Modeling Nonstationary Processes Through Dimension Expansion

Luke BORNN, Gavin SHADDICK, and James V. ZIDEK

In this article, we propose a novel approach to modeling nonstationary spatial fields. The proposed method works by expanding the geographic plane over which these processes evolve into higher-dimensional spaces, transforming and clarifying complex patterns in the physical plane. By combining aspects of multidimensional scaling, group lasso, and latent variable models, a dimensionally sparse projection is found in which the originally nonstationary field exhibits stationarity. Following a comparison with existing methods in a simulated environment, dimension expansion is studied on a classic test-bed dataset historically used to study nonstationary models. Following this, we explore the use of dimension expansion in modeling air pollution in the United Kingdom, a process known to be strongly influenced by rural/urban effects, amongst others, which gives rise to a nonstationary field.

KEY WORDS: Dimension expansion; Environmental processes; Nonstationary modeling; Spatial statistics.

1. INTRODUCTION

Recently, there has been great interest in using spatial statistical methods to model environmental processes, with the aim of both gaining an improved understanding of underlying processes and making predictions at locations where measurements of a process are not available. The majority of such methods make the assumption that the underlying process is stationary (Cressie 1993) which, for many environmental systems, may be untenable.

In this article, we focus on accurately explicating the nonstationary structure that often arises in measurements of atmospheric, agricultural, and other environmental systems. If these systems share one underlying theme it is complex spatial structures, being influenced by such features as topography, weather, and other environmental factors. For example, the air quality characteristics of cities are likely to be more similar than that of rural areas irrespective of their geographic proximity. Ideally we might model these effects directly; however, information on the underlying causes is often not routinely available. Hence, when modeling environmental systems there exists a need for a class of models that are more complex than those which rely on the assumption of stationarity.

In the field of atmospheric science, empirical orthogonal functions have been used to model a nonstationary process as the sum of a stationary isotropic process and a set of basis functions with random coefficients representing departures from nonstationarity (Nychka and Saltzman 1998; Nychka, Wikle, and Royle 2002). Current approaches to modeling nonstationary processes in the statistical literature broadly comprise those that (1) combine locally stationary processes to create an overall nonstationary process and (2) "image warping."

A number of approaches for handling nonstationarity assume that over small enough spatial domains, the effects of nonstationarity are negligible, and hence locally stationary models may be used. This concept is the basis of kernel approaches,

early examples of which can be found in Haas (1990a,b). The process–convolution approach (Higdon 1998; Higdon, Swall, and Kern 1999) relies on the notion that a wide range of stationary Gaussian processes may be expressed as a kernel convolved with a Gaussian white noise process, with the kernel being allowed to vary spatially to account for nonstationarity. The form of the kernel allows for a broad expression of potential covariance functions, with a Gaussian kernel corresponding to a Gaussian covariance function and other choices of kernel resulting in other correlation structures. Similarly, Fuentes (2001) suggested modeling the field as a weighted average of local stationary processes within a set of regions, an idea which was later extended to include a continuous convolution of stationary processes (Fuentes and Smith 2001). Various difficulties still remain in this class of models, including the lack of a complete and easily interpretable global model and the choice of local regions and details of the weight structures. An alternative approach proposed by Sampson and Guttorp (1992) is that of "image warping," the central idea of which is that a nonstationary process may be stationary in a deformed, or warped, version of geographic space. Multidimensional scaling (or related methods) can be used to find the deformed locations with a mapping between the original and deformed space found using, for example, a thin-plate spline.

The principal idea underlying the proposed method is that of embedding the original field in a space of higher dimension where it can be more straightforwardly described and modeled. Specifically, we shift the dimensionality of the problem from 2 or 3 dimensions to 4, 5, or more to recover stationarity in the process; we term our methodology "dimension expansion." Our starting point is that nonstationary systems may be represented as low-dimensional projections of high-dimensional stationary systems (see, e.g., Perrin and Meiring 2003). The method is superficially similar to that of image warping; however, it differs fundamentally in that here the locations in the geographic space are retained, with added flexibility obtained through the extra dimensions. In addition, it addresses one of the major issues with the image warping approach, namely folding of the space.

Luke Bornn is a PhD Candidate in the Department of Statistics at the University of British Columbia, 333-6356 Agricultural Road, Vancouver, BC, V6T 1Z2 Canada (E-mail: *l.bornn@stat.ubc.ca*). Gavin Shaddick is a Senior Lecturer in the Department of Mathematical Sciences at the University of Bath, Bath BA2 7AY (E-mail: *g.shaddick@bath.ac.uk*). James V. Zidek is a Professor in the Department of Statistics at the University of British Columbia, 333-6356 Agricultural Road, Vancouver, BC, V6T 1Z2 (E-mail: *jim@stat.ubc.ca*).

This occurs in image warping if the estimate of the function that transforms from geographical to deformed space is not injective. As a result of folding, two geographically distinct locations may be mapped to the same location, meaning the variation between them will be incorrectly treated as measurement error and small scale variation (i.e., the nugget), which is expressly appropriate only for collocated and other proximal monitoring sites. In such cases, mapping quantities such as prediction intervals becomes particularly challenging both in terms of implementation and interpretation.

The remainder of the article is organized as follows: Section 2 introduces the dimension expansion framework proposed here, including an illustrative example to demonstrate the fundamental concepts behind the approach. This example is then used to draw comparisons to image warping. In Section 3, dimension expansion is applied to two real-life examples. First, the solar radiation dataset originally used in Sampson and Guttorp (1992) and used as a test-bed in various more recent image warping articles is studied. Second, we study air pollution from 77 monitoring locations in the United Kingdom which show clear signs of nonstationarity. We highlight the ability of dimension expansion to accurately model such data as measured through cross-validated prediction error. Finally, Section 4 provides a discussion and suggestions for future developments.

2. DIMENSION EXPANSION

While an early work dealt primarily with stationary models (e.g. Cressie 1993), it is now generally recognized that many spatial processes $\{Y(\mathbf{x}) : \mathbf{x} \in \mathcal{S}\}$, $(\mathcal{S} \in \mathcal{R}^d)$ fail to satisfy this assumption. Environmental systems might exhibit behavior that looks locally stationary, yet when considered over large and heterogenous domains they very often exhibit nonstationarity. For ease of notation, we consider $Y(\mathbf{x})$ to be a (potentially nonstationary) mean-zero Gaussian process and place our emphasis on representing the nonstationary structure.

A principal task in spatial statistics is estimating a variogram model (or correlation function) to explain spatial dependencies. The theoretical variogram, defined as

$$2\gamma(\mathbf{x}_i, \mathbf{x}_j) = E(|Y(\mathbf{x}_i) - Y(\mathbf{x}_j)|^2),$$

is typically modeled using a parametric stationary variogram $\gamma_\phi(\mathbf{h})$ depending only on $\mathbf{h} = \mathbf{x}_i - \mathbf{x}_j$, the difference vector between locations, and the parameter(s) ϕ. If the field is nonstationary, such a model will be a misspecification. In response, we transform the set of observed spatial locations $\mathcal{S} \in \mathcal{R}^d$ into one of higher dimension $\mathcal{S}' \in \mathcal{R}^{d+p}$, where $p > 0$ and $\mathcal{S}$ is a subset of the dimensions of $\mathcal{S}'$. Put plainly, such an approach amounts to allowing extra dimensions for the observed locations $\mathbf{x}_1, \ldots, \mathbf{x}_s$, notated as $\mathbf{z}_1, \ldots, \mathbf{z}_s$ such that the field $Y([\mathbf{x}, \mathbf{z}])$ is stationary with a variogram model $\gamma_\phi([\mathbf{x}, \mathbf{z}_i] - [\mathbf{x}_j, \mathbf{z}_j])$. Here $[\mathbf{x}, \mathbf{z}]$ is the concatenation of the dimensions $\mathbf{x}$ and $\mathbf{z}$.

Perrin and Meiring (2003) explored this idea in the particular case where both the covariance function and the expansion from $\mathbf{x}$ to $[\mathbf{x}, \mathbf{z}]$ are known. In their motivating example, they consider the following stationary covariance on the plane:

$$\text{cov}(Y([\mathbf{x}_i, \mathbf{z}_i]), Y([\mathbf{x}_j, \mathbf{z}_j])) = \exp(-|\mathbf{x}_i - \mathbf{x}_j| - |\mathbf{z}_i - \mathbf{z}_j|).$$

By restricting to the set $\mathbf{z} = \mathbf{x}^2$ and defining $Y'(\mathbf{x}) = Y([\mathbf{x}, \mathbf{x}^2])$, the resulting covariance function on this reduced-dimension field is nonstationary, namely,

$$\text{cov}(Y'(\mathbf{x}_i), Y'(\mathbf{x}_j)) = \exp(-|\mathbf{x}_i - \mathbf{x}_j|) \exp(1 + |\mathbf{x}_i + \mathbf{x}_j|).$$

Perrin and Meiring (2003) then considered the reverse problem, proving that a nonstationary random field indexed by $\mathcal{R}^d$ (with moments of order greater than 2) can always be represented as second-order stationary in $\mathcal{R}^{2d}$. It is not, however, necessary to move from $\mathcal{R}^d$ to $\mathcal{R}^{2d}$ to obtain the existence of a stationary field. Consider a recent result of Perrin and Schlather (2007), which stated that a Gaussian random vector can always be interpreted as a realization of a stationary field in $\mathcal{R}^p$, $p \geq 2$, subject to moment constraints on the vector, namely, that all components have equal expectation with the covariance matrix having identical components on the diagonal. From this it is straightforward to state that, similarly, a realization of a Gaussian process in $\mathcal{R}^d$ may be interpreted as a realization of a stationary field in $\mathcal{R}^{d+p}$, $p \geq 2$ (similarly, subject to moment constraints), with the covariance function ignoring the initial d dimensions.

The above results show the existence of higher-dimensional stationary representations for nonstationary fields, yet in the vast majority of situations neither a nonstationary variogram, nor an analytic mapping to higher dimensions, is known. Here we construct a framework for using higher-dimensional representations to model nonstationary systems, with the goal of learning the latent dimensions nonparametrically from information contained within the data.

To learn the expanded, or latent, dimensions $\mathbf{z}_1, \ldots, \mathbf{z}_s$ we propose

$$\hat{\phi}, \mathbf{Z} = \underset{\phi, Z'}{\text{argmin}} \sum_{i<j} (v_{i,j}^* - \gamma_\phi(d_{i,j}([\mathbf{X}, \mathbf{Z}'])))^2, \qquad (1)$$

where v_{ij}^* estimates the spatial dispersion between sites i and j, for example,

$$v_{ij}^* = \frac{1}{|\tau|} \sum_\tau |Y(\mathbf{x}_i) - Y(\mathbf{x}_j)|^2,$$

with $\tau > 1$ indexing multiple observations of the system, the handling of which is considered in the discussion, and $d_{i,j}([\mathbf{X}, \mathbf{Z}])$ is the i, jth element of the distance matrix of the (augmented) locations $[\mathbf{X}, \mathbf{Z}]$. Once the matrix $\mathbf{Z} \in \mathcal{R}^s \times \mathcal{R}^p$ is found, a function f is built such that $f(\mathbf{X}) \approx \mathbf{Z}$. While a wide range of options exist, we follow Sampson and Guttorp (1992) in using thin-plate splines, here one for each dimension of $\mathbf{Z}$. Thin-plate splines are a mapping from one space to another such that the integral of the squared second-order derivatives of the mapping function is minimized. For $d = 2$, this corresponds to minimizing

$$\sum_{i=1}^s ||\mathbf{z}_i - f(\mathbf{x}_i)||^2 + \lambda_2 \int_{\mathcal{R}^2} \left[\left(\frac{\partial^2 f}{\partial^2 x_1}\right)^2 + 2\left(\frac{\partial^2 f}{\partial x_1 \partial x_2}\right)^2 \right. $$
$$\left. + \left(\frac{\partial^2 f}{\partial^2 x_2}\right)^2 \right] dx_1 dx_2,$$

and, therefore, the smoothing parameter λ_2 is analogous to λ_{IW}, the thin-plate spline parameter in the image warping framework. Setting $\lambda_2 = 0$ results in an interpolating spline, whereas
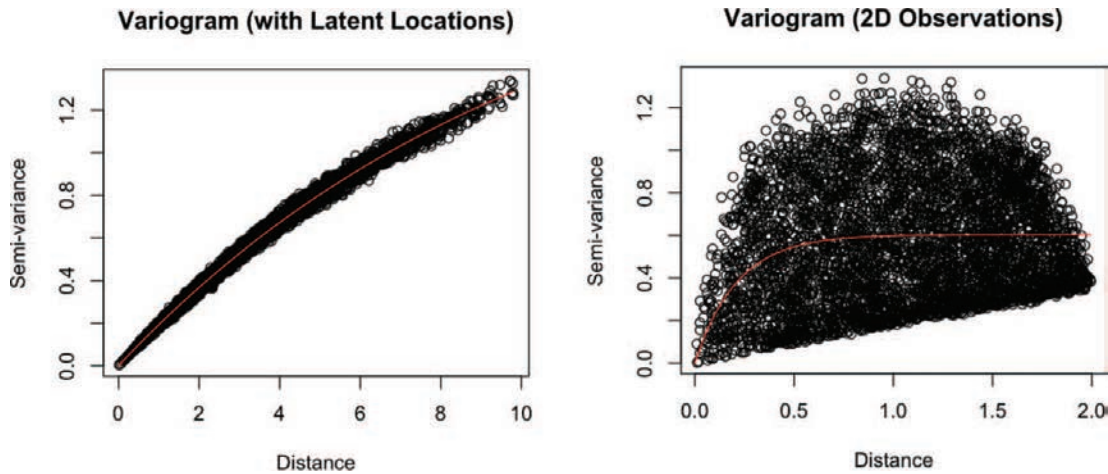
Figure 1. Empirical variograms from the original process (left) as well as a two-dimensional projection (right) on the illustrative ellipsoid example. A fitted exponential variogram is shown by the solid line. (The online version of this figure is in color.)

$\lambda_2 \to \infty$ results in the linear least squares fit. The nonlinear functions f are therefore linear combinations of basis functions centered at the locations $\mathcal{S} \in \mathcal{R}^d$. Once a model is built in the expanded space, f^{-1} will bring us from the manifold in $\mathcal{R}^{d+p}$ defined by $(\mathbf{X}, f(\mathbf{X})), \mathbf{X} \in \mathcal{R}^d$ back to the original space.

Due to our unique formulation, we have $f^{-1}(\mathbf{Z}) = \mathbf{X}$, and we need not be concerned with the difficulties associated with ensuring that f is bijective as in earlier approaches. Thus, we may view the originally observed locations $\mathbf{X}$ as a projection from a manifold within a higher-dimensional space, $[\mathbf{X}, \mathbf{Z}]$, in which the process is stationary. As an obvious (and direct) example, a process which is stationary given both geographical location and elevation may result in a nonstationary field given only longitude and latitude. Learning the latent dimensions (whether or not they have a physical meaning, such as elevation) means that a stationary model may be used in the expanded space.

In many situations, it is unclear how many additional dimensions are needed to accurately model the spatial field. One could use cross-validation or a model selection technique to determine the dimensionality of $\mathbf{Z}$; however, recognizing that (1) might result in overfitting the spatial dispersions, we would also like to regularize the estimation of $\mathbf{Z}$. As a result, we modify (1) by including a group lasso penalty term on $\mathbf{Z}$, where the groups are the dimensions of $\mathbf{Z}$ (Yuan and Lin 2006). The resulting objective function is

$$\hat{\phi}, Z = \underset{\phi, Z'}{\operatorname{argmin}} \sum_{i<j} (v_{i,j}^* - \gamma_\phi(d_{i,j}([\mathbf{X}, \mathbf{Z}'])))^2$$
$$+ \lambda_1 \sum_{k=1}^p ||\mathbf{Z}'_{\cdot,k}||_1, \qquad (2)$$

where $\mathbf{Z}'_{\cdot,k}$ is the kth column (dimension) of $\mathbf{Z}'$. As a consequence of this revised objection function, one need only determine a maximum number of dimensions p and the shrinkage parameter λ_1, whereupon the learned augmented dimensions $\mathbf{Z}$ will be both regularized toward zero and sparse in dimension. Hence λ_1 can be viewed as regularizing the estimation of $\mathbf{Z}$ and determining the dimension of the problem, whereas λ_2 controls the smoothness of the augmented dimensions; we

suggest learning both through cross-validation, although other model fit diagnostics or prior information may be used as well.

It is relatively straightforward to solve (2) using the gradient projection method of Kim, Kim, and Kim (2006), which conducts blockwise updates for group lasso with general loss functions. Here the blocks are the dimensions of $\mathbf{Z}$, and hence the optimization is efficient even for a large number of spatial locations. Optimization details are given in the appendix. For ease of exposition we use an exponential variogram,

$$\gamma_\phi(\mathbf{x}_1, \mathbf{x}_2) = \phi_1(1 - \exp(-||\mathbf{x}_1 - \mathbf{x}_2||/\phi_2)) + \phi_3,$$

which works well in the examples that follow, although the method applies analogously to other variograms.

2.1 Illustrative Example

We now present an illustrative example to help explain the concepts behind this proposed dimension expansion approach, as well as demonstrate the inability of image warping to handle complex nonstationarity. Specifically, we simulate a Gaussian process with $s = 100$ locations on a three-dimensional half-ellipsoid centered at $(0, 0, 0)$ such that the projection to the first two dimensions is a disk centered at the origin. Here, as throughout, distances are Euclidean. Figure 1 plots the empirical variograms for the original three-dimensional space as well as the two-dimensional projection, the latter of which results in a highly noisy empirical variogram cloud. Our goal is to recover the lost dimension through dimension expansion by optimizing (2) with $\lambda_1 = 0.5$, chosen to induce $\mathbf{Z}$ to have one dimension. Here, we calculate the matrix of empirical dispersions v_{ij}^* using 1000 realizations of the Gaussian process. Figure 2 shows the resulting learned locations as well as the corresponding empirical variogram, where we see that dimension expansion is capable of recovering the lost dimension, resulting in a variogram that closely reproduces the original.

2.2 Image Warping and Folding

In the image warping approach, Sampson and Guttorp (1992) employ nonmetric multidimensional scaling to move the locations along the geographic space, followed by fitting of the
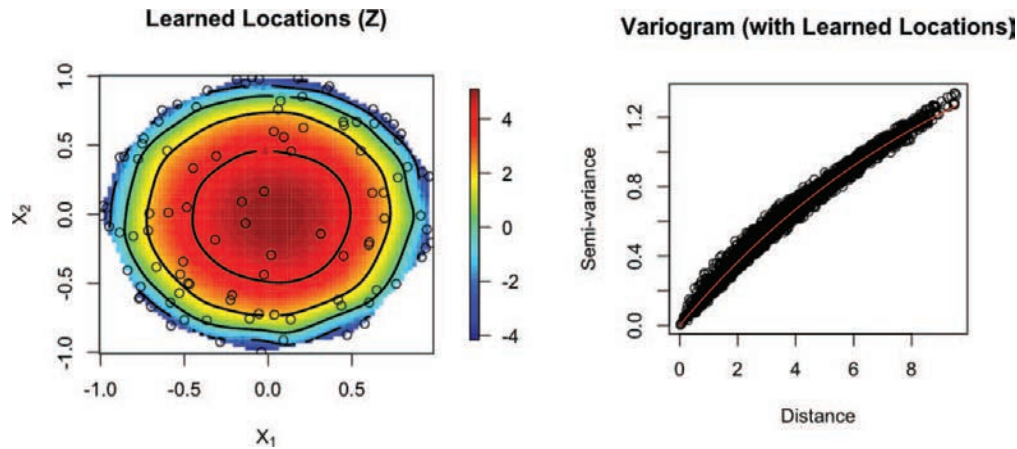
Figure 2. Learned latent locations (left) using $\lambda_2 = 10^{-4}$ as well as corresponding empirical variogram (right) after dimension expansion is applied. The fitted exponential variogram is shown by the solid line. (The online version of this figure is in color.)

variogram γ_ϕ using traditional variogram fitting methods. From this, a function f is found to go from the original to the warped locations, and back via f^{-1}. A number of adaptations of this approach have been proposed. Smith (1996) proposed modeling the covariance function as a linear combination of radial basis functions using maximum likelihood as suggested by Mardia and Goodall (1993). Monestiez and Switzer (1991) and Monestiez, Sampson, and Guttorp (1993) noted that the multistage algorithm of Sampson and Guttorp does not correspond to a unified optimization problem and instead propose finding the locations and fitting the variogram using a single-objective function, an approach also pursued by Meiring et al. (1997). It is worth noting that Monestiez and Switzer (1991) also explored mappings from $\mathcal{R}^2$ to $\mathcal{R}^3$ in the context of analyzing acid rain data, as the same-dimension mapping was incapable of describing the nonstationarity arising in the observed field. In a similar vein, Iovleff and Perrin (2004) proposed using simulated annealing to fit the spatial deformation model. Rather than imposing smoothness on the deformation through thin-plate splines, they used Delauney triangulation to constrain the mapping f from folding on itself. To acknowledge the uncertainty associated with the deformed locations, Damian, Sampson, and Guttorp (2001), Schmidt and O'Hagan (2003), and recently Schmidt, Guttorp, and O'Hagan (2011) have proposed Bayesian implementations of this approach, the latter additionally using observed covariate information to warp into higher dimensions.

As described in the introduction, the image warping framework can suffer from problems of folding, namely, of f not being bijective (see Zidek, Sun, and Le (2000) for a particularly extreme example of folding). Considering the illustrative example of Section 2.1, admittedly designed to be illustrative of such folding, we apply the image warping technique (Sampson and Guttrop 1992) with f modeled as a thin-plate spline. Because the image warping framework contains no term similar to λ_1 to regularize the warped locations, smoothing must be done through the thin-plate spline parameter λ_{IW} (analogous to λ_2 in the proposed dimension expansion framework). Figure 3 shows the warped grids and resulting empirical variograms for various settings of λ_{IW} applied to the ellipsoid example introduced in Section 2.1. We observe immediately that for a highly penalized f (corresponding to large λ_{IW}) the space does not fold; however,

the variogram fit is very poor. As λ_{IW} is relaxed to improve the fit, the space begins to fold, highlighting a potentially serious problem with the image warping framework—an issue which is addressed in the dimension expansion paradigm proposed here.

Also related to the proposed dimension expansion method are latent space models such as that proposed by Hoff, Raftery, and Handcock (2002). Here, latent dimensions are used to help learn a network of relationships between individuals. Recent work in the field of spatial statistics has also exploited latent dimensions to ensure valid cross-covariance functions in multivariate fields. Specifically, Apanasovich and Genton (2010) used latent dimensions for the different variables to build a class of valid cross-covariance functions.

3. APPLICATIONS

We now present two applications of dimension expansion applied to the modeling of nonstationary processes using two real datasets. The first uses the solar radiation data (Hay 1984) studied in the original image warping article of Sampson and Guttorp (1992). The second consists of measurements from a network of air pollution (black smoke) monitoring sites in the UK, further details of which can be found in Elliott et al. (2007).

3.1 Solar Radiation

The data of Hay (1984) includes measurements of solar radiation from 12 stations in the area surrounding Vancouver. Due to the location and elevation of station 1 (Grouse mountain), the field is inherently nonstationary, as exhibited by the sample variogram (Figure 4). Figure 4 shows the original and warped locations using Sampson and Guttorp's image warping approach with corresponding variogram plot. Image warping moves the locations (in particular the station at Grouse mountain, which is the northernmost location) to achieve something closer to stationarity. It is worth noting that overfitting may be controlled through the parameter λ_{IW} of the thin-plate spline.

Figures 4 and 5 show the results of applying the dimension expansion approach using $\lambda_1 = 0.5$ and $\lambda_1 = 0.2$, respectively, using a maximum number of dimensions of $p = 5$. The original locations are shown, as well as the added dimensions ($\mathbf{Z}$). With $\lambda_1 = 0.5$ (Figure 4, right), dimension expansion adds
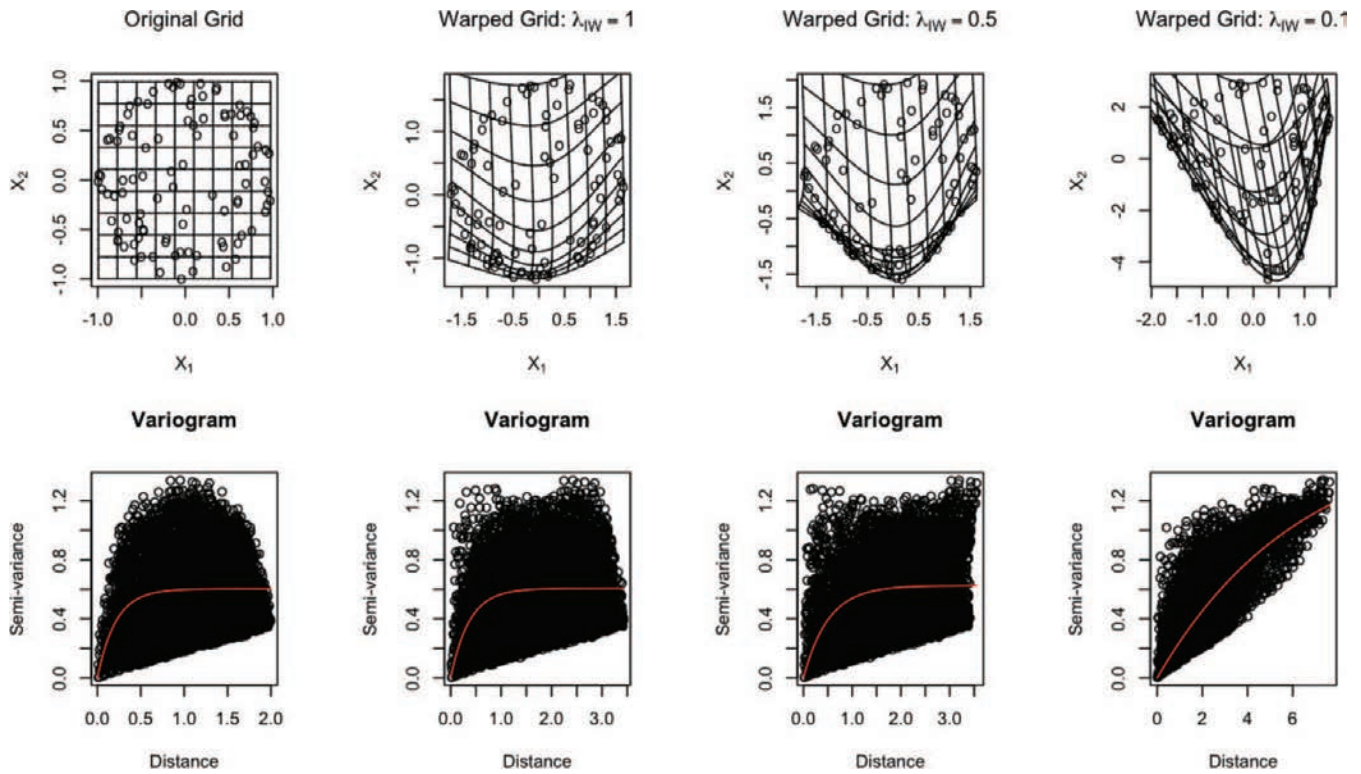
Figure 3. Warped grid of locations (top) and corresponding variograms (bottom) for various settings of the thin-plate spline parameter λ_{IW} using the image warping technique of Sampson and Guttorp (1992). (The online version of this figure is in color.)
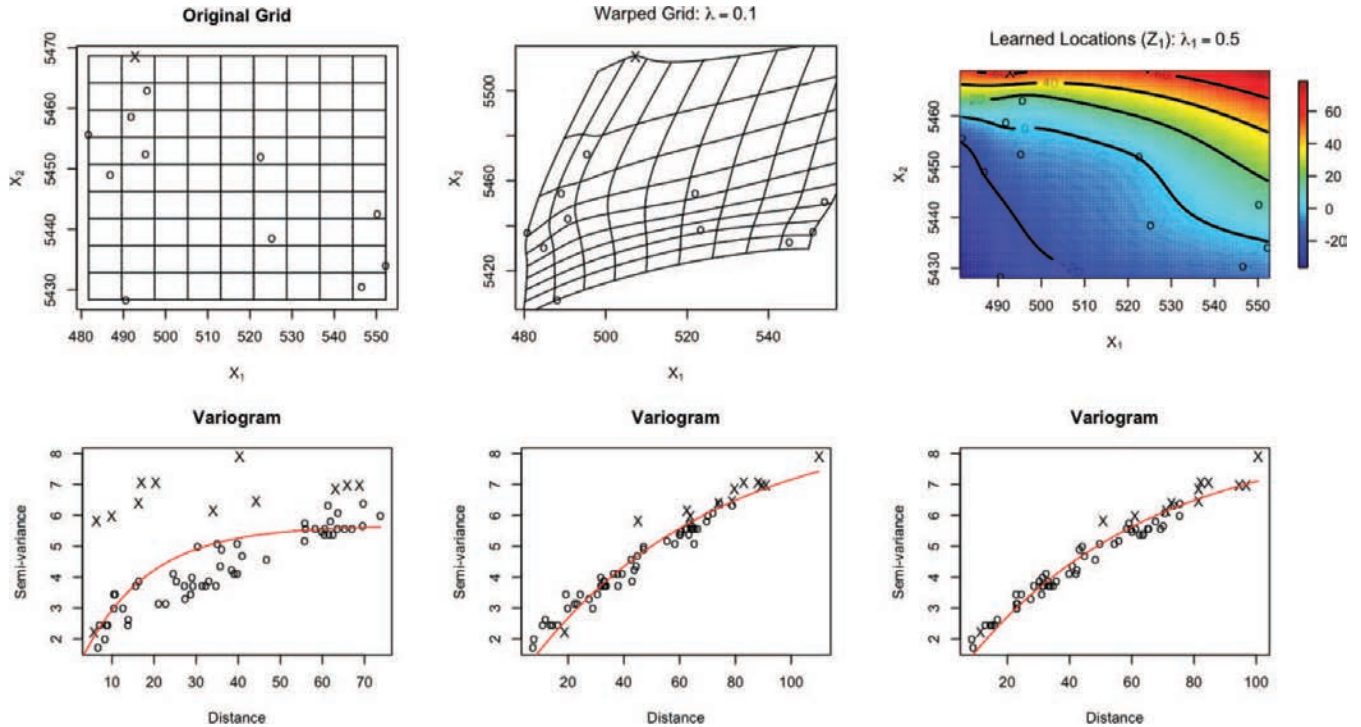

Figure 4. Original locations and empirical variogram for the solar radiation data (left); warped locations and associated empirical variogram using image warping with $\lambda_{IW} = 0.1$ (center); learned locations with associated empirical variogram using dimension expansion with $\lambda_1 = 0.5, \lambda_2 = 10^{-4}$ (right). The unit for the semivariance is $(\text{MJm}^{-2}\text{day}^{-1})^2$, and for distance is kilometer (km) (UTM coordinates, divided by 1000). The fitted variogram is shown by a solid line, and points associated with Grouse mountain (station 1) are highlighted using an "X." (The online version of this figure is in color.)
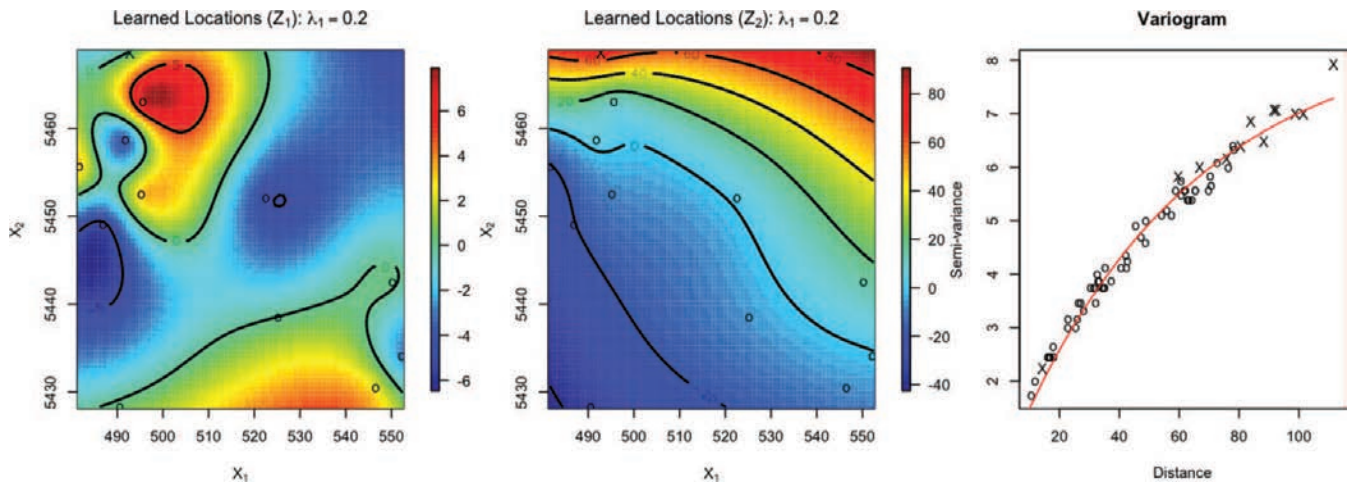
Figure 5. Dimension expansion of the solar radiation surface using $\lambda_1 = 0.2$, $\lambda_2 = 10^{-4}$. Z here is five dimensions, with Z_3, Z_4, and Z_5 set to zero as a result of the sparsity-inducing penalization. The first two panels show the learned locations, and the rightmost panel shows the associated empirical variogram (fitted variogram shown in red). The unit for the semivariance is $(MJm^{-2}day^{-1})^2$, and for distance is kilometer km. Points associated with Grouse mountain (station 1) highlighted using an "X." (The online version of this figure is in color.)

one additional dimension which primarily serves to push Grouse mountain out of the plane, reflecting the a priori suggestion that it is elevation that leads to the station's spurious correlation pattern. Interestingly, the contours of the learned dimension closely resemble the elevation contours of the mountains surrounding the Vancouver area. With $\lambda_1 = 0.2$ (Figure 5), two extra dimensions are used, and the fit of the parametric variogram improves marginally. We can, therefore, see how λ_1 influences the number of extra dimensions used, as well as the shrinkage in each dimension, to control the level of fit.

3.2 Air Pollution

The data consists of annual average concentrations of black smoke (μgm^{-3}) over a period of 16 years from 77 locations within the UK operating between April 1978 and March 1993 (inclusive) and was obtained from the Great Britain air quality archive (*www.airquality.co.uk*). Sites were selected in areas defined wholly or partially residential and measurements were aggregated to ward level (based on the 1991 census) using a geographical information system (Elliott et al. 2007). The majority of wards contained a single site, but where there were more than one, records were either joined together if the time periods did not overlap or averaged if time periods of operation were simultaneous. Due to similarities in levels of air pollution in urban locations, even if they are not geographically close, the field is known to be nonstationary. Specifically, we see in Figure 8 reduced empirical dispersions for distances around 280–290 km (the distance between London and Liverpool/Manchester), indicating that these urban centers are more similar than their distances would suggest. Our goal is to uncover and explore this nonstationarity through the dimension expansion framework.

We begin with cross-validation to learn the optimal setting of the parameters λ_1, λ_2 using (2) as described in Section 2. Figure 6 shows the resulting cross-validation RMSE for various parameter settings. We can see that moderate values of both λ_1 and λ_2 result in the best prediction performance. As λ_1 increases to its highest value ($10^{4.5}$), no dimension expansion

occurs, and hence λ_2 has no impact. From this, it is straightforward to see that the use of the original geographic space is a special case of the dimension expansion framework.

Using these parameter values, the dimensionally sparse optimization (2) used by dimension expansion leaves all but one dimension of **Z** set to zero. This dimension is shown in Figure 7, where we see a strong ridge connecting London, Birmingham, Liverpool, and Manchester. Hence, in the extra dimension, major cities are moved closer together while rural areas are pushed further away. The variograms, before and after the dimension
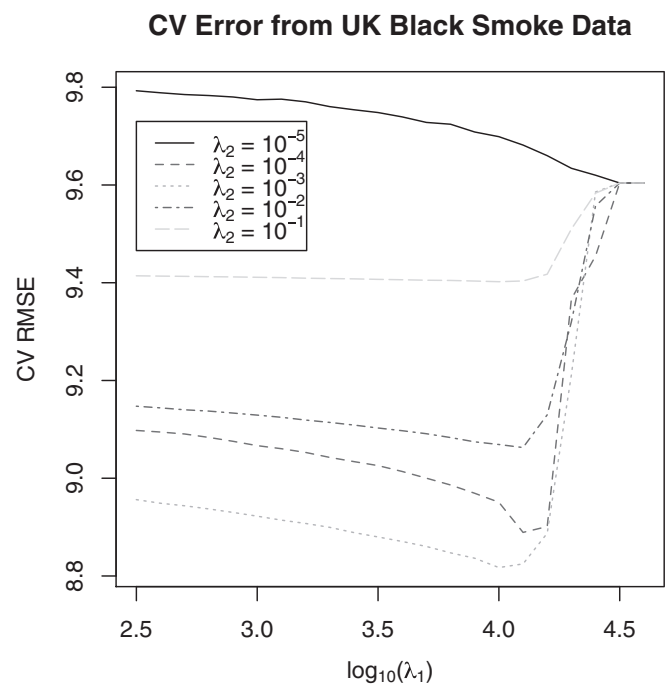

Figure 6. Leave-10-out cross-validated prediction error of dimension warping applied to the UK black smoke data. Here, we see optimal prediction for moderate values of both λ_1 and λ_2.
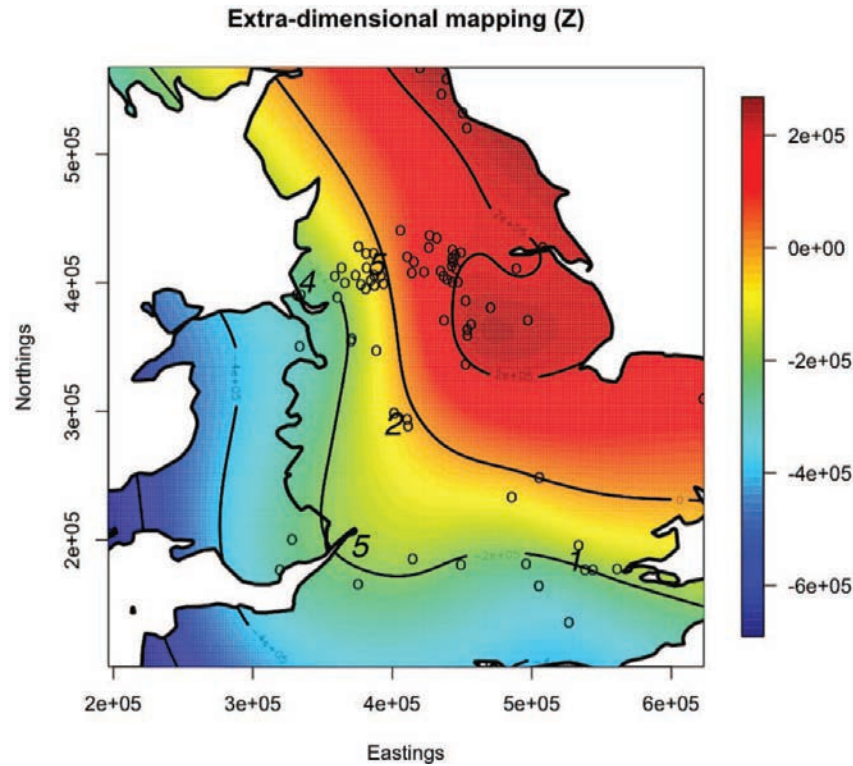
Extra-dimensional mapping (Z)

Figure 7. Coordinate surface of the learned dimension following dimension expansion. A strong ridge is visible connecting major cities indicating closer correlation between these locations than would be suggested in geographic space. The locations of a selection of major cities are shown: (1) London, (2) Birmingham, (3) Manchester, (4) Liverpool, and (5) Bristol. (The online version of this figure is in color.)

expansion, are shown in Figure 8, where we see no indications of nonstationarity after dimension expansion is applied.

4. DISCUSSION

By augmenting the dimensionality of the underlying geographic space, we have developed a highly flexible approach for handling the nonstationarity that often arises in environmental systems. While ostensibly similar to image warping, the proposed method avoids the issue of folding and allows one to model much more complex nonstationarity patterns through interdimensional expansions, allowing the user to perform nonparametric learning of the mapping function. In addition, through the use of a group lasso penalty, we are able to estimate the number of augmented dimensions, as well as regularize the optimization problem. Lastly, we have highlighted the practical application of the dimension expansion approach in three examples, two of which use data from observed environmental processes. It is worth noting that while we have developed the spatial model in terms of variograms, it could alternatively be
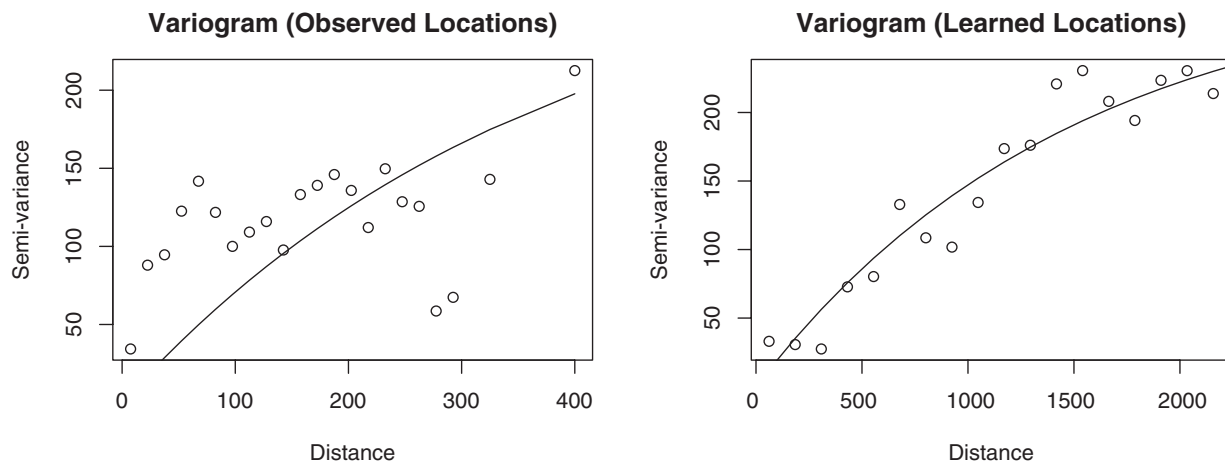

Figure 8. Binned empirical and fitted (solid line) variograms on the UK black smoke data, following dimension expansion. In the original geographic space, we see a dip in the empirical variogram at roughly 280 km, corresponding to the distance between London and Liverpool/Manchester. After dimension expansion is applied, the ridge between London and Liverpool/Manchester removes this effect of nonstationarity. The unit for the semivariance is $(\mu\text{gm}^{-3})^2$ and for distance is km.

expressed in terms of covariances; see, for example, Gneiting, Sasvári, and Schlather (2001) for a thorough comparison of the two approaches.

In general, models will comprise a spatial mean or "trend" term together with spatial covariance for deviations from this trend. It is desirable to maximally reflect the variation in the response using the mean function and thus known covariates, but inevitably the mean function will not be able to capture all of the spatial variation and thus residual spatial variation must be modeled specifically. When all relevant covariates are included in the mean term, it is commonly assumed that the resulting spatial term is stationary. However, as the Karhunen–Loeve expansion shows, the modeling of spatial trend and covariance are inseparable and misspecification of the former will induce a second-order distortion in the latter, thus violating any assumptions of stationarity in many cases. Due to the complexity of environmental processes, misspecification is inevitable because all relevant covariates can never be known or, even if known, observed. In the air pollution example presented here, concentrations in cities appeared to be more similar than that of rural areas irrespective of their geographic proximity. If available, it would be possible to incorporate a measure of rurality in the mean function, possibly produced using a geographical information system based on population density data. However, even if such information were available, stationarity would still not be guaranteed and so there is a need for methods such as that proposed here to allow nonstationary models to be constructed for the spatial process.

A Bayesian image warping approach which allows covariate effects to be included in the correlation structure has recently been suggested by Schmidt et al. (2011). By treating covariates as analogous to geographic coordinates, they warp the combined location-covariate space into a deformed space of the same dimension. To achieve computational efficiency, they consider a special case which restricts the form of the possible mapping function and assumes the spatial process to be a two-dimensional manifold with covariates treated as separate values at each location.

In practice, environmental data will often take the form of a number of measurements made over time at each location rather than true spatial replications per se. To try and isolate the purely spatial part of the process, the mean function may incorporate a temporal component, modeling underlying temporal patterns and allowing the possibility of time-varying covariates, or even space–time interactions. In the absence of such covariate information, it would be possible to consider the notion of time-varying nonstationarity structure. For instance, if one wants to study the changing impact of cities and industrial areas on air pollution levels, examining changes in stationarity over time would be a valuable way to understand these changes. The dimension expansion framework is also amenable to multivariate extensions. We are currently exploring a scenario whereby the dimension expansion functions and locations have a hierarchical structure, allowing the dimension expansion to vary for different variables, yet be tied together through the hierarchy.

We have demonstrated how the proposed approach can be used to perform predictions in the transformed, stationary space and mapped back to the original space. At present, the choice of the mapping, learning of latent locations, and prediction are performed in isolation. As the Sampson and Guttorp (1992)

approach was set within a Bayesian framework by Damian et al. (2001) and Schmidt and O'Hagan (2003), setting the current algorithmic approach within such an inferential framework would allow the inherent uncertainty to be accurately reflected in resulting inferences and this is the goal of current research.

A OPTIMIZATION OF EQUATION (2)

As with traditional multidimensional scaling, penalization functions of the form (1) do not have a unique maximum. However, the learned locations are unique up to rotation, scaling, and sign. The optimization problem (2) is more regularized, however, due to the presence of the l_1-norm. Specifically, not all rotations and scalings of the learned locations will have the same l_1-norm, and hence the resulting optimization is unique only up to sign and indices of zero/nonzero dimensions. For example, consider finding the location of a single latent location in two latent dimensions. The locations $(1, 0)$ and $(1/\sqrt{2}, 1\sqrt{2})$ fit the objection function (1) equally well, yet their l_1-norms are 1 and $\sqrt{2}$, respectively. Note that since the end goal is not to learn the dimensions, but rather to find an expanded space in which the process is stationary, the existence of multiple possible expansions is not important, so long as one of the projections in the equivalence class is found.

In our experience, traditional optimization procedures such as Nelder–Mead or the Broyden–Fletcher–Goldfarb–Shanno method (Nocedal and Wright 1999) work well for a moderate amount of locations ($s < 100$) and dimensions ($p < 3$). For larger problems, it may be necessary to use purpose-built optimization routines intended for generalized group lasso. Let $\Omega(\mathbf{U})$ be the first term in (2), where $\mathbf{U} = [\mathbf{X}, \mathbf{Z}]$. Then column k of the gradient matrix is

$$\nabla_k \Omega(\mathbf{U}) = \frac{2}{p} \mathbf{\Gamma} \circ \left(\mathbf{U}_{\cdot,k} \mathbf{1}_{p \times p} - \mathbf{1}_{p \times p} \mathbf{U}_{\cdot,k}^T \right) \mathbf{1}_{p \times 1},$$

where

$$\mathbf{\Gamma}_{i,j} = \left(\gamma_\phi(d_{ij}(\mathbf{U})) - v_{ij}^* \right) \frac{\partial \gamma_\phi}{\partial d_{ij}}.$$

Using this gradient information, the gradient projection algorithm of Kim et al. (2006) may be used to optimize (2). The algorithm proceeds as follows:

Initialize : $\mathbf{U}^0 = \mathbf{0}$, α : sufficiently small positive constant
for $t = 1, \ldots, T$ **do**
    Set $\mathbf{u} = \mathbf{U}^{t-1} - \alpha \nabla \Omega(\mathbf{U}^{t-1})$ and $\eta = \{1, \ldots, p\}$
    while $M_j > 0 \quad \forall j$ **do**
        For $j = 1, \ldots, p$

$$M_j = I(j \in \eta) \times \left(||\mathbf{u}_j|| + \frac{M - \sum_{j \in \eta} ||\mathbf{u}_j||}{|\eta|} \right)$$

        Set $\eta = \{j : M_j > 0\}$
    end
    Set $\mathbf{U}_{\cdot,j}^{t-1} = \mathbf{u}_j \frac{M_j}{||\mathbf{u}_j||}$ for $j = 1, \ldots, p$
end
Return $\mathbf{U}^T$

From this, one can alternate between optimizing the parameters of the variogram and the latent locations. Further algorithmic

details, such as the tuning of *M* and the setting of the algorithmic parameter α, can be found in Kim et al. (2006).

REFERENCES

Apanasovich, T., and Genton, M. (2010), "Cross-Covariance Functions for Multivariate Random Fields Based on Latent Dimensions," *Biometrika*, 97(1), 15–30. [284]

Cressie, N. (1993), *Statistics for Spatial Data*, New York: Wiley. [281,282]

Damian, D., Sampson, P., and Guttorp, P. (2001), "Bayesian Estimation of Semi-Parametric Non-Stationary Spatial Covariance Structures," *Environmetrics*, 12(2), 161–178. [283,288]

Elliott, P., Shaddick, G., Wakefield, J., Hoogh, C., and Briggs, D. (2007), "Long-Term Associations of Outdoor Air Pollution with Mortality in Great Britain," *Thorax*, 62(12), 1088–1094. [284,286]

Fuentes, M. (2001), "A High Frequency Kriging Approach for Non-Stationary Environmental Processes," *Environmetrics*, 12(5), 469–483. [281]

Fuentes, M., and Smith, R. (2001), A New Class of Nonstationary Spatial Models. Technical Report, Institute of Statistics Mimeo Series 2534, North Carolina State University, Raleigh, NC. [281]

Gneiting, T., Sasvári, Z., and Schlather, M. (2001), "Analogies and Correspondences Between Variograms and Covariance Functions," *Advances in Applied Probability*, 617–630. [287]

Haas, T. (1990a), "Kriging and Automated Variogram Modeling within a Moving Window," *Atmospheric Environment. Part A. General Topics*, 24(7), 1759–1769. [281]

——— (1990b), "Lognormal and Moving Window Methods of Estimating Acid Deposition," *Journal of the American Statistical Association*, 85(412), 950–963. [281]

Hay, J. (1984), "An Assessment of the Mesoscale Variability of Solar Radiation at the Earth's Surface," *Solar Energy*, 32(3), 425–434. [284]

Higdon, D. (1998), "A Process-Convolution Approach to Modelling Temperatures in the North Atlantic Ocean," *Environmental and Ecological Statistics*, 5(2), 173–190. [281]

Higdon, D., Swall, J., and Kern, J. (1999), "Non-Stationary Spatial Modeling," *Bayesian Statistics*, 6, 761–768. [281]

Hoff, P., Raftery, A., and Handcock, M. (2002), "Latent Space Approaches to Social Network Analysis," *Journal of the American Statistical Association*, 97(460), 1090–1098. [284]

Iovleff, S., and Perrin, O. (2004), "Estimating a Nonstationary Spatial Structure Using Simulated Annealing," *Journal of Computational and Graphical Statistics*, 13(1), 90–105. [283]

Kim, Y., Kim, J., and Kim, Y. (2006), "Blockwise Sparse Regression," *Statistica Sinica*, 16(2), 375. [283,288]

Mardia, K., and Goodall, C. (1993), "Spatial-Temporal Analysis of Multivariate Environmental Monitoring Data," *Multivariate Environmental Statistics*, 6, 347–385. [283]

Meiring, W., Monestiez, P., Sampson, P., and Guttorp, P. (1997), "Developments in the Modelling of Nonstationary Spatial Covariance Structure from Space-Time Monitoring Data," *Geostatistics Wollongong*, 96(1), 162–173. [283]

Monestiez, P., Sampson, P., and Guttorp, P. (1993), "Modelling of Heterogeneous Spatial Correlation Structure by Spatial Deformation," *Cahiers de Geostatistique*, 3, 1–12. [283]

Monestiez, P., and Switzer, P. (1991), Semiparametric Estimation of Nonstationary Spatial Covariance Models by Metric Multidimensional Scaling, Technical Report 165, Dept. of Statistics, Stanford University, available online at *http://statistics.stanford.edu/~ckirby/techreports/SIMS/SIMS%20165.pdf* [283]

Nocedal, J., and Wright, S. (1999), *Numerical Optimization*, Berlin: Springer Verlag. [288]

Nychka, D., and Saltzman, N. (1998), "Design of Air Quality Monitoring Networks," *Case Studies in Environmental Statistics*, 132, 51–76. [281]

Nychka, D., Wikle, C., and Royle, J. (2002), "Multiresolution Models for Nonstationary Spatial Sovariance Functions," *Statistical Modelling*, 2(4), 315. [281]

Perrin, O., and Meiring, W. (2003), "Nonstationarity in $\mathcal{R}^n$ is Second-Order Stationarity in $\mathcal{R}^{2n}$," *Journal of Applied Probability*, 40(3), 815–820. [281,282]

Perrin, O., and Schlather, M. (2007), "Can any Multivariate Gaussian Vector be Interpreted as a Sample from a Stationary Random Process?," *Statistics and Probability Letters*, 77(9), 881–884. [282]

Sampson, P., and Guttorp, P. (1992), "Nonparametric Estimation of Nonstationary Spatial Covariance Structure," *Journal of the American Statistical Association*, 87(417), 108–119. [281,282,283,284,288]

Schmidt, A., Guttorp, P., and O'Hagan, A. (2011), "Considering Covariates in the Covariance Structure of Spatial Processes," *Environmetrics*, 22(4), 487–500. [283,288]

Schmidt, A., and O'Hagan, A. (2003), "Bayesian Inference for Non-Stationary Spatial Covariance Structure via Spatial Deformations," *Journal of the Royal Statistical Society,* Series B, 65(3), 743–758. [283,288]

Smith, R. (1996), *Estimating Nonstationary Spatial Correlations*. Preprint, University of North Carolina. Available at *http://www.stat.unc.edu/postscripts/rs/nonstationary.ps*. [283]

Yuan, M., and Lin, Y. (2006), "Model Selection and Estimation in Regression with Grouped Variables," *Journal of the Royal Statistical Society,* Series B, 68(1), 49–67. [283]

Zidek, J., Sun, W., and Le, N. (2000), "Designing and Integrating Composite Networks for Monitoring Multivariate Gaussian Pollution Fields," *Journal of the Royal Statistical Society,* Series C, 49(1), 63–79. [284]