Computational Statistics and Data Analysis 187 (2023) 107799

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Additive dynamic models for correcting numerical model outputs

Yewen Chen[a,1], Xiaohui Chang[b,1], Fangzhi Luo[a], Hui Huang[c,*]

[a] *School of Mathematics, Sun Yat-sen University, Guangzhou, Guangdong 510275, China*
[b] *College of Business, Oregon State University, Corvallis, OR 97331, USA*
[c] *Center for Applied Statistics and School of Statistics, Renmin University of China, Beijing, 100872, China*

ARTICLE INFO

Article history:
Received 5 August 2022
Received in revised form 20 May 2023
Accepted 5 June 2023
Available online 13 June 2023

Keywords:
Numerical model output
Spatio-temporal bias correction
Additive partially linear model
Basis function approximation
Multi-resolution dynamic approach
Ensemble-based algorithm

ABSTRACT

Numerical air quality models are pivotal for the prediction and assessment of air pollution, but numerical model outputs may be systematically biased. An additive dynamic model is proposed to correct large-scale raw model outputs using data from other sources, including readings collected at ground monitoring networks and weather outputs from other numerical models. An additive partially linear model specification is employed for the nonlinear relationships between air pollutants and covariates. In addition, a multi-resolution basis function approximate is proposed to capture the different small-scale variations of biases, and a discretized stochastic integro-differential equation is constructed to characterize the dynamic evolution of the random coefficients at each spatial resolution. An expectation-maximization algorithm is developed for parameter estimation and a multi-resolution ensemble-based scheme is embedded to accelerate the computation. For statistical inference, a conditional simulation technique is applied to quantify the uncertainty of parameter estimates and bias correction results. The proposed approach is used to correct the biased raw outputs of $PM_{2.5}$ from the Community Multiscale Air Quality (CMAQ) system for China's Beijing-Tianjin-Hebei region. Our method improves the root mean squared error and continuous rank probability score by 43.70% and 34.76%, respectively. Compared to other statistical methods under different metrics, our model has advantages in both correction accuracy and computational efficiency.

1. Introduction

Numerical air quality models, such as the Nested Air Quality Prediction Modeling System (NAQPMS) (Wang et al., 2006) or the Community Multiscale Air Quality (CMAQ) model (Byun and Schere, 2006), have been extensively used in environmental science, especially air pollution forecast. Based on numerical models, the pollution concentration forecasts at fine spatial and temporal scales are generated by solving deterministic differential equations. However, numerical model outputs are usually biased. The reasons are manifold. Due to the expensive computation of the numerical models, the unknown parameters of the differential equations cannot be adjusted timely when new data become available (Xie et al., 2009). The differential equations in numerical models depend heavily on initial inputs such as meteorology variables, emission, trans-

---

[*] Corresponding author.
E-mail address: huangh89@mail.sysu.edu.cn (H. Huang).
[1] Equal contribution.

https://doi.org/10.1016/j.csda.2023.107799

portation dynamics, and ground characteristics (Byun and Schere, 2006; Bakar, 2012; Appel et al., 2017). The errors in initial inputs also introduce forecast bias. Therefore, calibration is critical for both the model parameters (Kennedy and O'Hagan, 2001; Higdon et al., 2008; Brynjarsdóttir and O'Hagan, 2014; Salter et al., 2019) and model outputs. While the more general framework for forecasting is to tune the input parameters of computer models than to calibrate numerical model outputs, the latter has gained attention in statistics; see Xie et al. (2009); Kang et al. (2010); Huang et al. (2017, 2019); Heinrich et al. (2021) and references therein. In this work, our motivation is to correct the bias of raw CMAQ outputs instead of the model parameters inputs, and hope to apply the corrected pollution map to other studies (e.g., to access the health impacts of air pollutants, see Jiang and Enki Yoo (2019)). To this end, we focus on building spatiotemporal statistical bias correction models for the raw CMAQ data along with other data sources. The proposed approach is a data-adaptive way to describe how the bias in raw data is developed, and provides a flexible and efficient framework to generate accurate pollution maps, especially for researchers outside the environmental science community who have no access to numeral modeling systems.

To improve the bias correction accuracy, a strategy commonly used is to combine the numerical model outputs with data from other sources. The most important data source in our study is reading from monitoring stations. Although monitoring stations are usually sparsely located and occasionally suffer from malfunctions, they provide direct measurements of actual pollutant concentrations and serve as a valuable reference for regional corrections.

In the literature, there are several methods to fuse model outputs with data from multiple sources. Under a point-to-block framework, Fuentes and Raftery (2005) proposed a fully model-based scheme, also known as the "Bayesian melding," where the grid-level model output is expressed as an integral of the underlying point-level process that is related to actual observations. The Bayesian melding approach provides a flexible Bayes procedure to account for the bias in numerical model outputs. However, for a large spatial domain, grid cells are too many to make the Bayesian melding computationally infeasible. To reduce the computational cost, McMillan et al. (2010) employed a latent process on the grid level rather than the point level and developed a spatio-temporal version of the Bayesian melding method. Different from the Bayesian melding approach, Berrocal et al. (2010a,b, 2012) proposed an alternative model-based strategy, known as "downscaling," where the outputs from the numerical model are downscaled from grid level to point level. To be specific, the numerical model outputs are downscaled as a site-specific explanatory variable, which is related to the site observational data via a spatially varying coefficients (SVC) regression model (Gelfand et al., 2003). The varying coefficients are assumed to be Gaussian processes via the coregionalization method (Schmidt and Gelfand, 2003), and then the correction in each grid cell can be computed by a stochastic integral over the corresponding grid cell; see Berrocal et al. (2010b) for more details.

Despite the widespread usage of the approaches mentioned earlier such as the Bayesian statistical downscaling approach (Berrocal et al., 2010b) for improving CMAQ-based outputs (Shaddick et al., 2018; Jiang and Enki Yoo, 2019; Mendy et al., 2019), it remains challenging to further improve the bias correction accuracy and computational efficiency (Vannitsem et al., 2021). Firstly, most models used in bias correction are performed in a linear regression framework (Guillas et al., 2008), which may not be able to accurately characterize the relationships between the response and other variables. For example, Liang et al. (2015) shows that the $PM_{2.5}$ (i.e., fine particulate matters with aerodynamic diameters less than 2.5μm) concentrations are highly nonlinear with respect to weather variables in China's Beijing. Secondly, classical space-time downscaling approaches such as the smoothed downscaler model (Berrocal et al., 2012) often adopt conditional autoregressive structures (e.g., AR(1)) to capture temporal correlation, while the evolution of spatio-temporal interactions is ignored. Lastly, the Bayesian methods used in earlier studies face computational difficulties for large datasets and hinder their broader applications.

In this work, we propose an additive dynamic model to correct raw numerical model outputs. An additive partially linear model is employed to capture the relationships between observations and numerical model outputs. We model the small-scale variations of the additive biases of model outputs using a multi-resolution basis function approximation. An effective discretized stochastic integro-differential equation (IDE) model defined on multi-resolution regular grid cells is used to characterize the dynamic evolution of the random coefficients. For parameter estimation, we develop an efficient algorithm by embedding a multi-resolution ensemble-based method in the expectation-maximization (EM) algorithm, which enables fast computations of the latent spatio-temporal process filtering, smoothing, and parameter estimation. Moreover, the uncertainties of parameter estimate and bias correction are quantified using a conditional simulation technique that is computationally efficient (Nychka et al., 2015). Our approach differs from earlier methods in that, it not only captures the nonlinear relationship between response and other variables, but also pictures the dynamic evolution of multi-resolution space-time interaction. In addition, the computational complexity of our method is not heavily affected by the data size, especially the number of grid cells of numerical model outputs. The proposed bias correction approach is demonstrated using the $PM_{2.5}$ concentration data in China's Beijing-Tianjin-Hebei (BTH) region along the CMAQ model outputs and the readings at the monitoring stations.

The rest of the paper proceeds as follows. We introduce the $PM_{2.5}$ datasets in Section 2, describe the proposed addictive dynamic correction model (ADCM) in Section 3, and explain the implementation of the method in Section 4. Section 5 investigates the ADCM's ability to capture nonlinear relationships and spatio-temporal patterns using a simulation study. In Section 6, using the data from the BTH region, we first compare the prediction performance of the ADCM with several widely used methods and then apply the ADCM to correct the raw CMAQ $PM_{2.5}$ outputs. We conclude with a discussion in Section 7. Technical details of the competing models can be found in the Supplementary Material. Data and code for this work are available online under https://github.com/ChenYW68/ADCM.

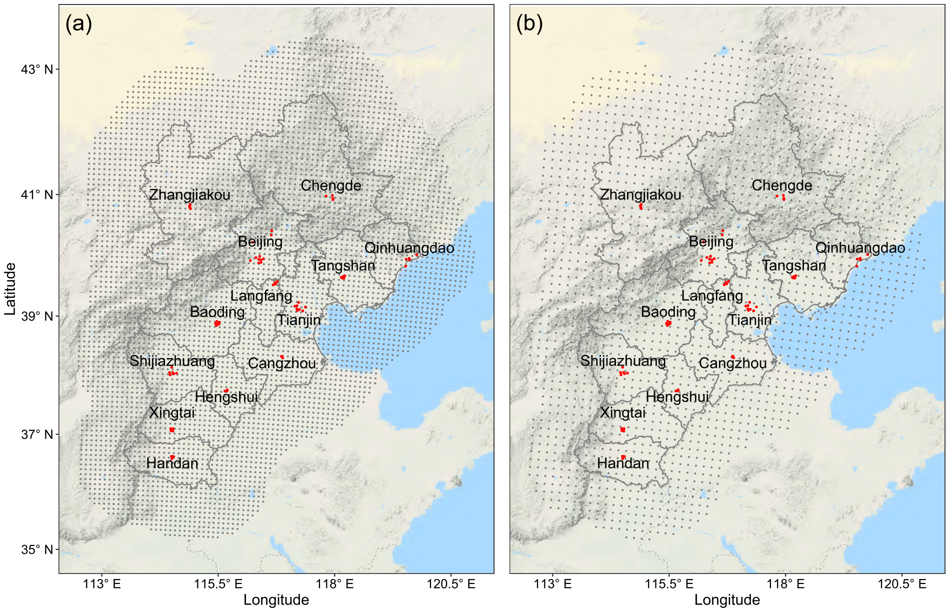

Fig. 1. Maps of the BTH region under different gridding systems with the locations of 68 monitoring stations marked by the red dots. (a) Map with the centroids of 5,587 9-km CMAQ grids (gray dots). (b) Map with the centroids of 2,141 15-km NAQPMS grids (gray dots). (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

2. Motivating applications

2.1. Data description

In the past decade, China launched a series of movements to reduce air pollution and carbon emissions, including the Air Pollution Prevention and Control Action Plan between 2013 and 2017 (China's State Council, 2013), the "Blue Sky Defense Battle" between 2018 and 2020 (China's State Council, 2018), and the long-term "Beautiful China" targets through 2035 (China's State Council, 2021). Even though the air quality in China has improved tremendously in recent years (Lu et al., 2020), there remain unresolved challenges, especially in assessing and controlling $PM_{2.5}$, in the BTH region, which is one of the most polluted areas in China with over 100 million residents.

We collect the daily data from November 1, 2015 to January 31, 2016 in the BTH region from three data sources. The first data set consists of $PM_{2.5}$ concentrations observed at 68 national monitoring stations in the region. The second data are CMAQ model outputs of the $PM_{2.5}$ concentration at a 9-km resolution from a total of 5,587 grid cells, where 2,499 grid cells are located in the BTH region and the rest are in the area adjoining the BTH region; see Fig. 1 (a) for the spatial distribution of the CMAQ grid cells. The third data are gridded meteorological variables, including temperature (°C), air pressure (Pa), and cumulative wind power (m/s), generated by the Weather Research and Forecasting (WRF) model (Skamarock et al., 2019) and available from the NAQPMS (Kong et al., 2020). The third dataset contains 2,141 gridded outputs at 15-km resolution with 963 grid cells inside the BTH region and 1,178 grid cells outside the region; see Fig. 1 (b) for the distribution of NAQPMS grid cells. The locations of 68 monitoring stations are also marked with red dots in Figs. 1 (a) and (b).

2.2. Multi-source data fusion

To connect the gridded data of the numerical models to the readings collected at the monitoring stations, we transform the numerical model outputs from grid level to point level using the inverse distance weighting method (Gotway et al., 1996). Specifically, for any location $\boldsymbol{s}$, the CMAQ $PM_{2.5}$ outputs within a circle centered at $\boldsymbol{s}$ with a radius of γ are used in the transformation to obtain a weighted CMAQ variable. In this work, we set γ to be about 50 km, corresponding to the spatial range of the $PM_{2.5}$ readings; see Fig. 2 (a). Compared to the CMAQ $PM_{2.5}$ outputs, we use a smaller circle with a radius of 15 km in the transformation procedure of the NAQPMS weather outputs because over-smoothed data can result in a less efficient estimation of nonlinear relationships between the $PM_{2.5}$ readings and other meteorological variables. In particular, the grid cells outside the region in Figs. 1 (a) and (b) are used to construct the weighted outputs of two numerical models near the boundaries of the BTH region.

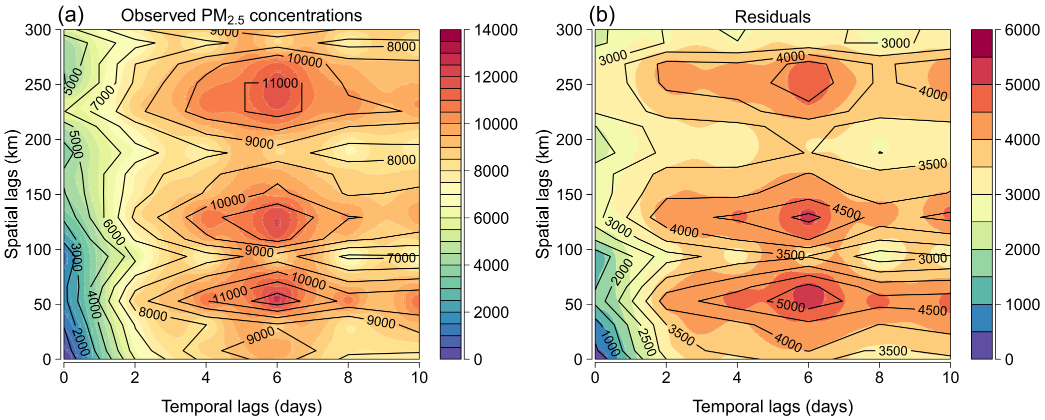

Fig. 2. Interpolated surface created using the R package MBA (Finley et al., 2017) and contour lines obtained using the R package gstat (Gräler et al., 2016) of empirical spatio-temporal variograms in the BTH region. (a) The observed $PM_{2.5}$ concentrations. (b) The residuals from the model $PM_{2.5} \sim 1 + x + \sum_{k=0}^{4} g_k(z_k)$, fitted by the R package mgcv (Wood, 2022), where the definitions of x, z_k, and $g_k(\cdot)$ can be found in Section 3.

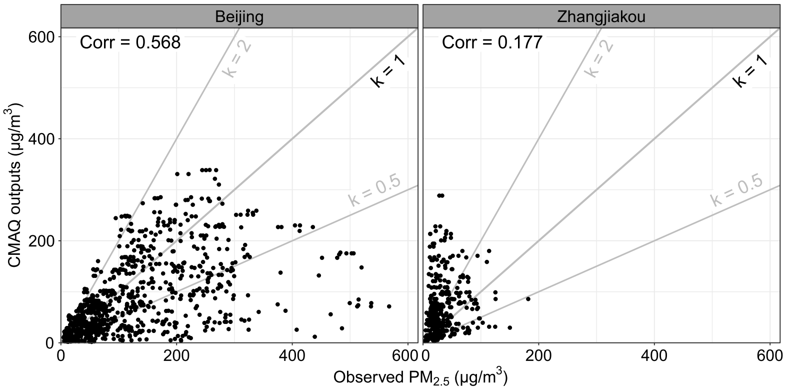

Fig. 3. Scatter plots of CMAQ $PM_{2.5}$ outputs versus observations at Beijing and Zhangjiakou, where "Corr" represents the Pearson correlation coefficient between CMAQ $PM_{2.5}$ outputs and observations. Three reference lines with slope $k = 0.5, 1,$ and 2 are colored in gray.

2.3. Data exploration

Our descriptive analyses demonstrate the necessity of correction of the CMAQ outputs in the BTH region. The scatter plots in Fig. 3 compare the CMAQ outputs with the observed $PM_{2.5}$ concentrations at two cities. It is evident that the CMAQ outputs underestimate the concentrations in Beijing but overestimate the concentrations in Zhangjiakou. Furthermore, the highest CMAQ output among all cities in the BTH region is 506.440 $\mu g/m^3$, which is much lower than the actual maximum observed $PM_{2.5}$ concentration of 735.625 $\mu g/m^3$.

The datasets involved in this work display several important data characteristics that require transformation and consideration. First of all, as the observed $PM_{2.5}$ concentrations show clear right skewness due to extreme values, we employ the square root transformation to both the observations and CMAQ outputs to stabilize the variance as seen in Sahu et al. (2006) and Sampson et al. (2013) (see Fig. S1 of the Supplementary Material), resulting in an increase in the correlation of 0.59 to 0.65. The moderately high correlation implies that a linear model is adequate to connect the $PM_{2.5}$ observations and the CMAQ outputs. Second, as seen from the empirical spatio-temporal variogram given in Fig. 2 (b), the residuals from observed data minus fixed effects contain ample spatial and temporal information. In particular, the spatial and temporal ranges are about 50 km and 5 days, respectively, suggesting a need to account for the spatio-temporal correlations in the bias correction models. Meanwhile, the empirical variogram exhibits a complex pattern that varies across different lags in space and/or time. This calls for appropriate methodologies for complicated spatio-temporal interactions. Third, the relationships between the $PM_{2.5}$ readings collected from monitoring stations and other variables (i.e., the time variable and four meteorological variables described in Section 2.1) are not linear, as illustrated in Figs. 4 (a)-(e). Last, there exhibits a significant spatio-temporal variability in the $PM_{2.5}$ concentrations of the BTH region. The spatial variability can be seen in Fig. 3 where the $PM_{2.5}$ concentrations in Beijing are much more variable than that in Zhangjiakou. The strong temporal variation is illustrated through the time series plot in Fig. 4 (a) where the $PM_{2.5}$ concentrations around December are much higher than other times of the year.

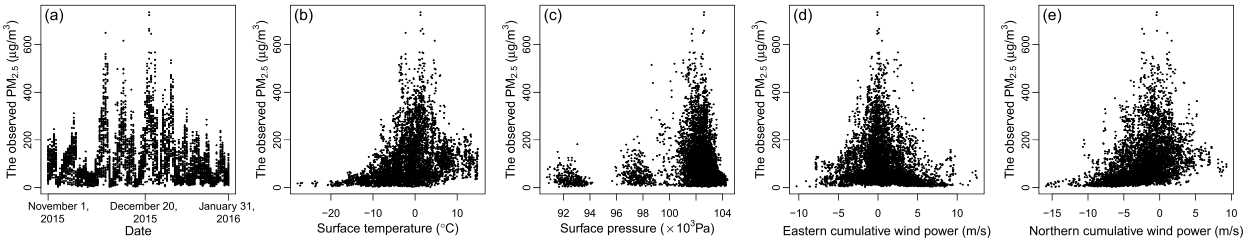

Fig. 4. (a) Time series plot of the observed $\mathrm{PM}_{2.5}$ concentrations; (b)-(e) Scatter plots of the observed $\mathrm{PM}_{2.5}$ concentration against surface temperature, surface pressure, and eastern and northern cumulative wind powers, respectively.

3. Methodology

Let $\mathcal{D}$ represent the BTH region that contains $n = 68$ monitoring stations, denoted by $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n$. At each monitoring station, the daily $\mathrm{PM}_{2.5}$ readings are recorded for a total of $N_t = 92$ days. Denote $y_t(\boldsymbol{s})$ as the square root of the reading of the monitoring station at location $\boldsymbol{s} \in \mathcal{D}$ and time point t. Based on the data explorations in Section 2.3, we connect $y_t(\boldsymbol{s})$ with the numerical model outputs through an additive partially linear model:

$$y_t(\boldsymbol{s}) = \beta_0 + \beta_1 x_t(C_{\boldsymbol{s}}) + g_0(t) + \sum_{\kappa=1}^{4} g_\kappa\left(z_{t,\kappa}(C_{\boldsymbol{s}})\right) + w_t(\boldsymbol{s}) + \varepsilon_t(\boldsymbol{s}), \tag{1}$$

where β_0 represents the intercept, $x_t(C_{\boldsymbol{s}})$ is the square root of the weighted average CMAQ outputs at location $\boldsymbol{s}$ and time point t as described in Section 2.2 with a corresponding regression coefficient β_1. The terms $z_{t,1}(C_{\boldsymbol{s}}), \cdots, z_{t,4}(C_{\boldsymbol{s}})$ represent four weighted weather outputs of NAQPMS, including surface temperature, surface pressure, eastern and northern cumulative wind powers. Along with the time variable t, these weighted outputs are connected to the readings through unknown smooth functions $g_0(\cdot), \cdots, g_4(\cdot)$. We define $w_t(\boldsymbol{s})$ as a zero-mean spatiotemporally correlated process to capture the small-scale variation of the additive bias of numerical model outputs, and $\varepsilon_t(\boldsymbol{s})$ is a zero-mean Gaussian white noise with variance σ^2.

For simplification, we let $z_{t,0}(C_{\boldsymbol{s}}) = t$ and $k = 0, 1, \cdots, 4$. Assume that the unknown smooth functions $g_k(\cdot)$ can be approximated by a basis expansion,

$$g_k\left(z_{t,k}(C_{\boldsymbol{s}})\right) = \sum_{j=0}^{J_k} \alpha_{k,j} \phi_{k,j}\left(z_{t,k}(C_{\boldsymbol{s}})\right), \tag{2}$$

where the $\left\{\phi_{k,j}(\cdot)\right\}$ are the basis functions with the corresponding basis coefficients of $\{\alpha_{k,j}\}$. To avoid identifiability issues, we assume $E\left\{g_k\left(z_{t,k}\right)\right\} = 0$ for each k, which leads to another constraint directly, that is $\sum_{i=1}^{n} \sum_{t=1}^{N_t} \sum_{j=1}^{J_k} \alpha_{k,j} \phi_{k,j}\left(z_{t,k}(C_{\boldsymbol{s}_i})\right) = 0$; see Section 5.4.1 of Wood (2017) for more details on this topic. Following (2), the model (1) can be rewritten as

$$y_t(\boldsymbol{s}) = \beta_0 + \beta_1 x_t(C_{\boldsymbol{s}}) + \sum_{k=0}^{4} \sum_{j=0}^{J_k} \alpha_{k,j} \phi_{k,j}\left(z_{t,k}(C_{\boldsymbol{s}})\right) + w_t(\boldsymbol{s}) + \varepsilon_t(\boldsymbol{s}). \tag{3}$$

Different from many other additive partially linear models, our model (3) is capable of fitting data with spatiotemporally correlated errors that are much more common and practical than independent errors. Moreover, the $w_t(\boldsymbol{s})$ process accounts for the dependence in the observations resulting in a more efficient estimation of the regression and basis coefficients compared to other methods; also refer to Table 1 obtained using a simulation study in Section 5.

To model $w_t(\boldsymbol{s})$, a widely used scheme is based on descriptive approaches (Wikle et al., 2019), where the joint spatio-temporal dependence is described using various covariance functions such as Gneiting class (Gneiting, 2002) under the stationarity assumption. However, these approaches would be computationally expensive for our study as they require the inversion of a (68 × 92) × (68 × 92) matrix in each iteration. In the literature, several strategies were adopted to reduce computational complexities, including fixed-rank approximations (Cressie and Johannesson, 2008; Cressie et al., 2010), sparse-matrix approximations (Furrer et al., 2006; Lindgren et al., 2011; Datta et al., 2016a,b), Gaussian predictive processes (Banerjee et al., 2008; Finley et al., 2012; Zhang et al., 2015, 2019), and others. In this work, we combine a basis function approximation of the process $w_t(\boldsymbol{s})$ presented in Section 3.1 with an IDE approach for the coefficients as dynamic processes given in Section 3.2. This integrated approach does not require solving the covariance of $w_t(\boldsymbol{s})$ explicitly and alleviates the computational burden.

3.1. *Multi-resolution basis function approximation of spatio-temporal processes*

The basis function approximation is a standard tool to model a spatial or spatio-temporal process due to its flexibility and computational efficiency. Many methods based on this framework have been developed, including the LatticeKrig (Nychka et al., 2015), the fixed-rank kriging (FRK, Cressie and Johannesson (2008)) approximation, the stochastic partial differential equation (SPDE, Lindgren et al. (2011)) approaches, the multi-resolution approximation (MRA, Katzfuss (2017); Jurek and Katzfuss (2021)), and references therein.

The monitoring station readings and numerical model outputs have different spatial supports, leading to differences in capturing spatial small-scale features. In our model, we use $w_t(\boldsymbol{s})$ in (3) to capture the spatial features at different scales. This is accomplished through a multi-resolution basis function approximation where the spatial variations at various scales are expressed in terms of the random coefficients at different resolutions. More specifically, we approximate process $w_t(\boldsymbol{s})$ through an R-resolution basis function

$$w_t(\boldsymbol{s}) = \sum_{r=1}^{R}\sum_{l=1}^{N_r} h_l^{(r)}(\boldsymbol{s}) v_t^{(r)}\big(B_l^r\big), \tag{4}$$

where $\big\{h_l^{(r)}(\boldsymbol{s})\big\}_{l=1}^{N_r}$ are two-dimensional spatial basis functions evaluated at location $\boldsymbol{s}$ and spatial knot B_l^r, and $\big\{v_t^{(r)}(B_l^r)\big\}_{l=1}^{N_r}$ are the corresponding N_r spatio-temporal random coefficients related to the spatial knots $\{B_l^r\}_{l=1}^{N_r}$. The knots $\big\{B_l^r\big\}_{l=1}^{N_r}$ are usually selected to be a discrete set of locations from the study region $\mathcal{D}$ and its extended area; more details are provided in Section 4.1.

Our model (4) is flexible enough to fit data with complex structures especially data with spatial and temporal variations. Earlier studies including Cressie et al. (2010), Ma and Kang (2020), and Zammit-Mangion and Cressie (2021) model the space-time process as a product of spatio-temporal basis functions (e.g., the tensor product of spatial and temporal basis functions) and some temporal random coefficients. In contrast, the space-time process $w_t(\boldsymbol{s})$ in (4) is a product of spatial basis functions and some spatio-temporal random coefficients. These random coefficients allow for variations in both space and time - a more realistic assumption for real-life datasets than the temporal random coefficients alone, as seen in Figs. 3 and 4 (a).

Model (4) is computationally efficient for datasets collected from a large number of observations in space (i.e., n is large) or large-scale numerical model outputs. Based on the multi-resolution approximation in (4), the computational complexity mainly depends on the total number of spatial knots (denoted by N_v, where $N_v = \sum_{r=1}^{R} N_r$) and is weakly related to the number of numerical model grid cells and the number of stations; see Section 4.2.2 for a more detailed discussion. In practice, the number of spatial knots is usually much less than that of the numerical model grid cells, e.g., 788 versus 5,587 in this work; see Section 4.1. We also assume the random coefficients at different resolutions are mutually and serially independent (i.e., when $r \neq r'$, $\operatorname{cov}\big(v_t^{(r)}(B_l^r), v_{t'}^{(r')}(B_{l'}^{r'})\big) = 0$ for all $\big(B_l^r, t\big)$ and $\big(B_{l'}^{r'}, t'\big)$, here $l' = 1, 2, \cdots, N_{r'}$ and $t' = 1, 2, \cdots, N_t$; similarly hereinafter), thus model (4) allows the number of spatial knots to exceed the spatial size of data (i.e., $N_v > n$) without much additional computational cost; see Section 4.2.2. Therefore, the proposed model can speed up parameter estimation significantly.

3.2. Dynamic specification for the random coefficients of basis function approximation

A dynamic specification is employed to capture the space-time interactions in our data. As seen in Fig. 2, the observed concentrations in the BTH region display lagged auto-correlation in both space and time. The IDE approach has gained considerable attention and becomes a popular tool for modeling space-time interactions; see Wikle (2002); Xu et al. (2005); Dewar et al. (2008); Cressie and Wikle (2011); Richardson et al. (2018); De Bézenac et al. (2019); Wikle et al. (2019); Zammit-Mangion and Wikle (2020). This approach admits a first-order Markov dependence in time and allows for spatial interactions by introducing a redistribution kernel $\mathcal{M}(\cdot,\cdot;\boldsymbol{\theta})$ that depends on parameters $\boldsymbol{\theta}$. Particularly, considering a spatio-temporal process with the continuous spatial index, say $\gamma_t(\boldsymbol{s})$, the stochastic IDE can be formulated as

$$\gamma_t(\boldsymbol{s}) = \int \mathcal{M}\big(\boldsymbol{s}, \boldsymbol{s}'; \boldsymbol{\theta}\big)\gamma_{t-1}(\boldsymbol{s}')d\boldsymbol{s}' + \eta_t\big(\boldsymbol{s}\big),$$

where $\eta_t\big(\boldsymbol{s}\big)$ represents random perturbations, called innovation.

Because the space-time random coefficients $v_t^{(r)}\big(B_l^r\big)$ in (4) are discretely defined with respect to space, we model the random coefficients for the rth resolution using a discretized stochastic IDE:

$$v_t^{(r)}\big(B_l^r\big) = \theta_1^{(r)} \sum_{l'=1}^{N_r} \mathcal{M}\big(B_l^r, B_{l'}^r; \theta_2^{(r)}\big) v_{t-1}^{(r)}\big(B_{l'}^r\big) + \eta_t^{(r)}\big(B_l^r\big), \tag{5}$$

where $\theta_1^{(r)} \neq 0$, the redistribution kernel $\mathcal{M}\big(B_l^r, B_{l'}^r; \theta_2^{(r)}\big)$ depends on the coordinates of location B_l^r, $B_{l'}^r$, and an unknown parameter $\theta_2^{(r)} > 0$. The innovation $\boldsymbol{\eta}_t^{(r)} = \Big(\eta_t^{(r)}\big(B_1^r\big), \cdots, \eta_t^{(r)}\big(B_{N_r}^r\big)\Big)$ follows a zero-mean Gaussian random field with a precision matrix $\boldsymbol{Q}^{(r)}$ that may vary with time, i.e., $\eta_t^{(r)} \sim \mathcal{N}\big(\boldsymbol{0}, \boldsymbol{Q}_t^{-1,(r)}\big)$. We also assume that $\boldsymbol{v}_0^{(r)} \sim \mathcal{N}\big(\boldsymbol{0}, \boldsymbol{Q}_0^{-1,(r)}\big)$, where $\boldsymbol{v}_0^{(r)} = \Big(v_0^{(r)}\big(B_1^r\big), \cdots, v_0^{(r)}\big(B_{N_r}^r\big)\Big)$. In the setting of (5), the random coefficients are also called states or processes.

This dynamic specification for the space-time random coefficients has several additional advantages over other approaches. Due to the use of the redistribution kernel, model (5) is more flexible in capturing the complex variations in space and time than the first-order autoregression imposed on the spatial dynamic downscaler model (Berrocal et al., 2012). While most statistical descriptive approaches such as the SVC need to iteratively estimate some unknown parameters for the joint spatial or spatio-temporal dependence structure (Cressie and Wikle, 2011; Wikle et al., 2019), our dynamic scheme computes only the conditional dependence making it much more efficient.

4. Model implementation

Let $\boldsymbol{y}_t = \left(y_t(\boldsymbol{s}_1), \dots, y_t(\boldsymbol{s}_n)\right)^T$ and $\boldsymbol{x}_{t,k'} = \left(x_{t,k'}(\boldsymbol{s}_1), \cdots, x_{t,k'}(\boldsymbol{s}_n)\right)^T$ where $k' = \{0, 1\}$, $x_{t,0}(\boldsymbol{s}_i) = 1$ for $t = 1, \cdots, N_t$, and $i = 1, \cdots, n$. Let $\boldsymbol{\alpha}_k$ be a J_k-dimensional column vector with α_{kj} in the jth entry, and $\boldsymbol{\phi}_{t,k}$ be an $n \times J_k$ matrix with $\phi_{k,j}\big(z_{t,k}(C_{\boldsymbol{s}_i})\big)$ in the (i, j)th element, for $j = 1, \cdots, J_k$. Let $\boldsymbol{I}_n$ be an $n \times n$ identity matrix. We express $\boldsymbol{h}_{\boldsymbol{s}}^{(r)} = \big(h_{l_1}^{(r)}(\boldsymbol{s}), \cdots, h_{N_r}^{(r)}(\boldsymbol{s})\big)$ and $\boldsymbol{v}_t^{(r)} = \Big(v_t^{(r)}\big(B_1^r\big), \cdots, v_t^{(r)}\big(B_{N_r}^r\big)\Big)$ in (4), and assume a block diagonal precision matrix $\boldsymbol{Q}_\iota = \mathrm{diag}\big(\boldsymbol{Q}_\iota^{(1)}, \cdots, \boldsymbol{Q}_\iota^{(R)}\big)$ for $\iota = 0, 1, \cdots, N_t$. We also stack each of the basis functions, the random coefficients, and the innovations, and obtain $\boldsymbol{h}_{\boldsymbol{s}} = \Big(\boldsymbol{h}_{\boldsymbol{s}}^{(1)}, \cdots, \boldsymbol{h}_{\boldsymbol{s}}^{(R)}\Big)^T$, $\boldsymbol{v}_t = \Big(\boldsymbol{v}_t^{(1)}, \cdots, \boldsymbol{v}_t^{(R)}\Big)^T$, and $\boldsymbol{\eta}_t = \Big(\boldsymbol{\eta}_t^{(1)}, \cdots, \boldsymbol{\eta}_t^{(R)}\Big)^T$. Thus, the complete ADCM can be written in a matrix and hierarchical form:

$$\begin{cases} \boldsymbol{y}_t = \boldsymbol{X}_t\boldsymbol{\beta} + \boldsymbol{\Phi}_t\boldsymbol{\alpha} + \boldsymbol{H}\boldsymbol{v}_t + \boldsymbol{\varepsilon}_t & \text{with } \boldsymbol{\varepsilon}_t \sim \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I}_n) \\ \boldsymbol{v}_t = \boldsymbol{M}_\theta\boldsymbol{v}_{t-1} + \boldsymbol{\eta}_t & \text{with } \boldsymbol{v}_0 \sim \mathcal{N}(\boldsymbol{0},\, \boldsymbol{Q}_0^{-1}) \text{ and } \boldsymbol{\eta}_t \sim \mathcal{N}(\boldsymbol{0},\, \boldsymbol{Q}_t^{-1}), \end{cases} \tag{6}$$

where $\boldsymbol{X}_t = \left(\boldsymbol{x}_{t,0}, \boldsymbol{x}_{t,1}\right)$, $\boldsymbol{\beta} = \left(\beta_0, \beta_1\right)^T$, $\boldsymbol{\Phi}_t = \left(\boldsymbol{\phi}_{t,0}, \cdots, \boldsymbol{\phi}_{t,4}\right)$ is an $n \times J$ basis design matrix with $J = j_0 + \cdots + j_4$, $\boldsymbol{\alpha} = \left(\boldsymbol{\alpha}_0^T, \cdots, \boldsymbol{\alpha}_4^T\right)^T$ is a $J \times 1$ vector of unknown basis coefficients, $\boldsymbol{H} = \left(\boldsymbol{h}_{\boldsymbol{s}_1}, \cdots, \boldsymbol{h}_{\boldsymbol{s}_n}\right)^T$, $\boldsymbol{\varepsilon}_t = \left(\varepsilon_t(\boldsymbol{s}_1), \dots, \varepsilon_t(\boldsymbol{s}_n)\right)^T$, and $\boldsymbol{M}_\theta = \mathrm{diag}\Big(\theta_1^{(1)}\boldsymbol{\mathcal{M}}_{\theta_2^{(1)}}, \cdots, \theta_1^{(R)}\boldsymbol{\mathcal{M}}_{\theta_2^{(R)}}\Big)$ is a block diagonal matrix with $\theta_1^{(r)}\boldsymbol{\mathcal{M}}_{\theta_2^{(r)}}$ at the rth block, here $\boldsymbol{\mathcal{M}}_{\theta_2^{(r)}}$ is an $N_r \times N_r$ matrix with $\mathcal{M}\big(B_l^r, B_{l'}^r; \theta_2^{(r)}\big)$ in its lth row and l'th column. The error $\boldsymbol{\varepsilon}_t$ and the innovation $\boldsymbol{\eta}_t$ are mutually and serially independent, which also holds for $\boldsymbol{\eta}_t$ and $\boldsymbol{v}_{t-1}$.

Recently, to give a specific form of dynamic spatio-temporal model, Bradley et al. (2015a) proposed to restrict $\boldsymbol{M}_\theta$ to the Moran's I (MI) basis function as an known propagator matrix; see also Bradley et al. (2015b, 2019). MI is especially useful for analyzing areal data and reducing dimensional space for a very large spatial observation n. However, for point-referenced spatio-temporal data such as the BTH $PM_{2.5}$ concentrations and a small n, the applicability and performance of MI may require further investigation. An alternative approach is to set $\mathcal{M}\big(B_l, B_{l'}; \theta_w\big)$ as physics-based operators based on a reaction-diffusion process (Hooten and Wikle, 2008; Hefley et al., 2017) or fluid dynamics (Wikle and Hooten, 2010) when the information of the physical process is available. For the ADCM, we hope that the proposed methodology can be applied to multiple pollutant datasets, not just the $PM_{2.5}$. Thus, the form of $\mathcal{M}\big(B_l, B_{l'}; \theta_w\big)$ is defined based on the results of data exploration while its dependent parameters are learned from the data. Specifically, for the BTH $PM_{2.5}$ data, we try to capture spatial interactions of the processes as the processes evolve over time while removing the spurious spatial interactions from large spatial distances, as pointed out by Wikle and Hooten (2010) that "some types of interaction make sense for some processes, and some do not." To achieve these goals, $\mathcal{M}\big(B_l, B_{l'}; \theta_w\big)$ is set as a Wendland function with compact support $W\big(d_{l,l'}; \theta_w\big)$ with the parameter $\theta_w > 0$, where $d_{l,l'} = \|B_l - B_{l'}\|$ and $W(d; \theta_w)$ is given by

$$W\big(d; \theta_w\big) = \begin{cases} \frac{1}{12}\big(1 - d/\theta_w\big)^3\big(1 + 3d/\theta_w\big) & \text{if } 0 \le d \le \theta_w, \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

This function expedites computation by creating sparse matrices. A similar Wendland function is also used in Nychka et al. (2015). More details on the Wendland functions can be found in Wendland (1995).

The implementation issues of the ADCM are studied in Sections 4.1–4.4, and the bias correction specification is given in Section 4.5.

4.1. Selection of basis functions and definition of spatial knots

Due to the seasonal behavior of the $PM_{2.5}$ concentrations and sparse distribution in the boundary of meteorological variables as illustrated in Fig. 4, we construct the basis design matrix $\boldsymbol{\Phi}_t$ in (6) from cyclic cubic-spline basis functions with equidistant knots. The basis function $h_l^{(r)}(\boldsymbol{s})$ in (4) is defined as the Wendland function in (7) and written as $h_l^{(r)}(\boldsymbol{s}) = W\big(d_{\boldsymbol{s},B_l^r}; c_h d_{\max}^H 2^{1-r}\big)$ where $d_{\boldsymbol{s},B_l^r} = \|\boldsymbol{s} - B_l^r\|$, $c_h \in (0, 1]$ is a constant, $d_{\max}^H = \max_{\boldsymbol{s},l}\big\{\|\boldsymbol{s} - B_l^r\|\big\}$, and 2^{1-r} is a scale factor (refer to Section 2.3 of Nychka et al. (2015)).

We develop a procedure to parameterize $\boldsymbol{Q}_\iota$ for $\iota = 0, 1, \cdots, N_t$ in (6) by introducing N_v spatial knots $\boldsymbol{B}$ where $\boldsymbol{B} = \big\{B_1^1, \cdots, B_{N_1}^1, \cdots, B_1^2, \cdots, B_{N_R}^R\big\}$. To capture spatial variation at various scales, we construct an R-resolution regular grid that covers the entire BTH region with the number of resolutions $R = 2$ (see Fig. 5). The details of the multi-resolution

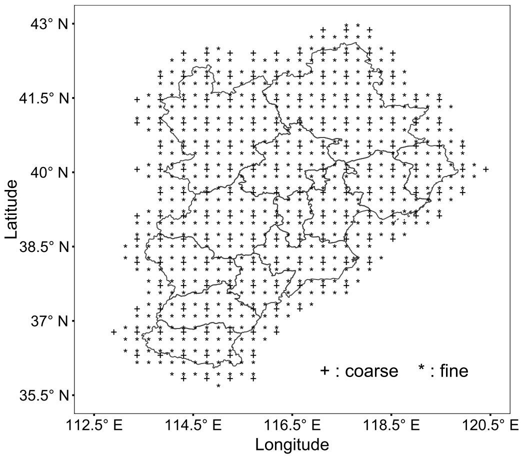

Fig. 5. Two different resolutions of the BTH region and its extended area ("+" = coarse and "*" = fine). The symbols represent the centroids of the rectangular grids with a total of 788 points with 168 points at the coarse resolution and 620 points at the fine resolution.

grid construction can be found in Nychka et al. (2015). Using the rectangular grid centroids as the knots $\boldsymbol{B}$, we obtain 2-resolution spatial Markov structures based on the relationships among the knots at each resolution. For the rth resolution, we subsequently model the initial process $\boldsymbol{v}_0^{(r)}$ and the innovation $\boldsymbol{\eta}_t^{(r)}$ as two Gaussian Markov random fields (GMRFs, Rue and Held (2005)) and specify their precision matrices as $\boldsymbol{Q}_0^{(r)} = \tau_0^{2,(r)}\big(\boldsymbol{G}^{(r)} + \zeta_0^{2,(r)}\boldsymbol{I}_{N_r}\big)$ and $\boldsymbol{Q}_t^{(r)} = \tau^{2,(r)}\big(\boldsymbol{G}^{(r)} + \zeta^{2,(r)}\boldsymbol{I}_{N_r}\big)$, respectively. The elements of matrix $\boldsymbol{G}^{(r)}$ are given by

$$G_{ll'}^{(r)} = \begin{cases} \text{the number of neighbors of knot } B_l^r & \text{if } l = l', \\ -1 & \text{if } B_l^r \text{ is adjacent to } B_{l'}^r, \\ 0 & \text{otherwise.} \end{cases}$$

As suggested by one reviewer, we can also set a precision matrix from a GMRF to $\boldsymbol{Q}_t^{(r)} = \tau^{2,(r)}\boldsymbol{I}$ with just one parameter $\tau^{2,(r)}$. When most of the spatial interactions can be captured through a well-defined redistribution kernel $\mathcal{M}(\cdot,\cdot;\boldsymbol{\theta})$, the simplest form of $\boldsymbol{Q}$ is preferred for the ADCM. The ADCM also allows other specifications for precision matrices such as conditional autoregressive (CAR) models (Besag, 1974; Sørbye and Rue, 2014; Bivand et al., 2015), the Leroux model (Leroux et al., 2000) for a generalization of the intrinsic CAR (ICAR, Besag et al. (1991)), and the simultaneous autoregressive (SAR, Anselin (1988)). See Cressie and Wikle (2011) and references for further detailed discussions on these specifications. Under the setting of the ADCM, a sensitivity analysis is provided in the Supplementary Material by using several different specifications of the precision matrices $\boldsymbol{Q}$.

4.2. Estimation and algorithm

Let $\boldsymbol{y}_{1:t} = \{\boldsymbol{y}_1, \cdots, \boldsymbol{y}_t\}$, $\boldsymbol{v}_{0:t} = \{\boldsymbol{v}_0, \cdots, \boldsymbol{v}_t\}$, and $\boldsymbol{\Theta} = \{\boldsymbol{\beta}, \boldsymbol{\alpha}, \sigma^2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\tau}^2, \boldsymbol{\zeta}, \boldsymbol{\tau}_0^2, \boldsymbol{\zeta}_0\}$ where $\boldsymbol{\theta}_1 = \big\{\theta_1^{(1)}, \cdots, \theta_1^{(R)}\big\}$, $\boldsymbol{\theta}_2 = \big\{\theta_2^{(1)}, \cdots, \theta_2^{(R)}\big\}$, $\boldsymbol{\tau}^2 = \big\{\tau^{2,(1)}, \cdots, \tau^{2,(R)}\big\}$, $\boldsymbol{\zeta} = \big\{\zeta^{(1)}, \cdots, \zeta^{(R)}\big\}$, $\boldsymbol{\tau}_0^2 = \big\{\tau_0^{2,(1)}, \cdots, \tau_0^{2,(R)}\big\}$, and $\boldsymbol{\zeta}_0 = \big\{\zeta_0^{(1)}, \cdots, \zeta_0^{(R)}\big\}$. Denote all observed data as $\boldsymbol{y} = \boldsymbol{y}_{1:N_t}$ and the random coefficients as $\boldsymbol{v} = \boldsymbol{v}_{0:N_t}$. We develop a multi-resolution ensemble Kalman smoother (MEnKS) scheme which is embedded in an EM algorithm (hereafter, MEnKS-EM) for parameter estimation. The filtering and smoothing of $\boldsymbol{v}$ are evaluated using the MEnKS algorithm first, and then the maximum likelihood estimation (MLE) of $\boldsymbol{\Theta}$ is carried out through the EM algorithm. Moreover, a conditional simulation method is employed to assess the uncertainty quantification of several quantities including the estimates of $\boldsymbol{\Theta}$ and the corrections of the numerical model outputs.

4.2.1. Likelihood function

The parameter estimation of the ADCM can be viewed as an MLE problem with missing data $\boldsymbol{v}$ and subsequently, solved using the EM algorithm (Dempster et al., 1977). We decompose the joint density of $(\boldsymbol{y}, \boldsymbol{v})$ into the following,

$$p\big(\boldsymbol{y}, \boldsymbol{v}|\boldsymbol{\Theta}\big) = \prod_{t=1}^{N_t} p\big(\boldsymbol{y}_t|\boldsymbol{v}_t, \boldsymbol{\beta}, \boldsymbol{\alpha}, \sigma^2\big) \prod_{t=1}^{N_t} p\big(\boldsymbol{v}_t|\boldsymbol{v}_{t-1}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\tau}^2, \boldsymbol{\zeta}\big) p\big(\boldsymbol{v}_0|\boldsymbol{\tau}_0^2, \boldsymbol{\zeta}_0\big). \tag{8}$$

Based on the distributional assumptions of $\boldsymbol{\varepsilon}_t$, $\boldsymbol{\eta}_t$ and $\boldsymbol{v}_0$ in (6), the log-likelihood of $(\boldsymbol{y}, \boldsymbol{v})$, up to an additive constant, is given by

$$\begin{aligned}2\log L\big(\boldsymbol{\Theta}; \boldsymbol{y}, \boldsymbol{v}\big) = & -nN_t \log\sigma^2 - \sum_{t=1}^{N_t} \frac{\big(\boldsymbol{y}_t - \boldsymbol{X}_t\boldsymbol{\beta} - \boldsymbol{\Phi}_t\boldsymbol{\alpha} - \boldsymbol{H}\boldsymbol{\upsilon}_t\big)^T\big(\boldsymbol{y}_t - \boldsymbol{X}_t\boldsymbol{\beta} - \boldsymbol{\Phi}_t\boldsymbol{\alpha} - \boldsymbol{H}\boldsymbol{\upsilon}_t\big)}{\sigma^2} + \\ & \sum_{t=1}^{N_t} \log|\boldsymbol{Q}_t| - \sum_{t=1}^{N_t} \big(\boldsymbol{\upsilon}_t - \boldsymbol{M}_\theta\boldsymbol{\upsilon}_{t-1}\big)^T \boldsymbol{Q}_t\big(\boldsymbol{\upsilon}_t - \boldsymbol{M}_\theta\boldsymbol{\upsilon}_{t-1}\big) + \\ & \log|\boldsymbol{Q}_0| - \sum_{t=1}^{N_t} \big(\boldsymbol{\upsilon}_0 - \boldsymbol{\mu}_0\big)^T \boldsymbol{Q}_0\big(\boldsymbol{\upsilon}_0 - \boldsymbol{\mu}_0\big). \qquad (9)\end{aligned}$$

4.2.2. The MEnKS-EM algorithm

We denote the conditional expectation of (9) with respect to $\boldsymbol{v}$ as the Q-function, i.e., $\mathrm{Q}\Big(\boldsymbol{\Theta}\big|\boldsymbol{\Theta}^{(\ell)}\Big) = E_{\boldsymbol{v}}\Big[2\log L\big(\boldsymbol{\Theta}; \boldsymbol{y}, \boldsymbol{v}\big)\big|\boldsymbol{y}, \boldsymbol{\Theta}^{(\ell)}\Big]$ where $\boldsymbol{\Theta}^{(\ell)}$ is the estimate of the (ℓ)th iteration. The EM algorithm alternates between the E-steps and the M-steps.

In the E-step, to compute the Q-function, we first obtain the filtering distribution $p\big(\boldsymbol{v}_t|\boldsymbol{y}_{1:t}, \boldsymbol{\Theta}^{(\ell)}\big)$ and the smoothing distribution $p\big(\boldsymbol{v}_t|\boldsymbol{y}_{1:u}, \boldsymbol{\Theta}^{(\ell)}\big)$ where $u > t$. The exact computation of the Q-function requires both filtering and smoothing distributions to be calculated exactly. Under a linear Gaussian assumption, although Kalman filter and Kalman smoother can infer these distributions exactly, their computations and storage easily become too expensive to be feasible (Katzfuss et al., 2016) as the number of spatial knots becomes large. This calls for approximating the Q-function. Several approximation methods have been proposed in the past including the stochastic EM (SEM) algorithm (Celeux, 1985; Ma and Kang, 2020) and the Monte Carlo EM (MCEM) algorithm (Wei and Tanner, 1990). Instead of the SEM and MCEM algorithms, we extend the EnKS algorithm proposed by Evensen and Van Leeuwen (2000) to a multi-resolution EnKS for approximating the Q-function of our proposed model, where the expectation of the random variables on the right-hand side of (9) is approximated by a sample or an ensemble from the distribution. More specifically, the ensemble is first generated from the distribution of initial process $\boldsymbol{v}_0$ and then propagated forward in time based on the dynamic model (5). When new data become available, the ensemble is updated using a method similar to the Kalman filter. The filtering and the smoothing distributions of $\boldsymbol{v}$ can then be evaluated using the updated ensemble approximately. See the Algorithm 1 for more details on the MEnKS.

In Algorithm 1, since N_e is usually less than N_r, two tapering functions are used to avoid the rank deficiency and defined using the Wendland function as seen in (7) (Wikle et al., 2019). Specifically, the spatial tapering function is formed by $\Gamma_{c_s}^{(r)}(l, l') = W\big(d_{l,l'}; c_s d_{\max}^B 2^{1-r}\big)$ where $\Gamma_{c_s}^{(r)}(l, l')$ represents the (l, l')th element of $\boldsymbol{\Gamma}_{c_s}^{(r)}$, $c_s \in (0, 1]$ is a given constant, and $d_{\max}^B = \max_{l,l'}\big\{\|\boldsymbol{B}_l - \boldsymbol{B}_{l'}\|\big\}$. The temporal tapering function is set to $\gamma_{c_t}(t', t) = W\big(d_t; c_t\big)$ where $d_t = t - t'$, and $c_t \in [1, N_t]$ is a known constant.

In the M-step, because the smoothness of the unknown functions $g_k(\cdot)$ in (1) usually needs to be controlled for a good fit of the data, a modification of the Q-function is required. Following Wood (2020), we first introduce a roughness penalty to the second term on the right-hand side of (9) and impose some identifiability constraints on (2). We then obtain a modified Q-function, denoted as $\widetilde{\mathrm{Q}}\Big(\boldsymbol{\Theta}\big|\boldsymbol{\Theta}^{(\ell)}\Big) = E_{\boldsymbol{v}}\Big[\big\{2\log L\big(\boldsymbol{\Theta}; \boldsymbol{y}, \boldsymbol{v}\big) - \sum_{k=0}^{4}\lambda_k\sum_{t=1}^{N_t}\big(\int g_k''(\boldsymbol{z}_{t,k})^2 dz + \boldsymbol{\alpha}_k^T\boldsymbol{\phi}_{t,k}^T\boldsymbol{1}_n\boldsymbol{1}_n^T\boldsymbol{\phi}_{t,k}\boldsymbol{\alpha}_k\big)\big\}\big|\boldsymbol{y}, \boldsymbol{\Theta}^{(\ell)}\Big]$, where $\{\lambda_k\}_{k=0}^4$ are the tuning parameters, g'' is the second derivative of g, and $\boldsymbol{1}_n$ is an n-dimensional column vector filled with ones. The estimate of $\boldsymbol{\alpha}$ can also be found analytically; see Appendix.

In summary, the MEnKS-EM algorithm is developed to solve the MLE of the ADCM iteratively. In the E-step, for a given set of parameters $\boldsymbol{\Theta}$, we first obtain the ensemble of the process $\boldsymbol{v}$ using the MEnKS in the Algorithm 1 and then evaluate the expectation, variance-covariance, and cross-covariance of $\boldsymbol{v}$ in the E-step of Algorithm 2. In the M-step, we update the estimation of $\boldsymbol{\Theta}$ by maximizing $\widetilde{\mathrm{Q}}$; see Appendix. The entire MEnKS-EM algorithm is also summarized in Algorithm 2.

The MEnKS-EM algorithm has several computational advantages. First, the sparse matrix $\boldsymbol{G}^{(r)}$ implies the sparsity of $\boldsymbol{Q}_\iota^{(r)}$, which leads to a fast Cholesky factorization and thus an efficient sampling (Rue and Held, 2005) in the forecast step of Algorithm 1. Second, Algorithm 1 is a forward-only process unlike the standard forward-backward method (e.g., the Kalman smoother). Third, the independence of the GMRFs of different resolutions introduces some block-diagonal matrices such as the cross-covariance in the Algorithm 1, the state transition matrix $\boldsymbol{M}_\theta$, and the precision matrix $\boldsymbol{Q}_\iota$ involved in Algorithm 2; these sparse structures speed up the parameter estimation. Lastly, for the usual case where $n \gg N_v$, we can quickly invert an $n \times n$ matrix related to the Kalman gain $\widehat{\boldsymbol{K}}_{t',t}$ in (U4) of the MEnKS using the Woodbury formula, i.e., $\big(\boldsymbol{H}\widehat{\boldsymbol{\Sigma}}_{t,t|t-1}\boldsymbol{H}^T + \widehat{\sigma}^2\boldsymbol{I}_n\big)^{-1} = \widehat{\sigma}^{-2}\boldsymbol{I}_n - \widehat{\sigma}^{-2}\boldsymbol{H}\boldsymbol{\Omega}^{-1}\boldsymbol{H}^T$, where $\boldsymbol{\Omega} = \widehat{\sigma}^2\widehat{\boldsymbol{\Sigma}}_{t,t|t-1}^{-1} + \boldsymbol{H}^T\boldsymbol{H}$. The matrix $\boldsymbol{\Omega}$ is sparse for a block-diagonal matrix $\widehat{\boldsymbol{\Sigma}}_{t,t|t-1}$ and a sparse $\boldsymbol{H}$. For a given $N_v \times N_e$ matrix $\boldsymbol{\eta}$, the linear system $\boldsymbol{\Omega}\boldsymbol{\xi} = \boldsymbol{\eta}$ for $\boldsymbol{\xi}$ can be solved efficiently due to a sparse Cholesky decomposition of $\boldsymbol{\Omega}$, where the forward and backward substituting algorithms are used sequentially. As a result, (U4) of the MEnKS can be modified to

Algorithm 1: Multi-resolution Ensemble Kalman smoother (MEnKS).

```
We use $\{\boldsymbol{v}_{t'|t}^{(e)}\}_{e=1}^{N_e}$ to denote $N_e$ ensemble members, which are approximate samples from $p(\boldsymbol{v}_{t'}|\boldsymbol{y}_{1:t}, \boldsymbol{\Theta})$.
Input: $\widehat{\boldsymbol{\Theta}}$ and residuals $\widehat{\boldsymbol{\varepsilon}}_t = \boldsymbol{y}_t - \boldsymbol{X}_t\widehat{\boldsymbol{\beta}} - \boldsymbol{\Phi}_t\widehat{\boldsymbol{\alpha}}$. Initialize $N_e$ ensemble members $\left\{\boldsymbol{v}_{0|0}^{(e)}\right\}_{e=1}^{N_e}$ where $\boldsymbol{v}_{0|0}^{(e)} = (\widetilde{\boldsymbol{v}}_{0|0}^{(1,e)}, \cdots, \widetilde{\boldsymbol{v}}_{0|0}^{(r,e)})^T$ with $\widetilde{\boldsymbol{v}}_{0|0}^{(r,e)} \sim$
        $\mathcal{N}\left(\boldsymbol{0}, \widehat{\boldsymbol{Q}}_0^{-1,(r)}\right)$, here $\widehat{\boldsymbol{Q}}_0^{(r)} = \widehat{\tau}_0^{2,(r)}\left(\boldsymbol{G}^{(r)} + \widehat{\zeta}_0^{2,(r)}\boldsymbol{I}_{N_r}\right)$.
begin
    for $t = 1, \cdots, N_t$ do
        for $e = 1, \cdots, N_e$ do
            Forecast step: for $r = 1, \cdots, R$ do
                $\widetilde{\boldsymbol{v}}_{t|t-1}^{(r,e)} = \widehat{\theta}_1^{(r)}\boldsymbol{\mathcal{M}}_{\widehat{\theta}_2^{(r)}}\widetilde{\boldsymbol{v}}_{t-1|t-1}^{(r,e)} + \boldsymbol{\eta}_t^{(r,e)}$ where $\boldsymbol{\eta}_t^{(r,e)} \sim \mathcal{N}\left(\boldsymbol{0}, \widehat{\boldsymbol{Q}}_t^{-1,(r)}\right)$ and $\widehat{\boldsymbol{Q}}_t^{(r)} = \widehat{\tau}^{2,(r)}\left(\boldsymbol{G}^{(r)} + \widehat{\zeta}^{2,(r)}\boldsymbol{I}_m\right)$.
            end
            Update step: for $t' = \max\{0, t - c_t\}, \cdots, t$ do
                for $r = 1, \cdots, R$ do
                    (U1) Update the cross-covariance matrix: $\widetilde{\boldsymbol{\Sigma}}_{t',t|t-1}^{(r)} = \frac{1}{N_e-1}\sum_{e=1}^{N_e}\left(\widetilde{\boldsymbol{v}}_{t'|t-1}^{(r,e)} - \widehat{\boldsymbol{v}}_{t'|t-1}^{(r)}\right)\left(\widetilde{\boldsymbol{v}}_{t|t-1}^{(r,e)} - \widehat{\boldsymbol{v}}_{t|t-1}^{(r)}\right)^T$
                         where $\widehat{\boldsymbol{v}}_{t'|t-1}^{(r)} = \frac{1}{N_e}\sum_{e=1}^{N_e}\widetilde{\boldsymbol{v}}_{t'|t-1}^{(r,e)}$.
                    (U2) Regularize the cross-covariance matrix: $\widehat{\boldsymbol{\Sigma}}_{t',t|t-1}^{(r)} = \widetilde{\boldsymbol{\Sigma}}_{t',t|t-1}^{(r)} \circ \boldsymbol{\Gamma}_{c_s}^{(r)}\gamma_{c_t}(t', t)$, where the "$\circ$"
                         denotes the element-wise multiplication, $\boldsymbol{\Gamma}_{c_s}^{(r)}$ and $\gamma_{c_t}(t', t)$ are defined using (7), $c_s$ and
                         $c_t$ are two fixed tapering parameters.
                end
                (U3) Obtain a complete cross-covariance: $\widehat{\boldsymbol{\Sigma}}_{t',t|t-1} = \text{diag}\left(\widehat{\boldsymbol{\Sigma}}_{t',t|t-1}^{(1)}, \cdots, \widehat{\boldsymbol{\Sigma}}_{t',t|t-1}^{(R)}\right)$.
                (U4) Update the smoothed ensemble:
                         $\boldsymbol{v}_{t'|t}^{(e)} = \boldsymbol{v}_{t'|t-1}^{(e)} + \widehat{\boldsymbol{K}}_{t',t}\left(\widehat{\boldsymbol{\varepsilon}}_t - \widetilde{\boldsymbol{y}}_t^{(e)}\right),$        (10)
                     where $\boldsymbol{v}_{t'|t}^{(e)} = (\widetilde{\boldsymbol{v}}_{t'|t}^{(1,e)}, \cdots, \widetilde{\boldsymbol{v}}_{t'|t}^{(r,e)})^T$, $\widehat{\boldsymbol{K}}_{t',t} = \widehat{\boldsymbol{\Sigma}}_{t',t|t-1}\boldsymbol{H}^T(\boldsymbol{H}\widehat{\boldsymbol{\Sigma}}_{t,t|t-1}\boldsymbol{H}^T + \widehat{\sigma}^2\boldsymbol{I}_n)^{-1}$, and $\widetilde{\boldsymbol{y}}_t^{(e)} =$
                     $\boldsymbol{H}\boldsymbol{v}_{t|t-1}^{(e)} + \boldsymbol{\varepsilon}_t^{(e)}$ is the pseudo observation with noise $\boldsymbol{\varepsilon}_t^{(e)} \sim \mathcal{N}(\boldsymbol{0}, \widehat{\sigma}^2\boldsymbol{I}_n)$.
            end
        end
    end
    return all ensemble members $\{\boldsymbol{v}_{t|T}^{(e)}\}$ at all time points.
end
```

$$\boldsymbol{v}_{t'|t}^{(e)} = \boldsymbol{v}_{t'|t-1}^{(e)} + \widehat{\sigma}^{-2}\widehat{\boldsymbol{\Sigma}}_{t,t'|t-1}\boldsymbol{H}^T\left\{\boldsymbol{I}_n\left(\widehat{\boldsymbol{\varepsilon}}_t - \widetilde{\boldsymbol{y}}_t^{(e)}\right) - \boldsymbol{H}\widehat{\boldsymbol{\xi}}\right\}, \tag{11}$$

where $\widehat{\boldsymbol{\xi}}$ is a solution of the system $\boldsymbol{\Omega}\boldsymbol{\xi} = \boldsymbol{H}^T\left(\widehat{\boldsymbol{\varepsilon}}_t - \widetilde{\boldsymbol{y}}_t^{(e)}\right)$ and obtained via functions `forwardsolve` and `backsolve` in the `R` package `spam` (Furrer and Sain, 2010). We find that the computation of both (11) and the system above is nearly 15 times faster than that of (U4) using a simulation with spatial size $n = 2{,}000$ in Section 5.

4.3. Selection of tuning parameters

Several tuning parameters need to be selected. First, the parameter c_h for $\boldsymbol{H}$ in Section 4.1 depends on the spatial distribution of the data, and data exploration may help in its selection, such as the empirical variogram in Fig. 2. In general, the value of the product $c_h d_{\max}^H$ related to the basis matrix $\boldsymbol{H}$ in Section 4.1 is at least equal to the spatial range of data. Since the monitoring stations in the BTH region are distributed sparsely in space and clustered in urban areas, to identify the bias of model outputs in rural areas, we would have to set the value of the product to be greater than the spatial range obtained from the empirical variogram. We set $c_h = 0.3$ in the data analysis of Section 6, corresponding to about 180 km. In the simulation study of Section 5, because the spatial size of the simulated data is greater than the number of monitoring stations and the simulated points are not clustered in space, we use a smaller $c_h = 0.2$. Second, for the two tapering parameters in Algorithm 1, c_s and c_t, the simulation studies and data analysis suggest that the ADCM performs well when these values are not too large. We set $c_s = 0.1$ for the BTH data, $c_s = 0.2$ for the simulation studies, and $c_t = 1$ for all cases. Lastly, when we increase the number of ensemble members N_e, we could achieve a higher accuracy of approximation but at the expense of computational costs. In the literature, $N_e = O(100)$ is considered to be a reasonable choice (Mitchell et al., 2002; Houtekamer and Zhang, 2016), and we use $N_e = 100$ here. It is worth mentioning that in real-world applications, the tuning parameters can be selected using cross-validation methods such as the procedure described in Section 6.1. Moreover, kernel-based bandwidth selection methods, such as those proposed by Sheather and Jones (1991) and Chaudhuri et al. (2017), may potentially be applied to explore the appropriate value for c_h.

Algorithm 2: The MEnKS-EM algorithm for the ADCM.

```
We let $\epsilon$ represent a relative error. The convergence threshold is set to 0.01 in this work.
Input : $\widehat{\boldsymbol{\Theta}}^{(0)}$, set the current data log-likelihood as CurrentLoglik $=\sum_{t=1}^{N_e}\log p\big(\boldsymbol{y}_t|\boldsymbol{v}_t,\boldsymbol{\beta}^{(0)},\boldsymbol{\alpha}^{(0)},\sigma^{2,(0)}\big)$ with $\boldsymbol{v}_t=\mathbf{0}$, and $\epsilon^{(0)}=|\text{CurrentLoglik}|$.
begin
    PastLoglik = CurrentLoglik and $\ell=1$.
    while $\epsilon^{(\ell-1)}>0.01$ or $\ell=1$ do
        1. Ensemble-step: Input $\widehat{\boldsymbol{\Theta}}^{(\ell-1)}$ to Algorithm 1 and obtain ensemble members $\big\{\boldsymbol{v}_{t|N_t}^{(e,\ell)}\big\}$.
        2. E-step: Based on $\big\{\boldsymbol{v}_{t|N_t}^{(e,\ell)}\big\}$, evaluate the expectation, variance-covariance, and cross-covariance of $\boldsymbol{v}$ by executing the following loops: for $r=1,\cdots,R$ do
            for $t=1,\cdots,N_t$ do
                $\widehat{\boldsymbol{v}}_{t|N_t}^{(r,\ell)}=\frac{1}{N_e}\sum_{e=1}^{N_e}\widetilde{\boldsymbol{v}}_{t|N_t}^{(r,e,\ell)},$
                $\widehat{\boldsymbol{\Sigma}}_{t,t|N_t}^{(r,\ell)}=\frac{1}{N_e-1}\sum_{e=1}^{N_e}\Big(\widetilde{\boldsymbol{v}}_{t|N_t}^{(r,e,\ell)}-\widehat{\boldsymbol{v}}_{t|N_t}^{(r,\ell)}\Big)\Big(\widetilde{\boldsymbol{v}}_{t|N_t}^{(r,e,\ell)}-\widehat{\boldsymbol{v}}_{t|N_t}^{(r,\ell)}\Big)^T\circ\boldsymbol{\Gamma}_{c_s}\gamma_{c_t}(t,t),$
                $\widehat{\boldsymbol{\Sigma}}_{t-1,t|N_t}^{(r,\ell)}=\frac{1}{N_e-1}\sum_{e=1}^{N_e}\Big(\widetilde{\boldsymbol{v}}_{t-1|N_t}^{(r,e,\ell)}-\widehat{\boldsymbol{v}}_{t-1|N_t}^{(r,\ell)}\Big)\Big(\widetilde{\boldsymbol{v}}_{t|N_t}^{(r,e,\ell)}-\widehat{\boldsymbol{v}}_{t|N_t}^{(r,\ell)}\Big)^T\circ\boldsymbol{\Gamma}_{c_s}\gamma_{c_t}(t-1,t).$
            end
        end
        Obtain three block diagonal matrices,
            $\boldsymbol{S}_{00}^{(\ell)}=\text{diag}\Big(\boldsymbol{S}_{00}^{(1,\ell)},\cdots,\boldsymbol{S}_{00}^{(R,\ell)}\Big),$
            $\boldsymbol{S}_{01}^{(\ell)}=\text{diag}\Big(\boldsymbol{S}_{01}^{(1,\ell)},\cdots,\boldsymbol{S}_{01}^{(R,\ell)}\Big),$ and
            $\boldsymbol{S}_{11}^{(\ell)}=\text{diag}\Big(\boldsymbol{S}_{11}^{(1,\ell)},\cdots,\boldsymbol{S}_{11}^{(R,\ell)}\Big),$
        where $\boldsymbol{S}_{00}^{(r,\ell)}=\sum_{t=1}^{N_t}\Big(\widehat{\boldsymbol{v}}_{t-1|N_t}^{(r,\ell)}\widehat{\boldsymbol{v}}_{t-1|N_t}^{T,(r,\ell)}+\widehat{\boldsymbol{\Sigma}}_{t-1,t-1|N_t}^{(r,\ell)}\Big)$, $\boldsymbol{S}_{01}^{(r,\ell)}=\sum_{t=1}^{N_t}\Big(\widehat{\boldsymbol{v}}_{t-1|N_t}^{(r,\ell)}\widehat{\boldsymbol{v}}_{t|N_t}^{T,(r,\ell)}+\widehat{\boldsymbol{\Sigma}}_{t-1,t|N_t}^{(r,\ell)}\Big)$, and $\boldsymbol{S}_{11}^{(r,\ell)}=\sum_{t=1}^{N_t}\Big(\widehat{\boldsymbol{v}}_{t|N_t}^{(r,\ell)}\widehat{\boldsymbol{v}}_{t|N_t}^{T,(r,\ell)}+\widehat{\boldsymbol{\Sigma}}_{t,t|N_t}^{(r,\ell)}\Big)$
        for $r=1,\cdots,R$.
        3. M-step: Update $\widehat{\boldsymbol{\Theta}}^{(\ell-1)}$ to $\widehat{\boldsymbol{\Theta}}^{(\ell)}$ using results provided in Appendix. Update several variables to
            $\text{CurrentLoglik}=\sum_{t=1}^{N_t}\log p\big(\boldsymbol{y}_t|\boldsymbol{v}_{t|N_t}^{(\ell)},\boldsymbol{\beta}^{(\ell)},\boldsymbol{\alpha}^{(\ell)},\sigma^{2,(\ell)}\big),$
            $\epsilon^{(\ell)}=\Big|\frac{\text{PastLoglike}-\text{CurrentLoglik}}{\text{PastLoglike}}\Big|,$
            $\text{PastLoglike}=\text{CurrentLoglik},$ and
            $\ell=\ell+1.$
    end
    return $\widehat{\boldsymbol{\Theta}}=\widehat{\boldsymbol{\Theta}}^{(\ell)}$ and $\boldsymbol{v}_{t|N_t}=\boldsymbol{v}_{t|N_t}^{(\ell)}$.
end
```

4.4. Prediction using conditional simulation

Following Nychka et al. (2015), we adopt a conditional simulation method to quantify the uncertainty of the parameter estimates and predictions. To be specific, we first use the Monte Carlo algorithm to obtain a sample of the conditional distribution $p\big(\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{v}|\boldsymbol{y},\widehat{\boldsymbol{\Theta}}\big)$, then generate a space-time predictive sample at any unobservable spatial location $\boldsymbol{s}_0$ and time point t. This conditional simulation can be carried out by the following three steps:

(a) Generate synthetic observations using $\boldsymbol{y}_t^*(\boldsymbol{s})=\boldsymbol{x}_t(\boldsymbol{s})\widehat{\boldsymbol{\beta}}+\boldsymbol{\Phi}_t(\boldsymbol{s})\widehat{\boldsymbol{\alpha}}+\boldsymbol{h}_{\boldsymbol{s}}^T\boldsymbol{v}_t^*+\varepsilon_t^*(\boldsymbol{s})$, where $\boldsymbol{v}_t^*$ are realizations of the space-time process $\boldsymbol{v}_t$ using the dynamic model (5), and $\varepsilon_t^*(\boldsymbol{s})\sim\mathcal{N}(0,\widehat{\sigma}^2)$.
(b) Estimate $\widehat{\boldsymbol{\Theta}}^*$ and $\widehat{\boldsymbol{v}}_t^*$ based on the synthetic observations $\boldsymbol{y}_t^*$.
(c) Compute $\widetilde{\boldsymbol{\Theta}}=\widehat{\boldsymbol{\Theta}}+\big(\widehat{\boldsymbol{\Theta}}-\widehat{\boldsymbol{\Theta}}^*\big)$ and $\widetilde{\boldsymbol{v}}_t=\widehat{\boldsymbol{v}}_t+\big(\boldsymbol{v}_t^*-\widehat{\boldsymbol{v}}_t^*\big)$, further generate the space-time predictive sample at location $\boldsymbol{s}_0$ and time point t by

$$y_t(\boldsymbol{s}_0)=\boldsymbol{x}_t(\boldsymbol{s}_0)\widetilde{\boldsymbol{\beta}}+\boldsymbol{\Phi}_t(\boldsymbol{s}_0)\widetilde{\boldsymbol{\alpha}}+\boldsymbol{h}_{\boldsymbol{s}_0}^T\widetilde{\boldsymbol{v}}_t+\widetilde{\varepsilon}_t(\boldsymbol{s}_0)\text{ with }\widetilde{\varepsilon}_t(\boldsymbol{s}_0)\sim\mathcal{N}(0,\widehat{\sigma}^2). \tag{12}$$

4.5. Bias correction specification

For a specific CMAQ grid cell, say C_0, we uniformly sample 50 spatial points in C_0 and denote them as $\{\mathbf{s}_0^{i'}\}_{i'=1}^{50}$. Then we repeat the conditional simulation N_c times and obtain N_c predictive samples for each $\mathbf{s}_0^{i'}$, that is

$$y_t^{(c)}\big(\mathbf{s}_0^{i'}\big) = \mathbf{x}_t\big(\mathbf{s}_0^{i'}\big)\widetilde{\boldsymbol{\beta}}^{(c)} + \boldsymbol{\Phi}_t\big(\mathbf{s}_0^{i'}\big)\widetilde{\boldsymbol{\alpha}}^{(c)} + \mathbf{h}^T_{\mathbf{s}_0^{i'}}\widetilde{\mathbf{v}}_t^{(c)} + \varepsilon_t^{(c)}\big(\mathbf{s}_0^{i'}\big), \tag{13}$$

where $c = 1, \cdots, N_c$. In this work, $N_c = 100$.

We denote the correction of CMAQ grid cell C_0 as $\tilde{y}_t(C_0)$. Then, the expectation and variance of $\tilde{y}_t(C_0)$ can be written as follows:

$$\mathrm{E}\big\{\tilde{y}_t(C_0)\big\} = \frac{1}{50N_c}\sum_{c=1}^{N_c}\sum_{i'=1}^{50} y_t^{(c)}\big(\mathbf{s}_0^{i'}\big),$$

$$\mathrm{Var}\big\{\tilde{y}_t(C_0)\big\} = \frac{1}{50N_c - 1}\sum_{c=1}^{N_c}\sum_{i'=1}^{50}\Big[y_t^{(c)}\big(\mathbf{s}_0^{i'}\big) - \mathrm{E}\big\{\tilde{y}_t(C_0)\big\}\Big]^2.$$

5. Simulation study

To investigate the capability of the ADCM to capture nonlinear relationships and spatio-temporal patterns, we perform a simulation study with data generated from a model with a nonseparable spatio-temporal covariance of the Gneiting class (Gneiting, 2002).

5.1. Simulation design

We set the spatial domain $\mathcal{D}'$ to be the BTH region with 2,499 equally spaced grids, which are located at the centroids of the regular CMAQ grids as shown in Fig. 6 (a) and denoted as $C_{i''}$ for $i'' = 1, \cdots, 2499$. The time domain $\mathcal{T}'$ is set to be $[0, 1]$ with 30 equally spaced time points. For any $\mathbf{s} \in \mathcal{D}'$ at time point $t \in \mathcal{T}'$, we use the following model to generate data,

$$y_t(\mathbf{s}) = \beta_0 + \beta_1 x_{1,t}(\mathbf{s}) + \beta_2 x_{2,t}(\mathbf{s}) + \tilde{g}_0(t) + \tilde{g}_1\big(z_t(\mathbf{s})\big) + w_t(\mathbf{s}) + \varepsilon_t(\mathbf{s}), \tag{14}$$

where the covariate $x_{1,t}(\mathbf{s})$ is generated from the standard normal distribution, $\mathbf{x}_{2,t}(\mathbf{s})$ is from a spatial Gaussian random field with mean $\mathbf{0}$ and an isotropic exponential covariance kernel, i.e., for location $\mathbf{s}$ and $\mathbf{s}'$, $\mathrm{Cov}\big(x_{2,t}(\mathbf{s}), x_{2,t}(\mathbf{s}')\big) = \exp\big(-10\|\mathbf{s} - \mathbf{s}'\|/d^{\mathbf{s}}_{\max}\big)$ with $d^{\mathbf{s}}_{\max} = \max_{\mathbf{s},\mathbf{s}'}\big\{\|\mathbf{s} - \mathbf{s}'\|\big\}$, $\tilde{g}_0(\cdot)$ and $\tilde{g}_1(\cdot)$ are two smooth functions, $z_t(\mathbf{s})$ is sampled from a uniform distribution with the interval $[0, 1]$, the spatio-temporal random process $\mathbf{w}_t(\mathbf{s})$ is mutually and serially independent of the error term $\boldsymbol{\varepsilon}_t(\mathbf{s})$, and the error term is set to $\varepsilon_t(\mathbf{s}) \sim N(0, \sigma^2)$. The random process $\mathbf{w}_t(\mathbf{s})$ follows a zero-mean stationary space-time Gaussian process with a kernel specified in the Gneiting class (Gneiting, 2002). For any $(\mathbf{s}, t)$ and $(\mathbf{s}', t')$, the covariance between $\mathbf{w}_t(\mathbf{s})$ and $\mathbf{w}_{t'}(\mathbf{s}')$ is given by

$$\mathrm{cov}\big(\mathbf{w}_t(\mathbf{s}), \mathbf{w}_{t'}(\mathbf{s}')\big) = \sigma_w^2\big(1 + \Psi_t(\tau)\big)^{-\delta/2}\Psi_s\Big(u/\sqrt{\Psi_t(\tau)}\Big), \tag{15}$$

where $\tau = |t - t'|$, $\sigma_w > 0, \delta \geq 2, \Psi_t(\tau) = \big(|\tau/\phi_t|^a + 1\big)^{-b/a}$ for $\phi_t > 0, a \in (0, 2]$ and $b > 0$, $u = \|\mathbf{s} - \mathbf{s}'\|, \Psi_s(u) = 2^{1-\nu}\Gamma^{-1}(\nu)\big(\sqrt{2\nu}u/\phi_s\big)^{\nu}K_\nu\big(\sqrt{2\nu}u/\phi_s\big)$ for $\phi_s > 0, \nu > 0$, $\Gamma(\cdot)$ is the Gamma function, and K_ν is the modified Bessel function of the second kind.

In (14), we set the regression coefficients to $\boldsymbol{\beta} = (15, 1, 1)$, the smooth function $\tilde{g}_0(t) = \tilde{g}_1(1 - t)$ with $\tilde{g}_1(z) = 2^5 z^{11}(1 - z)^6 + 10^4 z^3(1 - z)^{10}$, and the variance of $\varepsilon_t(s)$ is set to $\sigma^2 = 0.1$. The same smooth functions are used in the simulation study of generalized additive mixed models (Wood, 2017). For (15), we set $\sigma_w^2 = 1$, $\delta = 2, a = 1, b = 1.5$, $\phi_t = 0.2$ for Ψ_t, and $\nu = 3$, $\phi_s = 100$ for Ψ_s.

The simulation consists of the following steps:

Step 1: Sample $w_t(\mathbf{s})$ using (15), where $\mathbf{s} \in \mathcal{S} = \big\{C_1, \cdots, C_{2499}\big\}$, and $t = 0, 1/29, \cdots, 1$.

Step 2: Generate synthetic data $y_t(s)$ using (14).

Step 3: Sample n spatial points from $\mathcal{S}$ and denote as $\mathcal{S}^{(1)} = \big\{\mathbf{s}_1', \cdots, \mathbf{s}_n'\big\}$; see Fig. 6 (b) for an example of randomly sampled locations when $n = 100$. Let $\mathcal{S}^{(2)} = \mathcal{S}\backslash\mathcal{S}^{(1)}$. The training set is from $\mathcal{S}^{(1)}$ while the test set is from $\mathcal{S}^{(2)}$.

Step 4: Fit the ADCM using the training set through Algorithms 1 and 2, and generate predictions of the test data $y_t(s)$ in $\mathcal{S}^{(2)}$, denoted by $\widehat{y}_t(s)$.

Step 5: Repeat Steps $2 - 4$ to assess the estimates of the parameters (e.g., $\boldsymbol{\beta}$), the functions (i.e., $\tilde{g}_0$ and $\tilde{g}_1$), the predictions $\widehat{y}_t(s)$, and the recovered surface for the random process $w_t(s)$.

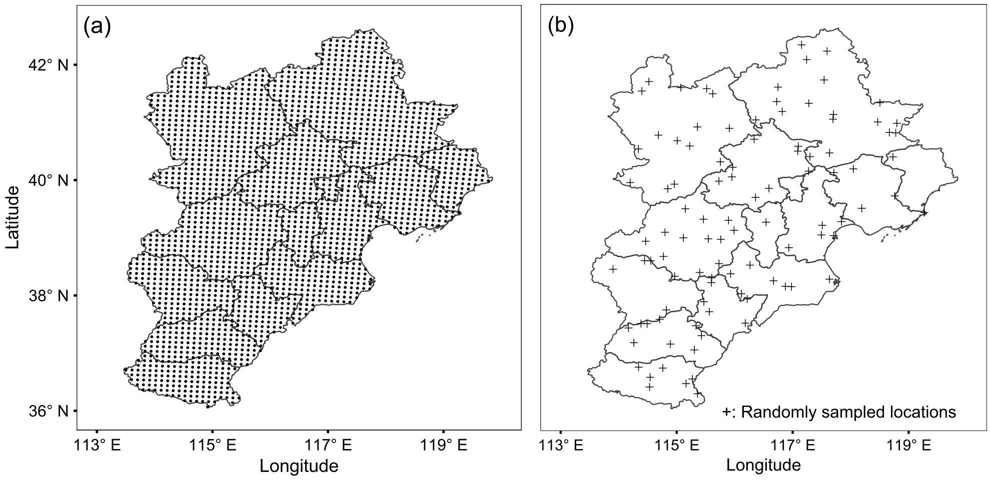

Fig. 6. Spatial sampling domain of the simulation study: (a) Centroids of 2,499 9-km CMAQ grids (dots). (b) 100 locations were randomly sampled from the BTH region (pluses).

Table 1
The biases, standard deviations (SD), and mean squared errors (MSEs) of the regression coefficients, nonlinear functions, and predictions when the training spatial sample size $n = 100$ and 2,000. The results are based on 200 independent simulations and obtained using two models: (1) the additive model (ADM) and (2) the proposed additive dynamic correction model (ADCM). The smallest MSE is in bold.

n	Model	Criteria	$\beta_0(=15)$	$\beta_1(=1)$	$\beta_2(=1)$	$\tilde{g}_0(t)$	$\tilde{g}_1(z)$	Prediction
100	ADM	bias	-0.00135	-0.00191	0.00003	0.11623	0.00033	-
		SD	0.07311	0.01231	0.02640	0.05521	0.06563	-
		MSE	0.00535	0.00016	0.00070	0.11937	0.00465	0.45723
	ADCM	Bias	-0.01229	-0.00026	0.00089	0.07226	0.00008	
		SD	0.06590	0.00762	0.01274	0.05537	0.05609	
		MSE	**0.00449**	**0.00006**	**0.00016**	**0.07545**	**0.00323**	**0.16814**
2,000	ADM	bias	0.00048	-0.00044	0.00246	0.15335	0.00002	-
		SD	0.00776	0.00278	0.02459	0.05028	0.01600	-
		MSE	**0.00006**	0.00001	0.00061	0.15642	0.00028	0.42697
	ADCM	Bias	-0.00056	-0.00017	0.00062	0.07427	0.00001	
		SD	0.01680	0.00139	0.00963	0.00725	0.01224	
		MSE	0.00028	**0.00000**	**0.00009**	**0.07432**	**0.00016**	**0.10709**

We choose the additive model (ADM) for comparison because it can capture nonlinear relationships and fit spatio-temporal processes through basis smoothing procedures (Wood, 2017). The ADM is similar to the ADCM except that the ADM's $w_t(\boldsymbol{s})$ in (14) is represented by a 3-dimensional smooth function, say g_2, and approximated by the thin plate spline smooth function (Wood, 2017). The ADM is formulated to $y_t(\boldsymbol{s}) = \beta_0 + \beta_1 x_{1,t}(\boldsymbol{s}) + \beta_2 x_{2,t}(\boldsymbol{s}) + \tilde{g}_0(t) + \tilde{g}_1\big(z_t(\boldsymbol{s})\big) + \tilde{g}_2(\boldsymbol{s}, t) + \varepsilon_t(\boldsymbol{s})$ and implemented through the `R` package `mgcv` (Wood, 2022).

The two smooth functions $\tilde{g}_0$ and $\tilde{g}_1$ are modeled by cyclic cubic-spline basis functions with equidistant knots and second-order penalties. The numbers of basis functions are set to 10 and 50 for $\tilde{g}_0$ and $\tilde{g}_1$, respectively. In step 5, the parameter estimates are evaluated using bias, standard deviation (SD), and mean squared error (MSE). For the time-invariant parameters, the definitions of these criteria are standard. For a smooth function, say $f(t)$, $\text{bias}_f = \int \big\{E(\widehat{f}) - f\big\}^2 dt$, $\text{SD}_f = \int \sqrt{E\big\{\widehat{f} - E(\widehat{f})\big\}^2} dt$ and $\text{MSE}_f = \int E\big(\widehat{f} - f\big)^2 dt$. To investigate the effect of the spatial sample size of the training set, we consider $n = 100$ and $n = 2{,}000$. Under each setting, we repeat the procedure 200 times and report the results in Table 1.

5.2. Simulation results

The results displayed in Table 1 reveal that the ADM and ADCM have minimal biases, standard deviations, and mean squared errors for the estimates of regression coefficients and nonlinear functions. Compared to the ADM, the ADCM in almost all cases yields more efficient estimates with lower standard deviations and produces even smaller MSEs for the estimates and the predictions. Based on the ADCM, the estimates of nonlinear functions match very well with the true functions when $n = 100$ as displayed in Fig. 7. The same results hold for $n = 2{,}000$.

In addition, our proposed basis approximation for $w_t(\boldsymbol{s})$ in (4) successfully captures most of the spatio-temporal patterns in the true spatio-temporal process. As an illustration, Fig. 8 demonstrates the excellent performance of the ADCM in

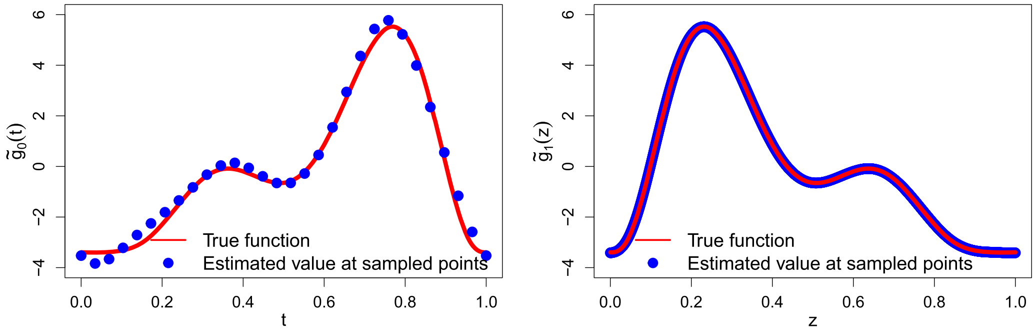

Fig. 7. The true and estimated nonlinear functions $g_0(t)$ (left) and $g_1(z)$ (right) when the spatial sample size $n = 100$.

recovering the true $w_t(s)$ from a simulated diffusion process for $n = 100$ and 2,000. As the training sample size increases, the predicted maps resemble the simulated map in the first row more. Even when $n = 100$, the ADCM does a remarkable job of characterizing the patterns of the random process $w_t(s)$. Moreover, the averaged root mean squared errors of the predictions for $w_t(s)$ for $n = 100$ and 2,000 are as low as 0.3672 and 0.2756, respectively, implying the validity of our proposed basis approximation of $w_t(\boldsymbol{s})$ in (4).

6. Data analysis

6.1. *Cross validation*

In this section, we evaluate the prediction performance of the ADCM using a regional cross-validation method called leave-one-city-out cross-validation (LOCOCV). Each time, observations from one of the thirteen cities are set aside for model testing and data from the remaining twelve cities are used to fit the model.

In addition to the ADM, we compare the proposed ADCM with three spatial or spatio-temporal models, including a spatial downscaler model with spatially varying coefficients (SVC, Gelfand et al. (2003)), a Bayesian spatio-temporal downscaler model with spatiotemporally varying coefficients (STVC, see Section 3.2 of Berrocal et al. (2010b)), and a Bayesian spatio-temporal model that allows for first-order spatiotemporal autoregression (STAR, see Section 4.2 of Blangiardo et al. (2013) or Section 7.2 of Blangiardo and Cameletti (2015)). The SVC models have been widely used in modeling air pollution in recent years and their applications in the fusion of data of different spatial support have been rather prevalent (Bass and Sahu, 2019; Jiang and Enki Yoo, 2019; Berrocal et al., 2020). In contrast to the SVC models, the STVC model allows the intercept and slope to vary with time, and the STAR model allows random processes to have a time-dynamic structure. In this work, the SVC model is implemented through maximum likelihood estimation methods (Dambon et al., 2021a) using the `R` package `varycoef` (Dambon et al., 2021b). For the spatiotemporal SVCs, the Gaussian process involved in the STVC or STAR is first linked to the GMRF (Rue and Held, 2005)) via a stochastic partial differential equation (SPDE, Lindgren et al. (2011)), and then two models can be estimated via the integrated nested Laplace approximation (INLA, Rue et al. (2009)) that is available through the `R` package `INLA` (Lindgren and Rue, 2015). In the recent 10 year, INLA has been shown as a powerful tool in dealing with large datasets in Bayesian analysis (Rue et al., 2017). Additional details of these competing models are provided in the Supplementary Material.

For model comparison, two commonly accepted criteria are adopted: the root mean squared error (RMSE) and the continuous rank probability score (CRPS, Gneiting and Raftery (2007); Jordan et al. (2019)). The CRPS is defined as CRPS $= \frac{1}{n_0 N_t}\sum_{t=1}^{N_t}\sum_{i=1}^{n_0} \mathrm{crps}\big(F_t(\boldsymbol{s}_i), y_t(\boldsymbol{s}_i)\big)$ with $\mathrm{crps}\big(F, y\big) = \sigma_y\left[\frac{1}{\sqrt{\pi}} - 2f\left(\frac{y-\mu_y}{\sigma_y}\right) - \frac{y-\mu_y}{\sigma_y}\left\{2\Phi\left(\frac{y-\mu_y}{\sigma_y}\right) - 1\right\}\right]$ for a normal predictive distribution F with mean μ_y and variance σ_y^2, where $f(\cdot)$ and $\Phi(\cdot)$ represent the probability density function and the cumulative distribution function of a standard normal random variable, respectively. In summary, a smaller RMSE or CRPS indicates more accurate, precise, or sharp predictions.

The summary results displayed in Table 2 show that the proposed ADCM beats all five competing models in terms of the average RMSE and CRPS, as seen in the last row of the table. Considering all cities individually, ADCM has the lowest RMSE for 8 cities, the second lowest RMSE for 2 cities, the lowest CRPS for 5 cities, and the second lowest CRPS for another 3 cities. The CMAQ outputs have the worst performance among all models suggesting an urgent need for correction of the numerical model outputs. Our ADCM improves the CMAQ outputs by 43.70% and 34.76% in the average RMSE and CRPS, respectively. Compared to smoothing-based approaches such as the ADM, the ADCM improves the performance by 14.04% in average RMSE and 10.07% in average CRPS. In the BTH region, the STVC or STAR generally predicts better than the SVC, implying that both the spatial and temporal structures are critical for improving the CMAQ outputs. The ADCM further improves the performance of the STVC or STAR by at least 14.09% in average RMSE and 11.55% in average CRPS, and the performance of the SVC by 28.30% and 23.66%, respectively. It is worth noting that the predictions at Zhangjiakou using SVC, STVC, or STAR are worse than the CMAQ outputs. This is because the $PM_{2.5}$ concentrations in Zhangjiakou were significantly lower than in its surrounding cities, and the absence of spatial interactions in the three models led to their poor predictions.

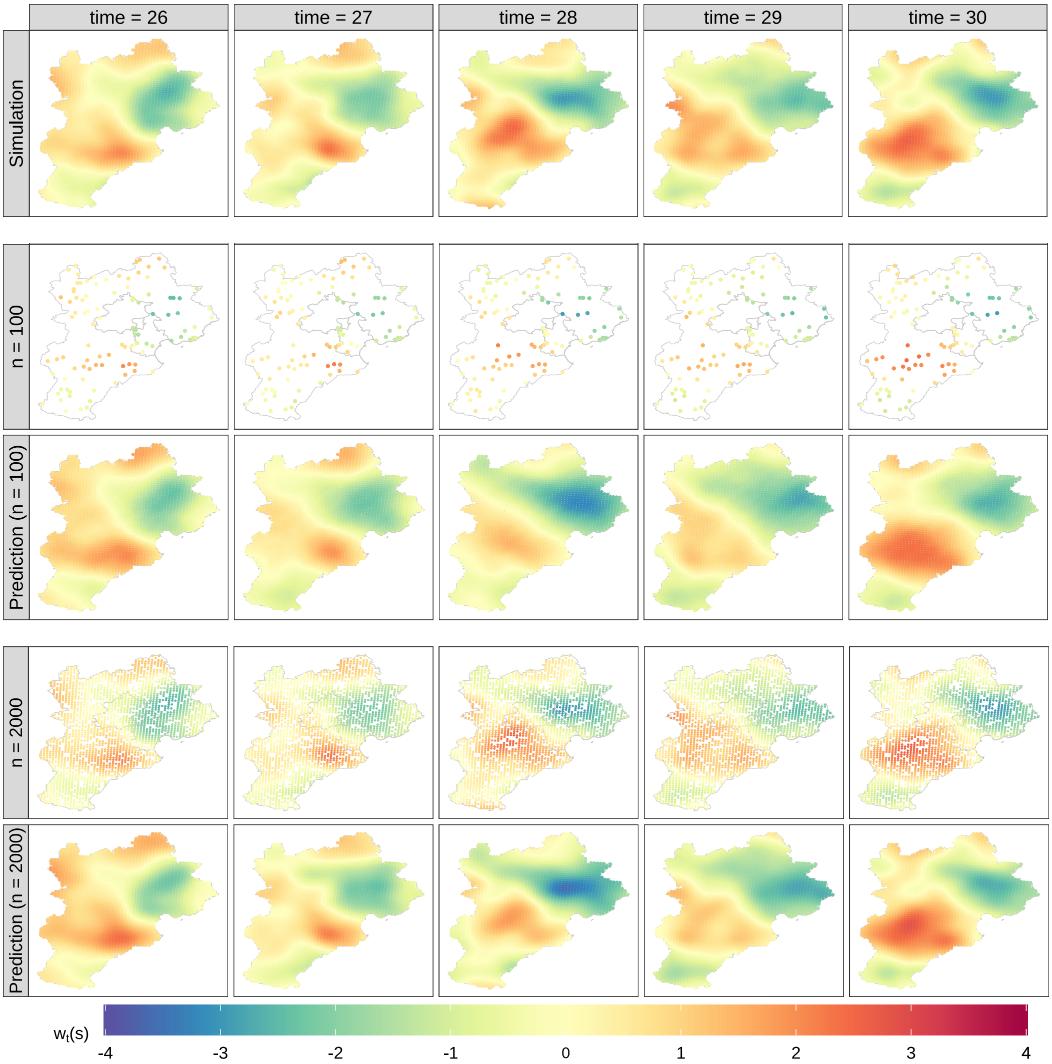

Fig. 8. Maps of the random process $w_t(s)$ in the Gneiting space-time covariance model. From top to bottom: simulated $w_t(s)$ map, map of 100 sampling locations, predicted $w_t(s)$ map based on the 100 locations, map of 2,000 sampling locations, and predicted $w_t(s)$ map based on the 2,000 locations. The time horizon is from time $= 26$ to time $= 30$.

This indicates that spatial interactions of the residuals from the model $y_t(\mathbf{s}) \sim \beta_0 + \beta_1 x_t(C_{\mathbf{s}}) + \sum_{k=0}^{4} \sum_{j=0}^{J_k} \alpha_{k,j} \phi_{k,j}\big(z_{t,k}(C_{\mathbf{s}})\big)$ are critical for improving the CMAQ outputs. Downscaler models with a single predictor (i.e., the CMAQ outputs) are often used for real-world applications (Berrocal et al., 2010b, 2020), we thus considered weighted CMAQ outputs as the single predictor for all models. These cross-validation results are presented in Table S1 of the Supplementary Material. Compared to other models, the ADCM with the single predictor improves the performance by at least 7.82% in average RMSE and 7.13% in average CRPS.

To further assess the prediction performance of the ADCM, we take Langfang as an example and present the ADCM predictions, the CMAQ model outputs, and the actual observations at Langfang in Fig. 9. Compared to the CMAQ outputs, the mean predictions of the ADCM are much closer to the true observations with almost all observations covered by the 95% confidence intervals of the ADCM. This visual comparison and the aforementioned numerical metrics provide substantial evidence for the effectiveness of the ADCM in correcting the CMAQ outputs in the BTH region.

6.2. *Correcting the raw CMAQ $PM_{2.5}$ outputs in the BTH region*

In this section, we use the ADCM to fit all $PM_{2.5}$ concentration data of the thirteen BTH cities and then correct the raw CMAQ outputs of $PM_{2.5}$ concentrations in the BTH region. Particularly, in the model fitting stage, we obtain the trend estimate with 95% confidence intervals over time; see Fig. 10 (a). This trend peaked at the end of December 2015, mostly due to coal-burning for heating in winter, leading to an increase in the emission of fine particulate matters such as $PM_{2.5}$ (Liang et al., 2016). The nonlinear relationships between the $PM_{2.5}$ concentrations and four meteorological variables are also

Table 2
Averaged RMSE and CRPS of $PM_{2.5}$ concentration predictions ($\mu g/m^3$) at 13 cities in the BTH region using LOCOCV from five models: (1) CMAQ model output, (2) spatially-varying coefficient downscaling model (SVC), (3) spatiotemporally-varying coefficient downscaling model (STVC), (4) first-order spatiotemporal autoregression (STAR), (5) additive model (ADM), and (6) proposed additive dynamic correction model (ADCM). The smallest RMSE and CRPS are in bold, and the second smallest ones are underlined. Daily data from November 1, 2015 to January 31, 2016 are considered.

City	RMSE						CRPS					
	CMAQ	SVC	STVC	STAR	ADM	ADCM	CMAQ	SVC	STVC	STAR	ADM	ADCM
Baoding	94.64	75.02	76.76	75.84	75.28	**73.96**	50.35	40.66	41.12	40.83	41.34	**39.65**
Beijing	92.52	62.42	58.75	62.50	67.48	**54.23**	45.82	32.89	**30.37**	30.97	35.88	30.83
Cangzhou	60.89	45.18	**31.90**	33.90	45.81	34.91	31.78	29.42	23.16	23.98	29.87	**22.80**
Chengde	41.62	62.39	25.23	**24.34**	30.28	26.83	22.66	30.73	15.55	**15.05**	16.69	15.29
Handan	66.09	57.61	53.94	57.04	54.60	**48.11**	37.58	36.33	34.99	36.30	34.26	**31.62**
Hengshui	91.19	60.31	57.32	57.16	59.19	**56.11**	46.21	33.01	**31.15**	31.26	33.69	31.82
Langfang	85.44	40.75	36.31	38.63	45.28	**36.21**	46.17	27.66	27.18	**26.37**	27.94	27.75
Qinhuangdao	46.19	80.10	36.60	45.96	30.44	**23.60**	24.90	42.81	21.16	26.37	17.10	**14.42**
Shijiazhuang	87.72	53.40	47.92	**46.25**	62.07	53.07	48.18	33.20	30.92	**30.17**	38.85	34.47
Tangshan	95.75	33.19	35.85	**28.42**	34.24	32.94	52.20	23.04	21.02	**19.60**	22.35	20.62
Tianjin	70.57	36.11	**29.93**	33.43	42.40	32.84	38.00	25.56	**23.49**	24.19	27.81	25.01
Xingtai	85.69	58.37	54.25	53.63	54.52	**52.98**	45.46	33.93	31.70	**30.89**	31.96	31.86
Zhangjiakou	64.42	106.78	102.09	86.92	42.00	**27.46**	33.94	58.04	65.12	50.00	21.82	**15.20**
Average	75.60	59.36	49.76	49.54	49.51	**42.56**	40.25	34.40	30.53	29.69	29.20	**26.26**

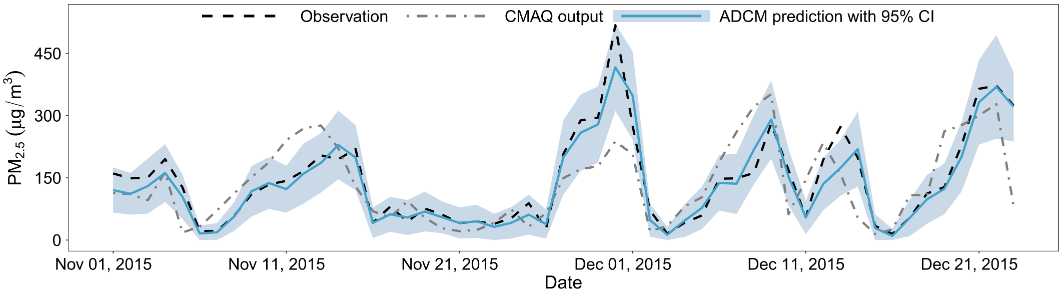

Fig. 9. Time series plot of observed $PM_{2.5}$ concentrations, the pre-correction CMAQ outputs, and the predicted $PM_{2.5}$ concentrations with conditionally simulated 95% confidence intervals (CI) obtained using the proposed ADCM model at Langfang.

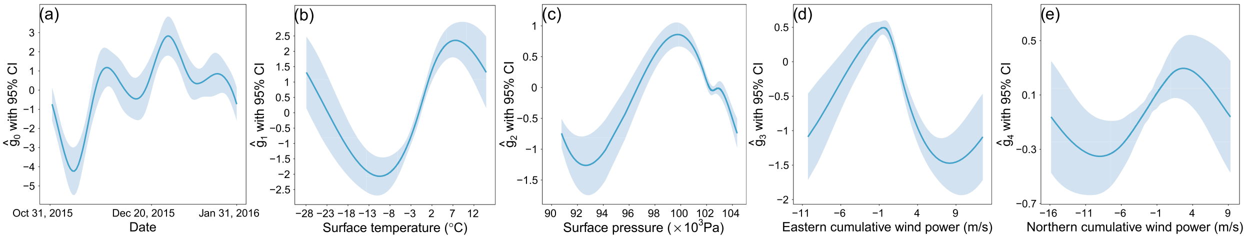

Fig. 10. Estimated nonlinear functions $\hat{g}(\cdot)$ with conditionally simulated 95% confidence intervals (CI). Figures (a)-(e) represent the estimates of functions for time, surface temperature, surface pressure, and eastern and northern cumulative wind powers, respectively.

estimated and presented in Fig. 10 (b)-(e) in turn. These results show that $PM_{2.5}$ readings from monitoring stations are nonlinearly affected by different weather conditions, consistent with the findings in Liang et al. (2015).

We follow the procedure described in Section 4.5 to carry out the bias correction. First, a sample of 50 locations in each CMAQ grid cell is randomly generated, resulting in a total of 124,950 spatial locations in the BTH region. We then apply the fitted ADCM to correct the space-time outputs of the CMAQ model for the entire BTH region. A severe air pollution episode in the BTH region from November 26, 2015 to December 1, 2015 is used to demonstrate the correction results. In Fig. 11, the average $PM_{2.5}$ concentration in each of the thirteen cities is marked by tiny solid squares. In general, a smooth transition from the cities (solid squares) to the rural areas implies an excellent correction. It is noticeable that the CMAQ outputs fail to match well with most of the average $PM_{2.5}$ concentration data in the cities. The ADCM corrections, on the other hand, produce a smooth transition from the cities to their surrounding areas, demonstrating the significant improvement made by the proposed method on the CMAQ outputs.

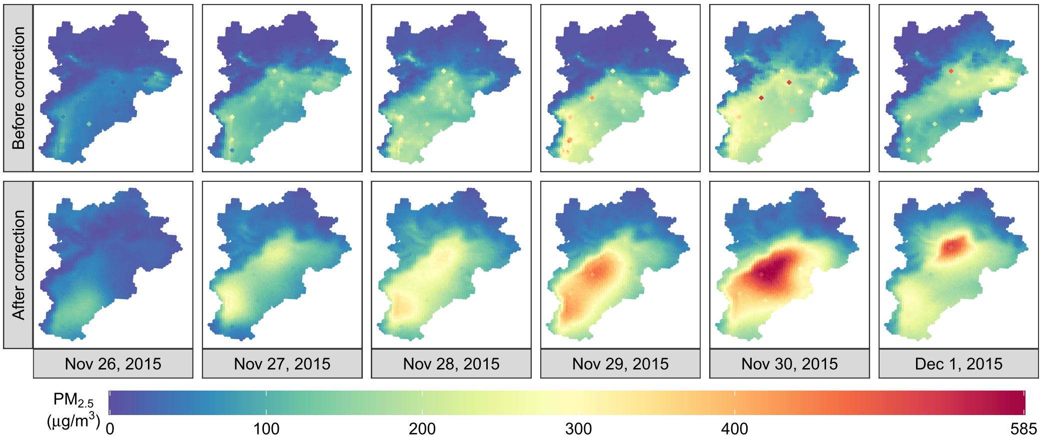

Fig. 11. CMAQ numerical model outputs before and after the ADCM correction from November 26 to December 1, 2015. The solid squares represent the average $PM_{2.5}$ levels at the monitoring stations.

6.3. *Computational efficiency*

To assess the computational efficiency of our method, we compare the running time of the ADCM with the benchmark SVC based on the MLE and STAR based on the INLA. Model fitting and corrections from two methods were run on a 10-core workstation with 128 GB of memory and a 3.8 GHz Intel Core i9-10900KF processor. In the model fitting stage, although the proposed ADCM takes more time than the SVC for a small number of spatial observations (i.e., n), it is much more advantageous when n is large because its computational complexity is not heavily affected by the size of the data (e.g., n), as discussed in Section 4.2.2. For example, when $n = 1{,}000$ and $N_t = 30$, the ADCM is almost 20 times faster than the SVC and 10 times faster than the STAR. Please refer to Table S2 of the Supplementary Material for other examples. In the correction stage, the ADCM is also more computationally efficient than the SVC. From a computational point of view, the flops operations of the ADCM using (12) mainly depend on the term $\boldsymbol{H}\boldsymbol{v}_t$. At an unobservable spatial point, the flops operations for producing a single predictive sample are no more than $O(c_h^2 N_\nu N_t)$ with $c_h \in (0, 1]$ and $N_\nu = 788$ for the BTH data, whereas the SVC requires $O(n^2 N_t)$ flops operations for a single predictive sample in the kriging (Chen and Stein, 2021). When correcting the raw CMAQ outputs in the BTH region, it took the SVC 457 seconds to generate a corrected sample of 124,950 spatial points, but only 169 seconds for the ADCM to do the same. Overall, the ADCM is a much more computationally efficient method than the SVC for correcting the CMAQ outputs in the BTH region. The source code of the ADCM is mostly written in the `R` language, thus its computational efficiency could be further improved through a low-level language such as `C`.

7. Discussion

In this work, we develop a novel method to correct large-scale numerical model raw outputs. The users can correct model outputs using datasets from multiple sources with different spatial supports and explore different small-scale variations of the bias field of the model outputs. The entire methodology can be completed quickly because the estimation, implementation, and correction are all carried out efficiently through well-designed procedures including the use of sparse matrices, efficient algorithms, and others. The simulation study presented in Section 5.1 illustrates that our model successfully captured nonlinear relationships and complex spatio-temporal patterns that are frequently observed in real-life data, such as a process with a nonseparable spatio-temporal covariance of the Gneiting class (Gneiting, 2002). The application study using the BTH $PM_{2.5}$ concentration data demonstrates that, compared to other popular methods, the proposed approach is not only more effective and computationally efficient in correcting the raw CMAQ model outputs, but also able to capture nonlinear relationships between the response and other variables, and model the space-time interactions in the air pollution data. The proposed ADCM is also robust for the different specifications of precision matrices for the IDE (5), including the CAR, Leroux model, and SAR, etc. as shown in Table S3 of the Supplementary Material.

Our study is subject to some limitations. Although the normality assumptions of the error and the innovation in (6) can greatly simplify the computations, this assumption may not be realistic and appropriate for air pollution data. While the dynamic approach provides a flexible framework to describe space-time interactions by specifying different redistribution kernels such as the Gaussian kernel or radial basis function, the selection of the kernel function is important for the approach and often largely depends on our understanding of data. Moreover, Theorem 1 of De Bézenac et al. (2019) showed the solution of the Gaussian-based IDE method corresponds to that of an advection-diffusion equation. However, it is currently unknown whether the solution of the Wendland-based IDE method also corresponds to the solution of a specific differential equation. Further investigation is required in future research to determine the precise differential equation associated with the Wendland-based IDE.

This approach can be conveniently extended to other versions and offers flexibility to meet a wide range of needs. For example, to model non-Gaussian data (Bradley et al., 2019) or data with even more elaborate relationships among the variables than that in this work, we can apply nonlinear operators of dynamic modeling that are induced by the laws of physics such as the Lorenz-96 evolution model (Lorenz, 1996) or employ fast sampling algorithms in a Bayesian framework (Riabiz et al., 2022). This is worth further investigation.

Acknowledgements

This research is partially supported by the National Natural Science Foundation of China (Grant No. 12231017, 12292984), and the MOE Project of Key Research Institute of Humanities and Social Sciences (Grant No. 22JJD910001).

Appendix A. M-step of EM algorithm of the ADCM

- Update $\boldsymbol{\alpha}$ to $\widehat{\boldsymbol{\alpha}}^{(\ell)}$ using the formula at the ℓth iteration:

$$\widehat{\boldsymbol{\alpha}}^{(\ell)} = \left(\sum_{t=1}^{N_t} \boldsymbol{\Phi}_t^T \boldsymbol{\Phi}_t + \boldsymbol{\Omega}_{t,\lambda}\right)^{-1} \sum_{t=1}^{N_t} \boldsymbol{\Phi}_t^T \left(\boldsymbol{y}_t - \boldsymbol{X}_t \widehat{\boldsymbol{\beta}}^{(\ell-1)} - \boldsymbol{H}\widehat{\boldsymbol{v}}_{t|N_t}^{(\ell)}\right),$$

  where $\boldsymbol{\Omega}_{t,\lambda}$ is a penalty matrix and depends on the tuning parameters, $\lambda_0, \cdots, \lambda_4$, which can be obtained by generalized cross validation.
- Update $\boldsymbol{\beta}$ to $\widehat{\boldsymbol{\beta}}^{(\ell)}$ using the formula:

$$\widehat{\boldsymbol{\beta}}^{(\ell)} = \left(\sum_{t=1}^{N_t} \boldsymbol{X}_t^T \boldsymbol{X}_t\right)^{-1} \sum_{t=1}^{N_t} \boldsymbol{X}_t^T \left(\boldsymbol{y}_t - \boldsymbol{\Phi}_t \widehat{\boldsymbol{\alpha}}^{(\ell)} - \boldsymbol{H}\widehat{\boldsymbol{v}}_{t|N_t}^{(\ell)}\right).$$

- Update σ^2 to $\widehat{\sigma}^{2,(\ell)}$ using the formula:

$$\widehat{\sigma}^{2,(\ell)} = \frac{1}{nN_t}\left[\sum_{t=1}^{N_t}\left(\widehat{\boldsymbol{\varepsilon}}_t^T \widehat{\boldsymbol{\varepsilon}}_t - 2\widehat{\boldsymbol{\varepsilon}}_t^T \boldsymbol{H}\widehat{\boldsymbol{v}}_{t|N_t}^{(\ell)}\right) + \mathtt{Tr}\left(\boldsymbol{H}^T \boldsymbol{H} \boldsymbol{S}_{11}^{(\ell)}\right)\right],$$

  where $\mathtt{Tr}(\boldsymbol{A})$ denotes the trace of the matrix $\boldsymbol{A}$.
- Update $\widehat{\theta}_1^{(r)}$ to $\widehat{\theta}_1^{(r,\ell)}$ using the formula:

$$\widehat{\theta}_1^{(r,\ell)} = \mathtt{Tr}\left(\boldsymbol{Q}^{(r,\ell)} \boldsymbol{\mathcal{M}}_{\theta_2^{(r)}}^{(\ell)} \boldsymbol{S}_{01}^{(r,\ell)}\right)\left\{\mathtt{Tr}\left(\boldsymbol{\mathcal{M}}_{\theta_2}^{(\ell)} \boldsymbol{Q}^{(r,\ell)} \boldsymbol{\mathcal{M}}_{\theta_2^{(r)}}^{(\ell)} \boldsymbol{S}_{00}^{(r,\ell)}\right)\right\}^{-1}.$$

- Update $\tau^{2,(r)}$ to $\widehat{\tau}^{2,(r,\ell)}$ using the formula:

$$\widehat{\tau}^{2,(r,\ell)} = N_r N_t \left\{\mathtt{Tr}\left(\boldsymbol{\Lambda}^{(r,\ell)} \boldsymbol{S}_{11}^{(r,\ell)} - 2\widehat{\theta}_1^{(r,\ell)} \boldsymbol{\Lambda}^{(r,\ell)} \boldsymbol{\mathcal{M}}_{\theta_2^{(r)}}^{(\ell)} \boldsymbol{S}_{01}^{(r,\ell)} + \widehat{\theta}_1^{2,(\ell)} \boldsymbol{\mathcal{M}}_{\theta_2^{(r)}}^{(\ell)} \boldsymbol{\Lambda}^{(r,\ell)} \boldsymbol{\mathcal{M}}_{\theta_2^{(r)}}^{(\ell)} \boldsymbol{S}_{00}^{(r,\ell)}\right)\right\}^{-1},$$

  where $\boldsymbol{\Lambda}^{(r,\ell)} = \boldsymbol{G}^{(r)} + \zeta^{2,(r,\ell)} \boldsymbol{I}_{N_r}$.
- Update $\tau_0^{2,(r)}$ to $\tau_0^{2,(r,\ell)}$ using the formula:

$$\widehat{\tau}_0^{2,(r,\ell)} = N_r \left[\mathtt{Tr}\left\{\boldsymbol{\Lambda}_0^{(r,\ell)}\left(\widehat{\boldsymbol{v}}_{0|N_t}^{(\ell)} \widehat{\boldsymbol{v}}_{0|N_t}^{T,(r,\ell)} + \widehat{\boldsymbol{\Sigma}}_{0|N_t}^{(r,\ell)}\right)\right\}\right]^{-1},$$

  where $\boldsymbol{\Lambda}_0^{(r,\ell)} = \boldsymbol{G}^{(r)} + \zeta_0^{2,(r,\ell)} \boldsymbol{I}_m$.
- Find $\widehat{\theta}_2^{(r,\ell)}$, $\widehat{\zeta}^{(r,\ell)}$, and $\widehat{\zeta}_0^{(r,\ell)}$ by maximizing the following three formulas:

$$\exp\left\{-\frac{1}{2}\widehat{\tau}^{2,(r,\ell)} \mathtt{Tr}\left(-2\widehat{\theta}_1^{(r,\ell)} \boldsymbol{\Lambda}^{(r,\ell)} \boldsymbol{\mathcal{M}}_{\theta_2^{(r)}} \boldsymbol{S}_{01}^{(r,\ell)} + \widehat{\theta}_1^{2,(r,\ell)} \boldsymbol{\mathcal{M}}_{\theta_2^{(r)}} \boldsymbol{\Lambda}^{(r,\ell)} \boldsymbol{\mathcal{M}}_{\theta_2^{(r)}} \boldsymbol{S}_{00}^{(r,\ell)}\right)\right\},$$

$$|\boldsymbol{G}^{(r)} + \zeta^{2,(r)} \boldsymbol{I}|^{\frac{N_t}{2}} \exp\left\{-\frac{1}{2}\mathtt{Tr}\left(\boldsymbol{S}_{11}^{(r,\ell)} - 2\widehat{\theta}_1^{(r,\ell)} \boldsymbol{\mathcal{M}}_{\widehat{\theta}_2^{(r)}}^{(\ell)} \boldsymbol{S}_{01}^{(r,\ell)} + \widehat{\theta}_1^{2,(r,\ell)} \boldsymbol{\mathcal{M}}_{\widehat{\theta}_2^{(r)}}^{(\ell)} \boldsymbol{\mathcal{M}}_{\widehat{\theta}_2^{(r)}}^{(\ell)} \boldsymbol{S}_{00}^{(r,\ell)}\right)\widehat{\tau}^{2,(r,\ell)} \zeta^{2,(r)}\right\},$$

$$|\boldsymbol{G}^{(r)} + \zeta_0^{2,(r)} \boldsymbol{I}|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}\mathtt{Tr}\left(\widehat{\boldsymbol{v}}_{0|N_t}^{(r,\ell)} \widehat{\boldsymbol{v}}_{0|N_t}^{T,(r,\ell)} + \widehat{\boldsymbol{\Sigma}}_{0|N_t}^{T,(r,\ell)}\right)\widehat{\tau}_0^{2,(r,\ell)} \zeta_0^{2,(r)}\right\}.$$

  The optimization problems above can be solved using the standard numerical optimization algorithms such as the Broyden–Fletcher–Goldfarb–Shanno (BFGS). In this work, we carry out the optimization using PORT routines in `nlminb` function in `R` environment.

Appendix B. Supplementary material

The “SuppMaterial.pdf” file in Supplementary Material describes the competing models discussed in this work, provides additional comparisons with simulated data, and gives a sensitivity analysis for the ADCM for different specifications of the precision matrices of IDE models.

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.csda.2023.107799.

References

Anselin, L., 1988. Spatial Econometrics: Methods and Models. Kluwer Academic Publishers, Dordrecht.

Appel, K.W., Napelenok, S.L., Foley, K.M., Pye, H.O.T., Hogrefe, C., Luecken, D.J., Bash, J.O., Roselle, S.J., Pleim, J.E., Foroutan, H., Hutzell, W.T., Pouliot, G.A., Sarwar, G., Fahey, K.M., Gantt, B., Gilliam, R.C., Heath, N.K., Kang, D., Mathur, R., Schwede, D.B., Spero, T.L., Wong, D.C., Young, J.O., 2017. Description and evaluation of the community multiscale air quality (CMAQ) modeling system version 5.1. Geosci. Model Dev. 10, 1703–1732.

Bakar, K.S., 2012. Bayesian analysis of daily maximum ozone levels. Ph.D. thesis. University of Southampton, Southampton, UK.

Banerjee, S., Gelfand, A.E., Finley, A.O., Sang, H., 2008. Gaussian predictive process models for large spatial data sets. J. R. Stat. Soc., Ser. B, Stat. Methodol. 70, 825–848.

Bass, M.R., Sahu, S.K., 2019. Dynamically updated spatially varying parameterizations of hierarchical Bayesian models for spatial data. J. Comput. Graph. Stat. 28, 105–116.

Berrocal, V.J., Gelfand, A.E., Holland, D.M., 2010a. A bivariate space-time downscaler under space and time misalignment. Ann. Appl. Stat. 4, 1942–1975.

Berrocal, V.J., Gelfand, A.E., Holland, D.M., 2010b. A spatio-temporal downscaler for output from numerical models. J. Agric. Biol. Environ. Stat. 15, 176–197.

Berrocal, V.J., Gelfand, A.E., Holland, D.M., 2012. Space-time data fusion under error in computer model output: an application to modeling air quality. Biometrics 68, 837–848.

Berrocal, V.J., Guan, Y., Muyskens, A., Wang, H., Reich, B.J., Mulholland, J.A., Chang, H.H., 2020. A comparison of statistical and machine learning methods for creating national daily maps of ambient $PM_{2.5}$ concentration. Atmos. Environ. 222, 117–130.

Besag, J., 1974. Spatial interaction and the statistical analysis of lattice systems. J. R. Stat. Soc., Ser. B, Methodol. 36, 192–225.

Besag, J., York, J., Mollié, A., 1991. Bayesian image restoration, with two applications in spatial statistics. Ann. Inst. Stat. Math. 43, 1–20.

Bivand, R., Gómez-Rubio, V., Rue, H., 2015. Spatial data analysis with R-INLA with some extensions. J. Stat. Softw. 63, 1–31.

Blangiardo, M., Cameletti, M., 2015. Spatial and Spatio-Temporal Bayesian Models with R-INLA. John Wiley & Sons, New York.

Blangiardo, M., Cameletti, M., Baio, G., Rue, H., 2013. Spatial and spatio-temporal models with R-INLA. Spat. Spatio-Tempor. Epidemiol. 4, 33–49.

Bradley, J.R., Holan, S.H., Wikle, C.K., 2015a. Multivariate spatio-temporal models for high-dimensional areal data with application to longitudinal employer-household dynamics. Ann. Appl. Stat. 9, 1761–1791.

Bradley, J.R., Wikle, C.K., Holan, S.H., 2015b. Spatio-temporal change of support with application to American community survey multi-year period estimates. Stat 4, 255–270.

Bradley, J.R., Wikle, C.K., Holan, S.H., 2019. Spatio-temporal models for big multinomial data using the conditional multivariate logit-beta distribution. J. Time Ser. Anal. 40, 363–382.

Brynjarsdóttir, J., O'Hagan, A., 2014. Learning about physical parameters: the importance of model discrepancy. Inverse Probl. 30, 114007.

Byun, D., Schere, K.L., 2006. Review of the governing equations, computational algorithms, and other components of the Models-3 Community Multiscale Air Quality (CMAQ) modeling system. Appl. Mech. Rev. 59, 51–77.

Celeux, G., 1985. The SEM algorithm: a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem. Comput. Stat. Q. 2, 73–82.

Chaudhuri, A., Kakde, D., Sadek, C., Gonzalez, L., Kong, S., 2017. The mean and median criteria for kernel bandwidth selection for support vector data description. In: 2017 IEEE International Conference on Data Mining Workshops (ICDMW). IEEE, pp. 842–849.

Chen, J., Stein, M.L., 2021. Linear-cost covariance functions for Gaussian random fields. J. Am. Stat. Assoc., 1–18.

China's State Council, 2013. The Action Plan for Air Pollution Prevention and Control. http://www.gov.cn/zwgk/2013-09/12/content_2486773.htm (in Chinese).

China's State Council, 2018. The Three-Year Action Plan for Winning the Blue Sky Defense Battle. http://www.gov.cn/xinwen/2018-07/03/content_5303212.htm (in Chinese).

China's State Council, 2021. The Long-Term “Beautiful China” Targets Through 2035. http://www.gov.cn/xinwen/2021-02/22/content_5588304.htm (in Chinese).

Cressie, N., Johannesson, G., 2008. Fixed rank kriging for very large spatial data sets. J. R. Stat. Soc., Ser. B, Stat. Methodol. 70, 209–226.

Cressie, N., Shi, T., Kang, E.L., 2010. Fixed rank filtering for spatio-temporal data. J. Comput. Graph. Stat. 19, 724–745.

Cressie, N., Wikle, C.K., 2011. Statistics for Spatio-Temporal Data. John Wiley & Sons, Hoboken, N.J.

Dambon, J.A., Sigrist, F., Furrer, R., 2021a. Maximum likelihood estimation of spatially varying coefficient models for large data with an application to real estate price prediction. Spat. Stat. 41, 100470.

Dambon, J.A., Sigrist, F., Furrer, R., 2021b. varycoef: an R package for Gaussian process-based spatially varying coefficient models. ArXiv preprint. arXiv:2106.02364.

Datta, A., Banerjee, S., Finley, A.O., Gelfand, A.E., 2016a. Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. J. Am. Stat. Assoc. 111, 800–812.

Datta, A., Banerjee, S., Finley, A.O., Hamm, N.A., Schaap, M., 2016b. Nonseparable dynamic nearest neighbor Gaussian process models for large spatio-temporal data with an application to particulate matter analysis. Ann. Appl. Stat. 10, 1286.

De Bézenac, E., Pajot, A., Gallinari, P., 2019. Deep learning for physical processes: incorporating prior scientific knowledge. J. Stat. Mech. Theory Exp. 2019, 124009.

Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc., Ser. B, Stat. Methodol. 39, 1–22.

Dewar, M., Scerri, K., Kadirkamanathan, V., 2008. Data-driven spatio-temporal modeling using the integro-difference equation. IEEE Trans. Signal Process. 57, 83–91.

Evensen, G., Van Leeuwen, P.J., 2000. An ensemble Kalman smoother for nonlinear dynamics. Mon. Weather Rev. 128, 1852–1867.

Finley, A., Banerjee, S., Hjelle, Ø., 2017. MBA: multilevel B-spline approximation. R package version 0.0-9. https://CRAN.R-project.org/package=MBA.

Finley, A.O., Banerjee, S., Gelfand, A.E., 2012. Bayesian dynamic modeling for large space-time datasets using Gaussian predictive processes. J. Geogr. Syst. 14, 29–47.

Fuentes, M., Raftery, A.E., 2005. Model evaluation and spatial interpolation by Bayesian combination of observations with outputs from numerical models. Biometrics 61, 36–45.

Furrer, R., Genton, M.G., Nychka, D., 2006. Covariance tapering for interpolation of large spatial datasets. J. Comput. Graph. Stat. 15, 502–523.

Furrer, R., Sain, S.R., 2010. spam: a sparse matrix R package with emphasis on MCMC methods for Gaussian Markov random fields. J. Stat. Softw. 36, 1–25.
Gelfand, A.E., Kim, H.J., Sirmans, C., Banerjee, S., 2003. Spatial modeling with spatially varying coefficient processes. J. Am. Stat. Assoc. 98, 387–396.
Gneiting, T., 2002. Nonseparable, stationary covariance functions for space-time data. J. Am. Stat. Assoc. 97, 590–600.
Gneiting, T., Raftery, A.E., 2007. Strictly proper scoring rules, prediction, and estimation. J. Am. Stat. Assoc. 102, 359–378.
Gotway, C.A., Ferguson, R.B., Hergert, G.W., Peterson, T.A., 1996. Comparison of kriging and inverse-distance methods for mapping soil parameters. Soil Sci. Soc. Am. J. 60, 1237–1247.
Gräler, B., Pebesma, E., Heuvelink, G., 2016. Spatio-temporal interpolation using gstat. R J. 8, 204–218.
Guillas, S., Bao, J., Choi, Y., Wang, Y., 2008. Statistical correction and downscaling of chemical transport model ozone forecasts over Atlanta. Atmos. Environ. 42, 1338–1348.
Hefley, T.J., Hooten, M.B., Russell, R.E., Walsh, D.P., Powell, J.A., 2017. When mechanism matters: Bayesian forecasting using models of ecological diffusion. Ecol. Lett. 20, 640–650.
Heinrich, C., Hellton, K.H., Lenkoski, A., Thorarinsdottir, T.L., 2021. Multivariate postprocessing methods for high-dimensional seasonal weather forecasts. J. Am. Stat. Assoc. 116, 1048–1059.
Higdon, D., Gattiker, J., Williams, B., Rightley, M., 2008. Computer model calibration using high-dimensional output. J. Am. Stat. Assoc. 103, 570–583.
Hooten, M.B., Wikle, C.K., 2008. A hierarchical Bayesian non-linear spatio-temporal model for the spread of invasive species with application to the Eurasian Collared-Dove. Environ. Ecol. Stat. 15, 59–70.
Houtekamer, P.L., Zhang, F., 2016. Review of the ensemble Kalman filter for atmospheric data assimilation. Mon. Weather Rev. 144, 4489–4532.
Huang, J., McQueen, J., Wilczak, J., Djalalova, I., Stajner, I., Shafran, P., Allured, D., Lee, P., Pan, L., Tong, D., et al., 2017. Improving NOAA NAQFC $PM_{2.5}$ predictions with a bias correction approach. Weather Forecast. 32, 407–421.
Huang, Y.N., Reich, B.J., Fuentes, M., Sankarasubramanian, A., 2019. Complete spatial model calibration. Ann. Appl. Stat. 13, 746–766.
Jiang, X., Enki Yoo, E.H., 2019. Modeling wildland fire-specific $PM_{2.5}$ concentrations for uncertainty-aware health impact assessments. Environ. Sci. Technol. 53, 11828–11839.
Jordan, A., Krüger, F., Lerch, S., 2019. Evaluating probabilistic forecasts with scoringrules. J. Stat. Softw. 90, 1–37.
Jurek, M., Katzfuss, M., 2021. Multi-resolution filters for massive spatio-temporal data. J. Comput. Graph. Stat. 30, 1–37.
Kang, D., Mathur, R., Rao, S.T., 2010. Real-time bias-adjusted O_3 and $PM_{2.5}$ air quality index forecasts and their performance evaluations over the continental United States. Atmos. Environ. 44, 2203–2212.
Katzfuss, M., 2017. A multi-resolution approximation for massive spatial datasets. J. Am. Stat. Assoc. 112, 201–214.
Katzfuss, M., Stroud, J.R., Wikle, C.K., 2016. Understanding the ensemble Kalman filter. Am. Stat. 70, 350–357.
Kennedy, M.C., O'Hagan, A., 2001. Bayesian calibration of computer models. J. R. Stat. Soc., Ser. B, Stat. Methodol. 63, 425–464.
Kong, L., Tang, X., Zhu, J., Wang, Z., Wu, H., Li, J., 2020. Developing high-resolution air quality reanalysis dataset over China for years 2013-2018 based on ensemble Kalman filter and surface observations from CNEMC. In: EGU General Assembly Conference Abstracts, p. 6848.
Leroux, B.G., Lei, X., Breslow, N., 2000. Estimation of disease rates in small areas: a new mixed model for spatial dependence. In: Statistical Models in Epidemiology, the Environment, and Clinical Trials. Springer, pp. 179–191.
Liang, X., Li, S., Zhang, S., Huang, H., Chen, S.X., 2016. $PM_{2.5}$ data reliability, consistency, and air quality assessment in five Chinese cities. J. Geophys. Res., Atmos. 121, 10–220.
Liang, X., Zou, T., Guo, B., Li, S., Zhang, H., Zhang, S., Huang, H., Chen, S.X., 2015. Assessing Beijing's $PM_{2.5}$ pollution: severity, weather impact, APEC and winter heating. Proc. R. Soc. A, Math. Phys. Eng. Sci. 471, 20150257.
Lindgren, F., Rue, H., 2015. Bayesian spatial modelling with R-INLA. J. Stat. Softw. 63, 1–25.
Lindgren, F., Rue, H., Lindström, J., 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. J. R. Stat. Soc., Ser. B, Stat. Methodol. 73, 423–498.
Lorenz, E.N., 1996. Predictability: a problem partly solved. In: Proc. Seminar on Predictability. ECMWF, pp. 1–18.
Lu, X., Zhang, S., Xing, J., Wang, Y., Chen, W., Ding, D., Wu, Y., Wang, S., Duan, L., Hao, J., 2020. Progress of air pollution control in China and its challenges and opportunities in the ecological civilization era. Engineering 6, 1423–1431.
Ma, P., Kang, E.L., 2020. Spatio-temporal data fusion for massive sea surface temperature data from MODIS and AMSR-E instruments. Environmetrics 31, e2594.
McMillan, N.J., Holland, D.M., Morara, M., Feng, J., 2010. Combining numerical model output and particulate data using Bayesian space-time modeling. Environmetrics 21, 48–65.
Mendy, A., Wilkerson, J., Salo, P.M., Weir, C.H., Feinstein, L., Zeldin, D.C., Thorne, P.S., 2019. Synergistic association of house endotoxin exposure and ambient air pollution with asthma outcomes. Am. J. Respir. Crit. Care Med. 200, 712–720.
Mitchell, H.L., Houtekamer, P.L., Pellerin, G., 2002. Ensemble size, balance, and model-error representation in an ensemble Kalman filter. Mon. Weather Rev. 130, 2791–2808.
Nychka, D., Bandyopadhyay, S., Hammerling, D., Lindgren, F., Sain, S., 2015. A multiresolution Gaussian process model for the analysis of large spatial datasets. J. Comput. Graph. Stat. 24, 579–599.
Riabiz, M., Chen, W.Y., Cockayne, J., Swietach, P., Niederer, S.A., Mackey, L., Oates, C.J., 2022. Optimal thinning of MCMC output. J. R. Stat. Soc., Ser. B, Stat. Methodol., 1–23.
Richardson, R., Kottas, A., Sansó, B., 2018. Bayesian non-parametric modeling for integro-difference equations. Stat. Comput. 28, 87–101.
Rue, H., Held, L., 2005. Gaussian Markov Random Fields: Theory and Applications. Chapman & Hall/CRC, Boca Raton, FL.
Rue, H., Martino, S., Chopin, N., 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J. R. Stat. Soc., Ser. B, Stat. Methodol. 71, 319–392.
Rue, H., Riebler, A., Sørbye, S.H., Illian, J.B., Simpson, D.P., Lindgren, F.K., 2017. Bayesian computing with INLA: a review. Annu. Rev. Stat. Appl. 4, 395–421.
Sahu, S.K., Gelfand, A.E., Holland, D.M., 2006. Spatio-temporal modeling of fine particulate matter. J. Agric. Biol. Environ. Stat. 11, 61–86.
Salter, J.M., Williamson, D.B., Scinocca, J., Kharin, V., 2019. Uncertainty quantification for computer models with spatial output using calibration-optimal bases. J. Am. Stat. Assoc. 114, 1800–1814.
Sampson, P.D., Richards, M., Szpiro, A.A., Bergen, S., Sheppard, L., Larson, T.V., Kaufman, J.D., 2013. A regionalized national universal kriging model using partial least squares regression for estimating annual $PM_{2.5}$ concentrations in epidemiology. Atmos. Environ. 75, 383–392.
Schmidt, A.M., Gelfand, A.E., 2003. A Bayesian coregionalization approach for multivariate pollutant data. J. Geophys. Res., Atmos. 108, 8783.
Shaddick, G., Thomas, M.L., Green, A., Brauer, M., van Donkelaar, A., Burnett, R., Chang, H.H., Cohen, A., Dingenen, R.V., Dora, C., et al., 2018. Data integration model for air quality: a hierarchical approach to the global estimation of exposures to ambient air pollution. J. R. Stat. Soc., Ser. C, Appl. Stat. 67, 231–253.
Sheather, S.J., Jones, M.C., 1991. A reliable data-based bandwidth selection method for kernel density estimation. J. R. Stat. Soc., Ser. B, Stat. Methodol. 53, 683–690.
Skamarock, W.C., Klemp, J.B., Dudhia, J., Gill, D.O., Liu, Z., Berner, J., Wang, W., Powers, J.G., Duda, M.G., Barker, D.M., et al., 2019. A Description of the Advanced Research WRF Model Version 4, vol. 145. National Center for Atmospheric Research, Boulder, CO, USA, p. 550.
Sørbye, S.H., Rue, H., 2014. Scaling intrinsic Gaussian Markov random field priors in spatial modelling. Spat. Stat. 8, 39–51.

Vannitsem, S., Bremnes, J.B., Demaeyer, J., Evans, G.R., Flowerdew, J., Hemri, S., Lerch, S., Roberts, N., Theis, S., Atencia, A., et al., 2021. Statistical postprocessing for weather forecasts: review, challenges, and avenues in a big data world. Bull. Am. Meteorol. Soc. 102, E681–E699.

Wang, Z.f., Xie, F.y., Wang, X.q., An, J., Zhu, J., 2006. Development and application of nested air quality prediction modeling system. Chin. J. Atmos. Sci. 30, 778–790 (in Chinese).

Wei, G.C., Tanner, M.A., 1990. A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. J. Am. Stat. Assoc. 85, 699–704.

Wendland, H., 1995. Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree. Adv. Comput. Math. 4, 389–396.

Wikle, C.K., 2002. A kernel-based spectral model for non-Gaussian spatio-temporal processes. Stat. Model. 2, 299–314.

Wikle, C.K., Hooten, M.B., 2010. A general science-based framework for dynamical spatio-temporal models. Test 19, 417–451.

Wikle, C.K., Zammit-Mangion, A., Cressie, N., 2019. Spatio-Temporal Statistics with R. Chapman & Hall/CRC, New York.

Wood, S.N., 2017. Generalized Additive Models: An Introduction with R. Chapman & Hall/CRC, New York.

Wood, S.N., 2020. Inference and computation with generalized additive models and their extensions. Test 29, 307–339.

Wood, S.N., 2022. Package 'mgcv' R package version 1.8-41. https://CRAN.R-project.org/package=mgcv.

Xie, H., Eheart, J.W., Chen, Y., Bailey, B.A., 2009. An approach for improving the sampling efficiency in the Bayesian calibration of computationally expensive simulation models. Water Resour. Res. 45, W06419.

Xu, K., Wikle, C.K., Fox, N.I., 2005. A kernel-based spatio-temporal dynamical model for nowcasting weather radar reflectivities. J. Am. Stat. Assoc. 100, 1133–1144.

Zammit-Mangion, A., Cressie, N., 2021. FRK: an R package for spatial and spatio-temporal prediction with large datasets. J. Stat. Softw. 98, 1–48.

Zammit-Mangion, A., Wikle, C.K., 2020. Deep integro-difference equation models for spatio-temporal forecasting. Spat. Stat. 37, 100408.

Zhang, B., Sang, H., Huang, J.Z., 2015. Full-scale approximations of spatio-temporal covariance models for large datasets. Stat. Sin., 99–114.

Zhang, B., Sang, H., Huang, J.Z., 2019. Smoothed full-scale approximation of Gaussian process models for computation of large spatial data sets. Stat. Sin. 29, 1711–1737.