Research Article

Hierarchical Bayesian Spatio-Temporal Modeling for PM$_{10}$ Prediction

Esam Mahdi (ID)**, Sana Alshamari** (ID)**, Maryam Khashabi** (ID)**, and Alya Alkorbi** (ID)

Department of Mathematics, Statistics and Physics, Qatar University, Doha, Qatar

Correspondence should be addressed to Esam Mahdi; emahdi@qu.edu.qa

Over the past few years, hierarchical Bayesian models have been extensively used for modeling the joint spatial and temporal dependence of big spatio-temporal data which commonly involves a large number of missing observations. This article represented, assessed, and compared some recently proposed Bayesian and non-Bayesian models for predicting the daily average particulate matter with a diameter of less than 10 (PM$_{10}$) measured in Qatar during the years 2016–2019. The disaggregating technique with a Markov chain Monte Carlo method with Gibbs sampler are used to handle the missing data. Based on the obtained results, we conclude that the Gaussian predictive processes with autoregressive terms of the latent underlying space-time process model is the best, compared with the Bayesian Gaussian processes and non-Bayesian generalized additive models.

1. Introduction

Many of the environmental data contain different scales of variability over space and time. For example, scientists from environmental and public health sciences are typically interested in modeling the evolving process of the air pollution during the time over specified locations. Such a stochastic process is often high-dimensional, large, and complicated with nonstationary structures, so the traditional statistical methods are hampered by the need of advanced statistical techniques to specify the spatio-temporal dependency. This can be quite practical with modern computers with high-level computational programming. The spatio-temporal modeling of PM$_{10}$ and PM$_{2.5}$ (particulate matter with diameters of less than 10 and 2.5 micrometers, respectively) is rapidly becoming an important component of most air quality studies [1–3]. Particulate matter (also called particle pollution) is a mixture of solid particles and liquid droplets found in the air as a result of dust, soot, dirt, smoke caused by road transportation, and complex chemical reactions in the atmosphere such as sulfur dioxide and nitrogen oxides. Exposure to particle pollution is a public health hazard and can cause

acute and chronic heart and lung diseases [4]. The larger the values of particulate matters (PM) are, the more harm on short and long terms of public health is. World Health Organization's (WHO) air quality guidelines recommend that the annual and 24-hour mean concentrations should not, respectively, exceed 20 and 50 microgram per cubic metre (μg/m^3) for PM$_{10}$, and 10 and 25 μg/m^3 for PM$_{2.5}$. Countries with fast-developed infrastructure such as Qatar usually suffer from relatively high levels of PM air pollutant. In this regard, the WHO classified the air quality in Qatar as poor and unsafe. In this paper, we focus our attention on modeling the PM$_{10}$ in Qatar because the PM$_{2.5}$ data is inaccessible. The most recent data indicates that country's annual average 24 hr PM$_{10}$ concentration levels were ranged from 126.69 μg/m^3 to 184.55 μg/m^3, which exceeds the recommended maximum of 50 μg/m^3 [4–6].

Several authors have developed spatio-temporal models for analyzing the ambient of air pollution. Research in this field is back in dates to Cressie [7] and Goodall and Mardia [8]. Cressie et al. [9] compared the performance of Markov-random field with the familiar geestatistical approach in prediction the PM$_{10}$ concentrations around the city of

Pittsburgh, United States of America. The authors did not model the joint temporal and spatial structure of observations. They first modeled the purely spatial structure of observations for one particular day, then they incorporated the temporal component in their final statistical modeling. Sun et al. [10] developed a spatial forecasting distribution for unmeasured daily log PM_{10} average concentration given from ten locations in Vancouver, Canada. At each monitoring site, Sun et al. [10] showed that the autoregressive model of order one (AR(1)) described quite well the daily log PM_{10} average values. The authors did not consider the Bayesian models in their approach. Golam Kibria et al. [11] proposed a multivariate spatial prediction methodology in a Bayesian approach that is suited for spatio-temporal data observed at a small set of ambient monitoring stations at successive time points. They demonstrated the usefulness of their approach by mapping $PM_{2.5}$ at monitoring sites with different start-up times in the city of Philadelphia, USA. Golam Kibria et al. [11] did not compare the performance of their model with other non-Bayesian models that are commonly used in literature. Shaddick and Wakefield [12] and Sahu and Mardia [13] used short-term spatio-temporal predictive analysis for modeling the $PM_{2.5}$ and PM_{10} concentration levels. Zidek et al. [3] presented predictive distributions on nonmonitored PM_{10} values in Vancouver, Canada. Smith et al. [14] proposed a spatio-temporal model to predict the weekly averages of particulate matter concentrations within three southeastern states in the USA. Sahu et al. [15] modeled the $PM_{2.5}$ by combining the rural background and the urban areas into one process. Cocchi et al. [1] developed a hierarchical Bayesian model for the daily average PM_{10} values. Pollice and Lasinio [2] developed a Bayesian-based kriging method for estimating the daily average PM_{10} concentration levels. Wikle et al. [16] provided an excellent review of classical and Bayesian approaches for analyzing spatio-temporal data. Although Taylor et al. [6], Ahmadi et al. [5], and others have studied the relationship between some environmental conditions and particulate matter levels assessing the air quality based on particulate matter levels in different locations in Qatar, they did not study the spatio-temporal variability of PM_{10}. The main objective of this research is to develop space-time models for daily PM_{10} air pollution levels in Qatar for the four years, 2016-2019 comparing the hierarchical Bayesian approach with other spatio-temporal recent methods. We develop spatial interpolation and forecasting model using iterative Markov chain Monte Carlo (MCMC) computation setup which is an effective method for modeling a data with large number of missing values [17]. To the best of our knowledge, this will be the first study in Qatar, and we hope that this research will be helpful to protect the environment and public health in Qatar.

The rest of this article is organized as follows: Section 2 provides a brief review of two-stage hierarchical Bayesian models that have been used for modeling spatio-temporal data. A numerical example is given in Section 3 to demonstrate that the Bayesian approach accurately predicts the daily average PM_{10} values with a large number of missing values comparing with non-Bayesian models. Finally, a conclusion is given in Section 4.

2. Hierarchical Spatio-Temporal Model

When data is collected at different points in space and time, we should use a model that can, simultaneously, describe the dependency structure coming from the three sources of variations: time variation, space variation, and joint variability between time and space. Such a model is called a space-time model (or spatio-temporal model, where spatio refers to space and temporal refers to time).

In this article, we develop hierarchical models to predict the daily PM_{10} concentration levels which vary over time and locations. The PM_{10} concentration levels do not often follow the normal distribution. Thus, we usually model these values on the square-root scale or we use the log-transformation to stabilize the variance and enforce normality and to stabilize the variance [18]. We consider the square-root scale to alleviate the departure from normality in our research data.

Let ℓ and t denote the two units of time where $\ell = 1, \cdots, r$ represents the longer unit (e.g., year), and $t = 1, \cdots, T_\ell$ represents the shorter unit (e.g., day). Let $Z_\ell(\mathbf{s}, t)$ denote the observed value of the PM_{10} concentration, after any necessary transformation, at a given location $\mathbf{s}$ and over a given discrete time t. We assume that the spatial location $\mathbf{s}$ is a two-dimensional vector describing the latitude-longitude (or equivalently northing and easting coordinates), and the time unit is typically hour, day, month, or year. We also assume that the $Z_\ell(\mathbf{s}, t)$ is observed at n monitoring sites denoted by $\mathbf{s}_i, i = 1, \cdots, n$ and at time points denoted by two indices ℓ and t so that the total number of observations is denoted by $N = n\sum_{\ell=1}^{r} T_\ell$. In this article, we denote all the missing data by $\mathbf{z}^\star$, whereas all the observed data will be denoted by $\mathbf{z}$.

The first stage of the hierarchy assumes that the observed values $\mathbf{Z}_{\ell t}$, where $\mathbf{Z}_{\ell t} = (Z_\ell(\mathbf{s}_1, t), \cdots, Z_\ell(\mathbf{s}_n, t))'$, can be decomposed into a true (latent) spatio-temporal process $\mathbf{Y}_{\ell t} = (Y_\ell(\mathbf{s}_1, t), \cdots, Y_\ell(\mathbf{s}_n, t))'$ with an error term $\varepsilon_{\ell t} = (\varepsilon_\ell(\mathbf{s}_1, t), \cdots, \varepsilon_\ell(\mathbf{s}_n, t))'$. More specifically, the data (or measurement error) model in the first stage of the hierarchy is

$$\mathbf{Z}_{\ell t} = \mathbf{Y}_{\ell t} + \varepsilon_{\ell t}, \ell = 1, \cdots, r, t = 1, \cdots, T_\ell. \tag{1}$$

The error term $\varepsilon_{\ell t}$ is assumed to be a Gaussian white noise process with mean zero and constant variance σ_ε^2, which is often called the nugget effect absorbing microscale variability.

The second stage assumes that the true process $\mathbf{Y}_{\ell t}$ has a systematic mean $\mu_{\ell t}$ and a spatio-temporal error term. The mean can be modeled based on the past values of the unobserved variable or/and based on some relevant covariates. Typically, $\mathbf{Y}_{\ell t}$ can be specified in the following formula:

$$\mathbf{Y}_{\ell t} = \mu_{\ell t} + \eta_{\ell t}, \tag{2}$$

where $\eta_{\ell t} = (\eta_\ell(s_1, t), \cdots, \eta_\ell(s_n, t))'$ is an spatio-temporal residual random intercept assumed to follow $\mathcal{N}(\mathbf{0}, C)$, where $C = \sigma_\eta^2 \mathscr{H}_\eta$, σ_η^2 is the site invariant spatial variance, and $\mathscr{H}_\eta$ is the spatial correlation matrix. In this article, we consider

four modelings for $\{\eta_{\ell t}\}$. The first one assumes that the random effects, $\eta_\ell(s_i, t) = 0$, for all locations s_i and times t. This implies that the model in (1)–(2) will be the simple regression model. The other models for $\{\eta_{\ell t}\}$ are assumed to be Gaussian process, independent over the time, which is specified in Sections 2.1, 2.2, 2.3, and 2.4.

2.1. Matérn Spatio-Temporal Covariance Function. For spatio-temporal modeling, we usually assume that the random effects process is a weakly stationary Gaussian with a zero mean and a valid isotropic covariance function. A valid covariance function implied that the covariance matrix is positive definite, and isotropic means that the separation vector between the two locations only depends on the distance and not on the direction. The class of the spatio-temporal covariance functions can be separable, product-sum, metric, and sum-metric [19]. In this paper, we use the separable covariance model which is simply the product of the pure spatial covariance function, $C_\mathbf{s}(\mathbf{h})$, by the pure temporal one, $C_t(u)$, given by

$$C(\mathbf{h}, u) = C_\mathbf{s}(\mathbf{h})C_t(u), \tag{3}$$

where $\mathbf{h} = \|\mathbf{s} - \mathbf{s}'\|$ is the separating spatial distance, and $u = |t - t'|$ is the temporal distance for any pair of points $(\mathbf{s}, t) \times (\mathbf{s}', t')$ in the spatial and temporal study domain.

The Matérn family provides a general choice of covariance functions. For each time t, the Matérn covariance is given by:

$$\begin{aligned} & C\left(w(\mathbf{s}, t), w\left(\mathbf{s}', t\right)\right) \\ &= \frac{\sigma_\eta^2}{2^{\kappa-1}\Gamma(\kappa)} \left(2\sqrt{\kappa}\|\mathbf{s} - \mathbf{s}'\|\phi\right)^\kappa K_\kappa \\ & \quad \cdot \left(2\sqrt{\kappa}\left(2\sqrt{\kappa}\|\mathbf{s} - \mathbf{s}'\|\phi\right)\phi\right), \quad \phi > 0, \kappa \geq 1, \end{aligned} \tag{4}$$

where $K_\kappa(\cdot)$ is the modified Bessel function of second kind of order κ which is a parameter controlling the smoothness of the realized random field [20], $\Gamma(\kappa)$ is the standard gamma function, and ϕ is a parameter which controls the decay rate in the correlation as the distance $\mathbf{h}$ increases. Popular special cases of Matérn model are (i) $\kappa = 0.5$ leads to exponential covariance function $C(\mathbf{h}) = \sigma_\eta^2 \exp(-\phi\mathbf{h})$ and (ii) Gaussian model, $C(\mathbf{h}) = \sigma_\eta^2 \exp(-\phi^2\mathbf{h}^2)$ when $\kappa \longrightarrow \infty$.

2.2. Gaussian Process Model Specification. Consider $\mu_{\ell t} = \mathbf{X}_{\ell t}\beta$ in (2), where $\ell = 1, \cdots, r, t = 1, \cdots, T_\ell$, $\mathbf{X}_{\ell t}$ is the $n \times (k+1)$ design matrix of spatially and temporally varying k-covariates and $\beta = (\beta_0, \beta_1, \cdots, \beta_k)'$ is the $k+1$-dimensional vector of regression coefficients. Thus, the Gaussian process (GP) two-stage model can be written as

$$\begin{aligned} \mathbf{Z}_{\ell t} &= \mathbf{Y}_{\ell t} + \varepsilon_{\ell t}, \\ \mathbf{Y}_{\ell t} &= \mathbf{X}_{\ell t}\beta + \eta_{\ell t}. \end{aligned} \tag{5}$$

We assume that the random effect process $\{\eta(\cdot)\}$ is independent from the white noise process $\{\varepsilon(\cdot)\}$. Note that some of the covariates may vary spatially and not temporally or vice versa.

One advantage of using Bayesian model is that we can use it to handle any missing data. This can be done by using the Markov chain Monte Carlo, MCMC, computation where the missing data is simulated from the $\mathcal{N}(\mathbf{Y}_{\ell t}, \sigma_\varepsilon^2)$ distribution defined by (5) at each MCMC iteration using Gibbs sampling. The Gibbs sampler requires that the full conditional distributions of the parameters $\theta = (\beta, \sigma_\varepsilon^2, \sigma_\eta^2, \phi, \kappa)$ are given in a closed form.

The logarithm of the joint posterior distribution of the missing data and the parameters in this case is given by:

$$\begin{aligned} & \log p(\theta, \mathbf{Y}, \mathbf{z}^\star \mid \mathbf{z}) \\ & \propto -\frac{N}{2} \log \sigma_\varepsilon^2 - \frac{1}{2\sigma_\varepsilon^2} \sum_{\ell=1}^r \sum_{t=1}^{T_\ell} (\mathbf{Z}_{\ell t} - \mathbf{Y}_{\ell t})'(\mathbf{Z}_{\ell t} - \mathbf{Y}_{\ell t}) \\ & \quad - \frac{1}{2} \sum_{\ell=1}^r T_\ell \log \left|\sigma_\eta^2 \mathscr{H}_\eta\right| - \frac{1}{2\sigma_\eta^2} \sum_{\ell=1}^r \sum_{t=1}^{T_\ell} \\ & \quad \cdot (\mathbf{Y}_{\ell t} - \mathbf{X}_{\ell t}\beta)' \mathscr{H}_\eta^{-1} (\mathbf{Y}_{\ell t} - \mathbf{X}_{\ell t}\beta) + \log p(\theta), \end{aligned} \tag{6}$$

where $p(\theta)$ is the prior distribution [21] and we refer the readers to Bakar and Sahu ([21], Appendix A) to obtain the Gibbs sampling using the full conditional distributions of θ.

2.3. Autoregressive Model Specification. The autoregressive process (AR) model in Equations (1) and (2) can be written in the hierarchical form (e.g., [18] model) as follows:

$$\begin{aligned} \mathbf{Z}_{\ell t} &= \mathbf{Y}_{\ell t} + \varepsilon_{\ell t}, \\ \mathbf{Y}_{\ell t} &= \rho\mathbf{Y}_{\ell t-1} + \mathbf{X}_{\ell t}\beta + \eta_{\ell t}, \end{aligned} \tag{7}$$

where $-1 < \rho < 1$ is parameter of the first-order autoregressive model and $\mu_\ell = \rho\mathbf{Y}_{\ell t-1} + \mathbf{X}_{\ell t}\beta$. Note that if $\rho = 0$, we get the GP model given by (5). Also, note that the autoregressive model requires to specify an independent spatial model with initial values of $\mathbf{Y}_{\ell 0}$ for each $\ell = 1, \cdots, r$, with mean μ_ℓ and covariance $\sigma_\ell^2 \mathscr{H}_0$ obtained from the Matérn covariance function in (4) with the same set of parameters. In this case, logarithm of the joint posterior distribution of the missing data and the parameters in this case is given by:

$$\begin{aligned} & \log p(\theta, \mathbf{Y}, \mathbf{z}^\star \mid \mathbf{z}) \\ & \propto -\frac{N}{2} \log \sigma_\varepsilon^2 - \frac{1}{2\sigma_\varepsilon^2} \sum_{\ell=1}^r \sum_{t=1}^{T_\ell} (\mathbf{Z}_{\ell t} - \mathbf{Y}_{\ell t})'(\mathbf{Z}_{\ell t} - \mathbf{Y}_{\ell t}) \\ & \quad - \frac{1}{2} \sum_{\ell=1}^r T_\ell \log \left|\sigma_\eta^2 \mathscr{H}_\eta\right| - \frac{1}{2\sigma_\eta^2} \sum_{\ell=1}^r \sum_{t=1}^{T_\ell} \\ & \quad \cdot (\mathbf{Y}_{\ell t} - \rho\mathbf{Y}_{\ell t-1} - \mathbf{X}_{\ell t}\beta)' \mathscr{H}_\eta^{-1} (\mathbf{Y}_{\ell t} - \rho\mathbf{Y}_{\ell t-1} - \mathbf{X}_{\ell t}\beta) \\ & \quad - \frac{1}{2} \sum_{\ell=1}^r \log \left|\sigma_\ell^2 \mathscr{H}_0\right| - \frac{1}{2} \sum_{\ell=1}^r \frac{1}{\sigma_\ell^2} (\mathbf{Y}_{\ell 0} - \mu_\ell)' \mathscr{H}_0^{-1} \\ & \quad \cdot (\mathbf{Y}_{\ell 0} - \mu_\ell) + \log p(\theta), \end{aligned} \tag{8}$$

Table 1: Summary statistics of the daily PM_{10} pollutant measured in $\mu g/m^3$, temperature (denoted by Temp) measured in degree Celsius, and relative humidity (RH) obtained by air pointer and the meteorological station at Qatar University site. n and SD stand for sample size and standard deviation, respectively. The % n.a. denotes the percentage of missing values in the data.

|	Year	n	% n.a.	Min	Max	Median	Mean	SD
	2016	348	4.92	18.35	659.49	76.95	102.63	89.23
	2017	257	2.28	66.22	1379.13	103.15	154.60	157.84
PM_{10}	2018	080	52.66	25.89	551.51	63.03	80.67	68.90
	2019	201	15.90	37.87	1542.57	72.29	117.46	157.24
	Overall	**886**	**10.96**	**18.35**	**1542.57**	**83.35**	**119.09**	**104.68**
	2016	363	0.82	8.76	39.08	28.32	27.85	6.89
	2017	221	15.97	10.86	41.67	33.10	29.77	7.67
Temp	2018	134	20.71	18.57	37.63	28.66	28.62	4.32
	2019	126	47.28	22.28	40.28	33.75	32.37	5.12
	Overall	**844**	**8.70**	**8.76**	**41.67**	**31.40**	**29.14**	**6.30**
	2016	363	0.82	15.13	92.77	56.34	53.67	18.31
	2017	221	15.97	18.16	86.94	52.12	50.57	16.74
RH	2018	134	20.71	23.40	89.05	57.74	55.19	16.86
	2019	126	47.28	20.83	80.29	46.40	48.39	15.99
	Overall	**844**	**8.70**	**15.13**	**92.77**	**50.81**	**52.31**	**16.66**
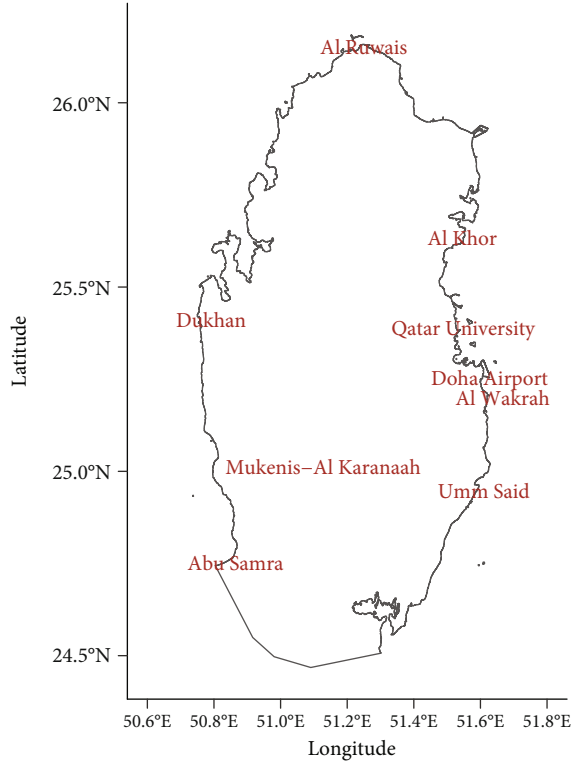
Figure 1: A map of the state of Qatar showing locations of the 9 monitoring sites.

where $p(\theta)$ is the prior distribution for the parameter $\theta = (\beta, \rho, \sigma_\varepsilon^2, \sigma_\eta^2, \phi, \kappa, \mu_\ell, \sigma_\ell^2)$ ([21], see Appendix B).

2.4. Gaussian Predictive Processes with Autoregressive Model Specification. We now introduce the case that the process $\{\eta_{\ell t}\}$ in (2) has a nonstationary covariance structure. We

Table 2: Summary statistics of the available temperature in Celsius, and relative humidity measured at 9 monitoring sites in Qatar; SD stands for standard deviation.

	Year	Min	Max	Median	Mean	SD
	2016	16.50	38.00	27.65	27.48	6.57
	2017	16.30	37.90	29.70	27.90	6.99
Temp	2018	16.40	37.80	28.90	28.21	6.44
	2019	18.80	37.60	28.50	28.00	6.42
	Overall	**16.30**	**38.00**	**28.30**	**27.90**	**6.59**
	2016	25.00	74.00	57.50	55.28	11.74
	2017	23.00	74.00	57.00	53.92	11.27
RH	2018	24.00	73.00	54.00	52.72	11.70
	2019	27.00	69.00	56.00	54.16	10.33
	Overall	**23.00**	**74.00**	**56.00**	**54.02**	**11.27**

follow Sahu and Mukhopadhyay [22] by choosing the integer m out of the sample prediction performance (crossvalidation) to get a smaller number of locations (denoted by knot), $\mathbf{S}_m^\star = (s_1^\star, \cdots, s_m^\star)'$. Given $\mathbf{S}_m^\star$, we assume that $\eta_\ell^\star = (\eta_\ell(s_1^\star, t), \cdots, \eta_\ell(s_m^\star, t))'$ is a Gaussian process with zero mean vector and covariance function given in (4). Thus, the non-stationary model is defined after replacing $\eta_\ell(s_i, t)$ in (2) by

$$\tilde{\eta}(s_i, t) = \mathrm{E}[\eta(s_i, t) \mid \eta_t^\star], \quad i = 1, \cdots, n, \qquad (9)$$

hence, $(\eta_{\ell t}, \eta_{\ell t}^\star)$ is a vector of $n + m$-independent realization at each time point ℓ and t with the same Gaussian process in (4). Define $\tilde{\eta}_\ell = (\tilde{\eta}_\ell(s_1, t), \cdots, \tilde{\eta}_\ell(s_n, t))'$, we have

$$\tilde{\eta}_t = C^\star(\phi, \kappa)\mathcal{H}_{\eta^\star}^{-1}(\phi, \kappa)\eta_t^\star, \qquad (10)$$
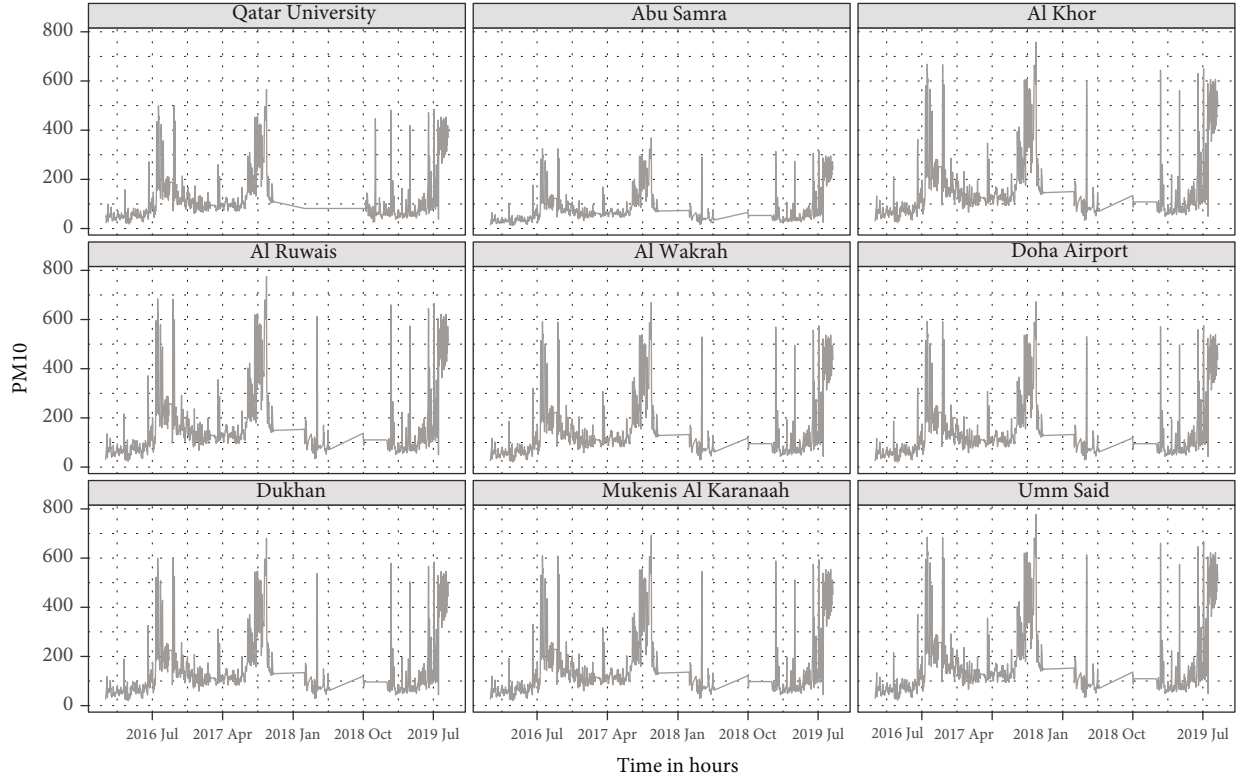
FIGURE 2: Daily PM_{10} concentrations observed in nine locations over the years 2016-2021.

where $C^{\star}(\phi, \kappa)$ is the $n \times m$ crosscorrelation matrix between η_t and $\eta_t^{\star}$ having elements $[C^{\star}]_{ij} = C(\|\mathbf{s}_i - \mathbf{s}_j^{\star}\|)$, $i = 1, \cdots, n$ and $j = 1, \cdots, m$, and $\mathcal{H}_{\eta^{\star}}(\phi, \kappa)$ is the $m \times m$ correlation matrix of $\eta_t^{\star}$ so that $[\mathcal{H}_{\eta^{\star}}]_{ij} = [C(\|\mathbf{s}_k^{\star} - \mathbf{s}_j^{\star}\|; \phi, \kappa)]$, for $k, j = 1, \cdots, m$. Clearly, the process $\{\tilde{\eta}_{\ell t}\}$ shows nonstationary structure with variance function that is given by

$$\mathrm{Var}(\tilde{\eta}_t) = C^{\star}(\phi, \kappa) \mathcal{H}_{\eta^{\star}}^{-1}(\phi, \kappa) C^{\star'}(\phi, \kappa). \qquad (11)$$

The advantage of using the nonstationary model in (10) is the flexibility in the $\tilde{\eta}_t$ surface which is based on $m \ll n$ linear functions of $\eta_t^{\star}$. When m is very small compared with n, this will lead to reduce the computational burden, especially for big data which is usually the case of the spatio-temporal data. Moreover, using nonstationary models usually provides more accurate results in prediction the nonstationary PM_{10} process. We specify the hierarchical Gaussian predictive processes (GPP) as follows:

$$\begin{aligned} \mathbf{Z}_{\ell t} &= \mathbf{Y}_{\ell t} + \varepsilon_{\ell t}, \\ \mathbf{Y}_{\ell t} &= \mathbf{X}_{\ell t}\beta + \tilde{\eta}_{\ell t}, \end{aligned} \qquad (12)$$

where $\tilde{\eta}_{\ell t}$ is given in (10). The process $\{\eta_{\ell t}^{\star}\}$, at the $\mathbf{S}_m^{\star}$ knots, can be modeled according to the autoregressive model

$$\eta_{\ell t}^{\star} = \rho \eta_{\ell t-1}^{\star} + \eta_{\ell t}, \qquad (13)$$

where $\eta_{\ell t}^{\star} \sim \mathcal{N}_m(\mathbf{0}, \Sigma_{\eta^{\star}})$ for $\ell = 1, \cdots, r$, $t = 1, \cdots, T_{\ell}$ and $\Sigma_{\eta^{\star}} = \sigma_{\eta}^2 \mathcal{H}_{\eta^{\star}}$. We assume that the initial conditions $\eta_{\ell 0}^{\star}$ has normal distribution with mean zero and covariance matrix $\Sigma_0 = \sigma_{\ell}^2 \mathcal{H}_0$, and both $\Sigma_{\eta^{\star}}, \Sigma_0$ can be obtained from the Matérn covariance function defined in (4), where $\Sigma_{\eta^{\star}}$ is an $m \times m$ matrix much lower dimensional than Σ_{η} of dimension $n \times n$.

In this case, logarithm of the joint posterior distribution of the missing data and the parameters in this case is given by:

$$\begin{aligned} \log &\, p(\theta, \mathbf{Y}, \mathbf{z}^{\star} \mid \mathbf{z}) \\ \propto &\, -\frac{N}{2} \log \sigma_{\varepsilon}^2 - \frac{1}{2\sigma_{\varepsilon}^2} \sum_{\ell=1}^{r} \sum_{t=1}^{T_{\ell}} (\mathbf{Z}_{\ell t} - \mathbf{Y}_{\ell t})'(\mathbf{Z}_{\ell t} - \mathbf{Y}_{\ell t}) \\ &\, -\frac{mN}{2} \log \left(\sigma_{\eta}^2\right) - \frac{N}{2} \log \left(|\mathcal{H}_{\eta^{\star}}|\right) \\ &\, -\frac{m}{2} \sum_{\ell=1}^{r} \log \left(\sigma_{\ell}^2\right) - \frac{r}{2} \log \left(|\mathcal{H}_0|\right) \\ &\, -\frac{1}{2\sigma_{\eta}^2} \sum_{\ell=1}^{r} \sum_{t=1}^{T_{\ell}} (\eta_{\ell t}^{\star} - \rho \eta_{\ell t-1}^{\star})' \mathcal{H}_{\eta^{\star}}^{-1} (\eta_{\ell t}^{\star} - \rho \eta_{\ell t-1}^{\star}) \\ &\, -\frac{1}{2} \sum_{\ell=1}^{r} \frac{1}{\sigma_{\ell}^2} \eta_{\ell 0}^{\star'} \mathcal{H}_0^{-1} \eta_{\ell 0}^{\star} + \log p(\theta), \end{aligned} \qquad (14)$$

where $p(\theta)$ is the prior distribution for the parameter $\theta = (\beta, \rho, \sigma_{\varepsilon}^2, \sigma_{\eta^{\star}}^2, \phi, \kappa, \sigma_{\ell}^2)$ and the Gibbs sampling procedure of these parameters can be obtained from the full conditional distributions required provided in the Appendix.
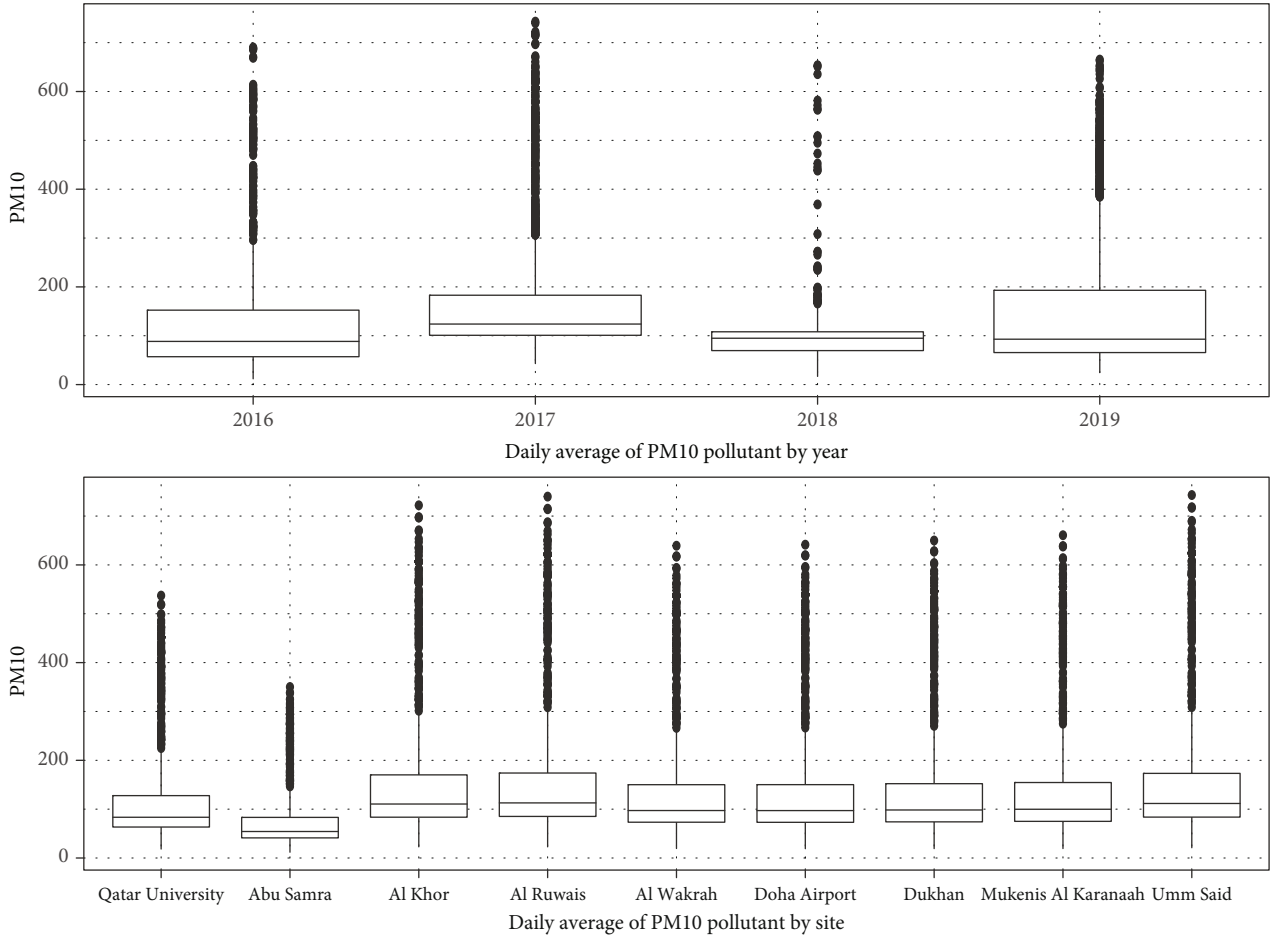
FIGURE 3: Boxplots of daily average of PM_{10} concentrations over the years 2016-2019 and by 9 locations in Qatar showing the quartiles; the whiskers show the farthest observation from the median.

3. Numerical Example

3.1. Data Description. The data used in this article is obtained from three different sources. The first one was collected by the air pointer and the meteorological station and is managed by the Environmental Science Center at Qatar University over the years 2016-2019. It has hourly pollutant including PM_{10} measured in $\mu g/m^3$, temperature, Temp, measured in degree Celsius, and relative humidity, RH, with several missing values.

First, we sort the data into an ascending order by date, and then, we impute the missing values of the PM_{10} by averaging the two nonmissing values before and after this missing value. When two or more successive missing values exist, we impute them by the corresponding monthly average. We do the same for the hourly missing values of temperature and relative humidity. After that, we aggregate the hourly data into daily data for which we fit our models. The amended data from this resource is irregular time series which started on the fourth of January 2016 and ended on the twenty-eighth of August 2019. Table 1 provides a summary statistics of this data. Results suggest that the PM_{10} levels increased during the time, and the majority of missing

observations (with more than 52%) have been occurred during the year 2018.

The second source of the data was the regular monthly observations obtained from nine meteorological stations over the years 2016-2019. The total number of observations is 432 where the stations are located in the following sites: Qatar University, Abu Samra, Al Khor, Al Ruwais, Al Wakrah, Doha Airport, Dukhan, Mukenis-Al Karanaah, and Umm Said. Figure 1 shows the map locations of 9 meteorological sites. Although this data provides the monthly average values of temperature in degree Celsius and relative humidity (RH), it does not include measures for any type of pollutant. Table 2 summarizes the descriptive statistics of this data. We use this data to interpolate the daily temperature and relative humidity by disaggregating technique at each location. We also use our own portable devices to collect daily PM_{10} values, as a third source of data, over the period of time from November 5th to December 31th of 2020 close to these sites. We use these devices to collect and simulate more daily PM_{10} concentrations over the aforementioned locations. In particular, we recorded the average of the PM_{10} taken randomly in the morning and evening of each day over this period. Then, we use this collected data to simulate a new data in each location.

TABLE 3: Posterior estimates and 95% credible interval for the parameters of the AR, GP, and GPP models fitted for PM_{10} concentration levels. SD stands for standard deviation.

Models	Parameters	Mean	Median	SD	95% credible interval
AR	β_0	1.3521	1.3519	0.0206	(1.3115, 1.3934)
	Temp	0.0001	0.0001	0.0004	(-0.0007, 0.0009)
	RH	0.0000	0.0000	0.0002	(-0.0004, 0.0003)
	ρ	0.0160	0.0145	0.0267	(-0.0328, 0.0734)
	σ_ε^2	0.0209	0.0208	0.0007	(0.0199, 0.0226)
	σ_η^2	0.0124	0.0123	0.0009	(0.0107, 0.0144)
	ϕ	0.0168	0.0168	0.0000	(0.0168, 0.0168)
GP	β_0	12.1303	12.4378	1.1060	(9.7744, 13.5797)
	Temp	-0.0173	-0.0267	0.0334	(-0.0590, 0.0538)
	RH	-0.0153	-0.0163	0.0041	(-0.0209, -0.0062)
	σ_ε^2	0.0052	0.0052	0.0001	(0.0050, 0.0054)
	σ_η^2	15.2262	12.8393	7.9727	(6.6616, 35.1414)
	ϕ	0.0032	0.0027	0.0018	(0.0011, 0.0069)
GPP	β_0	0.4792	0.4784	0.0692	(0.3425, 0.6131)
	Temp	-0.0028	-0.0028	0.0013	(-0.0053, -0.0003)
	RH	-0.0013	-0.0013	0.0003	(-0.0018, -0.0007)
	ρ	0.9704	0.9704	0.0025	(0.9654, 0.9754)
	σ_ε^2	0.0052	0.0052	0.0001	(0.0050, 0.0054)
	σ_η^2	1.7450	1.7453	0.0256	(1.6955, 1.7954)
	ϕ	0.0010	0.0010	0.0000	(0.0010, 0.0010)

TABLE 4: Validation statistics for the GP, AR, and GAM models.

Models		MSE	MAE	MAPE	rBIAS	rMSEP
Non-Bayesian	GAM	991.04	79.95	122.43	0.557	0.773
	AR	360.15	36.37	46.53	0.337	0.342
Spatio-temporal	GP	332.82	36.61	48.379	0.339	0.316
	GPP	202.04	16.61	28.351	0.103	0.096

Finally, we merge the simulated data with the previous one and obtained the data that we use in this article.

The final data has 1037 observations (see Figure 2) which has three variables (PM_{10}, temperature, and relative humidity) measured over irregular time over the years 2016-2020 in nine spatial locations.

The top panel in Figure 3 shows the distribution of the daily average PM_{10} concentrations over the years 2016-2020, whereas the bottom one in the same figure shows these averages by the nine sites. Clearly that the distribution of the PM_{10} concentrations, in all locations, are right skewed with very high extreme values. Thus, to stabilize the variance and reduce the departure from normality, we transform the original scale to square root scale.

3.2. Parameter Estimates, Model Validations, Prediction, and Comparison Results. The main objective of this research is to develop a hierarchical Bayesian model that can be used to

select and validate the best model which fits the daily average of the PM_{10} air pollutant levels over different locations in Qatar. We consider the two covariates, temperature and relative humidity, for spatial interpolation of the PM_{10} concentration at a new location and any time.

First, we use the MCMC algorithm with Gibbs sampler, based on 5,000 iterations discarding the first 1,000 values, to impute the missing data $z^\star$ as explained in Appendix, and then, we utilize this data to predict the values of $Z(s', t)$ at new location $s' \notin \{s_1, \cdots, s_n\}$. The posterior predictive distribution is

$$p\left(z\left(s', t\right) \mid z\right) = \int p\left(z\left(s', t\right) \mid S_m^\star, \eta^\star, \theta, z\right) \\ \cdot p(S_m^\star, \eta^\star, \theta \mid z) dS_m^\star d\eta^\star d\theta, \quad (15)$$

where $\theta = (\rho, \sigma_\varepsilon^2, \sigma_\eta^2, \phi)'$ are the model parameters and $\eta^\star = (\eta_1^\star, \cdots, \eta_n^\star)'$ (see [17]). We fit the Bayesian GP, AR, and GPP models described in Section 2 using the **spTimer** package [21, 23] and compare these models with the non-Bayesian generalized additive model (GAM) using the **R** package **mgcv** [24–26]. We use the crossvalidation method to evaluate the predictive performance of these models. Here, data at locations $(s_1, \cdots, s_m)$ where $m < n$ are used to fit the model while data at other sites $(s_{m+1}, \cdots, s_n)$ are used
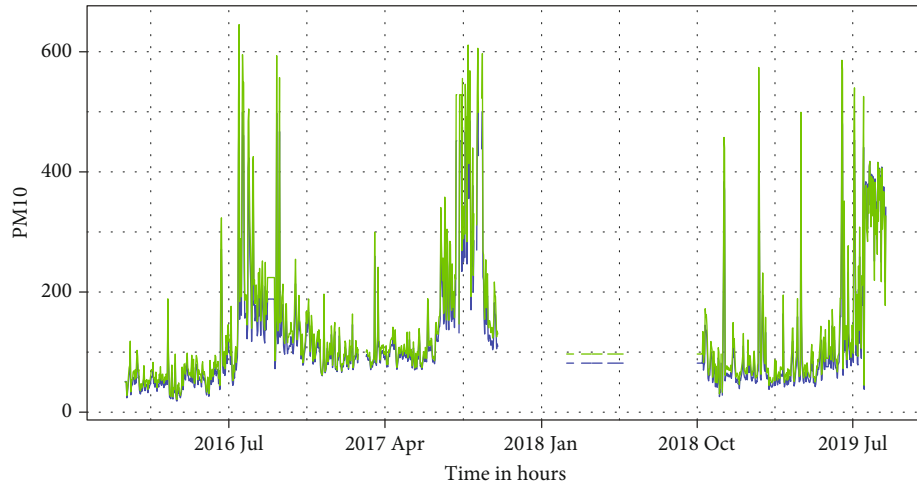
FIGURE 4: Time series plots of the true (blue) and fitted values (green) at one randomly chosen location (Qatar University).
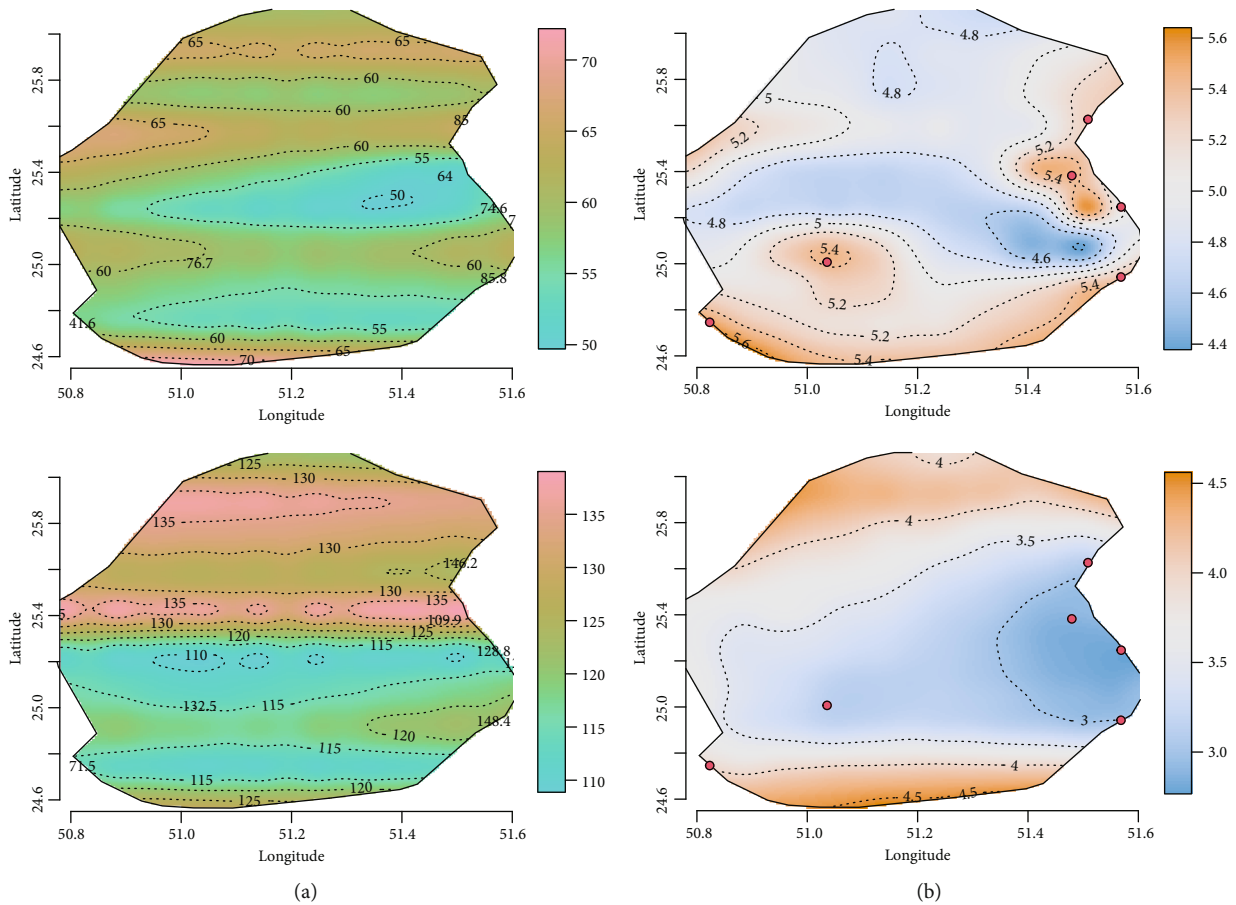

(a)

(b)

FIGURE 5: Spatially interpolated plots of the daily average of PM_{10} concentration levels (a) and their standard deviations (b) obtained from the AR model for the 29th of March 2018 (top graphs) and 15 May of 2019 (bottom graphs). Actual values and associated locations are superimposed in plots, respectively.

to assess the model using the mean squared error (MSE), mean absolute error (MAE), mean absolute prediction error (MAPE), relative bias (rBIAS), and relative mean separation (rMSEP) defined, respectively, as follows:

$$
\begin{aligned}
\text{MSE} &= \frac{1}{N}\sum_{i=1}^{N}(z_i - \widehat{z}_i)^2, \\
\text{MAE} &= \frac{1}{N}\sum_{i=1}^{N}|z_i - \widehat{z}_i|, \\
\text{MAPE} &= \frac{1}{N}\sum_{i=1}^{N}\left|\frac{(z_i - \widehat{z}_i)}{z_i}\right|, \\
\text{rBIAS} &= \frac{1}{N}\sum_{i=1}^{N}\frac{(z_i - \widehat{z}_i)}{\bar{z}}, \\
\text{rMSEP} &= \frac{1}{N}\sum_{i=1}^{N}\frac{(z_i - \widehat{z}_i)^2}{(\bar{\bar{z}} - z_i)},
\end{aligned}
\tag{16}
$$

where N is the total number of nonmissing observations, $\widehat{z}_i$ is the posterior predictive value of z_i, and $\bar{z}, \bar{\bar{z}} - z_i$ are the arithmetic means of $z_i, \widehat{z}_i$ for $i = 1, \cdots, N$.

Specifically, we use the data from Abu-Samra and Doha airport stations with $2 \times 1037 = 2074$ observations for validation purposes. The data from the remaining six stations with $7 \times 1037 = 7259$ total number of observations are used for model fitting (see Figure 1). The parameter estimates (posterior) for the Bayesian spatio-temporal models are given in Table 3. The 95% credible intervals for the parameters suggest that most of regression parameters for the AR and GP models are statistically significant, whereas the GPP model suggest that the variables are significant. In all models, the nugget effect σ_ε^2 is small. On the other hand, the spatial decay parameter $\phi = 0.0033$ for GP model and $\phi = 0.001$ for GPP model suggesting that the effective ranges will approximately be 909 and 3000 kilometers, respectively. These two values are unusual associated with very large spatial variance values (σ_η^2) suggesting that the PM_{10} concentrations in Qatar are not significantly different over the space.

After fitting the three models, we perform the crossvalidation and select the best one. Table 4 summarized the validation statistics for the GAM, GP, and AR models. Clearly, the MSE, MAE, MAPE, rBIAS, and rMSEP are smaller for the Bayesian spatio-temporal methods providing better predictive performance compared to the non-Bayesian additive model. For example, the Bayesian AP, GP, and GPP models reduced the MSE by about 63.7%, 66.4%, and 79.6%, respectively, compared with the non-Bayesian GAM model. We conclude from this table that the Bayesian spatio-temporal GPP model is the best model that can be used to predict the PM_{10} concentrations.

Figure 4 shows the time series plot of the true and fitted values for Qatar University location using the GPP model. We clearly see that the fitted values are very close to the true values. To further demonstrate the usefulness of the spatio-temporal GPP model for predicting the linear trend surfaces of the PM_{10} values with their standard values, we illustrate a prediction map for the 29th of March 2018 and 15 May of 2019 over a one-kilometer square grid (see Figure 5). Obviously, the graphs show that the GPP model correctly represented the PM_{10} concentration level over the space and time.

4. Conclusion

The potential of applying nonstationary hierarchical Bayesian spatio-temporal models in PM_{10} prediction with a large number of missing values is presented in this paper. The predicting model is developed by comparing the Gaussian predictive processes (GPP) with Gaussian processes (GP), autoregressive (AR) Bayesian models, and non-Bayesian generalized additive model (GAM) models using the datasets from the state of Qatar. The numerical results show that the GPP model outperforms other alternative models providing forecasting with good accuracy and interpretability. We applied the disaggregating technique and simulated a daily spatio-temporal PM_{10} data using the available and collected data. Then, we used the Markov chain Monte Carlo with Gibbs sampler to impute the missing data in real collected data. We believe that our statistical data analysis approach will give similar results for future available real data. In many applications, the support vector machine (SVM) algorithm has shown a superior forecasting performance compared with several evolutionary algorithms. This could be an interesting possible extension to this article where further research can be done by comparing the performance of the SVM algorithms with Bayesian models in predicting the daily average PM_{10} concentration levels.

Appendix

Full Conditional Distributions for Gibbs Sampling

In order to fit the model, we use Gibbs sampling [27] by repeatedly simulating the missing observations and parameter models from the full conditional distribution:

(1) **Sampling of missing values** $z^\star$ from its conditional distribution $\mathcal{N}(\mathbf{x}_{\ell t}'\beta + \tilde{\eta}_{\ell t}, \sigma_\varepsilon^2)$, $\ell = 1, 2, \cdots, r$, $t = 1, 2, \cdots, T_\ell$ at each iteration

(2) **Sampling** $\sigma_\varepsilon^2, \sigma_\eta^2$ and σ_ℓ^2 from the conditional Inverse Gamma distribution so that

$$
\begin{aligned}
\frac{1}{\sigma_\varepsilon^2} &\sim \Gamma\left(\frac{N}{2} + a, b + \frac{1}{2}\sum_{\ell=1}^{r}\sum_{t=1}^{T_\ell}(\mathbf{Z}_{\ell t} - \mathbf{Y}_{\ell t})'(\mathbf{Z}_{\ell t} - \mathbf{Y}_{\ell t})\right), \\
\frac{1}{\sigma_\eta^2} &\sim \Gamma\left(\frac{mN}{2} + a, b + \frac{1}{2}\sum_{\ell=1}^{r}\sum_{t=1}^{T_\ell}(\eta_{\ell t}^\star - \rho\eta_{\ell t-1}^\star)'\Sigma_{\eta^\star}^{-1}(\eta_{\ell t}^\star - \rho\eta_{\ell t-1}^\star)\right), \\
\frac{1}{\sigma_\ell^2} &\sim \Gamma\left(\frac{m}{2} + a, b + \frac{\eta_{\ell 0}^{\star\prime}\mathcal{H}_0^{-1}\eta_{\ell 0}^\star}{2}\right), \quad \ell = 1, \cdots, r.
\end{aligned}
\tag{A.1}
$$

(3) **Sampling of** $0 < \rho < 1$ from the full conditional distribution $\rho \sim \mathcal{N}(\Lambda\chi, \Lambda)$ where

$$\Lambda^{-1} = \sum_{\ell=1}^{r}\sum_{t=1}^{T_\ell} \eta^\star_{t-1} \Sigma^{-1}_{\eta^\star} \eta^\star_{\ell t-1} + 10^{-4}\mathbf{I}, \ \chi = \sum_{\ell=1}^{r}\sum_{t=1}^{T_\ell} \eta^\star_{t-1} \Sigma^{-1}_{\eta^\star} \eta^\star_{\ell t}, \quad (A.2)$$

where the sample values is restricted in the interval $(0, 1)$.

(4) **Sampling of** β from the full conditional distribution $\beta \sim \mathcal{N}(\Lambda\chi, \Lambda)$ where

$$\Lambda^{-1} = \frac{1}{\sigma_\varepsilon^2}\sum_{\ell=1}^{r}\sum_{i=1}^{T}\mathbf{X}_{\ell t}'\mathbf{X}_{\ell t} + 10^{-4}\mathbf{I}, \ \chi = \frac{1}{\sigma_\varepsilon^2}\sum_{\ell=1}^{r}\sum_{i=1}^{T}\mathbf{X}_{\ell t}'(\mathbf{Z}_{\ell t} - \tilde{\eta}_{\ell t}). \quad (A.3)$$

(5) **Sampling of** $\eta^\star_{\ell t}$ from the full conditional distribution $\eta^\star_{\ell t} \sim \mathcal{N}(\Lambda_{\ell t}\chi_{\ell t}, \Lambda_{\ell t})$, where

For $1 \le t \le T_\ell - 1$,

$$\Lambda^{-1}_{\ell t} = \frac{1}{\sigma_\varepsilon^2}A'A + \left(1 + \rho^2\right)\Sigma^{-1}_{\eta^\star},$$

$$\chi_{\ell t} = \frac{1}{\sigma_\varepsilon^2}A'(\mathbf{Z}_{\ell t} - \mathbf{X}_{\ell t}\beta) + \rho\Sigma^{-1}_{\eta^\star}\{\eta^\star_{\ell t-1} + \eta^\star_{\ell t+1}\}. \quad (A.4)$$

For $t = T_\ell$,

$$\Lambda^{-1}_{\ell t} = \frac{1}{\sigma_\varepsilon^2}A'A + \Sigma^{-1}_{\eta^\star}, \text{ and } \chi_{\ell t} = \frac{1}{\sigma_\varepsilon^2}A'(\mathbf{Z}_{\ell t} - \mathbf{X}_{\ell t}\beta) + \rho\Sigma^{-1}_{\eta^\star}\eta^\star_{\ell t-1}, \quad (A.5)$$

where $A = C^\star \mathcal{H}^{-1}_{\eta^\star}$.

(6) **Sampling of** $\eta^\star_{\ell 0}$ from the full conditional distribution $\eta^\star_{\ell 0} \sim \mathcal{N}(\Lambda_\ell\chi_\ell, \Lambda_\ell)$, where

$$\Lambda^{-1}_\ell = \rho^2\Sigma^{-1}_{\eta^\star} + \Sigma^{-1}_0, \text{ and } \chi_\ell = \rho\Sigma^{-1}_{\eta^\star}\eta^\star_{\ell 1} + \mu_\ell\Sigma^{-1}_0\mathbf{1}_m, \quad (A.6)$$

where $\mathbf{1}_m$ is a unit vector of dimension m.

(7) **Sampling of** ϕ. The full conditional distribution of ϕ is calculated based on Metropolis-Hastings algorithm from the prior and likelihood terms involving ϕ which is given by

$$\log\left\{p\left(\phi \mid \phi_{(-1)}\right)\right\}$$
$$= -\frac{1}{2}\left|\mathcal{H}_{\eta^\star}\right| - \frac{1}{2\sigma_\eta^2}\sum_{\ell=1}^{r}\sum_{t=1}^{T_\ell}(\mathbf{Y}_{\ell t} - \vartheta_{\ell t})' \quad (A.7)$$
$$\cdot \mathcal{H}^{-1}_{\eta^\star}(\mathbf{Y}_{\ell t} - \vartheta_{\ell t}) + \log(p(\phi)),$$

where $\vartheta_{\ell t} = \mathbf{X}_{\ell t}\beta + \tilde{\eta}_{\ell t}$ and $\phi_{(-1)}$ denotes all the data and parameters except for ϕ. We adopt the

Metropolis-Hastings algorithm to obtain the sample from (A.7) by considering the proposal distribution $q(\phi^{(p)})|\phi^{(c)})$ as the normal distribution on the log-scale for ϕ with the mean at the current value $\phi^{(c)}$ and the variance σ_p^2 with an acceptance rate between 15% and 40% as suggested by Gelman et al. [28]. The sample value $\phi^{(p)}$ is accepted with probability

$$\alpha\left(\phi^{(p)}, \phi^{(c)}\right) = \min\left\{1, \frac{p\left(\phi^{(p)} \mid \phi_{(-1)}\right)q\left(\phi^{(c)}\right)\mid\phi^{(p)}\right)}{p\left(\phi^{(c)} \mid \phi_{(-1)}\right)q\left(\phi^{(p)}\right)\mid\phi^{(c)}\right)}\right\}. \quad (A.8)$$

Data Availability

The data used in the article is obtained by the Environmental Science Center at Qatar University and can be made available upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] D. Cocchi, F. Greco, and C. Trivisano, "Hierarchical space-time modelling of PM$_{10}$ pollution," *Atmospheric Environment*, vol. 41, no. 3, pp. 532–542, 2007.

[2] A. Pollice and G. Jona Lasinio, "Spatiotemporal analysis of the PM10 concentration over the Taranto area," *Environmental Monitoring and Assessment*, vol. 162, no. 1-4, pp. 177–190, 2010.

[3] J. Zidek, L. Sun, N. Le, and H. Ozkaynak, "Contending with space-time interaction in the spatial prediction of pollution: Vancouver's hourly ambient PM10 field," *Environmetrics*, vol. 13, no. 5-6, pp. 595–613, 2002.

[4] World Health Organization, "Ambient air pollution: a global assessment of exposure and burden of disease," Technical report, Genf, Schweiz, Geneva, Switzerland, 2016.

[5] A. A. Ahmadi, P. Balakrishnan, K. Kakosimos, and I. Goktepe, "Determination of the levels of particulate matter 25 and 10 and their elemental composition in Qatar," in *Qatar Foundation Annual Research Conference Proceedings, Qatar Foundation Annual Research Conference Proceedings Volume 2018 Issue 1*, pp. 41–52, Qatar National Library, Doha, Qatar, 2018.

[6] C. Taylor, A. E. Yousi, and K. Mwitondi, "Statistical analysis of particulate matter data in Doha, Qatar," *WIT Transactions on Ecology and the Environment*, vol. 230, pp. 107–118, 2018.

[7] N. Cressie, "An approach to statistical spatial-temporal modeling of meteorological fields: comment," *Journal of the American Statistical Association*, vol. 89, 1994.

[8] C. Goodall and K. V. Mardia, "Challenges in multivariate spatio-temporal modeling," in *Proceedings of the XVIIth international biometric conference*, pp. 1–17, 1994.

[9] N. Cressie, M. S. Kaiser, M. J. Daniels, J. A. Lee, S. N. Lahiri, and L. H. Cox, "Spatial analysis of particulate matter in an urban environment," in *In geoENV II: Geostatistics for Environmental Applications*, pp. 41–52, Springer, 1999.

[10] L. Sun, J. V. Zidek, N. D. Le, and H. Ozkaynak, "Interpolating Vancouver's daily ambient PM10 field," *Environmetrics*, vol. 11, no. 6, pp. 651–663, 2000.

[11] B. M. G. Kibria, L. Sun, J. V. Zidek, and N. D. Le, "Bayesian spatial prediction of random spacetime fields with application to mapping $PM_{2.5}$ exposure," *Journal of the American Statistical Association*, vol. 97, no. 457, pp. 112–124, 2002.

[12] G. Shaddick and J. Wakefield, "Modelling daily multivariate pollutant data at multiple sites," *Journal of the Royal Statistical Society, Series C*, vol. 51, no. 3, pp. 351–372, 2002.

[13] S. K. Sahu and K. V. Mardia, "A Bayesian kriged Kalman model for short-term forecasting of air pollution levels," *Journal of the Royal Statistical Society, Series C*, vol. 54, no. 1, pp. 223–244, 2005.

[14] R. L. Smith, S. Kolenikov, and L. H. Cox, "Spatio-temporal modelling of $PM_{2.5}$ data with missing values," *Journal of Geophysical Research-Atmospheres*, vol. 108, 2003.

[15] S. K. Sahu, A. E. Gelfand, and D. M. Holland, "Spatio-temporal modeling of fine particulate matter," *Journal of Agricultural, Biological and Environmental Statistics*, vol. 11, no. 1, pp. 61–86, 2006.

[16] C. K. Wikle, A. Zammit-Mangion, and N. Cressie, *Spatio-Temporal Statistics with R*, Chapman and Hall/CRC, Boca Raton, FL, USA, 1st edition, 2019.

[17] S. Mukhopadhyay and S. K. Sahu, "A Bayesian spatiotemporal model to estimate long-term exposure to outdoor air pollution at coarser administrative geographies in England and Wales," *Journal of the Royal Statistical Society, Series A*, vol. 181, no. 2, pp. 465–486, 2018.

[18] S. K. Sahu, A. E. Gelfand, and D. M. Holland, "High resolution space-time ozone modeling for assessing trends," *Journal of the American Statistical Association*, vol. 102, no. 480, pp. 1221–1234, 2007.

[19] B. Gräler, E. Pebesma, and G. Heuvelink, "Spatio-temporal interpolation using gstat," *The R Journal*, vol. 8, no. 1, pp. 204–218, 2016.

[20] B. Matérn, *Spatial Variation*, Springer-Verlag, Berlin, 2nd edition edition, 1960.

[21] K. S. Bakar and S. K. Sahu, "spTimer: spatio-temporal Bayesian modeling using R," *Journal of Statistical Software*, vol. 63, no. 15, 2015.

[22] S. K. Sahu and S. Mukhopadhyay, "On generating a flexible class of anisotropic spatial models using gaussian predictive processes," Technical Report, University of Southampton, Southampton, United Kingdom, 2015.

[23] K. S. Bakar and S. K. Sahu, "Spatio-temporal Bayesian modeling," *R package version*, vol. 3, no. 3, 2020.

[24] J. Friedman, T. Hastie, and R. Tibshirani, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, New York, NY, USA, 3 edition edition, 2001.

[25] S. Wood, *Generalized Additive Models: An Introduction with R*, Chapman and Hall/CRC, London, United Kingdom, 2 edition edition, 2017.

[26] S. Wood, "MGCV: mixed GAM computation vehicle with automatic smoothness estimation," *R package version*, vol. 1, pp. 8–33, 2020.

[27] A. E. Gelfand and A. F. M. Smith, "Sampling-based approaches to calculating marginal densities," *Journal of the American Statistical Association*, vol. 85, no. 410, pp. 398–409, 1990.

[28] A. Gelman, G. O. Roberts, and W. R. Gilks, "Efficient metropolis jumping rules," in *In Bayesian Statistics 5*, J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, Eds., pp. 599–607, Oxford University Press, 1996.