Statistica Sinica **25** (2015), 369-384
doi:http://dx.doi.org/10.5705/ss.2013.262w

A PRACTICAL APPROACH TO SPATIO-TEMPORAL ANALYSIS

Huijing Jiang, Angela Schörgendorfer, Youngdeok Hwang and Yasuo Amemiya

IBM Thomas J. Watson Research Center

Abstract: This paper introduces a spatio-temporal statistical analysis approach appropriate for monitoring or managing a physical system in which measurements are taken over dense time resolution but at sparse locations. The proposed approach is designed for implementation in an automated and efficient operation with manual intervention required only for scenario analysis. The method is based on a modeling framework for complex predictor-response and spatio-temporal relationships, and issues model-based prediction intervals. To accommodate varying practical situations, the method also includes an automated decision criterion for choosing between parametric and nonparametric spatial covariance models. The approach is illustrated using a data center thermal management problem.

Key words and phrases: Empirical orthogonal function based prediction, goodness-of-fit test, monitoring network, nonparametric covariance matrix, spatio-temporal modeling.

1. Introduction

With recent advances in computation and data storage technology, data are often collected over automated monitoring networks. Service industries have been widely involved in such applications, including building energy management, performance analysis and forecasting for service branches, and public transportation planning. The interest in such applications lies in monitoring the operations, forecasting future behavior, issuing prediction at new locations, and providing decision support for remedial or proactive interventions.

In this paper, we consider scenarios in which data are collected from a network of monitoring sites with a fixed number of spatial locations. Deploying new monitoring stations or sensors often results in considerable additional cost, while the maintenance costs for existing sites are marginal. As a result, measurements are often taken over time with dense temporal resolution at sparse spatial locations.

In analyzing data for such applications, several challenges arise. First, computational methods need to be expeditious for repeated model fitting and forecasting the future values as new data arrives continuously, while accommodating complex spatial relationships with minimal human intervention for operation. Second, the model should be able to identify the factors affecting observations and hence enable prediction for hypothetical scenarios, so that prescribing actions may be taken to achieve a desired future change to better manage the system. Lastly, a flexible method is needed to select an appropriate spatial correlation structure. Although assuming spatial correlation of a certain parametric functional form facilitates computation, strong deviations from the assumed functional pattern may lead to inaccurate spatial prediction, invalid inferences and computational problems. The goal of this paper is to propose a spatio-temporal prediction model to address these challenges.

In the spatio-temporal statistics literature, a process $Y(\boldsymbol{s},t)$ observed over space and time is often modeled through

$$Y(\boldsymbol{s},t) = \mu(\boldsymbol{s},t) + Z(\boldsymbol{s},t), \tag{1.1}$$

where $\mu(\boldsymbol{s},t)$ captures the mean trend, $Z(\boldsymbol{s},t)$ is a mean zero Gaussian process with covariance function $C(\boldsymbol{s},\boldsymbol{s}';t,t')$, and $\boldsymbol{s}$ and t denote the space and time, respectively. Under this framework, we present a general modeling approach integrating a goodness-of-fit (GOF) test-based switching criterion which can automatically choose between parametric and nonparametric spatial models. In addition to the computation benefits, the separability assumption imposed in the spatio-temporal covariance model provides a flexibility to allow for any general form of spatial covariance incorporated in the model.

The existing work in this line can be generally grouped into two directions. The first focuses on developing valid spatio-temporal covariance functions for the error process, especially on valid non-separable spatio-temporal covariance functions (e.g., see Gneiting (2002); Stein (2005); Rodrigues and Diggle (2010); Fonseca and Steel (2011)). Excellent reviews of such modeling approaches can be found in Gneiting and Schlather (2002) and Gneiting, Genton, and Guttorp (2007). The spatio-temporal models developed in this direction view time as continuous rather than discrete and more emphasis is put on spatial prediction but less on forecasting future values.

The other research direction takes a dynamic modeling approach and explicitly considers a discrete time domain. It extends multivariate time-series models to spatio-temporal problems. Mardia et al. (1998) proposed a kriged Kalman filter approach which combines kriging and dynamic linear model for spatial interpolation and temporal forecasting, respectively. Cressie and Wikle (2011) advocated a dynamic spatio-temporal model (DSTM) which models spatial dependence via a set of spatial basis functions and the temporal autocorrelation through the evolution of state vectors. Nobrea, Sansob, and Schmidtc (2011) proposed a spatially varying autoregressive (AR) processes to allow AR coefficients to vary over space. Existing methods in this direction utilize Markov chain Monte Carlo as computational tools, which are not computationally affordable for our applications.

The proposed modeling approach is different from both existing research directions in the sense that it aims at issuing temporal forecasting and spatial prediction simultaneously with a fast and stable computation algorithm. The modeling and its computational algorithm are designed for automated and efficient operation with minimal manual intervention. External factors are incorporated as covariates in the model for system diagnosis and future scenario analysis. Our spatio-temporal model can also accommodate any spatial covariance structure flexibly, and our GOF test is applicable for any proposed structure including non-stationary cases (e.g., Sampson and Guttorp (1992), Higdon, Swall, and Kern (1999), Nychka, Wikle, and Royle (2002), Jun and Stein (2008)).

The remainder of the article is organized as follows. Section 2 introduces our model and describes the model fitting procedure. Section 3 derives the proposed decision criterion to choose between the modeling alternatives. Section 4 gives prediction method within the proposed framework. Section 5 illustrates the proposed method with a simulation study and an application from the information technology industry. We conclude with a short summary and discussion in Section 6.

2. Model

In our model, we consider a spatio-temporal process over discrete time and continuous space domain and hence notate $Y(\boldsymbol{s},t)$ in (1.1) as $Y_t(\boldsymbol{s})$. As in (1.1), the process is decomposed into mean trend and error process, where the mean trend is modeled with a set of covariates, $\mu(\boldsymbol{s},t) = \mu(\mathbf{x}_t(\boldsymbol{s}))$. The following spatio-temporal model is then

$$Y_t(\boldsymbol{s}) = \mu(\mathbf{x}_t(\boldsymbol{s})) + Z_t(\boldsymbol{s}), \tag{2.1}$$

where $Y_t(\boldsymbol{s})$ is the observed measurement at location $\boldsymbol{s} \in \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n\}$ and time $t = 1, \ldots, m$, $\mu(\mathbf{x}_t(\boldsymbol{s}))$ is a deterministic mean trend of q known factors $\mathbf{x}_t(\boldsymbol{s}) = (x_{1,t}(\boldsymbol{s}), \ldots, x_{q,t}(\boldsymbol{s}))'$ at location $\boldsymbol{s}$ and time t, and $Z_t(\boldsymbol{s})$ is a mean-zero space-time correlated random process.

The role of $\mu(\mathbf{x}_t(\boldsymbol{s}))$ is crucial in forecasting to incorporate impact of external factors, system settings, or seasonal trends that may happen in the future. The mean trend can be modeled flexibly, but a common model is a linear model

$$\mu(\mathbf{x}_t(\boldsymbol{s})) = \sum_{d=1}^{D} \beta_d \tilde{x}_{d,t}(\boldsymbol{s}), \tag{2.2}$$

where $\tilde{x}_{d,t}(\boldsymbol{s})$ is the dth regressor of $\mathbf{x}$ after an appropriate transformation at location $\boldsymbol{s}$ and time t, and β_d is the corresponding regression coefficient.

Despite its intuitive and simple form, it is not always straightforward to use (2.1) for forecasting purposes, because future $\mathbf{x}_t(\boldsymbol{s})$ may not be available. But, thanks to recent advances in computer modeling in engineering and science, these predictors can be obtained based on computational models. Moreover, since computer model outputs can be generated on an arbitrarily fine grid in space, we can assume that $\mathbf{x}_t(\boldsymbol{s})$ are available not only at limited locations but also at every desired location.

In our model, we assume that current value of the spatio-temporal process $Z_t(\boldsymbol{s})$ is a function of past values,

$$Z_t(\boldsymbol{s}) = \mathcal{M}(\{\epsilon_u(\boldsymbol{s})\}_{u<t}), \tag{2.3}$$

where $\mathcal{M}$ is a general class of models and $\{\epsilon_u(\boldsymbol{s})\}_{u<t}$ are spatially correlated errors at time points prior to t. A model for (2.3) allowing computational efficiency is an AR model of order L, given by

$$Z_t(\boldsymbol{s}) = \sum_{l=1}^{L} \rho_l Z_{t-l}(\boldsymbol{s}) + \epsilon_t(\boldsymbol{s}), \tag{2.4}$$

where ρ_l is the l-th autoregressive coefficient reflecting the correlation between Z_t and Z_{t-l}. The AR residual vectors $\boldsymbol{\epsilon}_t = (\epsilon_t(\boldsymbol{s}_1), \ldots, \epsilon_t(\boldsymbol{s}_n))' = (\epsilon_{t1}, \ldots, \epsilon_{tn})'$, with $E(\boldsymbol{\epsilon}_t) = \mathbf{0}$ and $\mathrm{Var}\,(\boldsymbol{\epsilon}_t) = \boldsymbol{\Sigma}$, are assumed to be spatially correlated but independent of each other. For simplicity, we illustrate our approach using (2.2) and (2.4), although it can be more generally applied under (2.1) and (2.3).

Under (2.4), the spatio-temporal process $\boldsymbol{Y} = (\boldsymbol{Y}_1', \ldots, \boldsymbol{Y}_m')'$ is of separable form $\mathrm{Var}\,(\boldsymbol{Y}) = \boldsymbol{\Gamma} \otimes \boldsymbol{\Sigma}$, where $\boldsymbol{\Sigma} = (\Sigma_{ij})$ is an $n \times n$ spatial covariance matrix, $\boldsymbol{\Gamma} = (\Gamma_{ij})$ is the $m \times m$ AR(L) covariance matrix whose elements are the autocovariance generating function $\gamma(t - t')$ for $t, t' = 1, \ldots, m$ (Brockwell and Davis (2002)) and $\otimes$ denotes Kronecker product. Using (2.4) is a practical way to handle $\boldsymbol{\Gamma}$ because it allows a straightforward forecasting without any complicated optimization or iterative computation. The separability assumption here not only facilitates the computation for forecasting, but also allows $\boldsymbol{\Sigma}$ to be any general form. As detailed in the next section, we proposed a goodness-of-fit test to automatically choose the form of $\boldsymbol{\Sigma}$.

Our modeling approach can be extended to other non-separable spatio-temporal model. For example, allowing AR coefficients to vary over space leads to a non-separable spatio-temporal model. Still, the test procedure in Section 3 can be applied to decide between a computationally efficient parametric model, and a flexible but computationally more demanding unstructured spatial model.

Under (2.2) and (2.4), the model can be fitted using a multi-step estimation procedure. Let $\boldsymbol{X} = (\mathbf{x}_1', \ldots, \mathbf{x}_m')'$ be the $mn \times q$ matrix containing the vector of covariates $\mathbf{x}_t' = (x_{1,t}, \ldots, x_{q,t})'$ in (2.1), and $\boldsymbol{Y} = (\boldsymbol{Y}_1', \ldots, \boldsymbol{Y}_m')'$, an mn-vector. First, assuming no spatio-temporal dependence, estimate $\hat{\boldsymbol{\beta}}$ and obtain the residuals $\hat{\boldsymbol{Z}} = \boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}$ using ordinary least squares (OLS). Then, assuming no spatial correlation, take a regression approach on $\hat{\boldsymbol{Z}}_t$ to estimate AR coefficients $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_L)$, to estimate the autocovariance matrix $\hat{\boldsymbol{\Gamma}}$. Once $\hat{\boldsymbol{\Gamma}}$ is available, one can obtain $\hat{\boldsymbol{\epsilon}}_1, \ldots, \hat{\boldsymbol{\epsilon}}_m$ where $\hat{\boldsymbol{\epsilon}}_t = (\hat{\epsilon}_{t1}, \ldots, \hat{\epsilon}_{tn})$ is the AR residual vector after removing temporal dependence. The spatial covariance $\hat{\boldsymbol{\Sigma}}$ then can be estimated based on $\hat{\boldsymbol{\epsilon}}_t$. The next step is to re-estimate $\hat{\boldsymbol{\Gamma}}$, accounting for this spatial dependence using generalized least squares (GLS). Finally, the regression parameters $\hat{\boldsymbol{\beta}}$ can be re-estimated by taking into account the spatio-temporal correlation in the data using GLS.

Although the spatial covariance $\hat{\boldsymbol{\Sigma}}$ is important in fitting this model, we have so far omitted its selection and estimation. A detailed discussion on this aspect follows.

3. Inference on Spatial Covariance

In this section, we describe the inference on spatial covariance, including a GOF test for the parametric models. The proposed approach can be applied generally to any parametric spatial covariance including the Matérn class. Despite its generality, it can be computationally intensive. Thus, we also present a two-step alternative test tailored for a popular class of covariance models.

Spatial covariance modeling is a topic that has been extensively studied; see, for example, Cressie (1993), Stein (1999), or Cressie and Wikle (2011). However, relatively little work has been done on a GOF test on spatial covariance models. We use large sample theory to develop a general test procedure.

Under our model, using the estimated residuals $\hat{\boldsymbol{\epsilon}}_t$ from the AR model over time, we can estimate the spatial covariance matrix $\boldsymbol{\Sigma}$ by a non-parametric method of moment estimator

$$\boldsymbol{S} = (s_{ij}),\ s_{ij} = \frac{1}{m}\sum_{t=1}^{m} \hat{\epsilon}_{ti}\hat{\epsilon}_{tj}, \tag{3.1}$$

for $i, j = 1, \ldots, n$. Let $\boldsymbol{s}_L$ be the vectorized lower triangular of $\boldsymbol{S}$ including the diagonal elements. Note that length of $\boldsymbol{s}_L$ is $N_L = n(n+1)/2$.

To test whether a particular spatial covariance model is adequate, we consider the hypotheses H_0: $\boldsymbol{\Sigma}$ follows a parametric spatial covariance versus H_1: $\boldsymbol{\Sigma}$ does not have the structure in H_0. Let $\hat{\boldsymbol{\Sigma}}_0$ be the estimated covariance matrix under a parametric spatial covariance model assumption, $\hat{\boldsymbol{\Sigma}}_L$ the vectorized lower triangular of $\hat{\boldsymbol{\Sigma}}_0$, and $\boldsymbol{\Lambda} = \mathrm{Var}\,(\hat{\boldsymbol{\Sigma}}_L)$. Then a test statistic for testing the hypotheses is given by

$$m\left(\hat{\boldsymbol{\Sigma}}_L - \boldsymbol{s}_L\right)' \hat{\boldsymbol{\Lambda}}^{-1}\left(\hat{\boldsymbol{\Sigma}}_L - \boldsymbol{s}_L\right). \tag{3.2}$$

Under H_0, for a sufficiently large m, the distribution of this test statistic can be approximated by χ^2 distribution. The alternative model always has N_L parameters while the null model has the number of parameters depending on the assumed model; the degrees of freedom of the χ^2 distribution is the difference. An intuitive estimator for $\boldsymbol{\Lambda}$ is available from its sample second moments. Let $\boldsymbol{\xi}_t = (\xi_{1t}, \ldots, \xi_{N_L t})'$ be the vectorized lower triangular of $\hat{\boldsymbol{\epsilon}}_t \hat{\boldsymbol{\epsilon}}_t'$, which is of length N_L. Then $\hat{\boldsymbol{\Lambda}}$ is available with its i,jth element, $\hat{v}_{ij}$, being

$$\frac{1}{m-1}\sum_{t=1}^{m}(\xi_{it}\xi_{jt} - \bar{\xi_i \xi_j})^2, \tag{3.3}$$

where $\bar{\xi_i \xi_j} = \sum_{t=1}^{m} \xi_{it}\xi_{jt}/m$. Note that the GOF test is only valid when the parameters associated with $\hat{\boldsymbol{\Sigma}}_L$ are estimated by minimizing (3.2). Although general, the computation can become prohibitively intensive because it is costly to minimize (3.2), and it requires $O(n^6)$ operations to compute (3.3). Thus we give an alternative procedure tailored for a powered exponential family covariance function. Here both parameter estimation and GOF test can be accomplished in a very efficient way.

3.1. Test for a powered exponential family

A common approach to modeling $\boldsymbol{\Sigma}$ is to assume $C(\boldsymbol{s}_i, \boldsymbol{s}_j) = \mathrm{Cov}\,(\epsilon_{ti}, \epsilon_{tj})$, where ϵ_{ti} and ϵ_{tj} are in (2.4), is a monotone function of the distance between two points $h_{ij} = \|\boldsymbol{s}_i - \boldsymbol{s}_j\|$. Errors are assumed to be more correlated for two locations that are closer to each other. A popular model in this class is the powered exponential spatial covariance function

$$C(\boldsymbol{s}_i, \boldsymbol{s}_j) = \begin{cases} \sigma^2 \exp(-\theta h_{ij}^p) & \text{if } h_{ij} > 0; \\ \sigma^2 + \tau^2 & \text{if } h_{ij} = 0, \end{cases} \tag{3.4}$$

for a given p (Cressie (1993)). This model has gained popularity due to a simple functional expression that allows an easy interpolation of covariances, and thus predictions, for unobserved locations. In the spatial statistics context, the parameters of this model are often estimated by a binned semi-variogram (Cressie (1993); Cressie and Wikle (2011)).

Despite its advantages, the parametric model in (3.4) often fails to fully describe the actual variability in many observed processes. Common problems include non-monotone behavior of the covariances over distance, non-homogeneity of variances across spatial location, and even negative correlations among locations. In a fixed network setting, the variance-covariance matrix of any spatial structure can be estimated by exploiting the repeated measurements over time. Taking advantage of such circumstances, an unstructured form of spatial covariance is able to handle a wider range of error structures without the potential issues from misspecification of the covariance model.

As such, the use of a restrictive model in (3.4) requires rigorous justification. The parametric model in (3.4) defines two key features on the spatial covariance structure: correlation decreasing with distance and constant variance over space. Accordingly, our proposed test diagnoses and validates the efficacy of the model by examining these characteristics.

3.2. Test for spatially decaying correlation

The first step of our testing procedure considers $C(\boldsymbol{s}_i, \boldsymbol{s}_j)$ for $h_{ij} > 0$. From (3.4), observe that

$$\log C(\boldsymbol{s}_i, \boldsymbol{s}_j) = \log \sigma^2 - \theta h_{ij}^p. \tag{3.5}$$

The parameters are collectively denoted by $\boldsymbol{\eta} = (\log \sigma^2, -\theta)$. When the covariance structure follows (3.4), θ is positive. We thus consider the hypotheses $H_0 : \theta \leq 0$ versus $H_1 : \theta > 0$. If we do not reject H_0, there is not enough evidence that (3.4) is suitable for the data and hence we decide to use an unstructured spatial covariance matrix. If the first test rejects H_0, we continue with the second test.

Let $\boldsymbol{s} = (s_i)$ be the vectorized lower triangular of $\boldsymbol{S}$ from (3.1) with the diagonal elements excluded and $\boldsymbol{h} = (h_i)$ be the corresponding vector of pairwise distances between the n locations. Note that $\boldsymbol{s}$ and $\boldsymbol{h}$ are vectors of length $N = n(n-1)/2$. Define an $N \times 2$ matrix $\boldsymbol{A} = [\mathbf{1}, \boldsymbol{h}^p]$, where $\mathbf{1}$ is the vector of ones. By definition, $s_1, \ldots, s_N$ are dependent of each other. Hence, we can estimate $\boldsymbol{\eta}$ using GLS by

$$\hat{\boldsymbol{\eta}} = (\widehat{\log \sigma^2}, -\hat{\theta}) = (\mathbf{A}'\boldsymbol{V}^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{V}^{-1}(\log \boldsymbol{s}), \tag{3.6}$$

where $\boldsymbol{V} = \mathrm{Var}\,(\log \boldsymbol{s})$. By the property of GLS estimators,

$$\hat{\boldsymbol{\eta}} \sim \mathrm{N}(\boldsymbol{\eta}, (\boldsymbol{A}'\boldsymbol{V}^{-1}\boldsymbol{A})^{-1}), \tag{3.7}$$

and hence the standard error for $\hat{\boldsymbol{\eta}}$ is $se(\hat{\boldsymbol{\eta}}) = \sqrt{\mathrm{diag}[(\boldsymbol{A}'\boldsymbol{V}^{-1}\boldsymbol{A})^{-1}]}$. Based on (3.7), one can calculate the test statistic $z_1 = \hat{\theta}/se(\hat{\theta})$ and conduct a test; if $z_1 < z_{1-\alpha_1}$ with given level of α_1, all following model fittings are performed using $\hat{\boldsymbol{\Sigma}} = \boldsymbol{S}$.

Although the sample second moment estimator in (3.3) can be used to estimate $\boldsymbol{V}$, the computational issue still exists. The computation is greatly facilitated by assuming normality of $\boldsymbol{\epsilon}_t$. Let $\boldsymbol{R} = (r_{ij})$ denote empirical correlations matrix, where $r_{ij} = s_{ij}/\sqrt{s_{ii}s_{jj}}$, for $i, j = 1, \ldots, n$, and $\boldsymbol{r} = (r_i)$ the vectorized lower triangular elements of $\boldsymbol{R}$. Under H_0 and normality, $\hat{\boldsymbol{V}} = 2(\boldsymbol{B} + \mathbf{1}_N\mathbf{1}_N')/m$, where $\boldsymbol{B}$ is an $N \times N$ diagonal matrix with its ith element $b_i = (r_i^{-2} - 1)/2$ for $i = 1, \ldots, N$. Let $\boldsymbol{b}^{-1} = \boldsymbol{B}^{-1}\mathbf{1}$ with ith element $b_i^{-1} = 2r_i^2/(1 - r_i^2)$ for $i = 1, \ldots, N$. It is straightforward to see

$$\hat{\boldsymbol{V}}^{-1} = \frac{m}{2}(\boldsymbol{B}^{-1} - c\boldsymbol{b}^{-1}\boldsymbol{b}^{-1'}), \tag{3.8}$$

with $c = 1/(1 + \mathbf{1}'\boldsymbol{B}^{-1}\mathbf{1})$. Calculation of (3.8) is straightforward as $\boldsymbol{B}$ is now a diagonal matrix. Once either (3.3) or (3.8) is available, the substitution principle allows computation of (3.6) with $\hat{\boldsymbol{V}}$. When normality is hard to justify, one can use robust covariance matrix estimators to simplify the computation.

3.3. Test for equal variances

The second test is related to the elements of the covariance matrix associated with $h_{ij} = 0$. Under (3.4), $\text{Var}(\epsilon_{it})$ is identical for all $\boldsymbol{s}_i$. Thus we consider H_0 : $\text{Var}(\epsilon_{1t}) = \ldots = \text{Var}(\epsilon_{nt})$ versus H_1: $\text{Var}(\epsilon_{it}) \neq \text{Var}(\epsilon_{jt})$ for some i, j. To test the equality, let $\boldsymbol{v} = (s_{11}, s_{22}, \ldots, s_{nn})'$ be the diagonal elements of $\boldsymbol{S}$ in (3.1), the vector of location-specific variances. Let $\boldsymbol{\Omega} = (\Omega_{ij})$ denote $\text{Var}(\boldsymbol{v})$, an $n \times n$ matrix with $\Omega_{ij} = \text{Cov}(s_{ii}, s_{jj})$. Note that $\hat{\boldsymbol{\Omega}}$ can be calculated similar to (3.3), and its computational burden is much less. When computational effort needs to be further reduced, one may again employ the normality assumption which leads to $\Omega_{ij} = 2\Sigma_{ij}^2/m$ for $i, j = 1, \ldots, n$. Then one can use $\hat{\boldsymbol{\Omega}} = (\hat{\Omega}_{ij})$, where $\hat{\Omega}_{ij} = 2s_{ij}^2/m$ and s_{ij} is defined in (3.1).

Under H_0, the distribution of the test statistic

$$z_2 = (\boldsymbol{v} - \mathbf{1}_n(\mathbf{1}'\hat{\boldsymbol{\Omega}}^{-1}\mathbf{1})^{-1}\mathbf{1}'\hat{\boldsymbol{\Omega}}^{-1}\boldsymbol{v})'\hat{\boldsymbol{\Omega}}^{-1}(\boldsymbol{v} - \mathbf{1}_n(\mathbf{1}'\hat{\boldsymbol{\Omega}}^{-1}\mathbf{1})^{-1}\mathbf{1}'\hat{\boldsymbol{\Omega}}^{-1}\boldsymbol{v})$$

can be approximated by χ^2_{n-1}. One can conduct a test based on z_2; if $z_2 > \chi^2_{n-1,1-\alpha_2}$ with given level of α_2, all subsequent calculations are performed using the empirical spatial covariance matrix. Otherwise, (3.4) are used.

3.4. Remarks on two tests

We choose the critical region of the first test such that all the associated parameters can be used directly for the computation of $\boldsymbol{\Sigma}$; when our procedure chooses the model in (3.4), it always yields a positive $\hat{\theta}$. An estimator of τ^2 is readily available as $\hat{\tau}^2 = \max\{0, n^{-1}\sum_{i=1,\ldots,n} s_{ii} - \exp(\log\hat{\sigma}^2)\}$. The unstructured covariance function does not restrict the parameters.

Although all correlation coefficients are to be positive under (3.4), in practice some correlations may be very close to zero or even negative, which in turn causes a problem in the estimation of (3.6). To handle such cases, we enforce a minimum correlation of δ by setting $s_{ij} = \delta\sqrt{s_{ii}s_{jj}}$ when $r_{ij} < \delta$. A prespecified small value, e.g., .01, can be used for δ, or more careful treatment based on m can be applied. Near zero or negative correlation coefficients already imply that the model specification in (3.4) is inadequate.

Lastly, we set the order of the two tests to be in the current sequence as we believe that the first test is of greater importance. Predictions at unobserved locations rely heavily on the decaying nature of the covariance with respect to the distance, and hence it is the more representative and fundamental characteristic of the model (3.4).

4. Prediction

In our study, we are interested in issuing forecasts not only at known monitoring sites but also at any locations in future time. At a new location to issue new predictions, denoted by $\boldsymbol{s}^*$, let $\boldsymbol{c}(\boldsymbol{s}^*)$ be the vector of the spatial covariance between $\boldsymbol{s}^*$ and the sample locations $(\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n)$. Similarly, let $\boldsymbol{\gamma}_{m+h} = (\gamma(m+h-1), \ldots, \gamma(h))'$ be the temporal covariance vector of length m. Let $\boldsymbol{c}_h(\boldsymbol{s}^*) = \boldsymbol{\gamma}_{m+h} \otimes \boldsymbol{c}(\boldsymbol{s}^*)$. The best linear unbiased predictor at $\boldsymbol{s}^*$ for future time $t = m + h$ is $\hat{Y}_{m+h}(\boldsymbol{s}^*) = \boldsymbol{\beta}'\mathbf{x}_{m+h}(\boldsymbol{s}^*) + \hat{Z}_{m+h}(\boldsymbol{s}^*)$, where $\hat{Z}_{m+h}(\boldsymbol{s}^*) = \boldsymbol{c}_h(\boldsymbol{s}^*)(\hat{\boldsymbol{\Gamma}} \otimes \hat{\boldsymbol{\Sigma}})^{-1}\hat{\boldsymbol{Z}}$ and $\hat{\boldsymbol{Z}}$ is the residual vector obtained after the model fitting in Section 2.

When the model in (3.4) is chosen from the tests in Section 3, $\boldsymbol{c}(\boldsymbol{s}^*)$ is directly available with the parameters estimated from (3.6). Otherwise, we use a nonparametric approach based on an empirical orthogonal function (EOF) method (Obled and Creutin (1986)). Specifically, we first perform an eigen-decomposition on the empirical covariance matrix of $\boldsymbol{S}$ in (3.1): $\boldsymbol{S} = \boldsymbol{\Phi\Lambda\Phi}'$ where $\boldsymbol{\Phi} = (\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_n)$ with $\boldsymbol{\phi}_k = (\phi_k(\boldsymbol{s}_i), \ldots, \phi_k(\boldsymbol{s}_i))'$ is the $n \times n$ matrix of eigenvectors, and $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ is the $n \times n$ eigenvalue matrix. Then we interpolate the eigenvectors at the prediction locations to obtain

$$\phi_k(\boldsymbol{s}^*) = \sum_{i=1}^{n} \frac{w_i(\boldsymbol{s}^*)\phi_k(\boldsymbol{s}_i)}{\sum_{j=1}^{n} w_j(\boldsymbol{s}^*)}, \qquad k = 1, \ldots, n, \tag{4.1}$$

where $w_i(\boldsymbol{s}^*)/\sum_{j=1}^{n} w_j(\boldsymbol{s}^*)$ is the weight for $\boldsymbol{s}_i$ (Munoz, Lesser, and Ramsey, 2008). We employ the inverse distance weighting function with $p \leq d$ for d-dimensional space (Shepard (1968)) to obtain $w_i(\boldsymbol{s}^*) = 1/\mathrm{dist}(\boldsymbol{s}^*, \boldsymbol{s}_i)^p$. The resulting spatial covariance vector is

$$\boldsymbol{c}(\boldsymbol{s}^*) = \Big(\sum_{k=1}^{n} \lambda_k \phi_k(\boldsymbol{s}^*)\phi_k(\boldsymbol{s}_1), \ldots, \sum_{k=1}^{n} \lambda_k \phi_k(\boldsymbol{s}^*)\phi_k(\boldsymbol{s}_n)\Big)'. \tag{4.2}$$

For more discussion on the use of EOF method, see Munoz, Lesser, and Ramsey (2008).

Under the normality assumption with known variance components, the variance of the prediction error is given by

$$\begin{aligned}&\mathrm{Var}\,[\hat{Y}_{m+h}(\boldsymbol{s}^*) - Y_{m+h}(\boldsymbol{s}^*)]\\ &= \left[\boldsymbol{X}'_{m+h}(\boldsymbol{s}^*) - \boldsymbol{c}'_h(\boldsymbol{s}^*)(\boldsymbol{\Gamma}\otimes\boldsymbol{\Sigma})^{-1}\boldsymbol{X}\right]\mathrm{Var}\,(\hat{\boldsymbol{\beta}})\left[\boldsymbol{X}'_{m+h}(\boldsymbol{s}^*) - \boldsymbol{c}'_h(\boldsymbol{s}^*)(\boldsymbol{\Gamma}\otimes\boldsymbol{\Sigma})^{-1}\boldsymbol{X}\right]'\\ &\quad -\boldsymbol{c}'_h(\boldsymbol{s}^*)(\boldsymbol{\Gamma}\otimes\boldsymbol{\Sigma})^{-1}\boldsymbol{c}_h(\boldsymbol{s}^*) + \mathrm{Var}\,(Y_{m+h}(\boldsymbol{s}^*)),\end{aligned}$$

where $\mathrm{Var}\,(Y_{m+h}(\boldsymbol{s}^*)) = \gamma(0)(\sigma^2+\tau^2)$ when the spatial parametric model (3.4) is applied, and $\gamma(0)\sum_{i=1}^{n}\lambda_k\phi_k^2(\boldsymbol{s}^*)$ for EOF-based spatial covariance function. Then, a symmetric $100(1-\alpha)\%$ prediction interval is given by

$$\hat{Y}_{m+h}(\boldsymbol{s}^*) \pm z_{1-\alpha/2}\sqrt{\mathrm{Var}\,[\hat{Y}_{m+h}(\boldsymbol{s}^*) - Y_{m+h}(\boldsymbol{s}^*)]},$$

where $z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of the standard normal distribution.

5. Numerical Study

In this section, the effectiveness of the proposed method is corroborated by a simulation study and then illustrated with a case study of thermal management in a data center.

5.1. Simulation

We conducted a simulation study to validate the efficacy of the two-step test in Section 3. To simulate the data, we considered the model in (2.2) having three predictor variables x_1, x_2, x_3, each being generated independently from $U(0,1)$, with $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)' = (2,2,1,1)'$; fifteen locations, $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_{15}$, were chosen from uniform $(0,10)^3$ and fixed during the simulations. The spatio-temporal process $Z_t(\boldsymbol{s})$ in (2.4) was assumed to be AR order $L = 3$, with coefficients $\alpha_1 = 0.5, \alpha_2 = 0.2, \alpha_3 = 0.1$. We generated the spatio-temporal error $Z_t(\boldsymbol{s}_j)$, for $t = 1, \ldots, m$, $j = 1, \ldots, 15$, with a spatial covariance $\boldsymbol{\Sigma}$ and normality assumption on $\boldsymbol{\epsilon}_t$. Data associated with 10 sites were used for model fitting while those from the remaining 5 sites were spared for performance evaluation.

We considered two scenarios: (1) $\boldsymbol{\Sigma}_1$ in the form of (3.4) with $m = 300$, and (2) a nonstationary spatial covariance matrix for $\boldsymbol{\Sigma}_2$ with $m =$1,000. The stationary $\boldsymbol{\Sigma}_1$ had equal variances across locations and spatial correlations that decayed in distance between locations. These assumptions were relaxed in scenario 2. We allowed the variances to vary over locations, and the monotone relationship between correlation and distance was disturbed by applying a local structure.

Table 1. The summarized results of simulation, where the numbers in parenthesis represent the standard deviation.

		RMSE				RMSPE
		β_0	β_1	β_2	β_3	
Scenario 1	Method I	0.168	0.093	0.079	0.095	1.507 (0.088)
	Method II	0.202	0.101	0.105	0.113	1.527 (0.104)
	Method III	0.185	0.102	0.093	0.103	1.516 (0.096)
Scenario 2	Method I	0.630	0.072	0.082	0.072	3.339 (1.294)
	Method II	0.222	0.052	0.060	0.057	2.142 (0.484)
	Method III	0.222	0.052	0.060	0.057	2.142 (0.484)

Specifically, for $\mathbf{\Sigma}_1$, we chose the parameters $\sigma = 1$, $\tau = 0.1$, $\theta = 1/4$ and $p = 2$. For $\mathbf{\Sigma}_2$, we first divided 15 locations into 6 and 9, and made a 15×15 covariance matrix $\boldsymbol{A}$ by letting the elements be 2.5 in a 6×6 block associated with the covariance within the first 6, with the remaining elements 0.5. Then a positive definite $\mathbf{\Sigma}_2$ was obtained by $\boldsymbol{A} + \mathbf{\Sigma}_1$. In this way, we could simulate departure of $\mathbf{\Sigma}_2$ from $\mathbf{\Sigma}_1$ in two key aspects that the tests in Section 3 examine.

We considered different ways to select the covariance model for the simulated data: (I) the method that assumes the stationary and parametric covariance model in (3.4); (II) the method that assumes the non-parametric model without the two-step procedure in Section 3; (III) our method that incorporates the flexibility with covariance switching from the two-step procedure. We replicated the simulations 100 times for each scenario and method, and compared the root mean square prediction error (RMSPE) and root mean square error (RMSE) of $\hat{\boldsymbol{\beta}}$ from the 100 replicates as a measure of overall stability and prediction accuracy.

The results are summarized in Table 1. Under scenario 1, method III does not lose much estimation and prediction accuracy compared to method I that uses the knowledge of the underlying spatial covariance structure. Method III also performs better, in both estimation and prediction than method II without the GOF test-based switching; our method often chooses the correct model and uses the same inference as method I, and the estimated covariance from (3.1) is reasonable even when it chooses a wrong model. Under scenario 2, method I performs considerably worse than method II and III; the model fitted with the incorrectly specified covariance model by using method I can lead to a large error as seen in standard deviation value of RMSPE, and method I produces less efficient coefficient parameter estimates.

These results indicate that our method has flexibility and can make appropriate adjustments in practical situations.

5.2. Industry application

In this section, we present a case study motivated by an industrial project aiming at better managing a data center thermal system and reducing the energy cost (Hamann et al. (2009)). Temperature is a key performance indicator for operating data center equipment in a reliable manner while avoiding excessive use of energy. In order to manage temperature in data centers, relevant environmental information is monitored via a sensor network. We build a predictive model to forecast future temperature distributions across the entire data center, based on hypothetical future settings of the cooling system. The goal is to utilize the predictive model to reduce energy consumption by the cooling system while ensuring safe operating temperatures throughout the center.

The layout of the data center in this case study is depicted in Figure 2. Servers and other equipments are mounted on racks above a raised floor, depicted as grey rectangles in the figure. The data center has alternating "cold aisles" and "hot aisles". The inlet side of a server faces a cold aisle, while the exhaust side faces a hot aisle. Four air conditioning units (ACUs) with large scale fans expel cool air into the plenum of the data center, thereby pressurizing it. Through perforated tiles located in the cold aisles, the cooled air is provided to the inlets of the servers. The heated exhaust air from the servers is returned to the ACUs via the ACU intake openings located in the hot aisles.

We considered three factors that affect the temperature at a given location in the data center: (i) temperature of the air supplied to the data center from each ACU outlet; (ii) airflow through the perforated tiles at the floor level, which determines how cooling air is distributed across the horizontal dimensions; (iii) the height of the location. In total, there are 105 thermal sensors distributed throughout the data center, marked as dots in Figures 2−4. The temperature data is collected in ten-minute intervals. We used data from 1,000 time points as training data to issue forecasts of the temperature distribution map for the entire data center, using the model in (2.2). All factors in the models were included as linear predictors without any transformation or higher order terms of factors. Computation using (3.3) was infeasible due to the sample size of $n = 105$, and therefore all tests were performed using the normality approximation in (3.8). Figure 1 shows a scatter plot of empirical covariances s_{ij} of (3.1) versus distance h_{ij}. Clearly, the observed relationship between covariances and distance does not fit (3.4) well, although there is an overall tendency that covariances decrease in absolute magnitude with respect to the distances. Negative covariances are also visible. These observations suggest that (3.4) may not be suitable for the data. With $\delta = 0.01$, our GOF test statistic in the first step was greater than $Z_{0.999}$ and we went on to the second test. The second test statistic was greater than $\chi^2_{104,0.999}$, suggesting that the assumption of variance homogeneity is invalid.

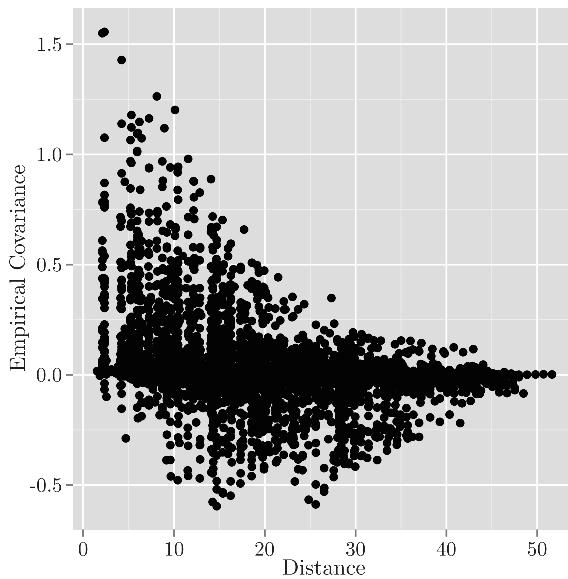

Figure 1. Plot of empirical spatial covariance and distances, where spatial covariances are calculated by (3.1).

Given that the ratio of the smallest and largest diagonal components of (3.1) is 137, the result is not surprising.

Figures 2−4 display the predicted temperature distribution map at time $t = 1,001$ based on the nonparametric spatial model, lower and upper prediction bounds at the 95% confidence level, respectively. The default system values were chosen for temperature and airflow for forecasting. As expected, we observed lower temperatures along the inlets where cold air is expelled through the perforated tiles. The temperature rose as height increased.

For comparison, we also applied the parametric model in (3.4) to predict the temperature distribution in the data center. For $t = 901, \ldots,$ 1,000, we randomly deleted one sensor's data, built the model by using the data obtained from the 1 to $t-1$ period, and obtained one-step ahead forecasting at the deleted location. Blindly applying parametric model gave an RMSPE of 30.775, while our method gave 5.534. The physical structures of the layout environment leads to a complicated underlying spatial process that cannot be fully modeled by parametric spatial models. Our approach detected the poor fit of the parametric model and switched to an alternative unstructured spatial model that successfully picked up the complicated spatial dependence pattern.

In an on-line monitoring framework, such temperature prediction maps are updated on an ongoing basis as new measurements arrive continuously. The predictive model can be used to explore the effect of changing the settings of

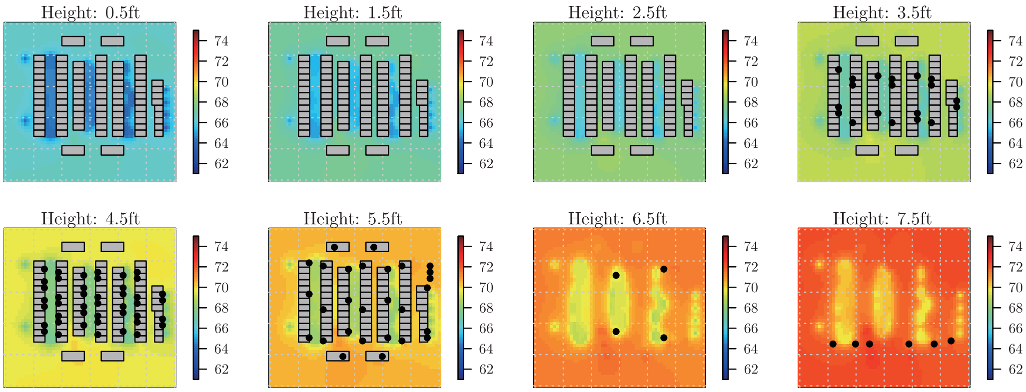

Figure 2. One-step ahead prediction of the temperature distribution map of the entire data center, where each subplot is a snapshot at a specified height.

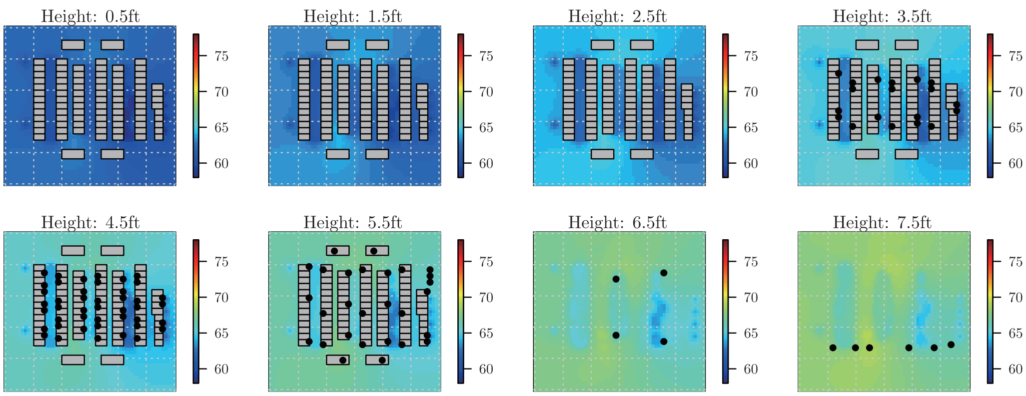

Figure 3. Lower prediction bound at 95% confidence level of one-step ahead prediction of the temperature distribution map of the entire data center, where each subplot is a snapshot at a specified height.

the cooling system on the operating temperature distribution. In particular, this predictive model may be tied in with an optimization framework that finds system settings to avoid over-cooling or over-heating by setting proper temperature of the air conditioning units.

6. Discussion

Analytics applied to data from monitoring networks is a growing trend in practice. An automated model fitting and forecast framework that can handle such data in a flexible and reliable manner is needed. We have introduced a framework that uses a generalized least squares approach and an empirical sample covariance matrix; with an automated test procedure, our method can detect the underlying nature well and can make appropriate adjustments in model fitting

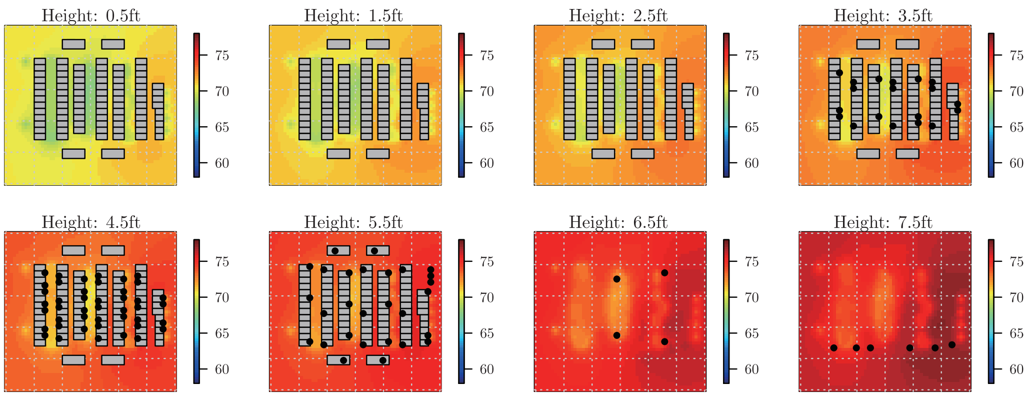

Figure 4. Upper prediction bound at 95% confidence level of one-step ahead prediction of the temperature distribution map of the entire data center, where each subplot is a snapshot at a specified height.

method for various practical situations. All required computation is designed to meet time and budget requirement simultaneously. As no procedure in our approach requires complicated optimization, the framework can be executed in an economical manner with minimal manual monitoring; it is already successfully implemented in some of our projects in the service industry.

Unlike the situations considered here, challenges will arise when the variable of interest is discrete, there is missing information in the dataset, or the location of the monitoring network changes over time. Research effort is needed to find procedures that can be run in a reliable and expeditious manner.

Acknowledgement

We would like to thank Hendrik Hamann at IBM Thomas J. Watson Research Center, and his world-wide MMT team for providing the problem and associated data center measurements used in this paper.

References

Brockwell, P. and Davis, R. (2002). *Introduction to Time Series and Forecasting*. 2nd edition. Springer, New York.

Cressie, N. (1993). *Statistics for Spatial Data*. Wiley-Interscience, New Jersey.

Cressie, N. and Wikle, C. K. (2011). *Statistics for Spatio-Temporal Data*. 1st edition. Wiley, New Jersey.

Fonseca, T. C. O. and Steel, M. F. J. (2011). A general class of nonseparable space-time covariance models. *Environmetrics* **22**, 224-242.

Gneiting, T. (2002). Nonseparable, stationary covariance functions for space-time data. *J. Amer. Statist. Assoc.* **2**, 590-600.

Gneiting, T., Genton, M. G., and Guttorp, P. (2007). Geostatistical space-time models, stationarity, separability and full symmetry. In *Statistical Methods for Spatio-Temporal Systems* (Edited by B. Finkenstadt, L. Held, and V. Isham), 151-175. Chapman & Hall, Boca Raton.

Gneiting, T. and Schlather, M. (2002). Space-time covariance models. In *Encyclopedia of Environmetrics* (Edited by A. El-Shaarawi and W. Piegorsch, Chichester) **4**, 2041-2045. Wiley.

Hamann, H., van Kessel, T., Iyengar, M., Chung, J.-Y., Hirt, W., Schappert, M., Claassen, A., Cook, J., Min, W., Amemiya, Y., Lopez, V., Lacey, L. and O'Boyle, M. (2009). Uncovering energy efficiency opportunities in data centers. *IBM Journal of Research and Development* **53**, 10:1-10:12.

Higdon, D., Swall, J. and Kern, J. (1999). Non-stationary spatial modeling. *Bayesian Statist.* **6**, 761-768.

Jun, M. and Stein, M. (2008). Nonstationary covariance models for global data. *Ann. Appl. Statist.* **2**, 1271-1289.

Mardia, K., Goodall, C., Redfern, E. and Alonso, F. (1998). The kriged Kalman filter. *Test* **7**, 217-282.

Munoz, B., Lesser, V. M. and Ramsey, F. L. (2008). Design-based empirical orthogonal function model for environmental monitoring data analysis. *Environmetrics* **19**, 805-817.

Nobrea, A., Sansob, B. and Schmidtc, A. (2011). Spatially varying autoregressive processes. *Technometrics* **53**, 310-321.

Nychka, D., Wikle, C. and Royle, K. (2002). Multiresolution models for nonstationary spatial covariance function. *Statist. Modelling* **2**, 315-332.

Obled, C. and Creutin, J. D. (1986). Some developments in the use of empirical orthogonal functions for mapping meteorological fields. *J. Climate and Appl. Meteorology* **25**, 1189-1204.

Rodrigues, A. and Diggle, P. (2010). A class of convolution-based models for spatio-temporal processes with non-separable covariance structure. *Scand. J. Statist.* **37**, 553-567.

Sampson, P. D. and Guttorp, P. (1992). Nonparametric estimation of nonstationary spatial covariance structure. *J. Amer. Statist. Assoc.* **87**, 108-119.

Shepard, D. (1968). A two-dimensional interpolation function for irregularly-spaced data. *Proceedings of the* 1968 *ACM National Conference*, 517-524.

Stein, M. L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging.* Springer, New York.

Stein, M. L. (2005). Space-time covariance functions. *J. Amer. Statist. Assoc.* **100**, 310-321.

IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598, U.S.A.

E-mail: huijiang@us.ibm.com

IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598, U.S.A.

E-mail: angelasch@gmail.com

IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598, U.S.A.

E-mail: yhwang@us.ibm.com

IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598, U.S.A.

E-mail: yasuo@us.ibm.com

(Received September 2013; accepted May 2014)