ENVIRONMETRICS
Environmetrics 2007; **18**: 681–695
Published online 8 May 2007 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/env.854

Separable approximations of space-time covariance matrices

Marc G. Genton[1,2*,†]

[1]*Department of Econometrics, University of Geneva, Bd du Pont-d'Arve 40, CH-1211 Geneva 4, Switzerland*
[2]*Department of Statistics, Texas A&M University, College Station, TX 77843-3143, USA*

SUMMARY

Statistical modeling of space-time data has often been based on separable covariance functions, that is, covariances that can be written as a product of a purely spatial covariance and a purely temporal covariance. The main reason is that the structure of separable covariances dramatically reduces the number of parameters in the covariance matrix and thus facilitates computational procedures for large space-time data sets. In this paper, we discuss separable approximations of nonseparable space-time covariance matrices. Specifically, we describe the nearest Kronecker product approximation, in the Frobenius norm, of a space-time covariance matrix. The algorithm is simple to implement and the solution preserves properties of the space-time covariance matrix, such as symmetry, positive definiteness, and other structures. The separable approximation allows for fast kriging of large space-time data sets. We present several illustrative examples based on an application to data of Irish wind speeds, showing that only small differences in prediction error arise while computational savings for large data sets can be obtained.

KEY WORDS: block Toeplitz matrix; kriging; Kronecker product; nonstationarity; positive definiteness; separable covariance; space-time stochastic process; stationarity

1. INTRODUCTION

Space-time stochastic processes have become a popular instrument in the statistician's toolkit for modeling observations from geophysical and environmental sciences. This work considers a space-time stochastic process $Z(\boldsymbol{s}, t)$, where $\boldsymbol{s} \in \mathbb{R}^d$, $d \geq 1$, and $t \in \mathbb{R}$. Usually, $d = 2$ for most practical applications. We assume that for all space-time coordinates in $\mathbb{R}^d \times \mathbb{R}$, the mean function is $\mathrm{E}(Z(\boldsymbol{s}, t)) = \mu(\boldsymbol{s}, t)$, the variance of Z is finite, and the nonstationary covariance of Z between the space-time coordinates $(\boldsymbol{s}, t)$ and $(\boldsymbol{s} + \boldsymbol{h}, t + u)$ exists for all $\boldsymbol{h} \in \mathbb{R}^d$ and $u \in \mathbb{R}$, and is given by

$$\mathrm{cov}(Z(\boldsymbol{s}, t), Z(\boldsymbol{s} + \boldsymbol{h}, t + u)) = C(\boldsymbol{s}, \boldsymbol{s} + \boldsymbol{h}, t, t + u) \tag{1}$$

The purely spatial and purely temporal nonstationary covariances of Z are given respectively by $C(\boldsymbol{s}, \boldsymbol{s} + \boldsymbol{h}, t, t)$ and $C(\boldsymbol{s}, \boldsymbol{s}, t, t + u)$.

*Correspondence to: M. G. Genton, Department of Econometrics, University of Geneva, Bd du Pont-d'Arve 40, CH-1211 Geneva 4, Switzerland.
†E-mail: Marc.Genton@metri.unige.ch; genton@stat.tamu.edu

Received 25 November 2006
Accepted 3 April 2007

We assume that Z is observed at $N = nm$ space-time coordinates $(\boldsymbol{s}_k, t_k)$, $k = 1, \ldots, N$, where n denotes the number of different spatial locations and m denotes the number of different temporal locations. The spatial locations could be irregularly or regularly (e.g., on a grid) placed, but the temporal locations are usually equidistant in time (e.g., hourly records). Typical data sets from geophysical and environmental sciences tend to be rich in time (i.e., long time series of, for example, pollutant concentrations, temperatures, wind speed) and rich or poor in space (i.e., many or few monitoring stations). Consequently, the number N of space-time coordinates is often very large and traditional statistical methods need to be scaled up. For instance, it is commonly of interest in a space-time analysis to perform optimal prediction of Z at an unobserved location $(\boldsymbol{s}_0, t_0)$, and actually at many such new locations in order to derive maps with associated prediction accuracy. The linear combination of the observations that minimizes the mean squared prediction error (MSPE) is called the simple kriging predictor of $Z(\boldsymbol{s}_0, t_0)$, see, for example, Cressie (1993), and is given by

$$\hat{Z}(\boldsymbol{s}_0, t_0) = \mu(\boldsymbol{s}_0, t_0) + \sum_{k=1}^{N} \lambda_k (Z(\boldsymbol{s}_k, t_k) - \mu(\boldsymbol{s}_k, t_k)) \tag{2}$$

The optimal weights $\lambda_1, \ldots, \lambda_N$ are obtained by solving an $N \times N$ linear system based on the space-time covariance (1) of the process Z. Evidently, when N is large, this is a computationally difficult task. This paper is about addressing this issue.

One approach recently proposed by Furrer *et al.* (2006) consists in tapering the covariance for prediction with large data sets. Their ingenious idea is to taper the covariance to zero beyond a certain range with an appropriate compactly supported covariance function (Gneiting, 2002a). The result is a sparse linear system that approximates the kriging linear system, and therefore can be solved very efficiently. The procedure is supported by asymptotic theory developed by Stein (1988, 1990) on the effect of misspecifying the covariance function. In the case of tapering, the misspecification is deliberate, and the asymptotic mean squared error resulting from kriging with the tapered covariance was shown, under specific conditions, to converge to the optimal error. Furrer *et al.* (2006) implemented their method on a large data set of monthly precipitation in the U.S. recorded at $n = 5909$ spatial locations. Temporally, the record spreads over $m = 1200$ months, and therefore the full size of this climatological space-time data set is $N = 7\,090\,800$. Even in the simple case of just the spatial kriging of the precipitation field on a fine grid of size 1000×1000, Furrer *et al.* (2006) reported computational gains in time of a factor over 560 for solving the linear system and over 110 for creating a whole kriged map.

An alternative, but possibly complementary, approach consists in the computation of a separable approximation of the nonseparable space-time covariance matrix. The main goal of this paper is to describe such a methodology. The nonstationary covariance (1) is space-time separable if it can be written as

$$C(\boldsymbol{s}, \boldsymbol{s} + \boldsymbol{h}, t, t + u) = C_S(\boldsymbol{s}, \boldsymbol{s} + \boldsymbol{h}) C_T(t, t + u) \tag{3}$$

where C_S is a purely spatial and C_T a purely temporal nonstationary covariance. In particular, a stationary covariance C is space-time separable if it can be written as

$$C(\boldsymbol{h}, u) = C_S(\boldsymbol{h}) C_T(u) \tag{4}$$

If the process Z itself is separable, that is $Z(\boldsymbol{s}, t) = Z_S(\boldsymbol{s}) Z_T(t)$, where Z_S is a purely spatial stochastic process with nonstationary covariance C_S and Z_T is a purely temporal stochastic process with nonstationary covariance C_T, and Z_S is independent of Z_T, then the property (3) holds. Indeed, we

DOI: 10.1002/env

have

$$\begin{aligned} C(\boldsymbol{s}, \boldsymbol{s}+\boldsymbol{h}, t, t+u) &= \mathrm{cov}(Z(\boldsymbol{s}, t), Z(\boldsymbol{s}+\boldsymbol{h}, t+u)) \\ &= \mathrm{cov}(Z_S(\boldsymbol{s})Z_T(t), Z_S(\boldsymbol{s}+\boldsymbol{h})Z_T(t+u)) \\ &= \mathrm{cov}(Z_S(\boldsymbol{s}), Z_S(\boldsymbol{s}+\boldsymbol{h}))\mathrm{cov}(Z_T(t), Z_T(t+u)) \\ &= C_S(\boldsymbol{s}, \boldsymbol{s}+\boldsymbol{h})C_T(t, t+u) \end{aligned}$$

However, the reverse is not true and is often a source of confusion in the literature. A separable covariance function C does not imply that the process Z is separable. Note also that if Z_S and Z_T are Gaussian processes, then their product is not.

A well-known shortcoming of separable covariance functions is that they do not allow for space-time interactions in the covariance, although deterministic space-time interactions in the process Z can be modeled through the mean function $\mu(\boldsymbol{s}, t)$; see Kyriakidis and Journel (1999), Cressie and Huang (1999), and Stein (2005) for detailed discussions on this topic. Nevertheless, statistical space-time modeling has been using separable covariance functions of the form (3) or (4) to a large extent. The main reason is that the structure of separable covariances dramatically reduces the number of parameters in the covariance matrix and thus facilitates computational procedures for large space-time data sets. However, another reason is that the covariance function (1) has to be positive definite due to the famed theorem of Bochner (1955). Valid (positive definite) parametric models are well-known in space and in time, but less so in space-time. The separable space-time structure in the covariances (3) or (4) allows for a simple construction of valid space-time parametric models. Recent efforts have been allocated to develop new parametric classes of nonseparable space-time covariance models; see, for example, Cressie and Huang (1999), Gneiting (2002b), Ma (2003), and Stein (2005), among others. Gneiting *et al*. (2007) present a survey of geostatistical space-time models. Simultaneously, a growing interest in developing formal testing procedures for separability has emerged. Most tests available in the literature are basically testing whether a parameter describing the separability in a parametric space-time covariance model is zero or not; see, for example, Shitan and Brockwell (1995), Guo and Billard (1998), Brown *et al*. (2000), and Genton and Koul (2007). Mitchell *et al*. (2005, 2006) proposed a likelihood ratio test for separability in the setting of multivariate repeated measures and suggested its application to space-time data sets that are rich in the time dimension. Matsuda and Yajima (2004), Scaccia and Martin (2005), and Fuentes (2006) suggested a test for separability of covariances based on spectral methods. Recently, Li *et al*. (2007) proposed a nonparametric test of various properties of space-time covariance functions, including separability. In this paper, however, we do not consider any test for separability. Instead, we suggest to use separable approximations of possibly nonseparable covariances.

Let $\Sigma \in \mathbb{R}^{N\times N}$ be the space-time covariance matrix of Z, that is $\Sigma_{ij} = C(\boldsymbol{s}_i, \boldsymbol{s}_j, t_i, t_j)$. If the space-time covariance C is separable, then there exist two covariance matrices, $S = (s_{ij}) \in \mathbb{R}^{n\times n}$ in space and $T = (t_{ij}) \in \mathbb{R}^{m\times m}$ in time, such that

$$\Sigma = S \otimes T = \begin{pmatrix} s_{11}T & s_{12}T & \dots & s_{1n}T \\ s_{21}T & s_{22}T & \dots & s_{2n}T \\ \vdots & \vdots & \ddots & \vdots \\ s_{n1}T & s_{n2}T & \dots & s_{nn}T \end{pmatrix} \tag{5}$$

DOI: 10.1002/env

where $s_{ij} = C_S(\boldsymbol{s}_i, \boldsymbol{s}_j)$, $t_{ij} = C_T(t_i, t_j)$, and $\otimes$ denotes the Kronecker product between two matrices; see, for example, Steeb (1997). The covariance matrices S and T are not unique since $S \otimes T = (aS) \otimes (\frac{1}{a}T)$ for any $a > 0$. This non-identifiability problem can be addressed by imposing a constraint, for example that the $(1, 1)$ entry of S or T be a 1. If the space-time covariance C is separable, then, given $\Sigma \in \mathbb{R}^{N \times N}$, an interesting question is how to determine the two matrices $S \in \mathbb{R}^{n \times n}$ and $T \in \mathbb{R}^{m \times m}$ satisfying (5). Furthermore, if the space-time covariance C is not separable, then we can ask to find the two matrices $S \in \mathbb{R}^{n \times n}$ and $T \in \mathbb{R}^{m \times m}$ that provide the 'best' separable (Kronecker product) approximation of $\Sigma \in \mathbb{R}^{N \times N}$. The solution to both questions has been presented from a linear algebra point of view by Van Loan and Pitsianis (1992) for a rectangular matrix Σ. They coined this question the nearest Kronecker product (NKP) problem, which we state here in the framework of a space-time covariance matrix. Note that we do not assume that the process Z is stationary. We denote the Frobenius norm of a matrix $\Sigma = (\sigma_{ij})$ by $\|\Sigma\|_F = (\sum_{i=1}^N \sum_{j=1}^N \sigma_{ij}^2)^{1/2}$.

Nearest Kronecker product for a space-time covariance matrix (NKPST) problem

Let $\Sigma \in \mathbb{R}^{N \times N}$ be a space-time covariance matrix with $N = nm$. Find two matrices $S \in \mathbb{R}^{n \times n}$ and $T \in \mathbb{R}^{m \times m}$ minimizing the Frobenius norm

$$\|\Sigma - S \otimes T\|_F \tag{6}$$

The solution to the NKPST problem is given by the singular value decomposition (SVD) of a permuted version of the matrix Σ and is presented in detail in Section 2. The remainder of the paper is set up as follows. In Section 2, we further define an index that quantifies the error of the separable approximation in the solution to the NKPST problem. We discuss properties of the space-time covariance matrix that are preserved in the solution to the NKPST problem, such as symmetry, positive definiteness, and other structures. We illustrate the separable approximation on a nonseparable space-time covariance function used in modeling Irish wind speed data. In Section 3, we discuss the computational gains associated with kriging with a separable covariance and describe the modeling of separable covariances. We derive an expression for the increase in mean squared prediction error in simple kriging due to the separable approximation and describe its application to the Irish wind speed data. We also investigate out-of-sample predictive performance of separable approximations of space-time covariances on this data set. We end the paper with a discussion in Section 4 of various extensions of the proposed methodology.

2. KRONECKER PRODUCT APPROXIMATIONS

In this section, we first describe the solution to the NKPST problem (6) based on the work of Van Loan and Pitsianis (1992) in linear algebra, and then characterize the properties of the solution based on typical features of space-time covariance matrices Σ. Recall that the Kronecker product $A \otimes B$ between two matrices $A \in \mathbb{R}^{n_1 \times n_2}$ and $B \in \mathbb{R}^{m_1 \times m_2}$ is a matrix of dimension $n_1 m_1 \times n_2 m_2$. It is an $n_1 \times n_2$ block matrix whose ij-th block is the $m_1 \times m_2$ matrix $a_{ij} B$. Recall also the vectorization operator which transforms a matrix $A \in \mathbb{R}^{n_1 \times n_2}$ into a vector $\mathrm{vec}(A) \in \mathbb{R}^{n_1 n_2}$ by stacking the columns of A on top of each other. The Kronecker product enjoys a very pleasant algebra, see, for example, Steeb (1997).

In the context of space-time data, the main computational advantage of separability is that the number of parameters in the covariance matrix is reduced to $n(n+1)/2 + m(m+1)/2 - 1$ (the -1 corresponds to the additional constraint to make the model identifiable) based on $S \in \mathbb{R}^{n \times n}$ and $T \in \mathbb{R}^{m \times m}$, compared to $N(N+1)/2$ for a matrix $\Sigma \in \mathbb{R}^{N \times N}$ based on a nonseparable covariance. In

DOI: 10.1002/env

addition, the three following properties involving the inverse, the determinant, and the vectorization of a matrix, respectively, are very attractive: $(S \otimes T)^{-1} = S^{-1} \otimes T^{-1}$; $|S \otimes T| = |S|^m |T|^n$; and $(S \otimes T)\text{vec}(\Lambda) = \text{vec}(T \Lambda S)$, where $\Lambda \in \mathbb{R}^{m \times n}$. In particular, the first and third properties are essential to reduce the number of space-time kriging operations for large data sets, see Section 3.1.

2.1. *Solution to the NKPST problem*

The solution to the NKPST problem is given by the singular value decomposition of a permuted version of the space-time covariance matrix $\Sigma \in \mathbb{R}^{N \times N}$. The main idea is to rearrange Σ into another matrix $\mathcal{R}(\Sigma) \in \mathbb{R}^{n^2 \times m^2}$ such that the sum of squares that arise in (6) is exactly the same as the sum of squares that arise in $\|\mathcal{R}(\Sigma) - \text{vec}(S) \otimes \text{vec}(T)^T\|_F$. Note that $\mathcal{R}(\Sigma)$ is no longer a square matrix, unless $n = m$. For illustration, consider the case of $n = 3$ spatial locations and $m = 2$ temporal locations, that is, $N = 6$. The Frobenius norm (6) can be rewritten as:

$$
\left\| \left(\begin{array}{cc|cc|cc}
\sigma_{11} & \boxed{\sigma_{12}} & \sigma_{13} & \sigma_{14} & \sigma_{15} & \sigma_{16} \\
\sigma_{21} & \sigma_{22} & \sigma_{23} & \sigma_{24} & \sigma_{25} & \sigma_{26} \\
\hline
\sigma_{31} & \sigma_{32} & \sigma_{33} & \sigma_{34} & \sigma_{35} & \sigma_{36} \\
\sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44} & \sigma_{45} & \sigma_{46} \\
\hline
\sigma_{51} & \sigma_{52} & \sigma_{53} & \sigma_{54} & \sigma_{55} & \sigma_{56} \\
\sigma_{61} & \sigma_{62} & \sigma_{63} & \sigma_{64} & \sigma_{65} & \sigma_{66}
\end{array}\right) - \begin{pmatrix} \boxed{s_{11}} & s_{12} & s_{13} \\ s_{21} & s_{22} & s_{23} \\ s_{31} & s_{32} & s_{33} \end{pmatrix} \otimes \begin{pmatrix} t_{11} & \boxed{t_{12}} \\ t_{21} & t_{22} \end{pmatrix} \right\|_F
$$

$$
= \left\| \left(\begin{array}{cccc}
\sigma_{11} & \sigma_{21} & \boxed{\sigma_{12}} & \sigma_{22} \\
\hline
\sigma_{31} & \sigma_{41} & \sigma_{32} & \sigma_{42} \\
\hline
\sigma_{51} & \sigma_{61} & \sigma_{52} & \sigma_{62} \\
\hline
\sigma_{13} & \sigma_{23} & \sigma_{14} & \sigma_{24} \\
\hline
\sigma_{33} & \sigma_{43} & \sigma_{34} & \sigma_{44} \\
\hline
\sigma_{53} & \sigma_{63} & \sigma_{54} & \sigma_{64} \\
\hline
\sigma_{15} & \sigma_{25} & \sigma_{16} & \sigma_{26} \\
\hline
\sigma_{35} & \sigma_{45} & \sigma_{36} & \sigma_{46} \\
\hline
\sigma_{55} & \sigma_{65} & \sigma_{56} & \sigma_{66}
\end{array}\right) - \begin{pmatrix} \boxed{s_{11}} \\ s_{21} \\ s_{31} \\ s_{12} \\ s_{22} \\ s_{32} \\ s_{13} \\ s_{23} \\ s_{33} \end{pmatrix} \otimes \begin{pmatrix} t_{11} & t_{21} & \boxed{t_{12}} & t_{22} \end{pmatrix} \right\|_F \tag{7}
$$

For instance, the term $\sigma_{12} - s_{11}t_{12}$ (boxed above) appears on both sides of the equality (7), and so do all other terms. It is then easily seen that defining the n^2 rows of $\mathcal{R}(\Sigma)$ as the transpose of the vectorized $m \times m$ blocks of Σ yields

$$
\|\Sigma - S \otimes T\|_F = \|\mathcal{R}(\Sigma) - \text{vec}(S) \otimes \text{vec}(T)^T\|_F,
$$

and also clearly $\|\Sigma\|_F = \|\mathcal{R}(\Sigma)\|_F$. Thus, the NKPST problem has been reduced to a rank-one approximation of a rectangular matrix, the solution of which is well-known (Golub and Van Loan, 1996, p. 70–71). It is based on the singular value decomposition of $\mathcal{R}(\Sigma)$, that is, $U^T \mathcal{R}(\Sigma) V = \Delta = \text{diag}(\delta_1, \ldots, \delta_q)$, where $U \in \mathbb{R}^{n^2 \times n^2}$ and $V \in \mathbb{R}^{m^2 \times m^2}$ are orthogonal matrices and the singular values satisfy $\delta_1 \geq \delta_2 \geq \cdots \geq \delta_q \geq 0$, with $q = \text{rank}(\mathcal{R}(\Sigma)) = \min\{m^2, n^2\}$. The solution to the NKPST problem is therefore given by

$$
\text{vec}(S) = \sqrt{\delta_1}\mathbf{u}_1 \qquad \text{and} \qquad \text{vec}(T) = \sqrt{\delta_1}\mathbf{v}_1 \tag{8}
$$

DOI: 10.1002/env

where $\mathbf{u}_1$ and $\mathbf{v}_1$ are the first column of U and V, respectively. This elegant solution to the NKPST problem of course results from the use of the Frobenius norm in (6) and the choice of other norms would lead to a computationally difficult optimization problem. Note also that the separable approximation (6) of a correlation matrix is not equal to the correlation matrix associated with the separable approximation (6) of a covariance matrix, although the difference is usually rather small.

A natural question is how 'good' is the approximation from the NKPST solution? To this end, we define a separability approximation error index $\kappa_\Sigma(S, T)$ of a matrix Σ approximated by the Kronecker product of two matrices S and T by

$$\kappa_\Sigma(S, T) = \frac{\|\Sigma - S \otimes T\|_F}{\|\Sigma\|_F}$$

The separability approximation error index takes values between zero (if Σ is separable) and $\sqrt{1 - 1/q}$, and is minimized by the solutions S and T of the NKPST problem. It can be rewritten as a function of the singular values $\delta_1 \geq \cdots \geq \delta_q \geq 0$ of $\mathcal{R}(\Sigma)$, in the form

$$\kappa_\Sigma(S, T) = \sqrt{\frac{\sum_{i=2}^q \delta_i^2}{\sum_{i=1}^q \delta_i^2}}$$

which is equal to zero if the matrix Σ is separable since $\delta_2 = \cdots = \delta_q = 0$ in that case. The upper bound of κ_Σ is essentially 1 for large space-time data sets.

2.2. *Structured covariance matrices*

A covariance matrix Σ based on a valid covariance function (1) is symmetric and positive definite. In addition, it has sometimes a patterned structure that can be exploited computationally in algorithms that require the evaluation and the inversion of Σ, see Zimmerman (1989). In particular, kriging can be performed more efficiently with a structured covariance matrix, for example, a block Toeplitz matrix or a banded matrix.

Recall that a matrix Σ is said to be of Toeplitz form if its entries are constant on each diagonal. It is block Toeplitz if its blocks are constant on each block diagonal. Toeplitz and block Toeplitz structures arise naturally in applications with gridded data and stationary processes. For instance, a spatial process with equidistant observations in $\mathbb{R}^1$ and with a stationary covariance function yields a symmetric Toeplitz covariance matrix. The same process on a regular grid in $\mathbb{R}^2$ with an isotropic stationary covariance function yields a symmetric block Toeplitz covariance matrix with Toeplitz blocks. Moreover, if there is a regular temporal component with stationary covariance in time, the resulting space-time covariance matrix has an additional block Toeplitz structure. If the covariance function is compactly supported, then the resulting matrix Σ will have a banded structure, that is, it will be sparse because certain entries are zero.

We know the Kronecker product has a pleasant algebra, and for example, if the matrices S and T are either symmetric, positive definite, banded, Toeplitz, non-negative, orthogonal, nonsingular, or lower/upper triangle, then $S \otimes T$ is either symmetric, positive definite, banded, block Toeplitz, non-negative, orthogonal, nonsingular, or lower/upper triangle. The question is whether the reverse is true? In other words, how are the properties of the solutions S and T of the NKPST problem related to those of Σ?

The answer is that properties of the space-time covariance matrix Σ, such as symmetry, positive definiteness, bandedness, non-negativity, and Toeplitz structures, are preserved by the solutions S and T of the NKPST problem. We summarize these results in the following proposition.

Propositon 1. *Let $S \in \mathbb{R}^{n\times n}$ and $T \in \mathbb{R}^{m\times m}$ be the solutions to the NKPST problem for $\Sigma \in \mathbb{R}^{N\times N}$. Then:*

DOI: 10.1002/env

(a) If Σ is symmetric positive definite, then S and T are symmetric positive definite.
(b) If Σ has bandwidth pm and each $m \times m$ block in Σ has bandwidth q or less, then S has bandwidth p and T has bandwidth q.
(c) If Σ is non-negative, then S and T are non-negative.
(d) If Σ is block Toeplitz, then S and T are Toeplitz matrices.

The proof of (a), (b), and (c) can be found in Van Loan and Pitsianis (1992), whereas the proof of (d) is given by Kamm and Nagy (2000).

2.3. Application to a covariance model for Irish wind speed data

The goal of this section is to illustrate how closely a covariance matrix based on a parametric nonseparable space-time covariance function can be approximated by a separable one. We consider a covariance model used by Gneiting *et al*. (2007) in a study of Irish wind speed. The data consists of time series of 6574 daily average wind speed recorded during the period 1961–1978 at $n = 11$ meteorological stations. The structure of the records is poor in space but rich in time, with a total of $N = 72\,314$ space-time coordinates. We refer to Haslett and Raftery (1989) for further information about this data set and to Gneiting (2002b), de Luna and Genton (2005), Stein (2005, 2006), and Gneiting *et al*. (2007) for subsequent analyses. In particular, we use the same square root transformation, estimated seasonal effect and spatially varying mean as in the latter paper, yielding a so-called data of *velocity measures*. In addition, we also use the same split of the data into a training period (1961–1970) and a testing period (1971–1978) in order to investigate out-of-sample predictive performance of our separable approximation in Section 3.

Gneiting *et al*. (2007) considered a stationary but generally nonseparable correlation function model of the form

$$C(\mathbf{h}, u) = \begin{cases} \frac{1}{1+a|u|^{2\alpha}} & \text{if } \mathbf{h} = \mathbf{0}, \\ \frac{1-\nu}{1+a|u|^{2\alpha}} \exp\left(\frac{-c\|\mathbf{h}\|}{(1+a|u|^{2\alpha})^{\beta/2}}\right) & \text{otherwise,} \end{cases} \tag{9}$$

where a and c are temporal and spatial nonnegative scale parameters, respectively, and $\alpha \in (0, 1]$ is a smoothness parameter. The parameter $\beta \in [0, 1]$ controls the space-time interaction and $\beta = 0$ yields a space-time separable correlation function, for which the spatial correlations at different temporal lags u are proportional to each other. The weighted least squares estimates of the parameters of (9) from the training data of velocity measures during 1961–1970 are $\hat{a} = 0.972$, $\hat{c} = 0.00128$, $\hat{\alpha} = 0.834$, $\hat{\nu} = 0.0415$, and $\hat{\beta} = 0.681$. In this fit, the usable range of lags was $\|\mathbf{h}\| \leq 450$ km in space and $|u| \leq 3$ days in time, the latter being motivated by one-day ahead forecasts at the stations. Therefore, we consider the covariance matrix $\Sigma \in \mathbb{R}^{N \times N}$ with $N = 11 \times 4 = 44$, based on the fitted model (9) and the station-specific spatial empirical variances of the velocity measures during the training period. Because $\hat{\beta} = 0.681 > 0$, the covariance matrix is nonseparable, and apparently rather 'highly' nonseparable due to $\hat{\beta}$ being larger than 1/2.

Next, we consider the NKPST approximation of the covariance matrix Σ. First, notice that Σ is a block Toeplitz matrix due to equispaced lags in time, but each block is not of Toeplitz form, due to the irregularly spaced coordinates of the meteorological stations. This structure is visualized in the top-left panel of Figure 1 by a contour plot of the entries of Σ. The NKPST approximation of Σ is depicted by a contour plot of its entries in the top-right panel of Figure 1. Notice the similarity between those two panels, which is confirmed by a small separability approximation error index of $\kappa_\Sigma(S, T) = 1.4\%$. This is due to the $q = \min\{4^2, 11^2\} = 16$ singular values of $\mathcal{R}(\Sigma)$ plotted in the bottom-left panel of Figure 1, where only the first singular value is large. We further investigate the sensitivity of the

DOI: 10.1002/env

separability approximation error index as a function of β in the bottom-right panel of Figure 1. The bold curve is for $m = 4$ and the other curves for larger matrices Σ with $m = 10$, 20, and 50. The vertical dashed line is at $\hat{\beta} = 0.681$. Overall, we can see that the separability approximation error index is not larger than 2%, even in the most nonseparable settings described previously, thus indicating a

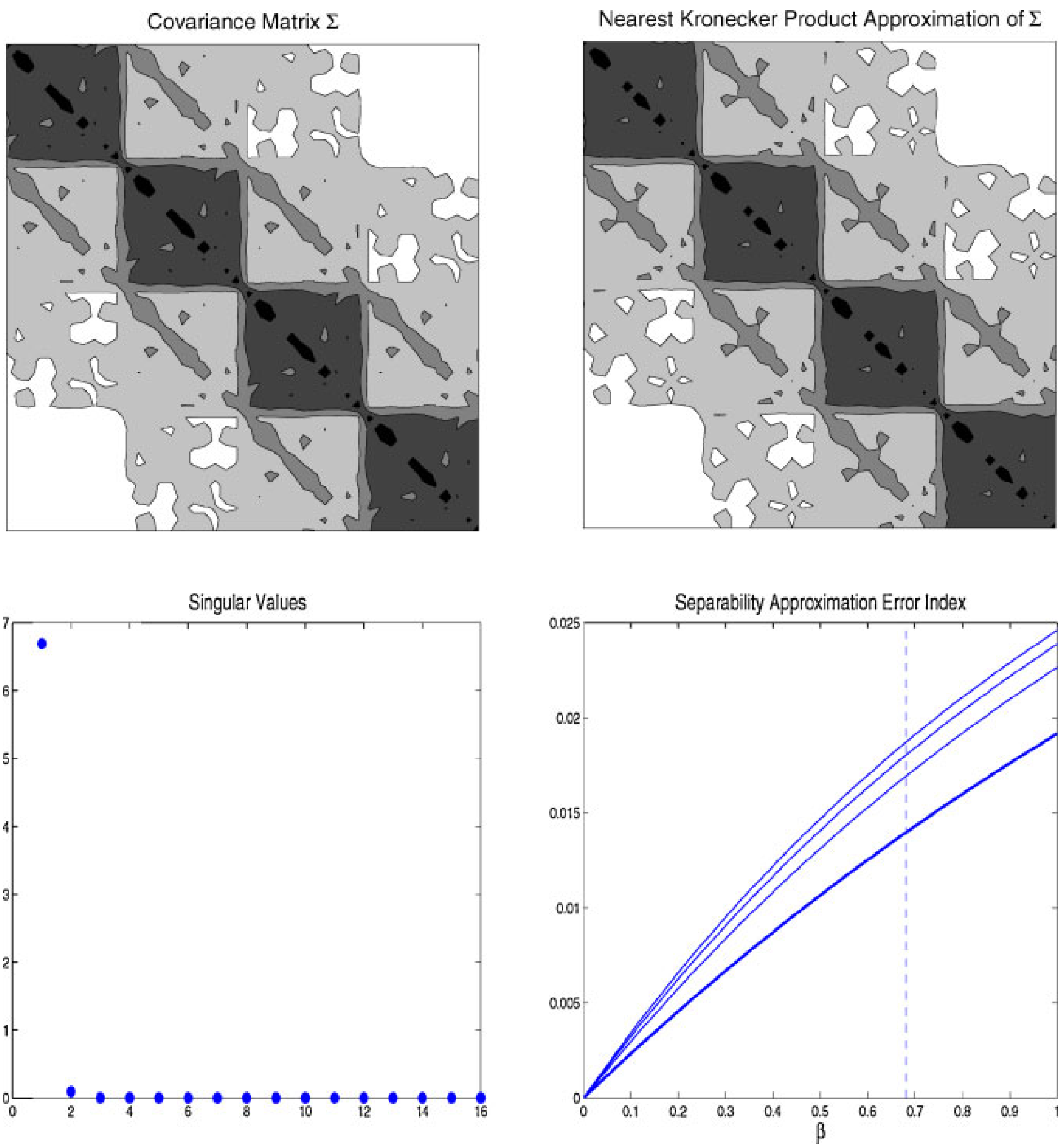

Figure 1. Covariance matrix Σ based on (9) for the Irish wind speed data estimated during the period 1961–1970. Top row: contour plots of the entries of Σ (left) and of the entries of its nearest Kronecker product approximation (right). Bottom row: singular values of $\mathcal{R}(\Sigma)$ (left) and separability approximation error index as a function of β (right). The bold curve is for $m = 4$ and the other curves for $m = 10$, 20, and 50. The vertical dashed line is at $\hat{\beta} = 0.681$

DOI: 10.1002/env

good NKPST approximation of Σ. Of course, this small index, based on the Frobenius norm, does not necessarily mean that the NKPST approximation is good in terms of mean squared prediction error in kriging. We investigate this question in Section 3.

3. KRIGING WITH SEPARABLE COVARIANCES

3.1. Separable kriging

From a computational point of view, the solution (8) to the NKPST problem only requires the largest singular value and corresponding singular vectors of $\mathcal{R}(\Sigma)$, not all of them. This can be performed very efficiently by means of the SVD Lanczos algorithm of Golub *et al*. (1981). Moreover, the explicit formation of $\mathcal{R}(\Sigma)$ is not necessary, and additional gains in computational efficiency can be obtained by exploiting the structure of the matrix Σ, for example, such as bandedness and/or a Toeplitz structure, see Kamm and Naggy (2000). This is important when dealing with large space-time data sets.

As described in the introduction, kriging requires solving an $N \times N$ linear system based on the covariance matrix Σ, a computationally difficult task for large N. However, the separability approximation of Σ by $S \otimes T$ leads to important gains in computational efficiency. Indeed, linear systems of the form

$$(S \otimes T)\boldsymbol{\lambda} = \mathbf{b}, \tag{10}$$

where $\boldsymbol{\lambda}$ is the vector of kriging weights and $\mathbf{b}$ is the vector of covariances between each observation and the new location, can be solved fast. For example, if $n = m$, then $\boldsymbol{\lambda}$ can be obtained in $O(m^3)$ flops (floating point operations) via the LU factorizations of S and T, see Van Loan and Pitsianis (1992). Without the exploitation of structure, an $m^2 \times m^2$ system would normally require $O(m^6)$ flops. Specifically, from the vectorization property listed at the beginning of Section 2, the system (10) is equivalent to $T\Lambda S = B$ with $\mathrm{vec}(\Lambda) = \boldsymbol{\lambda}$ and $\mathrm{vec}(B) = \mathbf{b}$, that is, $TY = B$ and $\Lambda S = Y$ with $Y \in \mathbb{R}^{m\times n}$. As an illustration, Table 1 reports necessary times (in seconds) to solve an $N \times N$ linear system in the

Table 1. Necessary times (in seconds) to solve an $N \times N$ linear system in Matlab for kriging at one new location based on an unstructured covariance matrix and a Kronecker structure. (Linux, 2.66 GHz Xeon processor with 16 Gbytes RAM)

$n = m$	N	Unstructured	Kronecker
16	256	0.0038	0.0002
25	625	0.0368	0.0003
36	1296	0.2513	0.0005
49	2401	1.4570	0.0007
64	4096	6.6984	0.0011
81	6561	25.4700	0.0018
100	10 000	—	0.0023
400	160 000	—	0.0830
900	810 000	—	0.7632
1600	2 560 000	—	3.9053
2500	6 250 000	—	14.3256
3600	12 960 000	—	41.3395
4900	24 010 000	—	102.0876
6400	40 960 000	—	—

DOI: 10.1002/env

numerical software package Matlab (on Linux, 2.66 GHz Xeon processor with 16 Gbytes RAM) for kriging at one new location. For simplicity, we set $n = m$. We compare kriging based on an unstructured covariance matrix with kriging based on a Kronecker structure. Table 1 shows that the computational gains arising from the Kronecker structure are important. A dash in the table indicates that there is a computer memory problem to solve the linear system. This occurs much earlier for the unstructured system than for the one with a Kronecker structure. Note also that the computation of a spatio-temporal map requires kriging at many new locations. In addition, further savings of time will result from the Kronecker structure when computing maps of kriging variances.

In order to perform kriging, the space-time covariance matrix Σ needs first to be estimated. Denote by $\hat{\Sigma}$ the estimate of Σ obtained either from space-time replicates or under a stationarity assumption. Then, compute the solutions $\hat{S}$ and $\hat{T}$ of the NKPST problem for $\hat{\Sigma}$. Fit a parametric spatial covariance model C_S to $\hat{S}$ and a parametric temporal covariance model C_T to $\hat{T}$. The final separable parametric space-time covariance model is $C = C_S C_T$. Observe that $\hat{S}$ and $\hat{T}$ capture the possible space-time nonseparability in $\hat{\Sigma}$.

3.2. MSPE with separable approximations

We investigate the NKPST approximation in terms of mean squared prediction error (MSPE) in kriging. Assume the random field Z is observed at N space-time coordinates, and for simplicity, that it has a zero mean function. The problem of interest is to predict Z at a new set of L arbitrary space-time coordinates, denoted by $\mathbf{Z}_0 = (Z(\boldsymbol{s}_{01}, t_{01}), \ldots, Z(\boldsymbol{s}_{0L}, t_{0L}))^T$ based on the vector of observations $\mathbf{Z} = (Z(\boldsymbol{s}_1, t_1), \ldots, Z(\boldsymbol{s}_N, t_N))^T$. We further assume that the vector $(\mathbf{Z}^T, \mathbf{Z}_0^T)^T$ has a multivariate normal joint distribution, $N_{N+L}(\mathbf{0}, \tilde{\Sigma})$, where $\tilde{\Sigma}$ is partitioned according to

$$\tilde{\Sigma} = \begin{pmatrix} \Sigma & \Sigma_{Z0} \\ \Sigma_{0Z} & \Sigma_{00} \end{pmatrix} \tag{11}$$

with $\Sigma \in \mathbb{R}^{N\times N}$, $\Sigma_{Z0} \in \mathbb{R}^{N\times L}$, $\Sigma_{0Z} \in \mathbb{R}^{L\times N}$, and $\Sigma_{00} \in \mathbb{R}^{L\times L}$. As mentioned in the introduction, the simple kriging predictor (2) of $\mathbf{Z}_0$ is

$$\hat{\mathbf{Z}}_0 = \Sigma_{0Z}\Sigma^{-1}\mathbf{Z},$$

with associated mean squared prediction error given by

$$\mathrm{MSPE}(\hat{\mathbf{Z}}_0) = \mathrm{tr}\left[\Sigma_{00} - \Sigma_{0Z}\Sigma^{-1}\Sigma_{Z0}\right],$$

where tr denotes the trace of a matrix. Suppose now that instead of $\tilde{\Sigma}$, we use $\tilde{\Omega} = S \otimes T$, where S and T are the solutions to the NKPST approximation of $\tilde{\Sigma}$. With the same partition of $\tilde{\Omega}$ as in (11), the simple kriging predictor of $\mathbf{Z}_0$ based on the separable covariance matrix $\tilde{\Omega}$, while the true covariance matrix is $\tilde{\Sigma}$, is

$$\hat{\mathbf{Z}}_0^* = \Omega_{0Z}\Omega^{-1}\mathbf{Z},$$

with associated mean squared prediction error given by

$$\text{MSPE}(\hat{\mathbf{Z}}_0^*) = \text{tr}\left[\Sigma_{00} + \Omega_{0Z}\Omega^{-1}\Sigma\Omega^{-1}\Omega_{Z0} - 2\Omega_{0Z}\Omega^{-1}\Sigma_{Z0}\right]$$

We define the relative difference in mean squared prediction error τ by the ratio

$$\tau = \frac{\text{MSPE}(\hat{\mathbf{Z}}_0^*) - \text{MSPE}(\hat{\mathbf{Z}}_0)}{\text{MSPE}(\hat{\mathbf{Z}}_0)} \tag{12}$$

which can be checked to be always positive, except when the covariance matrix $\tilde{\Sigma}$ is separable, in which case $\tau = 0$. Thus, τ measures the effect of using a separable approximation of the space-time covariance matrix $\tilde{\Sigma}$ in terms of mean squared prediction error.

As an illustration, we return to the Irish wind speed data considered in Section 2.3. Motivated by one-day ahead forecasts at the stations from time lags of three days, we consider the covariance matrix $\tilde{\Sigma}$ with $N = 33$ and $L = 11$ based on the fitted model (9) and the station-specific spatial empirical variances of the velocity measures during the training period. However, we allow β to vary in its admissible range and thus, $\tau = \tau(\beta)$. Figure 2 depicts the relative difference in mean squared prediction error $\tau(\beta)$ as a function of β. At the estimate $\hat{\beta} = 0.681$, represented by the vertical dashed line, we have $\tau(0.681) = 0.35\%$, which is small. Even at the most nonseparable model for this example, the relative difference in mean squared prediction error is only $\tau(1) = 0.72\%$. This suggests that there is little loss in terms of mean squared prediction error in using a separable approximation of the space-time covariance matrix used in this case.

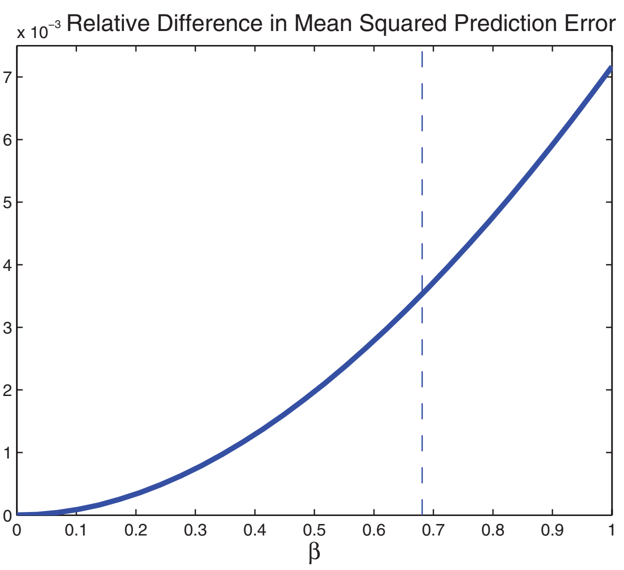

Figure 2. Relative difference in mean squared prediction error $\tau(\beta)$ as a function of β based on (9) for the Irish wind speed data estimated during the period 1961–1970. The vertical dashed line is at $\hat{\beta} = 0.681$

DOI: 10.1002/env

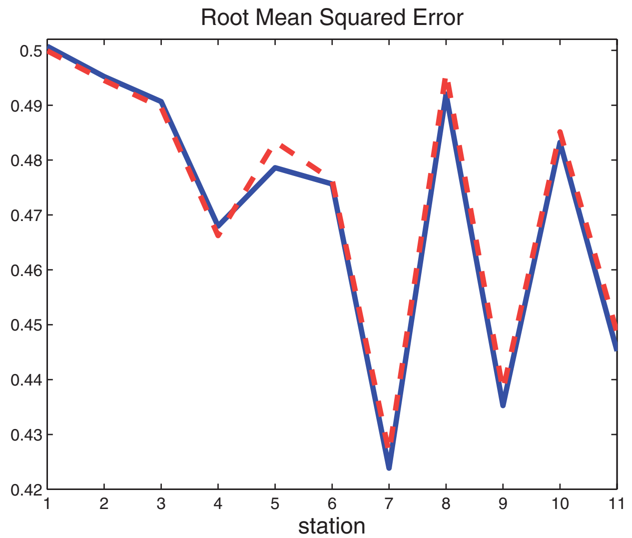

Figure 3. Root mean squared error at each of the 11 stations for the forecasts of Irish wind velocity measures during the 1971–1978 testing period based on the nonseparable covariance matrix (solid line) fitted during the 1961–1970 training period and its separable approximation (dashed line)

3.3. Predictive performance on Irish wind speed data

In order to assess the predictive performance of separable approximations of space-time covariance matrices, we consider time forward predictions of the Irish wind velocity measures during the 1971–1978 testing period based on the 1961–1970 training data. Specifically, we consider one-day ahead simple kriging forecasts at the stations from time lags of three days. We compare the fitted model (9) based on station-specific spatial empirical variances of the velocity measures during the training period, with its separable approximation. We compute root mean squared errors (RMSE) over the testing period of 2920 days for the forecasts based on these two space-time covariance models that were described in the top row of Figure 1.

Figure 3 depicts the RMSE at each of the 11 stations for the forecasts based on the nonseparable covariance matrix (solid line) and its separable approximation (dashed line). Overall, the differences in RMSE appear to be very small in this example. The RMSE for forecasts based on the separable approximation are slightly larger than those based on the nonseparable covariance for stations 5–9 (more so for station 5) but are even slightly smaller for stations 1–4. Thus, in this example, there is little loss in terms of RMSE when using a separable approximation of the space-time covariance matrix for kriging forecasts.

4. DISCUSSION AND EXTENSIONS

We have proposed a general methodology for computing separable approximations of space-time covariance matrices and illustrated that it results in small differences in prediction error while providing computational savings for large space-time data sets. Although we do not claim that separable

DOI: 10.1002/env

covariances should be used as a routine, we feel that existing nonseparable parametric covariance models do not produce as much nonseparability as one would think at first sight. We believe also that those ideas can be applied to many other settings, including likelihood-based problems and non-space-time frameworks. We now discuss briefly some extensions.

First, additional computational savings can be obtained by separable approximations in both space-time and space. Specifically, a separable approximation of the space-time covariance matrix Σ is obtained by minimization of $\|\Sigma - S \otimes T\|_F$, as described in this paper. Then, a separable approximation of the spatial covariance matrix $S \in \mathbb{R}^{n\times n}$ is obtained by minimization of $\|S - S_1 \otimes S_2\|_F$, where $S_1 \in \mathbb{R}^{n_1\times n_1}$ and $S_2 \in \mathbb{R}^{n_2\times n_2}$ are two spatial covariance matrices with $n = n_1 n_2$. Hence, we have the separable approximation $\Sigma \approx S_1 \otimes S_2 \otimes T$. Unfortunately, a direct minimization of $\|\Sigma - S_1 \otimes S_2 \otimes T\|_F$ does not have a closed-form SVD solution, see Van Loan (2000).

Another interesting direction consists in combining separable approximations with the tapering approach of Furrer *et al*. (2006). Indeed, in the NKPST problem, the optimization of (6) can be performed under constraints in order for the solution to have a prescribed structure, such as sparsity patterns or even a Toeplitz form. Those are imposed with constraints of the form

$$P_1^T \operatorname{vec}(S) = \mathbf{0} \qquad \text{and} \qquad P_2^T \operatorname{vec}(T) = \mathbf{0}$$

where $P_1 \in \mathbb{R}^{n^2\times p_1}$ and $P_2 \in \mathbb{R}^{m^2\times p_2}$ have full column rank. The solution of the constrained optimization problem is again obtained from an SVD decomposition, now of $Q_{P_1}^T \mathcal{R}(\Sigma) Q_{P_2}$, where Q_{P_1} and Q_{P_2} result from a QR decomposition of P_1 and P_2, see Van Loan and Pitsianis (1992). Thus, one can find the separable approximation of a nonseparable space-time covariance matrix while, for example, imposing a banded Toeplitz structure on the temporal covariance matrix T. In that case, it is expected that a combination of computational gains, reported in Table 1 for Kronecker structures and in Furrer *et al*. (2006) for tapering, can be achieved.

Sometimes, the covariance matrices Σ and T (or Σ and S) are given, for instance if the covariance matrix in time (or in space) is known. Then, the optimization of (6) is a frozen factor NKPST problem, that is, it reduces to linear least squares that can be solved easily.

When not only the first singular value δ_1 of $\mathcal{R}(\Sigma)$ is large, then the separable approximation will not be accurate and we can consider approximations with sums of Kronecker products by minimizing

$$\left\| \Sigma - \sum_{i=1}^{r} S_i \otimes T_i \right\|_F$$

the solution of which is given by

$$\operatorname{vec}(S_i) = \sqrt{\delta_i}\,\mathbf{u}_i \qquad \text{and} \qquad \operatorname{vec}(T_i) = \sqrt{\delta_i}\,\mathbf{v}_i$$

for $i = 1, \ldots, r$, $r \leq q$, based on the notation in Section 2.1. Kriging linear systems with this sum of Kronecker products structure can still be solved more efficiently than those without pattern. For example when $r = 2$, the generalized Schur decomposition triangularizes simultaneously the matrices S_1 and S_2, and T_1 and T_2, leading to a generalized Sylvester equation problem, see Gardiner *et al*. (1992).

When dealing with several variables evolving in space and time, that is, multivariate space-time stochastic processes, cokriging approaches are used. However, it is well-known that cokriging is computationally very costly already when considering a few variables. The separable approximation

DOI: 10.1002/env

techniques described in this paper can be used in this context too, in order to reduce the computational burden of cokriging.

Finally, separable approximations can also be used directly on the data, rather than on a covariance matrix. Indeed, separable process approximation can be achieved by the same SVD technique described in Section 2.1 since it is not restricted to square matrices.

ACKNOWLEDGEMENTS

The author thanks the editor and a referee for suggestions that improved this paper, as well as Professor Charles Van Loan for various discussions on Kronecker products. This work was partially supported by NSF grants DMS-0504896 and CMG ATM-0620624.

REFERENCES

Bochner S. 1955. *Harmonic Analysis and the Theory of Probability*. University of California Press: Berkeley and Los Angeles.

Brown P, Kåresen K, Roberts G, Tonellato S. 2000. Blur-generated nonseparable space-time models. *Journal of the Royal Statistical Society Series B* **62**: 847–860.

Cressie N. 1993. *Statistics for Spatial Data*. Wiley: New York.

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

de Luna X, Genton MG. 2005. Predictive spatio-temporal models for spatially sparse environmental data. *Statistica Sinica* **15**: 547–568.

Furrer R, Genton MG, Nychka D. 2006. Covariance tapering for interpolation of large spatial datasets. *Journal of Computational and Graphical Statistics* **15**: 502–523.

Fuentes M. 2006. Testing for separability of spatial-temporal covariance functions. *Journal of Statistical Planning and Inference* **136**: 447–466.

Gardiner J, Wette MR, Laub AJ, Amato JJ, Moler CB. 1992. Algorithm 705: A FORTRAN-77 software package for solving the Sylvester matrix equation $AXB^T + CXD^T = E$. *ACM Transactions on Mathematical Software* **18**: 232–238.

Genton MG, Koul HL. 2007. Minimum distance inference in unilateral autoregressive lattice processes. *Statistica Sinica*, in press.

Gneiting T. 2002a. Compactly supported correlation functions. *Journal of Multivariate Analysis* **83**: 493–508.

Gneiting T. 2002b. Nonseparable, stationary covariance functions for space-time data. *Journal of the American Statistical Association* **97**: 590–600.

Gneiting T, Genton MG, Guttorp P. 2007. Geostatistical space-time models, stationarity, separability and full symmetry. In *Statistics of Spatio-Temporal Systems. Monographs in Statistics and Applied Probability*. Finkenstaedt B, Held L, Isham V (eds). Chapman & Hall/CRC Press, Boca Raton, Florida, 151–175.

Golub G, Luk F, Overton M. 1981. A block Lanzcos method for computing the singular values and corresponding singular vectors of a matrix. *ACM Transactions on Mathematical Software* **7**: 149–169.

Golub G, Van Loan C. 1996. *Matrix Computations* (3rd edn). The Johns Hopkins University Press: London.

Guo J-H, Billard L. 1998. Some inference results for causal autoregressive processes on a plane. *Journal of Time Series Analysis* **19**: 681–691.

Haslett J, Raftery AE. 1989. Space-time modelling with long-memory dependence: Assessing Ireland's wind-power resource. *Applied Statistics* **38**: 1–50.

Kamm J, Nagy JG. 2000. Optimal Kronecker product approximation of block Toeplitz matrices. *SIAM Journal on Matrix Analysis and Applications* **22**: 155–172.

Kyriakidis PC, Journel AG. 1999. Geostatistical space-time models: a review. *Mathematical Geology* **31**: 651–684.

Li B, Genton MG, Sherman M. 2007. A nonparametric assessment of properties of space-time covariance functions. *Journal of the American Statistical Association*, **102**: 736–744.

Ma C. 2003. Families of spatio-temporal stationary covariance models. *Journal of Statistical Planning and Inference* **116**: 489–501.

Matsuda Y, Yajima Y. 2004. On testing for separable correlations of multivariate time series. *Journal of Time Series Analysis* **25**: 501–528.

Mitchell M, Genton MG, Gumpertz M. 2005. Testing for separability of space-time covariances. *Environmetrics* **16**: 819–831.

Mitchell M, Genton MG, Gumpertz M. 2006. A likelihood ratio test for separability of covariances. *Journal of Multivariate Analysis* **97**: 1025–1043.

Scaccia L, Martin RJ. 2005. Testing axial symmetry and separability of lattice processes. *Journal of Statistical Planning and Inference* **131**: 19–39.
Shitan M, Brockwell P. 1995. An asymptotic test for separability of a spatial autoregressive model. *Communications in Statistics—Theory and Methods* **24**: 2027–2040.
Steeb W-H. 1997. *Matrix Calculus and Kronecker Product with Applications and C++ Programs*. World Scientific: Singapore.
Stein ML. 1988. Asymptotically efficient prediction of a random field with a misspecified covariance function. *The Annals of Statistics* **16**: 55–63.
Stein ML. 1990. Uniform asymptotic optimality of linear prediction of a random field using an incorrect second-order structure. *The Annals of Statistics* **18**: 850–872.
Stein ML. 2005. Space-time covariance functions. *Journal of the American Statistical Association* **100**: 310–321.
Stein ML. 2006. Statistical methods for regular monitoring data. *Journal of the Royal Statistical Society Series B* **67**: 667–687.
Van Loan CF. 2000. The ubiquitous Kronecker product. *Journal of Computational and Applied Mathematics* **123**: 85–100.
Van Loan CF, Pitsianis N. 1992. Approximation with Kronecker products. In *Linear Algebra for Large Scale and Real-Time Applications*, De Moor BLR, Moonen MS, Golub GH (eds). Kluwer Publications, Dordrecht, The Netherlands, 293–314.
Zimmerman D. 1989. Computationally exploitable structure of covariance matrices and generalized covariance matrices in spatial models. *Journal of Statistical Computation and Simulation* **32**: 1–15.

DOI: 10.1002/env