Atmospheric Environment 72 (2013) 60–69

Contents lists available at SciVerse ScienceDirect

Atmospheric Environment

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

Time-space Kriging to address the spatiotemporal misalignment in the large datasets

Dong Liang [a], Naresh Kumar [b,*]

[a] Department of Epidemiology, University of Iowa, Iowa City, IA 52242, USA
[b] Department of Epidemiology and Public Health, University of Miami, 1425 NW 10 Ave, Suite 308C, Miami, FL 33136, USA

HIGHLIGHTS

- Novel method of interpolation across space and time.
- Computationally efficient method to estimate exposure at a given location and time.
- Hierarchical time-space Kriging.

ARTICLE INFO

Article history:
Received 30 July 2012
Received in revised form
10 February 2013
Accepted 19 February 2013

Keywords:
Time-space Kriging
Spatiotemporal hierarchical model
Gaussian Markov Random Fields
Nonstationarity
Bayesian computation
Fine particulate matter $PM_{2.5}$

ABSTRACT

This paper presents a Bayesian hierarchical spatiotemporal method of interpolation, termed as Markov Cube Kriging (MCK). The classical Kriging methods become computationally prohibitive, especially for large datasets due to the $O(n^3)$ matrix decomposition. MCK offers novel and computationally efficient solutions to address spatiotemporal misalignment, mismatch in the spatiotemporal scales and missing values across space and time in large spatiotemporal datasets. MCK is flexible in that it allows for non-separable spatiotemporal structure and nonstationary covariance at the hierarchical spatiotemporal scales. Employing MCK we developed estimates of daily concentration of fine particulates matter ≤ 2.5 μm in aerodynamic diameter ($PM_{2.5}$) at 2.5 km spatial grid for the Cleveland Metropolitan Statistical Area, 2000 to 2009. Our validation and cross-validation suggest that MCK achieved robust prediction of spatiotemporal random effects and underlying hierarchical and nonstationary spatiotemporal structure in air pollution data. MCK has important implications for environmental epidemiology and environmental sciences for exposure quantification and collocation of data from different sources, available at different spatiotemporal scales.

1. Introduction

Blurring disciplinary boundaries and converging time and space domains are important developments for interdisciplinary multi-level (or hierarchical) research. While such research can help us understand complex patterns of socio-physical phenomena and the processes that operate at hierarchical spatiotemporal scales and shape them, it requires data from multiple sources and poses four major challenges. First, the size of these data is large. For example, aerosol optical depth (AOD) data for Cleveland MSA between 2000 and 2009 included 2.3 million data points. Second, these data suffer from spatiotemporal misalignment because the location and time stamps of these data do not align well (see Fig. 1a; (Kumar, 2012)). For example, to investigate health effects of air pollution we need air pollution and health data. The location and time stamp of air pollution data (monitored at sparsely distributed EPA sites) does not perfectly align (or correspond) with the location and time stamp of (aggregated and disaggregated or individual level) health data. Thus, environmental data need to be imputed at the spatiotemporal stamp of health data in order to address the misalignment problem. Third, spatiotemporal scales/resolutions of these data are different (Fig. 1b). This means that spatiotemporal scales at which these data are aggregated and reported are not the same. For example air pollution data are monitored at sparsely distributed sites every day (every hour at some sites), and mortality data are aggregated by county and day. Fourth, these data suffer from missing values across geographic space and time. For example, there are systematic gaps in air pollution monitoring; likewise there are systematic gaps in the retrieval of AOD from satellites due to cloud cover and cloud contamination.

* Corresponding author. Tel.: +1 305 243 4854.
E-mail addresses: dong.m.liang@gmail.com (D. Liang), nkumar@med.miami.edu (N. Kumar).
URL: http://eph.ccs.miami.edu

1352-2310/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.atmosenv.2013.02.034

The analysis of time-space varying datasets that come from different sources requires that these data are: a) aligned with respect to location and time, b) arranged on the same spatiotemporal scales and c) missing values are filled. For example, we need to estimate exposure using the existing air pollution data at the spatiotemporal scale of mortality data in order to evaluate the association between air pollution and mortality. Finest spatial resolution of mortality data is point location (i.e. street address of decedents) and the temporal scale is the date of mortality. Daily exposure estimates are needed several days prior to the date of death (for time-lagged exposure) at the location of residence (and potentially at all other locations where decedents have spent some time) for each case or these data need to be aggregated to coarser spatiotemporal scale. Likewise, the spatiotemporal scales of different environmental datasets are not the same. Thus, imputing one environmental dataset at the spatiotemporal scales of other environmental dataset is critically important to collocate different environmental datasets. If adequate data points, spread across geographic space and time, are available, different methods of interpolation can be employed to impute value at a given location and time. Among these methods, time-space Kriging is an attractive option because it minimizes the mean squared prediction errors among linear unbiased predictors.

Although time-space Kriging is a relatively newer development, spatial Kriging has been in practice for a while. Given a random process $\{Z(\boldsymbol{s}):\boldsymbol{s}\in D\}$; where D is the spatial domain and $\boldsymbol{s}$ is the location represented by a pair of coordinates, Kriging relies on the assumption of spatial stationarity (i.e. constant variance within domain D), which assumes that the process satisfies the condition $E(Z(\boldsymbol{s}))=\mu$, for all $\boldsymbol{s}\in D$ and $cov(Z(\boldsymbol{s}_1),Z(\boldsymbol{s}_2))=C(\boldsymbol{s}_1-\boldsymbol{s}_2)$ for all $\boldsymbol{s}_1,\boldsymbol{s}_2\in D$. Kriging requires an understanding of how a quantity varies with respect to those observed at the neighboring sites (located at different distance intervals) to develop an empirical model to estimate covariance. Based on the covariance function, Kriging has been used extensively for developing continuous surface (i.e. prediction at every location within the study domain) of environmental variables, such as distribution of air pollution, minerals, soil characteristics and meteorological conditions (Cressie, 1993, 2008).

The extension of spatial Kriging to the spatiotemporal domain is straightforward. Time-space Kriging requires proper specification of spatiotemporal covariance functions. According to the classical theory, separable and stationary covariance assumptions require that the process satisfies the condition $E(Z(\boldsymbol{s},t))=\mu$, for all $\boldsymbol{s}\in D$, $t\in[1,T]$ and $cov(Z(\boldsymbol{s}_1,t_1),Z(\boldsymbol{s}_2,t_2))=C(\boldsymbol{h},u)=C_s(\boldsymbol{h})C_t(u)$ where $\boldsymbol{h}=\boldsymbol{s}_1-\boldsymbol{s}_2$ and $u=t_1-t_2$ for all $\boldsymbol{s}_1,\boldsymbol{s}_2\in D$ and $t_1,t_2\in\{1,\ldots,T\}$; C_s and C_t are spatial and temporal only stationary covariance functions. Satisfying these assumptions can be difficult because the inherent differences in spatial and temporal scales of data are likely to produce nonstationary covariance when time and space domains converge. To develop robust time-space Kriging model, we face three important challenges, especially for large datasets: a) non-separable covariance across time and space, b) nonstationary covariance at multiple spatiotemporal scales, and c) computational issues. Researchers have begun to address some of these challenges.

Time-space Kriging requires the specifications of spatial, temporal and non-separable spatiotemporal covariance. Spatial and temporal covariance can be constructed using spatial and temporal trends of the data separately. Non-separable spatiotemporal covariance emerges due to the convergence of spatial and temporal domains. Researchers suggest the use of product sum model (De Cesare et al., 2001) and integration of spectral densities (Cressie and Huang, 1999) to address non-separable spatiotemporal

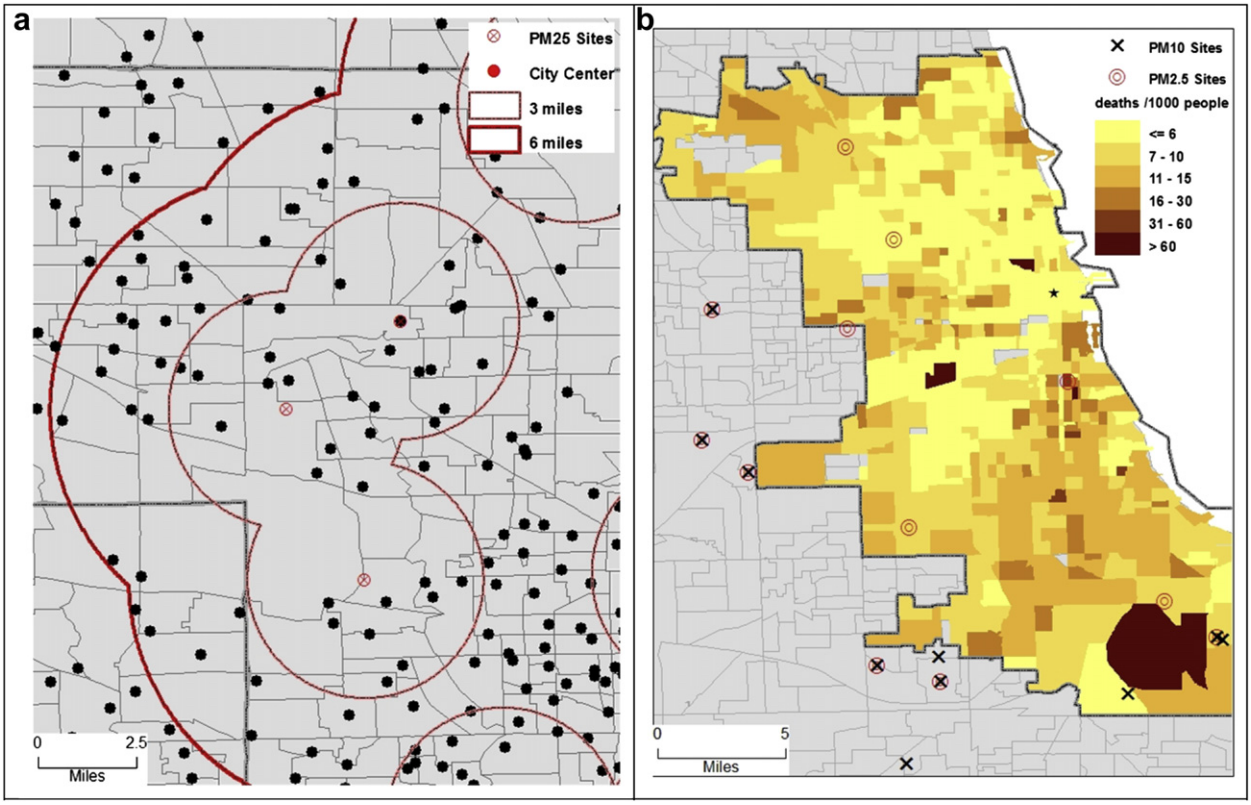

Fig. 1. a. An example of *Spatial Misalignment* in the locations of EPA data monitored at point locations and the point locations of health data, where one might need exposure. b: An example of *Mismatch in the geographic scales* of EPA data (at point locations) and mortality data (aggregated at Census Tract level). Source: Kumar et al., 2011a,b. Data were acquired from the Chicago Department of Public Health.

covariance. Since the rate of spatiotemporal trend can vary regionally, seasonally, and across local spatiotemporal subdomains, the convergence of spatiotemporal domains also results in nonstationarity covariance at multiple spatiotemporal scales. The first order (at global scale) nonstationarity can be handled by incorporating covariates and/or non-linear spatiotemporal trends (De Iaco et al., 2002; Haas, 1995). The effectiveness of such an approach largely depends on the robustness of covariates and/or spatiotemporal trends incorporated into the model.

Given the inherent regional and seasonal structures in the environmental data, nonstationarity needs to be modeled at multiple spatiotemporal scales separately. For example, diurnal variability in air pollution can be different from seasonal variability; likewise, air pollution variation within a neighborhood (within a few miles) can be different from that within and between regions, such as within and between urban and sub-urban areas (Kumar et al., 2011a). Given this hierarchical spatiotemporal structure in environmental data, the second-order nonstationarity at multiple spatiotemporal scales is important for time-space Kriging. Nonstationarity spatiotemporal covariance can be specified by basis function approximations and Kalman filtering (Cressie et al., 2010). Among other alternatives, a computationally efficient hierarchical model can be used to address nonstationarity at multiple spatiotemporal scales (Fuentes et al., 2005; Sahu et al., 2006).

Time-space Kriging also poses computational challenges due to *the big n problem* (Banerjee et al., 2004). Researchers address this challenge through approximating a large process by its realizations on a small set of knots (or subset of data) (Banerjee et al., 2008). An alternative approach involves fitting a Gaussian Markov Random Fields (GMRF) to the original process on a lattice in order to utilize sparse matrix algorithms (Hartman and Hossjer, 2008). This approach is useful for regularly spaced data where projection onto a lattice can be made. But it is of little use for irregular spatiotemporal data (such as deviating path of Terra and Aqua satellite every day, before repeating every 16th day). Cressie and Johannesson present an efficient solution by approximation of covariance using linear combinations of small number of low-dimensional deterministic (or basis) functions (2008).

In this paper, we present a flexible and computationally efficient Markov spatiotemporal Cube (or voxel) Kriging (MCK) to address the above outlined challenges. MCK simultaneously addresses modeling and computation challenges. The flexibility of the model lies in the fact that it incorporates varying covariance at multiple spatiotemporal scales and separable and non-separable spatiotemporal covariance. MCK is a Bayesian hierarchical model where GMRF priors are utilized to achieve the model richness, such as the use of multi-level interactions across space and time. We can also use Bayesian methodology to incorporate covariates, time-space heterogeneity, and physically meaningful prior information. MCK also utilizes structures within GMRF priors for developing an efficient solution to *the big n problem*. To demonstrate the application of this method, we utilize satellite based air pollution data (on fine particulate matter $\leq$2.5 μm in aerodynamic diameter ($PM_{2.5}$)) for Cleveland MSA from 2000 to 2009. These data exhibit strong spatial dependency, spatial nonstationarity, weak temporal autocorrelation and extensive gaps across time and space. We also created artificial gaps in the dataset for cross-validation. These gaps can be treated as the locations and times for which exposure estimates are needed. The remainder of this paper is organized into four sections. The first section presents the statistical formulation of MCK, followed by the second section that details the implementation of MCK. The third section includes results and the fourth section discusses our results with the relevant literature.

2. Statistical model

Let $\{Z(s,t)\}$ denote the observed Gaussian process defined over $D \times \{1,2,\ldots,T\}$ where $D \subset R^2$ denotes the spatial domain and t indexes discrete time stamps. The term cube (voxel) in the context of this paper refers to as tessellation (or partitioning) of the data points into spatiotemporal subsets. We partition D into small areas, termed as regions. Likewise, we partition time T into L subsets, known as temporal intervals. Utilizing the tessellation, we decompose the spatiotemporal domain hierarchically into sub-spatiotemporal domains, termed as spatiotemporal cubes. We then utilize hierarchical Gaussian Markov Random Fields (GMRF) to capture spatiotemporal interactions at *intra-* and *inter*-spatiotemporal cube level. The highly structured GMRF model also facilitates Bayesian estimation for large dataset.

In practice, the process Z is only observed at n spatiotemporal locations denoted by $\{(\boldsymbol{s_i},t_i), i = 1,\ldots,n\}$. Define $\boldsymbol{Z_t} = [Z(\boldsymbol{s}_1,t),\ldots,Z(\boldsymbol{s_{nt}},t)]^T$; where n_t represents observed data at time t. Define the m spatiotemporal locations of health data. We term as the query points where predictions are needed using n_t data points. Our model aims to predict the un-observable at these m spatiotemporal locations with minimum error.

The first stage of the model takes the form

$$\boldsymbol{Z}_t = \boldsymbol{X}_t^T\boldsymbol{\beta} + \boldsymbol{v}_t + \boldsymbol{e}_t + \varepsilon_{\boldsymbol{t}}$$

where $\boldsymbol{X}_t = [X(s_1,t)^T,\ldots,X(\boldsymbol{s_{nt}},t)^T]^T$ denote the $n_t \times p$ design matrix for covariates, The random vector $\boldsymbol{v}_t$ and $\boldsymbol{e}_t$ model between cube and within cube variation, and ϵ_t denote measurement errors that are white noises.

Our specification of cube model involves selection of R spatial locations over D (the study area). Voronoi tessellation can be employed to partition spatial domain D into non-overlapping regions (R). Likewise, temporal domain can be partitioned into non-overlapping L time intervals. The spatiotemporal domain is partitioned into $R \times L$ cubes. Partitioning spatiotemporal domain into ($R \times L$) cubes is integral component to make MCK computationally efficient and to capture nonstationarity at local and cube levels. Different approaches are suggested for partitioning spatial domain, which includes: minimizing variance within sub-regions and variance maximization between sub-regions (Kumar et al., 2011b), overlaying a balanced sample of points (Stevens and Olsen, 2004) and minimizing prediction errors using model based approaches (Diggle and Lophaven, 2006; Zhu and Stein, 2006; Zimmerman, 2006). Borrowing the strengths from the literature on spatial partitioning, we propose overlaying systematic grid onto the study domain D to partition it into R regions (termed as spatiotemporal cubes), and then regularly space time points on each grid cell. The systematic partitioning of geographic space reduces spatial autocorrelation and facilitate parameter estimation (Stevens and Olsen, 2004).

Let $r(s_i)$ denote the region identifier for location s_i and $l(t_i)$ the interval for time stamp t_i. Utilizing the tessellation, we can decompose the process $v(s_i,t_i)$ additively (Schmid and Held, 2004) as

$$v(s_i,t_i) = \phi(r(s_i)) + \theta(l(t_i)) + \delta(r(s_i),l(t_i))$$

where the term ϕ represents un-observed regional effects, denoting factors affecting observations only over space; the term θ models un-observed temporal effects; the term δ captures residual stochastic variations that can interact across spatiotemporal cubes. Details on each of these spatiotemporal and local effects are provided in Supplementary Online Material (SOM): Appendix – A.

For regional effects ϕ, we adopt a conditional autoregressive prior (Besag et al., 1991). For the regional temporal trends, we adopt a first-order random walk GMRF (Clayton, 1996). For the interaction effects δ, we construct a GMRF on the spatiotemporal domain by using the Kronecker product of the structure matrices of ϕ and $\boldsymbol{\theta}$ (Clayton, 1996; Rue and Held, 2006).

$$\pi(\boldsymbol{\delta}|\tau_\delta) \propto \tau_\delta^{\frac{n_\theta n_\phi}{2}} \exp\Big\{ -\frac{\tau_\delta}{2}\boldsymbol{\delta}^T\Big(\boldsymbol{Q}_\theta \otimes \boldsymbol{Q}_\phi\Big)\boldsymbol{\delta}\Big\}$$

where τ_δ denotes the unknown scale parameters, $n_\phi = R - 1$ and $n_\theta = L - 1$ denote the ranks of structure matrices for regional effect (Q_ϕ) and temporal trend (Q_θ) and $\otimes$ denotes the Kronecker product operator.

Local effects $\mathbf{e_t}$ represent variation within a cube. We assume that within the cube, temporal dynamic are negligible. We partition each cube using Voronoi tessellation. We model the within cube variation using ξ_t, an $n_t + m_t$ random vector. We propose intrinsic CAR model on ξ_t, where adjacency is defined through sharing a boundary between the Voronoi polygons. We assume that

$$e_t = K_t\xi_t \quad \text{where} \quad K_t = \left[I_{n_t} 0_{n_t \times m_t}\right]$$

where K_t maps the data to local effects. In the general case, the above model is specified across each cube. We allowed the CAR scaling parameter to be independent across cubes, enforcing local smoothing.

We specify a spatiotemporal CAR prior on the local effects due to computational and modeling issues of large datasets. First, this conditional specification allows efficient computation. Instead of iteratively factoring dense matrices, our implementation involves diagonalization of the fixed adjacency matrices only once during the setup. The resulting Bayesian computation becomes feasible for large datasets. Second, the accuracy of GMRF approximation to Gaussian field over regular grid has been validated for stationary Gaussian field (Rue and Tjelmeland, 2002). The correlation approximated by the nearest neighbor structure is not restricted to stationary or separable family, similar to basis function approximations (Cressie and Johannesson, 2008).

We allow for cube specific scale parameters (τ_{rl}) for local spatial effects ($\boldsymbol{\xi_{rl}}$) and white noise (λ_{rl}). While such richly parameterized model allows for local non-parametric smoothing of the data to predict missing values, the model requires efficient algorithms to derive Bayesian inference. The Bayesian computation details are outlined in the implementation section. All scale parameters are assigned hyper-priors to complete the Bayesian model specification. We adopt semi-conjugate Gamma priors for all scale parameters

$$\tau_x \sim Gamma\left(\frac{n_x}{2}, \frac{n_x}{2s_x^2}\right) \quad \text{where} \quad x \text{ stands for } \theta, \phi, \delta, \tau_{rl}, \lambda_{rl}$$

where n_x is small so that the hyper prior is uninformative and s_x is chosen such that resulting prior with the 95% credible sets are wide enough to cover possible range of the scale parameters. These hyper-priors are chosen to be proper, yet uninformative, which allows data to dominate the inference.

3. Implementation

3.1. Tessellation of spatiotemporal domain

The EPA data are monitored at sparsely distributed seven sites in the Cleveland MSA. Thus, these data are not sufficient to estimate spatial co-variance function. Therefore, we utilize 2.3 million data points of aerosol optical depth (AOD) retrieved using NASA's satellite between 2000 and 2009 and indirect estimates of airborne particulate $\leq$2.5 μm in aerodynamic diameter ($PM_{2.5}$) using AOD data. AOD was retrieved at 2 km spatial resolution using the same algorithm as for the standard 10 km AOD from NASA (Levy et al., 2007), but with more restrictive criteria (see Kumar et al. (2011a) for details). AOD data represent ~10:30 AM for Terra and ~1:30 PM Aqua satellites local time, respectively. A hybrid approach was employed to develop $PM_{2.5}$ (Kumar et al., 2011a). Both AOD and $PM_{2.5}$ datasets were used to implement and validate MCK.

Given the large dataset and nonstationary spatiotemporal covariation in AOD and $PM_{2.5}$, partitioning these data into non-overlapping spatiotemporal cubes and identifying intra- and inter-cube interactions becomes important. A balance needs to be achieved between computational efficiency and spatiotemporal scales of decomposition because the finer scale decomposition will results in a large number of spatiotemporal cubes and will increase the computation burden. Given the strong spatial autocorrelation in AOD and $PM_{2.5}$ data, we decompose spatial domain by overlaying a 10 × 10 systematic grid. This overlay ensures a spatially balanced representation of these data. Because of high diurnal variability in these data, we do not partition these data into sub-temporal domains, and utilize the finest available temporal scale of these data, i.e. every day.

3.2. Markov chain Monte Carlo algorithm

We propose a hierarchical model to predict value for the query points (i.e. points across space and time where predictions are needed)

$$\boldsymbol{Z}_{mt} = \boldsymbol{X}_{mt}\boldsymbol{\beta} + \theta_t^* 1_{n_t} + \boldsymbol{K}_{\phi,t}\left(\phi^* + \boldsymbol{\delta}_t^*\right) + \boldsymbol{K}_{\xi,t}\xi_t + \varepsilon_{\boldsymbol{t}}$$

where $\boldsymbol{Z}_{mt}$ denotes the underlying true values; the '*' above a given parameter indicates appropriate sum-to-zero constraint and $\boldsymbol{X_{mt}}$ is $m_t \times p$ design matrix. The K matrices are mapping matrices of appropriate dimension. Specifically, $\boldsymbol{K}_{\varphi,t}$ is a $m_t \times R_t$ adjacency matrix where $K_{\varphi,t}[\boldsymbol{s}_{it}^*, r(\boldsymbol{s}_{it}^*)] = 1$ and zero elsewhere. Let $N_t = \sum_{r_t=1}^{R_t} N_{t,r_t}$ denote total local parameters at time t, $\boldsymbol{K}_{\xi,t}$ is an $m_t \times N_t$ adjacency matrix where missing data are associated with the corresponding underlying truth. The error process has diagonal covariance matrix $\boldsymbol{W}(\boldsymbol{\lambda})$. Our prediction rely on the distribution $\pi(\boldsymbol{Z_m}|\boldsymbol{Z})$. Let $\boldsymbol{\Theta}$ denote all random effects and $\boldsymbol{T}$ all variance parameters in the model.

$$\pi(\boldsymbol{Z_m}|\boldsymbol{Z}) = \int\int \pi(\boldsymbol{Z_m}|\boldsymbol{\Theta}, \mathbf{T}, \boldsymbol{Z})\pi(\boldsymbol{\Theta}, \mathbf{T}|\boldsymbol{Z})\mathrm{d}\nu(\boldsymbol{\Theta})\mathrm{d}\nu(\mathbf{T})$$

The above integral can be approximated as a Monte Carlo mixture of the form

$$\widehat{\pi}(\boldsymbol{Z_m}|\boldsymbol{Z}) = \frac{1}{G}\sum_{g=1}^{G} \pi\Big(\boldsymbol{Z_m}\Big|\boldsymbol{\Theta}^{(g)}, \mathbf{T}^{(g)}, \boldsymbol{Z}\Big)$$

where g denotes the iteration number.

Ideally, marginalized posterior $\pi(\boldsymbol{T}|\boldsymbol{Z})$ are appropriate after integrating the random effects (Rue and Held, 2006). This integration requires factoring of a large covariance matrix of large dimension $(n+m)\times(n+m)$, where n denotes observed data points and m query points (where predictions are needed). In addition to large data dimension, such factorization needs to satisfy numerous identifying constraints, which makes the inverse covariance matrix dense and numerical integration of random effects challenging. Thus, we employ block Gibbs sampler to explore the posterior

density. Details of the Gibbs sampler are included in SOM: Appendix – B.

4. Data and results

4.1. Data

For the implementation and validation of MCK we utilize three different datasets: aerosol optical depth (AOD) and $PM_{2.5}$ data at 2 km spatial resolution from December 2007 to January 2009 for Cleveland MSA (Fig. 2a), and data simulated using hierarchical nonstationarity assumption. The methodologies for the extraction of 2 km AOD and prediction of $PM_{2.5}$ using a hybrid approach is detailed elsewhere (Kumar et al., 2011a; Levy et al., 2007). Assuming these data represent the best estimates of AOD and $PM_{2.5}$, a systematic grid, as shown in Fig. 2b, was developed using the proposed MCK.

Validation was performed on the 2008 data with one-month overlap on both ends. Thus, the final dataset included all data points between December 2007 and January 2009. Exploratory analysis provided insight into the important characteristics of these data. First, there were systematic spatiotemporal gaps in these data, because AOD retrievals were not possible every day due to cloud cover and contamination. For example, over a span of 14 months, there were no observations for about 41% of the days. Second, the rate of spatiotemporal variability in AOD and $PM_{2.5}$ was not the same, for example temporal variability in both AOD and $PM_{2.5}$ was greater than the spatial variability. Given the static emission sources and dynamic physical and chemical atmospheric processes, the differences in the rate of spatial and temporal variability were obvious. Moreover, meteorological fields (such as wind velocity, wind direction and relative humidity) can transport aerosols and fine particulates to greater distances within short time span. Therefore, AOD and $PM_{2.5}$ data are likely to show greater temporal heterogeneity and stronger spatial structure.

4.2. Results

Markov Cube Kriging (MCK) was implemented on AOD and $PM_{2.5}$ for the year 2008 ($n \sim 269$ k and 255 k, respectively). Since both AOD and $PM_{2.5}$ were highly skewed, both were transformed to a natural log scale to stabilize variance. For cross-validation purposes, 10% of the data points (15 k and 22 k, respectively were randomly left out and MCK was employed to predict values at the left out points. The study area was partitioned into 100 regions by overlaying a 10×10 grid on Cleveland MSA. This resulted in two levels spatial hierarchy, individual point locations of 2 km AOD and $PM_{2.5}$ and their grouping within 100 regions. The implementation and validation of MCK excluded hierarchical temporal structure because of significant temporal variability within weeks. Cubes that showed a sample standard deviation of 0.1 or less were modeled using exchangeable measurement errors only. Initial values for regional effects were calculated from the observed regional means and the local effects were initialized by maximizing the marginalized density $\pi(u_{r_l}|\boldsymbol{r}_{\boldsymbol{r}_l})$; where $\boldsymbol{r}_{\boldsymbol{r}_l}$ denotes the observed data centered on the regional means.

The hyper-priors for the model were chosen to be proper, yet vague. The ranges of the $\ln(PM_{10})$ and $\ln(PM_{2.5})$ were $\sim$8.45 and $\sim$4.09 $\ln(\mu g\ m^{-3})$ respectively. The scale parameter λ_{r_l} were assigned *Gamma*(*1,0.1*) prior such that the 95% prior credible set for the standard deviation parameters were (0.16, 1.98) on the natural log scale of $\mu g\ m^{-3}$, this interval was wide enough to contain candidate values for the parameters. The local spatial scales $\lambda_{\phi_2}(r_l, l)$ were assigned *Gamma*(*1,1*) hyper-prior. The hyper-priors for inter-cube scales parameters were $\tau_\phi \sim$ *Gamma*(*0.1,0.1*), $\tau_\theta \sim$ *Gamma*(*0.1,0.1*) and $\tau_\delta \sim$ *Gamma*(*0.5,0.2*). One MCMC chain was run for 2100 iterations, discarding the initial 100 as burn-in. The convergence diagnostic was performed using R package coda.

4.3. Cross validation

The robustness of MCK was evaluated using cross validation, in which we Krigged the values for 10% data points skipped randomly. The Krigged (or predicted) values were then compared against the observed values. We ran several Monte Carlo experiments by holding off 10% of the data. We define the cross validation residual $RCV_{js} = Z(s_j, t_j) - E(Z(s_j, t_j)|Z)$ in experiment s. Following Haas (1995), the prediction accuracy was assessed by $bias_s = 1/m \sum_{j=1}^{m} RCV_{js}/\overline{Z_s}$. The average residual over the average of the actual data when and where prediction is needed is denoted by $\overline{Z_s}$. We also used the t-test statistic for H_0:$RCV_{js} = 0$ using $bias_s$ for $s = 1,\ldots,S$. Both $bias_s$ and the t-statistic should be small for unbiased predictors. The accuracy of prediction uncertainty $SE(\widehat{Z}_{js})$

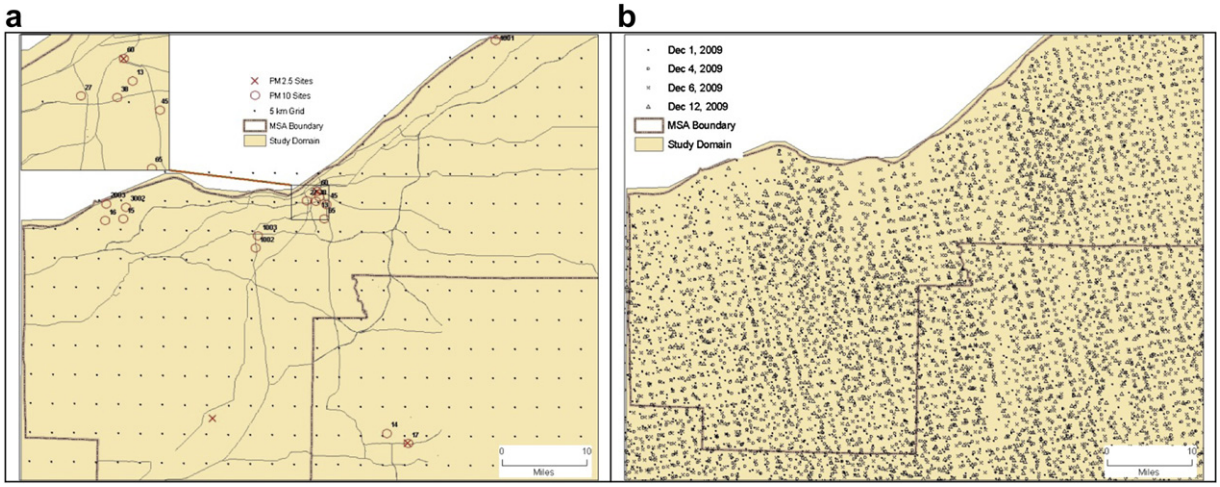

Fig. 2. a: Study Area – Cleveland Metropolitan Statistical Area (MSA), $PM_{2.5}$ and PM_{10} monitoring stations and 2.5 km spatial grid overlaid on the study area. In the proposed study, daily $PM_{2.5}$ was predicted for each grid point from 2000 to 2009. b: Location of the 2 km AOD; using an empirical relationship between PM monitored at EPA sites (in (a)) and satellite based AOD data, the daily $PM_{2.5}$ was predicted for all locations and days AOD was available between 2000 and 2009 (see Kumar et al., 2011a,b for details).

can be assessed by $ratio_s = 1/m\sum_{j=1}^{m} SE^2(\widehat{Z}_{js})/MSE_s$ where $MSE_s = 1/m\sum_{j=1}^{m} RCV_{js}^2$. For unbiased predictors, $ratio_s$ should be close to 1. Let $\widetilde{R}CV_{js} = RCV_{js}/SE(\widehat{Z}_{js})$ denote the standardized cross validation residuals and $var(z_s)$ denote the sample variance of $\widetilde{R}CV_{js}$ for dataset s. The sample variance $var(z_s)$ should also be close to 1 if the standard errors are negligibly biased.

We also implemented a naive spatiotemporal prediction method to predict values at the query points (i.e. 10% of the randomly skipped data points). This involved averaging all observed data points within the chosen spatiotemporal lags; we used 0.10° and 10 days to search for the neighboring data points around a given query point. Table 1 shows that MCK generates robust predictions as compared to the naïve average method. The overall bias is 0.0074 ± 0.0019 and −0.012 ± 0.0019 for ln($PM_{2.5}$) and ln(AOD) respectively. The t test statistics of the null hypothesis that cross validation residuals equal zero are −4.40 and 2.59, respectively. Thus, MCK prediction is slightly negatively biased. The estimated standard errors for ln(AOD) are too small, representing 82% of the MSE. The negative bias for ln(AOD) indicates mis-specification of hyper-priors for local effects. With sparse data, hyper-priors assert little smoothing. Thus, the standard errors are under-estimated. Figs. 3 and 4 show the validation of MCK and naïve averaging method. As expected, MCK outperforms the naïve averaging method in terms of both bias and MSE.

Our analysis suggests that the RMSE for ln($PM_{2.5}$) and ln(AOD) were 0.10 ± 0.0044 and 0.30 ± 0.02, respectively (Table 1 and Figs. 3–4). Both were significant smaller than those estimated using the naïve averaging method. Figs. 3–4 suggests that MCK performs poorly for extremely low values, but for most data points within mean ± 2 standard deviation, (SD) predicted values match nicely with the observed values. About 95% and 93% of the observed (randomly left-out) data points were within 95% intervals of the predicted values for ln($PM_{2.5}$) and ln(AOD) respectively.

4.4. Computational performance

MCK is computationally intensive, for example, a prediction of 2000 and 2500 values (of AOD and $PM_{2.5}$, respectively) using 40,000 and 53,000 observed data points took 4.96 and 8.26 CPU hours, respectively. This time included knot selection and simulation. The experiment was run on a PC with 2.00 GHZ CPU and 24 GB RAM. In principle, MCK can handle much larger number of query points at negligible additional cost because knot selection and regional parameters are estimated only once.

4.5. Application

To demonstrate the application of MCK we developed a systematic grid of daily $PM_{2.5}$ for five different days for three different seasons (Fig. 5). MCK nicely fills the continuous surface. Some caution must be used for the prediction of extremely high values, especially when observed data are very sparse across geographic space. Similar to the prediction surfaces, $PM_{2.5}$ exposure can be computed for any day and location. As the cross-validation suggests most predictions were within 95% confidence intervals. Therefore, eliminating prediction beyond mean ± 2 SD MCK offers a robust solution for developing time-space resolved estimates of air pollution exposure.

4.6. Simulation

Since the $PM_{2.5}$ and AOD data we used for implementation of MCK do not fully follow the MCK assumption and lacked temporal hierarchical structure, a simulation study was performed to evaluate the robustness of MCK. Simulation was conducted for predicting estimates for regular grid and irregularly spaced points. Nine independent datasets ($S = 9$) were simulated using the nonstationary process as detailed below. The spatial domain was a 50 × 50 regular grid. The temporal domain was 18 regularly spaced time points. We divided the spatial domain into 10 × 10 regions, and region was sub-divided overlaying a 5 × 5 grid. Within each region, a non-separable product sum variogram model for the spatiotemporal process

$$\gamma(\boldsymbol{h}, u; parms) = \gamma_s(\boldsymbol{h}; nug_s, H_s, sill_s) + \gamma_t(u; nug_t, H_t, sill_t) - \kappa\gamma_s(\boldsymbol{h}; nug_s, H_s, sill_s) \times \gamma_t(u; nug_t, H_t, sill_t)$$

where *nug*, *H* and *sill* denote the nugget, range and sill parameters (parms) for spatial and temporal variograms. We allowed for nonstationarity by varying the spatial and temporal covariance parameters across regions.

$$nug_s \sim UNIF(0.4, 0.6) \quad H_s \sim UNIF(30, 70) \quad sill_s \sim UNIF(1.5, 2.5)$$

$$nug_t \sim UNIF(1.5, 2.5) \quad H_t \sim UNIF(1, 19) \quad sill_t \sim UNIF(3.5, 4.5)$$

and the interaction parameter $\kappa \sim UNIF(0, [max(nug_s + sill_s, nug_t + sill_t)]^{-1})$ so that the resulting variogram is defined (De Iaco et al., 2001).

Through partitioning the spatial domain, we simulated the spatiotemporal field independent across regions. However, partitioning created artificial edge effects for data points at the boundaries of the regions. Thus, we smoothed the data spatially at each time point using a Gaussian kernel with bandwidth of one lattice unit. We also included large-scale spatiotemporal trend with a non-linear spatial trend

$$M_s(r, c) = \left(0.5 + 0.1r + 0.1c + 0.5r^2 + 0.7c^2 + 0.5rc\right)^{-1}$$

where r and c denote the standardized row and column numbers and seasonal temporal trend

Table 1
Mean and 95% margin of errors (based on 9 Monte Carlo samples) of MCK and AVG using observed ln(AOD), predicted ln($PM_{2.5}$) and simulated data. The results are based holding off 10% of the data. The statistics are proportion of average predicted minus data values over the average of the data values (bias); the Mean Square prediction Error (MSE); the t test statistic (t) for hypothesis H_0: $\mu_b = 0$; the sample variance of standardized cross-validation residuals (VarZ); and the ratio of average square standard errors over MSE (ratio).

	ln(AOD)		ln($PM_{2.5}$)		Simulated	
	MCK	AVG	MCK	AVG	MCK	AVG
Bias	−0.012 ± 0.0019	−0.002 ± 0.0025	0.0074 ± 0.0019	0.0076 ± 0.0034	−0.00043 ± 0.0048	−0.076 ± 0.015
MSE	0.30 ± 0.02	0.85 ± 0.038	0.10 ± 0.0044	0.34 ± 0.038	0.34 ± 0.0039	20 ± 0.068
t	−4.40	−0.53	2.59	1.47	−0.06	−3.21
VarZ	1.5 ± 0.055		1.00 ± 0.018		2.40 ± 0.041	
Ratio	0.82 ± 0.023		1.00 ± 0.014		0.81 ± 0.015	

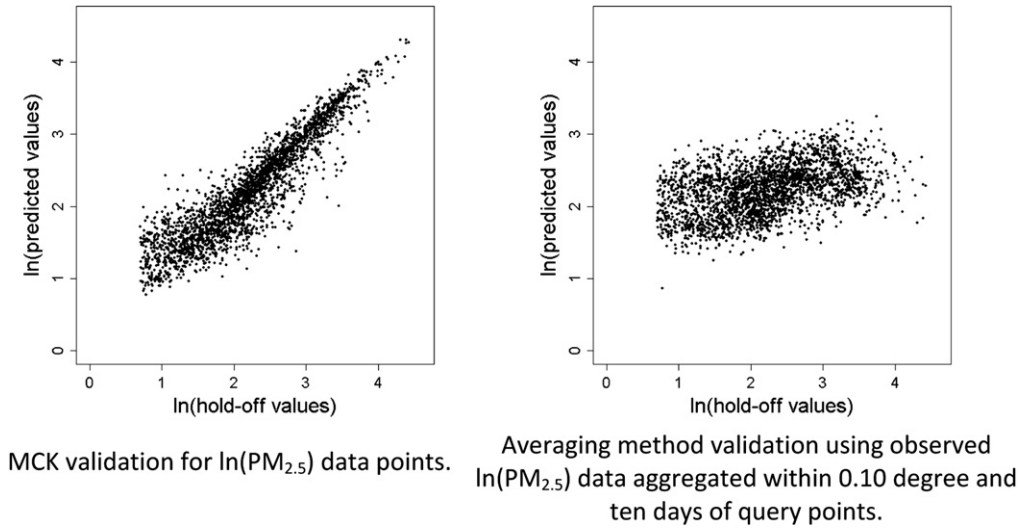

Fig. 3. Validation of MCK and spatiotemporal averaging methods using observed ln($PM_{2.5}$) data between December 2007 and January 2009, Cleveland MSA. 10% of the data points were held-off and values were predicted at these points using MCK method and averaging method.

$$M_t(t) = 4\sin\left(\frac{2\pi t}{6}\right) + 4\cos\left(\frac{2\pi t}{6}\right)$$

where t denotes the discrete time point between 1 and 18. As a result, we simulated S datasets with seasonal and non-linear spatial trends, as well as nonstationary and non-separable local process.

Fig. 6a show the comparison of observed (from the simulated dataset) and predicted values. There are hardly any differences between observed and predicted values. The overall bias was -0.0004 ± 0.0048 and MSE was 0.034 ± 0.0039. The t-test statistic of the null hypothesis H_0:$RCV_{js} = 0$ was 0.06. This means that given the model assumptions (in the observed dataset), MCK can provide robust prediction across geographic space and time and can enable us to develop robust exposure estimates at the spatiotemporal scales of health datasets, and offers an attractive solutions for collocating environmental datasets available at different spatiotemporal scales.

We also conducted a simulation study at irregularly spaced 2500 points within spatiotemporal cubes, rather than on 50×50 grid. Our assumption of nonstationarity and non-separable spatiotemporal covariance function remained the same as in the previous analysis. The data ranged between -10 and 10. Tenfold cross-validation yielded bias = 0.004, root mean square error = 0.43 and correlation 0.99. Like the simulation, MCK simulation at irregularly spaced data points yielded robust estimates (Fig. 6b).

5. Discussion

Time-space Kriging is an important method of prediction that can address multiple problems arising from the convergence of time and space domains – misalignment problem, missing values and mismatch in the spatiotemporal resolutions (Kumar, 2012). This paper presents a theoretical framework of a computationally efficient MCK procedure for predicting a systematic grid of time-space varying quantities. Our analysis suggests that the MCK handled both nonstationarity and large gaps effectively and efficiently.

MCK is a Bayesian hierarchical approach and addresses both the modeling and computational challenges posed by Kriging for large spatiotemporal dataset. We specify time-space covariance function conditionally by regionalization and decomposition of the

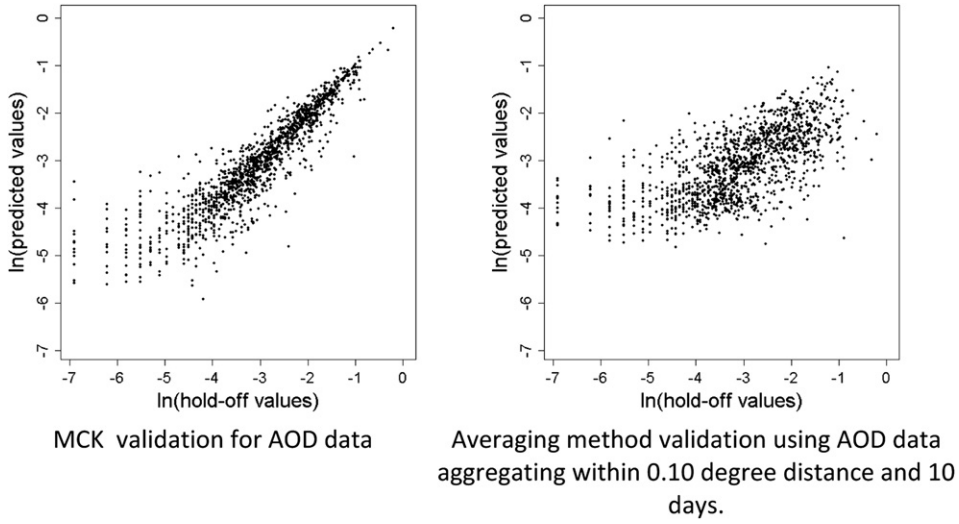

Fig. 4. Validation of MCK and spatiotemporal averaging methods using observed ln(AOD) data between December 2007 and January 2009 for Cleveland MSA. 10% of the data points were held-off and values were predicted at these points using MCK method and averaging method.

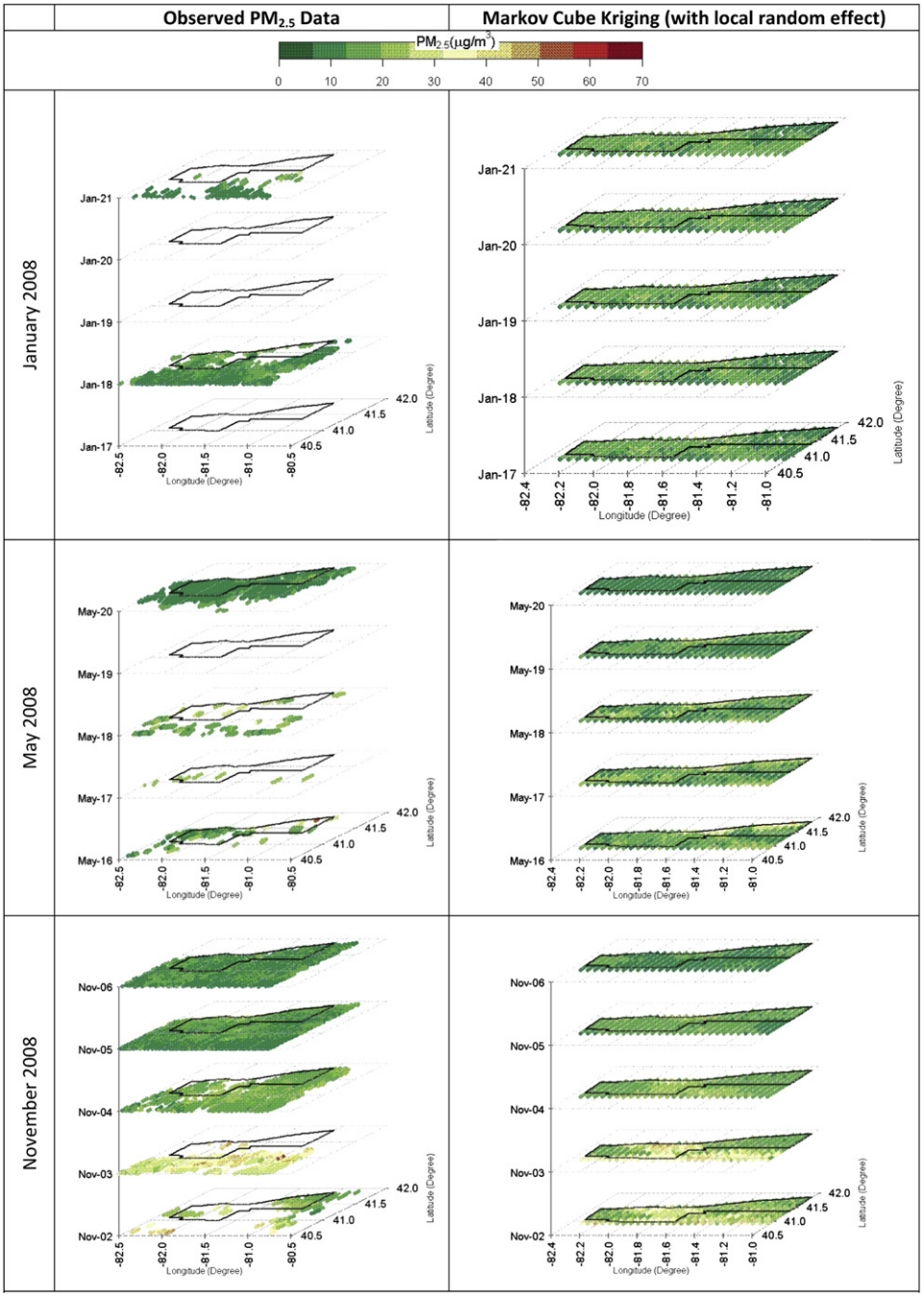

Fig. 5. Observed $PM_{2.5}$ data points and predicted $PM_{2.5}$ surface in Cleveland, MSA 2008.

underlying process into temporal, regional and local effects. In addition, the induced marginal covariance is not limited to the separable or stationary family. Computationally, we utilize initial decompositions of various GMRF structure matrices and advances in sampling from high dimensional GMRFs with linear constraints (Rue, 2001). The computation involves matrices of fixed dimension chosen by the modeler, independent of the number of observations. Therefore, we can analyze large data and offer a worked out example of overcoming the big *n* problem.

Despite the above advantages, a number of limitations remain. First, full Bayesian inference by MCK is still computationally intensive when performing Kriging on large dataset that consists of large geographical area span across greater time interval without losing local spatiotemporal details. This issue can be addressed by approximation methods such as integrated nested Laplace approximation of posterior marginal (Rue et al., 2009). Parallel computing is another solution to address the computation problem. Second, the choice of tessellation over time and space is still a challenging issue. We need to determine the time and spatial locations in the dataset that lead to best predictive performance. Some researchers suggest varying locations over time and space, especially for MCMC sampling (Banerjee et al., 2010). Third, the current implementation does not make full use of the temporal hierarchy in the model. However, such hierarchy will be useful for process(es) that exhibit both seasonal and local temporal trends. In this paper, we demonstrate the application of MCK to predict $PM_{2.5}$

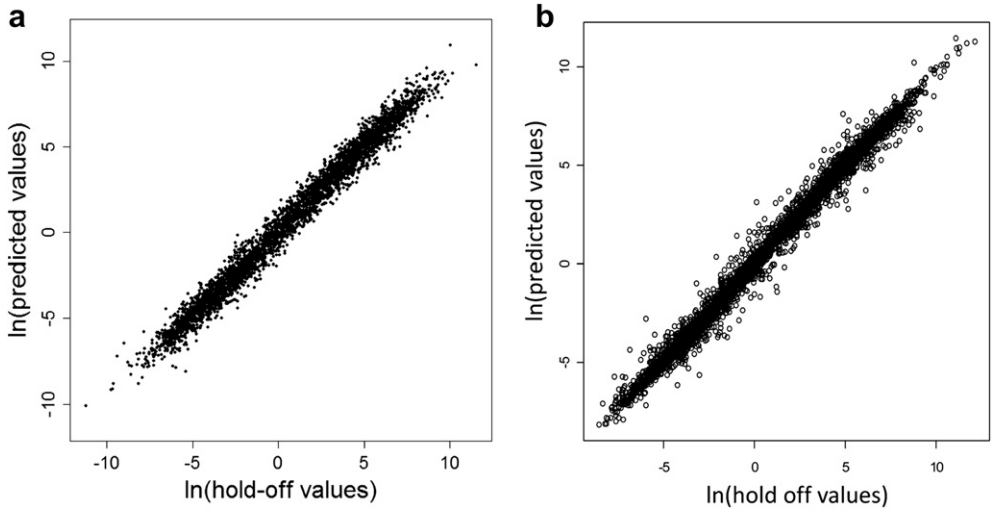

Fig. 6. a: MCK validation using the simulated data at *regularly spaced* data points, simulated from spatiotemporal process with large spatiotemporal trends and nonstationary covariance function. Shown is 10% random subset of the 9 datasets. b: MCK validation using simulated data at *irregularly spaced* data points, simulated from spatiotemporal process with large spatiotemporal trends and nonstationary covariance function. Shown is 10% random subset of the 9 datasets.

and AOD at a given location and time, assuming the health data are at point locations. Nonetheless, MCK application can be extended further to predict estimates for at any geographic units units and temporal intervals. It can be achieved either by computing an estimates at the centroid of the areal unit (as demonstrated for the 2.5 km grid) or by predicting estimates at many points (simulated randomly) within the selected areal unit and time interval and averaging the predictions at the simulated locations.

Our modeling approach has the potential to improve time-space resolved estimates of $PM_{2.5}$, which is vital for air quality monitoring. As typical with other environmental processes, air pollution data come from multiple sources. The EPA monitoring stations collect hourly data for longer periods; but the monitoring stations are sparsely distributed. Therefore, these data from EPA alone are not sufficient to quantify population exposure to air pollution. The satellite based AOD data (corrected for meteorological conditions and spatiotemporal structure) is increasingly being recognized as a surrogate measure of air quality (Kumar et al., 2007; Kumar et al., 2011a). Although AOD data have comprehensive spatial coverage, these data suffer from cloud contamination and include systematic temporal gaps. The hierarchical framework of MCK offers a consistent approach to combine data from these two sources (EPA and NASA) to develop time-space resolved estimates of air pollution exposure. Not only will these data be critically important to develop population exposure at unprecedented spatiotemporal scales (of health dataset), but also these data can be used for air quality surveillance and management.

With the increasing availability of multi-resolution spatiotemporal data from multiple sources, the use of time-space Kriging is likely to increase in the near future. We live in the era of geospatial information revolution in which convergence of spatiotemporal data and information from multiple sources and multiple sensors (with varying spatiotemporal scales) is inevitable. While this convergence is important to develop an understanding of the complex socio-physical patterns and process that shape them, increasing data dimension and data size make it difficult to address the problem of spatiotemporal misalignment, mismatch in the spatiotemporal resolution/scales and gaps (or missing values) across space and time. As demonstrated in this paper, the computationally efficient MCK method holds the key to address these problems. This paper also calls for further research on computationally efficient time-space Kriging models and algorithms that minimize computation time and capture hierarchical (local, regional and global) nonstationary spatiotemporal structures in the data.

Acknowledgement

This research was supported by NIH (R21 ES014004-01A2) and EPA (R833865). We would like to thank two anonymous referees for providing us with constructive comments and suggestions that allowed us to improve the quality of the initial submission.

Appendices A and B. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.atmosenv.2013.02.034.

R Library: MCK library is at http://www.ccs.miami.edu/~nkumar/MCK_0.0-1.zip.

References

Banerjee, S., Carlin, B.P., Gelfand, A.E., 2004. Hierarchical Modeling and Analysis for Spatial Data. Chapman & Hall/CRC, Boca Raton, Florida.

Banerjee, S., Finley, A.O., Waldmann, P., Ericsson, T., 2010. Hierarchical spatial process models for multiple traits in large genetic trials. Journal of the American Statistical Association 105, 506–521.

Banerjee, S., Gelfand, A.E., Finley, A.O., Sang, H., 2008. Stationary process approximation for the analysis of large spatial datasets. Journal of the Royal Statistical Society Series B-Statistical Methodology 70, 825–848.

Besag, J., York, J., Mollie, A., 1991. Bayesian image-restoration, with 2 applications in spatial statistics. Annals of the Institute of Statistical Mathematics 43, 1–20.

Clayton, D.G., 1996. Generalized linear mixed models. In: Gilks, W.R., Richardson, S., Spiegelhalter, D.J. (Eds.), Markov Chain Monte Carlo in Practice. Chapman & Hall, London, pp. 275–301.

Cressie, N., 1993. Statistics for Spatial Data. John Wiley & Sons, INC, New York.

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

Cressie, N., Johannesson, G., 2008. Fixed rank Kriging for very large spatial data sets. Journal of the Royal Statistical Society Series B-Statistical Methodology 70, 209–226.

Cressie, N., Shi, T., Kang, E.L., 2010. Fixed rank filtering for spatio-temporal data. Journal of Computational and Graphical Statistics 19, 724–745.

De Cesare, L., Myers, D.E., Posa, D., 2001. Product-sum covariance for space-time modeling: an environmental application. Environmetrics 12, 11–23.

De Iaco, S., Myers, D.E., Posa, D., 2001. Space-time analysis using a general product-sum model. Statistics & Probability Letters 52, 21–28.

De Iaco, S., Myers, D.E., Posa, D., 2002. Space-time variograms and a functional form for total air pollution measurements. Computational Statistics & Data Analysis 41, 311–328.

Diggle, P., Lophaven, S., 2006. Bayesian geostatistical design. Scandinavian Journal of Statistics 33, 53–64.
Fuentes, M., Chen, L., Davis, J.M., Lackmann, G.M., 2005. Modeling and predicting complex space-time structures and patterns of coastal wind fields. Environmetrics 16, 449–464.
Haas, T.C., 1995. Local prediction of a spatio-temporal process with an application to wet sulfate deposition. Journal of the American Statistical Association 90, 1189–1199.
Hartman, L., Hossjer, O., 2008. Fast kriging of large data sets with Gaussian Markov random fields. Computational Statistics & Data Analysis 52, 2331–2349.
Kumar, N., 2012. Uncertainty in the relationship between criteria pollutants and low birth weight in Chicago. Atmospheric Environment 49, 171–179.
Kumar, N., Chu, A., Foster, A., 2007. An empirical relationship between $PM_{2.5}$ and aerosol optical depth in Delhi metropolitan. Atmospheric Environment 41, 4492–4503.
Kumar, N., Chu, A.D., Foster, A.D., Peters, T., Willis, R., 2011a. Satellite remote sensing for developing time and space resolved estimates of ambient particulate in Cleveland, OH. Aerosol Science and Technology 45, 1090–1108.
Kumar, N., Liang, D., Linderman, M., Chen, J., 2011b. An Optimal Spatial Sampling for Demographic and Health Surveys. SSRN Working Paper – 1808974.
Levy, R.C., Remer, L.A., Mattoo, S., Vermote, E.F., Kaufman, Y.J., 2007. Second-generation operational algorithm: retrieval of aerosol properties over land from inversion of moderate resolution imaging spectroradiometer spectral reflectance. Journal of Geophysical Research-atmospheres 112, D13211.
Rue, H., 2001. Fast sampling of Gaussian Markov random fields. Journal of the Royal Statistical Society Series B-Statistical Methodology 63, 325–338.
Rue, H., Held, L., 2006. Gaussian Markov Random Fields Theory and Applications. Chapman & Hall/CRC, Boca Raton.
Rue, H., Martino, S., Chopin, N., 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. Journal of the Royal Statistical Society Series B-Statistical Methodology 71, 319–392.
Rue, H., Tjelmeland, H., 2002. Fitting Gaussian Markov random fields to Gaussian fields. Scandinavian Journal of Statistics 29, 31–49.
Sahu, S.K., Gelfand, A.E., Holland, D.M., 2006. Spatio-temporal modeling of fine particulate matter. Journal of Agricultural Biological and Environmental Statistics 11, 61–86.
Schmid, V., Held, L., 2004. Bayesian extrapolation of space-time trends in cancer registry data. Biometrics 60, 1034–1042.
Stevens, D.L., Olsen, A.R., 2004. Spatially balanced sampling of natural resources. Journal of the American Statistical Association 99, 262–278.
Zhu, Z.Y., Stein, M.L., 2006. Spatial sampling design for prediction with estimated parameters. Journal of Agricultural Biological and Environmental Statistics 11, 24–44.
Zimmerman, D.L., 2006. Optimal network design for spatial prediction, covariance parameter estimation, and empirical prediction. Environmetrics 17, 635–652.