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

Esam Mahdi, Sana Alshamari, Maryam Khashabi, and Alya Alkorbi

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

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

Received 30 May 2021; Revised 29 July 2021; Accepted 23 August 2021; Published 13 September 2021

Academic Editor: Wei-Chiang Hong

Copyright © 2021 Esam Mahdi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The publication of this article was funded by Qatar National Library.

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 ($\mu$g/m$^3$) for PM$_{10}$, and 10 and 25 $\mu$g/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 24hr PM$_{10}$ concentration levels were ranged from 126.69 $\mu$g/m$^3$ to 184.55 $\mu$g/m$^3$, which exceeds the recommended maximum of 50 $\mu$g/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 geostatistical approach in prediction the PM$_{10}$ concentrations around the city of
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 $\ell$ 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 $Z_{\ell}(s,t)$ denote the observed value of the PM$_{10}$ concentration, after any necessary transformation, at a given location $s$ and over a given discrete time $t$. We assume that the spatial location $s$ 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 $Z_{\ell}(s,t)$ is observed at $n$ monitoring sites denoted by $s_i, i = 1, \cdots, n$ and at time points denoted by two indices $\ell$ 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 $\mathbf{z}^\prime$, whereas all the observed data will be denoted by $\mathbf{z}$.

The first stage of the hierarchy assumes that the observed values $Z_{\ell t}$, where $Z_{\ell t} = (Z_{\ell}(s_1, t), \cdots, Z_{\ell}(s_n, t))^T$, can be decomposed into a true (latent) spatio-temporal process $Y_{\ell t} = (Y_{\ell}(s_1, t), \cdots, Y_{\ell}(s_n, t))^T$ with an error term $\varepsilon_{\ell t} = (\epsilon_{\ell 1}(s_1, t), \cdots, \epsilon_{\ell n}(s_n, t))^T$. More specifically, the data (or measurement error) model in the first stage of the hierarchy is

$$Z_{\ell t} = Y_{\ell t} + \varepsilon_{\ell t}, \ell = 1, \cdots, r, t = 1, \cdots, T_\ell.$$  \hspace{1cm} (1)

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

The second stage assumes that the true process $Y_{\ell t}$ 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, $Y_{\ell t}$ can be specified in the following formula:

$$Y_{\ell t} = \mu_{\ell t} + \eta_{\ell t},$$ \hspace{1cm} (2)

where $\eta_{\ell t} = (\eta_{\ell 1}(s_1, t), \cdots, \eta_{\ell n}(s_n, t))^T$ is a spatio-temporal residual random intercept assumed to follow $N(\mathbf{0}, C)$, where $C = \sigma_\eta^2 \mathcal{H}_n$, $\sigma_\eta^2$ is the site invariant spatial variance, and $\mathcal{H}_n$ is the spatial correlation matrix. In this article, we consider...
four modelings for \( \{ \eta_{t} \} \). The first one assumes that the random effects, \( \eta_{s}(s, t) = 0 \), for all locations \( s \) and times \( t \). This implies that the model in (1)–(2) will be the simple regression model. The other models for \( \{ \eta_{t} \} \) 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, \( C_{s}(u) \), by the pure temporal one, \( C_{t}(u) \), given by

\[
C(h, u) = C_{s}(h)C_{t}(u),
\]

where \( h = \|s - s'\| \) is the separating spatial distance, and \( u = |t - t'| \) is the temporal distance for any pair of points \((s, t) \times (s', t')\) 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:

\[
C(w(s, t), w(s', t)) = \frac{\sigma^2_{\eta}}{2^{\nu - 1}\Gamma(\nu)} \left(2\sqrt{\nu}\|s - s'\| \phi\right)^{\nu} K_{\nu}(2\sqrt{\nu}\|s - s'\| \phi), \quad \phi > 0, \kappa \geq 1,
\]

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

2.2. Gaussian Process Model Specification. Consider \( \mu_{t} = X_{t}\beta \) in (2), where \( \ell = 1, \cdots, r, t = 1, \cdots, T_{t} \), \( X_{t} \) is the \( n \times (k + 1) \) design matrix of spatially and temporally varying \( k - \)covariates and \( \beta = (\beta_{1}, \beta_{2}, \cdots, \beta_{k}) \) is the \( k + 1 \)-dimensional vector of regression coefficients. Thus, the Gaussian process (GP) two-stage model can be written as

\[
\begin{align*}
Z_{t} &= Y_{t} + \varepsilon_{t}, \\
Y_{t} &= X_{t}\beta + \eta_{t}.
\end{align*}
\]

We assume that the random effect process \( \{ \eta_{t}(\cdot) \} \) is independent from the white noise process \( \{ \varepsilon_{t}(\cdot) \} \). 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 \( N'(Y_{t}, \sigma^2_{t}) \) distribution defined by (5) at each MCMC iteration using Gibbs sampling. The Gibbs sampler requires that the full conditional distributions of the parameters \( \theta = (\beta, \sigma^2_{\eta}, \sigma^2_{\varepsilon}, \sigma^2_{\f}, \phi, \kappa) \) 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:

\[
\log p(\theta, Y, z' | z) = -\frac{N}{2} \log \sigma^2_{\varepsilon} - \frac{1}{2\sigma^2_{\eta}} \sum_{t=1}^{T_{t}} (Z_{t} - Y_{t})^\prime (Z_{t} - Y_{t}) - \frac{1}{2} \sum_{t=1}^{T_{t}} T_{t} \log \sigma^2_{\f} - \frac{1}{2\sigma^2_{\f}} \sum_{t=1}^{T_{t}} \sum_{\ell=1}^{r} (Y_{t\ell} - X_{t\ell}\beta)^\prime \f^{-1}(Y_{t\ell} - X_{t\ell}\beta) + \log p(\theta),
\]

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 \( \theta \).

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:

\[
\begin{align*}
Z_{t} &= Y_{t} + \varepsilon_{t}, \\
Y_{t} &= \rho Y_{t-1} + X_{t}\beta + \eta_{t},
\end{align*}
\]

where \( -1 < \rho < 1 \) is parameter of the first-order autoregressive model and \( \mu_{t} = \rho Y_{t-1} + X_{t}\beta \). 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 \( Y_{0} \) for each \( \ell = 1, \cdots, r \), with mean \( \mu_{t} \) and covariance \( \sigma^2_{\f} \f^{-1} \) 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:

\[
\log p(\theta, Y, z' | z) = -\frac{N}{2} \log \sigma^2_{\varepsilon} - \frac{1}{2\sigma^2_{\eta}} \sum_{t=1}^{T_{t}} (Z_{t} - Y_{t})^\prime (Z_{t} - Y_{t}) - \frac{1}{2} \sum_{t=1}^{T_{t}} T_{t} \log \sigma^2_{\f} - \frac{1}{2\sigma^2_{\f}} \sum_{t=1}^{T_{t}} \sum_{\ell=1}^{r} (Y_{t\ell} - X_{t\ell}\beta)^\prime \f^{-1}(Y_{t\ell} - X_{t\ell}\beta) - \frac{1}{2} \sum_{t=1}^{T_{t}} \log \sigma^2_{\f} - \frac{1}{2} \sum_{t=1}^{T_{t}} \sigma^2_{\f} (Y_{0\ell} - \mu_{t})^\prime \f^{-1}(Y_{0\ell} - \mu_{t}) + \log p(\theta),
\]
where $\rho(\theta)$ is the prior distribution for the parameter $\theta = (\beta, \rho, \sigma^2, \sigma^2_\eta, \phi, \mu_t, \sigma^2_t)$ (21), see Appendix B).

2.4. Gaussian Predictive Processes with Autoregressive Model Specification. We now introduce the case that the process $\{\eta_t\}$ in (2) has a nonstationary covariance structure. We 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), $S_m = (s_1^*, \ldots, s_m^*)$. Given $S_m$, we assume that $\eta^*_t = (\eta^*_t(s_1^*, t), \ldots, \eta^*_t(s_m^*, t))'$ is a Gaussian process with zero mean vector and covariance function given in (4). Thus, the nonstationary model is defined after replacing $\eta_t(s, t)$ in (2) by

$$\hat{\eta}(s, t) = E[\eta(s, t) | \eta^*_t], \quad i = 1, \ldots, n,$$

hence, $(\eta_t^*, \eta^*_t)$ is a vector of $n + m$-independent realization at each time point $t$ and $t$ with the same Gaussian process in (4). Define $\eta^*_t = (\eta^*_t(s_1^*, t), \ldots, \eta^*_t(s_m^*, t))'$, we have

$$\eta^*_t = C^*(\phi, \kappa) \mathcal{N}(\phi, \kappa) \eta^*_t, \quad (10)$$

**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 | 384 | 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 |
| 2018 | 480 | 52.56  | 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 |

| Year | $n$ | % n.a. | Min | Max | Median | Mean | SD |
|------|-----|--------|-----|-----|--------|------|----|
| 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 |
| 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  | 15.13 | 92.77 | 56.34  | 53.67  | 18.31 |

| Year | $n$ | % n.a. | Min | Max | Median | Mean | SD |
|------|-----|--------|-----|-----|--------|------|----|
| 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 |
| 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 $C^* (\phi, \kappa)$ is the $n \times m$ crosscorrelation matrix between $\eta_i$ and $\eta_j$ having elements $[C^*]_{ij} = C(||s_i - s_j||)$, $i = 1, \ldots, n$ and $j = 1, \ldots, m$, and $\mathcal{H}_m (\phi, \kappa)$ is the $m \times m$ correlation matrix of $\eta_i$ so that $[\mathcal{H}_m]_{ij} = [C(||s_i - s_j|| ; \phi, \kappa)]$, for $k = 1, \ldots, m$. Clearly, the process $\{\hat{\eta}_t\}$ shows nonstationary structure with variance function that is given by

$$\text{Var}(\hat{\eta}_t) = C^* (\phi, \kappa) \mathcal{H}_m^{-1} (\phi, \kappa) C^* (\phi, \kappa).$$  

(11)

The advantage of using the nonstationary model in (10) is the flexibility in the $\hat{\eta}_t$ surface which is based on $m \ll n$ linear functions of $\eta^t$. 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:

$$Z_{t, t} = Y_{t, t} + e_{t, t},$$

$$Y_{t, t} = X_{t, t} \beta + \eta_{t, t},$$  

(12)

where $\hat{\eta}_{t, t}$ is given in (10). The process $\{\eta^*_{t, t}\}$, at the $S_m^*$ knots, can be modeled according to the autoregressive model

$$\eta^*_{t, t} = \rho \eta^*_{t, t-1} + \eta_{t, t},$$  

(13)

where $\eta^*_{t, t} \sim \mathcal{N}_m (0, \Sigma^*)$ for $\ell = 1, \ldots, r, t = 1, \ldots, T$, and $\Sigma^* = \sigma^2 \mathcal{H}_m$. We assume that the initial conditions $\eta^*_{t, 0}$ has normal distribution with mean zero and covariance matrix $\Sigma_0 = \sigma_0^2 \mathcal{H}_0$, and both $\Sigma^*, \Sigma_0$ can be obtained from the Matérn covariance function defined in (4), where $\Sigma^*$ is an $m \times m$ matrix much lower dimensional than $\Sigma_0$ 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:

$$\log p(\theta, Y, z^* | z) = - \frac{N}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{t=1}^r \sum_{i=1}^{T_t} (Z_{t, t} - Y_{t, t})' (Z_{t, t} - Y_{t, t}) - \frac{mN}{2} \log \left( \sigma_0^2 \right) - N \log \left( |\mathcal{H}_m| \right) - \frac{m}{2} \sum_{t=1}^r \log \left( \sigma^2 \right) - \frac{r}{2} \log \left( |\mathcal{H}_0| \right) - \frac{1}{2\sigma^2} \sum_{t=1}^r \sum_{i=1}^{T_t} (\eta^*_{t, t} - \rho \eta^*_{t, t-1})' \mathcal{H}_m^{-1} (\eta^*_{t, t} - \rho \eta^*_{t, t-1}) - \frac{1}{2} \sum_{t=1}^r \frac{1}{\sigma_t^2} \eta^*_{t, t} \mathcal{H}_0^{-1} \eta^*_{t, 0} + \log p(\theta),$$

(14)

where $p(\theta)$ is the prior distribution for the parameter $\theta = (\beta, \rho, \sigma^2, \sigma_0^2, \phi, \kappa, \sigma_t^2)$ and the Gibbs sampling procedure of these parameters can be obtained from the full conditional distributions required provided in the Appendix.

![Figure 2: Daily PM$_{10}$ concentrations observed in nine locations over the years 2016-2021.](image-url)

- Doha Airport
- Al Khor
- Umm Said
- Al Wakrah
- Mukenis Al Karanaah
- Abu Samra
- Al Ruwais
- Doha Airport
- Umm Said

Time in hours
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.
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 $z^*$ 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

$$p(z(s', t) | z) = \int p(z(s', t) | S_m^*, \eta^*, \theta, z) \cdot p(S_m^*, \eta^*, \theta | z) dS_m^* d\eta^* d\theta,$$

(15)

where $\theta = (\rho, \sigma^2_x, \sigma^2_y, \phi)$ are the model parameters and $\eta^* = (\eta^*_1, \cdots, \eta^*_m)$ (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

| Models | Parameters | Mean | Median | SD | 95% credible interval |
|--------|------------|------|--------|----|----------------------|
| AR     | $\beta_0$  | 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)    |
|        | $\rho$     | 0.0160 | 0.0145 | 0.0267 | (-0.0328, 0.0734)    |
|        | $\sigma^2_x$ | 0.0209 | 0.0208 | 0.0007 | (0.0199, 0.0226)     |
|        | $\sigma^2_y$ | 0.0124 | 0.0123 | 0.0009 | (0.0107, 0.0144)     |
|        | $\phi$     | 0.0168 | 0.0168 | 0.0000 | (0.0168, 0.0168)     |
| GP     | $\beta_0$  | 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)   |
|        | $\sigma^2_x$ | 0.0052 | 0.0052 | 0.0001 | (0.0050, 0.0054)     |
|        | $\sigma^2_y$ | 15.2262 | 12.8393 | 7.9727 | (6.6616, 35.1414)    |
|        | $\phi$     | 0.0032 | 0.0027 | 0.0018 | (0.0011, 0.0069)     |
| GPP    | $\beta_0$  | 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)   |
|        | $\rho$     | 0.9704 | 0.9704 | 0.0025 | (0.9654, 0.9754)     |
|        | $\sigma^2_x$ | 0.0052 | 0.0052 | 0.0001 | (0.0050, 0.0054)     |
|        | $\sigma^2_y$ | 1.7450 | 1.7453 | 0.0256 | (1.6955, 1.7954)     |
|        | $\phi$     | 0.0010 | 0.0010 | 0.0000 | (0.0010, 0.0010)     |

| 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 |
| GP | 332.82 | 36.61 | 48.37 | 0.339 | 0.316 |
| GPP | 202.04 | 16.61 | 28.35 | 0.103 | 0.096 |
Figure 4: Time series plots of the true (blue) and fitted values (green) at one randomly chosen location (Qatar University).

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 29$^{th}$ 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:

\[
\begin{align*}
\text{MSE} &= \frac{1}{N} \sum_{i=1}^{N} (z_i - \bar{z}_i)^2, \\
\text{MAE} &= \frac{1}{N} \sum_{i=1}^{N} |z_i - \bar{z}_i|, \\
\text{MAPE} &= \frac{1}{N} \sum_{i=1}^{N} \left(\frac{|z_i - \bar{z}_i|}{\bar{z}_i}\right), \\
r\text{BIAS} &= \frac{1}{N} \sum_{i=1}^{N} \left(\frac{z_i - \bar{z}_i}{\bar{z}_i}\right), \\
r\text{MSEP} &= \frac{1}{N} \sum_{i=1}^{N} \left(\frac{(z_i - \bar{z}_i)^2}{\bar{z}_i}\right),
\end{align*}
\]

where \( N \) is the total number of nonmissing observations, \( \bar{z}_i \) is the posterior predictive value of \( z_i \), and \( \bar{z}, \bar{z}_2 \) are the arithmetic means of \( z_i, \bar{z}_i \) for \( i = 1, \ldots, N \).

Specifically, we use the data from Abu-Samra and Doha airport 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 \( \sigma^2_n \) 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 (\( \sigma^2_n \)) suggesting that the PM10 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 PM10 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 PM10 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 PM10 concentration level over the space and time.

4. Conclusion

The potential of applying nonstationary hierarchical Bayesian spatio-temporal models in PM10 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 PM10 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 PM10 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** \( z \) from its conditional distribution \( N(x_\theta^0 + \eta_0t, \sigma^2_z) \), \( \ell = 1, 2, \ldots, r, t = 1, 2, \ldots, T_\ell \) at each iteration

2. **Sampling** \( \sigma^2_n, \sigma^2_\eta, \sigma^2_\ell \) from the conditional Inverse Gamma distribution so that

\[
\begin{align*}
\frac{1}{\sigma^2_n} &\sim \Gamma\left(\frac{N}{2} + a, \frac{b}{2} + \frac{1}{2} \sum_{\ell=1}^{r} T_\ell (Z_{\ell t} - Y_{\ell t})' (Z_{\ell t} - Y_{\ell t})\right), \\
\frac{1}{\sigma^2_\eta} &\sim \Gamma\left(\frac{mN}{2} + a, \frac{b}{2} + \frac{1}{2} \sum_{\ell=1}^{r} \sum_{t=1}^{T_\ell} (\eta_{\ell t} - \rho \eta^*_{\ell t})' \Sigma^{-1}_\eta (\eta_{\ell t} - \rho \eta^*_{\ell t})\right), \\
\frac{1}{\sigma^2_\ell} &\sim \Gamma\left(\frac{m}{2} + a, \frac{b}{2} + \frac{\eta^*_{\ell 0} \mathcal{M}^{-1}_\eta \eta^*_{\ell 0}}{2}\right), \quad \ell = 1, \ldots, r.
\end{align*}
\]
where the sample values is restricted in the interval (0, 1).

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

$$\Lambda^{-1} = \frac{1}{\sigma_{\epsilon}^2} \sum_{t=1}^{T} \sum_{i=1}^{r} \chi_{ti} X_{ti} + 10^{-4} \mathbf{1}, \quad \chi_t = \frac{1}{\sigma_{\epsilon}^2} \sum_{i=1}^{r} \sum_{t=1}^{T} X_{ti} (Z_t - \bar{\eta}_t).$$

(A.3)

(5) Sampling of $\eta_t^*$ from the full conditional distribution $\eta_t^* \sim \mathcal{N}(A\chi_t, \Lambda_t)$, where

For $1 \leq t \leq T_t - 1$,

$$\Lambda_t^{-1} = \frac{1}{\sigma_{\epsilon}^2} A^t \Lambda + (1 + \rho^2) \sigma_{\eta^{-1}},$$

$$\chi_t = \frac{1}{\sigma_{\epsilon}^2} A^t (Z_t - X_t \beta + \rho \sigma_{\eta^{-1}} (\eta_{t-1}^* + \eta_t^*)).$$

For $t = T_t$,

$$\Lambda_t^{-1} = \frac{1}{\sigma_{\epsilon}^2} A^t \Lambda + \sigma_{\eta^{-1}}, \quad \chi_t = \frac{1}{\sigma_{\epsilon}^2} A^t (Z_t - X_t \beta + \rho \sigma_{\eta^{-1}} \eta_{t-1}^*),$$

(A.5)

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

(6) Sampling of $\eta_0^*$. The full conditional distribution of $\eta_0^*$ is calculated based on Metropolis-Hastings algorithm from the prior and likelihood terms involving $\phi$ which is given by

$$\log \left\{ p(\phi | \phi_{(-1)}^*) \right\} = -\frac{1}{2} [\mathcal{H}_{\eta'} - \frac{1}{2\sigma_{\eta}^2} \sum_{k=1}^{T_k} (Y_k - \theta_k)']' \mathcal{H}_{\eta'}^{-1} (Y_k - \theta_k) + \log (p(\phi)).$$

(A.7)

where $\theta_k = X_k \beta + \bar{\eta}_k$ and $\phi_{(-1)}^*$ denotes all the data and parameters except for $\phi$. 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 $\phi$ with the mean at the current value $\phi^{(c)}$ and the variance $\sigma^2_p$ with an acceptance rate between 15% and 40% as suggested by Gelman et al. [28]. The sample value $\phi^{(p)}$ is accepted with probability

$$\alpha(\phi^{(p)}, \phi^{(c)}) = \min \left\{ 1, \frac{p(\phi^{(p)} | \phi_{(-1)}^*) q(\phi^{(c)} | \phi_{(-1)}^*)}{p(\phi^{(c)} | \phi_{(-1)}^*) q(\phi^{(p)} | \phi_{(-1)}^*)} \right\}.$$ 

(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

The authors are grateful to the academic editor, Hong Wei-Chiang, and three anonymous peer reviewers for their comments and suggestions. This research was supported by the Qatar National Research Fund with a grant number UREP25-010-1-003.

### 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. Ozekaynak, “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. Kakosimou, 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. Mvitondi, “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.