvar.M is flexible enough.

We now perform spatio-temporal prediction at validation space-time coordinates through kriging with the estimated three candidate covariance models to achieve cross-validation in all the four simulation cases. Specifically, we predict Z at the space-time coordinates in $\mathcal{V}_i$ and $\mathcal{V}_f$ to obtain the interpolation and forecast of Z, respectively. Under Gaussian process framework, the predictive distribution of any unobserved $Z(\mathbf{s}^0, t^0)$ is the conditional Gaussian distribution where conditioning is over all the

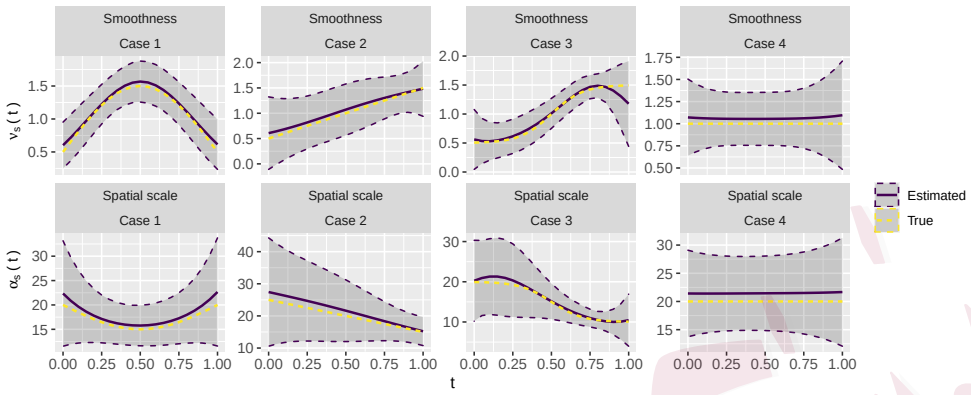

Figure 3: Casewise comparison of the pointwise average of the estimated functions $\hat{\alpha}_s(t)$ and $\hat{\nu}_s(t)$ from the candidate Tvar.M model with their corresponding true functions of the data generating Tvar.M model. The average is taken over the 100 simulation runs. The 95% pointwise-interval is also shown in grey bands.

observed $Z(\mathbf{s}, t)$. Kriging provides us with prediction value $\hat{Z}(\mathbf{s}^0, t^0)$ and prediction variance $\hat{\sigma^2}_{\mathbf{s}^0,t^0}$ of the unobserved $Z(\mathbf{s}^0, t^0)$, which, under Gaussian process assumption, defines the predictive distribution of $Z(\mathbf{s}^0, t^0)$ as $\mathcal{N}\big(\hat{Z}(\mathbf{s}^0, t^0), \hat{\sigma^2}_{\mathbf{s}^0,t^0}\big)$. The predictive distribution enables us to construct $p-$prediction intervals (p-PI) for $Z(\mathbf{s}^0, t^0)$ as $\big(\hat{Q}_{\frac{1-p}{2}}(\mathbf{s}^0, t^0), \hat{Q}_{\frac{1+p}{2}}(\mathbf{s}^0, t^0)\big)$, where $\hat{Q}_p(\mathbf{s}^0, t^0)$ denote the p quantile of $\mathcal{N}\big(\hat{Z}(\mathbf{s}^0, t^0), \hat{\sigma^2}_{\mathbf{s}^0,t^0}\big)$, and by construction, the $p-$PI includes the true value of $Z(\mathbf{s}^0, t^0)$ with probability $0 < p < 1$. For a thorough assessment of prediction quality, the accuracy of the predicted value, prediction variance and the $p-$PI should be evaluated. Accordingly, we consider the following commonly used metrics to quantify

prediction performance in our cross-validation: (i) root mean squared error (RMSE), (ii) mean continuous ranked probability score (mCRPS), (iii) mean logarithmic score (mLogS) (Gneiting and Raftery, 2007), (iv) goodness statistic (G) (Deutsch, 1997), (iv) accuracy plot and (v) the average width plot (Qadir et al., 2021). While the RMSE considers accuracy of only the prediction value, mCRPS and mLogS consider both the prediction value and prediction variance to assess the prediction quality. Lower values of RMSE, mCRPS and mLogS indicate superior predictions. The remaining other metrics G, accuracy plot and the average width plot explore the accuracy of the $p-$PI. In particular, $G \in [0, 1]$ quantifies coverage accuracy of the $p-$PI, the accuracy plot visualizes the coverage accuracy through scatter plot of theoretical vs. empirical coverage of the $p-$PI over $p \in (0, 1)$, and average width plot display the width of the $p-$PI as a function of $p \in (0, 1)$. As a rule, higher value of G and points closer to the identity line in the accuracy plot indicates better coverage; and for a fixed coverage accuracy, narrower p-PI is preferred. We compute all these metrics on $\mathcal{V}_i$ and $\mathcal{V}_f$ to provide a comprehensive juxtaposition of the three candidate models in terms of their predictive power.

Figure 4 shows the casewise boxplots for the RMSE, mCRPS, mLogS and G, computed over $\mathcal{V}_i$ for all the three candidate models and Figure 5

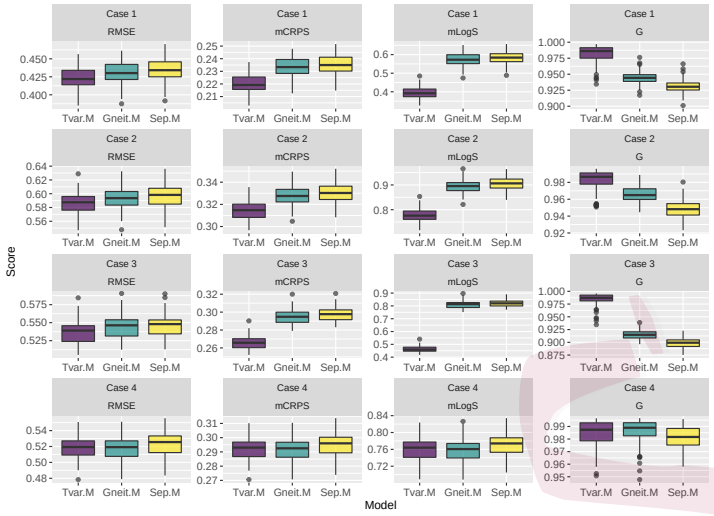

Figure 4: Boxplots of 100 simulation run-based RMSE, mCRPS, mLogS and G, computed over the interpolation validation set ($\mathcal{V}_i$).

shows the same set of boxplots which are computed over $\mathcal{V}_f$ instead. Figure 6 shows the corresponding casewise accuracy plots and average width plots over $\mathcal{V}_i$ and Figure 7 shows those plots for $\mathcal{V}_f$. In terms of interpolation accuracy, the Tvar.M model significantly outperforms the other two candidate models in Cases 1–3 as it produces noticeably higher G and lower RMSE, mCRPS and mLogs against Gneit.M and Sep.M (see Figure 4). In addition, the Tvar.M model exhibits the highest accuracy of interpolation $p-$PI (see accuracy plot in Figure 6) with narrowest p-PI width (see average width plot in Figure 6) among the three candidate models for Cases 1–3. These results are expected since the true underlying spatio-temporal dependence

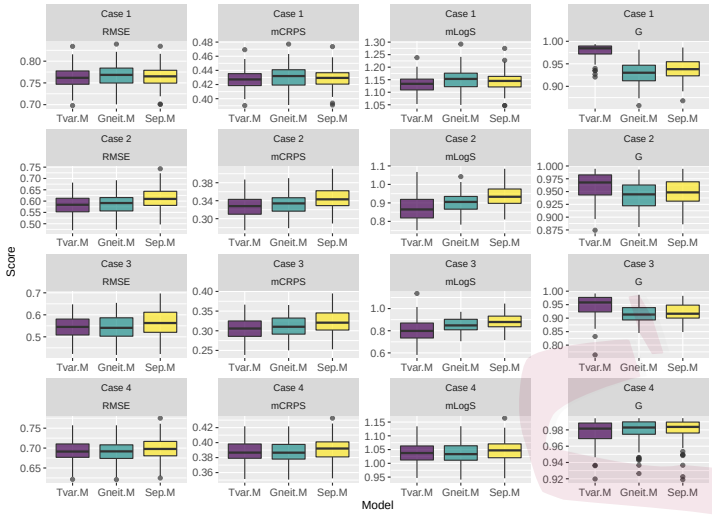

Figure 5: Boxplots of 100 simulation run-based RMSE, mCRPS, mLogS and G, computed over the forecasting validation set ($\mathcal{V}_f$).

of the simulated data in Cases 1–3 is time-varying, and such dependence can be satisfactorily captured only by the Tvar.M among the three candidate models. Furthermore, between Gneit.M model and Sep.M model, the former exhibits better interpolation accuracy, although only slightly, in terms of RMSE, mCRPS and mLogS, but strongly, in terms of G (see Figure 4) for Cases 1–3. The candidate Tvar.M model does not exhibit any improvement in interpolation accuracy over Gneit.M in any of the assessment metrics for Case 4, which is not surprising since the true underlying spatio-temporal dependence in Case 4 is not time-varying. Also, the interpolation accuracy of Sep.M model in Case 4 is lowest among the three candidate models, and

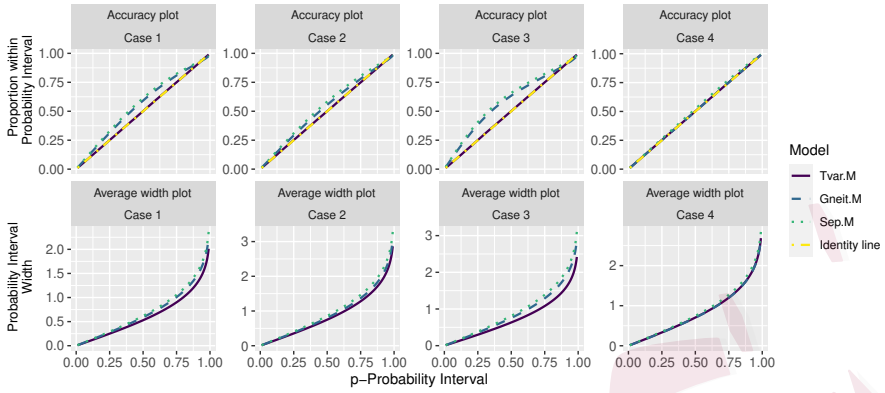

Figure 6: The accuracy plot and the average width plot computed over the validation set for interpolation ($\mathcal{V}_i$) and averaged over the 100 simulation runs.

this is attributed to the high degree of nonseparability ($\beta = 0.8$) in the simulated data. In terms of forecasting accuracy, the improvements by Tvar.M against other candidate models are clearly observed in Cases 1–3, on all the metrics (see Figure 5 and Figure 7), except for RMSE in which the improvement are less evident. By and large, these results endorse the use of Tvar.M against Gneit.M and Sep.M for modeling spatio-temporal data, as the Tvar.M can potentially lead to improved predictions.

5. Data Analysis

In this section, we continue with the data analysis started in Section 2, using the same notations as defined therein. As mentioned earlier in Section 2, it is crucial to consider a class of time-varying spatio-temporal models to

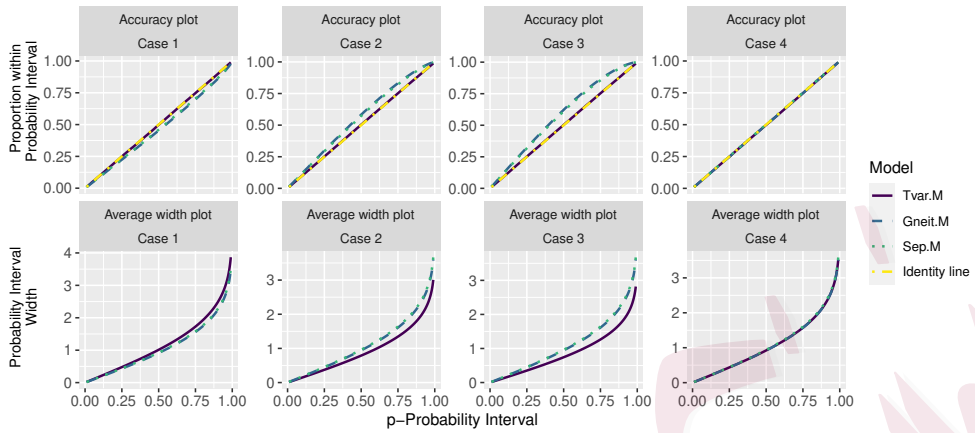

Figure 7: The accuracy plot and the average width plot computed over the validation set for forecasting ($\mathcal{V}_f$), and averaged over the 100 simulation runs.

adequately capture the dependence structure of $\epsilon(\mathbf{s}, t)$. In line with this, we have developed the class of time-varying spatio-temporal covariance models (3.3) in Section 3, thoroughly tested its performance on the simulated data in Section 4, and we now illustrate its application for the $PM_{2.5}$ prediction in the context of Gaussian process modeling. Accordingly, we employ the proposed time-varying class of spatio-temporal covariance functions to model $\epsilon(\mathbf{s}, t)$ as a zero-mean Gaussian process. Moreover, to explore the relative suitability of the proposed model for this dataset, we also consider the Gneiting-Matérn class and the separable class of models as candidates for comparison.

We fit the candidate models: Tvar.M, Gneit.M, and Sep.M, indepen-

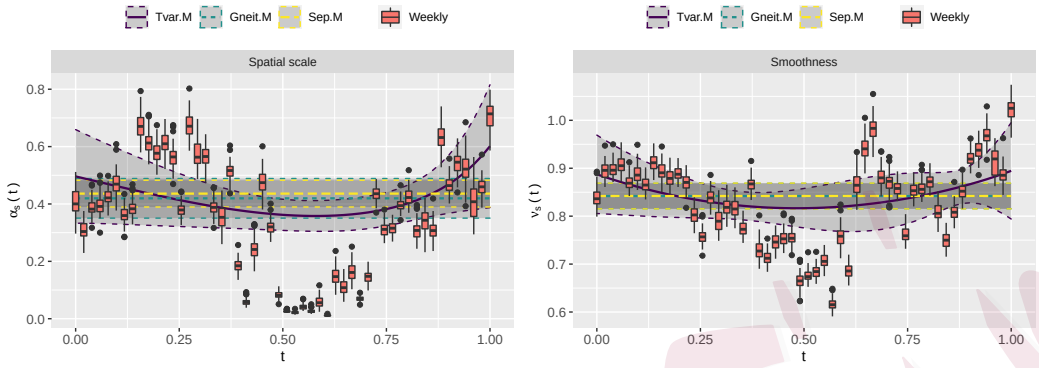

Figure 8: The estimated mean functions $\alpha_s(t)$ (left) and $\nu_s(t)$ (right) along with their respective 95% pointwise-intervals. These mean functions and intervals are overlaid on boxplots representing the independent weekly maximum likelihood estimates of α and ν obtained from the covariance function $\sigma^2\mathrm{M}(\mathbf{h} \mid \alpha, \nu)$. The mean function and 95% pointwise-intervals are calculated based on the pointwise mean and standard deviation of the estimated functions derived from the 100 training sets.

dently over the 100 training sets using the RCL estimation on the $\epsilon(\mathbf{s}, t)$ data. For the Tvar.M model, we need to select appropriate values for $n.\alpha$ and $n.\nu$ to determine the degree of the polynomials in the temporally varying functions $\alpha_s(t)$ and $\nu_s(t)$, respectively. To determine appropriate values for $n.\alpha$ and $n.\nu$, we perform a stepwise polynomial regression-based exploratory analysis. The detailed procedure for this analysis is provided in Supplementary Material Section S5. Based on the results of the exploratory

analysis, we decide to set $n.\alpha = 5$ and $n.\nu = 5$ in the Tvar.M model.

Next, we perform the RCL estimation for the three candidate models independently on the 100 training sets. Figure 8 displays the mean and 95% pointwise-interval of the estimated functions $\alpha_s(t)$ and $\nu_s(t)$, respectively, obtained from the three candidate models based on the 100 training sets. The boxplots from Figure 2 are also included in Figure 8 for the reference. In the Tvar.M model, the mean values of the estimated functions, $\alpha_s(t)$ and $\nu_s(t)$, manifest a general trend that echoes the weekly estimates of α and ν. In particular, the estimated values peak at the boundaries of $\mathcal{D}_t$ and diminish towards the center, suggesting a pattern reminiscent of seasonal variations. Yet, it's essential to acknowledge that the estimated $\alpha_s(t)$ and $\nu_s(t)$ don't perfectly correspond with the nuanced fluctuations observed in the respective weekly estimates. This deviation likely stems from the joint estimation of all parameters, leading to a constrained and smoother behavior that helps prevent discontinuities in spatio-temporal predictions. This suggests that the Tvar.M model captures the temporally evolving nature of the spatio-temporal process $\epsilon(\mathbf{s}, t)$, which can potentially lead to improved spatio-temporal predictions. The other two candidate models, Gneit.M and Sep.M, do not incorporate the time-varying dependence of $\epsilon(\mathbf{s}, t)$ due to their inherent theoretical limitations. As a result, they estimate constant

values for $\alpha_s(t)$ and $\nu_s(t)$ throughout the entire time period.

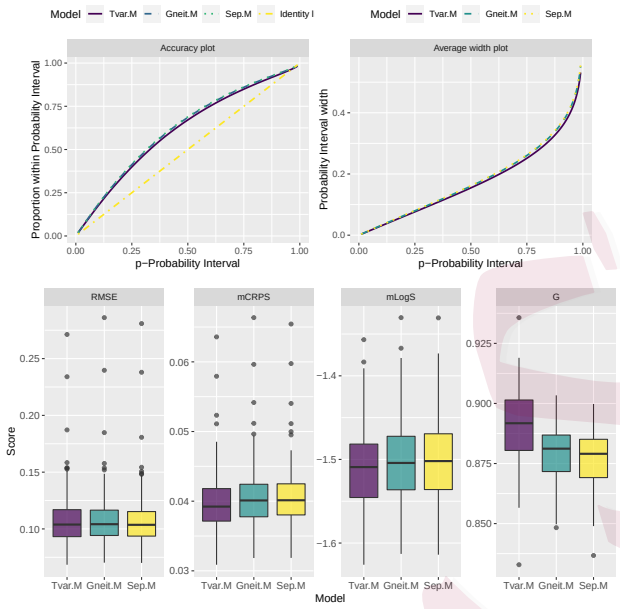

Figure 9: Prediction quality assessment metrics based on the 100 validation sets for the three candidate models. The top row shows the accuracy plot and the average width plot. The bottom row shows the boxplots of RMSE, mCRPS, mLogS and G based on the 100 runs.

While the estimated candidate model, Tvar.M, acceptably conform to the underlying time-varying spatio-temporal dependence of $\epsilon(\mathbf{s}, t)$ (see Figure 8), its usefulness in terms of spatio-temporal prediction still needs to be validated empirically. To this end, we now examine and compare the prediction performance of all the three candidate models for this dataset through a cross-validation study, similar to the one conducted in Section

4. We perform spatio-temporal prediction of $\epsilon(\mathbf{s}, t)$ across the 100 validations sets, through Kriging with their respective estimated candidate models. We evaluate the prediction performance of the three candidate models: Tvar.M, Gneit.M, and Sep.M, on the 100 validation sets, following the prediction quality assessment metrics described in Section 4. These metrics are computed based on the models estimated on the corresponding 100 training sets. Figure 9 shows the boxplots of RMSE, mCRPS, mLogS and G based on the 100 validation sets, for the three candidate models. Additionally, the corresponding accuracy plots and the average width plots are also included in Figure 9.

The accuracy plot and average width plot clearly indicate that among the three candidate models, Tvar.M provides the best prediction intervals in terms of calibration and sharpness. Furthermore, the joint assessment of RMSE, mCRPS, mLogS, and G reveals that Tvar.M outperforms both Gneit.M and Sep.M. Specifically, in terms of mCRPS, mLogS, and G, Tvar.M demonstrates superior performance, while the boxplots of RMSE show that all three candidate models exhibit similar point prediction performance. These findings align with the empirical findings of Fuglstad et al. (2015), highlighting the improvement offered by the nonstationary model, Tvar.M, over the stationary models, Gneit.M and Sep.M, is primarily due

to the improved uncertainty quantification, since the point prediction performance is similar across the three candidate models.

Overall, these results provide substantial support for the modeling of spatio-temporal processes using proposed class of time-varying spatio-temporal covariance model over stationary models. The time-varying model effectively captures subtle nonstationarity features, leading to improved prediction accuracy.

6. Discussion

In this article, we have developed a time-varying class of spatio-temporal models which includes the commonly used Gneiting-Matérn class and separable Matérn class of models as particular cases. Through our simulation study and the $PM_{2.5}$ data application, we have established modeling advantages of our proposed class. Although, the proposed model is stationary in space and nonstationary in time, at the cost of additional complexity it can easily be made nonstationary in space as well by modifying the time-varying functions $\alpha_s(t)$ and $\nu_s(t)$ to the space-time-varying functions $\alpha_s(\mathbf{s}, t)$ and $\nu_s(\mathbf{s}, t)$, respectively. Specifically, if we replace $\alpha_s(t) > 0$ and $\nu_s(t) > 0$ with functions $\alpha_s(\mathbf{s}, t) > 0$ and $\nu_s(\mathbf{s}, t) > 0$, respectively, in Theorem 1, the resulting model is still valid and is nonstationary both in space and time.

In addition to the potential prediction improvements by using a non-stationary model against stationary model, it also reveals some important subtle features of the spatio-temporal data. Moreover, it can help us to determine the sources of nonstationarity which can be then infused in simpler models that can lead to equally well performance (Fuglstad et al., 2015). The statistical analysis of the $PM_{2.5}$ data revealed its time-varying spatio-temporal dependence in which the spatial scale and smoothness are generally lower in the middle of the year, (i.e in spring and summer), and attain peaks near the beginning and the end of the year (i.e., fall and winter). Such a spatio-temporal dependence is paramatrically modeled by the proposed class of time-varying model, and the result of which is an improvement in spatio-temporal predictions of $PM_{2.5}$.

A particular downside of the proposed model is that the time-varying functions $\alpha_s(t)$ and $\nu_s(t)$ can be sometimes misleading for the time point t which is far from the training time-periods. For instance, suppose that the estimated $\nu_s(t)$ is linearly increasing in training time-period, then that would not necessarily mean that the process smoothness will be extremely high for an extremely far time point. The estimated $\nu_s(t)$ in Case 3 of simulation study illustrate this particular downside, since the estimated function $\nu_s(t)$ becomes increasingly misleading as t moves away from the

training time period. Additionally, the proposed model disregards the space-time asymmetry which is a commonly inherited feature in spatio-temporal datasets, therefore, introducing space-time asymmetry to the proposed class is plausible extension to this work. Lastly, it is desirable to extend the proposed class for multivariate setting, a possible direction for which is the development of multivariate analgous of Bernstein functions $\psi_{i,j}(w),\ i,j = 1,\ldots,p, p \geq 1$, and using it to appropriately redefine (3.3).

Supplementary Materials

Section S1 gives the proof for Theorem 1. Section S2 discusses RCL estimation. Section S3 discusses properties of the proposed model. Section S4 provides an extended discussion on the simulation study. Section S5 details the exploratory analysis for the polynomial order selection and its implementation for the simulation study and the data application.

References

Abramowitz, M. and I. A. Stegun (1965). *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, Volume 55. Courier Corporation.

Amoatey, P., H. Omidvarborna, and M. Baawain (2018). The modeling and health risk assessment of pm2.5 from tema oil refinery. *Human and Ecological Risk Assessment: An*

International Journal 24(5), 1181–1196.

Bell, M. L., F. Dominici, K. Ebisu, S. L. Zeger, and J. M. Samet (2007). Spatial and temporal variation in pm2. 5 chemical composition in the united states for health effects studies. *Environmental health perspectives 115*(7), 989–995.

Bhatia, R. and T. Jain (2015). On some positive definite functions. *Positivity 19*(4), 903–910.

Bornn, L., G. Shaddick, and J. V. Zidek (2012). Modeling nonstationary processes through dimension expansion. *Journal of the American Statistical Association 107*(497), 281–289.

Bourotte, M., D. Allard, and E. Porcu (2016). A flexible class of non-separable cross-covariance functions for multivariate space–time data. *Spatial Statistics 18*, 125 – 146. Spatial Statistics Avignon: Emerging Patterns.

Chen, W., M. G. Genton, and Y. Sun (2021). Space-time covariance structures and models. *Annual Review of Statistics and Its Application 8*(1), null.

Cressie, N. (1993). *Statistics for Spatial Data*. Wiley, New york.

Cressie, N. and H.-C. Huang (1999). Classes of nonseparable, spatio-temporal stationary covariance functions. *Journal of the American Statistical Association 94*(448), 1330–1340.

Dawson, J., P. Adams, and S. Pandis (2007). Sensitivity of pm2. 5 to climate in the eastern us: a modeling case study. *Atmos. Chem. Phys 7*, 4295–4309.

De Iaco, S., D. Myers, and D. Posa (2002). Nonseparable space-time covariance models: Some parametric families. *Mathematical Geology 34*(1), 23–42.

de Oliveira, B. F. A., E. Ignotti, P. Artaxo, P. H. do Nascimento Saldiva, W. L. Junger, and S. Hacon (2012). Risk assessment of pm2.5 to child residents in brazilian amazon region with biofuel production. *Environmental Health 11*(1), 64.

Deutsch, C. V. (1997). Direct assessment of local accuracy and precision. *Geostatistics Wollongong 96*, 115–125.

Dominici, F., R. D. Peng, M. L. Bell, L. Pham, A. McDermott, S. L. Zeger, and J. M. Samet (2006). Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. *JAMA 295*, 1127–1134.

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

Fuglstad, G.-A., D. Simpson, F. Lindgren, and H. Rue (2015). Does non-stationary spatial data always require non-stationary random fields? *Spatial Statistics 14*, 505–531.

Garg, S., A. Singh, and F. Ramos (2012). Learning non-stationary space-time models for environmental monitoring. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Volume 26.

Gneiting, T. (2002). Nonseparable, stationary covariance functions for space–time data. *Journal of the American Statistical Association 97*(458), 590–600.

Gneiting, T., M. G. Genton, and P. Guttorp (2006). Geostatistical space-time models, stationarity, separability, and full symmetry. *Monographs On Statistics and Applied Probability 107*, 151.

REFERENCES

Gneiting, T., W. Kleiber, and M. Schlather (2010). Matérn cross-covariance functions for multivariate random fields. *Journal of the American Statistical Association 105*, 1167–1177.

Gneiting, T. and A. E. Raftery (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association 102*, 359–378.

Huang, H.-C. and N.-J. Hsu (2004). Modeling transport effects on ground-level ozone using a non-stationary space–time model. *Environmetrics 15*(3), 251–268.

Ip, R. H. and W. K. Li (2015). Time varying spatio-temporal covariance models. *Spatial Statistics 14*, 269–285.

Li, L., J. Zhang, W. Qiu, J. Wang, and Y. Fang (2017). An ensemble spatiotemporal model for predicting pm2. 5 concentrations. *International journal of environmental research and public health 14*(5), 549.

Ma, C. (2002). Spatio-temporal covariance functions generated by mixtures. *Mathematical geology 34*(8), 965–975.

Matérn, B. (1986). *Spatial Variation* (2nd ed.). Berlin:Springer-Verlag.

Myers, D. E. and A. Journel (1990). Variograms with zonal anisotropies and noninvertible kriging systems. *Mathematical Geology 22*(7), 779–785.

Pope III, C. A. and D. W. Dockery (2006). Health effects of fine particulate air pollution: Lines that connect. *Journal of the Air & Waste Management Association 56*, 709–742.

Porcu, E., P. Gregori, and J. Mateu (2006). Nonseparable stationary anisotropic space–time covariance functions. *Stochastic Environmental Research and Risk Assessment 21*(2), 113–122.

Porcu, E., J. Mateu, and M. Bevilacqua (2007). Covariance functions that are stationary or nonstationary in space and stationary in time. *Statistica Neerlandica 61*(3), 358–382.

Qadir, G. A., C. Euán, and Y. Sun (2020). Flexible modeling of variable asymmetries in cross-covariance functions for multivariate random fields. *Journal of Agricultural, Biological and Environmental Statistics*.

Qadir, G. A. and Y. Sun (2020). Semiparametric estimation of cross-covariance functions for multivariate random fields. *Biometrics*, 1–14.

Qadir, G. A., Y. Sun, and S. Kurtek (2021). Estimation of spatial deformation for nonstationary processes via variogram alignment. *Technometrics 0*(ja), 1–28.

Shand, L. and B. Li (2017). Modeling nonstationarity in space and time. *Biometrics 73*(3), 759–768.

Sigrist, F., H. R. Künsch, and W. A. Stahel (2012). A dynamic nonstationary spatio-temporal model for short term prediction of precipitation. *The Annals of Applied Statistics 6*(4), 1452 – 1477.

Stein, M. L. (1999). *Interpolation of spatial data: Some theory for kriging*. Springer-Verlag New York.

REFERENCES

Stein, M. L. (2005). Space–time covariance functions. *Journal of the American Statistical Association 100*(469), 310–321.

Stroud, J. R., P. Müller, and B. Sansó (2001). Dynamic models for spatiotemporal data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 63*(4), 673–689.

Van Den Berg, C. and G. Forst (2012). *Potential theory on locally compact abelian groups*, Volume 87. Springer Science & Business Media.

Xiao, L., Y. Lang, and G. Christakos (2018). High-resolution spatiotemporal mapping of pm2. 5 concentrations at mainland china using a combined bme-gwr technique. *Atmospheric Environment 173*, 295–305.

Xu, H. and P. Gardoni (2018). Improved latent space approach for modelling non-stationary spatial–temporal random fields. *Spatial Statistics 23*, 160–181.

Zhang, G., C. Ding, X. Jiang, G. Pan, X. Wei, and Y. Sun (2020). Chemical compositions and sources contribution of atmospheric particles at a typical steel industrial urban site. *Scientific Reports 10*(1), 7654.

Zhao, X., W. Zhou, L. Han, and D. Locke (2019). Spatiotemporal variation in pm2. 5 concentrations and their relationship with socioeconomic factors in china's major cities. *Environment international 133*, 105145.

Heidelberg Institute for Theoretical Studies; E-mail: ghulam.qadir@h-its.org

King Abdullah University of Science and Technology; E-mail: ying.sun@kaust.edu.sa