Kriging Models for Linear Networks and non-Euclidean Distances: Cautions and Solutions

Jay M. Ver Hoef

Marine Mammal Laboratory, NOAA Fisheries Alaska Fisheries Science Center
7600 Sand Point Way NE, Seattle, WA 98115
tel: (907) 347-5552 E-mail: jay.verhoef@noaa.gov

January 20, 2018

Summary

1. There are now many examples where ecological researchers used non-Euclidean distance metrics in geostatistical models that were designed for Euclidean distance, such as those used for kriging. This can lead to problems where predictions have negative variance estimates. Technically, this occurs because the spatial covariance matrix, which depends on the geostatistical models, is not guaranteed to be positive definite when non-Euclidean distance metrics are used. These are not permissible models, and should be avoided.

2. I give a quick review of kriging and illustrate the problem with several simulated examples, including locations on a circle, locations on a linear dichotomous network (such as might be used for streams), and locations on a linear trail or road network. I re-examine the linear-network distance models from Ladle et al. (2017b) and show that they are not guaranteed to have a positive-definite covariance matrix.

3. I introduce the reduced-rank method, also called a predictive-process model, for creating valid spatial covariance matrices with non-Euclidean distance metrics. It has an additional advantage of fast computation for large data sets.

4. I reanalyzed the data of Ladle et al. (2017b), showing that fitted models that used linear-network distance in geostatistical models, both with and without a nugget effect, had negative variances, poor predictive performance compared reduced-rank methods, and had improper coverage for the prediction intervals. The reduced-rank approach using linear-network distances provided a class of permissible models that had better predictive performance and proper coverage for the prediction intervals, and could be combined with Euclidean-distance models to provide the best overall predictive performance.

KEY WORDS: spatial statistics, geostatistics, prediction, reduced-rank methods, predictive process models

INTRODUCTION

There are now several examples in the ecological literature where, for spatial prediction like kriging, non-Euclidean distances were used in autocorrelation models developed under a Euclidean distance assumption. This leads to a problem where prediction variances may be negative, and generally leads to unreliable standard errors for prediction. My objective is to help ecologists understand the problem and avoid this mistake. I introduce a class of models that are easy to construct, based on linear mixed models, that perform well and guarantee that prediction standard errors will be positive.

A Quick Review of Kriging

Kriging is a method for spatial interpolation, beginning as a discipline of atmospheric sciences in Russia, of geostatistics in France, and appearing in English in the early 1960's (Gandin, 1963; Matheron, 1963; Cressie, 1990). Kriging is attractive because it produces both predictions and prediction standard errors, providing uncertainty estimates for the predictions. Predictions and their standard errors are obtained after first estimating parameters of the kriging model. The ordinary kriging model, which we will feature here, is,

$$Y_i = \mu + Z_i + \varepsilon_i, \qquad \text{eqn 1}$$

where Y_i is a spatial random variable at location i, $i = 1, 2, \ldots, n$, with constant mean μ (the fixed effect), a zero-mean spatially-autocorrelated error Z_i, and independent random error ε_i. For the set $\{Z_i; i = 1, \ldots, n\}$, the spatial distance among locations is used to model autocorrelation among the random errors. Spatial autocorrelation is the tendency for spatial variables to co-vary, either in a similar fashion, or opposite from each other. The most commonly observed type of

spatial autocorrelation manifests as higher positive correlation among variables at sites closer together than among those at sites farther apart. These tendencies are captured in autocorrelation and covariance matrices.

Let $\mathbf{R}$ be an autocorrelation matrix among spatial locations. All of the diagonal elements of $\mathbf{R}$ are ones. The off-diagonal element in the ith row and jth column of $\mathbf{R}$ is the correlation, which lies between -1 and 1, between variables at site i and site j. Then a covariance matrix $\mathbf{C} = \sigma_p^2\mathbf{R}$ is just a scaled autocorrelation matrix that includes an overall variance, σ_p^2. In constructing kriging models, practitioners often include a "nugget" effect, which is an independent (uncorrelated) random effect, ε_i, in eqn 1 with variance σ_0^2. The nugget effect is often ascribed to measurement error, or microscale variation, at a scale finer than the closest measurements in the data set. Constructing a full covariance matrix for a kriging model generally yields

$$\mathbf{\Sigma} = \mathbf{C} + \sigma_0^2\mathbf{I} = \sigma_p^2\mathbf{R} + \sigma_0^2\mathbf{I}, \qquad \text{eqn 2}$$

where $\sigma_p^2 \geqslant 0$ is called the partial sill, $\sigma_0^2 \geqslant 0$ is the nugget effect, and $\mathbf{I}$ is the identity matrix (a diagonal matrix of all ones). The total variance is $\sigma_p^2 + \sigma_0^2$. The off-diagonal elements of $\mathbf{R}$ are obtained from models that generally decrease as distance increases, with a few that also oscillate. Several autocorrelation models (Chiles & Delfiner, 1999, p. 80–93), based on Euclidean distance,

$d_{i,j}$, between sites i and j, are

$$\begin{aligned}
\rho_e(d_{i,j}) &= \exp(-d_{i,j}/\alpha), \\
\rho_s(d_{i,j}) &= [1 - 1.5(d_{i,j}/\alpha) + 0.5(d_{i,j}/\alpha)^3]\mathcal{I}(d_{i,j} < \alpha), \\
\rho_g(d_{i,j}) &= \exp(-(d_{i,j}/\alpha)^2), \\
\rho_c(d_{i,j}) &= 1/(1 + (d_{i,j}/\alpha)^2), \\
\rho_h(d_{i,j}) &= (\alpha/d_{i,j})\sin(d_{i,j}/\alpha)\mathcal{I}(d_{i,j} > 0) + \mathcal{I}(d_{i,j} = 0),
\end{aligned} \qquad \text{eqn 3}$$

where distances are scaled by $\alpha \geqslant 0$, called the range parameter. $\mathcal{I}(a)$ is an indicator function, equal to one if the argument a is true, otherwise it is zero.

Examples of the autocorrelation models in eqn 3, scaled with a partial sill, $\sigma_p^2 = 2$, and a nugget effect, $\sigma_0^2 = 1$, are shown in Figure 1a. The exponential model, $\rho_e(d_{i,j})$, is commonly used, and a special case of the Matern model that approaches zero autocorrelation asymptotically (Figure 1c). The spherical model, $\rho_s(d_{i,j})$, also is common, attaining exactly zero autocorrelation at α (Figure 1d). Both the exponential and spherical models decrease rapidly near the origin, for short distances, whereas the Gaussian model, $\rho_g(d_{i,j})$, decreases more slowly near the origin. The Gaussian model occurs as a limiting case for the smoothness parameter of the Matern model, and creates very smooth spatial surfaces. The Cauchy model, $\rho_c(d_{i,j})$, is similar to the Gaussian, but approaches zero autocorrelation very slowly. Finally, the hole effect model, $\rho_h(d_{i,j})$, allows for negative autocorrelation in a dampened oscillating manner. These models highlight different features of autocorrelation models, and they will be used throughout this paper. Many more models are given in Chiles & Delfiner (1999, p. 80–93). Autocorrelation is generally controlled by α, which must estimated from real data. However, it is useful to vary α through simulated data, and even for real distance data, to understand its effect on covariance models, which I do in Figures 1c,d, and also in Figures 2 and 3.

Kriging is often expressed in terms of semivariograms rather than autocorrelation models. Semivariograms model the variance of the *difference* among variables. If Y_i and Y_j are random variables at spatial locations i and j, respectively, a semivariogram is defined as $\gamma(d_{i,j}) \equiv \mathrm{E}(Y_i - Y_j)^2/2$, where E is expectation. All of the models in eqn 3 can be written as semivariograms,

$$\gamma_m(d_{i,j}) = \sigma_p^2(1 - \rho_m(d_{i,j})), \qquad \text{eqn 4}$$

where $m = e, s, g, c$, or h for exponential, spherical, Gaussian, Cauchy, or hole effect, respectively. Figure 1b shows semivariograms that are equivalent to the models in Figure 1a. A matrix of semivariogram values among spatial locations can be written in terms of eqn 2,

$$\mathbf{\Gamma} = (\sigma_0^2 + \sigma_p^2)\mathbf{I} - \mathbf{\Sigma}.$$

Autocorrelation needs to be estimated from data. Empirical semivariograms have been used since the origins of kriging. First, all pairwise distances are binned into distance classes, $\mathcal{D}_k = [h_{k-1}, h_k)$, where $0 \leqslant h_0 < h_1$ and $h_{k-1} < h_k$ for $k = 1, 2, \ldots, K$, that partition the real line into mutually exclusive and exhaustive segments that cover all distances in the data set. Then the empirical semivariogram is,

$$\hat{\gamma}(h_k) = \frac{1}{2N(\mathcal{D}_k)} \sum_{d_{i,j} \in \mathcal{D}_k} (y_i - y_j)^2, \qquad \text{eqn 5}$$

for all possible pairs of i and j, and $k = 1, \ldots, K$, where $y_1, ..., y_n$ are the observed data, h_k is a representative distance (often the average or midrange) for a distance bin $\mathcal{D}_k$, and $N(\mathcal{D}_k)$ is the number of distinct pairs in $\mathcal{D}_k$. Empirical semivariograms have desirable estimation properties (it is an unbiased estimator, Cressie, 1993, p. 71) because, substituting eqn 1 into the semivariogram definition, μ cancels, obviating the need to estimate it. To estimate autocorrelation, one of the

models in eqn 3, in semivariogram form, eqn 4, can be fit to $\hat{\gamma}(h_k)$ as a function of h_k, often using weighted least squares (WLS) or a modification that puts increased weight near the origin (CWLS) (Cressie, 1985). This concept is generalized by restricted maximum likelihood (REML, Patterson & Thompson, 1971, 1974), which can be used for autocorrelation in regression models with several covariates and regression coefficients (for REML applied to spatial models, see, e.g., Cressie, 1993, p. 93). In addition, using REML eliminates the arbitrary binning of distances for semivariogram estimation. Although REML was originally derived assuming normality, REML can be viewed as unbiased estimating equations (Heyde, 1994; Cressie & Lahiri, 1996), so normality is not required to estimate covariance parameters. Later, I will use WLS, CWLS, and REML for estimation, and full details are given in Supporting Information. No matter how the parameters are estimated, I focus on covariance matrices $\mathbf{\Sigma}$ (eqn 2), rather than semivariogram matrices, because $\mathbf{\Sigma}$ is more readily understood in the broader context of statistical models.

After covariance parameters are estimated from the data, kriging can produce spatial predictions (interpolations) at any locations where data were not collected. Kriging provides best-linear-unbiased predictions (BLUP) in the sense of minimizing the expected squared error between linear combinations of the data as predictors, and the predictand, subject to unbiasedness (on average). The ordinary kriging predictor, in terms of the covariance matrix (Schabenberger & Gotway, 2005, p.33), is $\hat{Y}_{n+\ell} = \boldsymbol{\lambda}'\boldsymbol{Y}$, where

$$\boldsymbol{\lambda}' = \left(\mathbf{c} + \mathbf{1}\frac{1 - \mathbf{1}'\mathbf{\Sigma}^{-1}\mathbf{c}}{\mathbf{1}'\mathbf{\Sigma}^{-1}\mathbf{1}}\right)' \mathbf{\Sigma}^{-1}, \qquad \text{eqn 6}$$

for M predictions with locations indexed by $n+\ell$, $\ell = 1, 2, \ldots, M$. Here, $\mathbf{1}$ is a vector of ones, and $\mathbf{c}$ has, as its ith element, $\sigma_p^2 \rho_m(d_{i,n+\ell})$, where m is the same model (one of those in eqn 3) that was

used in $\mathbf{\Sigma}$. The prediction variance (the expected squared error that was minimized) is given by

$$\text{var}(\hat{Y}_{n+\ell} - Y_{n+\ell}) = \text{E}(\hat{Y}_{n+\ell} - Y_{n+\ell})^2 = (\sigma_p^2 + \sigma_0^2) - \mathbf{c}'\mathbf{\Sigma}^{-1}\mathbf{c} + \frac{(1 - \mathbf{1}'\mathbf{\Sigma}^{-1}\mathbf{c})^2}{\mathbf{1}'\mathbf{\Sigma}^{-1}\mathbf{1}}, \qquad \text{eqn 7}$$

where the first equality occurs due to the unbiasedness condition ($\boldsymbol{\lambda}'\mathbf{1} = 1$) imposed by the kriging method (e.g., Cressie, 1993, p. 120-121).

The Problem

One of the properties shared by all models in eqn 3 is that, when $d_{i,j}$ is Euclidean distance, the covariance matrix in eqn 2 is guaranteed to be positive definite for all possible spatial configurations of points (in 3 dimensions or less) and all possible parameter values: $\sigma_p^2 \geqslant 0, \sigma_0^2 \geqslant 0$, and $\alpha \geqslant 0$ (one of σ_p^2 or σ_0^2 must be greater than zero). It is important for $\mathbf{\Sigma}$ to be positive definite because many estimators and predictors in statistics are linear functions of the data, the kriging predictor being one of them. That is, let $\boldsymbol{\omega}$ be a nonnull vector of weights and $\boldsymbol{Y}$ be a vector of random variables with covariance matrix $\mathbf{\Sigma}$. Then an estimator or predictor $\hat{T} = \boldsymbol{\omega}'\boldsymbol{Y}$ will have variance

$$\text{var}(\hat{T}) = \boldsymbol{\omega}'\mathbf{\Sigma}\boldsymbol{\omega}, \qquad \text{eqn 8}$$

which is guaranteed to be positive only if $\mathbf{\Sigma}$ is positive definite (Guillot et al., 2014). Requiring $\mathbf{\Sigma}$ to be positive definite is the matrix analog of requiring a variance parameter to be positive.

The prediction variance (eqn 7) involves the variance of a difference between a linear combination of data at observed locations, with weights given by eqn 6, and the prediction location, so $\boldsymbol{\omega}' = (\boldsymbol{\lambda}', -1)$. We can add the covariances between prediction location and data locations (denoted $\mathbf{c}$ in eqn 6) to $\mathbf{\Sigma}$, call it $\mathbf{\Sigma}^*$, and eqn 8 must hold for $\mathbf{\Sigma}^*$ as well. That is, more generally, let $\mathbf{\Sigma}_{o,o}$ be the covariance matrix among the observed locations, $\mathbf{\Sigma}_{o,p}$ be the covariance

matrix between the observed and prediction locations, and $\mathbf{\Sigma}_{p,p}$ be the covariance matrix among the prediction locations. Then

$$\mathbf{\Sigma}^* = \begin{pmatrix} \mathbf{\Sigma}_{o,o} & \mathbf{\Sigma}_{o,p} \\ \mathbf{\Sigma}'_{o,p} & \mathbf{\Sigma}_{p,p} \end{pmatrix} \quad \text{eqn 9}$$

must be positive definite when making predictions at unobserved locations. A simple example is given in the Supplementary Material. For another example, Guillot et al. (2014) demonstrate that the triangle model (not given in eqn 3), which is only valid in one dimension, yields negative prediction variances when used with Euclidean distances based on locations in two-dimensions.

It is also worth noting that if any square submatrix of $\mathbf{\Sigma}^*$ (eqn 9) (formed by removing full columns and rows with corresponding indexes) is not positive definite, then neither is the larger matrix. The implications are that, if the observed data have a covariance matrix that is not positive definite, then $\mathbf{\Sigma}^*$ will not be positive definite. However, even if the observed data (eqn 9) have a covariance matrix that is positive definite, there is no guarantee that the larger matrix, $\mathbf{\Sigma}^*$, will be positive definite without a proper model to ensure it.

The simplest way to check whether a matrix is positive definite is to check the eigenvalues of that matrix. A covariance matrix $\mathbf{\Sigma}$ should be composed of real values, and it should be symmetric. Then

$$\mathbf{\Sigma} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}' \quad \text{eqn 10}$$

is called the spectral decomposition of $\mathbf{\Sigma}$, where each column of $\mathbf{Q}$ contains an eigenvector, and the corresponding eigenvalue is contained in $\mathbf{\Lambda}$, which is a diagonal matrix. Substituting eqn 10 into eqn 8 gives

$$\text{var}(\hat{T}) = \mathbf{v}'\mathbf{\Lambda}\mathbf{v} = \sum_{i=1}^{n} v_i^2 \lambda_i$$

where $\mathbf{v} = \mathbf{Q}'\boldsymbol{\omega}$. Because $v_i^2 \geqslant 0$, $\text{var}(\hat{T})$ is guaranteed to be positive as long as all λ_i are greater

than zero and at least one v_i^2 is greater than zero. So, if the smallest eigenvalue of $\mathbf{\Sigma}$ is greater than zero, then $\mathbf{\Sigma}$ is positive definite.

Now consider using the models in eqn 3 for cases where $d_{i,j}$ is non-Euclidean. For example, let 11 spatial locations occur at equal distances on a circle (Figure 2a). Let distance be defined as the shortest path distance, so that two adjacent points have distance $2\pi/11$, and the maximum distance between any two points is $10\pi/11$. The 11×11 distance matrix was used with autocorrelation models in eqn 3, and the minimum eigenvalue is plotted as a function of α in Figure 2b. Notice that as the range parameter α increases, the hole-effect, Gaussian, and Cauchy models have a minimum eigenvalue that is less than zero, so for these values of α, the matrix is not positive definite, and cannot be a covariance matrix. This example illustrates another problem because although the exponential model and spherical model are valid models for all range values, this is true only if 11 points are equidistant apart. There is no guarantee that the exponential and spherical model will provide positive-definite covariance matrices for other sample sizes and other spatial configurations. Later, I will discuss more general approaches for developing models for all spatial configurations and all values of the range parameter.

Another example is provided by the spatial locations at the nodes of a dichotomous network (Figure 2c). The distance between each location and the nearest node is exactly one, and there are $2^7 - 1$ locations. Again, let distance be defined as the shortest path between any two locations, so the maximum distance between two terminal locations is $2 \times 6 = 12$. Using the 127×127 distance matrix with the autocorrelation models in eqn 3 for various α values showed that all models yielded minimum eigenvalues below zero except the exponential model (Figure 2d). The hole effect model illustrates how erratic the positive-definite condition can be, where small changes in α cause wild swings on whether the covariance matrix is positive definite. An argument on why the exponential model is always positive definite for the dichotomous

network situation is given by Ver Hoef & Peterson (2010).

Finally, consider the 25 locations in Figure 2e. This is representative of a road or trail system on a perfectly regular grid. Again, consider the shortest path distance between any two points. First, consider the situation where sites are only connected by the solid lines. In that case, sites one and two are not connected directly, but rather the distance between them is 3 (through sites 6 and 7). Using the 25×25 distance matrix with the autocorrelation models in eqn 3 for various α values shows that none of the models are positive definite for all α (Figure 2f). A variation occurs if we let the sites with dotted lines be connected, as well as those with solid lines. In this case, the exponential model remains positive definite for all values of α, and an explanation is provided by Curriero (2006).

In Figure 2 I illustrate that, in a variety of situations, models that guarantee positive-definite covariance matrices for any spatial configuration, and any range value $\alpha > 0$, when using Euclidean distance, no longer guarantee positive-definite matrices when using linear-network distances. Similarly, one might wonder why we do not use empirical covariances in $\mathbf{\Sigma}$. That is, let the i, j entry in $\mathbf{\Sigma}$ be $(y_i - \hat{\mu})(y_j - \hat{\mu})$, where $\hat{\mu}$ is the average of all y_i. Again, there is no guarantee that $\mathbf{\Sigma}$ will be positive definite. If it is not, then what is the analyst to do? Geostatistics has a long tradition of only considering models that guarantee positive-definite matrices (Journel & Huijbregts, 1978, p. 161). For example, Webster & Oliver (2007, p. 80) call them "authorized" models, while Goovaerts (1997, p. 87) calls them "permissible" models. All of the models in eqn 3 are permissible for Euclidean distance in three dimensions or less, but they are clearly not generally permissible for linear-network distances.

Literature Review

There are now many examples where autocovariance models, such as those in eqn 3, have been used incorrectly with non-Euclidean distances, and they have been roundly criticized (Curriero, 2006). For example, for streams, impermissible models have been used by Cressie & Majure (1997) and Gardner et al. (2003), who substituted in-stream distance for Euclidean distance, and in fact this same idea was inappropriately recommended in Okabe & Sugihara (2012). Alternatively, permissible models that guarantee positive-definite covariance matrices were developed (based on spatial moving averages, a spatially continuous analog of moving-average models in time series) by Ver Hoef et al. (2006), Cressie et al. (2006) and Ver Hoef & Peterson (2010).

For roads and trails, impermissible models have been used by Shiode & Shiode (2011), Selby & Kockelman (2013) and Ladle et al. (2017b), who substitute network-based distance for Euclidean distance. However, the exponential is a permissible model for a perfect grid using Manhattan distance (as described for Figure 2e); see Curriero (2006). I provide a more general approach based on reduced-rank radial-basis functions below.

In estuaries, shortest-path distances were incorrectly used to replace Euclidean distance in Little et al. (1997), Rathbun (1998), and Jensen et al. (2006), which yielded impermissible models. Instead, permissible models based on reduced-rank radial-basis functions were given by Wang & Ranalli (2007).

There has been a great deal of interest in kriging over the surface of the earth, which is an approximate sphere. Kriging on geographical coordinates can create distortions, yet such applications have appeared (Ecker & Gelfand, 1997; Kaluzny et al., 1998), which have been criticized (Banerjee, 2005). Most research has centered on geodesic, or great-circle distance. If geodesic distance is substituted for Euclidean distance for the models in eqn 3, only the exponential and spherical models are permissible (Gneiting, 2013). Note that distance is

measured in radians, and restricted to the interval $[0, \pi]$.

For an interesting ecological application, Bradburd et al. (2013) propose an extension of a powered exponential, also called a stable geostatistical model, that combines Euclidean distance with ecological or genetic distance. Then Guillot et al. (2014) show how the stable model can be used with geodesic (great circle) distances, but only if the power parameter of the stable model is restricted, and they also discuss ways of "gluing" geographical distances and environmental distances to create permissible models.

The literature given above, with many examples, shows that replacing Euclidean distance with some other metric that makes more physical sense is intuitively appealing, but yields impermissible models that do not guarantee positive-definite covariance matrices. To further illustrate the issues with a real example, I re-analyze the data in Ladle et al. (2017b).

REANALYSIS OF LADLE ET AL. (2017)

Prior to a reanalysis of Ladle et al. (2017b), I summarize their analysis. I then review several general approaches to spatial models for non-Euclidean distance metrics. Finally, I introduce the reduced-rank method that I ultimately use on the data of Ladle et al. (2017b).

Review of Ladle et al. (2017)

Ladle et al. (2017b) provide an interesting study of human activity along a linear network of trails in a portion of Alberta's Rocky Mountains. They analyzed both motorised and non-motorised activities; see Figure 1 in Ladle et al. (2017b) for the trails and study area. Their on-line spatial locations are shown here in Figure 4. They use a two-stage analysis, first fitting a mixed-effects logistic regression model to the presence of any activity during hourly increments. The fixed effects in their models include rainfall, date, time of day, etc. Random effects for spatial location

and time were also included, and estimated as best linear unbiased predictions (BLUPs). These BLUPs were subsequently used in a second stage of analysis as spatial data. Linear network distance among BLUPs was used in place of Euclidean distance, and ordinary kriging was used to predict BLUPs at unsampled locations along the linear network; see Figure 4 in Ladle et al. (2017b). In all that follows, I will re-analyze only the non-motorised data from Ladle et al. (2017b), using the estimated BLUP values and the linear-network and Euclidean distance matrices that they provided as on-line data.

The main objective of this paper, and my prior review, is to show that substitution of non-Euclidean distance metrics into autocorrelation models derived for Euclidean distance can create covariance matrices that are not positive definite. For the particular case of Ladle et al. (2017b), using their linear-network distance matrix in the models given in eqn 3 showed that none of the models are permissible beyond a certain α value (Figure 3a). On the other hand, using the Euclidean distance matrix provided by Ladle et al. (2016), all models yield positive-definite covariance matrices at all values of $\alpha > 0$ (Figure 3b), which simply verifies that they are permissible models. Note that the fitted exponential model had $\hat{\alpha} = 14.2$ km in Ladle et al. (2017b) for nonmotorised variables, which yielded a positive-definite covariance matrix because $\alpha < 28.2$ km had all positive eigenvalues (Figure 3a). The (incorrectly) fitted spherical models in Ladle et al. (2017b) (see Ladle et al., 2017a) had estimated range parameters > 40 km, which would not yield positive-definite covariance matrices because $\alpha > 15.9$ had negative eigenvalues (Figure 3a).

Review of Non-Euclidean Distance Models

Several approaches can be used for creating spatial models in novel situations, whether for non-Euclidean distances or other situations. The first is the spatial moving average, also called a

process convolution and autoconvolution. The spatial moving-average approach is very similar to a moving-average model in time series, except that the random variables that are "smoothed" are continuous in space (also known as a white noise process). This approach has been used for flexible variogram modeling (Barry & Ver Hoef, 1996), multivariable (cokriging) models (Ver Hoef & Barry, 1998; Ver Hoef et al., 2004), nonstationary models (Higdon, 1998; Higdon et al., 1999), stream network models (Ver Hoef et al., 2006; Cressie et al., 2006; Ver Hoef & Peterson, 2010), models on the sphere (Gneiting, 2013), and spatio-temporal models (Wikle, 2002; Conn et al., 2015). Using the moving-average approach requires solving integrals to obtain the autocorrelation function, or approximating the integrals. For example, the integrals are tractable for stream networks when purely dichotomous branching occurs (Ver Hoef et al., 2006), however they are not tractable for more general linear networks.

The use of bivariate splines over complex spatial domains is an area of active research, beginning with Ramsay (2002), which includes Wang & Ranalli (2007) and soap-film smoothing (Wood et al., 2008), with recent improvements (Sangalli et al., 2013; Miller & Wood, 2014). Approximating locations within irregular boundaries by a wire mesh introduces neighbor-based methods, also known as lattice-based methods, such as integrated nested Laplace approximation (INLA, Rue et al., 2009). At the limit of a very dense mesh, these methods are an approximation to a spatial partial difference equation (SPDE, Lindgren et al., 2011), that can allow for barriers and complex spatial domains (Bakka et al., 2016). Another approach using wire meshes is given by McIntyre & Barry (2017).

There are many connections among the methods given above, and I do not attempt a complete review. The approach that I will feature is a reduced-rank idea, also called a dimension-reduction (Wikle & Cressie, 1999) and spatial radial-basis (Lin & Chen, 2004; Hefley et al., 2016) method. It is closely related to splines, and handles non-Euclidean topology and has

computational advantages. This is a very general method, and the one that I will use to re-analyze the data of Ladle et al. (2017b). It has been mostly featured as a method for big data sets (e.g. Wikle & Cressie, 1999; Ruppert et al., 2003; Cressie & Johannesson, 2008; Banerjee et al., 2008). I will use this method for models using linear network distances, which I describe next.

Reduced-Rank Methods for Non-Euclidean Distances

The reduced-rank models are a special case of linear mixed models, so I provide a quick review. In fact, eqn 1 is a special case of a mixed model. A mixed model is often written as

$$\boldsymbol{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\boldsymbol{\nu} + \boldsymbol{\varepsilon}, \qquad \text{eqn 11}$$

where $\mathbf{X}$ is a design matrix with covariates, $\boldsymbol{\beta}$ is a vector of regression parameters, $\mathbf{W}$ is a random-effects design matrix, $\boldsymbol{\nu}$ is a vector of zero-mean random effects with variance σ_p^2, and $\text{var}(\boldsymbol{\varepsilon}) = \sigma_0^2\mathbf{I}$. In statistical textbooks, $\mathbf{W}$ in eqn 11 often contains dummy variables (zeros or ones) that indicate some factor level of the random effect. However, $\mathbf{W}$ can also contain covariates, in which case $\boldsymbol{\nu}$ contains random effects for the slope of a line, illustrating that there are no restrictions on the types of values (continuous or categorical) contained in $\mathbf{W}$. For the linear mixed model, eqn 11, recall that

$$\text{var}(\boldsymbol{Y}) = \sigma_p^2\mathbf{W}\mathbf{G}\mathbf{W}' + \sigma_0^2\mathbf{I}, \qquad \text{eqn 12}$$

where $\mathbf{G}$ is the correlation matrix for $\boldsymbol{\nu}$. Classically, for mixed models, random effects are assumed independent, so $\mathbf{G} = \mathbf{I}$, and then $\text{var}(\boldsymbol{Y}) = \sigma_p^2\mathbf{W}\mathbf{W}' + \sigma_0^2\mathbf{I}$.

For the reduced-rank models, let $\mathbf{D}$ denote a matrix of Euclidean distances among locations and $\mathbf{L}$ denote a matrix of linear-network distances. Let $\mathbf{R}_{m,\mathbf{A},\alpha}$ be a spatial autocorrelation

matrix, where $m = e, s, g, c$, or h, for exponential, spherical, Gaussian, Cauchy, or hole effect, respectively, for one of the models in eqn 3, $\mathbf{A}$ is a distance matrix, either $\mathbf{D}$ or $\mathbf{L}$, and α is the range parameter for one of the models in eqn 3. For example, $\mathbf{R}_{e,\mathbf{L},\alpha} = \exp(-\mathbf{L}/\alpha)$. Then let $\mathbf{R}^r_{m,\mathbf{A},\alpha}$ be the matrix where some of the columns of $\mathbf{R}_{m,\mathbf{A},\alpha}$ are kept as "knots", and all other columns have been removed; hence the term "reduced-rank." For example, for the Ladle et al. (2016) data, there are 239 locations, so $\mathbf{R}_{m,\mathbf{A},\alpha}$ is 239×239. I will reduce it to just 120 columns, so $\mathbf{R}^r_{m,\mathbf{A},\alpha}$ is 239×120, and the reason for 120 knots is discussed later.

The reduced-rank method requires the selection of knots. In general, knots can be placed anywhere, and not only at the observed locations. I used K-means clustering (MacQueen, 1967) on the spatial coordinates to create 120 groups. Because K-means clustering minimizes within-group variance while maximizing among-group variance, the centroid of each group tends to be regularly spaced; i.e. it is a space-filling design (e.g. Ver Hoef & Jansen, 2015). Then, the knots were moved to the nearest observed location. The original knot locations are shown in blue, and then moved to the red circles in Fig. 4. It will be useful to have the matrix of Euclidean distances among knots only, which is a subset of the rows and columns of $\mathbf{D}$, and I denote the knot-to-knot distances as $\mathbf{D}^k$.

Now consider the following random-effects model as a special case of eqn 11,

$$\boldsymbol{Y} = \mathbf{1}\mu + [\mathbf{R}^r_{m,\mathbf{A},\alpha}]\boldsymbol{\nu} + \boldsymbol{\varepsilon}, \qquad \text{eqn 13}$$

In eqn 13, I have replaced $\mathbf{W}$ with $\mathbf{R}^r_{m,\mathbf{A},\alpha}$, and there are no covariates in $\mathbf{X}$, so $\mathbf{X}$ is a vector of ones, and I will assume that $\text{var}(\boldsymbol{\nu}) = [\mathbf{R}_{m,\mathbf{D}^k,\eta}]^{-1}$. A broad introduction to spatial basis functions, and rank reduction, for ecologists is given by Hefley et al. (2016).

The innovations for reduced-rank spatial models in eqn 13 occur because: 1) we use

correlation models of distance in the random-effects design matrix, essentially $\mathbf{W} = \mathbf{R}^r_{m,\mathbf{A},\alpha}$, and 2) we also allow the random effects $\boldsymbol{\nu}$ to be spatially autocorrelated using the *inverse* covariance matrix from one of the models in eqn 3. The model in eqn 13 must have a positive-definite covariance matrix, so I assume Euclidean distance will be used for the distance among knots. In that case, eqn 13 leads to the following covariance matrix,

$$\boldsymbol{\Sigma} = \sigma^2_p \mathbf{R}^r_{m,\mathbf{A},\alpha}[\mathbf{R}_{m,\mathbf{D}^k,\eta}]^{-1}[\mathbf{R}^r_{m,\mathbf{A},\alpha}]' + \sigma^2_0\mathbf{I} \qquad \text{eqn 14}$$

Note that $\mathbf{R}^r_{m,\mathbf{A},\alpha}$ and $\mathbf{R}_{m,\mathbf{D}^k,\eta}$ could have different model forms (e.g., m could be exponential from eqn 3 for $\mathbf{R}^r_{m,\mathbf{A},\alpha}$, while m is spherical from eqn 3 for $\mathbf{R}_{m,\mathbf{D}^k,\eta}$). Also note that $\mathbf{A}$ could be $\mathbf{D}$, $\mathbf{L}$, or some other matrix based on any number of distance metrics. The construction in eqn 14 is very flexible, and several comments are pertinent:

1. Strictly speaking, the covariance matrix in eqn 14 is guaranteed to be positive definite only if $\sigma^2_0 > 0$. This is no different than mixed models, eqn 11, where recall that the variance was $\sigma^2_{\mathrm{p}}\mathbf{WGW}' + \sigma^2_0\mathbf{I}$.

2. Note that the inverse of a positive-definite matrix will also be positive definite, so $[\mathbf{R}_{m,\mathbf{D}^k,\eta}]^{-1}$ is positive definite as long as Euclidean distance $\mathbf{D}^k$ is used, which ensures that $\sigma^2_{\mathrm{p}}\mathbf{R}^r_{m,\mathbf{A},\alpha}[\mathbf{R}_{m,\mathbf{D}^k,\eta}]^{-1}[\mathbf{R}^r_{m,\mathbf{A},\alpha}]'$ is nonnegative definite.

3. It might seem unusual to model the covariance among the knots as the inverse $[\mathbf{R}_{m,\mathbf{D}^k,\eta}]^{-1}$. The reasons for the inverse are complex (Banerjee et al., 2008), but there is an intuitive explanation. Suppose that the reduced-rank matrix is based on Euclidean distance, that is, let $\mathbf{A} = \mathbf{D}$, so we have $\mathbf{R}^r_{m,\mathbf{D},\alpha}$. Now, let the knots increase in number until the knots become exactly the same as the observed locations. Then, $\mathbf{R}^r_{m,\mathbf{D},\alpha}$ becomes $\mathbf{R}_{m,\mathbf{D},\alpha}$, the full covariance matrix, and $[\mathbf{R}_{m,\mathbf{D}^k,\eta}]^{-1}$ becomes $[\mathbf{R}_{m,\mathbf{D},\alpha}]^{-1}$ (note that because they have the

same model type and distance matrix, η is equivalent to α), the inverse of the full covariance matrix. The inverse cancels one of the full covariance matrices, so in eqn 14, $\sigma^2_{\mathrm{p}}\mathbf{R}_{m,\mathbf{D},\alpha}[\mathbf{R}_{m,\mathbf{D},\alpha}]^{-1}[\mathbf{R}_{m,\mathbf{D},\alpha}]' = \sigma^2_{\mathrm{p}}\mathbf{R}_{m,\mathbf{D},\alpha}$, which is the $n \times n$ symmetric covariance matrix without any reduction in rank. By using the inverse, the formulation in eqn 14 allows us to recover a typical covariance matrix as the knots become equal to the observed locations. My approach will be that $\mathbf{G}$ in eqn 12 is $[\mathbf{R}_{m,\mathbf{D}^k,\eta}]^{-1}$, but note that any other positive-definite matrix could be used for $\mathbf{G}$, including $\mathbf{G} = \mathbf{I}$.

4. It is not necessary to use reduced rank. The full covariance matrices in eqn 14 could be used, including the inverse of the Euclidean-distance covariance matrix sandwiched between the linear-distance covariance matrices, but see the next item.

5. In addition to allowing non-Euclidean distances in the random-effects design matrix, $\mathbf{R}^r_{m,\mathbf{A},\alpha}$, there is a computational advantage to using rank reduction in eqn 14. Notice that $\mathbf{\Sigma}$ is a 239×239 matrix, and likelihood-based methods (such as maximum likelihood, or restricted maximum likelihood) require the inverse of $\mathbf{\Sigma}$. Computing matrix inverses is computationally expensive, and grows exponentially with the dimension of the matrix (as a cube of the number of locations). However, the reduced-rank formulation allows an inverse of $\mathbf{\Sigma}$ that is reduced to the size of the rank reduction by using the Sherman-Morrison-Woodbury result (Sherman & Morrison, 1949; Woodbury, 1950); see an excellent review by Henderson & Searle (1981). In our case, if we choose 120 knots, then the inverse would be for a 120×120 matrix rather than a 239×239 matrix. The number of knots is a decision based on speed versus precision. Generally, the model will perform better with more knots, and a good guideline is to use as many knots as is computationally feasible. I used 120 knots, approximately half of all 239 locations, to illustrate that a

reduction in rank still works well.

In what follows, I will always choose a single model form, m, across all 3 components of $\mathbf{R}^r_{m,\mathbf{A},\alpha}[\mathbf{R}_{m,\mathbf{D}^k,\eta}]^{-1}[\mathbf{R}^r_{m,\mathbf{A},\alpha}]'$, and I will always use the linear-network distance matrix $\mathbf{L}$ for $\mathbf{A}$, but allow the autocorrelation parameter α to be different from η. For example, the reduced-rank exponential model that uses linear-network distance has a covariance matrix

$$\mathbf{\Sigma} = \sigma_p^2 \mathbf{R}^r_{e,\mathbf{L},\alpha}[\mathbf{R}_{e,\mathbf{D}^k,\eta}]^{-1}[\mathbf{R}^r_{e,\mathbf{L},\alpha}]' + \sigma_0^2 \mathbf{I}. \qquad \text{eqn 15}$$

For this covariance matrix, there are 4 parameters to estimate; σ_p^2, α, η, and σ_0^2. In what follows, I fit all reduced-rank models using REML.

Reanalysis of the Ladle et al. (2017) Data

The reanalysis of Ladle et al. (2017b) is given in Table 1. The data were downloaded from the Dryad Repository http://dx.doi.org/10.5061/dryad.62t17. To evaluate models, I use four criteria, the first being AIC (Akaike, 1973; Burnham & Anderson, 2002), which assumes that the data were distributed as a multivariate normal likelihood with a spatial covariance matrix (for an example using spatial models, see Hoeting et al., 2006). AIC was only used when fitting with REML.

The rest of the criteria are based on leave-one-out crossvalidation. Let $\mathbf{y}_{-i}$ be the vector of observed data with the ith observation removed. Then, using $\mathbf{y}_{-i}$ and the estimated covariance matrix, with the ith row and column removed, the ith observation is predicted, denoted as $\hat{Y}_i$, with eqn 6, and its prediction standard error, denoted as $se(\hat{Y}_i)$, is estimated with (the square root of) eqn 7. The correlation was computed on the set of pairs $\{(y_i, \hat{Y}_i); i = 1, \ldots, n\}$ for all i and reported as Corr in Table 1. Root-mean-squared prediction error (RMSPE, Table 1) was computed as the square root of the mean of $(y_i - \hat{Y}_i)^2$ for all i. The coverage of the 90%

prediction interval (CI90, Table 1) was the proportion of times that the interval $[\hat{Y}_i - 1.645se(\hat{Y}_i),\ \hat{Y}_i + 1.645se(\hat{Y}_i)]$ contained the true value y_i for all i.

First, I consider the fitted exponential model reported in Ladle et al. (2017b) (the first row in Table 1). The fitted model, which did not have a nugget effect, along with the empirical semivariogram, are shown as the dashed line for the exponential model in Figure 5a. Of particular interest is the fact that the CI90 for the model in Ladle et al. (2017b) covers the true value only 69.9% of the time (Table 1). This is due to the lack of a nugget effect. The covariance matrix is forcing high autocorrelation among sites that are close together, and hence the prediction variance assumes prediction is better than it really is, which results in prediction standard errors that are estimated to be too small. When semivariograms are fitted without a nugget effect, they should be checked carefully for fitting and prediction instabilities. Models without nugget can lead to computational instability when inverting the covariance matrix (Diamond & Armstrong, 1984; Posa, 1989; O'Dowd, 1991; Ababou et al., 1994). If the modeler insists on excluding the nugget effect (as often occurs when using kriging to approximate deterministic computer models, e.g. Martin & Simpson, 2005), a small nugget effect can be added to the diagonal (e.g. $1 \times 10^{-}6$ was used in Booker et al. (1999)) to improve computational stability. Problems can occur due to model type (Gaussian autocorrelation is the worst) and the arrangement of the spatial locations, when "near duplicate" locations can cause apparently singular matrices for computational purposes (Bivand et al., 2008, p. 220).

I fit all other models in eqn 3, both with and without a nugget effect, where linear-network distance was used in place of Euclidean distance. These form rows 2-10 in Table 1. REML was not used to fit these models because REML depends on the inverse of the covariance matrix, which was unstable for these models because their covariance matrices were not positive definite. For models without a nugget, CWLS, which adds weight to empirical semivariogram values with

smaller distances, provided poor fits due to the lack of congruence between the model being forced to zero at the origin, and the empirical semivariogram values. Thus, all models without a nugget effect were fitted by WLS, and all models with a nugget effect were fitted with CWLS (Table 1, Figure 5a). The results show that, other than the exponential model, all fitted models without a nugget effect had negative eigenvalues and, when using cross-validation, produced substantial numbers of negative values for prediction standard errors when using eqn 7 (31 for spherical, 97 for Gaussian, 121 for Cauchy, and 125 for Hole-effect). Adding a nugget effect helped, but only exponential, spherical, and Cauchy models had positive-definite covariance matrices. However, CI90 for all three models were well below the nominal 90% level. Of particular interest is the hole-effect model with a nugget effect. It would appear to have the best fit visually (Figure 5a), yet even when a nugget effect is included, it produced a cross-validation prediction with a negative prediction standard error (Table 1).

All models in eqn 3 were fitted with both CWLS and REML using Euclidean distance (Table 1, Figure 5b). As expected, all had positive-definite covariance matrices. In all cases, models fitted with REML outperformed those same model types when fitted with CWLS; that is, the exponential model fitted with REML had lower RMPSE than the exponential model fitted with CWLS, and models fitted with REML had CI90 closer to 90% than those same models fitted with CWLS.

Four models in Table 1 used the reduced-rank approach, based on exponential, spherical, Gaussian, and Cauchy autocorrelation models in eqn 3 as used in eqn 14 (the hole-effect model always performed poorly, so was eliminated). The estimated covariance parameters for each of the models are shown in Table 1. Note that all reduced-rank models outperformed all other models in terms of RMSPE, and they also had lower AIC than their Euclidean distance counterparts. CI90 for the reduced-rank models was always above 88%, so very close to the nominal 90%. Not only

were the reduced-rank models the best performers, they were all completely permissible and computationally faster than the Euclidean distance models. There was little actual difference among the reduced-rank models in performance.

The results in Table 1 show a clear advantage for the reduced-rank linear-network-distance models, but the actual gain in performance is rather small. That is, prediction intervals are valid for both Euclidean distance and reduced-rank models, but the reduced-rank models have prediction standard errors that are about 2% shorter than those for Euclidean distance. Next, I discuss Euclidean distance and network distance models in more detail.

Euclidean Distance versus Linear Network Distance

Representing a road, stream, etc., as a linear network in ecology, such as the trail network analyzed above, is a mathematical topology that is embedded in 2-D (or 3-D) Euclidean space. As such, variables measured on linear networks may be influenced by processes and patterns that operate strictly within the linear network, but also processes and patterns that operate in Euclidean space. For example, human activity on trails might be affected by slope, aspect, vegetation, a beautiful view, etc., that operate more in 2-D space than linear-network space. On the other hand, travel times from parking areas will affect human activity, and operate purely within linear-network space. My view, and those of others, (Dale & Fortin, 2010; Peterson et al., 2013) is that linear networks embedded in 2-D space have a duality. Moreover, a pattern occurring on one (say the linear network), can, and often will, be captured in the other (say Euclidean) purely through the correlation between their distances. For example, Figure 6 shows a scatter plot of Euclidean distances and linear network distances for all pair-wise sites in the data from Ladle et al. (2017b). In this case, it will be very difficult to see a large advantage in linear-network distance models over Euclidean distance models, or vice versa, which is confirmed by Table 1.

Nevertheless, we can model both linear-network distance and Euclidean distance simultaneously as a variance component model. Consider a combination of eqn 1 and eqn 13, where the reduced-rank construction is added, rather than replacing Euclidean distance, so

$$\boldsymbol{Y} = \mathbf{1}\mu + \boldsymbol{Z} + [\mathbf{R}^r_{m,\mathbf{A},\alpha}]\boldsymbol{\nu} + \boldsymbol{\varepsilon},$$

where the random effect $\boldsymbol{Z}$ has a Euclidean distance covariance matrix. For example, I fit a model that has a covariance matrix

$$\boldsymbol{\Sigma} = \sigma^2_{\text{Euc}}\mathbf{R}_{s,\mathbf{D},\phi} + \sigma^2_p\mathbf{R}^r_{s,\mathbf{L},\alpha}[\mathbf{R}_{s,\mathbf{D}^k,\eta}]^{-1}[\mathbf{R}^r_{s,\mathbf{L},\alpha}]' + \sigma^2_0\mathbf{I},$$

where $\mathbf{R}_{s,\mathbf{D},\phi}$ is an autocorrelation matrix based a spherical model with full Euclidean distance matrix $\mathbf{D}$ among all sites, range parameter ϕ, and σ^2_{Euc} is the Euclidean distance variance component. The fitted model parameters are shown as the last two rows in Table 1, with the first row the linear network distance component, and the last row the Euclidean distance component. Combining both linear network distance and Euclidean distance provided the best predictions overall, with the lowest RMSPE and good CI90. According to AIC = 903.35, the variance component model does not warrant estimating the two extra parameters because AIC was lower for exponential, spherical, and Cauchy reduced-rank-only models, however cross-validation summaries indicated otherwise. A variance component approach, combining covariance models based on linear networks, with those based on Euclidean distance, was also recommended for stream network models (Ver Hoef & Peterson, 2010), and is an intuitively appealing idea that puts both components in the model and lets the data decide on their relative contributions.

DISCUSSION AND CONCLUSIONS

I have shown that a reduced-rank method can be used to create permissible models that guarantee positive-definite covariance matrices for spatial models using linear-network distance. The reduced-rank method is very flexible for various spatial topologies and distance metrics, and also has computational advantages. For the data from Ladle et al. (2017b), there was a distinct benefit, by lowering RMSPE and AIC, for linear-network distance over Euclidean distance models, but the best model combined both distance metrics (Table 1). For the reduced-rank models, consideration must be given to the number and placement of knots (Ruppert et al., 2003; Gelfand et al., 2012), which continues to be an area of active research.

While it is possible to fit impermissible models (Table 1) and then check the fitted model to ensure that the covariance matrix is positive definite, this practice is discouraged in traditional geostatistics. For example, note that some models (Table 1) happened to have positive-definite covariance matrices for the specific set of locations and estimated α values, resulting in cross-validation predictions that had positive variance estimates. However, as discussed for eqn 9, when predicting at locations where data were not collected, a larger covariance matrix must be considered. This can be computationally expensive or impossible to check (it is computationally expensive to compute eigenvalues) if there are thousands of prediction locations, as there were in Ladle et al. (2017b). Much simpler, and safer, is to choose permissible models/methods that guarantee positive-definite covariance matrices for all spatial configurations and model parameter values.

The reduced-rank methods are not the only approach for developing models for non-Euclidean distance metrics, as I reviewed earlier. The larger point of Ladle et al. (2017b) is important. Scientists are realizing that Euclidean distance may not represent ecologically-relevant

distance. New methods using non-Euclidean distance provide exciting research opportunities, but it requires collaboration between statisticians and ecologists to ensure statistical models have appropriate properties.

ACKNOWLEDGMENTS

The project received financial support from the National Marine Fisheries Service, NOAA. The findings and conclusions in the paper of the author does not necessarily represent the views of the reviewers nor the National Marine Fisheries Service, NOAA. Any use of trade, product, or firm names do not imply an endorsement by the U.S. Government.

DATA AND CODE ACCESSIBILITY

Original data from Ladle et al. (2017b) were made available at the Dryad Repository http://dx.doi.org/10.5061/dryad.62t17. An `R` (R Core Team, 2017) package called `KrigLinCaution` was created that contains all data, code, and analyses. This manuscript was created using `knitr` (Xie, 2014, 2015, 2016), and the manuscript combining LaTeX and `R` code is also included in the package. The package can be downloaded at https://github.com/jayverhoef/KrigLinCaution.git, with instructions for installing the package.

References

Ababou, R., A. C. Bagtzoglou and E. F. Wood, 1994. On the condition number of covariance matrices in kriging, estimation, and simulation of random fields. Mathematical Geology **26**(1):99–133.

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In

B. Petrov & F. Csaki, editors, Second International Symposium on Information Theory, pages 267–281. Akademiai Kiado, Budapest.

Bakka, H., J. Vanhatalo, J. Illian, D. Simpson and H. Rue, 2016. Accounting for physical barriers in species distribution modeling with non-stationary spatial random effects. arXiv preprint arXiv:1608.03787 .

Banerjee, S., 2005. On geodetic distance computations in spatial modeling. Biometrics **61**(2):617–625.

Banerjee, S., A. E. Gelfand, A. O. Finley and H. Sang, 2008. Gaussian predictive process models for large spatial data sets. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **70**(4):825–848.

Barry, R. P. and J. M. Ver Hoef, 1996. Blackbox kriging: Spatial prediction without specifying variogram models. Journal of Agricultural, Biological, and Environmental Statistics **1**:297–322.

Bivand, R. S., E. J. Pebesma and V. Gomez-Rubio, 2008. Applied Spatial Data Analysis with R. Springer, NY. URL `http://www.asdar-book.org/`.

Booker, A. J., J. Dennis Jr, P. D. Frank, D. B. Serafini, V. Torczon and M. W. Trosset, 1999. A rigorous framework for optimization of expensive functions by surrogates. Structural Optimization **17**(1):1–13.

Bradburd, G. S., P. L. Ralph and G. M. Coop, 2013. Disentangling the effects of geographic and ecological isolation on genetic differentiation. Evolution **67**(11):3258–3273.

Burnham, K. P. and D. R. Anderson, 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. Springer-Verlag Inc, New York.

554 Chiles, J.-P. and P. Delfiner, 1999. Geostatistics: Modeling Spatial Uncertainty. John Wiley &
555 Sons, New York.

556 Conn, P. B., D. S. Johnson, J. M. Ver Hoef, M. B. Hooten, J. M. London and P. L. Boveng, 2015.
557 Using spatiotemporal statistical models to estimate animal abundance and infer ecological
558 dynamics from survey counts. Ecological Monographs **85**(2):235–252.

559 Cressie, N., 1985. Fitting models by weighted least squares. Journal of the International
560 Association for Mathematical Geology **17**:563–586.

561 Cressie, N., 1990. The origins of kriging. Mathematical Geology **22**:239–252.

562 Cressie, N., J. Frey, B. Harch and M. Smith, 2006. Spatial prediction on a river network. Journal
563 of Agricultural, Biological, and Environmental Statistics **11**(2):127–150.

564 Cressie, N. and G. Johannesson, 2008. Fixed rank kriging for very large spatial data sets. Journal
565 of the Royal Statistical Society, Series B **70**(1):209–226.

566 Cressie, N. and S. N. Lahiri, 1996. Asymptotics for REML estimation of spatial covariance
567 parameters. Journal of Statistical Planning and Inference **50**:327–341.

568 Cressie, N. and J. J. Majure, 1997. Spatio-temporal statistical modeling of livestock waste in
569 streams. Journal of Agricultural, Biological, and Environmental Statistics pages 24–47.

570 Cressie, N. A. C., 1993. Statistics for Spatial Data, Revised Edition. John Wiley & Sons, New
571 York.

572 Curriero, F. C., 2006. On the use of non-euclidean distance measures in geostatistics.
573 Mathematical Geology **38**(8):907–926.

574 Dale, M. and M.-J. Fortin, 2010. From graphs to spatial graphs. Annual Review of Ecology,
575 Evolution, and Systematics **41**(1):21.

576 Diamond, P. and M. Armstrong, 1984. Robustness of variograms and conditioning of kriging
577 matrices. Mathematical Geology **16**(8):809–822.

578 Ecker, M. D. and A. E. Gelfand, 1997. Bayesian variogram modeling for an isotropic spatial
579 process. Journal of Agricultural, Biological, and Environmental Statistics pages 347–369.

580 Gandin, L. S., 1963. Objective Analysis of Meteorological Fields, volume 242.
581 Gidrometeorologichoskoe Izdatel'stvo (GIMIZ), Leningrad, (translated by Israel Program for
582 Scientific Translations Jerusalem, 1965).

583 Gardner, B., P. J. Sullivan and A. J. Lembo Jr., 2003. Predicting stream temperatures:
584 geostatistical model comparison using alternative distance metrics. Canadian Journal of
585 Fisheries and Aquatic Sciences **60**:344–351.

586 Gelfand, A. E., S. Banerjee and A. O. Finley, 2012. Spatial design for knot selection in
587 knot-based dimension reduction models. Spatio-Temporal Design: Advances in Efficient Data
588 Acquisition pages 142–169.

589 Gneiting, T., 2013. Strictly and non-strictly positive definite functions on spheres. Bernoulli
590 **19**(4):1327–1349.

591 Goovaerts, P., 1997. Geostatistics for Natural Resources Evaluation. Oxford University Press,
592 New York, NY.

593 Guillot, G., R. L. Schilling, E. Porcu and M. Bevilacqua, 2014. Validity of covariance models for
594 the analysis of geographical variation. Methods in Ecology and Evolution **5**(4):329–335.

595 Hefley, T. J., K. M. Broms, B. M. Brost, F. E. Buderman, S. L. Kay, H. R. Scharf, J. R. Tipton,
596 P. J. Williams and M. B. Hooten, 2016. The basis function approach for modeling
597 autocorrelation in ecological data. Ecology page Wiley Online Library.

598 Henderson, H. and S. R. Searle, 1981. On deriving the inverse of a sum of matrices. SIAM Review
599 **50**:53–60.

600 Heyde, C. C., 1994. A quasi-likelihood approach to the REML estimating equations. Statistics &
601 Probability Letters **21**:381–384.

602 Higdon, D., 1998. A process-convolution approach to modelling temperatures in the North
603 Atlantic Ocean (Disc: P191-192). Environmental and Ecological Statistics **5**:173–190.

604 Higdon, D., J. Swall and J. Kern, 1999. Non-stationary spatial modeling. In J. M. Bernardo,
605 J. O. Berger, A. P. Dawid & A. Smith, editors, Bayesian Statistics 6 – Proceedings of the Sixth
606 Valencia International Meeting, pages 761–768. Clarendon Press [Oxford University Press].

607 Hoeting, J. A., R. A. Davis, A. A. Merton and S. E. Thompson, 2006. Model selection for
608 geostatistical models. Ecological Applications **16**(1):87–98.

609 Jensen, O. P., M. C. Christman and T. J. Miller, 2006. Landscape-based geostatistics: a case
610 study of the distribution of blue crab in chesapeake bay. Environmetrics **17**(6):605–621.

611 Journel, A. G. and C. W. Huijbregts, 1978. Mining Geostatistics. Academic Press, London, UK.

612 Kaluzny, S. P., S. C. Vega, T. P. Cardoso and A. A. Shelly, 1998. Analyzing geostatistical data.
613 In S+SpatialStats: Users Manual for Windows and UNIX, pages 67–109. Springer New York,
614 New York, NY.

615 Ladle, A., T. Avgar, M. Wheatley and M. S. Boyce, 2017a. Corrigendum. Methods in Ecology
616 and Evolution URL http://dx.doi.org/10.1111/2041-210X.12905.

617 Ladle, A., T. Avgar, M. Wheatley and M. S. Boyce, 2017b. Predictive modelling of ecological
618 patterns along linear-feature networks. Methods in Ecology and Evolution **8**(3):329–338.

619 Lin, G.-F. and L.-H. Chen, 2004. A spatial interpolation method based on radial basis function
620 networks incorporating a semivariogram model. Journal of Hydrology **288**(3):288–298.

621 Lindgren, F., H. Rue and J. Lindström, 2011. An explicit link between Gaussian fields and
622 Gaussian Markov random fields: the stochastic partial differential equation approach. Journal
623 of the Royal Statistical Society: Series B (Statistical Methodology) **73**(4):423–498.

624 Little, L. S., D. Edwards and D. E. Porter, 1997. Kriging in estuaries: as the crow flies, or as the
625 fish swims? Journal of Experimental Marine Biology and Ecology **213**(1):1–11.

626 MacQueen, J. B., 1967. Some methods for classification and analysis of multivariate observations.
627 In L. M. L. Cam & J. Neyman, editors, Proc. of the fifth Berkeley Symposium on Mathematical
628 Statistics and Probability, volume 1, pages 281–297. University of California Press.

629 Martin, J. D. and T. W. Simpson, 2005. Use of kriging models to approximate deterministic
630 computer models. AIAA journal **43**(4):853–863.

631 Matheron, G., 1963. Principles of geostatistics. Economic Geology **58**:1246–1266.

632 McIntyre, J. and R. P. Barry, 2017. A lattice-based smoother for regions with irregular
633 boundaries and holes. Journal of Computational and Graphical Statistics, **early**
634 **on-line**:http://dx.doi.org/10.1080/10618600.2017.1375935.

635 Miller, D. L. and S. N. Wood, 2014. Finite area smoothing with generalized distance splines.
636 Environmental and ecological statistics **21**(4):715–731.

637 O'Dowd, R., 1991. Conditioning of coefficient matrices of ordinary kriging. Mathematical
638 Geology **23**(5):721–739.

639 Okabe, A. and K. Sugihara, 2012. Spatial Analysis Along Networks: Statistical and
640 Computational Methods. John Wiley & Sons.

641 Patterson, H. and R. Thompson, 1974. Maximum likelihood estimation of components of
642 variance. In Proceedings of the 8th International Biometric Conference, pages 197–207.
643 Biometric Society, Washington, DC.

644 Patterson, H. D. and R. Thompson, 1971. Recovery of inter-block information when block sizes
645 are unequal. Biometrika **58**:545–554.

646 Peterson, E. E., J. M. Ver Hoef, D. J. Isaak, J. A. Falke, M.-J. Fortin, C. Jordan, K. McNyset,
647 P. Monestiez, A. S. Ruesch, A. Sengupta et al., 2013. Stream networks in space: concepts,
648 models, and synthesis. Ecology Letters **16**:707–719.

649 Posa, D., 1989. Conditioning of the stationary kriging matrices for some well-known covariance
650 models. Mathematical Geology **21**(7):755–765.

651 R Core Team, 2017. R: A Language and Environment for Statistical Computing. R Foundation
652 for Statistical Computing, Vienna, Austria. URL `http://www.R-project.org`.

653 Ramsay, T., 2002. Spline smoothing over difficult regions. Journal of the Royal Statistical
654 Society: Series B (Statistical Methodology) **64**(2):307–319.

655 Rathbun, S. L., 1998. Spatial modelling in irregularly shaped regions: kriging estuaries.
656 Environmetrics **9**(2):109–129.

657 Rue, H., S. Martino and N. Chopin, 2009. Approximate bayesian inference for latent gaussian
658 models by using integrated nested laplace approximations. Journal of the Royal Statistical
659 Society: Series B (Statistical Methodology) **71**(2):319–392. URL
660 `http://dx.doi.org/10.1111/j.1467-9868.2008.00700.x`.

661 Ruppert, D., M. P. Wand and R. J. Carroll, 2003. Semiparametric Regression. Campbridge
662 University Press.

663 Sangalli, L. M., J. O. Ramsay and T. O. Ramsay, 2013. Spatial spline regression models. Journal
664 of the Royal Statistical Society: Series B (Statistical Methodology) **75**(4):681–703.

665 Schabenberger, O. and C. A. Gotway, 2005. Statistical Methods for Spatial Data Analysis.
666 Chapman Hall/CRC, Boca Raton, Florida.

667 Selby, B. and K. M. Kockelman, 2013. Spatial prediction of traffic levels in unmeasured locations:
668 applications of universal kriging and geographically weighted regression. Journal of Transport
669 Geography **29**:24–32.

670 Sherman, J. and W. J. Morrison, 1949. Adjustment of an inverse matrix corresponding to changes
671 in the elements of a given column or a given row of the original matrix. Annals of
672 Mathematical Statistics **20**:621.

673 Shiode, N. and S. Shiode, 2011. Street-level spatial interpolation using network-based IDW and
674 ordinary kriging. Transactions in GIS **15**(4):457–477.

675 Ver Hoef, J. M. and R. P. Barry, 1998. Constructing and fitting models for cokriging and
676 multivariable spatial prediction. Journal of Statistical Planning and Inference **69**:275–294.

677 Ver Hoef, J. M., N. Cressie and R. P. Barry, 2004. Flexible spatial models for kriging and
678 cokriging using moving averages and the fast Fourier transform (FFT). Journal of
679 Computational and Graphical Statistics **13**(2):265–282.

680 Ver Hoef, J. M. and J. K. Jansen, 2015. Estimating abundance from counts in large data sets of
681 irregularly-spaced plots using spatial basis functions. Journal of Agricultural, Biological, and
682 Environmental Statistics **20**:1–27.

683 Ver Hoef, J. M. and E. Peterson, 2010. A moving average approach for spatial statistical models
684 of stream networks (with discussion). Journal of the American Statistical Association **105**:6–18.

685 Ver Hoef, J. M., E. E. Peterson and D. Theobald, 2006. Spatial statistical models that use flow
686 and stream distance. Environmental and Ecological Statistics **13**(1):449–464.

687 Wang, H. and M. G. Ranalli, 2007. Low-rank smoothing splines on complicated domains.
688 Biometrics **63**:209–217.

689 Webster, R. and M. A. Oliver, 2007. Geostatistics for Environmental Scientists. John Wiley &
690 Sons, Chichester, England.

691 Wikle, C. K., 2002. A kernel-based spectral model for non-gaussian spatio-temporal processes.
692 Statistical Modelling **2**(4):299–314.

693 Wikle, C. K. and N. Cressie, 1999. A dimension-reduced approach to space-time kalman filtering.
694 Biometrika **86**(4):815–829.

695 Wood, S. N., M. V. Bravington and S. L. Hedley, 2008. Soap film smoothing. Journal of the
696 Royal Statistical Society: Series B (Statistical Methodology) **70**(5):931–955. URL
697 `http://dx.doi.org/10.1111/j.1467-9868.2008.00665.x`.

Woodbury, M. A., 1950. Inverting modified matrices. Memorandum Report 42, Statistical Research Group, Princeton N.J.

Xie, Y., 2014. knitr: a comprehensive tool for reproducible research in R. In V. Stodden, F. Leisch & R. D. Peng, editors, Implementing Reproducible Computational Research, pages 3 – 32. Chapman and Hall/CRC. URL http://www.crcpress.com/product/isbn/9781466561595. ISBN 978-1466561595.

Xie, Y., 2015. Dynamic Documents with R and knitr. 2nd edition. Chapman and Hall/CRC, Boca Raton, Florida. URL http://yihui.name/knitr/. ISBN 978-1498716963.

Xie, Y., 2016. knitr: A General-Purpose Package for Dynamic Report Generation in R. URL http://yihui.name/knitr/. R package version 1.15.1.

Table 1: Model fits and cross-validations statistics using the non-motorised data found in Ladle et al. (2017b). Models are given in eqn 3, and Y in the RR column indicates the reduced-rank version. The distance matrix used (Lin for linear, Euc for Euclidean) has column heading Dis. Meth column is fitting method, either WLS, CWLS, or REML, as described in Supplementary Material. Parameter estimates are given with column headings indicating parameter, using notation from eqn 2, eqn 3, and eqn 14. A blank indicates it was not part of the model. The column heading PD has a Y if the fitted covariance matrix was positive definite, otherwise it is blank. The Nnv column shows the number of negative prediction standard errors from cross-validation. On the right are Akaike Information Criteria (AIC) and summary statistics from cross-validation, showing Corr, the correlation between true and predicted values, root-mean-squared prediction errors (RMSPE), and proportion of times that the 90% prediction interval covered the true value (CI90). The last two rows, below the solid line, are a single variance component model including a reduced rank component, and a Euclidean distance component.

Model	RR	Dis	Meth	σ_p^2	α	η	σ_0^2	PD	Nnv	AIC	Corr	RMSPE	CI90
Exp		Lin		5.1	14.2			Y	0		0.639	1.594	0.699
Exp		Lin	CWLS	4.9	28.6		1.1	Y	0		0.672	1.483	0.866
Sph		Lin	WLS	4.8	36.4				31				
Sph		Lin	CWLS	3.6	43.7		1.2	Y	0		0.659	1.507	0.858
Gau		Lin	WLS	4.7	15.7				97				
Gau		Lin	CWLS	3.2	22.3		1.8		0		0.603	1.692	0.782
Cau		Lin	WLS	5.1	12.1				121				
Cau		Lin	CWLS	4.1	21.8		1.7	Y	0		0.613	1.593	0.828
Hol		Lin	WLS	4.2	7.9				125				
Hol		Lin	CWLS	2.5	8.9		1.8		1				
Exp		Euc	CWLS	4.6	15.9		1.0	Y	0		0.664	1.496	0.883
Sph		Euc	CWLS	3.6	30.0		1.3	Y	0		0.665	1.492	0.883
Gau		Euc	CWLS	3.1	15.1		1.8	Y	0		0.640	1.537	0.866
Cau		Euc	CWLS	3.9	14.1		1.7	Y	0		0.654	1.512	0.879
Hol		Euc	CWLS	2.5	6.2		1.9	Y	0		0.617	1.573	0.866
Exp		Euc	REML	3.0	11.4		1.4	Y	0	906.71	0.665	1.492	0.900
Sph		Euc	REML	3.3	27.9		1.5	Y	0	905.23	0.668	1.488	0.887
Gau		Euc	REML	2.2	9.0		1.8	Y	0	907.02	0.663	1.496	0.891
Cau		Euc	REML	2.7	9.5		1.8	Y	0	906.52	0.661	1.499	0.900
Hol		Euc	REML	2.0	5.7		2.3	Y	0	918.30	0.621	1.567	0.912
Exp	Y	Lin	REML	1.6	12.5	3.4	1.3	Y	0	901.61	0.674	1.475	0.891
Sph	Y	Lin	REML	1.4	26.0	9.4	1.3	Y	0	902.21	0.678	1.468	0.887
Gau	Y	Lin	REML	1.2	10.7	3.6	1.3	Y	0	905.76	0.671	1.481	0.883
Cau	Y	Lin	REML	1.5	9.8	3.5	1.3	Y	0	901.58	0.674	1.476	0.891
Sph	Y	Lin	REML	0.8	16.7	10.1	1.4	Y	0	903.35	0.686	1.453	0.895
Sph	N	Euc		2.0	27.6								

FIGURES

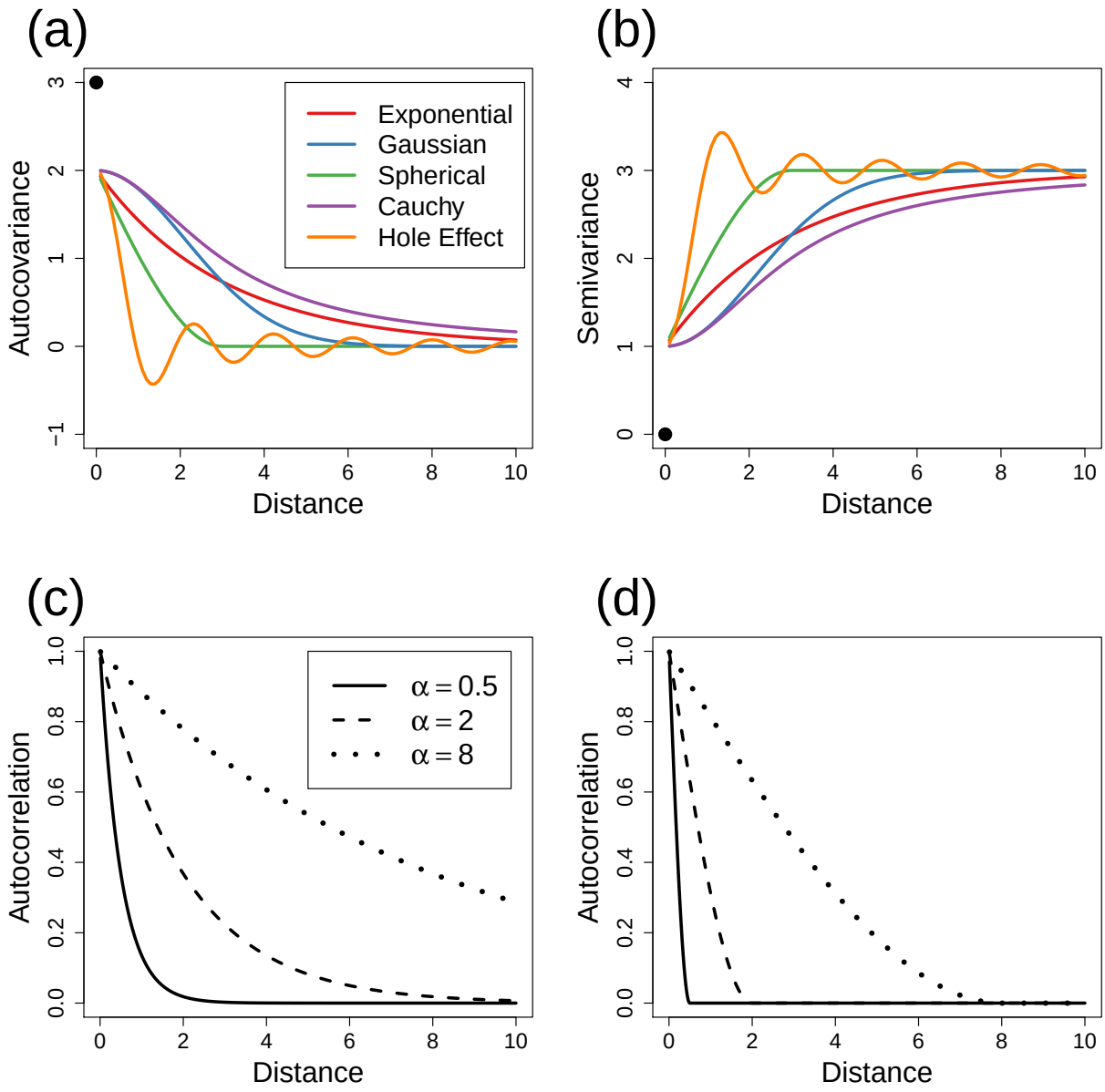

Figure 1: Autocorrelation models. (a) Autocovariance functions for various models, with a partial sill of 2 and a nugget effect of 1. (b) The same models as in (a), except represented as semivariogram models. Note that the black dots indicate a discontinuity of the fitted model at the origin due to the nugget effect, where the model "jumps" to the black dots when distance is exactly 0. Effect of the range parameter α on the (c) exponential model, and (d) spherical model.

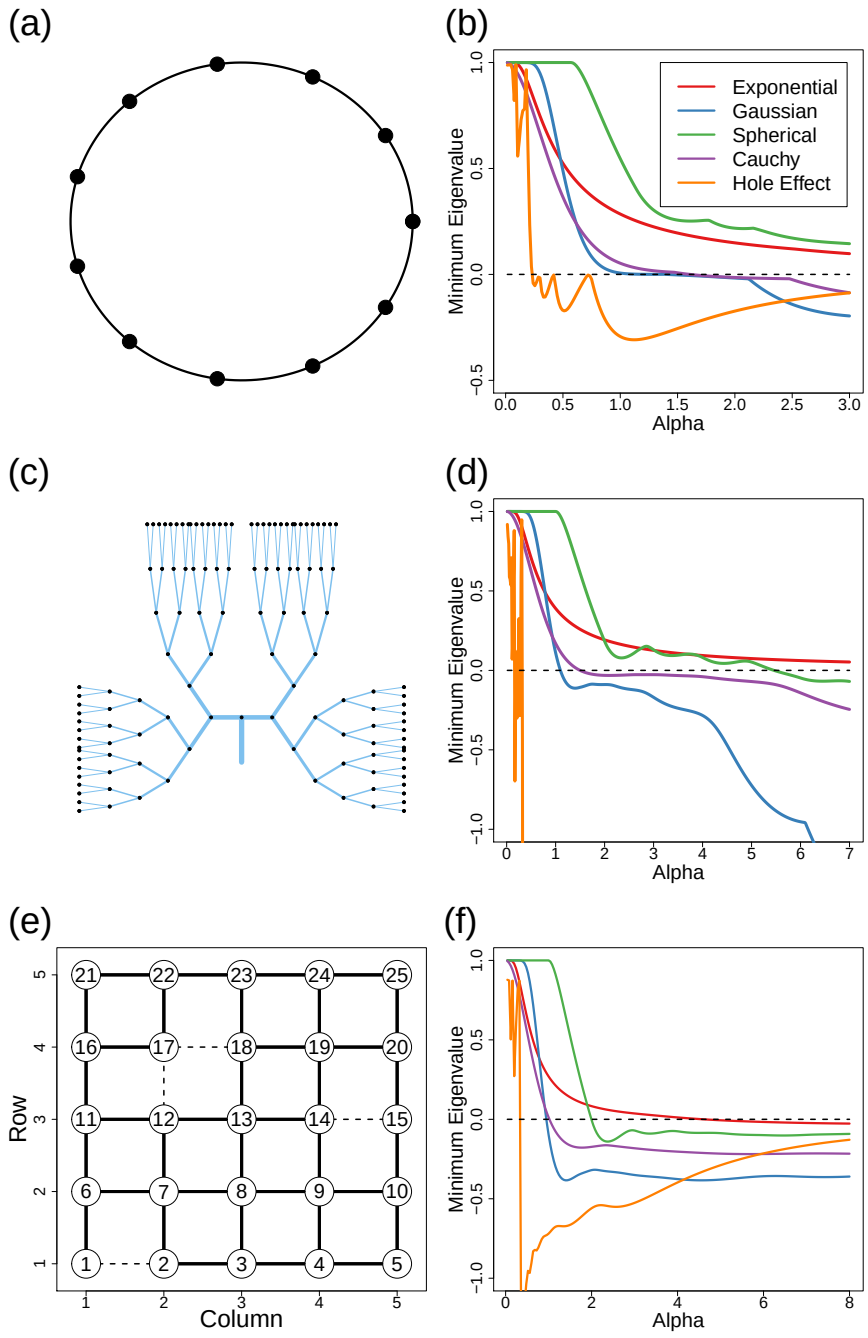

Figure 2: Cautionary examples. (a) 11 spatial locations on a circle are shown with solid black dots. (b) Minimum eigenvalue as a function of α for various autocorrelation models using distances on the circle. (c) A dichotomous branching network (stream) with 127 spatial locations (black dots) at the node of each branch. (d) Minimum eigenvalue as a function of α for various autocorrelation models using in-stream distance only. (e) 25 spatial locations on a grid network, where a perfect lattice includes the dashed line, but an irregular lattice includes only the solid lines. (f) Minimum eigenvalue as a function of α for various autocorrelation models using shortest path distances along the irregular lattice.

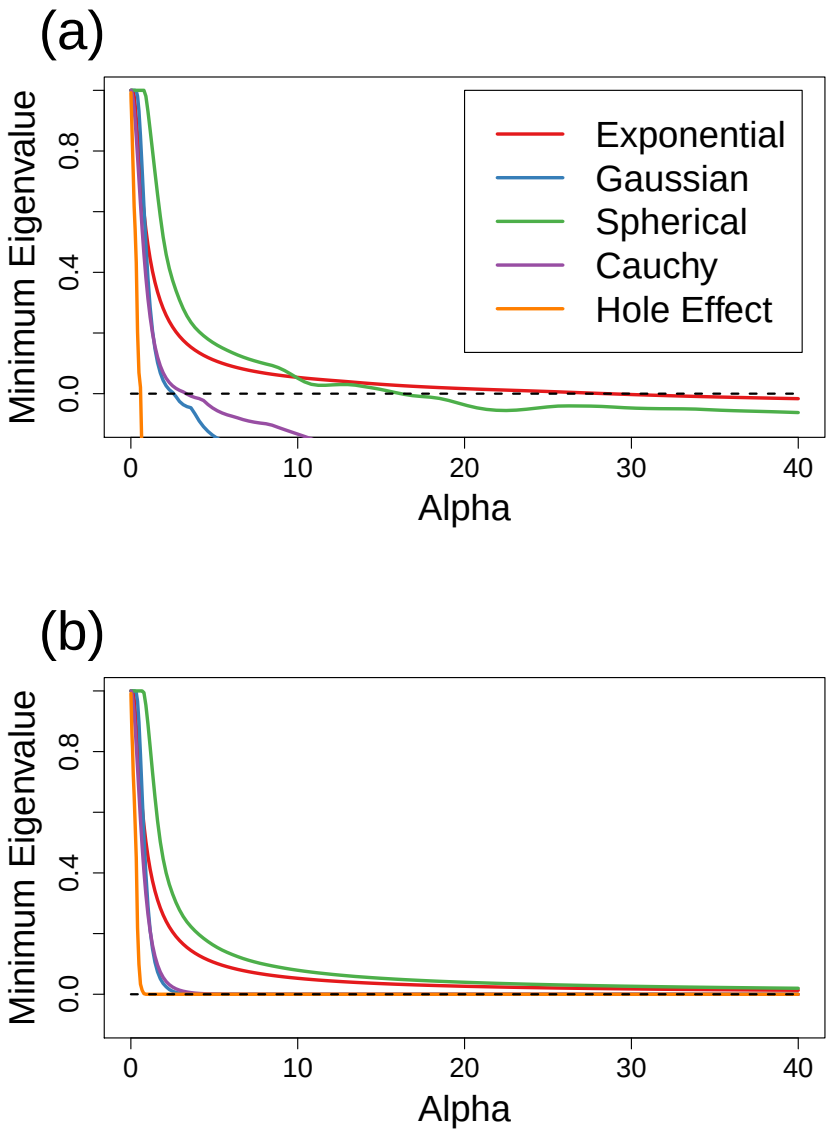

Figure 3: Minimum eigenvalues as a function of α for various autocorrelation models for the Ladle et al. (2017) data set. (a) Using linear distances (kilometers) among cameras. (b) Using Euclidean distances (kilometers) among cameras.

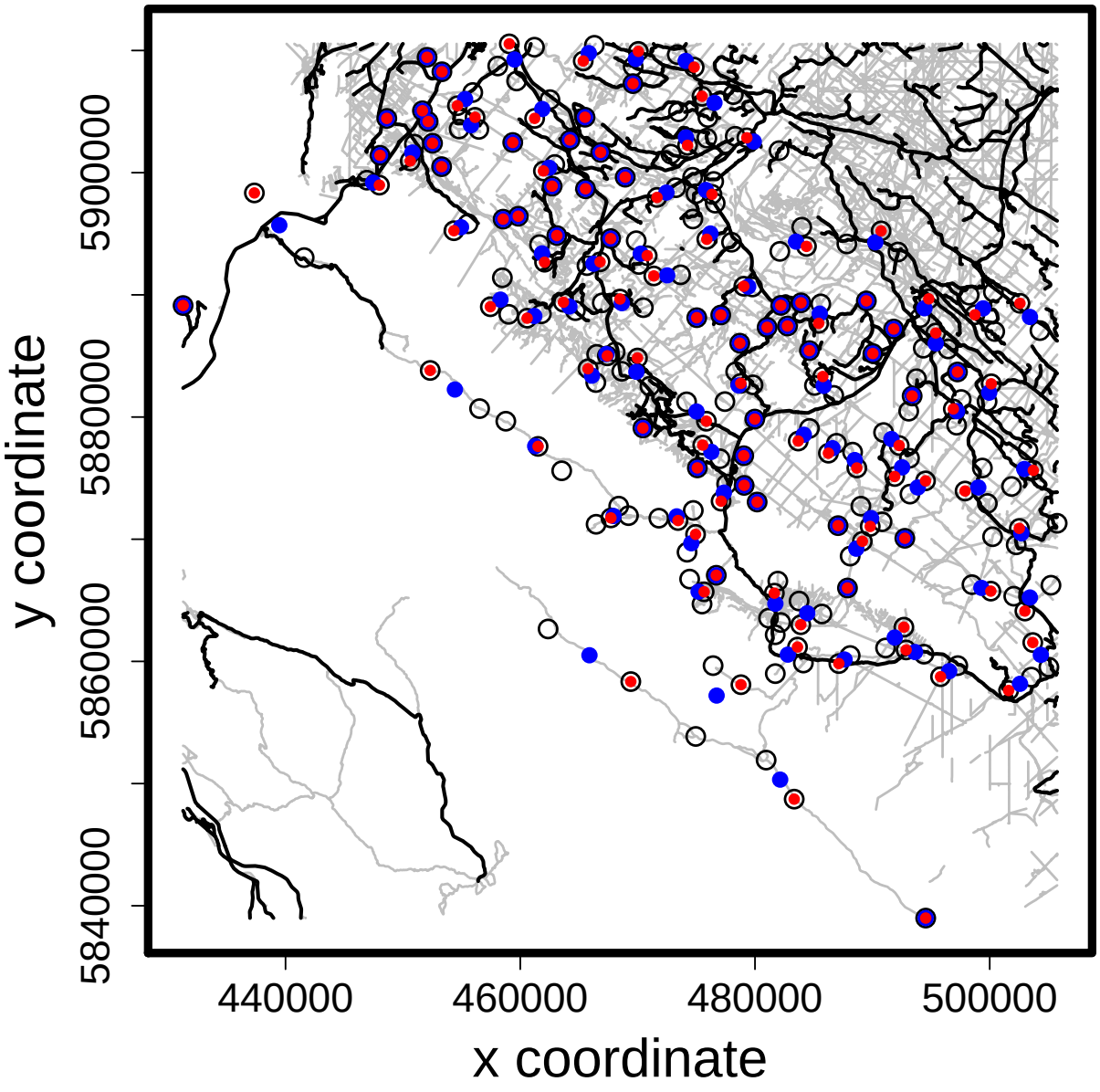

Figure 4: Study area from on-line data (similar to Fig. 1 in Ladle et al. 2017). Linear network consists of roads (black lines) and trails (gray lines). Spatial locations are open circles, and knot locations (used for reduced-rank methods) are shown as solid red circles. Initially, k-means on x- and y-coordinates created 120 clusters with center locations given by solid blue circles, and then these were moved to nearest actual locations (solid red circles). Note that there is some discrepancy between the map in Ladle et al. (2017) and the on-line data especially along the western and southern borders. All analyses in this paper used the on-line data.

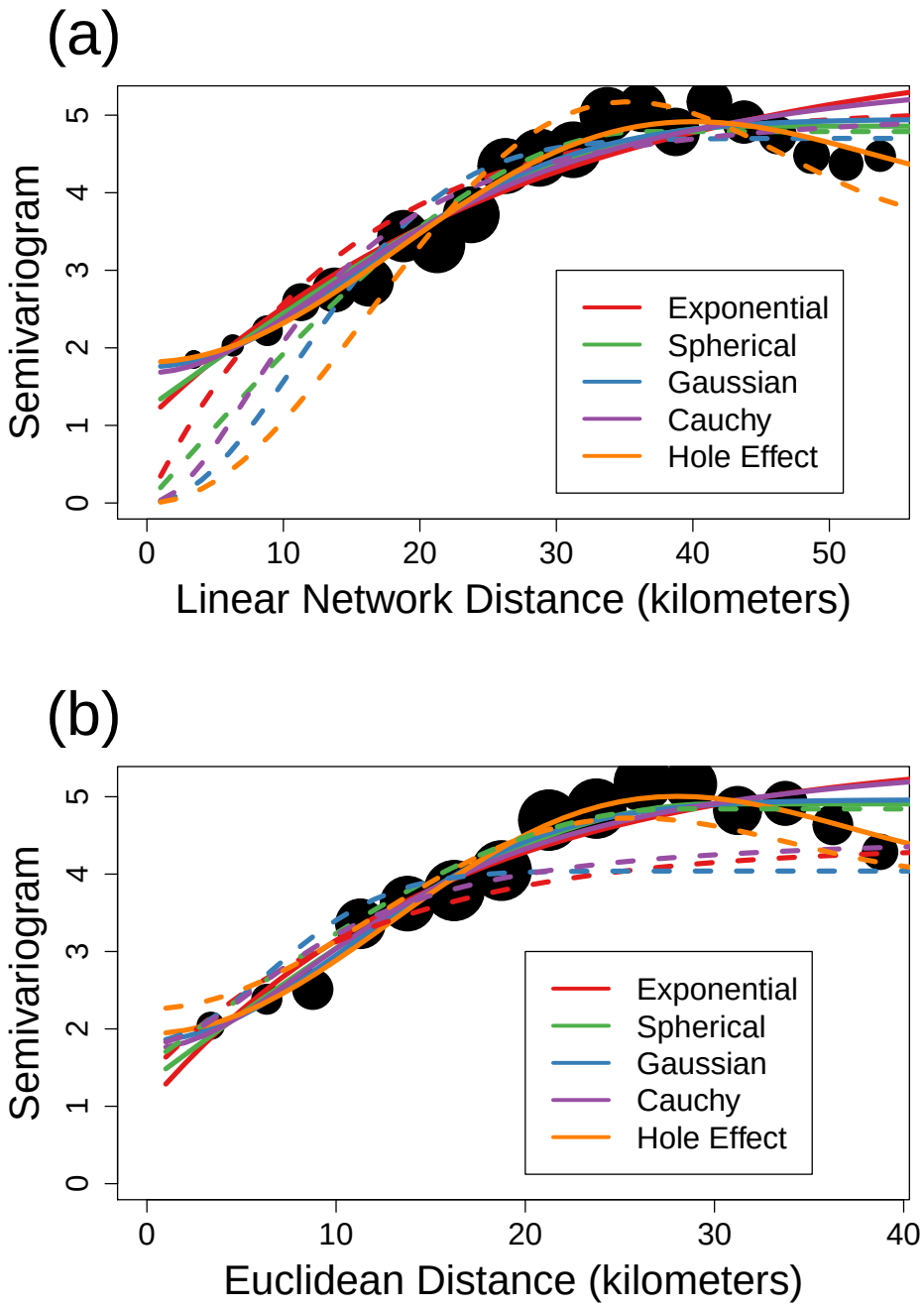

Figure 5: Empirical semivariograms with various fits. The solid black circles are empirical semivariogram values in distance classes, with size proportional to number of pairs of points in each distance class. (a) Linear network distances, where the dashed lines are fitted models without a nugget effect using WLS, and the solid lines are fitted models with a nugget effect using CWLS. (b) Euclidean distances, where the solid lines use CWLS, and the dashed lines use REML (which are not actually fit to the empirical semivariograms)

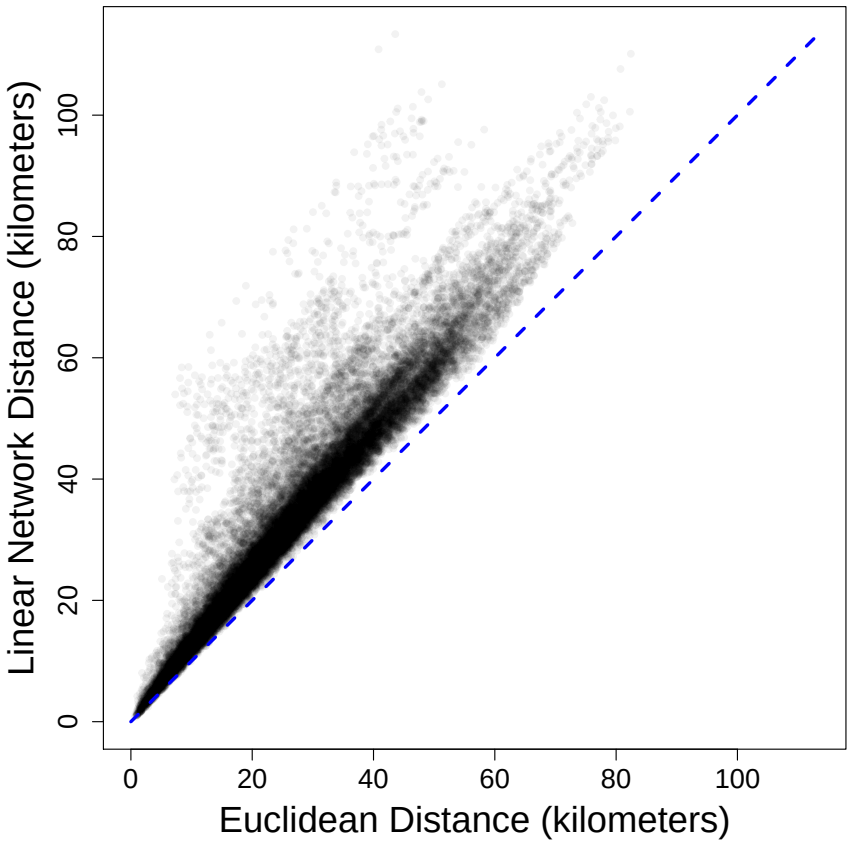

Figure 6: Scatter plot of Euclidean distance versus linear network distance for real data example. The points are semitransparent to reveal a strong correlation between distance metrics.

SUPPLEMENTAL MATERIAL

Estimation Methods

I use two methods to fit theoretical semivariograms eqn 3 to empirical semivariograms eqn 5. The first is simple weighted least squares. To show the dependence of the theoretical semivariogram on parameters, write any of the models, eqn 3, in semivariogram form with a nugget effect, $\gamma(h_k|\boldsymbol{\theta}) = \sigma_0^2 + \sigma_p^2(1 - \rho_m(h_k|\alpha))$, where $\boldsymbol{\theta} = (\sigma_p^2, \sigma_0^2, \alpha)$. Then the weighted least squares estimator of $\boldsymbol{\theta}$ is,

$$\hat{\boldsymbol{\theta}}_{WLS} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \sum_{k=1}^{K} [N(\mathcal{D}_k)](\hat{\gamma}(h_k) - \gamma(h_k|\boldsymbol{\theta}))^2.$$

Cressie's weighted least squares estimate of $\boldsymbol{\theta}$ is,

$$\hat{\boldsymbol{\theta}}_{CWLS} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \sum_{k=1}^{K} [N(\mathcal{D}_k)] \left(\frac{\hat{\gamma}(h_k)}{\gamma(h_k|\boldsymbol{\theta})} - 1 \right)^2.$$

REML does not use an empirical semivariogram. Rather, let $\mathbf{y}$ be a vector of observed data of length n, $\mathbf{X}$ a fixed effects design matrix with n rows and p linearly independent columns, $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ an $n \times n$ covariance matrix, in the same order as the data, that depends on distances between observations, and a set of parameters, as given in eqn 2. Note that I show the dependence of $\boldsymbol{\Sigma}$ on $\boldsymbol{\theta}$ with a subscript. Then REML estimates are given by

$$\begin{aligned} \hat{\boldsymbol{\theta}}_{REML} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \; [&(\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}_g)' \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} (\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}_g) + \log(|\boldsymbol{\Sigma}_{\boldsymbol{\theta}}|) + \\ &\log(|\mathbf{X}' \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \mathbf{X}|) + (n - p)\log(2\pi)], \end{aligned} \qquad \text{eqn S.1}$$

where

$$\boldsymbol{\beta}_g = (\mathbf{X}' \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \boldsymbol{Y} \qquad \text{eqn S.2}$$

is the generalized least squares estimator of $\boldsymbol{\beta}$. Note that for our case, $\mathbf{X} = \mathbf{1}$, where $\mathbf{1}$ is a vector of all 1s, and $\boldsymbol{\beta} = \mu$, a scalar.

Simple Example on Negative Variances from Improper Covariance Matrices

For a very simple, worked example in `R` on how a covariance matrix that is not positive definite can lead to negative variances, consider the 4 locations in a linear network shown in Figure S1.

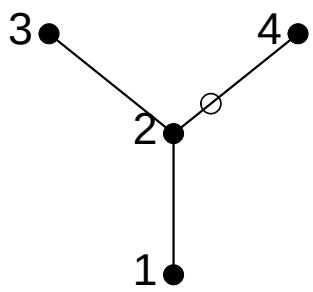

Figure S1: A simple 4-location network, where each location is given by a solid circle numbered from 1 to 4, along with a prediction location, shown by the open circle.

Let the linear distance between each connected location be 1 unit, so the distance matrix among the 4 locations, numbered sequentially for the rows and columns, is

```
linDmat = rbind(
  c(0,1,2,2),
  c(1,0,1,1),
  c(2,1,0,2),
  c(2,1,2,0))
```

$$\mathbf{D} = \begin{pmatrix} 0 & 1 & 2 & 2 \\ 1 & 0 & 1 & 1 \\ 2 & 1 & 0 & 2 \\ 2 & 1 & 2 & 0 \end{pmatrix}$$

I will use the Gaussian autocorrelation model, eqn 3, with $\sigma_p^2 = 1$, $\alpha = 3$, and a small nugget effect, $\sigma_0 = 0.01$.

```
Sig = exp(-(linDmat/3)^2) + diag(rep(0.01, times = 4))
```

$$\mathbf{\Sigma} = \begin{pmatrix} 1.010 & 0.895 & 0.641 & 0.641 \\ 0.895 & 1.010 & 0.895 & 0.895 \\ 0.641 & 0.895 & 1.010 & 0.641 \\ 0.641 & 0.895 & 0.641 & 1.010 \end{pmatrix} \quad \text{eqn S.3}$$

The spectral decomposition, $\mathbf{\Sigma} = \mathbf{Q\Lambda Q}'$ (eqn 10) is

```
Lambda = diag(eigen(Sig)$values)
Q = eigen(Sig)$vectors
```

$$\mathbf{\Lambda} = \begin{pmatrix} 3.328 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.369 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.369 & 0.000 \\ 0.000 & 0.000 & 0.000 & -0.026 \end{pmatrix} \quad \mathbf{Q} = \begin{pmatrix} -0.480 & 0.000 & 0.816 & -0.321 \\ -0.556 & -0.000 & 0.000 & 0.831 \\ -0.480 & -0.707 & -0.408 & -0.321 \\ -0.480 & 0.707 & -0.408 & -0.321 \end{pmatrix} \quad \text{eqn S.4}$$

The eigenvectors, $\mathbf{v}_i; i = 1, \ldots, 4$, in $\mathbf{Q} = [\mathbf{v}_1|\mathbf{v}_2|\mathbf{v}_3|\mathbf{v}_4]$ are orthonormal, which means that $\mathbf{v}_i'\mathbf{v}_j = 0$ if $i \neq j$, but $\mathbf{v}_i'\mathbf{v}_i = 1$.

```
t(Q[,1]) %*% Q[,4]

##               [,1]
## [1,] 2.775558e-17

t(Q[,4]) %*% Q[,4]

##      [,1]
## [1,]    1
```

Now, consider 4 random variables, $\boldsymbol{Y} = \{Y_1, Y_2, Y_3, Y_4\}$. The linear combination $\mathbf{v}_4'\boldsymbol{Y} = -0.321Y_1 + 0.831Y_2 - 0.321Y_3 - 0.321Y_4$ is a perfectly valid construction, and must have a positive variance. However, if $\boldsymbol{Y}$ has covariance matrix $\mathbf{\Sigma}$ in eqn S.3, then $\text{var}(\mathbf{v}_4'\boldsymbol{Y}) = \mathbf{v}_4'\mathbf{\Sigma}\mathbf{v}_4 = -0.026$, which is the 4th eigenvalue,

```
v4 = Q[,4]
t(v4) %*% Sig %*% v4

##              [,1]
## [1,] -0.02611639
```

which is not a valid variance, so $\mathbf{\Sigma}$ in eqn S.3 is not a valid covariance matrix.

To show how this works for kriging, consider predicting the location shown with the open circle in Figure S1, which is 3/10 of the way from location 2 to location 4. Then the distance from the 4 locations with solid circles in Figure S1 to the prediction location is the vector $(1.3, 0.3, 1.3, 0.7)$, and the covariances between the prediction location and the 4 locations with solid circles in Figure S1 is

```
cvec = exp(-(c(1.3, 0.3, 1.3, 0.7)/3)^2)
cvec

## [1] 0.8287989 0.9900498 0.8287989 0.9470111
```

Using eqn 7, the prediction variance of the location with the open circle, using data from the locations with the solid black circles, would be computed as

```
(1 + 0.01) - t(cvec) %*% solve(Sig) %*% cvec +
  (1 - (sum(solve(Sig) %*% cvec))^2)/sum(solve(Sig))

##             [,1]
## [1,] -0.0425027
```

which is negative, so we see that the larger matrix, where $\mathbf{\Sigma}$ is appended with covariances that include the prediction location, eqn 9, is not a valid covariance matrix.