

Computationally Simple Anisotropic Lattice Covariograms

Dean Koch · Subhash Lele · Mark A. Lewis

Received: date / Accepted: date

Abstract When working with contemporary spatial ecological datasets, statistical modellers are often confronted by two major challenges: (I) the need for covariance

Dean Koch is PhD candidate in the Department of Mathematical and Statistical Sciences, University of Alberta (UofA), Edmonton, Canada T6G 2R3 (email: dkoch@ualberta.ca); Subhash R. Lele is Professor of Mathematical and Statistical Sciences at UofA (email: slele@ualberta.ca); Mark A. Lewis (MAL) is Professor of Mathematical and Statistical Sciences and Biological Sciences at UofA (email: mark.lewis@ualberta.ca). The authors thank the Lewis Research Group for providing expert advice and feedback. MAL is also grateful for support through NSERC and the Canada Research Chair Program. This research was supported by a grant to MAL from the Natural Science and Engineering Research Council of Canada (Grant No. NET GP 434810-12) to the TRIA Network, with contributions from Alberta Agriculture and Forestry, Foothills Research Institute, Manitoba Conservation and Water Stewardship, Natural Resources Canada-Canadian Forest Service, Northwest Territories Environment and Natural Resources, Ontario Ministry of Natural Resources and Forestry, Saskatchewan Ministry of Environment, West Fraser and Weyerhaeuser.

Dean Koch

Department of Mathematical and Statistical Sciences, University of Alberta, 11324 89 Ave NW, Edmonton, AB, T6G 2J5, Canada.

E-mail: dkoch@ualberta.ca

models with the flexibility to accommodate directional patterns of anisotropy; and (II) the computational effort demanded by high-dimensional inverse and determinant problems involving the covariance matrix \vec{V} . In the case of rectangular lattice data, the spatially separable covariogram is a longstanding but underused model that can reduce arithmetic complexity by orders of magnitude. We examine a class of covariograms for stationary data that extends the separable model through affine coordinate transformations, providing a far greater flexibility for handling anisotropy than that offered by the standard approach of using geometric anisotropy to extend an isotropic model. This motivates our development of an extremely fast estimator of the orientation of the axes of range anisotropy on spatial lattice data, and a powerful visual diagnostic for nonstationarity. In a case study, we demonstrate how these tools can be used to analyze and predict forest damage patterns caused by outbreaks of the mountain pine beetle.

Keywords bisymmetry · covariogram · geometric anisotropy · lattice · separable covariance

1 Introduction

In the analysis of large scale spatial ecological data, researchers frequently encounter special statistical challenges that preclude the use of more traditional models. Examples include incomplete data (Nakagawa and Freckleton 2008); extremely large datasets (Simpson et al. 2012); complex dependencies among model residuals (Legendre 1993); an inability to replicate measurements; and complex underlying ecolog-

ical systems that obscure the relationship between covariate and response (Inchausti 1998).

A large number of spatial regression methods have been developed to address these challenges, including generalized least squares (GLS), the autoregressive model family, generalized additive mixed models, as well as fully Bayesian approaches. In fact some authors (eg. Beale et al. 2010; Hawkins 2012) suggest the sheer diversity of methods in the literature and resulting paralysis-by-analysis has been an impediment to the widespread adoption of spatial techniques by ecologists. A framework that is both immediately intuitive and easy to implement will therefore be useful to researchers outside the field of statistics.

We believe the geostatistical approach is appropriate for this role. Geostatistics, whose name reflects early origins in the mining industry, has over time grown into a quite general and mature spatial statistical framework (see eg. Cressie 1992; Banerjee et al. 2014). Its theory is built from regionalized random variables $Z(\vec{s})$ whose properties depend on a location index \vec{s} that varies continuously through some spatial domain \mathcal{D} . Covariances among the n sample points $Z(\vec{s}_i)$ are explicitly specified by the *covariogram* (also known as a covariance function or kernel), which maps coordinate pairs \vec{s}_i, \vec{s}_j to the entries of the $n \times n$ covariance matrix \vec{V} . In applications, covariograms that decrease with separation distance can serve as convenient models for data exhibiting spatial autocorrelation (SAC).

While this transparent representation of covariance has intuitive appeal, it can lead to computational difficulties with large sample sizes (n), owing to the $O(n^3)$ complexity of important matrix computations involving \vec{V} (Simpson et al. 2012). If,

however, the sample sites form a rectangular lattice, the structure of \vec{V} can often be simplified and these computational difficulties largely avoided. Such sampling designs are increasingly common in ecology today with the widespread adoption of remote sensing methods (Wulder et al. 2006).

Computational shortcuts relying on a gridded layout of sampling locations, such as Kronecker product decompositions of \vec{V} for separable models (Martin 1979), are often tailored for the parsimonious scenario of isotropic and stationary covariance – *ie.* where the covariance structure lacks directionality and location-dependence. However, departures from isotropy and stationarity are the norm in ecological data analysis, particularly when the model covers large geographical areas (Dale and Fortin 2014). Some of this may be due to unaccounted-for trends, which can be remedied by improving the mean model. More commonly, however, they are an important and intrinsic feature of the error-generating process, such as dispersal (in models of ecological populations). This demands a refinement of the covariance model.

Geometric anisotropy, in which a transformation (\vec{A}) of coordinates deforms the circular contours of an *isotropic* model, is one of the more common refinements for introducing directionality into covariograms. However, this characterization of anisotropy has several drawbacks. It requires a rather inflexible assumption that \vec{A}^{-1} produces an isotropic covariance function (*ie.* contours of constant covariance are assumed to be ellipses). Sill anisotropy is therefore neglected (Zimmerman 1993). Moreover, in practice, the angle of anisotropy is unknown, so it must be somehow inferred from the data. Finally, by transforming the coordinate system of a separable

covariogram, one typically loses the Kronecker product structure in \vec{V} that would otherwise be exploited for computational efficiency.

In this paper, we explore the problem of introducing anisotropy (both range and sill) into *separable* covariance models using the transformation \vec{A} . This leads to a family of covariograms with improved flexibility for handling departures from isotropy, compared with the standard approach of geometric anisotropy. We show how a careful subsampling of gridded datapoints allows modellers to retain the computational advantages of the Kronecker product under a rotated coordinate system. This is critically important in applications, as it allows the covariogram to be applied to large and/or densely sampled geographical areas, wherein the sample size of the gridded survey is too large to (feasibly) use standard $O(n^3)$ complexity techniques for variable selection, model fitting, prediction, *etc.*

This idea leads to a novel and fast estimator of the direction of range anisotropy in geometrically anisotropic data. By applying this estimator in a blockwise fashion to spatial data, we show how the (computationally fast) separable covariogram can be repurposed as a graphical tool for studying nonstationary covariance structures. Though we emphasize ecological data and SAC, the computational simplicity of these models makes them attractive in broader applications.

1.1 Why model spatial autocorrelation?

SAC describes the extent to which the random components of measurements that are near in space tend to be more (or less) similar than expected for a spatially distant pair. This phenomenon is extremely common in ecological studies, where the data-

generating process is often driven by environmental factors shared among nearby sites, and/or demographic processes intrinsically tied to distance, such as aggregation or dispersal (Beale et al. 2010).

It is widely accepted that neglecting SAC (in favour of an independence assumption) amounts to pseudoreplication, and leads to precision issues for inference (Legendre 1993). Though some authors dispute the importance of this misspecification, it is clear that SAC should be examined in the course of model development (Keitt et al. 2002). Examples from our research area include Klutsch *et al.* (2009), who used SAC as a model diagnostic in a study of environmental predictors for outbreaks of the mountain pine beetle, and Robertson *et al.* (2009), who improved their outbreak model with a simple SAC model.

The autocorrelation patterns themselves are sometimes of scientific interest. For example, Aukema *et al.* (2008) used nonparametric covariance functions to study how spatial synchrony drives different stages of beetle outbreaks. SAC may illuminate features of the underlying ecological system that generates a dataset. We will touch on this idea in a case study of beetle outbreak data in Sect. 4.

1.2 Covariograms in practice

In explicit covariance models, all n^2 entries of \vec{V} must be specified, of which up to $n(n+1)/2$ can be distinct (by symmetry). Since these values are typically unknown, \vec{V} is constructed using a parametric covariogram $c : (\vec{s}_i, \vec{s}_j; \vec{\theta}) \mapsto \text{Cov} \{Z(\vec{s}_i), Z(\vec{s}_j)\}$, and the parameters $\vec{\theta}$ inferred from the data. Many functional forms have been proposed for c , but some care is required to ensure a well defined distribution (Guillot et al.

2014). In particular c (and \vec{V}) must be symmetric and positive definite (SPD). The reader is directed to Roberts et al. (2013) for a discussion of design principles and admissible forms for c .

In spatial ecology, modellers tend to choose covariograms that are a functions of separation distance $d_{ij} = \|\vec{s}_i - \vec{s}_j\|$ only, ignoring directionality and position. This reflects the assumption that the $Z(\vec{s}_i)$ are drawn from a second-order stationary (SOS) and isotropic random field (Chen et al. 2019); that is, the covariance structure is invariant to translations and/or rotations of \mathcal{D} . A popular example is the Whittle-Mátern (WM) covariogram which, in 2-dimensional (2D) space, can be written:

$$c(d_{ij}; \sigma, \lambda, \nu) = \sigma^2 \left(2^{1-\nu} / \Gamma(\nu) \right) (2\sqrt{\nu}d_{ij}/\lambda)^\nu H_\nu(2\sqrt{\nu}d_{ij}/\lambda), \quad (1)$$

where H_ν is the modified Bessel function of the second kind of order ν . Figure 1 (top-left) illustrates the type of spatial patterns generated by the WM. This covariogram has many names, and many desirable properties, as chronicled in Guttorp and Gneiting (2006). Given its ubiquity and importance in statistics, we will make use of the WM later on as a reference model for generating data in our simulation studies.

Isotropy is however seldom justified, except as a means to a parsimonious model. A more robust model should allow directionality, and one of the simplest ways of building this into a covariogram is to assume *geometric anisotropy*. This extends the isotropic SOS covariogram by applying an affine transformation \vec{A} to the coordinate system, and measuring distances by $d_{ij} := \|\vec{A}(\vec{s}_i - \vec{s}_j)\|$. In 2D, \vec{A} can be understood as the product of a diagonal scaling matrix $\vec{S}(s^x, s^y)$, and a rotation matrix $\vec{R}_{-\alpha}$: Circular contours of constant covariance get mapped to ellipses whose axes are stretched by

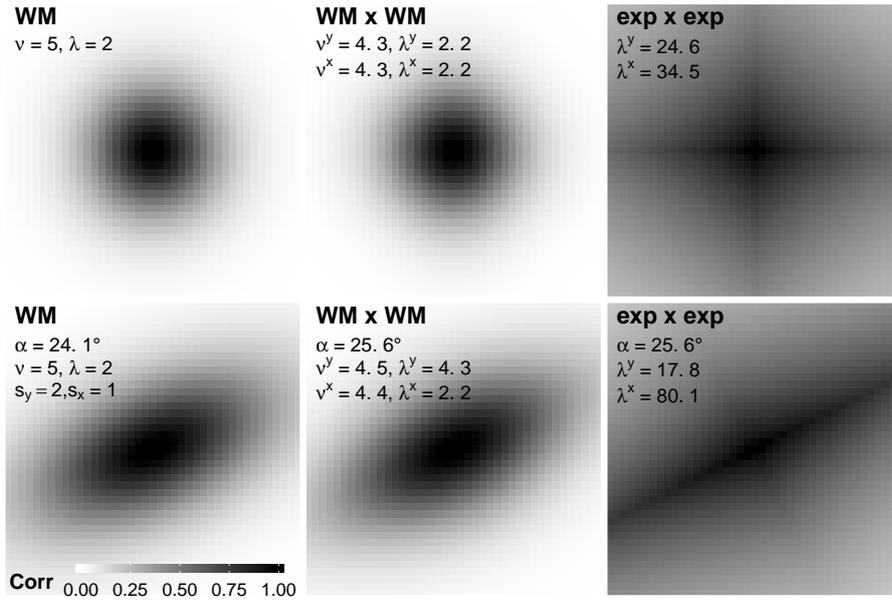


Fig. 1 Examples of covariograms in 2D illustrated by heatmaps of the correlation with the central point. The two leftmost panels illustrate isotropy (top) and geometric anisotropy (bottom). The others illustrate separable (top) and rotated product (bottom) covariograms, as introduced in Sect. 2.1-3. The two rightmost panels are examples of separable kernels fitted in simulations in Sect. 2.2, 3.3

factors s^x, s^y , and which are oriented along the counterclockwise rotation of the x, y axes by angle α . Figure 1 (bottom-left) illustrates the resulting pattern of SAC.

1.3 Computations with covariance matrices

Among the simplest implementations of the covariogram in regression is GLS. This extends ordinary least squares (OLS) on the residuals vector \vec{Z} , when \vec{V} is given. The model is:

$$\vec{Y} = \vec{\beta}\vec{X} + \vec{Z} \quad \text{where} \quad \mathbb{E}(\vec{Z} | \vec{X}) = \vec{0} \quad \text{Cov}(\vec{Z} | \vec{X}, \vec{\theta}) = \vec{V}(\vec{\theta}) \quad (2)$$

GLS uses the Cholesky factor \vec{L} of \vec{V} to define a transformed response $\vec{Y} = \vec{L}^{-1}\vec{Y}$ for which the problem of estimating $\vec{\beta}$ reduces to OLS. However since \vec{V} is seldom known, a parametric form $\vec{V} = \vec{V}(\vec{\theta})$ is often proposed, and $\vec{\theta}$ estimated by numerical likelihood maximization. For example, when Z is multivariate normal (MVN), we minimize:

$$-\log \mathcal{L}(\vec{\theta}, \vec{\beta} \mid \vec{X}, \vec{Y} = \vec{y}) \propto \log |\vec{V}(\vec{\theta})| + (\vec{y} - \vec{\beta}\vec{X})^* \vec{V}(\vec{\theta})^{-1} (\vec{y} - \vec{\beta}\vec{X}). \quad (3)$$

Note that here (and throughout the paper) we use an asterisk to denote transposes.

When n is large, evaluations of (3) can be computationally demanding because $\vec{V}(\vec{\theta})$ is usually dense, and analytic forms for $\vec{V}(\vec{\theta})^{-1}$ and $|\vec{V}(\vec{\theta})|$ are rarely available. Thus each time the optimizer adjusts $\vec{\theta}$, it must solve a factorization problem with arithmetic complexity $O(n^3)$. This problem arises in universal kriging interpolation methods (Simpson et al. 2012), as well as in more sophisticated extensions of GLS such as spatial generalized estimating equations (Dormann 2009); spatial generalized linear mixed models (Heagerty and Lele 1998); and Bayesian MCMC based techniques (Banerjee et al. 2014).

The autoregressive model family (CAR, SAR) avoids this problem by defining \vec{V} implicitly. A weights matrix is used to specify partial correlations rather than covariances, thus constructing the precision matrix \vec{V}^{-1} directly. Ver Hoef et al. (2018) and Beale et al. (2010) make compelling cases for the autoregressive approach. There are drawbacks, however. The implied correlation structure in \vec{V} is often unclear (Wall 2004); and some unintuitive restrictions on the weights are required to ensure a valid joint distribution for \vec{Z} .

Authors preferring the more intuitive covariogram approach have developed various structured forms for \vec{V} that avoid the large- n difficulties (Banerjee et al. 2014, chap. 12). These include Gaussian predictive process models (Banerjee et al. 2008), approximate likelihood (Vecchia 1988; Stein et al. 2004), fixed-rank kriging methods (Cressie and Johannesson 2008), kernel convolution and stochastic partial differential equation based methods (Simpson et al. 2012), as well as low-rank or sparse approximations of \vec{V} (Ambikasaran et al. 2016), such as covariance tapering (Kaufman et al. 2008). Following Genton (2007) we will be interested in the highly patterned forms of \vec{V} generated by a *separable* covariance model.

1.4 Paper outline

We begin with a review of separable covariance structures in Sect. 2, demonstrating in a simulation study their flexibility as surrogates for more common isotropic models. In Sect. 3 we propose a broader class of covariograms incorporating range anisotropy, and use them to develop a novel estimator of the angle of geometric anisotropy. Sect. 4 demonstrates the method on a spatial ecological dataset, to reveal nonstationary patterns of covariance in mountain pine beetle damage patterns.

2 Covariograms on the rectangular lattice

A random field is called SOS (or weakly stationary) when its covariogram has the form $c(\vec{s}_i - \vec{s}_j; \vec{\theta})$, and the data are detrended ($\mathbb{E}\{Z(\vec{s}_i)\} = \mu$), for all $\vec{s}_i, \vec{s}_j \in \mathcal{D}$ (Cressie 1992, sec. 2.3). We focus on the application of this model to 2D lattice data, where

the random field has been sampled at a fixed set of locations $\mathcal{G} \subset \mathbb{R}^2$ that together form a spatially regular $n^y \times n^x$ rectangular grid (with sample size $n = n^x n^y$).

For notational convenience we assume that $\vec{Z} := \{Z(\vec{s}_1), \dots, Z(\vec{s}_n)\}^*$ is in column-vectorized order, so that $\text{Vec}_{n^y}^{-1}(\vec{Z})$ recovers the natural $n^y \times n^x$ matrix representation for the data (*ie.* as a raster image). This ordering introduces computationally useful patterns in \vec{V} . For example, in an important early paper, Zimmerman (1989) showed that for SOS models on \mathcal{G} , the covariance matrix \vec{V} is block-Toeplitz with Toeplitz blocks (BTTB). He concluded with an algorithm that reduces the $\mathcal{O}(n^3)$ arithmetic complexity of solving \vec{V}^{-1} by a factor of n^x . Dietrich (1993) subsequently showed that the Cholesky factor of \vec{V} can be computed for the same complexity cost, providing a shortcut for the determinant and quadratic form in (3). In a similar vein, Jun and Stein (2008) showed how, for axially symmetric processes, the SOS property introduces a block-circulant structure which can be exploited using the Discrete Fourier Transform (DFT) to reduce the complexity of factorizations by a factor of $(n^x)^2$.

In the appendix (Online Resource 1), we show that a related property, bisymmetry, can be exploited to directly block-diagonalize \vec{V} into a pair of $n/2$ -dimensional blocks, thereby reducing algorithmic complexity by a factor of four, while avoiding potential issues of numerical rounding errors associated with FFTs and Cholesky factorizations. However, with large enough n , constant factor improvements like these are of little consequence (numerical stability notwithstanding) and even the $\mathcal{O}(n^x)$ and $\mathcal{O}((n^x)^2)$ improvements of Dietrich (1993) and Jun and Stein (2008) may be inadequate to make computations feasible. In that case we suggest that modellers consider an *a priori* assumption of separability in order to exploit computational efficiency in the

well-known algebra of Kronecker products (Van Loan 2000). In Sect. 2.2 we provide some justification for the robustness of these models and in Sect. 3 we extend them to include a parametric model for anisotropy.

2.1 Separable SOS covariograms

The idea of separability is to disentangle the x and y component distances by applying 1D covariograms (say c^x and c^y) to each component separately, before taking their product. Thus if $\vec{s}_i = (x_i, y_i)$, a SOS separable spatial covariogram can be written:

$$c(\vec{s}_i - \vec{s}_j) = \sigma^2 c^x(x_i - x_j) c^y(y_i - y_j), \quad (4)$$

where the marginal variance parameters from c^x and c^y have been combined into σ^2 (as they are not separately identifiable). The resulting covariance matrix decomposes into a Kronecker product $\vec{V} = \vec{V}^x \otimes \vec{V}^y$. Whereas the $(n \times n)$ matrix \vec{V} has one row per sample site, matrices \vec{V}^x and \vec{V}^y ($n^x \times n^x$ and $n^y \times n^y$, respectively) have only one row per grid line; They are themselves covariance matrices, for a pair of 1D processes with covariograms σc^x and σc^y . Indeed this is how separable covariance was originally formalized by Martin (1979), though he presented it in the framework of autoregression.

Martin recognized a number of desirable computational properties in (4), and we will mention some of them before moving to less familiar results. In brief, most of the matrix algebraic computations on \vec{V} that arise in spatial inference and prediction can be applied instead to the lower-dimensional components \vec{V}^x and \vec{V}^y . This includes the inverse and determinant; as well as matrix-vector multiplications; and the SVD,

Cholesky, and spectral decompositions. For example, the negative log-likelihood for an observation of $\vec{Z} \sim \mathcal{N}(\vec{\mu}, \vec{V})$ is proportional to:

$$n^y \log|\vec{V}^x| + n^x \log|\vec{V}^y| + (\vec{Z} - \vec{\mu})^* \text{vec} \left\{ (\vec{V}^y)^{-1} \text{vec}_{n^y}^{-1} (\vec{Z} - \vec{\mu}) (\vec{V}^x)^{-1} \right\}. \quad (5)$$

Comparing with (3), this reduces the arithmetic complexity from $\mathcal{O}\{(n^x n^y)^3\}$ to $\mathcal{O}\{(n^x)^3 + (n^y)^3\}$. Computer memory requirements are also reduced; Only the components \vec{V}^x and \vec{V}^y must be stored in memory, and never the full covariance matrix \vec{V} . Moreover since \vec{V}^x and \vec{V}^y are bisymmetric, they can be block-diagonalized to further speed computations by a factor of four (as described in Online Resource 1).

Formula (5) is often exploited in analyses of spatio-temporal datasets, where a Kronecker product of spatial and temporal covariance matrices is commonly viewed as the simplest baseline model (Genton 2007). It is also well-established in pattern recognition applications of machine learning (Wilson et al. 2014). We are however aware of very few examples in the applied statistics literature of *spatially* separable covariograms.

Statisticians may prefer non-separable isotropic models like the WM (or its geometric anisotropy extension) for reasons of parsimony. However the more computationally attractive separable covariogram seems to mimic these standard models quite well, as we demonstrate next in a simulation study. In the appendix (Online Resource 1), we address computational aspects of marginal distributions, since these are characterized by submatrices of \vec{V} that lack separability.

2.2 Simulation study comparing separable and isotropic covariograms

Given the scarcity of empirical results on the comparative performance of separable spatial covariograms, we sought to evaluate their flexibility in a simulation study. Using simulated spatial datasets with a known ("true") covariogram, we fit the linear model (2) using five different candidate covariogram families; Only one of these five contained the true covariogram (as a special case), while the other four served to illustrate the (more commonplace) scenario of a covariogram misspecification.

The simulations were set up as follows: For each of 300 replicates, we generated a pair of 40×40 lattice datasets from the linear model (2), corresponding to two different types of Gaussian spatial error terms: The first (\vec{Z}_{iso}) was generated from the isotropic WM covariogram ($\nu = 5, \lambda = 2$; Figure 1, top-left); and the second (\vec{Z}_{sep}) from the separable (and highly anisotropic) product of two 1D WMs ($\nu^x = 5, \lambda^x = 1, \nu^y = 5, \lambda^y = 3$). In each replicate we drew regression coefficients $\vec{\beta}$ uniformly at random from $(-1, 1)$, computed the linear predictor $\vec{X}\vec{\beta}$, then added the spatial errors (\vec{Z}_{iso} and \vec{Z}_{sep}) to form a pair of simulated response vectors (\vec{y}_{iso} and \vec{y}_{sep}). The design matrix (\vec{X}), comprised four covariates: two independent standard normal variates (X_1 and X_2); and two MVN variates (X_3 and X_4) drawn from the isotropic WM with strong SAC ($\nu = 6, \lambda = 4$). Thus we produced 300 different *pairs* of simulated datasets, of the form $\left\{ \left\{ \vec{X}, \vec{Z}_{\text{iso}} \right\}, \left\{ \vec{X}, \vec{Z}_{\text{sep}} \right\} \right\}_k$, for $k = 1, \dots, 300$.

We fitted $\hat{\beta}, \hat{\theta}$ to each dataset by maximum likelihood using the model (2), using each of the following five candidate covariance models: OLS (independence); exponential; WM; a separable product of 1D exponentials; and a separable product of 1D WMs. The log-likelihood function was optimized numerically using the Hooke-Jeeves

algorithm (Hooke and Jeeves 1961), as implemented in the R package *dfoptim* (Varadhan et al. 2016). Notice that in each replicate, one of the five candidates covariogram families included, as a special case, the true covariogram for \vec{Z}_{iso} , while the others did not (and similarly, for \vec{Z}_{sep}). This facilitated a comparison of precision and predictive ability among the candidate covariograms. We estimated the root mean-squared prediction error (RMSPE) for each of the models trained on \vec{Z}_{iso} and \vec{Z}_{sep} , by randomly generating a new (test) dataset from the same distribution, selecting half of the points to condition over (uniformly at random), and predicting on the other half.

These results, and the errors in parameter inference, are summarized in Figure 2. For brevity we plot only the regression parameter estimate of the first covariate, X_1 , and omit the OLS errors. Unsurprisingly these OLS errors were extremely large by comparison, making it difficult to discern differences among the boxplots for the other four models. Results for the autocorrelated covariates (X_3 and X_4) showed little difference among the five candidate models.

Although the exponential is nested in the WM family ($\nu = 0.5$), it did relatively poorly in nearly all of our simulations, outperforming only the naive OLS estimates. Nevertheless, the separable product of exponentials produced surprisingly robust MLEs for both types of spatial error. Indeed, when the data were generated from an isotropic WM covariogram, both of the separable covariograms were about as precise in their estimates of β_1 as the correct one (Figure 2, bottom-left). Meanwhile, in datasets generated from separable covariograms (Figure 2, bottom-right), both isotropic models performed noticeably worse than the separable ones.

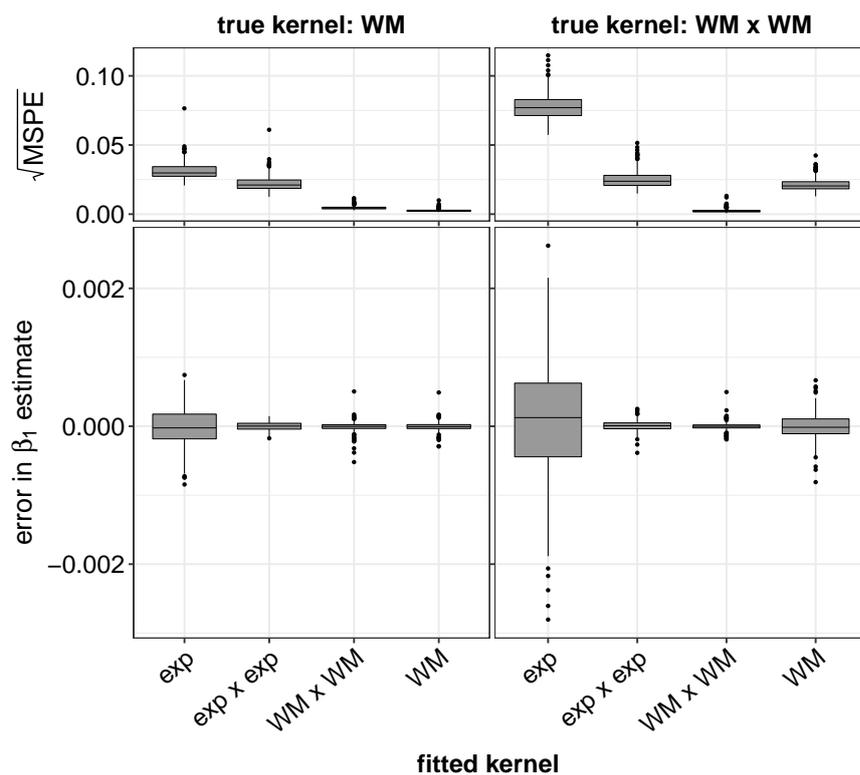


Fig. 2 Inference and prediction on simulated data ($n = 40 \times 40$) from a linear model with either of two covariograms: the WM (left); and a separable product of 1D WMs (right). In each of 300 replicates, we fit MLEs for 5 covariogram families. We plot errors in one of the estimated regression parameters (bottom, corresponding to covariate X_1); and RMSPEs in predictions on a second simulated dataset (top)

The RMSPE results were also favourable to separable covariograms. In datasets from the isotropic WM (Figure 2, top-left), the separable product of 1D WMs did nearly as well as the correct model. The separable product of 1D exponentials performed worse than the correct model, but better than the isotropic exponential. This was surprising given the markedly anisotropic nature of the fitted covariograms from this family (Figure 1, top-right).

These results suggest a remarkable flexibility in separable products of the 1D WM. In our simulations they did well to approximate their isotropic counterparts, but the converse was not true. Moreover, the likelihood maximization problem was far simpler with separability. The reduction in computational complexity is borne out even on this relatively small dataset, where separable models fit around 400X faster than nonseparable ones.

3 Product anisotropic covariograms

In Sect. 1.2 we saw that isotropic covariograms can be generalized to incorporate directionality by a modification of the coordinate system, called geometric anisotropy. What happens if we do the same with separable covariograms?

Separable covariograms are already equipped to handle some degree of range anisotropy, since c^x and c^y can be assigned different range parameters. However, the directionality is constrained to align with the coordinate system, making this approach relatively inflexible. This motivates an extension that we will call the *product anisotropic covariogram* (PAC), in which we compose a separable covariogram with an affine transformation (\vec{A}) of coordinates. Using the notation of Sect. 1.2, we write the general 2D spatial PAC as a function of the transformed coordinate differences:

$$c(\vec{s}_i - \vec{s}_j) = \sigma^2 c^x(\tilde{x}_i - \tilde{x}_j) c^y(\tilde{y}_i - \tilde{y}_j) \text{ where } (\tilde{x}_i - \tilde{x}_j, \tilde{y}_i - \tilde{y}_j)^* = \vec{A}(\vec{s}_i - \vec{s}_j). \quad (6)$$

This formulation was briefly outlined by Allard et al. (2016), who called it *componentwise anisotropy*. Some examples are illustrated in Figure 1 (middle and righthand panes). Note that while geometric anisotropy assumes that \vec{A}^{-1} leads to an isotropic

process, the PAC model assumes it leads to a separable process (which is seldom isotropic). However, if we take the simulation results from Sect. 2.2 as an indication that the (unrotated) exponential PAC adequately approximates the WM, then it follows that its rotated analogue should adequately approximate geometric anisotropy. We propose that the former can serve as a computationally efficient surrogate in situations where the latter is a reasonable model.

This approach allows far more flexibility in covariance structure than does standard geometric anisotropy. For example, c^x and c^y need not be from the same covariogram family; the contours of constant covariance are not restricted to ellipses; and sill anisotropy is naturally accommodated in addition to range anisotropy. These are highly desirable features since, as discussed in Zimmerman (1993), subtly different types of anisotropy can have important consequences for predictions. However, one potential weakness of the PAC deserves mention here: if c^x and c^y have different asymptotic behaviour near zero, then the resulting covariogram is inadmissible as a model for a *fractal* random field (Allard et al. 2016). This result, and its implications for the physical realism of the model, are explained further in the discussion.

The covariance matrix defined by (6) is BTTB (and bisymmetric), by the SOS assumption. However, it will usually *not* be separable. Moreover, since the angle of anisotropy is seldom known *a priori*, we have an additional model parameter to estimate, introducing additional computational complexity. Thus in generalizing the separable covariogram we seem to have lost its main selling point of computational efficiency. However, for MVN data – whose marginal covariances are simply submatrices of \vec{V} – it turns out we can partition the data into subsets whose marginal

covariance *is* separable. This idea can be used to solve both problems, as we show next.

3.1 Special transformations of the coordinate system

The key insight here is to consider transformations \vec{A} for which α , s^x , and s^y satisfy:

$$\alpha = \arctan(\alpha^y/\alpha^x), \quad \text{where } \alpha^x, \alpha^y \in \mathbb{Z}^+, \text{ and } s^x = s^y = \|(\alpha^x, \alpha^y)\|. \quad (7)$$

The scaling $s = \|(\alpha^x, \alpha^y)\|$ ensures that the entries of \vec{A} are integers, so that the transformed coordinates lie in the square lattice \mathbb{Z}^2 . Thus if the original coordinate system \mathcal{G} is an $N^y \times N^x$ regular rectangular grid then we can always choose a subset of locations $\mathcal{G}_{\vec{k}} \subset \mathcal{G}$ that forms a $n^y \times n^x$ regular rectangular grid with respect to the *transformed* coordinates, provided N^x and N^y are large enough. Specifically, in order for the transformed coordinates to not land outside of the original domain \mathcal{G} , it is necessary that:

$$N^x - 1 \geq \alpha^x(n^x - 1) + \alpha^y(n^y - 1) \quad \text{and} \quad N^y - 1 \geq \alpha^y(n^x - 1) + \alpha^x(n^y - 1). \quad (8)$$

For example, Figure 3 (left) shows how $\alpha = \arctan(1/2)$ produces a 3×4 rotated inner subgrid within an 8×9 outer grid. Notice for this choice of $n^y = 3$ and $n^x = 4$, the bounds (8) are tight, and thus the outer boundaries of $\mathcal{G}_{\vec{k}}$ are perfectly inscribed by the outer boundaries of \mathcal{G} (black lines).

It is fairly straightforward (though tedious) to find the indexing vector \vec{k} that pulls this rotated subgrid from \mathcal{G} in column-vectorized order, so we state the general case here without proof and refer the reader to the appendix (Online Resource 1) for a derivation. Let \vec{V} be the covariance matrix for MVN vector \vec{Z} , as in (6), with \vec{A}

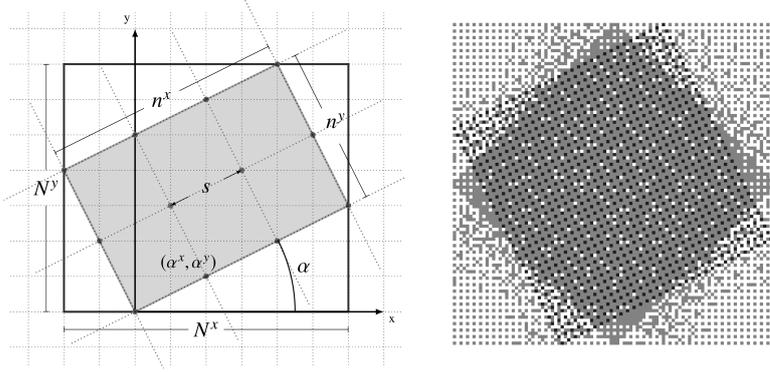


Fig. 3 Special transformations of lattice coordinates. The left diagram illustrates the geometry of a rotation by $\alpha = \arctan(1/2)$ and a scaling of $s = \sqrt{5}$. The shaded rectangle is the rotated $n^y \times n^x$ subgrid $\mathcal{G}_{\vec{k}}$, which lies on the (unrotated) $N^y \times N^x$ grid \mathcal{G} . The right panel shows $\mathcal{G}_{\vec{k}}$ (black cells) on a much larger domain. The gray and black cells together make up the full set of subgrids corresponding to the 12 special angles in (12). Our proposed range anisotropy estimator uses these shaded cells to fit covariograms to each of the 12 angles, and their complement (white cells) to estimate error.

defined as in (7). Assuming the subgrid dimensions $n^y \times n^x$ satisfy (8), we define the indexing vector \vec{k} associated with rotation α by:

$$\left[\text{vec}_{n^y}^{-1}(\vec{k}) \right]_{pq} = 1 + \alpha^y (n^x - 1) + (N^y \alpha^y + \alpha^x)(p - 1) + (N^y \alpha^x - \alpha^y)(q - 1). \quad (9)$$

where the indices $p \in \{1, \dots, n^y\}$ and $q \in \{1, \dots, n^x\}$ refer to the rows and columns (respectively) of the $n^y \times n^x$ raster (matrix) representation of the rotated subgrid $\mathcal{G}_{\vec{k}}$.

Referring to Figure 3 (left), if the unshaded rectangle is \mathcal{G} , then the indexing \vec{k} selects the rotated (3×4) rectangular lattice $\mathcal{G}_{\vec{k}}$ of grey points, whose data vector we denote by $\vec{Z}[\vec{k}]$. For example, the point located on the lower left corner of $\mathcal{G}_{\vec{k}}$ is the third element of $\vec{Z}[\vec{k}]$. This corner point lies on the third row ($p = 3$) of the first column ($q = 1$) of $\mathcal{G}_{\vec{k}}$; Therefore by (9), its index in the full domain \mathcal{G} is given by

$1 + 1(4 - 1) + (8 + 2)(3 - 1) + (16 - 1)(0) = 24$, *ie.* it is the 24th element of the data vector \vec{Z} associated with the full lattice \mathcal{G} .

By construction, \vec{k} generates the column-vectorized ordering of points with respect to the *rotated* coordinate system. The covariance matrix $\vec{V}_{\vec{k}}$ (of $\vec{Z}[\vec{k}]$) is therefore a Kronecker product (of $n^x \times n^x$ and $n^y \times n^y$ matrices), similar to that of \mathcal{G} except with smaller dimensional component matrices, and the spacing of grid lines increased by a factor of s .

Notice that, as a consequence of Pick's theorem, it is possible to partition the entire integer lattice \mathbb{Z}^2 into s^2 disjoint subsets of the form $\vec{A}^T \vec{s} + \vec{\tau}_k$, for $k = 1, \dots, s^2$, where $\vec{s} \in \mathbb{Z}^2$, and $\vec{\tau}_k$ is an integer-valued translation vector (Oliveira et al. 2005). From each subset, one may select a rectangular grid $\mathcal{G}_{\vec{k}^{(k)}}$ of sample sites that lies in \mathcal{G} , by a suitable choice of inner subgrid dimensions n^x and n^y .

For example, with reference to Figure 3 (left), suppose we start by defining a smaller 3×3 subgrid $\mathcal{G}_{\vec{k}^{(1)}}$ with the same lower-left corner point as the 3×4 example $\mathcal{G}_{\vec{k}}$. Its indexing in the full domain \mathcal{G} is $\vec{k}^{(1)}$, as defined by (9) with $n^x = 3$ instead of $n^x = 4$. Now suppose we translate this subgrid upwards by one unit. This yields a different subgrid of points, say $\mathcal{G}_{\vec{k}^{(2)}}$, whose indexing in \mathcal{G} is $\vec{k}^{(2)} = \vec{k}^{(1)} - 1$. Yet another can be found by translating one (or two) units to the right, or equivalently by adding $N^y = 8$ (or $2N^y = 16$) to $\vec{k}^{(1)}$. In total, one can find $s^2 = 5$ (disjoint) subgrids of \mathcal{G} in this manner, each with the same orientation, and (by stationarity) the same separable covariance matrix $\vec{V}_{\vec{k}^{(1)}}$.

3.2 Applications of product anisotropic covariograms

An immediate application for the ideas of Sect. 3.1 is in analyses where a MVN with geometric anisotropy is a suitable model, and the angle α is known or has been estimated (eg. from directional semivariograms). If N is large enough that computational complexity becomes an issue, we suggest using a PAC with a nearby special angle of the form (7). Separability can then be exploited over the subsets $\mathcal{G}_{\vec{k}^{(k)}}$.

Of course when using one of the $\mathcal{G}_{\vec{k}^{(k)}}$ and discarding points from $\mathcal{G} \setminus \mathcal{G}_{\vec{k}^{(k)}}$, we lose efficiency. However much of this efficiency can be recovered by using a composite marginal likelihood function \mathcal{L}_C (Lindsay 1988) that combines information from all s^2 subsets:

$$\mathcal{L}_C \left(\vec{\mu}, \vec{\theta} \mid \left\{ Z_i; \vec{s}_i \in \cup_{k=1}^{s^2} \mathcal{G}_{\vec{k}^{(k)}} \right\} \right) = \prod_{k=1}^{s^2} \mathcal{L} \left(\vec{\mu}, \vec{\theta} \mid \vec{Z}[\vec{k}^{(k)}] \right). \quad (10)$$

This can be evaluated using (5), with the factorization of $\vec{V}_{\vec{k}}$ reused for all s^2 terms.

Typically the direction of range anisotropy α will be unknown and so must be estimated from the data. We propose a simple cross-validation-like method for this estimation problem. The idea is to assemble a set of test angles α_j ($j = 1, \dots, m$), each of the form (7), that define a suite of candidate PAC models. We then fit each of these candidates to its corresponding rotated subgrid $\mathcal{G}_{\vec{k}}(\alpha_j)$ as defined in (9). The model-fitting points $\cup_{j=1}^m \mathcal{G}_{\vec{k}}(\alpha_j)$ do not cover all of \mathcal{G} , so from the unused portion we can select at random a test set, $\mathcal{G}_{\text{pred}}$ to predict over, conditional on the remaining data. Figure 3 (right) illustrates this partition of \mathcal{G} into test and training data.

Having estimated the prediction errors for each α_j , a preferred angle can be chosen by lowest RMSPE. Alternatively one can compute the circular mean of the

α_j 's, inversely weighted by their RMSPE; Letting ω_j denote the j^{th} weight, the circular mean is defined as:

$$\hat{\alpha}_{\bar{\omega}} = (1/2) \arg \left\{ \sum_{j=1}^m \omega_j (\cos(2\alpha_j) + i \sin(2\alpha_j)) \right\}. \quad (11)$$

This simply maps each angle to a vector on the unit circle, scaling lengths according to RMSPE, before finding the angle of the resulting vector sum. Note that because covariance functions are symmetric, we cannot distinguish between α and $\alpha \pm \pi$. Hence we double each $\alpha_j \in [0, \pi)$ before mapping it to the unit circle in equation (11), dividing the final result by 2 to return to $[0, \pi)$.

The more conventional method of investigating the angle of range anisotropy involves studying empirical directional (co)variograms for ad-hoc sets of angles and spatial lags (Sherman 2010). This method is both computationally fast and intuitive, and remains an important part of model selection. However, as an informal graphical diagnostic it suffers from issues of subjective interpretation (Guan et al. 2004). By comparison our method requires very little calibration on the part of the user. It also appears quite robust to model misspecification, as we demonstrate in the next section.

3.3 Simulation study for range anisotropy detection

We examined the performance of the range anisotropy direction estimator of the preceding section using simulated data. As we are mainly interested in situations where the covariance structure is unknown, we considered a situation of model misspecification: simulating data from a WM covariogram with geometric anisotropy ($\nu = 3$, $\lambda = 1$, $s^y/s^x = 2$), but using exponential PACs to fit each candidate angle.

We used the experimental setup of Sect. 2.2: a linear model with two autocorrelated covariates and two independent ones, on a 40×40 spatial domain. We tested 50 values of $\alpha \in [0, \pi)$ for the true spatial error covariogram (chosen uniformly at random). For each of these angles we replicated the simulation with new data 50 times, following the procedure outlined in the previous section to construct two angle estimates per replicate. Our candidate models comprised the 12 angles of the form (7) for which the interpoint distance $s \leq 5$, or:

$$\alpha_j = \{\arctan(\alpha^y/\alpha^x) \mid \alpha^y, \alpha^x \in \{1, \dots, 4\}, \|(\alpha^x, \alpha^y)\| \leq 5\} \cup \{0\}. \quad (12)$$

From this list we constructed the subsets $\mathcal{G}(\alpha_j)$ using equation (9). In the case of $\alpha_j = 0$, we defined $\mathcal{G}(0)$ by simply omitting all even-numbered gridlines from \mathcal{G} . For the nonzero angles, note that the dimensions $N^y \times N^x$ of the full dataset can be decremented by discarding outer rows/columns, and the dimensions $n^y \times n^x$ of the subgrids $\mathcal{G}(\alpha_j)$ adjusted as needed, until the dimensional constraints (8) are met. In a more cautious implementation, (9) can be adjusted to ensure equal sample sizes in each $\mathcal{G}(\alpha_j)$. However we found this had little impact on our simulation results.

Because we used the same covariogram family for c^x and c^y (the 1D exponential), the model for $\mathcal{G}(\alpha_j)$ simultaneously tests both α_j and $\alpha_j + \pi/2$. Our test set (12) therefore encompasses 24 angles, whose positions on the (mod π) compass rose are indicated by the gray bars in Figure 4 (left). For a given fitted covariogram, we distinguished α_j and $\alpha_j + \pi/2$ by taking the larger of the two fitted range parameters ($\hat{\lambda}^x$ or $\hat{\lambda}^y$) to indicate the major axis direction.

In the final step, we selected from the unused data (white cells in Figure 3, right) a subset of size $\sqrt{n^x n^y} = 40$ to set aside as a conditioning set, and predicted the

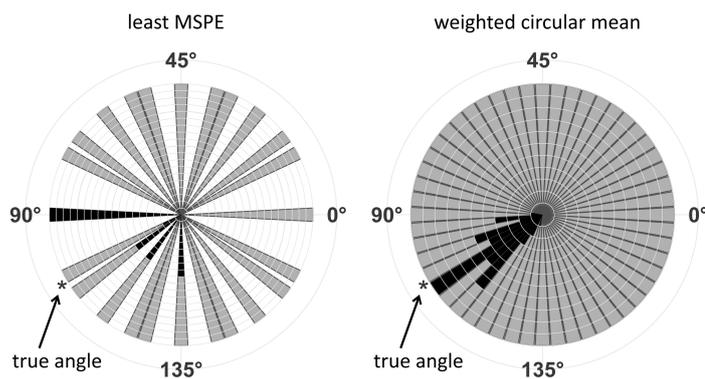


Fig. 4 Rose diagrams summarizing estimates of the angle of range anisotropy over 50 simulation replicates. Datasets were generated from a model with a geometrically anisotropic WM covariogram, oriented at angle $\approx 105^\circ$. Estimates by least RMSPE select the best performing angle from 24 specially chosen candidates (grey bins, left). Weighted circular mean (right) combines information from all 24 candidates to form a continuous estimate

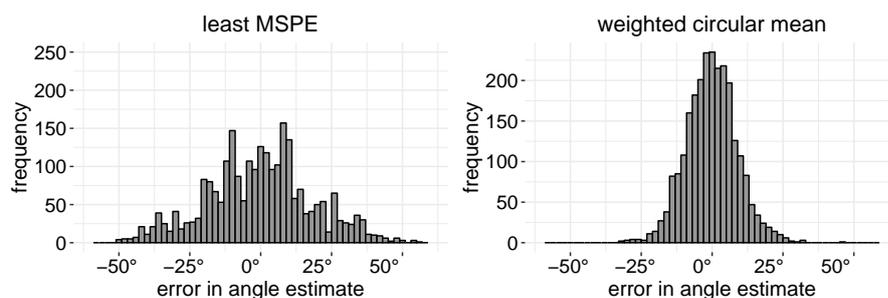


Fig. 5 Errors in two estimators of the angle of range anisotropy for data generated from a model of geometric anisotropy (WM covariogram). Left and right histograms are the pooled results over the same set of simulations, with 50 repetitions for each of 50 randomly chosen true angles

remaining points $\mathcal{G}_{\text{pred}}$ under each of the 12 candidate models. We then determined the angle with least RMSPE, and the weighted circular mean $\hat{\alpha}_{\omega}$, and recorded the error (mod π) in each case. Figure 4 plots the results for one of the 50 tested angles. In Figure 5 we show the pooled errors over all 2500 simulations.

While Figure 5 suggests that both estimators are reasonably unbiased, least-RMSPE tended to favour angles with higher sampling density (specifically $\alpha_j = 0, \pi/4, \pi/2$), leading to a multimodal error distribution. Moreover, the individual histograms often exhibited an interesting (but unwanted) dip near the true angle, as in Figure 4, left.

Weighted circular mean performed far better. Its pooled error distribution appears unimodal, as do the individual histograms, and there was noticeably less variance. Even for the worst performing simulation angle (in terms of error variance), $\hat{\alpha}_{\omega}$ fell within 15° of the true angle in 70% of the repetitions. Over all simulations it was within 45° of the true angle 99.1% of the time, and 77% of the time it was within 15° (Figure 5). This is remarkable given that our estimator derives from only 12 candidate angles. Recall that an empirical rose plot over 12 evenly spaced angles in $[0, \pi)$ would have a detection tolerance of 15° . Intuitively, one would expect the precision of $\hat{\alpha}_{\omega}$ to increase with sampling density – both in terms of the number of candidate angles, and the resolution of the grid – however this remains a topic of future research.

By design, the computing resources needed to estimate α are quite reasonable for large sample sizes. A desktop PC required only around 1-3 seconds with the 40×40 sample size. Moreover the algorithm is completely automated. Unlike the empirical rose plot there is no need to pick an ad-hoc collection of spatial lags or contour levels, nor does the output rely on any kind of subjective visual inspection.

Needless to say, these estimators are only meaningful if it is reasonable to assume a SOS process with range anisotropy. A number of nonparametric tests can detect departures from isotropy (Weller and Hoeting 2016), but we are aware of few such

tests for stationarity. Modellers will more often seek to detrend the data using a carefully constructed mean function. This is not always possible however, and sometimes it is informative to study the nonstationary covariance structure itself, particularly with ecological data.

For example, Sampson and Guttorp (1992) described how nonstationarity can be explored visually using smooth nonlinear deformations of the spatial domain. We will do something similar with $\hat{\alpha}_{\vec{\omega}}$. Suppose that the spatial process over the $N^y \times N^x$ domain \mathcal{G} is nonstationary, but exhibits local stationarity on the scale of a much smaller $n^y \times n^x$ subgrid. We propose using our range anisotropy detection method repeatedly on a sliding window of size $n^y \times n^x$ that moves across \mathcal{G} , estimating at each position the angle $\hat{\alpha}_{\vec{\omega}}$ and the range parameters $\hat{\lambda}^x, \hat{\lambda}^y$. These values define a pair of orthogonal vectors for each location. Plotted together, these depict graphically how the covariance changes through space, much like a biorthogonal grid. This idea is demonstrated in the following case study.

4 Case study: mountain pine beetle damage

We applied our angle detection method to analyse damage to pine forests caused by the mountain pine beetle (MPB). Populations of this tree-killing bark beetle have in recent decades grown to unprecedented levels, leading to an epidemic of mortality in pines throughout its vast native habitat in Western North America. The economical and ecological consequences of the epidemic will be severe and long-lasting (Dhar et al. 2016).

In response, a large body of research has sought to reveal the factors that give rise to MPB outbreaks and allow them to spread, including microclimate, altitude, pine density and proximity to infested stands (Duan et al. 2011). Nevertheless, the large-scale dispersal habits of the MPB are difficult to assay, and remain poorly understood. We applied our methods here to better understand the movements of this forest pest, in the hope that spatially explicit predictions of future outbreaks can be improved.

Monitoring efforts by the Canadian province of British Columbia (BC) are a source of unusually detailed and comprehensive spatial data on MPB activity (Westfall 2005). These data comprise yearly sketch maps of the severity of damage by the beetle (% of pines killed per hectare). We rasterized these maps to produce regular gridded data covering almost the entire treed area of the province, at a one hectare resolution.

In the preliminary analysis we looked at a 300×300 subset, using the linear model (2) with an exponential PAC. To avoid the complications of temporal dependence, we only fitted the spatial process from a single year, $t = 2007$ (around the peak of the pine beetle epidemic in Southern BC). Our response variable $Y_t(\vec{s}_i)$ is the logit-transformed beetle damage measurement for site \vec{s}_i in year t , after adding a small offset ϵ_t to adjust for zeroes (Warton and Hui 2011) (Figure 6, left). The design matrix \vec{X} comprised 29 covariates – mostly climate and weather related – known to influence MPB attack dynamics (The full list can be found in the online supplement.)

To begin we fit the covariogram to the full 300×300 domain by maximum likelihood, given an initial set of OLS estimates for $\vec{\beta}$. We then used GLS, as described below equation (2), to obtain $\hat{\vec{\beta}}_{GLS}$, and refitted the covariogram using the updated regression parameters. Next, to examine nonstationarity we constructed a 12×12

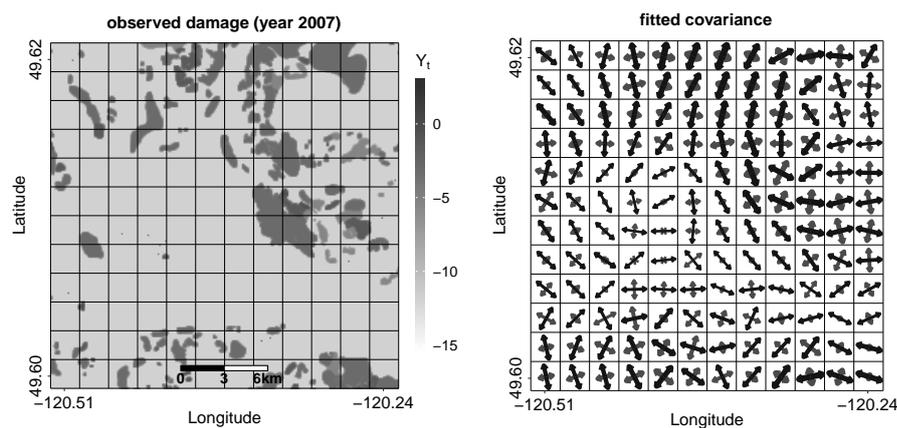


Fig. 6 Beetle damage data from 2007 (left) used to fit a nonstationary covariance structure (right). Arrows indicate the directions $\hat{\alpha}_{\vec{\omega}}$ identified by a blockwise application of the angle-detection technique of Sect. 3.1 to model residuals. Black arrows indicate the major axis direction (largest $\hat{\lambda}$) and arrow thickness indicates the magnitude of the sill ($\hat{\sigma}$)

layout of spatial blocks, each of size 80×80 , with an overlap of 60 cells in each direction. Within each block (and with $\vec{\beta}$ fixed to $\hat{\beta}_{GLS}$) we fitted the linear model (2) using exponential PACs corresponding to each of the angles in (12). We estimated the within-block angle by $\hat{\alpha}_{\vec{\omega}}$ using the method of Sect. 3.3, picking the nearest special angle from the set (12) and using its corresponding fitted covariogram to predict over the unseen data in that block. We then compared RMSPE values of these blockwise predictions with those of the separable model fitted to the full domain.

Lastly, to gauge future predictive ability we used the 2007 (blockwise) models to estimate damage in the year 2008 (Figure 7, left), conditional on a subset of the response data from that year. The conditioning set comprised a random subsample of 20% of the points from the non-overlapping 20×20 subsets at the center of each block (black grid in Figures 6 and 7). We then predicted on the unobserved points to compute

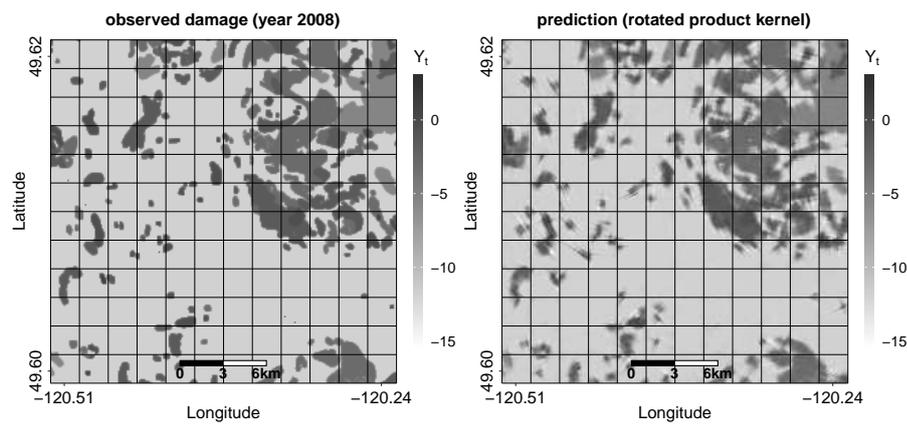


Fig. 7 Observed pine beetle damage in a 30 x 30 km area of Southern BC in year 2008 (left). A random sample of 20% of these datapoints was used to predict the remainder (right), using a linear model with rotated product anisotropic covariograms fitted blockwise to previous-year data

RMSPE. This arrangement ensured that the spatial locations of the predictions lay entirely within the subset used to fit the covariogram of the previous year. Again, we formulated blockwise predictions from both the PAC (within block) and the ordinary separable covariogram (full domain) models, and compared prediction error.

The fitted values of $\hat{\alpha}_{\omega}$ (Figure 6, right) revealed an interesting pattern of range anisotropy varying through space, indicating non-stationarity. The likely physical explanation for these patterns is that they are driven by attributes of the topography and forest structure in our study area that have not been incorporated into the mean model; Clusters of MPB damage will naturally track corridors of pine-rich forestland, and may be skewed in the direction of prevailing winds during the summer dispersal period (Duan et al. 2011). However, large parts of the study area exhibited a consistent directionality and effective range, suggesting that an assumption of local stationarity is reasonable over these areas. This is reflected in the within-block predictions for

the model fitting year, for which the product covariogram (RMSPE = 0.026) outperformed the (full domain) separable covariogram (RMSPE = 0.027). Unsurprisingly, both spatial models improved on the nonspatial OLS predictions (RMSPE = 0.032), highlighting the strong SAC in these data.

If the same spatial covariance structure persists to some degree between years, we can expect to improve 2008 predictions by incorporating information from the blockwise estimated covariograms from the previous year. This was indeed the case, with OLS estimates producing much larger errors in next-year predictions (RMSPE = 0.069) when compared with kernel-based predictions from the separable (RMSPE = 0.044) and PAC (RMSPE = 0.046) models.

Note that overall, the separable covariogram outperformed the blockwise PAC on these next-year test data. This could be a consequence of the decreased sampling resolution induced by (9) (see discussion), or of year-to-year changes in the covariance structure. The PAC, however, performed far better on blocks heavily damaged by the pine beetle, producing the lowest MSPE in 40 of the 64 blocks (63%) for which damage extended over 25% or more of the area. Nevertheless, in spite of the variability in $\hat{\alpha}_{\vec{\omega}}$, both spatial models performed adequately. The predicted sizes and shapes of the beetle damaged areas in 2008 appear quite reasonable given the sparsity of the conditioning set.

Computationally, this analysis was very simple. The entire process - including GLS, anisotropy estimation, covariance model fitting, and prediction, for both models on all 144 blocks - was completed by an ordinary desktop PC in about 7 minutes. By contrast it took closer to an hour to fit a standard model of geometric anisotropy

to a *single* block, even with α known. Moreover, the blockwise approach is easily parallelized, and thus if needed it could be sped up even further by using a cluster of computers. In summary, while the RMSPE performance of the PAC was comparable to the more standard nonseparable alternatives, it allowed model-fitting to complete with remarkable computational speed and a minimal computer memory footprint.

5 Discussion

The computational complexity involved with explicit representations of covariance can be a formidable obstacle. Building upon work by Zimmerman (1989), we have argued for the unconventional solution of using covariograms that are separable in space. These models are not without their drawbacks. Stein (2005) noted issues related to ridges along the coordinate axes that lead to undesirable correlations in linear predictors. These ridges are visible as orthogonal patterns in the plot of the exponential product covariogram in Figure 1 (top-right), for example, as well as in the predicted values in Figure 7. Moreover, separable covariograms are never mathematically isotropic (except in the special, but problematic case of the Gaussian covariogram).

Lastly, if c_x and c_y have different asymptotic behaviour near the origin, the resulting PAC will be incompatible with fractal random field representations of the underlying physical process. The behaviour of a covariogram near the origin, or its *regularity*, is closely tied to (probabilistic) notions of continuity and differentiability in the spatial process (Chilès and Delfiner 2012). A detailed discussion of this relationship is beyond the scope of this paper, so we direct the interested reader to work by Matheron (1962), Rathbun and Stein (2000), and Davies and Hall (1999).

However, in the interest of transparency and completeness, we note that in Allard et al. (2016) it was proven that if the regularity parameters of c_x and c_y are different (eg. two WM covariograms with different shape parameters), the resulting random field cannot be fractal. As fractals are a popular mathematical representation of spatial patterns in ecology (Goodchild and Mark 1987), some modellers may view such PACs as physically unrealistic.

In many applications, however, we believe the foregoing issues have minimal impact on prediction and inference, and that the drawbacks are outweighed by the substantial reduction in computational complexity. In ecology, this complexity often prevents a SAC-corrected analysis in the first place – here it bears repeating the observation of Keitt et al. (2002), that "making any correction is more important than quibbling about which correction to make". Our simulations suggests that separable WM product covariograms are adequate surrogates for more conventional isotropic models. Thus we suggest them as a simpler alternative to the more sophisticated approximations developed in Genton (2007) and Hirano (2014).

On simulating data exhibiting range anisotropy with a known direction, we found that separable covariograms substantially improved MSPE compared with isotropic ones, in spite of a model misspecification. This kind of flexibility will be desirable to modellers with reason to doubt the assumption of isotropy in their data, a common situation in ecology. Note that the separable product of WM covariograms limits to a Gaussian (as $\nu \rightarrow \infty$), so not only is this extremely common model well approximated, it is generalized to include a range of heavier-tailed alternatives. Readers interested in separable approximations are directed to (Wilson et al. 2014), who showed that *any*

SOS covariogram can be approximated to arbitrary precision using sums of separable covariograms.

While the direction of range anisotropy will rarely be known *a priori*, we showed in Sect. 3 how separability leads to a fast estimator of this angle, $\hat{\alpha}_{\vec{v}}$. The simulation results in Sect. 3.3 demonstrated a satisfactory level of accuracy and precision, in spite of model misspecification. We believe this method will nicely complement more standard data exploration techniques such as windrose plots, and formal tests of anisotropy, such as in Guan et al. (2004). It could be used to automate the analysis of a large number of datasets; to identify specific angles and lags to study in more detail using directional (co)variograms; and to provide an objective verification of the conclusions of the graphical analyst.

Note that because the dimensional constraints in (8) preclude large values of α^x, α^y , not all angles of the form (7) can be feasibly tested in a given domain \mathcal{G} . Moreover the interpoint distance s of points in the rotated subgrid increases with both α^x and α^y , making its sampling layout increasingly grainy, and hindering the detection of small-scale covariances over the data in $\mathcal{G}_{\vec{r}}$. For example, in our case study the largest interpoint distance was $s = 5$. At a one-hectare resolution this was acceptable, given that the clusters of beetle damage of greatest concern were much larger than 500 metres. However, depending on the application, modellers may need to upsample their raster data, or shrink the set of candidate angles (12), until the scale of interest is smaller than the largest s . Alternatively one could modify the composite likelihood function in (10) to incorporate information on both the large and small scale, much like the hybrid method proposed by Varin et al. (2011).

By building $\hat{\alpha}_{\omega}$ into a sliding-window estimator, we revealed a remarkably smooth pattern of directional dependence resembling a vector field (Figure 6) in the pine beetle damage dataset. This figure is reminiscent of the graphical depictions of nonstationarity in the kernel-convolution covariance models of Higdon (1998) and Paciorek and Schervish (2006), for example. Future extensions of our approach might look at combining these methodologies, for example, by replacing the local Gaussian convolution kernel in the Higdon (1998) model with the more flexible PAC covariogram.

An ecological explanation for these spatial patterns would be another interesting topic for further research, as it could shine a light on the dispersal habits of the beetle and assist in future predictions. In future work we hope to explore the connection between local estimates $\hat{\alpha}_{\omega}$ and local covariates such as wind direction or connectivity of forest corridors - both likely drivers of directionality in pine beetle damage patterns (Aukema et al. 2006).

Though $\hat{\alpha}_{\omega}$ served as an exploratory tool in our demonstration, it could be useful in other roles. For example, covariance plots like Figure 6 might lead to a graphical means of model selection similar to that described in Das et al. (2002); candidate covariate sets can be compared in terms of fitted covariograms, with the aim of finding a minimal set of explanatory variables that yields a stationary process. Another interesting avenue of research would be to construct a predictor that incorporates information from nonlocal covariance estimates. For example, one could take an average of the blockwise predictions, weighted by distance to the block centroid, to obtain a smoothed prediction surface, similar to the method of Higdon (1998).

We have throughout this paper used the WM covariogram in demonstrations because it is extremely common in spatial statistics. However in future research it will be important to compare against other covariance families to get a more complete picture of the robustness of the product-form alternatives proposed here. Although our empirical results are promising, a healthy skepticism of $\hat{\alpha}_{\omega}$ is wise until its statistical properties are investigated more formally in a theoretical setting.

Until then we would simply argue that separable PACs are a viable means of accounting for SAC, while speeding up analyses by many orders of magnitude on large- n problems. Though our focus here is spatial, we remind the reader that Kronecker product decompositions also apply to separable spatio-temporal covariance matrices. It is our hope that by adopting this trick in the *spatial* domain, practitioners can continue to use the easily interpreted geostatistical model in the rapidly developing world of big data.

References

- Allard D, Senoussi R, Porcu E (2016) Anisotropy Models for Spatial Data. *Math Geosci* 48(3):305–328, DOI 10.1007/s11004-015-9594-x
- Ambikasaran S, Foreman-Mackey D, Greengard L, Hogg DW, O’Neil M (2016) Fast Direct Methods for Gaussian Processes. *IEEE Trans Pattern Anal Mach Intell* 38(2):252–265, DOI 10.1109/TPAMI.2015.2448083, 1403.6015
- Aukema BH, Carroll AL, Zhu J, Raffa KF, Sickley TA, Taylor SW (2006) Landscape level analysis of mountain pine beetle in British Columbia, Canada: Spatiotemporal

- development and spatial synchrony within the present outbreak. *Ecography (Cop)* 29(3):427–441, DOI 10.1111/j.2006.0906-7590.04445.x
- Aukema BH, Carroll AL, Zheng Y, Zhu J, Raffa KF, Dan Moore R, Stahl K, Taylor SW (2008) Movement of outbreak populations of mountain pine beetle: Influences of spatiotemporal patterns and climate. *Ecography (Cop)* 31(3):348–358, DOI 10.1111/j.0906-7590.2007.05453.x
- Banerjee S, Gelfand AE, Finley AO, Sang H (2008) Gaussian predictive process models for large spatial data sets. *J R Stat Soc Ser B Stat Methodol* 70(4):825–848, DOI 10.1111/j.1467-9868.2008.00663.x
- Banerjee S, Carlin BP, Gelfand AE (2014) *Hierarchical Modeling and Analysis for Spatial Data*. CRC Press, Boca Raton, FL, USA, DOI 10.1201/b17115
- Beale CM, Lennon JJ, Yearsley JM, Brewer MJ, Elston DA (2010) Regression analysis of spatial data. *Ecol Lett* 13(2):246–264, DOI 10.1111/j.1461-0248.2009.01422.x
- Chen J, Yang B, Yang L (2019) To BE or not to BE, that is the question. *Nat Biotechnol* 37(5):520–522, DOI 10.1038/s41587-019-0119-x
- Chilès JP, Delfiner P (2012) *Geostatistics: Modeling Spatial Uncertainty: Second Edition*, 2nd edn. John Wiley & Sons, Hoboken, NJ, DOI 10.1002/9781118136188
- Cressie N (1992) *Statistics for Spatial Data*, vol 4. John Wiley & Sons, New York, NY, USA, DOI 10.1111/j.1365-3121.1992.tb00605.x
- Cressie N, Johannesson G (2008) Fixed rank kriging for very large spatial data sets. *J R Stat Soc Ser B Stat Methodol* 70(1):209–226, DOI 10.1111/j.1467-9868.2007.00633.x

- Dale MR, Fortin MJ (2014) *Spatial analysis: A guide for ecologists*, second edition. Cambridge University Press, Cambridge, UK, DOI 10.1017/CBO9780511978913
- Das A, Lele SR, Glass GE, Shields T, Patz J (2002) Modelling a discrete spatial response using generalized linear mixed models: Application to Lyme disease vectors. *Int J Geogr Inf Sci* 16(2):151–166, DOI 10.1080/13658810110099134
- Davies S, Hall P (1999) Fractal analysis of surface roughness by using spatial data. *J R Stat Soc Ser B Stat Methodol* 61(1):3–37, DOI 10.1111/1467-9868.00160
- Dhar A, Parrott L, Heckbert S (2016) Consequences of mountain pine beetle outbreak on forest ecosystem services in western Canada. *Can J For Res* 46(8):987–999, DOI 10.1139/cjfr-2016-0137
- 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
- Dormann CF (2009) Response to comment on "methods to account for spatial autocorrelation in the analysis of species distributional data: A review". *Ecography (Cop)* 32(3):379–381, DOI 10.1111/j.1600-0587.2009.05907.x
- Duan JJ, Taylor PB, Fuester RW (2011) Biology and Life History of *Balcha indica*, an Ectoparasitoid Attacking the Emerald Ash Borer, *Agrilus planipennis*, in North America. In: Safranyik L, Wilson B (eds) *J Insect Sci*, vol 11, Canadian Forest Service, Victoria, Canada, pp 1–9, DOI 10.1673/031.011.12701
- Genton MG (2007) Separable approximations of space-time covariance matrices. *Environmetrics* 18(7):681–695, DOI 10.1002/env.854

- Goodchild MF, Mark DM (1987) The Fractal Nature of Geographic Phenomena. *Ann Assoc Am Geogr* 77(2):265–278, DOI 10.1111/j.1467-8306.1987.tb00158.x
- Guan Y, Sherman M, Calvin JA (2004) A nonparametric test for spatial isotropy using subsampling. *J Am Stat Assoc* 99(467):810–821, DOI 10.1198/016214504000001150
- Guillot G, Schilling RL, Porcu E, Bevilacqua M (2014) Validity of covariance models for the analysis of geographical variation. *Methods Ecol Evol* 5(4):329–335, DOI 10.1111/2041-210X.12167, 1311.4136
- Guttorp P, Gneiting T (2006) Studies in the history of probability and statistics XLIX on the Matérn correlation family. *Biometrika* 93(4):989–995, DOI 10.1093/biomet/93.4.989
- Hawkins BA (2012) Eight (and a half) deadly sins of spatial analysis. *J Biogeogr* 39(1):1–9, DOI 10.1111/j.1365-2699.2011.02637.x
- Heagerty PJ, Lele SR (1998) A composite likelihood approach to binary spatial data. *J Am Stat Assoc* 93(443):1099–1111, DOI 10.1080/01621459.1998.10473771
- Higdon D (1998) A process-convolution approach to modelling temperatures in the North Atlantic Ocean. *Environ Ecol Stat* 5(2):173–190, DOI 10.1023/A:1009666805688
- Hirano T (2014) Pseudo Best Estimator by a Separable Approximation of Spatial Covariance Structures. *J Japan Stat Soc* 44(1):43–71, DOI 10.14490/jjss.44.43
- Hooke R, Jeeves TA (1961) “Direct Search” Solution of Numerical and Statistical Problems. *J ACM* 8(2):212–229, DOI 10.1145/321062.321069

- Inchausti P (1998) *The Ecological Detective: Confronting Models with Data*. Ray Hilborn, Marc Mangel, vol 73. Princeton University Press, Princeton, NJ, USA, DOI 10.1086/420265
- Jun M, Stein ML (2008) Nonstationary covariance models for global data. *Ann Appl Stat* 2(4):1271–1289, DOI 10.1214/08-AOAS183
- Kaufman CG, Schervish MJ, Nychka DW (2008) Covariance tapering for likelihood-based estimation in large spatial data sets. *J Am Stat Assoc* 103(484):1545–1555, DOI 10.1198/016214508000000959
- Keitt TH, Bjørnstad ON, Dixon PM, Citron-Pousty S (2002) Accounting for spatial pattern when modeling organism-environment interactions. *Ecography (Cop)* 25(5):616–625, DOI 10.1034/j.1600-0587.2002.250509.x
- Klutsch JG, Negrón JF, Costello SL, Rhoades CC, West DR, Popp J, Caissie R (2009) Stand characteristics and downed woody debris accumulations associated with a mountain pine beetle (*Dendroctonus ponderosae* Hopkins) outbreak in Colorado. *For Ecol Manage* 258(5):641–649, DOI 10.1016/j.foreco.2009.04.034
- Legendre P (1993) Spatial autocorrelation: trouble or new paradigm? *Ecology* 74(6):1659–1673, DOI 10.2307/1939924
- Lindsay BG (1988) Composite likelihood methods. *Contemp Math* 80(1):221–239, DOI 10.1090/conm/080/999014
- Martin RJ (1979) A subclass of lattice processes applied to a problem in planar sampling. *Biometrika* 66(2):209–217, DOI 10.1093/biomet/66.2.209
- Matheron G (1962) *Traité de Géostatistique Appliquée*, vol 1. Editions Technip-Paris

- Nakagawa S, Freckleton RP (2008) Missing inaction: the dangers of ignoring missing data. *Trends Ecol Evol* 23(11):592–596, DOI 10.1016/j.tree.2008.06.014
- Oliveira FMM, Dantas RT, Furtado DA, Nascimento JWB, Medeiros AN (2005) Parâmetros de conforto térmico e fisiológico de ovinos Santa Inês, sob diferentes sistemas de condicionamento. *Rev Bras Eng Agrícola e Ambient* 9(4):631–635, DOI 10.1590/s1415-43662005000400029
- Paciorek CJ, Schervish MJ (2006) Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics* 17(5):483–506, DOI 10.1002/env.785
- Rathbun SL, Stein ML (2000) *Interpolation of Spatial Data: Some Theory for Kriging*, vol 95. Springer-Verlag, New York, NY, DOI 10.2307/2669494
- Roberts S, Osborne M, Ebden M, Reece S, Gibson N, Aigrain S (2013) Gaussian processes for time-series modelling. *Phil Trans R Soc A* 371(1984):20110550
- Robertson C, Farmer CJ, Nelson TA, MacKenzie IK, Wulder MA, White JC (2009) Determination of the compositional change (1999–2006) in the pine forests of British Columbia due to mountain pine beetle infestation. *Environ Monit Assess* 158(1–4):593–608, DOI 10.1007/s10661-008-0607-9
- Sampson PD, Guttorp P (1992) Nonparametric estimation of nonstationary spatial covariance structure. *J Am Stat Assoc* 87(417):108–119, DOI 10.1080/01621459.1992.10475181
- Sherman M (2010) *Spatial Statistics and Spatio-Temporal Data: Covariance Functions and Directional Properties*. John Wiley & Sons, West Sussex, UK, DOI 10.1002/9780470974391

- Simpson D, Lindgren F, Rue H (2012) In order to make spatial statistics computationally feasible, we need to forget about the covariance function. *Environmetrics* 23(1):65–74, DOI 10.1002/env.1137
- Stein ML (2005) Space-time covariance functions. *J Am Stat Assoc* 100(469):310–321, DOI 10.1198/016214504000000854
- Stein ML, Chi Z, Welty LJ (2004) Approximating likelihoods for large spatial data sets. *J R Stat Soc Ser B Stat Methodol* 66(2):275–296, DOI 10.1046/j.1369-7412.2003.05512.x
- Van Loan CF (2000) The ubiquitous Kronecker product. *J Comput Appl Math* 123(1):85–100
- Varadhan R, Borchers HW, Varadhan MR (2016) Package ‘dfoptim’
- Varin C, Reid N, Firth D (2011) An overview of composite likelihood methods. *Stat Sin* 21(1):5–42
- Vecchia AV (1988) Estimation and Model Identification for Continuous Spatial Processes. *J R Stat Soc Ser B* 50(2):297–312, DOI 10.1111/j.2517-6161.1988.tb01729.x
- 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
- Wall MM (2004) A close look at the spatial structure implied by the CAR and SAR models. *J Stat Plan Inference* 121(2):311–324, DOI 10.1016/S0378-3758(03)00111-3

- Warton DI, Hui FK (2011) The arcsine is asinine: The analysis of proportions in ecology. *Ecology* 92(1):3–10, DOI 10.1890/10-0340.1
- Weller ZD, Hoeting JA (2016) A review of nonparametric hypothesis tests of isotropy properties in spatial data. *Stat Sci* 31(3):305–324, DOI 10.1214/16-STS547, 1508.05973
- Westfall J (2005) 2005 Summary of Forest Health Conditions in British Columbia. Tech. rep., British Columbia Ministry of Forests and Range, Forest Practices Branch, Victoria
- Wilson AG, Gilboa E, Nehorai A, Cunningham JP (2014) Fast kernel learning for multidimensional pattern extrapolation. In: *Adv Neural Inf Process Syst*, vol 4, pp 3626–3634
- Wulder MA, Dymond CC, White JC, Leckie DG, Carroll AL (2006) Surveying mountain pine beetle damage of forests: A review of remote sensing opportunities. *For Ecol Manage* 221(1-3):27–41, DOI 10.1016/j.foreco.2005.09.021
- 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
- Zimmerman DL (1993) Another look at anisotropy in geostatistics. *Math Geol* 25(4):453–470, DOI 10.1007/BF00894779