Published in final edited form as:

Biometrics. 2017 September ; 73(3): 759–768. doi:10.1111/biom.12656.

Modeling Nonstationarity in Space and Time

Lyndsay Shand and Bo Li

Department of Statistics, University of Illinois at Urbana-Champaign, U.S.A

Summary

We propose to model a spatio-temporal random field that has nonstationary covariance structure in both space and time domains by applying the concept of the dimension expansion method in Bornn et al. (2012). Simulations are conducted for both separable and nonseparable space-time covariance models, and the model is also illustrated with a streamflow dataset. Both simulation and data analyses show that modeling nonstationarity in both space and time can improve the predictive performance over stationary covariance models or models that are nonstationary in space but stationary in time.

Keywords

Dimension expansion; Nonstationarity; Space-time random field

1. Introduction

Interests in spatio-temporal data analysis are rising as such data has become more abundant in agriculture, atmospheric, hydrological and other environmental sciences. For example, the amount of precipitation at any given hour and unmonitored location depends on the precipitation in the surrounding areas prior to that hour. Understanding the underlying structures of spatio-temporal processes largely relies on covariance and variogram modeling techniques, the majority of which assume for simplicity that the process is stationary. This assumption, however, can be violated as real environmental spatio-temporal processes are very often inherently nonstationary in both space and time. For example, precipitation measures can greatly vary if taken prior to or during a moving storm as well as at locations outside the edges or at the eye of the storm.

The nonstationarity can be observed in either or both the space and time domains. Let $Y(\mathbf{s}, t)$ denote a space-time random field, where $(\mathbf{s}, t) \in \mathbb{R}^d \times \mathbb{R}$. The process $Y(\mathbf{s}, t)$ is nonstationary if either $E\{Y(\mathbf{s}, t)\}$ varies or the covariance $C\{Y(\mathbf{s}_i, t_i), Y(\mathbf{s}_j, t_j)\}$ for $(\mathbf{s}_i, t_i)$ and $(\mathbf{s}_j, t_j)$ in $\mathbb{R}^d \times \mathbb{R}$ depends on locations $\mathbf{s}_i$ and $\mathbf{s}_j$ and/or time points t_i and t_j rather than only their space-time lag. We focus on the nonstationarity that rises only from the heterogeneous correlation between observations separated by the same space-time lags, assuming constant mean and variance. This assumption is reasonable as large-scale spatially and temporally varying means and variances are often removed before analyzing the

dependence structure of $Y(\mathbf{s}, t)$. Nonstationarity can be visually diagnosed by examining the variability of the empirical covariances at the same space-time lags. Large variability indicates nonstationarity. A more rigorous test for stationarity is seen in Jun and Genton (2012). Another indication of nonstationarity can be derived from the physical formulation of the process. For instance, if a random process such as precipitation is generated from complex climate dynamics and affected by many environmental factors, then it is most likely nonstationary.

Modeling nonstationarity has been investigated for years and has recently received more attention due to high demands in practice. Most available models focus only on the nonstationarity in the spatial domain. For example, Sampson and Guttorp (1992) and Guttorp et al. (1994) proposed a nonstationary modeling framework based on deformation techniques (see also Schmidt and O'Hagan, 2003; Aberg et al., 2005). Higdon (1998) and Higdon et al. (1999) developed a flexible nonstationary model through the convolution of stationary spatial processes with spatially varying kernels (see also Risser and Calder, 2015; Wikle, 2002). Fuentes (2001, 2002) reframed the convolution method by allowing the latent process to be spatially dependent and Paciorek and Schervish (2004) generalized the kernel convolution approach to a class of nonstationary covariance functions. Bornn et al. (2012) proposed a different idea to model nonstationarity which assumed the nonstationary process is a projection of a stationary process in higher dimensions.

Recently, with the onset of big data, the spatial random effects model (e.g. Cressie and Johannesson, 2008; Katzfuss, 2013; Nychka et al., 2015) and the predictive process model (e.g. Banerjee et al., 2008; Eidsvik et al., 2012; Ren and Banerjee, 2013) were developed to make the computation feasible while carrying the ability to model nonstationarity. The main idea of those two types of models is to represent the full-rank random processes using reduced-rank basis functions or predictive processes. Although these models show big advantages in capturing large-scale correlation, Stein (2014) demonstrated possible statistical inefficiency of low rank models for spatial interpolation. The class of Markov random field models (Bolin and Lingren, 2011) constructed from nested statistical partial differential equations serves as another computationally efficient method to capture nonstationary features in spatial data.

Compared to the vast literature for modeling nonstationarity in space, nonstationarity in both space and time is studied much less extensively. Ma (2002) derived nonstationary space-time covariance functions by applying certain kernels to stationary covariance functions. Garg et al. (2012) extended the idea of convolution processes in Higdon et al. (1999) and Paciorek and Schervish (2004) to model a nonstationary spatio-temporal Gaussian process. Set in a Bayesian framework, Sigrist et al. (2012) proposed a dynamic nonstationary spatio-temporal model for short term precipitation by linking the advection parameter, describing the horizontal transport of rainfall, of the convolution kernel to an external wind vector, measured independently of precipitation. Stroud et al. (2001) achieved nonstationarity in time by allowing the regression coefficients in a locally weighted mixture model to temporally vary. Huang and Hsu (2004) extended Wikle and Cressie (1999) to develop a space-time Kalman filter where the spatio-temporal covariance function depends on covariates.

We propose modeling nonstationarity in both space and time for a spatio-temporal process by applying the dimension expansion technique described in Bornn et al. (2012). Spatio-temporal data often exhibit temporal correlation in addition to spatial and both may not be stationary. It is thus important to build a flexible space-time model which accommodates nonstationarity in both domains. Unlike previously proposed nonstationary space-time models, dimension expansion allows us to make use of the plethora of stationary covariance models available. Numerous studies have been conducted on developing stationary space-time covariance models, e.g., Gneiting (2002), Stein (2005), Fonesca and Steel (2011) and Choi et al. (2013). Our model is demonstrated to have advantages over the nonstationary spatial model in Bornn et al. (2012) as well as over a combination of this nonstationary spatial model with a stationary temporal model. Compared to the low rank models, our model focuses on fine-scale correlation.

The paper is organized as follows: Section 2 details the dimension expansion technique, investigates model fitting and proposes an approach to deal with computation for large datasets. Section 3 studies the properties of the nonstationary model and compares it to other models through simulated data, while Section 4 applies the nonstationary model to a spatio-temporal streamflow dataset. Finally, Section 5 provides a conclusion and discussion.

2. Dimension Expansion in Space and Time

Perrin and Meiring (2003) showed that any nonstationary random field in $\mathbb{R}^n$, with moments of at least order 2, can be interpreted as a lower-dimensional representation of a second-order stationary random field in a higher dimensional space $\mathbb{R}^{2n}$, provided that the components of the random vector have identical expectations and variances. Further they established that a bijective mapping exists between the two spaces. Perrin and Schlather (2007) generalized this idea and proved that any Gaussian random vector can be interpreted as a sample from a stationary random function on a graph in $\mathbb{R}^d$, $d \geq 2$, subject to the same moment constraints as in Perrin and Meiring (2003). An intuitive example is that of suppressing the elevation from a stationary 3- d temperature process, leading to a nonstationary 2- d process.

Based on the above theoretical results, Bornn et al. (2012) justified their dimension augmentation procedure by stating that a realization of a nonstationary Gaussian process in $\mathbb{R}^d$ may be interpreted as a realization of a stationary field in $\mathbb{R}^{d+p}$ for $p > 0$ under appropriate moment constraints. Since a spatio-temporal random field can be considered a random field in $\mathbb{R}^d \times \mathbb{R}$, it follows that under appropriate moments constraints, the nonstationary space-time process $Y(\mathbf{s}, t)$ with $(\mathbf{s}, t) \in \mathbb{R}^d \times \mathbb{R}$ can be represented by a stationary process $Y([\mathbf{s}, \mathbf{z}], [t, \mathbf{w}])$ with $[\mathbf{z}, \mathbf{w}] \in \mathbb{R}^p \times \mathbb{R}^q$ for $p+q > 0$, that is, the process Y is stationary in the space $\mathbb{R}^{d+p} \times \mathbb{R}^{1+q}$ although nonstationary in $\mathbb{R}^d \times \mathbb{R}$. Then, after estimating the augmented dimensions $\mathbf{z}$ and $\mathbf{w}$, any valid stationary space-time covariance model in $\mathbb{R}^{d+p} \times \mathbb{R}^{1+q}$ will be applicable to $Y([\mathbf{s}, \mathbf{z}], [t, \mathbf{w}])$. Following traditional spatio-temporal modeling, we treat the time and spatial domains separately to reflect their different characteristics.

2.1 Review of dimension expansion in space

Let $Y(\mathbf{s}), \mathbf{s} \in \mathbb{R}^d$ be a nonstationary spatial process, and suppose $\mathbf{z} \in \mathbb{R}^p, p > 0$ are the latent p dimensions such that $Y([\mathbf{s}, \mathbf{z}])$ is stationary with semivariogram $\gamma_\theta(\cdot)$ defined as

$$\gamma_\theta([\mathbf{s}_i, \mathbf{z}_i] - [\mathbf{s}_j, \mathbf{z}_j]) = \frac{1}{2} E \left[\{Y([\mathbf{s}_i, \mathbf{z}_i]) - Y([\mathbf{s}_j, \mathbf{z}_j])\}^2 \right],$$

where $[\mathbf{s}_i, \mathbf{z}_i]$ is the i th element of $[\mathbf{S}, \mathbf{Z}]$, the concatenation of the dimensions $\mathbf{S}$ and $\mathbf{Z}$. One can model the nonstationarity in $Y(\mathbf{s})$ by a stationary variogram model defined for $Y[\mathbf{s}, \mathbf{z}]$. Note any valid stationary variogram model is a legitimate choice for $\gamma(\cdot)$. Bornn et al. (2012) proposed to estimate the latent dimensions $\mathbf{z}$ using lasso-penalized least squares:

$$\hat{\theta}, \mathbf{z} = \underset{\theta, \mathbf{z}}{\operatorname{argmin}} \sum_{i < j} [\nu_{i,j} - \gamma_\theta\{d_{i,j}([\mathbf{S}, \mathbf{Z}])\}]^2 + \lambda_{s,1} \sum_{k=1}^p \|\mathbf{Z}_{\cdot k}\|_1, \quad (1)$$

where $\nu_{i,j}$ are the moment estimates of $\gamma_\theta(\cdot)$, $d_{i,j}([\mathbf{S}, \mathbf{Z}])$ is the i, j th element of the distance matrix of the augmented locations $[\mathbf{S}, \mathbf{Z}]$, $\mathbf{Z}_{\cdot k}$ is the k th column (dimension) of $\mathbf{Z}$ and $\|\cdot\|_1$ is the L_1 norm. In their model fitting, time points at different locations are treated as replicates to compute the empirical spatial variogram. The purpose of the lasso group penalty parameter, $\lambda_{s,1}$, is to regularize the estimation of $\mathbf{z}$ and prevent an overfitting of the spatial dimensions by controlling the dimension sparsity of $\mathbf{z}$. Finally, a bijective function $f(\cdot)$ between $\mathbf{s}$ and $\mathbf{z}$ can be established using thin-plate splines based on the estimates of $\mathbf{z}$.

2.2 Extend dimension expansion to space-time

In addition to spatial correlation, spatio-temporal data often exhibit correlation in time, which can also be nonstationary. Nevertheless, it is still quite common that nonstationary spatio-temporal models only address spatial nonstationarity. Allowing dimension expansion in both the space and time domains, however, allows us to relax all stationary assumptions while enjoying already established stationary models. We now illustrate how to enable the rich class of stationary models to capture nonstationary features in data.

We will use two classes of stationary space-time covariance models to exemplify this approach. These two classes of models will also be used in our simulation studies in Section 3. The first one is the space-time separable covariance function. The separable model is often an oversimplification of the dependence structure in real data, yet it is widely used for its simplicity and computational efficiency. Without loss of generality, we choose the exponential covariance function for both the space and time domain:

$$C(h, u) = \sigma^2 \exp(-\phi_s h) \exp(-\phi_t u), \quad (2)$$

where h and u are the Euclidian distances of the spatial lag $\mathbf{h} \in \mathbb{R}^{d+p}$ and temporal lag $\mathbf{u} \in \mathbb{R}^{1+q}$, respectively, and ϕ_s and ϕ_t are range parameters in space and time domains. Suppose $([\mathbf{s}_i, \mathbf{z}_i], [t_i, \mathbf{w}_i])$ and $([\mathbf{s}_j, \mathbf{z}_j], [t_j, \mathbf{w}_j])$ are two locations in the expanded space, then

$$\mathbf{h} = [\mathbf{s}_i - \mathbf{s}_j, \mathbf{z}_i - \mathbf{z}_j], \mathbf{u} = [t_i - t_j, \mathbf{w}_i - \mathbf{w}_j],$$

and $h = \|\mathbf{h}\|$, $u = \|\mathbf{u}\|$, where $\|\mathbf{a}\|$ is the Euclidean norm of a vector $\mathbf{a}$. Although model (2) is obviously a stationary model in $\mathbb{R}^{d+p} \times \mathbb{R}^{1+q}$, it can model the nonstationary correlation structure of the process Y in a lower dimensional space, e.g., in $\mathbb{R}^d \times \mathbb{R}$.

The second example is the class of nonseparable space-time models. These models are more flexible though usually involve more parameters and are more computationally challenging. Gneiting (2002) proposed a general form for nonseparable covariance functions, and we adopted one particular form for our illustration:

$$C(h, u) = \begin{cases} \sigma^2(a|u|^{2\alpha} + 1)^{-1}, & \text{if } h=0 \\ \sigma^2(a|u|^{2\alpha} + 1)^{-1} \exp\left\{\frac{-ch}{(a|u|^{2\alpha} + 1)^{\beta/2}}\right\}, & \text{otherwise,} \end{cases} \quad (3)$$

where h and u follow the notation in model (2), a and c are nonnegative scaling parameters of time and space respectively, $\alpha \in (0, 1]$ is a smoothing parameter and σ^2 is the variance of the process. The parameter $\beta \in [0, 1]$ measures the space-time interaction where a larger β indicates a stronger interaction or weaker separability between space and time components. Model (3) is developed for stationary processes, but with distance lags $\mathbf{u} \in \mathbb{R}^{1+q}$ and $\mathbf{h} \in \mathbb{R}^{d+p}$, it is capable of modeling nonstationary processes observed in lower dimensional spaces. Although dimension expansion applies to all covariance structures, separable or nonseparable, a separability test, e.g. Li et al. (2007), can help choose an appropriate model.

2.3 Learning latent dimensions and model fitting

Let $\boldsymbol{\theta}$ denote the vector of parameters in the stationary space-time covariance function. The nonstationary model using expanded dimension(s) has three sets of unknown parameters: $\boldsymbol{\theta}$, $\mathbf{z}$, and $\mathbf{w}$. To reflect that $\mathbf{z}$ and $\mathbf{w}$ essentially form two continuous processes, we denote them as $\mathbf{z}(\mathbf{s})$ and $\mathbf{w}(t)$. Note that the dimensions of $\mathbf{z}(\mathbf{s})$ and $\mathbf{w}(t)$, p and q , also remain unknown. A simultaneous estimation for all parameters at once is obviously not ideal from a numerical point of view. We therefore propose to estimate the parameters in two steps. First we estimate $\mathbf{z}(\mathbf{s})$ and $\mathbf{w}(t)$ using penalized least squares and then we estimate $\boldsymbol{\theta}$ using likelihood method based on the estimated latent dimensions.

To estimate $\mathbf{z}(\mathbf{s})$ and $\mathbf{w}(t)$, it seems natural to extend the estimation method in (1) into the space-time context so that we have

$$\begin{aligned} \hat{\boldsymbol{\theta}}, \mathbf{z}(\mathbf{s}), \mathbf{w}(t) = & \operatorname{argmin}_{\boldsymbol{\theta}, \mathbf{z}, \mathbf{w}} \sum_{i < j} \left[\hat{C} \{(\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)\} \right. \\ & \left. - C_{\boldsymbol{\theta}}(\|\mathbf{s}_i - \mathbf{s}_j, \mathbf{z}(\mathbf{s}_i) - \mathbf{z}(\mathbf{s}_j)\|, \|t_i - t_j, \mathbf{w}(t_i) - \mathbf{w}(t_j)\|) \right]^2 \\ & + \lambda_{s,1} \sum_{k=1}^p \|\mathbf{Z}_{\cdot k}\|_1 + \lambda_{t,1} \sum_{k=1}^q \|\mathbf{W}_{\cdot k}\|_1, \end{aligned} \quad (4)$$

where $\hat{C}(\cdot)$ is an empirical estimate of the covariance function. However, estimating $\mathbf{z}(\mathbf{s})$ and $\mathbf{w}(t)$ simultaneously is not accessible in our case as there are simply no replicates to compute $\hat{C}(\cdot)$ unless there are repeat measurements at each $(\mathbf{s}_i, t_i)$. To make the estimation procedure more feasible, we propose to estimate $\mathbf{z}(\mathbf{s})$ and $\mathbf{w}(t)$ separately.

Treating the spatial locations as replicates for a given time point, we first estimate the temporal covariance function using moment estimator:

$$\hat{C}(t_i, t_j) = \frac{1}{|D|} \sum_{\mathbf{s} \in D} \{Y(\mathbf{s}, t_i)Y(\mathbf{s}, t_j)\},$$

where $|D|$ is the cardinality of the spatial domain D . Then we estimate $\mathbf{w}(t)$ by minimizing the following lasso penalized least squares:

$$\hat{\boldsymbol{\theta}}_t, \mathbf{w} = \underset{\boldsymbol{\theta}_t, \mathbf{w}}{\operatorname{argmin}} \sum_{i < j} [\hat{C}(t_i, t_j) - C_{\boldsymbol{\theta}_t} \{\|t_i - t_j, \mathbf{w}(t_i) - \mathbf{w}(t_j)\|\}]^2 + \lambda_{t,1} \sum_{k=1}^q \|\mathbf{w}_{\cdot k}\|_1, \quad (5)$$

where $\boldsymbol{\theta}_t$ is the vector of parameters only involved in the pure temporal covariance function. The estimation of $\mathbf{z}(\mathbf{s})$ follows Bornn et al. (2012) by treating the observations at different time points as replicates for a given location. After we estimate the point estimates $\mathbf{z}(\mathbf{s}_i)$ and $\mathbf{w}(t_j)$ for $i = 1, \dots, n$ and $j = 1, \dots, T$, we then build thin-plate splines $g_s(\mathbf{s}) \approx \mathbf{z}$ and $g_t(t) \approx \mathbf{w}$, with smoothing parameters $\lambda_{s,2}$ and $\lambda_{t,2}$ respectively, to obtain the continuous processes of $\mathbf{z}(\mathbf{s})$ and $\mathbf{w}(t)$. Another example of using thin-plate splines in modeling nonstationary covariance structures can be seen in Sampson and Guttorp (1992). The choice of all tuning parameters is determined by cross validation for which the target is to minimize the prediction errors. Finally, we reestimate $\boldsymbol{\theta}$ using the maximum likelihood (ML) method based on the observed dimension and estimated latent dimensions $\{[\mathbf{s}, \mathbf{z}], [t, \mathbf{w}]\}$. Although estimates for $\boldsymbol{\theta}$ are obtained in (1) and (5), only the estimates for $\mathbf{w}$ and $\mathbf{z}$ are kept from these steps. ML estimates for $\boldsymbol{\theta}$ are preferred since the MLE is asymptotically most efficient.

The separate estimation of latent dimensions in space and time may seem suboptimal, especially when fitting a nonseparable covariance model. We carried out a small simulation estimating $\mathbf{w}$ and $\mathbf{z}$ together by replacing $\hat{C}\{(\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)\}$ in (4) with $Y(\mathbf{s}_i, t_i)Y(\mathbf{s}_j, t_j)$ so that all interactions between space and time are taken into account. Due to the large number of unknown parameters to be estimated and great variability in each individual pair, the simultaneous estimates of $\mathbf{w}$ and $\mathbf{z}$ performed worse than our proposed method in addition to being overly time consuming. The deteriorated performance of estimators that attempt to use all individual pairs directly is also observed in Choi et al. (2013).

Additionally we explored two other methods to estimate $\mathbf{w}$ and $\mathbf{z}$: an iterative procedure, performed by estimating $\mathbf{w}$ given $\mathbf{z}$ and then $\mathbf{z}$ given $\mathbf{w}$ using (4) until convergence, and a 2-step hybrid procedure where we estimate $\mathbf{z}$ using (1) and then $\mathbf{w}$ given $\mathbf{z}$ using (4). Although we found both methods performed comparably to our current method, the relative computational efficiency makes the current method preferable. Finally, when data contain outliers, we can consider a more robust estimation method by replacing the classical

empirical variogram and covariance estimates in (1) and (5) respectively with more robust estimates (e.g. Cressie and Hawkins, 1980). A simulation study not presented here shows that the robust estimator may improve the prediction when outliers are present.

2.4 Reducing computational burden for a large dataset

Dimension expansion can be computationally expensive as the number of spatial locations, n , or time points, T increases. Bornn et al. (2012) mentioned that more complex optimization methods are necessary when n exceeds 100 and the number of estimated dimensions, p is greater than 3. To tackle the computational issue without resorting to complex optimization methods, we propose to use a small subset of spatial locations and time points to estimate $\mathbf{z}(\mathbf{s})$ and $\mathbf{w}(t)$ respectively. Specifically, we take a subset of n_0 locations and T_0 time points and then estimate $\mathbf{z}(s_i)$ and $\mathbf{w}(t_j)$ for $i = 1, \dots, n_0$ and $j = 1, \dots, T_0$ based on this subset. Then we build thin-plate splines $g_t(\cdot)$ and $g_s(\cdot)$, with respective smoothing parameters $\lambda_{s,2}$, and $\lambda_{t,2}$, only based on these point estimates and apply those functions to the whole domain to obtain the continuous processes $\mathbf{z}(\mathbf{s})$ and $\mathbf{w}(t)$. Tuning parameters $\lambda_{t,1}$, $\lambda_{s,1}$, $\lambda_{t,2}$, and $\lambda_{s,2}$ are chosen by minimizing the drop-one hundred MSPE of the $n_0 \times T_0$ sampled observations. The subset of spatial locations can be randomly sampled from the spatial domain, whereas we suggest the T_0 time points to be selected strategically by first sectioning the T time points into blocks and then sampling a fraction of time points from each block. We recommend to always include the first and last time points in the sample to avoid extrapolation when using thin-plate splines to estimate $\mathbf{w}(t)$. In the case of a large number of spatial locations over a larger spatial domain, a similar sampling strategy may also need to be implemented to ensure that the selected sample has adequate spatial coverage. Note that in the case of fitting a separable space-time covariance function, one can take advantage of properties of the Kronecker product to further reduce the computation.

3. Simulation study

3.1 Basic setup

We perform simulation studies to demonstrate the advantages of using our method when the data are nonstationary in both space and time. Two space-time covariance functions are used to simulate spatio-temporal processes. One is the separable model in (2) and the other is the nonseparable model in (3). For each model type, we set parameters at different values to eliminate the effect of particular parameters on the simulation results. Specifically, as shown in Table 1, we set $\phi_t = \phi_s = 0.3, 0.5, 0.8$ in the separable model (2) to simulate three processes representing strong, fair and weak spatial and temporal correlations, respectively. To simulate a variety of processes with different properties using the nonseparable model (3), we set smoothing parameter $\alpha = 0.5$ and set $\beta = 0.5, 1$ to represent weak and strong space-time interactions. We also set $a = 2, 7$ and $c = 1, 2$ to represent strong and weak temporal and spatial correlation, respectively. We set $\sigma^2 = 1$ for all separable and nonseparable models.

At each parameter setting, 200 zero-mean Gaussian spatio-temporal processes, $\{Y(\mathbf{s}_i, \mathbf{t}_j)\}$, $i = 1, \dots, 30, j = 1, \dots, 30$, are simulated at 30 randomly chosen locations and 30 time points.

Here we set spatial locations $\mathbf{s}_i = (s_i^1, s_i^2, s_i^3) \in \mathbb{R}^3$ and time locations $\mathbf{t}_j = (t_j^1, t_j^2) \in \mathbb{R}^2$, that is, we generate stationary Gaussian random processes in $\mathbb{R}^3 \times \mathbb{R}^2$. The time points t_j^1 are equally spaced between 0 and 1 and the latent dimension t_j^2 are generated as a quadratic function of t_j^1 . The spatial locations (s_i^1, s_i^2) are generated uniformly on a unit circular region centered at (0, 0) and the latent dimension s_i^3 is generated such that (s_i^1, s_i^2, s_i^3) are centered on a three-dimensional half-ellipsoid. Then we project the processes onto the reduced dimensional space $\mathbb{R}^2 \times \mathbb{R}$ by retaining only (s_i^1, s_i^2) and t_j^1 in their spatial and temporal domains. The spatio-temporal processes in the $\mathbb{R}^2 \times \mathbb{R}$ space exhibit nonstationarity.

In the simulations, all parameters are estimated using (1) and (5). Note that in the cross validation for identifying the tuning parameters for the estimation of $\mathbf{z}(s)$ and $\mathbf{w}(t)$, we again propose to substitute the nonseparable covariance function by its corresponding separable form to ease the computation. Since we will eventually discard the covariance parameter estimates yielded during the estimation of latent dimension and use ML to estimate $\boldsymbol{\theta}$ of the whole space-time covariance function based on the recovered latent dimensions, the approximations of $\boldsymbol{\theta}$ obtained when estimating latent dimensions have very little effect on covariance modeling. Through another small simulation study we found that the prediction errors using our method are comparable to those using the full nonseparable model to estimate the tuning parameters and latent dimensions.

To evaluate the strength of modeling nonstationarity in both space and time, we compare its predictive performance to several other models. The nonstationary separable model allows us to compare it with the (i) true model as a reference; (ii) stationary space-time model which simply applies a stationary model to the observed dimensions; (iii) Bornn et al.'s method that treats the temporal replicates as independent, hereinafter referred to as BSZ. This model takes advantage of the latent dimensions in the spatial domain but ignores the correlation in time; (iv) Bornn et al.'s method in combination with a stationary temporal model, hereinafter referred to as BSZ+ST. This model also considers the latent dimensions in the spatial domain but not in the time domain. The nonstationary nonseparable model only allows us to compare with models (i), (ii) and (iv). We use the mean squared prediction error (MSPE) and log scores of drop-one predictions as measures for the predictive performance:

$$\text{MSPE} = \frac{1}{nT} \sum_{i=1}^n \sum_{j=1}^T \left\{ Y(\mathbf{s}_i, t_j) - \hat{Y}(\mathbf{s}_i, t_j) \right\}^2,$$

$$\text{Log Score} = \frac{1}{nT} \sum_{i=1}^n \sum_{j=1}^T \left[\log \left(2\pi \hat{\sigma}_{ij}^2 \right) + \left\{ Y(\mathbf{s}_i, t_j) - \hat{Y}(\mathbf{s}_i, t_j) \right\}^2 / \hat{\sigma}_{ij}^2 \right]$$

where $\hat{\sigma}_{ij}^2$ is the prediction variance for $Y(\mathbf{s}_i, t_j)$.

Figure 1 shows an example of the empirical covariance in the expanded dimensional space using the estimated latent dimensions compared to that in the reduced dimensional space and true dimensions, based on a randomly chosen simulated dataset using the separable model. The high variability of empirical covariances at each spatial lag in Figure 1 (c) and (d) largely indicates possible nonstationarity. In contrast, tight bands of points exhibited in Figure 1 (a) and (b) or (e) and (f) are a sign of stationarity. The euclidean distances u and h in Figure 1(a–f) are calculated using the respective dimensions from each of the three settings. It is seen that the goodness-of-fit of the MLE of the covariance model in the expanded dimensional space is much improved over the fit in the reduced dimensional space.

3.2 Numerical results

Table 1 reports the MSPE and log score corresponding to each parameter setting for the separable model and Table 2 for the nonseparable model. In the separable case, we see that across all parameter combinations, MSPE and log score are the smallest for the true model, but the nonstationary model we propose is the most comparable. The BSZ+ST model is the third best while the stationary and BSZ models perform the worst. Improvements in MSPE and log score of the nonstationary model over others are similar across all weak, moderate and strong spatial and temporal correlations. It is interesting to note that the BSZ model performs worse than the stationary model when there is temporal correlation present, and the discrepancy is more obvious when the temporal correlation is stronger. This indicates that the BSZ model can be insufficient for a spatio-temporal data set.

For the nonseparable case, we again see that the nonstationary model is most comparable to the true model in terms of both MSPE and log score. We also see that the nonstationary model consistently outperforms the BSZ+ST across all parameter combinations except when there is low temporal correlation and strong spatial correlation as with $a = 7$ and $c = 1$. However, the improvement is less prominent compared to the separable case, in particular when comparing the log scores. In the case of $a = 7$, $c = 1$, the two models perform very similarly. This result is not surprising as when the temporal correlation is relative weaker than the spatial correlation, modeling of the temporal correlation, whether stationary or nonstationary, becomes less important. Therefore, the BSZ+ST model is nearly as sufficient as the nonstationary model in this case. The stationary model again consistently performs the worst across all parameters combinations. The pattern of the results seems insensitive to the value of the separability parameter.

To explore the efficiency of the sampling method for larger datasets described in Section 2.4, 100 spatio-temporal processes are simulated at $n = 100$ spatial locations and $T = 100$ time points using the separable covariance function (2) with $\sigma^2 = 1$ and $\phi_t = \phi_s = 0.5$. For each simulated dataset, we sampled 900 of the observations by randomly selecting 30 spatial locations from the 100 locations, and 30 time points from the 100 time points. We also sampled 2,500 observations by sampling 50 spatial locations and 50 time points. Details of the random selection of locations and time points are described in Section 2.4. Then using the 900, 2,500 and all of the observations, we estimate the latent dimensions $\mathbf{w}(t)$ and $\mathbf{z}(s)$, fit the nonstationary model (2) and compute the MSPE and log score of 100 randomly

dropped observations from the whole dataset. The results are reported in Table 3. We see a gradual improvement in prediction when more and more observations are used to estimate the latent dimensions. Even with only 900 of the observations, the nonstationary model has significantly reduced the MSPE of the stationary model by 14%. With 2,500 of the observations, the nonstationary model reduces the MSPE of the stationary model by 25%. Figure 2 compares the true latent dimensions s_{3j} and t_{2j} , $i = 1, \dots, n$, $j = 1, \dots, T$ with the estimated dimensions $\mathbf{w}$ and $\mathbf{z}$ using only 900 observations and those found using all data points. It is seen that the estimates using 900 observations are already similar to those using the whole data set and both share the same pattern as the true latent dimensions. These results show that strategically taking a subset of the data could be used to efficiently apply dimension expansion to a large space-time dataset.

4. Application to streamflow data

To further illustrate that our nonstationary space-time covariance model can be useful in practice, we apply it to a streamflow dataset. Daily average discharge of streams from 28 stations in Northern Illinois and Wisconsin during June through October 2013 was obtained from the US Geological Survey, see Figure 3a for a station map. The discharge, measured in ft^3/s , is the volume of water passing through any given point in the stream. Due to of lake and urbanization effects near Madison, WI and Lake Koshkonong on Rock River, Rock River stations below Lake Koshkonong and with extreme drainage areas were excluded from consideration. A stream is a form of catchment or drainage basin, i.e. an extent of land where surface water from rain, snow melt, etc. converges and drains to a lower elevation. Thus, the discharge at a station depends on the water level at an unknown point of higher elevation upstream within the drainage basin of that station. Considering upstream elevation as a latent dimension and only observing latitude and longitude coordinates of stations, we apply dimension expansion in an attempt to mimic the effect of elevation upstream of each station.

We first take a log transformation of the data and remove the station mean of 2013. We then apply dimension expansion with specified separable exponential model (2). The nonstationary model expands time and space dimensions to $[t^1, w] \in \mathbb{R}^2$ and $[s^1, s^2, z] \in \mathbb{R}^3$, $i = 1, \dots, n$, $j = 1, \dots, T$, while the stationary model uses only the observed dimensions $t^1 \in \mathbb{N}$ and $[s^1, s^2] \in \mathbb{R}^2$. To learn the additional dimensions, every 3rd day within this time frame was used to reduce the number of time points from 151 to 51 and to reduce the sample size from 4,228 to 1,428. Additional dimensions for time and space, $\mathbf{w}(t)$ and $\mathbf{z}(\mathbf{s})$ respectively, were estimated with optimal tuning parameters: $\lambda_{t,1} = 4$, $\lambda_{s,1} = 0.08$, $\lambda_{t,2} = 10e^{-5}$, and $\lambda_{2,s} = 10e^{-4}$, which were chosen by minimizing the drop-one hundred MSPE.

Figure 3b shows the estimated latent dimensions, $\mathbf{w}(t)$ and $\mathbf{z}(\mathbf{s})$. The relatively large magnitudes of $\mathbf{w}$ and $\mathbf{z}$ indicate that their influence is not negligible, meaning we would expect a great prediction improvement by our model over the stationary model. We then fit the model using the estimated latent dimensions and compare the MSPE and log score of drop-one hundred predictions of our nonstationary model with BSZ+ST, BSZ and stationary models. To make all models comparable, the latter three models follow the basic form of (2). Our nonstationary model performs best with MSPE of 0.131 and log score of 0.216 while

the stationary model performs the worst with MSPE of 0.165 and log score of 1.15. The BSZ+ST model is the second best with MSPE of 0.149 and log score of 0.471 and BSZ performs similarly with MSPE 0.140 and log score of 0.514. Although the MSPEs of all models except for the stationary case seem similar, the log score is notably smaller for the nonstationary model. This indicates that modeling nonstationarity in both space and time can improve the prediction variance estimation for certain spatio-temporal data.

The density plot of the streamflow data shows a weak sign of outliers so we additionally carry out the robust estimation by employing the robust variogram estimator proposed by Cressie and Hawkins (1980). The results are very similar to the above suggesting robust estimation may not be necessary for this dataset. We also applied the nonstationary spacetime models in Ma (2002) and Garg et al. (2012) to this data set. The drop-100 MSPE from those two methods are 0.264 and 0.495 and log scores 1.572 and 543.8, respectively. We conjecture that their poor performance could be because the model in Ma (2002) is too parsimonious for this data as it only involves a few parameters, while the model in Garg et al. (2012) may have overfitting issues as it involves many parameters but no penalty on the exibility of the model is discussed in their model fitting procedure.

5. Discussion

Many spatio-temporal datasets in the natural sciences exhibit nonstationary dependence structures in both space and time due to various reasons, although most existing space-time covariance models assume second order stationarity. Dimension expansion in both space and time offers an easy way to allow for nonstationarity while still enjoying the abundance of traditional stationary covariance models. We investigate the estimation of parameters and latent dimensions in space and time for separable and nonseparable space-time covariance models and illustrate our method in simulation studies and a real data application. It is seen that our method can significantly improve prediction compared to using observed dimensions when the data are nonstationary. Therefore, the comparison may give rise to a crude way to test for stationarity, an alternative to the test proposed by Jun and Genton (2012).

There is no definitive answer to the question of how many latent dimensions are sufficient to attain stationarity. In practice, the number of expanded dimensions p and q should be chosen as the fewest possible to construct a stationary space. This can be explored by a try-error strategy, gradually increasing p and q . The lasso penalty in equations (1) and (5) will impose the coordinates of unnecessary additional dimensions to be close to zero, indicating the best choice. In our experience, we have found one to two extra dimensions to be sufficient.

Our paper mainly focuses on modeling the covariance structure at fine resolution and we illustrate the method using small datasets. Dimension expansion can be computationally expensive as the number of extra dimensions, p or q , or the number of space-time locations, n or T increases. For large spatial data, Bornn et al. (2012) suggested the gradient projection method of Kim et al. (2006) for the optimization. Without resorting to more complex optimization algorithms, we propose a strategic approach to reduce the computation by estimating the latent dimensions using only a subset of observations. This method shows

great potential in practice. Another promising approach to deal with computation for large data is the binned method proposed by Kang et al. (2010), which divides the entire data set into bins, treating each bin as an observation and the locations within each bin as replicates.

The identifiability of latent dimensions is not rigorously investigated here. Bornn et al. (2012) briefly discussed the conditions under which the precise latent dimensions are identifiable, but they emphasized that dimension expansion mainly focuses on approximating a real process rather than the estimation of the latent dimensions. We agree that the main merit of using latent dimensions lies in their practical advantages and that the estimation method is imperfect. Furthermore, we note that the estimation of latent dimensions depends on the choice of the covariance model. Given a different choice, the estimates of latent dimensions will likely be different. Hence the method of dimension expansion is essentially an attempt of approximating the nonstationarity based on a given covariance model.

Acknowledgments

The authors thank the editor, the associate editor, and the referees for constructive suggestions that have improved the content and presentation of this article. The authors also thank Dr. Luke Bornn and Mr. Sahil Garg for helpful discussions on this work, Dr. Thomas M. Over for providing the data, and James Balamuta for his help on data visualization. We acknowledge partial support from NSF grants DPP-1418339 and AGS-1602845 and NIH grant R56.

References

- Aberg S, Lindgren F, Malmberg A, Holst J, Holst U. An image warping approach to spatio-temporal modeling. *Environmetrics*. 2005; 16:833–848.
- Banerjee S, Gelfand AE, Finley AO, Sang H. Gaussian predictive models for large spatial datasets. *Journal of the Royal Statistical Society Series B*. 2008; 70:825–848.
- Bolin D, Lingren F. Spatial models generated by nested stochastic partial differential equations, with an application to global ozone mapping. *AOAS*. 2011; 5:532–550.
- Bornn L, Shaddick G, Zidek JV. Modeling nonstationary processes through dimension expansion. *Journal of the American Statistical Association*. 2012; 107:281–289.
- Choi I, Li B, Wang X. Nonparametric estimation of spatial and space-time covariance functions. *Journal of Agriculture, Biological and Environmental Statistics*. 2013; 18:611–630.
- Cressie N, Hawkins DM. Estimation of the variogram. *Journal of the International Association for Mathematical Geology*. 1980; 12:115–125.
- Cressie N, Johannesson G. Fixed rank kriging for very large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2008; 70:209–226.
- Eidsvik J, Finley AO, Banerjee S, Rue H. Approximate bayesian inference for large spatial datasets using predictive process models. *Computational Statistics and Data Analysis*. 2012; 56:1362–1380.
- Fonesca TC, Steel MFJ. A general class of nonseparable space-time covariance models. *Environmetrics*. 2011; 22:224–242.
- Fuentes M. A high frequency kriging approach for non-stationary environmental processes. *Environmetrics*. 2001; 12:469–483.
- Fuentes M. Spectral methods for nonstationary spatial processes. *Biometrika*. 2002; 89:197–210.
- Garg, S., Singh, A., Ramos, F. Learning non-stationary space-time models for environmental monitoring; *Proceedings of the 26th AAAI Conference on Artificial Intelligence*; 2012.
- Gneiting T. Nonseparable, stationary covariance functions for space-time data. *Journal of the American Statistical Association*. 2002; 97:590–600.

- Guttorp P, Meiring W, Sampson PD. A space-time analysis of ground-level ozone data. *Environmetrics*. 1994; 5:241–254.
- Higdon D. A process-convolution approach to modeling temperatures in the north atlantic ocean. *Environmental and Ecological Statistics*. 1998; 5:173–190.
- Higdon, DM., Swall, J., Kern, J. Non-stationary spatial modeling. In: JMB, et al., editors. *Bayesian Statistics*. Vol. 6. Oxford University Press; 1999. p. 761–768.
- Huang H, Hsu N-J. Modeling transport effects on ground-level ozone using a non-stationary space-time model. *Environmetrics*. 2004; 15:251–268.
- Jun M, Genton MG. A test for stationarity of spatio-temporal random fields on planar and spherical domains. *Statistica Sinica*. 2012; 22:1737–1764.
- Kang EL, Cressie N, Shi T. Using temporal variability to improve spatial mapping with application to satellite data. *Canadian Journal of Statistics*. 2010; 38:271–289.
- Katzfuss M. Bayesian nonstationary spatial modeling for very large datasets. *Environmetrics*. 2013; 24:189–200.
- Kim Y, Kim J, Kim Y. Blockwise sparse regression. *Statistica Sinica*. 2006; 16:375–390.
- Li B, Genton MG, Sherman M. A nonparametric assessment of properties of space-time covariance functions. *JASA*. 2007; 102:736–744.
- Ma C. Spatio-temporal covariance functions generated by mixtures. *Mathematical Geology*. 2002; 34:965–975.
- Nychka D, Bandyopadhyay S, Hammerling D, Lindgren F, Sain S. A multiresolution gaussian process model for the analysis of large spatial datasets. *Journal of Computational and Graphical Statistics*. 2015; 24:579–599.
- Paciorek CJ, Schervish MJ. Nonstationary covariance functions for gaussian process regression. *Advances in Neural Information Processing Systems* 16. 2004; 16:273–280.
- Perrin O, Meiring W. Nonstationarity in $\mathbb{R}^n$ is second-order stationary in $\mathbb{R}^{2n}$. *Journal of Applied Probability*. 2003; 40:815–820.
- Perrin O, Schlather M. Can any multivariate gaussian vector be interpreted as a sample from a stationary random process? *Statistics & Probability Letters*. 2007; 77:881–884.
- Ren Q, Banerjee S. Hierarchical factor models for large spatially misaligned datasets: A low-rank predictive process approach. *Biometrics*. 2013; 69:19–30. [PubMed: 23379832]
- Risser MD, Calder C. Regression-based covariance functions for nonstationary modeling. *Environmetrics*. 2015; 26
- Sampson PD, Guttorp P. Nonparametric estimation of nonstationary spatial covariance. *Journal of the American Statistical Society*. 1992; 87:108–119.
- Schmidt AM, O'Hagan A. Bayesian inference for non-stationary spatial covariance structure via spatial deformations. *JRSSB*. 2003; 65:743–758.
- Sigrist F, R Künsch H, Stahel WA. A dynamic nonstationary spatio-temporal model for short term prediction of precipitation. *AOAS*. 2012; 6:1452–1477.
- Stein ML. Space-time covariance functions. *JASA*. 2005; 100:310–321.
- Stein ML. Limitations on low rank approximations for covariance matrices of spatial data. *Spatial Statistics*. 2014; 8:1–19.
- Stroud JR, Müller P, Sansó B. Dynamic models for spatio-temporal data. *Journal of the Royal Statistical Society: Series B*. 2001; 63:673–689.
- Wikle CK. A kernel-based spectral model for non-gaussian spatio-temporal processes. *Statistical Modeling: An International Journal*. 2002; 2:299–314.
- Wikle CK, Cressie N. A dimension-reduced approach to space-time kalman filtering. *Biometrika*. 1999; 86:815–829.

Figure 1. Empirical temporal and spatial covariance plots with (a)–(b): true dimensions; (c) – (d): reduced dimensions; and (e) – (f): the expanded dimensions using the estimated latent dimensions. The lines are the fitted exponential model using the maximum likelihood method.

Figure 2.

Upper: true latent dimensions in time (left panel) and space (right panel); Middle: estimated latent dimensions using 900 of observations. The solid points in the left (right) panel represent the sampled time (spatial) locations; Bottom: estimated latent dimensions using all observations.

(a) Stream station locations in Illinois and Wisconsin.

(b) Estimated latent temporal dimensions w (upper) and spatial dimensions z (lower).

Table 1

Comparison between nonstationary models with other models using separable space-time covariance function.

Model	Dimensions	$\phi_t = \phi_s = 0.3$		$\phi_t = \phi_s = 0.5$		$\phi_t = \phi_s = 0.8$	
		MSPE (SE)	LS (SE)	MSPE (SE)	LS (SE)	MSPE (SE)	LS (SE)
True	$(t^1, t^2), [s^1, s^2, s^3]$	0.043 (0.0006)	-0.676 (0.0097)	0.106 (0.0006)	0.261 (0.0097)	0.222 (0.0028)	1.046 (0.01)
NonStat	$(t^1, \mathbf{w}), [s^1, s^2, \mathbf{z}]$	0.065 (0.0013)	-0.072 (0.0178)	0.106 (0.0013)	0.771 (0.0178)	0.274 (0.0042)	1.364 (0.0149)
BSZ+ST	$(t^1, [s^1, s^2, \mathbf{z}])$	0.073 (0.0014)	0.176 (0.0168)	0.167 (0.0029)	1.018 (0.0167)	0.313 (0.0049)	1.629 (0.0151)
BSZ	$[s^1, s^2, \mathbf{z}]$	0.271 (0.0055)	1.472 (0.0188)	0.408 (0.0067)	1.921 (0.0193)	0.555 (0.009)	2.206 (0.0181)
Stationary	$(t^1, [s^1, s^2])$	0.111 (0.0019)	0.594 (0.0174)	0.235 (0.0019)	1.352 (0.0174)	0.409 (0.006)	1.913 (0.0145)

* LS represents log score. In the column of Dimensions, $\mathbf{z}$ is the estimated dimension for s^3 and $\mathbf{w}$ is the estimated dimension for t^2 .

Table 2

Comparison between nonstationary models with other models using nonseparable space-time covariance function.

β	Model	$a = 2, c = 2$		$a = 2, c = 1$		$a = 7, c = 2$		$a = 7, c = 1$	
		MSPE (SE)	LS (SE)	MSPE (SE)	LS (SE)	MSPE (SE)	LS (SE)	MSPE (SE)	LS (SE)
0.5	True	0.520 (0.006)	2.037 (0.012)	0.393 (0.005)	1.729 (0.013)	0.722 (0.009)	2.435 (0.011)	0.534 (0.007)	2.102 (0.012)
	Nonstat	0.570 (0.008)	2.204 (0.015)	0.493 (0.008)	2.045 (0.018)	0.772 (0.010)	2.552 (0.013)	0.626 (0.009)	2.317 (0.015)
	BSZ+ST	0.601 (0.008)	2.285 (0.013)	0.512 (0.008)	2.098 (0.017)	0.794 (0.010)	2.559 (0.013)	0.631 (0.009)	2.291 (0.014)
	Stationary	0.650 (0.008)	2.396 (0.013)	0.604 (0.008)	2.313 (0.014)	0.853 (0.010)	2.670 (0.012)	0.811 (0.008)	2.612 (0.011)
	True	0.533 (0.006)	2.069 (0.012)	0.418 (0.005)	1.800 (0.013)	0.730 (0.009)	2.450 (0.011)	0.547 (0.007)	2.133 (0.012)
	Nonstat	0.587 (0.008)	2.242 (0.014)	0.519 (0.009)	2.107 (0.018)	0.775 (0.009)	2.552 (0.012)	0.640 (0.004)	2.348 (0.014)
1	BSZ+ST	0.612 (0.008)	2.305 (0.013)	0.532 (0.008)	2.138 (0.016)	0.803 (0.010)	2.568 (0.012)	0.641 (0.009)	2.311 (0.014)
	Stationary	0.656 (0.008)	2.407 (0.013)	0.620 (0.009)	2.343 (0.015)	0.857 (0.010)	2.675 (0.012)	0.790 (0.010)	2.583 (0.013)

* The dimensions of each model and the notation of LS follow those in Table 1.

Table 3

Comparison between nonstationary models estimated with different amounts of data, stationary models, and true models. The nonstationary data are generated based on the separable model (2) with parameters $\sigma^2 = 1$, $\phi_t = \phi_s = 0.5$ and the expanded dimensions described in Section 3.

Model	MSPE (SE)	Log Score (SE)
Stationary	0.457 (0.007)	2.168 (0.020)
Nonstationary ₉₀₀	0.402 (0.010)	1.840 (0.026)
Nonstationary _{2,500}	0.344 (0.004)	1.649 (0.015)
Nonstationary _{all}	0.253 (0.004)	1.311 (0.019)
True	0.235 (0.003)	1.109 (0.013)

Nonstationary_{*n*} is the nonstationary model fitted with *n* selected observations.