Efficient Calibration of Numerical Model Output Using Sparse Hierarchical Dynamic Models

Abstract

Numerical air quality models, such as Community Multiscale Air Quality (CMAQ) system, play a critical role in characterizing pollution levels at fine spatial and temporal scales. Nevertheless, numerical model outputs may systematically overestimate or underestimate pollutants concentrations due to various reasons. In this work, we propose an hierarchal dynamic model that can be implemented to calibrate grid-level CMAQ outputs using point-level observations from sparsely located monitoring stations. Under a Bayesian framework, our model presents a flexible quantification of uncertainties by considering deep hierarchies for key parameters, which can also be used to describe the dynamic nature of data structural changes. In addition, we adopt several newly-emerged techniques, including the triangulation of the domain, tapering-based Gaussian kernel, sparse Gaussian graphical model, variational Bayes, and ensemble Kalman smoother, that significantly speed up the whole calibration procedure. Our approach is illustrated using daily $PM_{2.5}$ datasets of China's Beijing-Tianjin-Hebei region, which contains 68 monitoring stations and is covered by 2499 CMAQ 9-km grids. In contrast to existing methods, our model gives more accurate calibration results in most of the grids with higher computation efficiencies. This allows us to provide an effective calibration tool for large-scale numerical model outputs and generate better high-resolution maps of pollutants.

1. Introduction

High-quality, high-resolution, and large-scale air pollution maps are pivotal to accurately assess local air pollution exposures.[1,2] In recent years, numerical model systems, such as the Community Multiscale Air Quality (CMAQ)[3] and the Nested Air Quality Prediction Modeling System (NAQPMS),[4] combined with other systems for meteorological fields and emissions have been widely used in the production of air pollution forecast maps. Pollutant concentrations in each grid cell are usually estimated by solving a large deterministic dynamic system, which involves diffusion equations, chemical transportations, movements of the atmosphere, and other components.[3,5,6] The complexity of the systems renders these models valuable for forecasting on large spatial domains with no missing values, but may also introduce prediction biases through errors in input parameters or initial conditions (e.g., emission source inventory) especially under extreme conditions.[7,8] As an illustration, Figure 1 compares the CMAQ $PM_{2.5}$ predictions with the actual observations in China's Beijing-Tianjin-Hebei (BTH) region. Notice that the CMAQ overestimates the $PM_{2.5}$ concentrations in Zhangjiakou's winter and Tangshan's summer and yet underestimates the concentrations in Baoding's summer.
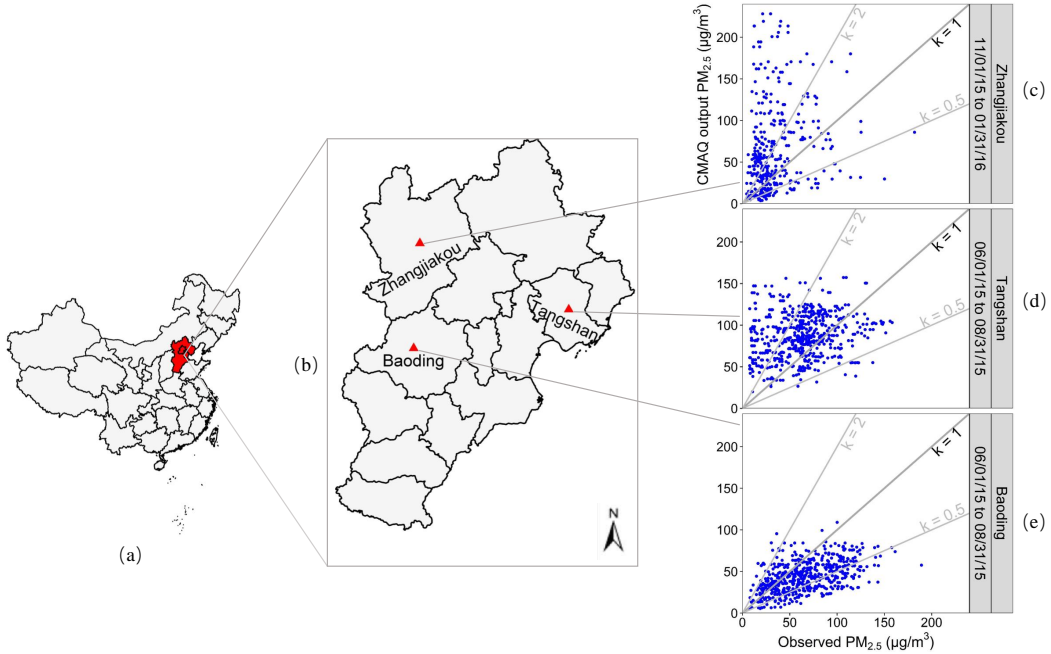

Figure 1. (a) Map of China; (b) zoomed-in map of Beijing-Tianjin-Hebei (BTH) region; (c)-(e) scatter plots of CMAQ $PM_{2.5}$ predictions versus actual observations at Zhangjiakou, Tangshan, and Baoding in different seasons.

Data assimilation, also known as data fusion, is one of the classical methods to improve CMAQ-based model forecasts; see Talagrand and Courtier[9], Evensen[10], Lopez-Restrepo et al.[11], Wu et al.[12] and ref-

erences therein. (Move elsewhere or get rid of this line?) For studies that are not interested in predicting pollutants in the future, researchers only need to calibrate existing pollution maps [13,14], numerical model outputs are frequently fused with datasets from other sources, especially observations from monitoring networks as they reflect the actual pollution levels at the local scale. [15–17] Monitoring stations are usually sparsely distributed in space and clustered in urban areas, but observations collected from the stations still provide reliable references for large-scale calibrations. [18–20]

In the literature, calibrating numerical model outputs using measurements from monitoring stations accurately and efficiently remains as a challenge for large-scale datasets. To improve spatial prediction accuracy, early studies such as Kennedy and O'Hagan [21] adopted a Bayesian approach to account for the uncertainties in numerical model parameters and performed calibration using conventional linear interpolation or "kriging." Because this method requires one dimension for each uncertainty parameter, the computation quickly becomes expensive for large-scale datasets. Fuentes and Raftery [18] assumed a "true" latent point-level process that drives both the grid-level outputs and point-level observations, and developed a "Bayesian melding" framework to connect the two data sources, by which the bias in the numerical model outputs can be estimated and corrected. Based on a spatiotemporal version of Kennedy and O'Hagan [21], McMillan et al. [22] calibrated CMAQ $PM_{2.5}$ outputs using Markov chain Monte Carlo (MCMC) techniques, which have been the standard tools for Bayesian inferences but faced computational difficulties for large-scale datasets. On the other hand, Berrocal et al. [15] and Berrocal et al. [16] proposed several downscaling approaches based on spatially varying coefficient (SVC) models, [23] where the numerical model outputs were explanatory variables and "downscaled" to explain the point-level observations. Under a Bayesian setting, once the varying coefficients are fitted under Gaussian assumptions, calibrations in each grid-cell can be simulated and integrated from posterior distributions, but the SVC models are also subject to computational difficulties for very large datasets. For spatio-temporal datasets, the dynamical state-space models are frequently adopted as exact inferences can be made using Kalman filter [24] under the linear Gaussian assumption, but they also become computationally infeasible when the state dimension is large. [25]

To reduce computational complexities, several approaches have been proposed for large space-time datasets. In the statistics literature, the common strategies are to approximate the covariance matrix using fixed-rank approximation, [26,27] sparse-matrix approximation, [28–32] or Gaussian Predictive Process approximation. [33,34] These reduced-rank methods usually fail to accurately capture the local dependence structure [35] or are unable to maintain the sparse structure over time. [36] In the geoscience community, ensemble-based methods, such as the ensemble Kalman filter (EnKF) [10] and the ensemble Kalman smoother (EnKs), [37] have been frequently used in many different applications [38–42] in which an ensemble that is propagated over time and update with new observations is used to approximate the results of Kalman filter. However, in practice,

the ensemble size is generally smaller than the state space and the states are typically confined to some spatial grids, which leads to a rank deficiency, spurious correlations,[43–45] and estimation bias.

In this work, we propose a hierarchical model to calibrate numerical model outputs. Under a Bayesian framework, we quantify uncertainties of input parameters with variational Bayes (VB) methods and evaluate a dynamic scheme of the true latent process with spatio-temporal tapering-based EnKs. At the deeper layer of the hierarchical model, we connect the latent process to the numerical model outputs in a similar fashion as the downscaler models. Compared to the SVC models, this approach improves computational efficiency through a series of computational techniques including domain triangulation, tapering-based Gaussian kernel, sparse Gaussian graphical model, and a transformation matrix, which are all described in more details in Section 2.3. We demonstrate this methodology using the CMAQ model outputs and $PM_{2.5}$ measurements collected from the BTH regions between 2015 and 2016 as seen in Figure 1. Under different assessment criteria, the proposed method outperforms all competing methods in calibration accuracy in most of the grids.

2. Materials and Methods

2.1. Beijing-Tianjin-Hebei region. The air quality of North China has drawn worldwide attention since the record-high haze events in January 2013.[46] The Beijing-Tianjin-Hebei (BTH) region, one of the largest industrial bases in China, have suffered from severe air pollution for decades. In recent years, the major pollutant in this region is fine particulate matters, or $PM_{2.5}$. With the launch of a series of movements including the short-term Air Pollution Prevention and Control Action Plan between 2013 and 2017,[47] "Blue Sky Defense Battle" between 2018 and 2020,[48] and the long-term "Beautiful China" targets through 2035,[49] significant air quality improvements have been made. Taking Beijing as an example, since 2013, SO_2 emissions have fell by 83%, NOx by 43%, VOCs by 42%, and $PM_{2.5}$ by 59%.[50] The densely populated BTH region consists of thirteen cities with more than 100 million residents and accounts for nearly 9% of China's GDP. Therefore, pollution control in this region is extremely important for balancing public health and economic growth.[51]

To illustrate our calibration model, we employ two sources of $PM_{2.5}$ data in the BTH region. The first data are CMAQ model outputs at 9-km scale with a total of 2,499 CMAQ grid cells that cover the BTH region. The second data are readings from 68 national monitoring stations, which are mostly clustered in urban areas, especially the city centers. For the two datasets, we collected hourly $PM_{2.5}$ concentrations from warm season (from June 1, 2015 to August 31, 2015) and cold season (from November 1, 2015 to January 31, 2016). The locations of CMAQ grid cells and monitoring stations are indicated in Figure 2 (a).
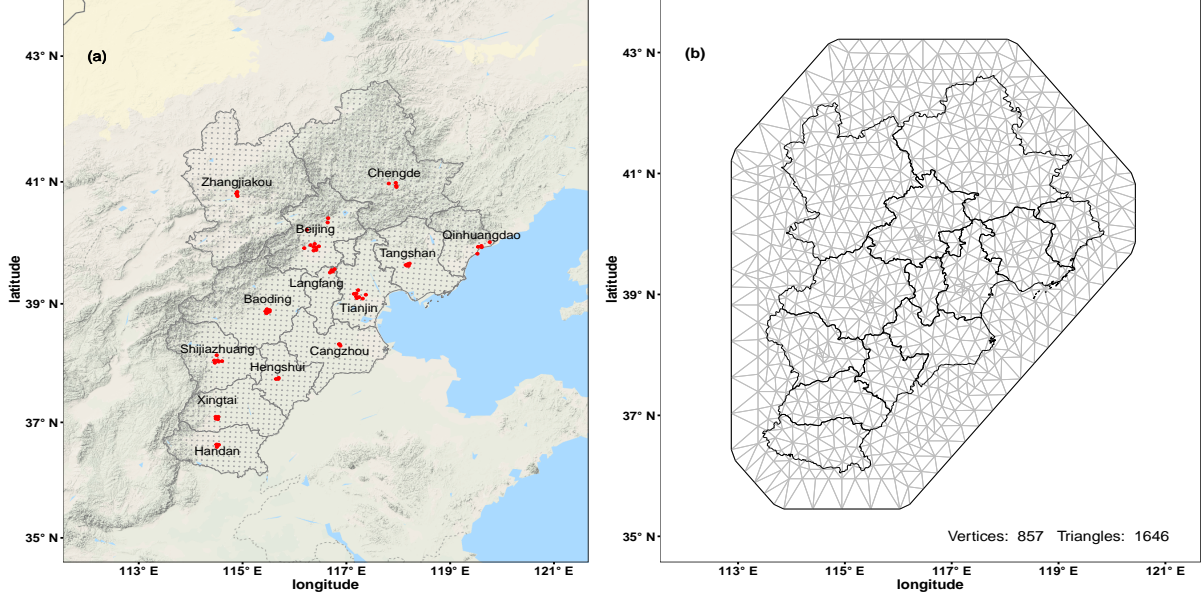
Figure 2. (a) Locations of 68 monitoring stations (red dots) and centroids of 2,499 9-km CMAQ grid cells (gray dots) in the BTH region; (b) Triangulated mesh of the BTH region and extended area with 857 vertices and 1,646 triangles.

2.2. Data preprocessing. Hourly observations from the monitoring stations consist of missing data. We first averaged the observations to daily data and imputed missing values using the ordinary kriging;[52] see Supporting Information (SI) Table S1 for the summary information. Due to the right skewness of the data, we then transformed the daily observations using square root that was also consistent with earlier studies on the $PM_{2.5}$ concentrations in the US.[53–55] Detailed comparisons between the square root and the logarithmic transformations, and between $PM_{2.5}$ observations and CMAQ model output can be found in SI Figures S1-S2 and S3-S4, respectively. Thus, we model the $PM_{2.5}$ observations on the square root scale and then present predictions on the original scale for ease of interpretation and comparison.

2.3. Hierarchical dynamic calibration model (HDCM). To connect point-level observations and grid-level CMAQ outputs, we develop an hierarchical dynamic space-time model. Let $\boldsymbol{D}$ be the spatial domain of interest that consists of n point-level observational sites. In our analysis of the BTH region, $n = 68$. We further assume that $\boldsymbol{D}$ can be covered by m grids, that could be either a regular grid network of the CMAQ outputs or a triangulated mesh with m vertices.

Denote $\boldsymbol{y}_t$ as the point-level observations of n monitoring stations at time t, which is driven by the true underlying process $\boldsymbol{w}_t$ with some mean-zero measurement error $\boldsymbol{\varepsilon}_t$. The first level our model, i.e., the data model, is

$$\boldsymbol{y}_t = \boldsymbol{w}_t + \boldsymbol{\varepsilon}_t. \tag{1}$$

Here we further assume a Gaussian distribution for the error such that $\boldsymbol{\varepsilon}_t \sim \mathcal{N}_n(\boldsymbol{0}, \boldsymbol{R})$, where $\boldsymbol{R} = \sigma^2 \boldsymbol{I}_n$ for

some $\sigma > 0$ and $\boldsymbol{I}_n$ is the $n \times n$ identity matrix. It is worth noting that this assumption can be relaxed for other distributions with more complex covariance structures.

In the second level our model, i.e., the process model, the latent process $\boldsymbol{w}_t$ can be further decomposed into two parts: an n-dimensional point-level additive offset term $\boldsymbol{\mu}_t$ and an m-dimensional grid-level spatiotemporal process $\boldsymbol{v}_t$ transformed by a predetermined $n \times m$ mapping matrix $\boldsymbol{H}$. In particular, the model is

$$\boldsymbol{w}_t = \boldsymbol{\mu}_t + \boldsymbol{H}\boldsymbol{v}_t. \tag{2}$$

where the offset term $\boldsymbol{\mu}_t$ plays a key role in fusing multiple data sources even when the spatial resolutions of the data sources are different and is useful in incorporating effects from the covariates; see Section 2.3.2 for more details. In this work, $\boldsymbol{v}_t$ is a spatial Gaussian field evaluated at some predetermined triangulated grids.

2.3.1 Triangulation of the region. A triangulated mesh, instead of a regular grid, of the region is employed for our model setup. The monitoring stations in the BTH region, and those from many other real-life applications, are irregularly scattered in space, making it impractical to interpolate station locations to the nearest grid point under a regular grid setup. Following Lindgren et al. [29], we use a triangulation scheme to represent the irregular BTH region by dividing it into a set of non-intersecting triangles where any two triangles meet in at most one common edge or vertex. We also extend the region to 100 km beyond its boundary and implement finer triangles inside the region and coarse triangles outside the region to avoid any boundary effects. To overcome the trade-off between accuracy of the spatial representation of the region and the computation costs, we divided the BTH region into 1,646 triangles with 857 vertices in total. The triangulated mesh is illustrated in Figure 2 (b).

Need a subsubsection title here? Alternatively, we can remove the title of the triangulation of the region. Let $\boldsymbol{v}_t = (v_t(B_1), \cdots, v_t(B_m))'$. The grid-level dynamic process $\boldsymbol{v}_t$ can be defined on either the vertices or the centroids of the triangles. We follow the finite element approximation of Gaussian Markov random fields described in Lindgren et al. [29] and define $\boldsymbol{v}_t$ on the vertices of the mesh in Figure 2 (b), and hence, $m = 857$.

For the dynamic process $\boldsymbol{v}_t$ at time t, we impose a stochastic model that evolves dynamically in time:

$$v_t(B_s) = \int_{\boldsymbol{D}} \mathcal{M}\{v_{t-1}(B_x); \boldsymbol{\theta}\} \circ \boldsymbol{\Gamma}_{B_s} dB_x + \eta_t(B_s) \tag{3}$$

where $\mathcal{M}$ could be a predefined linear or non-linear function of $\boldsymbol{v}_{t-1}$ with some time-independent input parameters $\boldsymbol{\theta}$, the symbol "$\circ$" denotes the Hadamard product (also known as the element-wise multiplication), $\boldsymbol{\Gamma}_{B_s}$ is a tapered correlation vector [28] in the spatial domain $\boldsymbol{D}$, its role is to remove some spatial "useless" information and reduce unnecessary process propagation, so as to improve the accuracy of model and the efficiency of matrix calculation (see SI Text S4), and $\boldsymbol{\eta}_t$ is a the Gaussian Markov random field in spatial

6

statistics used to capture remaining spatial information from improper specification of $\mathcal{M}$ and be mutually and serially independent of the error ε_t in eq 1.

(a)

(b)

$$v_t(B_0) = \mathcal{M}(\boldsymbol{v}_{t-1}; \boldsymbol{\theta}) \circ \boldsymbol{\Gamma}_{B_0} + \eta_t(B_0)$$
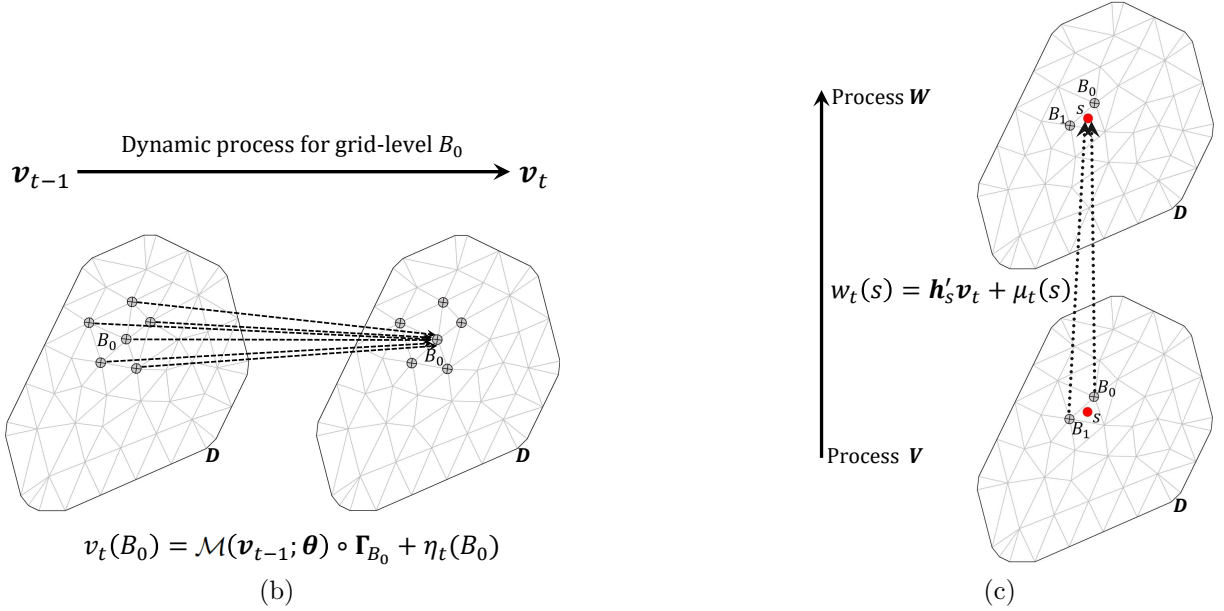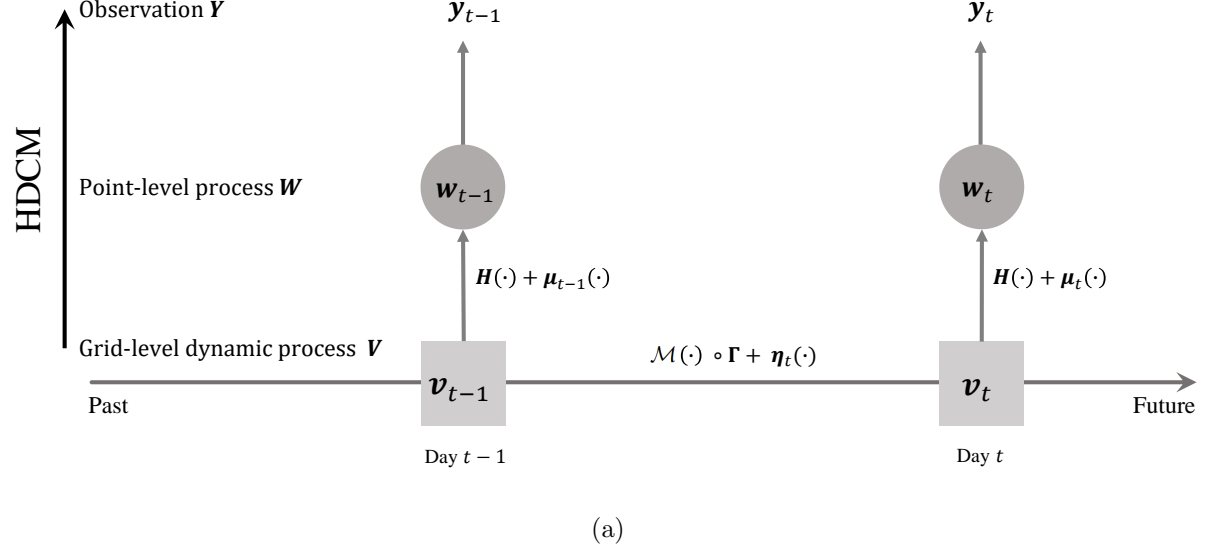
(c)

Figure 3. Cartoon illustration of the hierarchical dynamic calibration model (HDCM). (a) The general framework of the HDCM. The direction of the arrow on the bottom indicates the passage of time. The direction of the arrow on the left represents the hierarchical structure of the HDCM: from the grid-level dynamic process $\boldsymbol{V}$ to the point-level process $\boldsymbol{W}$ and then to the point-level observations $\boldsymbol{Y}$. (b) Zoomed-in diagram of the transition from $\boldsymbol{v}_{t-1}$ to $\boldsymbol{v}_t$. At time t, the process at vertex B_0, $v_t(B_0)$, is constructed from a weighted linear combination of $v_{t-1}(\cdot)$ at the vertices surrounding B_0 with weights defined by function $\mathcal{M} \circ \boldsymbol{\Gamma}_{B_0}$ and an "instantaneous" spatial random error $\eta_t(B_0)$. (c) Zoomed-in diagram of the transition from $\boldsymbol{v}_t$ to $\boldsymbol{w}_t$. At time t, the point-level process $w_t(s)$ at spatial location s (marked by the red dot) is obtained from the grid-level process $v_t(B_0)$ and $v_t(B_1)$ via a mapping factor $\boldsymbol{h}'_s$ and an offset term $\mu_t(s)$. Only vertices B_0 and B_1 are considered for s because all other vertices are considered too far from s from the definition of $\boldsymbol{h}_s$.

In this work, we assume that this recursive system of the stochastic dynamic model initiates with an m-dimensional normal vector $\boldsymbol{v}_0$, such that $\boldsymbol{v}_0 \sim \mathcal{N}_m \left(\boldsymbol{0}, \boldsymbol{Q}_0^{-1}\right)$, where $\boldsymbol{Q}_0 = \tau_0^2 \left(\boldsymbol{G} + k_0^2 \boldsymbol{I}_m\right)$, $\boldsymbol{G}$ is the Laplacian matrix[57] defined on the triangulated mesh (see eq S2 for its detailed definition), $\tau_0 > 0$, and $k_0 > 0$. A linear function is considered for $\mathcal{M}$ with $\boldsymbol{\theta} = (\theta_1, \theta_2)$, such that $\mathcal{M}(v_{t-1}(B_i); \theta_1, \theta_2) = \theta_1 \sum_{j=1}^m \left\{ v_{t-1}(B_j) \exp\left(-\frac{\|B_i - B_j\|^2}{\theta_2}\right)\right\}$, where $\theta_1 \neq 0$, $\theta_2 > 0$ is a scale parameter, and $\|B_i - B_j\|$ denotes the Euclidean distance with respect to the spatial coordinates B_i and B_j of grids or vertices i and j. We also model $\boldsymbol{\eta}_t$ using the undirected Gaussian graphical model[58] that is the Gaussian Markov random field. For simplicity, $\boldsymbol{\eta}_t \sim \mathcal{N}_m \left(\boldsymbol{0}, \boldsymbol{Q}^{-1}\right)$, where $\boldsymbol{Q} = \tau^2 \left(\boldsymbol{G} + k^2 \boldsymbol{I}_m\right)$, $\tau > 0$, and $k > 0$. Additional details can be referred to SI Text S1. The HDCM (eq 1-3) is also illustrated using a cartoon in Figure 3. Once all the model parameters are estimated, we will be able to make use of the numerical model output to characterize $PM_{2.5}$ information at locations with no available data (see SI Text S2). In the following sections, we will discuss several challenges in building, fitting, and assessing the model, and present the corresponding solutions.

2.3.2. Data fusion with CMAQ output $PM_{2.5}$ data and the meteorological variables. Because the CMAQ does not incorporate any monitoring observations into the model and yet both the CMAQ model output and the $PM_{2.5}$ observations are similar (see SI Figure S3, S4), we employ the CMAQ model output as a covariate for the observational data and denote the model output at time t as $\tilde{\boldsymbol{x}}_t$. This is known as data fusion[18]. Note the spatial resolution of the model output does not align well with that of the data. The most popular data fusion technique uses the model output from the CMAQ grid cell nearest to the site at each site; see Berrocal et al.[15] and Berrocal et al.[16]. However, there can be multiple nearest grid cells when the site falls exactly on the boundary of a cell. Also, as all sites in the same grid cell (the area of each cell is 81 km^2 in our data) are assigned with the same CMAQ model output, this technique fails to integrate all available information provided by the data. To improve on the conventional data fusion procedure, we first consider all grid cells that are within a certain range from site s (we use 30 km here), and then calculate the CMAQ covariate $x_{1,t}(s)$ as an inverse distance weighted average of the CMAQ model output at these cells scaled by the inverse distance between the centroid of each cell and site s.

Earlier research works have shown that $PM_{2.5}$ concentrations in the BTH region are heavily influenced by meteorological variables including air pressure, temperature, dew point, and the cumulative wind power.[46,59] The proposed calibration model allows us to conveniently include these variables and intercept terms as other covariates $x_{2,t}(s), x_{3,t}(s), \ldots$. Lastly, all covariates enter the HDCM as $\boldsymbol{x}_t(s) = (x_{1,t}(s), \cdots, x_{p,t}(s))$ for the additive offset term $\boldsymbol{\mu}_t$ in eq 2:

$$\mu_t(s) = \boldsymbol{x}_t(s)\boldsymbol{\beta}. \tag{4}$$

2.3.3. Bayesian framework. There are many parameters involved in this hierarchical space-time

model. We employ a Bayesian approach to estimate the parameters. All parameters other than the $\boldsymbol{V}$ are saved in $\boldsymbol{\Theta}$, that is $\boldsymbol{\Theta} = \{\boldsymbol{\beta}, \sigma^2, \tau_0^2, \tau^2, k_0^2, k^2, \theta_1, \theta_2\}$.

From the Bayes' theorem, the conditional distribution of the parameters given the point-level observations $p(\boldsymbol{\Theta}, \boldsymbol{V}|\boldsymbol{y})$, i.e., the posterior distribution, can be expressed as

$$p(\boldsymbol{\Theta}, \boldsymbol{V}|\boldsymbol{y}) \propto L(\boldsymbol{y}; \boldsymbol{\Theta}, \boldsymbol{V})\pi(\boldsymbol{\Theta})\pi(\boldsymbol{V}), \tag{5}$$

where $L(\boldsymbol{y}; \boldsymbol{\Theta}, \boldsymbol{V})$ is the likelihood conditional on the parameters, and the expert knowledge or the prior belief is incorporated into the prior distributions $\pi(\boldsymbol{\Theta})$ and $\pi(\boldsymbol{V})$. The log-likelihood can be written as:

$$l(\boldsymbol{y}; \boldsymbol{\Theta}, \boldsymbol{V}) = \log L(\boldsymbol{y}; \boldsymbol{\Theta}, \boldsymbol{V}) = -\frac{nT \log \sigma^2}{2} - \frac{\sum_{t=1}^{T}\sum_{k=1}^{n}\left[y_t(s) - \boldsymbol{x}_t(s)\boldsymbol{\beta} - \boldsymbol{h}_j'\boldsymbol{v}_t\right]^2}{2\sigma^2} + \text{constant},$$

where $\boldsymbol{h}_j'$ represents the jth row of the mapping matrix $\boldsymbol{H}$ for $j = 1, 2, \cdots, n$. For the second term on the right side of eq 5, all parameters in $\boldsymbol{\Theta}$ are assigned weakly informative priors as given in Table **??**. Using the recursive nature of the stochastic dynamic model for $\boldsymbol{v}_t$, the last term of eq 5 can be expressed as:

$$\pi(\boldsymbol{V}) = \pi(\boldsymbol{v}_0)\prod_{t=1}^{T} p(\boldsymbol{v}_t|\boldsymbol{v}_{t-1}). \tag{6}$$

From eq 3 and the distributional assumption for $\boldsymbol{\eta}_t$, we also have

$$p(\boldsymbol{v}_t|\boldsymbol{v}_{t-1}) \propto \frac{\det(\boldsymbol{Q}^{-1})}{2} \exp\left\{-\frac{[\boldsymbol{v}_t - \mathcal{M}(\boldsymbol{v}_{t-1}; \boldsymbol{\theta}) \circ \boldsymbol{\Gamma}]'\boldsymbol{Q}[\boldsymbol{v}_t - \mathcal{M}(\boldsymbol{v}_{t-1}; \boldsymbol{\theta}) \circ \boldsymbol{\Gamma}]}{2}\right\}. \tag{7}$$

Under this Bayesian framework, we are able to combine all the structural relationships described by the data model, the process model, the time-dependent model for the grid-level spatio-temporal process, and the linear relationship between the CMAQ model output and measurements from monitoring stations into the likelihood function , see SI Text S3. We also minimize any influence from the prior distributions by employing non-informative priors shown in SI Table S2, and then full inferences of the parameters can then be made based on the posterior distributions of the parameters.

2.3.4. Variational Bayesian and ensemble Kalman methods. The joint posterior distribution $p(\boldsymbol{\Theta}, \boldsymbol{V}|\boldsymbol{y})$ demands some highly efficient computational techniques for our space-time hierarchical model under the Bayesian setting. The dynamic nature of the grid-level process $\boldsymbol{V}$ and the expensive matrix inversion due to the size of the grid exacerbate the computational challenges. Instead of using the conventional Markov chain Monte Carlo (MCMC) algorithm, we adopt a Variational Bayesian (VB) method that is fast and can be easily adapted for the HDCM described above. An ensemble Kalman Smoother (EnKS) and a tapering-based approximation technique are also employed here. To perform statistical inference, we would assume all parameters in $\{\boldsymbol{\Theta}, \boldsymbol{V}\}$ are mutually conditionally independent given $\boldsymbol{y}$.

We carry out an approximate inference on the posterior distribution of $\boldsymbol{\Theta}$ using the VB method. VB

estimation is a fast alternative to MCMC for performing approximate Bayesian inference, especially for large datasets. It approximates intractable integrals arising in Bayesian inference using optimization. Based on a lower bound on the marginal likelihood that can be computed quickly, we then derive a variational approximation for the posterior distribution; see Blei et al.[60] and references therein for a further detailed discussion on this topic. Additional technical details involved in the VB estimation for the HDCM can be referred to SI Text S5. The excellent performance of this approach has been verified through our simulation studies and in Ren et al.[61].

On the contrary, we approximate the posterior distribution of $\boldsymbol{V}$ using an ensemble-based Kalman smoother. Although the exact form of the posterior can be calculated using the Kalman filter under a linear function $\mathcal{M}$ in eq 3, storing and inverting large matrices quickly become expensive and infeasible when m and/or n is large. Thus, approximations are needed. The ensemble Kalman filter (EnKF) developed for geophysical data assimilation by Evensen[10] is an approximate version of the Kalman filter in which the distribution of the state vectors (i.e., $\boldsymbol{V}$ in our model) is approximated by a sample or an "ensemble" from the distribution. The ensemble is then propagated forward through time and updated when new data become available. This ensemble-based representation is a form of dimension reduction that leads to computational feasibility even for very high-dimensional systems. The ensemble Kalman smoother (EnKS) proposed by Evensen and Van Leeuwen[37] is an extension of the EnKF to a smoother formulation. For numerical model output calibration problems, instead of including only the past and present observations as in the EnKF, the EnKS uses all of the past, present, and future observational data $\boldsymbol{y}_t$ for $t = 1, \cdots, T$, to calibrate the CMAQ output at each time $t \in \{1, \cdots, T\}$ and consequently provides more accurate predictions than the EnKF. The EnKS is also computationally efficient as it is a forward-only process. It works particularly well for data with multimodality and skewness properties that are shared by many real-life data applications including ours. Using a spatio-temporal tapering, we regularize the sample covariance matrix of the smoothing ensemble to avoid its rank deficiency and spurious correlations that are documented in Katzfuss et al.[42] and verified through our simulation studies. More specifically, we multiply the sample covariance matrix by a sparse matrix from the Wendland tapering function.[62] More details are provided in the SI.

3. Results and Discussion

3.1. Regional cross validation. To examine the performance of the HDCM, we carry out a regional cross validation to compare the model predictions against the observational data. Each time, we consider one prefecture-level city as a *test* city and the remaining twelve cities as the *training* cities. At each test city, we first use the PM$_{2.5}$ concentrations measured at the monitoring stations of the remaining 12 training

cities and the CMAQ model output of the entire BTH region to fit the model, and then predict the daily measurements at the monitoring stations of that test city. This procedure is known as leave-one-area-out cross validation in the literature.

3.2. Model comparison. To assess the predictions of the proposed HDCM and calibrate the CMAQ model output, we perform regional cross validations using the HDCM and CMAQ, along with several popular models in the field including the universal kriging, a random forest-based method advocated by atmospheric scientists[63–65], and a downscaling calibration model allowing for spatially-varying coefficients (hereinafter referred as SVC),[23] see SI Text S9 for detailed description of these models. The SVC is implemented through the `spBayes` package in R.[66] Four meteorological variables (air pressure, temperature, dew point, and cumulative wind power) are employed as the additional covariates for the aforementioned models except for the CMAQ. The random forest model also considers some meteorological and land-use covariates.

3.2.1. Numerical comparison. Different model predictions are first compared against the actual observations using numerical measures including root mean squared error (RMSE) and continuous rank probability score (CRPS). The CPRS is developed to evaluate probabilistic forecasts by accounting for both calibration and sharpness of the forecasts, with the calibration analogous to accuracy and sharpness to precision in point estimation[67]. It is defined as

$$\text{crps}(F_t, y_t) = \int_{-\infty}^{\infty} \left[F_t(z) - \mathbb{1}(z \le y_t) \right]^2 dz, \tag{8}$$

where F_t is the cumulative distribution function of the predictive distribution, y_t is the observation, and $\mathbb{1}(z \le y_t)$ denotes the indicator function that is one if $z \le y_t$ and zero otherwise. Under normality, we derive a measure of the model prediction performance using the space-time average of all crps values calculated in eq 8.

$$\text{CRPS} = \frac{1}{nT} \sum_{t=1}^{T} \sum_{s=1}^{n} \text{crps}(F_t(s), y_t(s)). \tag{9}$$

A smaller CRPS value indicates a more accurate and precise prediction. The RMSE and CRPS are summarized in Table 1 for the warm season and Table 2 for the cold season.

Tables 1 and 2 show that the proposed HDCM tops the list of five competitive models with the best overall prediction performance using either RMSE and CRPS for both seasons, followed by universal kriging, SVC, and the random forest model, while CMAQ is at the bottom of the list indicating an urgent need to calibrate CMAQ numerical model output. Not only does the HDCM yield the best predictions averaged across all cities, but also offers the most superior results for almost every city. For example, CMAQ beats all models except HDCM for Qinghuangdao in the summer of 2015. Cold-season predictions are generally worse than warm-season predictions for the same model at each city due to the large fluctuations in winter

11

air quality. In comparison to CMAQ, the HDCM significantly improves RMSE by 46.35% and CRPS by 36.78% in the cold season, and 37.74% and 41.08% in the warm season, respectively. Compared to the SVC that has gained considerable popularity in the modeling of air pollution, [23,66,68,69] the HDCM reduces both RMSE and CRPS by as much as 26% in the winter and 17% in the summer. Similar results are observed using other measures such as FAC2 (i.e., the fraction of predictions within a factor of two, that is a robust measure as it is not overly influenced by outliers) and the Pearson correlation coefficient (Corr) and are documented in SI Table S3 and Table S4, providing substantial evidence for the effectiveness of the HDCM in the BTH region.

Table 1. Mean of daily root mean squared error (RMSE) and continuous rank probability score (CRPS) for $PM_{2.5}$ concentration predictions (µg/m^3) calculated for 13 cities using five models during the warm season: (1) CMAQ numerical model output, (2) universal kriging, (3) random forest model, (4) spatially-varying coefficient downscaling model (SVC), and (5) proposed HDCM. Models (2)-(5) include CMAQ numerical model output and some meteorological variables as their covariates discussed in Text S8. The warm season is from June 1, 2015, to August 31, 2015.

| City	RMSE					CRPS				
	CMAQ	UK	RF	SVC	HDCM	CMAQ	UK	RF	SVC	HDCM
Baoding	36.25	32.22	25.29	34.58	20.40	27.93	15.91	13.95	16.77	10.75
Beijing	36.87	22.07	30.02	23.39	21.40	26.37	12.39	17.11	12.73	11.75
Cangzhou	34.41	19.76	21.57	20.99	18.76	28.16	10.98	12.13	11.84	10.16
Chengde	25.04	18.82	20.34	19.44	17.82	15.83	10.25	10.79	10.62	9.93
Handan	34.31	22.55	28.11	22.46	23.70	26.67	12.92	15.99	12.85	13.37
Hengshui	44.14	19.77	26.32	20.86	18.36	36.56	11.25	14.68	11.82	10.54
Langfang	27.52	13.77	22.53	13.45	15.30	21.45	9.49	12.40	9.22	10.24
Qinhuangdao	24.15	28.03	26.35	25.76	23.39	17.18	15.30	15.80	14.26	12.22
Shijiazhuang	34.22	28.74	31.55	28.41	26.69	25.85	16.15	18.37	15.86	14.79
Tangshan	43.65	29.72	27.84	28.81	23.99	35.01	16.73	16.68	16.31	13.51
Tianjin	24.24	19.01	23.60	18.85	15.53	18.27	11.04	13.42	11.08	9.28
Xingtai	40.07	30.05	32.94	32.56	21.22	31.10	16.78	19.10	18.22	11.88
Zhangjiakou	24.02	35.54	18.03	36.24	17.61	18.26	19.21	10.11	19.37	9.82
Mean	32.99	24.62	25.73	25.06	20.54	25.28	13.72	14.66	13.92	11.53

Table 2. Mean of daily root mean squared error (RMSE) and continuous rank probability score (CRPS) for $PM_{2.5}$concentration predictions (µg/m^3) calculated for 13 cities using five models during the cold season. Same definitions as in Table 1. The cold season is from November 1, 2015, to January 31, 2016.

City	RMSE					CRPS				
	CMAQ	UK	RF	SVC	HDCM	CMAQ	UK	RF	SVC	HDCM
Baoding	94.64	67.93	75.45	73.08	71.03	50.36	39.18	40.61	40.92	38.79
Beijing	92.52	68.95	63.58	66.41	54.51	46.19	35.21	32.64	34.01	28.73
Cangzhou	60.99	47.50	41.40	46.93	32.28	31.87	31.02	23.86	30.21	21.99
Chengde	41.62	38.77	65.89	47.15	24.79	22.69	20.76	34.98	24.11	14.54
Handan	66.09	68.47	54.94	68.69	44.23	37.60	41.79	31.81	41.53	27.47
Hengshui	91.19	57.95	80.29	63.09	48.67	46.21	32.08	38.82	33.30	30.42
Langfang	85.44	32.75	52.44	35.81	37.56	46.14	26.62	29.86	26.79	29.37
Qinhuangdao	46.19	79.02	45.67	80.97	31.79	24.91	40.52	25.68	42.66	16.35
Shijiazhuang	87.72	48.50	60.96	44.91	48.56	47.99	33.72	35.87	31.62	32.51
Tangshan	95.75	38.34	60.68	47.20	25.89	51.84	25.41	34.60	29.80	19.80
Tianjin	70.74	38.22	49.84	44.78	30.17	38.08	25.43	28.46	29.40	23.34
Xingtai	85.79	63.19	74.17	64.72	49.37	45.57	35.94	39.12	36.37	31.37
Zhangjiakou	64.42	78.29	54.55	90.37	28.63	33.03	41.22	28.15	45.39	15.61
Mean	75.62	55.99	59.99	59.55	40.57	40.19	32.99	32.65	34.32	25.41

3.2.2. Graphical comparison. Besides the aforementioned numerical metrics, several graphs are also employed to assess the performance of the models. For Bayesian models including the SVC and HDCM, we calculate the pointwise median predictions and construct the 95% credible intervals at each city. These prediction results along with the CMAQ numerical model output and the actual observations are portrayed in Figure 4. For a clearer illustration, we only present two cities using a month of data, and other cities and time periods follow very similar patterns. The median predictions from the HDCM are generally closer to the actual observations than the other two models with significantly narrower credible intervals, especially during the cold season, implying that the HDCM's predictions are both accurate and precise. While the CMAQ numerical model output underestimates Beijing's air pollutant concentrations in the summer and overestimates Tangshan's concentrations in the winter, the HDCM is able to calibrate the model output quite well.
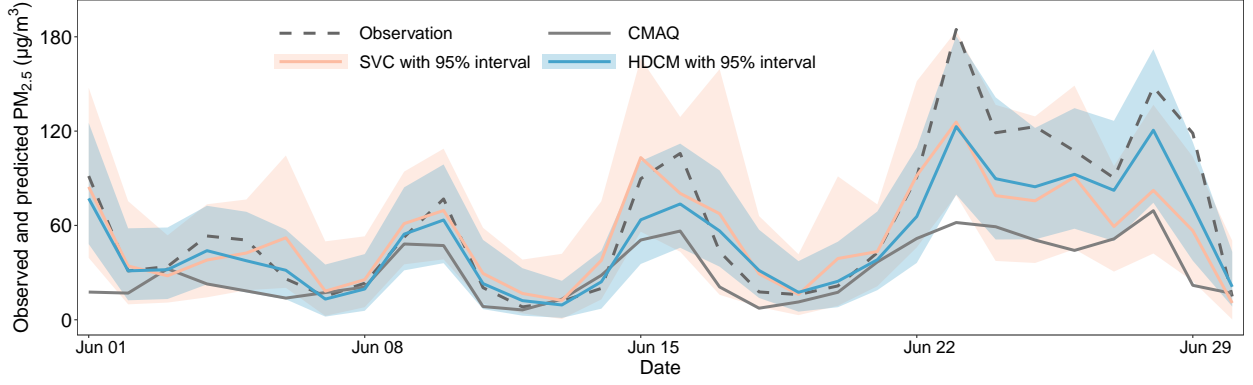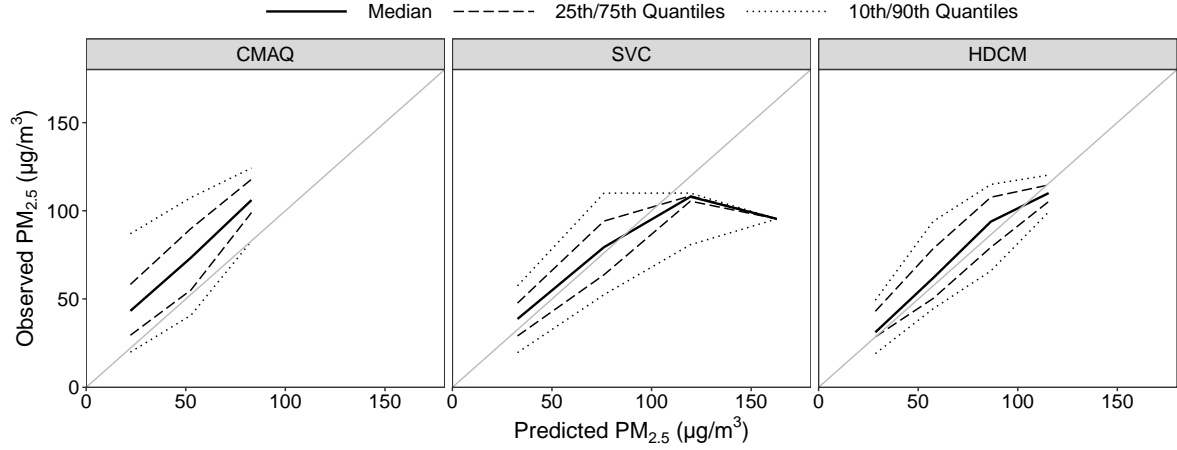
(a) Beijing from June 1, 2015, to June 30, 2015.
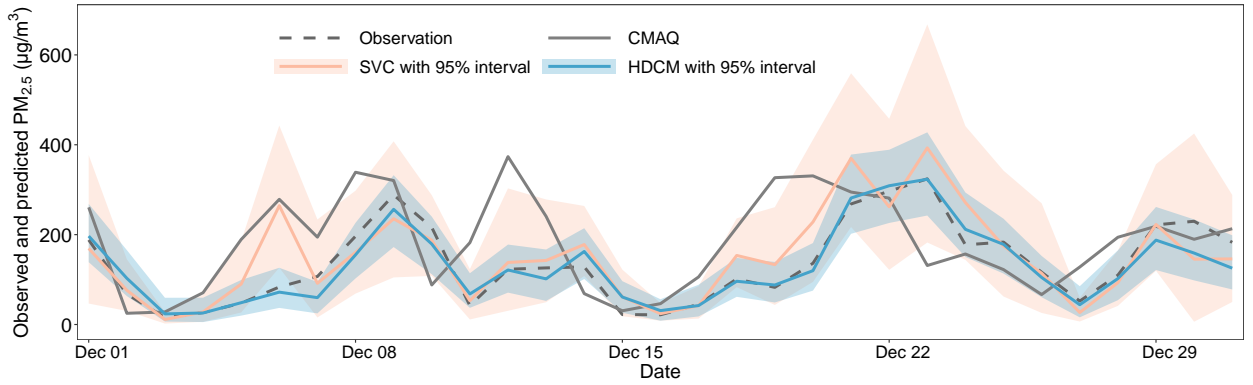
(b) Tangshan from December 1, 2015, to December 31, 2015.

Figure 4. Time series plots of observed $PM_{2.5}$ concentrations and daily predictions (median and 95% credible intervals) obtained from CMAQ, SVC, and HDCM at Beijing and Tangshan.

We are also concerned with the joint distribution of the observations and predictions obtained from various models, known as the conditional bias. While the unconditional prediction bias is examined through RMSE or time series plot like Figure 4, the conditional bias can be investigated using unconditional quantile plots developed in Murphy et al.[70]. The conditional quantile plots characterize the conditional distributions of the observations given the predictions and are helpful for model calibration refinement; interested readers are referred to Chapter 8 of Wilks[71] for more discussion on this topic. In Figure 5, compared to CMAQ that exhibits a large degree of underforecasting and SVC that is overforecasting at Baoding and Cangzhou, the HDCM shows a minimal amount of conditional bias in its forecasts.
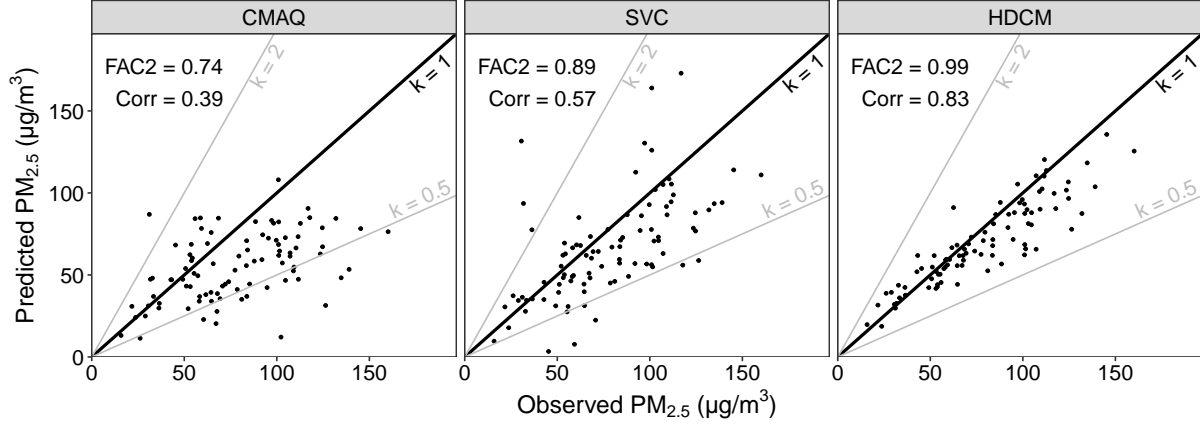
(a) Baoding from June 1, 2015, to August 31, 2015.
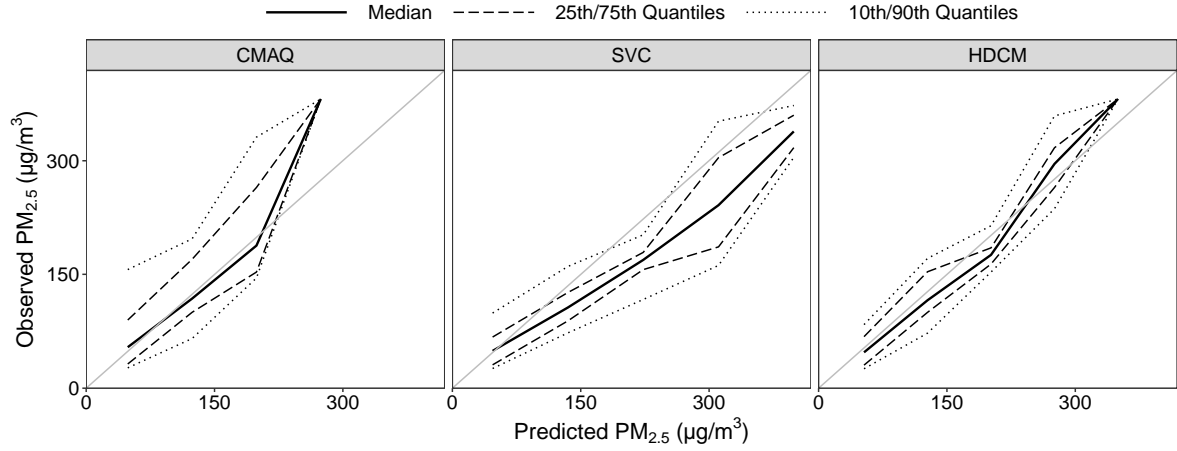
(b) Cangzhou from November 1, 2015, to January 31, 2016.

Figure 5. Conditional quantile plot of three-month daily mean predictions obtained from CMAQ, SVC, and HDCM with the observed $PM_{2.5}$ concentrations at Baoding and Cangzhou. The smoothed quantiles from the conditional distributions are drawn in relation to the 1 : 1 line.

Scatter plots of the predictions against the observations are the last set of graphical displays to assess the model performance. A perfect model would have FAC2 = 1 and Corr = 1. As displayed in Figure 6, almost all HDCM predictions are between 0.5 and 2 times the observation values with an FAC2 value as high as 0.99 at Xingtai in the summer and 0.93 at Handan in the winter. The correlations are 0.83 and 0.89 at the two cities, respectively. In contrast to the HDCM, CMAQ and SVC have not done an equally excellent job in predicting the $PM_{2.5}$ concentrations with lower FAC2 values and correlations.

(a) Xingtai from June 1, 2015, to August 31, 2015.
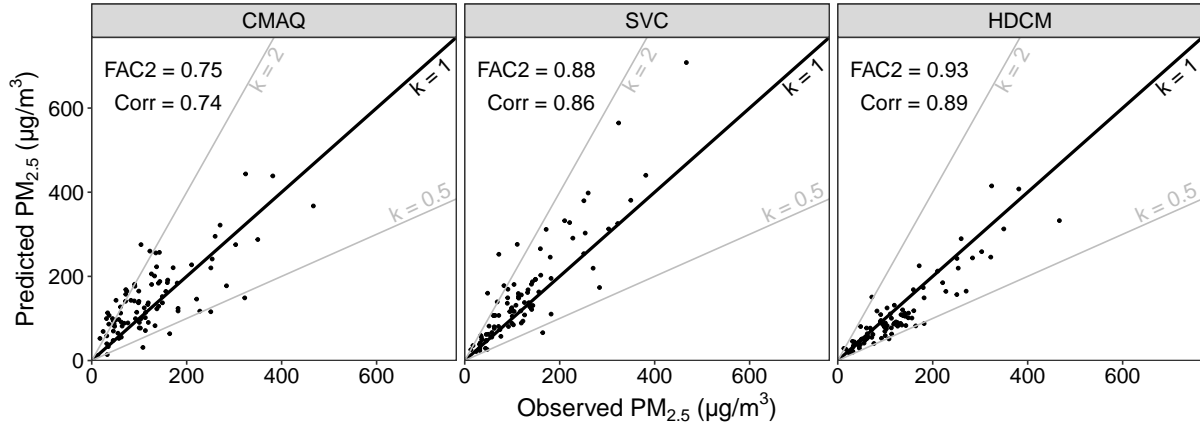
(b) Handan from November 1, 2015, to January 31, 2016.

Figure 6. Scatter plots of three-month daily mean predictions of CMAQ, SVC, and HDCM against observed PM$_{2.5}$ concentrations.

3.3. Space-time calibration. We now perform a space-time calibration of the CAMQ numerical model output for the entire BTH region using the SVC and HDCM models. Because the CMAQ data we acquired consist of only the average PM$_{2.5}$ concentration at each of the 2,499 CMAQ grid cells of the region. To ensure a fair comparison with the space-time calibrations using the SVC and HDCM models, we first generate multiple spatial locations around the centroid of each grid cell, interpolate PM$_{2.5}$ concentrations at each spatial location based on the CMAQ data, and then use the newly generated space-time data at all locations as our pre-calibration CMAQ model output. In this study, we generate a total of 12,787 spatial locations in the BTH region.

Figure 7 displays the calibration results using the SVC and HDCM models from December 17 to December 22, 2015. The average PM$_{2.5}$ measurements across all monitoring stations in each of the 13 cities are also marked in all plots. The smoother the transition from the cities to their surrounding areas, the better the overall calibration results of the corresponding model. It is evident that HDCM offers the most continuous

16

spatial map compared to the other two models, especially on December 22, 2015 when both CMAQ and SVC clearly underestimates the air pollution levels in the southeastern part of the BTH region. While the CMAQ model output tends to either overestimate or underestimate the pollutant concentrations at all cities in the cold season and underestimates almost all cities in the warm season (see SI Figure S15), the HDCM proves to be an effective calibration tool for the CMAQ model output.
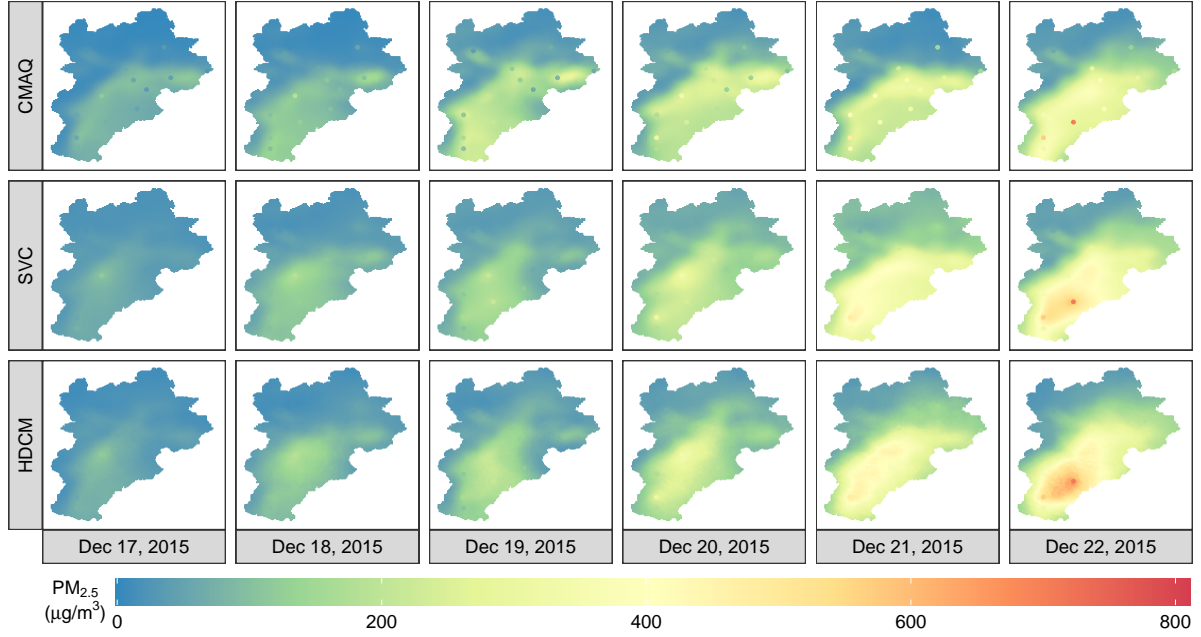

Figure 7. CMAQ numerical model output and calibration results of the SVC and HDCM from December 17 to December 22, 2015. The solid dots represent the average $PM_{2.5}$ levels of all monitoring stations in each of the 13 cities.

We also compare the computational efficiency of using the HDCM versus the SVC for the calibration process. All model estimation and bayesian calibration were run on a 8-core workstation with a 3.8 GHz Intel Core i7-9800X and 128 GB RAM. Although it took $4,291$ seconds to fit the HDCM models for the warm and cold seasons as opposed to 899 seconds for the SVC, the calibration step of the HDCM took considerably less time with just 52 seconds versus $47,448$ seconds for the SVC with $10,000$ iterations after a burn-in of $10,000$ that is crucial for the MCMC algorithm.[23] Therefore, the HDCM calibrates the CMAQ numerical model output at the BTH region better than other competitive models without increasing the computation costs.

17

4. Limitations(not rewritten)

There are some limitations to consider in our study. Firstly, for all methods, including competing models and our model, the prediction results of some extreme heavy pollutions are not so satisfactory (e.g., the sixth column in Figure 7), which implies that we need further understanding of transmission processes of $PM_{2.5}$ and to consider those processes by combining with relevant events. Secondly, HDCM can not automatically capture nonlinear patterns from data, whereas the environmental datasets show that $PM_{2.5}$ values are highly nonlinear with respect to the meteorological variables.[46] Moreover, for some linear SVC models, there is significant confounding both among and between covariates and spatial/spatio-temporal effects; that is why the SVC models typically get worse as additional variables sequentially coming into the model.[72] On the other hand, in terms of our model, although interpretations of nonlinearity patterns and the issues around confounding among those factors remain to be better addressed, the data analysis of China's BTH region shows that HDCM has the potential to reduce the bias between ground observations and CMAQ outputs by considering deeper hierarchies for spatio-temporal pattern, and demonstrates that HDCM has the ability to partially alleviate the aforementioned spatial confounding problems compared to SVC.

Associated Content

Acknowledgement

Please use "The authors thank . . . " rather than "The authors would like to thank . . . ".

The author thanks Mats Dahlgren for version one of achemso, and Donald Arseneau for the code taken from cite to move citations after punctuation. Many users have provided feedback on the class, which is reflected in all of the different demonstrations shown in this document.

Supporting Information Available

References

(1) Li, Y.; Zhu, Y.; Yin, W.; Liu, Y.; Shi, G.; Han, Z. Prediction of high resolution spatial-temporal air pollutant map from big data sources. *International Conference On Big Data Computing And Communications, Taiyuan, China* **2015**, 273–282.

(2) Guan, Y.; Johnson, M. C.; Katzfuss, M.; Mannshardt, E.; Messier, K. P.; Reich, B. J.; Song, J. J. Fine-scale spatiotemporal air pollution analysis using mobile monitors on Google Street View vehicles. *J. Am. Stat. Assoc.* **2020**, *115*, 1111–1124.

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

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

(5) Borrego, C.; Miranda, A. I.; Eds, *Air pollution modeling and its application XIX*; Springer Science & Business Media, 2008.

(6) 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.; .et al, Description and evaluation of the Community Multiscale Air Quality (CMAQ) modeling system version 5.1. *Geosci. Model. Dev.* **2017**, *10*, 1703–1732.

(7) Kang, D.; Mathur, R.; Trivikrama Rao, S. Assessment of bias-adjusted $PM_{2.5}$ air quality forecasts over the continental United States during 2007. *Geosci. Model. Dev.* **2010**, *3*, 309–320.

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

(9) Talagrand, O.; Courtier, P. Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory. *Q. J. Roy. Meteor. Soc.* **1987**, *113*, 1311–1328.

(10) Evensen, G. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.:Oceans.* **1994**, *99*, 10143–10162.

(11) Lopez-Restrepo, S.; Yarce, A.; Pinel, N.; Quintero, O.; Segers, A.; Heemink, A. Forecasting PM_{10} and $PM_{2.5}$ in the Aburrá Valley (Medellín, Colombia) via EnKF based data assimilation. *Atmos. Environ.* **2020**, *232*, 117507.

(12) Wu, H.; Zheng, X.; Zhu, J.; Lin, W.; Zheng, H.; Chen, X.; Wang, W.; Wang, Z.; Chen, S. X. Improving $PM_{2.5}$ Forecasts in China Using an Initial Error Transport Model. *Environ. Sci. Technol.* **2020**, *54*, 10493–10501.

(13) Crooks, J. L.; Özkaynak, H. Simultaneous statistical bias correction of multiple $PM_{2.5}$ species from a regional photochemical grid model. *Atmos. Environ.* **2014**, *95*, 126–141.

(14) Jiang, X.; Enki Yoo, E.-H. Modeling Wildland Fire-Specific $PM_{2.5}$ Concentrations for Uncertainty-Aware Health Impact Assessments. *Environ. Sci. Technol.* **2019**, *53*, 11828–11839.

(15) Berrocal, V. J.; Gelfand, A. E.; Holland, D. M. A spatio-temporal downscaler for output from numerical models. *J. Agr. Biol. Envir. St.* **2010**, *15*, 176–197.

(16) Berrocal, V. J.; Gelfand, A. E.; Holland, D. M. A bivariate space-time downscaler under space and time misalignment. *Ann. Appl. Stat.* **2010**, *4*, 1942.

(17) Wang, M.; Sampson, P. D.; Hu, J.; Kleeman, M.; Keller, J. P.; Olives, C.; Szpiro, A. A.; Vedal, S.; Kaufman, J. D. Combining land-use regression and chemical transport modeling in a spatiotemporal geostatistical model for ozone and $PM_{2.5}$. *Environ. Sci. Technol.* **2016**, *50*, 5111–5118.

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

(19) Guillas, S.; Bao, J.; Choi, Y.; Wang, Y. Statistical correction and downscaling of chemical transport model ozone forecasts over Atlanta. *Atmos. Environ.* **2008**, *42*, 1338–1348.

(20) Cheng, F. Y.; Feng, C. Y.; Yang, Z. M.; Hsu, C. H.; Chan, K. W.; Lee, C. Y.; Chang, S. C. Evaluation of real-time $PM_{2.5}$ forecasts with the WRF-CMAQ modeling system and weather-pattern-dependent bias-adjusted $PM_{2.5}$ forecasts in Taiwan. *Atmos. Environ.* **2021**, *244*, 117909.

(21) Kennedy, M. C.; O'Hagan, A. Bayesian calibration of computer models. *J. R. Stat. Soc. B* **2001**, *63*, 425–464.

(22) McMillan, N. J.; Holland, D. M.; Morara, M.; Feng, J. Combining numerical model output and particulate data using Bayesian space-time modeling. *Environmetrics* **2010**, *21*, 48–65.

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

(24) Kalman, R. E. A new approach to linear filtering and prediction problems. *J. Basic Eng.* **1960**, *82*, 35–45.

(25) Katzfuss, M.; Stroud, J. R.; Wikle, C. K. Understanding the ensemble Kalman filter. *Am. Stat.* **2016**, *70*, 350–357.

(26) Cressie, N.; Johannesson, G. Fixed rank kriging for very large spatial data sets. *J. R. Stat. Soc. B* **2008**, *70*, 209–226.

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

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

(29) Lindgren, F.; Rue, H.; Lindström, J. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *J. R. Stat. Soc. B* **2011**, *73*, 423–498.

(30) Zhang, B.; Sang, H.; Huang, J. Z. Full-scale approximations of spatio-temporal covariance models for large datasets. *Stat. Sinica.* **2015**, 99–114.

(31) Datta, A.; Banerjee, S.; Finley, A. O.; Gelfand, A. E. Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. *Journal of the American Statistical Association* **2016**, *111*, 800–812.

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

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

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

(35) Sang, H.; Huang, J. Z. A full scale approximation of covariance functions for large spatial data sets. *J. R. Stat. Soc. B* **2012**, *74*, 111–132.

(36) Jurek, M.; Katzfuss, M. Multi-resolution filters for massive spatio-temporal data. *J. Comput. Graph. Stat.* **2021**, 1–37.

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

(38) Houtekamer, P. L.; Mitchell, H. L.; Pellerin, G.; Buehner, M.; Charron, M.; Spacek, L.; Hansen, B. Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations. *Mon. Weather Rev.* **2005**, *133*, 604–620.

(39) Stroud, J. R.; Stein, M. L.; Lesht, B. M.; Schwab, D. J.; Beletsky, D. An ensemble Kalman filter and smoother for satellite data assimilation. *J. Am. Stat. Assoc.* **2010**, *105*, 978–990.

(40) Buehner, M.; McTaggart-Cowan, R.; Beaulne, A.; Charette, C.; Garand, L.; Heilliette, S.; Lapalme, E.; Laroche, S.; Macpherson, S. R. M.; Morneau, J.; Zadra, A. Implementation of deterministic weather forecasting systems based on ensemble-variational data assimilation at Environment Canada. Part I: The global system. *Mon. Weather Rev.* **2015**, *143*, 2532–2559.

(41) Zhang, F.; Weng, Y. Predicting hurricane intensity and associated hazards: A five-year real-time forecast experiment with assimilation of airborne Doppler radar observations. *Bull. Am. Meteorol. Soc.* **2015**, *96*, 25–33.

(42) Katzfuss, M.; Stroud, J. R.; Wikle, C. K. Ensemble Kalman methods for high-dimensional hierarchical dynamic space-time models. *J. Am. Stat. Assoc.* **2019**, 1–43.

(43) Houtekamer, P. L.; Mitchell, H. L. Data assimilation using an ensemble Kalman filter technique. *Mon. Weather Rev.* **1998**, *126*, 796–811.

(44) Furrer, R.; Bengtsson, T. Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants. *J. Multivar. Anal.* **2007**, *98*, 227–255.

(45) Houtekamer, P. L.; Zhang, F. Review of the ensemble Kalman filter for atmospheric data assimilation. *Mon. Weather Rev.* **2016**, *144*, 4489–4532.

(46) Liang, X.; Zou, T.; Guo, B.; Li, S.; Zhang, H.; Zhang, S.; Huang, H.; Chen, S. X. Assessing Beijing's $PM_{2.5}$ pollution: severity, weather impact, APEC and winter heating. *P. Roy. Soc. A-Math. Phy.* **2015**, *471*, 20150257.

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

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

(49) China's State Council. The long-term "Beautiful China" targets through 2035 (in Chinese). `http://www.gov.cn/xinwen/2021-02/22/content_5588304.htm`, 2021.

(50) Lu, X.; Zhang, S.; Xing, J.; Wang, Y.; Chen, W.; Ding, D.; Wu, Y.; Wang, S.; Duan, L.; Hao, J. Progress of air pollution control in China and its challenges and opportunities in the ecological civilization era. *Engineering* **2020**,

(51) Zhang, L.; Shao, J.; Lu, X.; Zhao, Y.; Hu, Y.; Henze, D. K.; Liao, H.; Gong, S.; Zhang, Q. Sources and processes affecting fine particulate matter pollution over north China: an adjoint analysis of the Beijing APEC period. *Environ. Sci. Technol.* **2016**, *50*, 8731–8740.

(52) Cressie, N.; Wikle, C. K. *Statistics for spatio-temporal data*; John Wiley & Sons, 2011.

(53) Smith, R. L.; Kolenikov, S.; Cox, L. H. Spatiotemporal modeling of $PM_{2.5}$ data with missing values. *J. Geophys. Res-Atmos.* **2003**, *108*.

(54) Sahu, S. K.; Gelfand, A. E.; Holland, D. M. Spatio-temporal modeling of fine particulate matter. *J. Agr. Biol. Envir. St.* **2006**, *11*, 61.

(55) Sampson, P. D.; Richards, M.; Szpiro, A. A.; Bergen, S.; Sheppard, L.; Larson, T. V.; Kaufman, J. D. A regionalized national universal kriging model using Partial Least Squares regression for estimating annual $PM_{2.5}$ concentrations in epidemiology. *Atmos. Environ.* **2013**, *75*, 383–392.

(56) Lindgren, F.; Rue, H. Bayesian spatial modelling with R-INLA. *J. Stat. Softw.* **2015**, *63*, 1–25.

(57) Bolin, D.; Wallin, J.; Lindgren, F. Latent Gaussian random field mixture models. *Comput. Stat. Data. An.* **2019**, *130*, 80–93.

(58) Rue, H.; Held, L. *Gaussian Markov random fields: theory and applications*; Chapman and Hall/CRC, 2005; pp 15–130.

(59) Wan, Y.; Xu, M.; Huang, H.; Xi Chen, S. A spatio-temporal model for the analysis and prediction of fine particulate matter concentration in Beijing. *Environmetrics* **2021**, *32*, e2648.

(60) Blei, D. M.; Kucukelbir, A.; McAuliffe, J. D. Variational inference: A review for statisticians. *J. Am. Stat. Assoc.* **2017**, *112*, 859–877.

(61) Ren, Q.; Banerjee, S.; Finley, A. O.; Hodges, J. S. Variational Bayesian methods for spatial data analysis. *Comput. Stat. Data. An.* **2011**, *55*, 3197 – 3217.

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

(63) Wang, Y.; Du, Y.; Wang, J.; Li, T. Calibration of a low-cost $PM_{2.5}$ monitor using a random forest model. *Environ. Int.* **2019**, *133*, 105161.

(64) Zhao, C.; Wang, Q.; Ban, J.; Liu, Z.; Zhang, Y.; Ma, R.; Li, S.; Li, T. Estimating the daily $PM_{2.5}$ concentration in the Beijing-Tianjin-Hebei region using a random forest model with a $0.01° \times 0.01°$ spatial resolution. *Environ. Int.* **2020**, *134*, 105297.

(65) Berrocal, V. J.; Guan, Y.; Muyskens, A.; Wang, H.; Reich, B. J.; Mulholland, J. A.; Chang, H. H. A comparison of statistical and machine learning methods for creating national daily maps of ambient $PM_{2.5}$ concentration. *Atmos. Environ.* **2020**, *222*, 117130.

(66) Finley, A. O.; Banerjee, S. Bayesian spatially varying coefficient models in the spBayes R package. *Environ. Modell. Softw.* **2020**, *125*, 104608.

(67) Gneiting, T.; Raftery, A. E. Strictly proper scoring rules, prediction, and estimation. *J. Am. Stat. Assoc.* **2007**, *102*, 359–378.

(68) Hamm, N.; Finley, A.; Schaap, M.; Stein, A. A spatially varying coefficient model for mapping PM10 air quality at the European scale. *Atmos. Environ.* **2015**, *102*, 393–405.

(69) Bass, M. R.; Sahu, S. K. Dynamically Updated Spatially Varying Parameterizations of Hierarchical Bayesian Models for Spatial Data. *J. Comput. Graph. Stat.* **2019**, *28*, 105–116.

(70) Murphy, A. H.; Brown, B. G.; Chen, Y.-S. Diagnostic verification of temperature forecasts. *Weather Forecast.* **1989**, *4*, 485–501.

(71) Wilks, D. S. *Statistical Methods in the Atmospheric Sciences*; Academic Press, 2011.

(72) Hefley, T. J.; Hooten, M. B.; Hanks, E. M.; Russell, R. E.; Walsh, D. P. The Bayesian group lasso for confounded spatial data. *J. Agric. Biol. Environ. Stat.* **2017**, *22*, 42–59.