Supplementary material

Title: Computationally Simple Anisotropic Lattice Covariograms
Journal: Environmental and Ecological Statistics
Authors: Dean C. Koch · Subhash R. Lele · Mark A. Lewis
Affiliation: Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Canada T6G 2R3. email: dkoch@ualberta.ca

A Matrix symmetries

All covariance matrices are symmetric about their diagonal, but a regular arrangement of sampling locations can lead to additional useful structures in $\vec{V}$:

Definition 1 **(Toeplitz structure)** $\vec{A}$ is *Toeplitz* when its entries satisfy $\vec{A}[i,j] = \vec{A}[i+k, j+k]$ for all feasible k. Equivalently, the entries along its k-diagonals are all equal. $\vec{A}$ is called *block-Toeplitz* when it has identical blocks along all of its block-diagonals.

Definition 2 **(The exchange matrix $\vec{J}_m$)** Also called the counteridentity, $\vec{J}_m$ is the $m \times m$ matrix with 1's on the counterdiagonal and zeroes everywhere else.

$$\underset{m\times m}{\vec{J}_m} = \begin{pmatrix} 0 & \dots & 0 & 1 \\ \vdots & & 1 & 0 \\ 0 & \iddots & & \vdots \\ 1 & 0 & \dots & 0 \end{pmatrix}$$

Note $\vec{J}_m$ is symmetric and orthogonal. When left-multiplied by matrix $\vec{A}$, it reverses the order of the columns in $\vec{A}$, and when right-multiplied it reverses the order of the rows. This leads to the following symmetry definitions:

Definition 3 **(Persymmetry and bisymmetry)** $\vec{A}$ is *persymmetric* if $\vec{J}_m\vec{A}\vec{J}_m = \vec{A}^T$. This is symmetry about the counterdiagonal. If $\vec{A}$ is symmetric then $\vec{A}\vec{J}_m$ and $\vec{J}_m\vec{A}$ are both persymmetric, and vice versa. $\vec{A}$ is *bisymmetric* if it is both symmetric and persymmetric.

Definition 4 **(Centrosymmetry)** $\vec{A}$ is *centrosymmetric* if $\vec{J}_m\vec{A}\vec{J}_m = \vec{A}$. This is symmetry about the centre of the matrix. Centrosymmetric matrices need not be symmetric, but when they are, bisymmetry is implied. If $\vec{J}_m\vec{A}\vec{J}_m = -\vec{A}$ then $\vec{A}$ is called *skew-centrosymmetric*.

B Kronecker products

the Kronecker product of $m \times n$ matrix $\vec{A}$ (with entries a_{ij}) and $p \times q$ matrix $\vec{B}$ is the matrix:

$$\underbrace{\vec{A} \otimes \vec{B}}_{mp\times nq} := \begin{pmatrix} a_{11}\vec{B} & \dots & a_{1n}\vec{B} \\ \vdots & & \vdots \\ a_{m1}\vec{B} & \dots & a_{mn}\vec{B} \end{pmatrix}. \tag{1}$$

Kronecker products admit the following useful identities (Iserles et al., 1990):

Proposition 1 **(Determinant)** *If $\vec{A}$ is order m and $\vec{B}$ is order n, then $|\vec{A} \otimes \vec{B}| = |\vec{A}|^n |\vec{B}|^m$.*

Proposition 2 **(Inverse)** *If $\vec{A}$ and $\vec{B}$ are nonsingular, then $\left(\vec{A} \otimes \vec{B}\right)^{-1} = \vec{A}^{-1} \otimes \vec{B}^{-1}$.*

Proposition 3 (Mixed product) *If the products $\vec{A}\vec{C}$ and $\vec{B}\vec{D}$ are conformable, then*

$$\left(\vec{A}\otimes\vec{B}\right)\left(\vec{C}\otimes\vec{D}\right)=\left(\vec{A}\vec{C}\otimes\vec{B}\vec{D}\right).$$

Proposition 4 (Decompositions of Kronecker products) *If $\vec{A}=\vec{U}_A\vec{C}_A\vec{V}_A$ and $\vec{B}=\vec{U}_B\vec{C}_B\vec{V}_B$,*

$$\vec{A}\otimes\vec{B}=\left(\vec{U}_A\otimes\vec{U}_B\right)\left(\vec{C}_A\vec{V}_A\otimes\vec{C}_B\vec{V}_B\right)=\left(\vec{U}_A\otimes\vec{U}_B\right)\left(\vec{C}_A\otimes\vec{C}_B\right)\left(\vec{V}_A\otimes\vec{V}_B\right).$$

Proposition 5 (Matrix equation representation) *If the product $A\vec{B}C$ is conformable, then*

$$vec\left(A\vec{B}C\right)=\left(\vec{C}^T\otimes\vec{A}\right)vec(\vec{B}).$$

C Indexing via selection matrices

We use the square-bracket notation of Poularikas (2014) to refer to submatrices and subvectors: If $\vec{A}$ is $M\times N$, and $\vec{\kappa}$ indexes a subset of rows and $\vec{\gamma}$ a subset of columns, then we write $\vec{A}[\vec{\kappa},\vec{\gamma}]$ for the corresponding submatrix. However we also allow duplications and re-orderings of rows and/or columns in $\vec{A}[\vec{\kappa},\vec{\gamma}]$ when this is indicated in $\vec{\kappa}$ and/or $\vec{\gamma}$. This concept is likely familiar already to users of the R programming language, in which arrays can be manipulated by this syntax. Let us make this precise:

Definition 5 (Selection matrices) If $\vec{\gamma}$ is a length-n vector with entries γ_i drawn from $\{1,\ldots,N\}$, and $\vec{e}_i^N$ is the i^{th} standard basis vector of $\mathbb{R}^N$, the $N\times n$ selection matrix $\vec{S}_{\vec{\gamma}}$ is:

$$\vec{S}_{\vec{\gamma}}:=\left(\begin{array}{cccc}\vec{e}_{\gamma_1}^N & \vec{e}_{\gamma_2}^N & \ldots & \vec{e}_{\gamma_n}^N\end{array}\right) \tag{2}$$

Thus if $\vec{\kappa}$ (length-m) has entries drawn from $\{1,\ldots,M\}$, and $\vec{\gamma}$ (length-n) has entries drawn from $\{1,\ldots,N\}$, then we write $\vec{A}[\vec{\kappa},\vec{\gamma}]$ to mean the $m\times n$ matrix product $\vec{S_\kappa}^T\vec{A}\vec{S_\gamma}$.

Using square-bracket notation we can also express (1) as a Hadamard (entrywise) product. Writing $\mathbb{1}_m$ for the length-m vector of ones, and letting $\vec{m}=(1,\ldots m)^T$ denote the column vector of the first m positive integers (and similar for $\vec{n}$, $\vec{p}$, and $\vec{q}$), we write

$$\vec{A}\otimes\vec{B}=\vec{A}[\vec{m}\otimes\mathbb{1}_p,\vec{n}\otimes\mathbb{1}_q]\odot\vec{B}[\mathbb{1}_m\otimes\vec{p},\mathbb{1}_n\otimes\vec{q}]. \tag{3}$$

This will be convenient when dealing with submatrices of Kronecker products, such as in the marginal covariance formula of Section E. We present this idea in a lemma:

Lemma 1 (Selections) *Suppose $\vec{A}$ is $m\times n$; and $\vec{B}$ is $p\times q$; where $M=mp$, $N=nq$. If $\vec{\gamma}_r$ (length-r) has entries drawn from $\{1,\ldots M\}$, and $\vec{\gamma}_c$ (length-c) from $\{1,\ldots N\}$, then*

$$(\vec{A}\otimes\vec{B})[\vec{\gamma}_r,\vec{\gamma}_c]=\vec{A}[(\vec{m}\otimes\mathbb{1}_p)[\vec{\gamma}_r],(\vec{n}\otimes\mathbb{1}_q)[\vec{\gamma}_c]]\odot\vec{B}[(\mathbb{1}_m\otimes\vec{p})[\vec{\gamma}_r],(\mathbb{1}_n\otimes\vec{q})[\vec{\gamma}_c]]. \tag{4}$$

Proof This is immediate from (3) on recognizing that the square-bracket indexing admits a composition rule $\vec{X}[\vec{\kappa}_r,\vec{\kappa}_c][\vec{\eta}_r,\vec{\eta}_c]=\vec{X}[\vec{\kappa}_r[\vec{\eta}_r],\vec{\kappa}_c[\vec{\eta}_c]]$, as well as a distributive rule over Hadamard products $\left(\vec{X}\odot\vec{Y}\right)[\vec{\kappa},\vec{\eta}]=\vec{X}[\vec{\kappa},\vec{\eta}]\odot\vec{Y}[\vec{\kappa},\vec{\eta}]$. These properties become obvious upon switching to the selection matrix representation in (2).

D Distance matrices for rectangular grids

We will focus on covariance models for lattice data where the random field variable $Z(\vec{s})$ is observed at a fixed set of locations $\mathscr{G}=\{\vec{s}_k\}_{k=1,\ldots n}\subset\mathbb{R}^2$ that together form a spatially regular $n^y\times n^x$ rectangular grid in the Cartesian plane (where $n=n^xn^y$). Assuming our random vector $\vec{Z}:=(Z(\vec{s}_1),\ldots Z(\vec{s}_n))^*$ is in column-vectorized order, the position (i,j) in the grid is mapped to the index $k=\mathrm{vec}_{n^y}(i,j):=i+n^y(j-1)$ in $\vec{Z}$.

For convenience we will assume that $\mathscr{G}$ has been scaled/translated to lie on the integer lattice $\mathbb{Z}^2$ with its bottom-left corner on the origin, so that its point coordinates are given by $\vec{s}_k = (x_k, y_k) = (j-1, n^y - 1)$.

In this configuration, pairwise distances between points in $\mathscr{G}$ can be succinctly represented using Kronecker products. To see this, number the grid lines of $\mathscr{G}$ using vectors $\vec{g^x} = (1, \dots n^x)^*$ and $\vec{g^y} = (n^y, \dots 1)^*$, so that if $\vec{s}_k$ is at the $(i,j)^{th}$ position in the grid, its x and y coordinates are $x_k = \left[(\mathbb{1}_{n^y})(\vec{g^x})^*\right]_{i,j}$ and $y_k = \left[(\vec{g^y})(\mathbb{1}_{n^x})^*\right]_{i,j}$ (where $\mathbb{1}_m$ is the length-m column vector of ones). Vectorizing these outer products, one obtains the full set of coordinates for $\mathscr{G}$ (in column-vectorized order) as the vectors $\vec{x} = (x_1, \dots x_n)^* = \vec{g^x} \otimes \mathbb{1}_{n^y}$ and $\vec{y} = (y_1, \dots y_n)^* = \mathbb{1}_{n^x} \otimes \vec{g^y}$. Denote the matrix of differences among the x coordinates as $\left[\vec{\delta}^x\right]_{i,j} := x_i - x_j$, and similar for y. Then a straightforward application of Property 3 yields:

$$\vec{\delta}^x = \vec{x}\mathbb{1}_n^* - \mathbb{1}_n\vec{x}^* = \left(\vec{g^x}\mathbb{1}_{n^x}^* - \mathbb{1}_{n^x}\vec{g^x}^*\right) \otimes (\mathbb{1}_{n^y}(\mathbb{1}_{n^y})^*), \tag{5}$$

$$\vec{\delta}^y = \vec{y}\mathbb{1}_n^* - \mathbb{1}_n\vec{y}^* = (\mathbb{1}_{n^x}(\mathbb{1}_{n^x})^*) \otimes \left(\vec{g^y}\mathbb{1}_{n^y}^* - \mathbb{1}_{n^y}\vec{g^y}^*\right). \tag{6}$$

This reveals a $(n^x \times n^x)$ pattern of $(n^y \times n^y)$ blocks in both matrices: in $\vec{\delta}^x$ the i,j^{th} block is $(i-j)I$; and in $\vec{\delta}^y$ it is $\vec{g^y}\mathbb{1}_{n^y}^* - \mathbb{1}_{n^y}\vec{g^y}^*$. These matrices provide an algebraically convenient decomposition of $\vec{s}_i - \vec{s}_j$ into its projections along the x and y axes, leading naturally to the idea of separability. They also allow us to write the matrix of pairwise Euclidean distances $\left[\vec{D}\right]_{ij} = d_{ij}$ as the (entrywise) square root of:

$$\vec{D} \odot \vec{D} = \vec{\delta}^x \odot \vec{\delta}^x + \vec{\delta}^y \odot \vec{\delta}^y. \tag{7}$$

D.1 Structure in covariance matrices for nonseparable SOS models

Note that both of the coordinate difference matrices in (5)-(6) are block-Toeplitz with Toeplitz blocks (BTTB). This is because $\vec{g}\mathbb{1}_m^* - \mathbb{1}_m\vec{g}^*$ expresses the differences between adjacent entries of the length-m vector $\vec{g}$, and therefore becomes Toeplitz when the entries of $\vec{g}$ are regularly spaced (eg. as when $\mathscr{G}$ is a regular grid).

In an SOS isotropic model, where covariance is defined by applying the covariogram function c to the vector $\vec{s}_i - \vec{s}_j = (\left[\vec{\delta}^x\right]_{i,j}, \left[\vec{\delta}^y\right]_{i,j})^*$, the covariance matrix $\vec{V}$ inherits the BTTB structure. This is because the entrywise application of any function to a matrix (for instance to $\vec{\delta}^x$ and $\vec{\delta}^y$) preserves all of its unsigned symmetry structures. Thus the entrywise squares, sums, and square-roots implied in (7) are all structure-preserving, which means that $\vec{D}$ inherits the BTTB property from $\vec{\delta}^x$ and $\vec{\delta}^y$, and $\vec{V}$ inherits it from $\vec{D}$ in turn. These arguments also apply with geometric anisotropy, where the $\vec{\delta}^x$ and $\vec{\delta}^y$ terms in (7) are replaced by linear combinations $s^x(\cos(\alpha)\vec{\delta}^x + \sin(\alpha)\vec{\delta}^y)$ and $s^y(\cos(\alpha)\vec{\delta}^y - \sin(\alpha)\vec{\delta}^x)$. Linear combinations preserve BTTB structure, and thus $\vec{V}$ is also BTTB.

The BTTB result was proven in more generality by Zimmerman (1989). However by emphasizing more directly how matrix symmetry patterns emerge from the regular spacing of grid lines, we believe our description has pedogogical value to readers unfamiliar with lattice designs. Moreover (5)-(7) holds more generally for irregular grid lines (with spacings defined in $\vec{g}^x$, $\vec{g}^y$) and so could prove useful in expressing the block structure in $\vec{V}$ common to any SOS model for gridded data.

E Covariance matrices for marginal distributions

Applications often call for marginal distributions over subsets of the data. Examples include cross-validation; conditional expectation, such as in Kriging; and inference with missing data. We will consider the Gaussian case, for which subsets of $\vec{Z}$ are themselves Gaussian with covariances given as submatrices of $\vec{V}$. Let $\vec{\kappa}$ denote the indexing vector for the subset of interest, so that $\vec{Z}_{\vec{\kappa}} = \vec{Z}[\vec{\kappa}]$ is the subset, and $\vec{V}_{\vec{\kappa}} = \vec{V}[\vec{\kappa}, \vec{\kappa}]$ is its covariance matrix.

Generally speaking $\vec{V}_{\vec{\kappa}}$ inherits neither separability, BTTB structure, nor bisymmetry from $\vec{V}$, unless the subset in question forms a regular rectangular grid (note that the angle estimator from Section 3 of

the main text exploits this exception.) Thus $\vec{V}_{\vec{\kappa}}$ may be more difficult to work with than $\vec{V}$, in spite of its smaller dimensionality. However if the subset is not too small, one can use a shortcut suggested in Ver Hoef et al. (2018): Leting $\vec{\kappa}^c$ index the observations not in $\vec{\kappa}$, the determinant and inverse of $\vec{V}_{\vec{\kappa}}$ can be computed using:

$$\vec{V}_{\vec{\kappa}}^{-1} = \vec{V}^{-1}[\vec{\kappa},\vec{\kappa}] - \left(\vec{V}^{-1}[\vec{\kappa},\vec{\kappa}^c]\right)\left(\vec{V}^{-1}[\vec{\kappa}^c,\vec{\kappa}^c]\right)^{-1}\left(\vec{V}^{-1}[\vec{\kappa}^c,\vec{\kappa}]\right) \tag{8}$$

$$|\vec{V}_{\vec{\kappa}}| = |\vec{V}||\vec{V}^{-1}[\vec{\kappa}^c,\vec{\kappa}^c]|. \tag{9}$$

When it is computationally cheap to evaluate the inverse and determinant of $\vec{V}$, but not $\vec{V}_{\vec{\kappa}}$, these formulae can lead to substantial speedups; If m is the length $\vec{\kappa}$ and n the length of $\vec{Z}$, we trade a problem of dimension m for one of dimension $n-m$ (the length of $\vec{\kappa}^c$).

These formulae are proved as follows. Suppose $\{1,\dots n\}$ is partitioned into sets of size m and $n-m$, identified by the index vectors $\vec{\kappa}$ and $\vec{\kappa}^c$, respectively. Then if $\vec{V}$ is any $n\times n$ matrix, the permutation $\vec{P} = (\,\vec{S}_{\vec{\kappa}}\;\vec{S}_{\vec{\kappa}^c}\,)$ takes $\vec{V}$ and its inverse to the block forms:

$$\vec{P}^T\vec{V}\vec{P} = \begin{pmatrix} \vec{V}_{\vec{\kappa}} & \vec{V}[\vec{\kappa},\vec{\kappa}^c] \\ \vec{V}[\vec{\kappa}^c,\vec{\kappa}] & \vec{V}_{\vec{\kappa}^c} \end{pmatrix},\ \vec{P}^T\vec{V}^{-1}\vec{P} = \begin{pmatrix} \vec{V}^{-1}[\vec{\kappa},\vec{\kappa}] & \vec{V}^{-1}[\vec{\kappa},\vec{\kappa}^c] \\ \vec{V}^{-1}[\vec{\kappa}^c,\vec{\kappa}] & \vec{V}^{-1}[\vec{\kappa}^c,\vec{\kappa}^c] \end{pmatrix} \tag{10}$$

Writing $[\vec{A}]_{ij}$ for the four blocks of a square partitioned matrix $\vec{A}$, the matrix $\vec{A}/[\vec{A}]_{22} := [\vec{A}]_{11} - [\vec{A}]_{12}[\vec{A}]_{22}^{-1}[\vec{A}]_{21}$ is called the *Schur complement of* $[\vec{A}]_{22}$ *in* $\vec{A}$. Ver Hoef et al. (2018) note that $[\vec{A}]_{11}^{-1} = \vec{A}^{-1}/[\vec{A}^{-1}]_{22}$, and suggest this identity as a shortcut to invert a block of a covariance matrix, after having computed its full inverse. Thus equation (8) follows from letting $\vec{A} = \vec{P}^T\vec{V}\vec{P}$ and writing its inverse in the form (10). Equation (9) follows from the identity $|\vec{A}/\vec{B}| = |\vec{A}|/|\vec{B}|$ for Schur complements.

E.1 A Bayesian perspective

A more general version of (8) due to Pukelsheim is presented in Glaser and Friston (2003, pp. 447-452), in the context of likelihood functions for hierarchical models. Before discussing applications, let us revisit this Bayesian perspective. Suppose we put an improper prior on the mean of $\vec{Z}$. In particular consider the normal hierarchical model,

$$\vec{Z}\mid\vec{\mu} \sim N(\vec{X}_{\vec{\kappa}^c}\vec{\mu},\vec{V}) \;\text{ with }\; \mu_i \overset{iid}{\sim} \text{unif}(-\infty,\infty), \tag{11}$$

where the $n\times n$ design matrix $\vec{X}_{\vec{\kappa}^c}$ has a one at each diagonal entry indexed in $\vec{\kappa}^c$, and is zero otherwise. Using (2) we can write $\vec{X}_{\vec{\kappa}^c} = \vec{S}_{\vec{\kappa}^c}(\vec{S}_{\vec{\kappa}^c})^T$ where the $N\times(N-n)$ selection matrix $\vec{S}_{\vec{\kappa}^c}$ is generated from the length-$(N-n)$ indexing vector for unobserved datapoints.

The likelihood function $\mathscr{L}(\vec{V}\mid\vec{Z}=\vec{z})$ is then equal to $\mathscr{L}(\vec{V}\mid\vec{Y}=\vec{y})$ for the model $\vec{Y}\sim N(\vec{0},\vec{V}_{\vec{\kappa}})$, over all $\vec{z}$ for which $\vec{z}[\vec{\kappa}] = \vec{y}$. This is because the effect of $\vec{X}_{\vec{\kappa}^c}\vec{\mu}$ is to add noise to the mean-zero version of $\vec{Z}$, but only to those components indexed in $\vec{\kappa}^c$. As the variance of this noise increases, the corresponding entries in the precision matrix for (11) shrink, and in the limit of unbounded variance they vanish.

The improper Bayesian posterior likelihood completely ignores these components of $\vec{Z}$, as their values are washed out by the noise of the flat prior. From an information-theoretical perspective, it is equivalent to simply omit the noisy, unreliable components. Indeed the likelihood function for (11) contains exactly those formulae (8)-(9) that specify $\vec{Z}_{\vec{\kappa}}$.

E.2 Applications of the marginal covariance formula

Equations (8)-(9) will speed evaluations of the MVN likelihood function (for a marginal distribution) whenever it is computational simpler to find the inverse and determinant of $\vec{V}$ and $\vec{V}^{-1}[\vec{\kappa}^c,\vec{\kappa}^c]$ instead of $\vec{V}_{\vec{\kappa}}$. This happens when n is large and the number $(n-m)$ of points omitted from $\vec{\kappa}$ is relatively small. For example if we consider only the highest order (cubic) terms in the arithmetic complexity, then computation time is reduced whenever $\vec{V}$ (but not $\vec{V}_{\vec{\kappa}}$) is bisymmetric and $(n-m)/n < \sqrt[3]{(m/n)^3 - 1/4}$, or

(approximately) $(n-m)/n < 0.37$. If $\vec{V}$ (but not $\vec{V}_{\vec{\kappa}}$) is separable and we assume for simplicity that $\mathscr{G}$ is a *square* grid, the bound becomes $(n-m)/n < \sqrt[3]{(m/n)^3 - 2/n^{(3/2)}}$. Asymptotically (for large n), the bound is $(n-m)/n < 0.5$.

An obvious application is in analyses where the locations $\mathscr{D}$ lie on a lattice, but do not form a complete rectangular subset. In this case one simply formulates $\vec{V}$ over a grid $\vec{G}$ covering the bounding box for the data, by introducing empty cells and indexing them in $\vec{\kappa}^c$. Other applications include cross-validation and leave-one-out estimators, where subsets of the data are withheld and the marginal distribution over the remainder is of interest. For example Ver Hoef *et al.* (2018) implemented (8)-(9) in a predictions over unobserved data. More generally these formulae can simplify computations of the conditional expectation when a partial observation $\vec{Z}_{\vec{\kappa}}$ is used to inform estimates of the unobserved components $\vec{Z}_{\vec{\kappa}^c}$, since the conditional distribution of $\vec{Z}_{\vec{\kappa}^c}$ given $\vec{Z}_{\vec{\kappa}} = \vec{z}[\vec{\kappa}]$ is:

$$\mathscr{N}\left\{\vec{\mu}[\vec{\kappa}^c] + \vec{V}^*_{\vec{\kappa}\vec{\kappa}^c}\vec{V}^{-1}_{\vec{\kappa}}\left(\vec{z}[\vec{\kappa}] - \vec{\mu}[\vec{\kappa}]\right), \vec{V}_{\vec{\kappa}^c} - \vec{V}^*_{\vec{\kappa}\vec{\kappa}^c}\vec{V}^{-1}_{\vec{\kappa}}\vec{V}_{\vec{\kappa}\vec{\kappa}^c}\right\}, \tag{12}$$

where $\vec{V}_{\vec{\kappa}^c}$ is the marginal covariance matrix for $\vec{Z}_{\vec{\kappa}^c}$, and $\vec{V}_{\vec{\kappa}\vec{\kappa}^c}$ is the cross covariance of $\vec{Z}_{\vec{\kappa}}$ and $\vec{Z}_{\vec{\kappa}^c}$. Note that in the case of separable covariance, the computationally prohibitive step of forming the $N \times N$ matrix $\vec{V} = \vec{V}^x \otimes \vec{V}^y$ can be avoided. The required matrices can easily be computed directly from $\vec{V}^x$ and $\vec{V}^y$ using the identity (4), which leads to

$$\vec{V}_{\vec{\kappa}\vec{\kappa}^c} = \vec{V}[\vec{\kappa}, \vec{\kappa}^c] = \vec{V}^x[\vec{\kappa}^x, \vec{\kappa}^x] \odot \vec{V}^y[\vec{\kappa}^y, \vec{\kappa}^y], \tag{13}$$

$$\vec{V}_{\vec{\kappa}^c} = \vec{V}[\vec{\kappa}^c, \vec{\kappa}^c] = \vec{V}^x[\vec{\kappa}^x, (\vec{\kappa}^x)^c] \odot \vec{V}^y[\vec{\kappa}^y, (\vec{\kappa}^y)^c], \tag{14}$$

where the sites indexed in $\vec{\kappa}$ have x-coordinates $\vec{g}^x[\vec{\kappa}^x]$ and y-coordinates $\vec{g}^y[\vec{\kappa}^y]$, and the $\vec{\kappa}^x$, $\vec{\kappa}^y$ can be found using the inverse vectorization operator: $(\vec{\kappa}^x[i], \vec{\kappa}^y[i]) = \mathrm{vec}^{-1}_{n^y}(\vec{\kappa}[i])$,

$$\mathrm{vec}^{-1}_{n^y}(k) = (i_k, j_k) = \left(k - n^y\left(\left\lceil \frac{k}{n^y} \right\rceil - 1\right), \left\lceil \frac{k}{n^y} \right\rceil\right).$$

E.3 Symmetry structures in covariance matrices for lattice data

When $\vec{V}$ is numerically singular, both the FFT-based methods of Jun and Stein (2008) and the BTTB-based algorithm of Dietrich (1993) can become inaccurate due to numerical rounding errors (Iserles et al., 1990, sec. 2.7). This problem is not uncommon with large-n covariance models. For example the popular Gaussian covariogram (the $\nu \to \infty$ limit of the WM) is known for generating matrices with extremely large condition numbers. It may be prudent in this situation to avoid explicit inversions and Cholesky factorizations in favour of more robust and numerically stable methods. A common workaround involves the singular value decomposition (SVD) (Neumaier, 1998), however this has $\mathscr{O}(n^3)$ complexity.

We can nevertheless speed computations by appealing to *bisymmetry* in $\vec{V}$. This is symmetry about both the diagonal and counterdiagonal. More precisely, if $\vec{J}_n$ is the $n \times n$ *exchange matrix* (with ones on the counterdiagonal and zeros otherwise), then $\vec{V}$ is bisymmetric if $\vec{V} = \vec{J}_n\vec{V}\vec{J}_n$ and $\vec{V} = \vec{V}^*$. The SOS assumption implies bisymmetry in $\vec{V}$, since the BTTB property implies *persymmetry* (symmetry about the counterdiagonal).

For bisymmetric covariance matrices $\vec{V}$, a unitary similarity transformation cuts the dimensionality of factorizations in half, speeding up computations by a factor of four. We present the even-n case here, and refer readers to Abu-jeib (2002) for the odd case. Suppose that $n = 2m$, and $\vec{V}$ is $n \times n$ and bisymmetric. Writing $\vec{V}_a$ for the $(m \times m)$ top-left block and $\vec{V}_b$ for the bottom-left block of $\vec{V}$, we have the block-diagonalization:

$$\vec{H}_n\vec{V}\vec{H}^*_n = \begin{pmatrix} \vec{V}_a + \vec{J}_m\vec{V}^*_b & \\ & \vec{V}_a - \vec{V}_b\vec{J}_m \end{pmatrix}, \text{ where } \vec{H}_n = \frac{1}{\sqrt{2}}\begin{pmatrix} I & J_m \\ -J_m & I \end{pmatrix}. \tag{15}$$

For example let us consider the simple case of $m = 2$ (a 2×2 layout of gridpoints, with $n = 4$). A general covariance matrix for this layout, given the SOS assumption (but not necessarily isotropy) is:

$$\vec{V} = \begin{pmatrix} a & b & c & d \\ b & a & e & c \\ c & e & a & b \\ d & c & b & a \end{pmatrix} = \begin{pmatrix} \vec{V}_a & \vec{V}_b^* \\ \vec{V}_b & \vec{V}_a \end{pmatrix}, \text{ where } \vec{V}_a = \begin{pmatrix} a & b \\ b & a \end{pmatrix} \text{ and } \vec{V}_b = \begin{pmatrix} c & e \\ d & c \end{pmatrix}.$$

One easily verifies that we obtain the desired form:

$$\vec{H}_4 \vec{V} \vec{H}_4^* = \frac{1}{2} \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & -1 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a & b & c & d \\ b & a & e & c \\ c & e & a & b \\ d & c & b & a \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} a+d & b+c & 0 & 0 \\ b+c & a+e & 0 & 0 \\ 0 & 0 & a-e & b-c \\ 0 & 0 & b-c & a-d \end{pmatrix}.$$

The diagonal blocks of $\vec{H}_n \vec{V} \vec{H}_n^*$ have two important properties: They are symmetric, as a consequence of persymmetry in $\vec{V}_b$ and symmetry in $\vec{V}_a$; and the union of their spectra is the spectrum of $\vec{V}$, since $\vec{H}_n$ is orthogonal and $\vec{H}_n \vec{V} \vec{H}_n^*$ block-diagonal. This implies that both blocks inherit positive definiteness from $\vec{V}$, and hence their Cholesky factors exist. It also implies that their condition numbers are as good or better than that of $\vec{V}$.

Equation (15) therefore stably converts an n-dimensional factorization problem in $\vec{V}$ into a pair of $n/2$-dimensional ones in $\vec{H}_n \vec{V} \vec{H}_n^*$. In statistical applications, the $\mathscr{O}(n^3)$ complexity of the relevant factorization (*eg.* SVD, spectral, Cholesky) is therefore reduced by a factor of $2^3/2 = 4$. Moreover since $\vec{H}_n$ has only two nonzero elements per row/column, the overhead of performing this transformation is $\mathscr{O}(n^2)$, a negligible cost compared to the factorization step when n is large.

E.4 A geometrical perspective on bisymmetry

The action of $\vec{H}_n$ on vectorized lattice data from Gaussian random fields leads to an interesting theoretical aside. Since $\vec{V}_a + \vec{J}_m \vec{V}_b$ and $\vec{V}_a - \vec{J}_m \vec{V}_b^*$ are SPD, we can view them as covariance matrices for a pair of m-dimensional random vectors, $\vec{Z}_1$ and $\vec{Z}_2$. Indeed by standard MVN theory, if $\vec{Z} \sim \mathscr{N}\left(\vec{\mu}, \vec{V}\right)$ then $\vec{H}_n \vec{Z} := (\vec{Z}_1^*, \vec{Z}_2^*)^* \sim \mathscr{N}\left(\vec{H}_n \vec{\mu}, \vec{H}_n \vec{V} \vec{H}_n^*\right)$.

When vector $\vec{Z}$ represents a raster image, the transformed components $\vec{Z}_1$ and $\vec{Z}_2$ have a simple geometrical interpretation. Let $\vec{Z}^F$ denote the transformation of the raster $\vec{Z}$ by a left-right, and an up-down reflection. Then $\vec{Z}_1$ is simply the left half of the superimposition $\vec{Z}^F + \vec{Z}$. Similarly $\vec{Z}_2$ is the right half of the superimposition $\vec{Z}^F - \vec{Z}$.

Despite a striking symmetry between the left and right halves of $\vec{H}_n \vec{Z}$, they are statistically independent under the Gaussian model, since by (15) their cross-covariance matrices are zero. While Gaussian data can always be linearly transformed into independent subsets using eigenvectors of $\vec{V}$, (15) shows that for SOS lattice data, half of these eigenvectors are symmetric and the other half are skew-symmetric (Abu-jeib, 2002).

E.5 Indexing of the rotated sublattice

Here we derive the indexing vector $\vec{\kappa}$ for the rotated subgrid in $\mathscr{G}$ described in Section 3.1. Without loss of generality we will assume that N^x, N^y, n^x, and n^y have been suitably chosen to satisfy the boundary constraints in equation (8) of the main text.

Begin by applying the transformation $\vec{A}^T$ to the coordinates $\vec{s}_k$ of points in the $n^y \times n^x$ sublattice $\mathscr{G}_0$ with grid line vectors $\vec{g}_0^x = (0, \ldots n^x - 1)^T$ and $\vec{g}_0^y = (n^y - 1, \ldots 0)^T$. By construction, these new coordinates,

$$\vec{A}^T \mathscr{G}_0 = \left\{ \tilde{\vec{s}}_k = \vec{A}^T \vec{s}_k ; k = \mathrm{vec}_{n^y}(p,q), p = 1, \ldots n^y, q = 1, \ldots n^x \right\}$$

are a subset of $\mathbb{Z}^2$ that forms a regular rectangular grid in the rotated coordinate system $\mathscr{D}_\alpha = \left\{ \vec{R}_\alpha \vec{v}_k ; \vec{v}_k \in \mathbb{R}^2 \right\}$, with inter-point spacing $s = \|(\alpha^x, \alpha^y)\|$. Thus if we apply equation (6) of the main

text to the $\tilde{\vec{s}}_k$, we recover a separable kernel, since $\vec{A}\vec{A}^T = I$. It follows that after rescaling the coordinate difference matrices $\vec{\delta}^x$ and $\vec{\delta}^y$ by the new inter-point distance s, we obtain

$$c(\tilde{\vec{s}}_i - \tilde{\vec{s}}_j) = \sigma^2 c^x(s\left[\vec{\delta}^x\right]_{ij}) c^y(s\left[\vec{\delta}^y\right]_{ij})$$

which implies the Kronecker product representation.

Now let us track one of the grid points along the transformation. The position $\vec{s}_k = (q-1, n^y - p)^T$ of the $(p,q)^{th}$ element in $\mathscr{G}_0$ gets mapped to the new position $\tilde{\vec{s}}_k = \vec{A}^T\vec{s}_k = (\alpha^x(q-1) - \alpha^y(n^y - p), \alpha^x(n^y - p) + \alpha^y(q-1))^T$. In order to find the row/column index in $\vec{V}$ corresponding to $\tilde{\vec{s}}_k \in \vec{A}^T\mathscr{G}_0$, we must identify the element of $\mathscr{G}$ that coincides with $\tilde{\vec{s}}_k$. However, following the transformation by $\vec{A}^T$, some of the leftmost points of $\vec{A}^T\mathscr{G}_0$ will have fallen out of $\mathscr{G}$ (they have negative x coordinates). So, we simply translate all of the points $\vec{A}^T\mathscr{G}_0$ to the right by distance $\alpha^y(n^y - 1)$ to ensure that the least x coordinate becomes zero. Note that such translations of coordinates have no bearing on the covariance matrix for points in $\vec{A}^T\mathscr{G}_0$ under the SOS assumption.

Following this translation, the new positions are $(\tilde{x}_{pq}, \tilde{y}_{pq}) = (\alpha^x(q-1) + \alpha^y(p-1), \alpha^x(n^y - p) + \alpha^y(q-1))^T$, where $p = 1, \ldots n^y$, $q = 1, \ldots n^x$. With reference to $\mathscr{G}$, we identify their row and column indices as $\tilde{p} = N^y - \tilde{y}_{pq}$ and $\tilde{q} = 1 + \tilde{x}_{pq}$. This simplifies to $\tilde{p} = 1 + \alpha^x(p-1) - \alpha^y(q-1) + \alpha^y(n^x - 1)$ and $\tilde{q} = 1 + \alpha^x(q-1) + \alpha^y(p-1)$. The vectorized index for the element of $\mathscr{G}$ in row $\tilde{p}$ and column $\tilde{q}$ is $\text{vec}^{-1}_{N^y}(\vec{\kappa})[p,q] = \tilde{p} + N^y(\tilde{q} - 1)$. Substituting the above expressions for $\tilde{p}$ and $\tilde{q}$, and collecting terms, we obtain:

$$\left[\text{vec}^{-1}_{n^y}(\vec{\kappa})\right]_{ij} = 1 + \alpha^y(n^x - 1) + (N^y\alpha^y + \alpha^x)(p-1) + (N^y\alpha^x - \alpha^y)(q-1) \tag{16}$$

$$\implies \vec{V}_{\vec{\kappa}} = \vec{V}[\vec{\kappa}, \vec{\kappa}] = \sigma^2 c^x\left(s(\vec{n}^x\mathbb{1}^*_{n^x} - \mathbb{1}_{n^x}\vec{n}^{x*})\right) \otimes c^y\left(s(\vec{n}^y\mathbb{1}^*_{n^y} - \mathbb{1}_{n^y}\vec{n}^{y*})\right) \tag{17}$$

where $\vec{n^x}$ is the length-n^x vector $(1, 2, \ldots n^x)^*$ (representing the grid lines of $\mathscr{G}_0$), and similar for y.

F Mountain pine beetle covariates

Category	Covariates	Notes
topography	altitude, slope, aspect, lakes indicator	
host quality	tree density, pine density, stand age	baseline 2001 estimates from Beaudoin et al. (2014), with pine density adjusted for cumulative mortality
beetle activity	beetle pressure (200, 400, 800, 1600, 3200 metre radius)	a (WM) kernel convolution of the product of the previous year beetle damage and the host pine density. Radius indicates the effective range of the kernel
	beetles/host (200, 400, 800, 1600, 3200 metre radius)	beetle pressure divided by pine density (attack density)
microclimate	seasonal precipitation (winter and spring); precipitation as snow; degree days (below 0, below 18); seasonal temperatures: mean in coldest winter month; min, max, & averages (spring, winter)	weather station data interpolated using ClimateBC software (Wang et al., 2007)

Table 1 Covariates included in the linear regression model for mountain pine beetle damage patterns.

References

Abu-jeib I (2002) Centrosymmetric matrices: properties and an alternative approach. Can Appl Math Q 10(4):429–445

Beaudoin A, Bernier PY, Guindon L, Villemaire P, Guo XJ, Stinson G, Bergeron T, Magnussen S, Hall RJ (2014) Mapping attributes of Canada's forests at moderate resolution through kNN and MODIS imagery. Can J For Res 44(5):521–532, DOI 10.1139/cjfr-2013-0401

Dietrich CR (1993) Computationally Efficient Cholesky Factorization Of A Covariance Matrix With Bedk Toeplitz Structure. J Stat Comput Simul 45(3-4):203–218, DOI 10.1080/00949659308811481

Glaser D, Friston K (2003) Variance Components, vol 391. John Wiley & Sons, Hoboken, NJ, USA, DOI 10.1016/B978-012264841-0/50041-X

Iserles A, Golub GH, Loan CFV (1990) Matrix Computations, vol 74. JHU Press, Baltimore, MD, USA, DOI 10.2307/3619868

Jun M, Stein ML (2008) Nonstationary covariance models for global data. Ann Appl Stat 2(4):1271–1289, DOI 10.1214/08-AOAS183

Neumaier A (1998) Solving ill-conditioned and singular linear systems: A tutorial on regularization. SIAM Rev 40(3):636–666, DOI 10.1137/S0036144597321909

Poularikas A (2014) Matrix Analysis. Cambridge University Press, New York, NY, USA, DOI 10.1201/b17464-12

Ver Hoef JM, Peterson EE, Hooten MB, Hanks EM, Fortin MJ (2018) Spatial autoregressive models for statistical inference from ecological data. Ecol Monogr 88(1):36–59, DOI 10.1002/ecm.1283

Wang T, Hamann A, Aitken S, Spittlehouse D (2007) A program to generate climate normal , decade , annual , seasonal and monthly data for genecology and climate change studies in British Columbia . For Sci (Figure 1):6

Zimmerman DL (1989) Computationally Exploitable Structure of Covariance Matrices and Generalized Covariance Matrices in Spatial Models. J Stat Comput Simul 32(1-2):1–15, DOI 10.1080/00949658908811149