

₁ Kriging Models for Linear Networks and non-Euclidean Distances:
₂ Cautions and Solutions

₃ 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

₄ November 12, 2017

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. (2017) 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. (2017), 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.

27

28

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

31

32 **INTRODUCTION**

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

40 **A Quick Review of Kriging**

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

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

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

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

55 Let \mathbf{R} be an autocorrelation matrix among spatial locations. All of the diagonal elements of
56 \mathbf{R} are ones. The off-diagonal element in the i th row and j th column of \mathbf{R} is the correlation, which
57 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
58 just a scaled autocorrelation matrix that includes an overall variance, σ_p^2 . In constructing kriging
59 models, practitioners often include a “nugget” effect, which is an independent (uncorrelated)
60 random effect, ε_i , in eqn 1 with variance σ_0^2 . The nugget effect is often ascribed to measurement
61 error, or microscale variation, at a scale finer than the closest measurements in the data set.
62 Constructing a full covariance matrix for a kriging model generally yields

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

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

67 $d_{i,j}$, between sites i and j , are

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

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

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

85 Kriging is often expressed as semivariograms rather than autocorrelation models.
 86 Semivariograms model the variance of the *difference* among variables. If Y_i and Y_j are random
 87 variables at spatial locations i and j , respectively, a semivariogram is defined as
 88 $\gamma(d_{i,j}) \equiv E(Y_i - Y_j)^2/2$, where E is expectation. All of the models in eqn 3 can be written as
 89 semivariograms,

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

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

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

93 Autocorrelation needs to be estimated from data. Empirical semivariograms have been used
 94 since the origins of kriging. First, all pairwise distances are binned into distance classes,
 95 $\mathcal{D}_k = [h_{k-1}, h_k)$, where $0 \leq h_0 < h_1$ and $h_{k-1} < h_k$ for $k = 1, 2, \dots, K$, that partition the real line
 96 into mutually exclusive and exhaustive segments that cover all distances in the data set. Then the
 97 empirical semivariogram is,

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

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

103 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
 104 weighted least squares (WLS) or a modification that puts increased weight near the origin
 105 (CWLS) (Cressie, 1985). This concept is generalized by restricted maximum likelihood (REML,
 106 Patterson & Thompson, 1971, 1974), which can be used for autocorrelation in regression models
 107 with several covariates and regression coefficients (for REML applied to spatial models, see, e.g.,
 108 Cressie, 1993, p. 93). In addition, using REML eliminates the arbitrary binning of distances for
 109 semivariogram estimation. Although REML was originally derived assuming normality, REML
 110 can be viewed as unbiased estimating equations (Heyde, 1994; Cressie & Lahiri, 1996), so
 111 normality is not required to estimate covariance parameters. Later, I will use WLS, CWLS, and
 112 REML for estimation, and full details are given in Supporting Information. No matter how the
 113 parameters are estimated, I focus on covariance matrices Σ (eqn 2), rather than semivariogram
 114 matrices, because Σ is more readily understood in the broader context of statistical models.

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

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

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

123 used in Σ . The prediction variance (the expected squared error that was minimized) is given by

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

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

126 **The Problem**

127 One of the properties shared by all models in eqn 3 is that, when $d_{i,j}$ is Euclidean distance, the
128 covariance matrix in eqn 2 is guaranteed to be positive definite for all possible spatial
129 configurations of points (in 3 dimensions or less) and all possible parameter values:

130 $\sigma_p^2 \geq 0, \sigma_0^2 \geq 0$, and $\alpha \geq 0$ (one of σ_p^2 or σ_0^2 must be greater than zero). It is important for Σ to be
131 positive definite because many estimators and predictors in statistics are linear functions of the
132 data, the kriging predictor being one of them. That is, let ω be a nonnull vector of weights and \mathbf{Y}
133 be a vector of random variables with covariance matrix Σ . Then an estimator or predictor
134 $\hat{T} = \omega' \mathbf{Y}$ will have variance

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

135 which is guaranteed to be positive only if Σ is positive definite (Guillot et al., 2014). Requiring Σ
136 to be positive definite is the matrix analog of requiring a variance parameter to be positive.

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

142 matrix between the observed and prediction locations, and $\Sigma_{p,p}$ be the covariance matrix among
 143 the prediction locations. Then

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

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

148 It is also worth noting that if any square submatrix of Σ^* (eqn 9) (formed by removing full
 149 columns and rows with a corresponding index) is not positive definite, then neither is the larger
 150 matrix. The implications are that, if the observed data have a covariance matrix that is not
 151 positive definite, then Σ^* will not be positive definite. However, even if the observed data (eqn 9)
 152 have a covariance matrix that is positive definite, there is no guarantee that the larger matrix,
 153 Σ^* , will be positive definite without a proper model to ensure it.

154 The simplest way to check whether a matrix is positive definite is to check the eigenvalues
 155 of that matrix. A covariance matrix Σ should be composed of real values, and it should be
 156 symmetric. Then

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

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

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

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

161 than zero and at least one v_i^2 is greater than zero. So, if the smallest eigenvalue of Σ is greater
162 than zero, then Σ is positive definite.

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

176 Another example is provided by the spatial locations at the nodes of a dichotomous network
177 (Figure 2c). The distance between each location and the nearest node is exactly one, and there
178 are $2^7 - 1$ locations. Again, let distance be defined as the shortest path between any two
179 locations, so the maximum distance between two terminal locations is $2 \times 6 = 12$. Using the
180 127×127 distance matrix with the autocorrelation models in eqn 3 for various α values showed
181 that all models yielded minimum eigenvalues below zero except the exponential model
182 (Figure 2d). The hole effect model illustrates how erratic the positive-definite condition can be,
183 where small changes in α cause wild swings on whether the covariance matrix is positive definite.
184 An argument on why the exponential model is always positive definite for the dichotomous

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

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

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

206 **Literature Review**

207 There are now many examples where autocovariance models, such as those in eqn 3, have been
208 used incorrectly with non-Euclidean distances, and they have been roundly criticized (Curriero,
209 2006). For example, for streams, impermissible models have been used by Cressie & Majure (1997)
210 and Gardner et al. (2003), who substituted in-stream distance for Euclidean distance, and in fact
211 this same idea was inappropriately recommended in Okabe & Sugihara (2012). Alternatively,
212 permissible models that guarantee positive-definite covariance matrices were developed (based on
213 a spatial moving averages, a spatially continuous analog of moving average models in times series)
214 by Ver Hoef et al. (2006), Cressie et al. (2006) and Ver Hoef & Peterson (2010).

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

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

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

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

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

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

241 REANALYSIS OF LADLE ET AL. (2017)

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

245 Review of Ladle et al. (2017)

246 Ladle et al. (2017) provide an interesting study of human activity along a linear network of trails
247 in a portion of Alberta’s Rocky Mountains. They analyzed both motorised and non-motorised
248 activities; see Figure 4a for the trails and study area. They use a two stage analysis, first fitting a
249 mixed-effects logistic regression model to the presence of any activity during hourly increments.
250 The fixed effects in their models include rainfall, date, time of day, etc. Random effects for spatial
251 location and time were also included, and estimated as best linear unbiased predictions (BLUPs).

252 These BLUPs were subsequently used in a second stage of analysis as spatial data. Linear
253 distance among BLUPs was used in place of Euclidean distance, and ordinary kriging was used to
254 predict BLUPs at unsampled locations along the linear network; see Figure 4 in Ladle et al.
255 (2017). In all that follows, I will re-analyze only the non-motorised data from Ladle et al. (2017),
256 using the estimated BLUP values and the linear-network and Euclidean distance matrices that
257 they provided.

258 The main objective of this paper, and my prior review, is that substitution of
259 non-Euclidean distance metrics into autocorrelation models derived for Euclidean distance can
260 create covariance matrices that are not positive definite. For the particular case of Ladle et al.
261 (2017), using their linear-network distance matrix in the models given in eqn 3 showed that none
262 of the models are permissible beyond a certain α value (Figure 3a). On the other hand, using the
263 Euclidean distance matrix provided by Ladle et al. (2016), all models yield positive-definite
264 covariance matrices at all values of $\alpha > 0$ (Figure 3b), which simply verifies that they are
265 permissible models. Note that the fitted exponential model had $\hat{\alpha} = 14245$ in Ladle et al. (2017)
266 for nonmotorised variables, which yielded a positive-definite covariance matrix because $\alpha < 28224$
267 had all positive eigenvalues (Figure 3a). The (incorrectly) fitted spherical models in Ladle et al.
268 (2017) had estimated range parameters $> 40,000$, which would not yield positive-definite
269 covariance matrices because $\alpha > 15876$ had negative eigenvalues (Figure 3a).

270 **Review of Non-Euclidean Distance Models**

271 Several approaches can be used for creating spatial models in novel situations, whether for
272 non-Euclidean distances or other situations. The first is the spatial moving average, also called a
273 process convolution and autoconvolution. The spatial moving average approach is very similar to
274 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 known as 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. (2017). It has been mostly featured as a method for big data sets (e.g.

²⁹⁹ Wikle & Cressie, 1999; Ruppert et al., 2003; Cressie & Johannesson, 2008; Banerjee et al., 2008).

³⁰⁰ I will use this method for models using linear network distances, which I describe next.

³⁰¹ **Reduced-Rank Methods for Non-Euclidean Distances**

³⁰² The reduced-rank models are a special case of linear mixed models, so I provide a quick review.

³⁰³ In fact, eqn 1 is a special case of a mixed model. A mixed model is often written as

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

³⁰⁴ where \mathbf{X} is a design matrix with covariates, $\boldsymbol{\beta}$ is a vector of regression parameters, \mathbf{W} is a

³⁰⁵ random-effects design matrix, $\boldsymbol{\nu}$ is a vector of zero-mean random effects with variance σ_p^2 , and

³⁰⁶ $\text{var}(\boldsymbol{\varepsilon}) = \sigma_0^2\mathbf{I}$. In statistical textbooks, \mathbf{W} in eqn 11 often contains dummy variables (zeros or

³⁰⁷ ones) that indicate some factor level of the random effect. However, \mathbf{W} can also contain

³⁰⁸ covariates, in which case $\boldsymbol{\nu}$ contains random effects for the slope of a line, illustrating that there

³⁰⁹ are no restrictions on the types of values (continuous or categorical) contained in \mathbf{W} . For the

³¹⁰ linear mixed model, eqn 11, recall that

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

³¹¹ where \mathbf{G} is the correlation matrix for $\boldsymbol{\nu}$. Classically, for mixed models, random effects are

³¹² assumed independent, so $\mathbf{G} = \mathbf{I}$, and then $\text{var}(\mathbf{Y}) = \sigma_p^2\mathbf{WW}' + \sigma_0^2\mathbf{I}$.

³¹³ For the reduced-rank models, let \mathbf{D} denote a matrix of Euclidean distances among locations

³¹⁴ and \mathbf{L} denote a matrix of linear network distances. Let $\mathbf{R}_{m,\mathbf{A},\alpha}$ be a spatial autocorrelation

³¹⁵ matrix, where $m = e, s, g, c$, or h , for exponential, spherical, Gaussian, Cauchy, or hole effect,

³¹⁶ 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

317 range parameter for one of the models in eqn 3. For example, $\mathbf{R}_{e,\mathbf{L},\alpha} = \exp(-\mathbf{L}/\alpha)$. Then let
 318 $\mathbf{R}_{m,\mathbf{A},\alpha}^r$ be the matrix where some of the columns of $\mathbf{R}_{m,\mathbf{A},\alpha}$ are kept as “knots”, and all other
 319 columns have been removed; hence the term “reduced-rank.” For example, for the Ladle et al.
 320 (2016) data, there are 239 locations, so $\mathbf{R}_{m,\mathbf{A},\alpha}$ is 239×239 . I will reduce it to just 120 columns,
 321 so $\mathbf{R}_{m,\mathbf{A},\alpha}^r$ is 239×120 .

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

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

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

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

335 The innovations for reduced-rank spatial models in eqn 13 occur because: 1) we use
 336 correlation models of distance in the random effects design matrix, essentially $\mathbf{W} = \mathbf{R}_{m,\mathbf{A},\alpha}^r$, and
 337 2) we also allow the random effects $\boldsymbol{\nu}$ to be spatially autocorrelated using the *inverse* covariance

338 matrix from one of the models in eqn 3. The model in eqn 13 must have a positive-definite
 339 covariance matrix, so I assume Euclidean distance will be used for the distance among knots. In
 340 that case, eqn 13 leads to the following covariance matrix,

$$\Sigma = \sigma_p^2 \mathbf{R}_{m,\mathbf{A},\alpha}^r [\mathbf{R}_{m,\mathbf{D}^k,\eta}]^{-1} [\mathbf{R}_{m,\mathbf{A},\alpha}^r]' + \sigma_0^2 \mathbf{I} \quad \text{eqn 14}$$

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

- 345 1. Strictly speaking, the covariance matrix in eqn 14 is guaranteed to be positive definite only
 346 if $\sigma_0^2 > 0$. This is no different than mixed models, eqn 11, where recall that the variance was
 347 $\sigma_p^2 \mathbf{W} \mathbf{G} \mathbf{W}' + \sigma_0^2 \mathbf{I}$.
- 348 2. Note that the inverse of a positive-definite matrix will also be positive definite, so
 349 $[\mathbf{R}_{m,\mathbf{D}^k,\eta}]^{-1}$ is positive definite as long as Euclidean distance \mathbf{D}^k is used. That ensures that
 350 $\sigma_p^2 \mathbf{R}_{m,\mathbf{A},\alpha}^r [\mathbf{R}_{m,\mathbf{D}^k,\eta}]^{-1} [\mathbf{R}_{m,\mathbf{A},\alpha}^r]'$ is nonnegative definite.
- 351 3. It might seem unusual to model the covariance among the knots as the inverse $[\mathbf{R}_{m,\mathbf{D}^k,\eta}]^{-1}$.
 352 The reasons for the inverse are complex (Banerjee et al., 2008), but there is an intuitive
 353 explanation. Suppose that the reduced-rank matrix is based on Euclidean distance, that is,
 354 let $\mathbf{A} = \mathbf{D}$, so we have $\mathbf{R}_{m,\mathbf{D},\alpha}^r$. Now, let the knots increase in number until the knots
 355 become exactly the same as the observed locations. Then, $\mathbf{R}_{m,\mathbf{D},\alpha}^r$ becomes $\mathbf{R}_{m,\mathbf{D},\alpha}$, the full
 356 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
 357 same model type and distance matrix, η is equivalent to α), the inverse of the full
 358 covariance matrix. The inverse cancels one of the full covariance matrices, so in eqn 14,

359 $\sigma_p^2 \mathbf{R}_{m,\mathbf{D},\alpha} [\mathbf{R}_{m,\mathbf{D},\alpha}]^{-1} [\mathbf{R}_{m,\mathbf{D},\alpha}]' = \sigma_p^2 \mathbf{R}_{m,\mathbf{D},\alpha}$, which is the $n \times n$ symmetric covariance matrix
 360 without any reduction in rank. By using the inverse, the formulation in eqn 14 allows us to
 361 recover a typical covariance matrix as the knots become equal to the observed locations. My
 362 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
 363 matrix could be used for \mathbf{G} , including $\mathbf{G} = \mathbf{I}$.

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

367 5. In addition to allowing non-Euclidean distances in the random-effects design matrix,
 368 $\mathbf{R}_{m,\mathbf{A},\alpha}^r$, there is a computational advantage to using rank reduction in eqn 14. Notice that
 369 Σ is a 239×239 matrix, and likelihood based methods (such as maximum likelihood, or
 370 restricted maximum likelihood) require the inverse of Σ . Computing matrix inverses is
 371 computationally expensive, and grows exponentially with the dimension of the matrix (as a
 372 cube of the number of locations). However, the reduced-rank formulation allows an inverse
 373 of Σ that is reduced to the size of the rank reduction by using the
 374 Sherman-Morrison-Woodbury result (Sherman & Morrison, 1949; Woodbury, 1950); see an
 375 excellent review by Henderson & Searle (1981). In our case, if we choose 120 knots, then the
 376 inverse would be for a 120×120 matrix rather than a 239×239 matrix.

377 In what follows, I will always choose a single model form, m , across all 3 components of
 378 $\mathbf{R}_{m,\mathbf{A},\alpha}^r [\mathbf{R}_{m,\mathbf{D}^k,\eta}]^{-1} [\mathbf{R}_{m,\mathbf{A},\alpha}^r]'$, and I will always use the linear network distance matrix \mathbf{L} for \mathbf{A} ,
 379 but allow the autocorrelation parameter α to be different from η . For example, the reduced-rank

380 exponential model that uses linear network distance has a covariance matrix

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

381 For this covariance matrix, there are 4 parameters to estimate; σ_p^2 , α , η , and σ_0^2 . In what follows,

382 I fit all reduced-rank models using REML.

383 Reanalysis of the Ladle et al. (2016) Data

384 The reanalysis of Ladle et al. (2017) is given in Table 1. The data were downloaded from the

385 Dryad Repository <http://dx.doi.org/10.5061/dryad.62t17>. To evaluate models, I use four criteria,

386 the first being AIC (Akaike, 1973; Burnham & Anderson, 2002), which assumes that the data were

387 distributed as a multivariate normal likelihood with a spatial covariance matrix (for an example

388 using spatial models, see Hoeting et al., 2006). AIC was only used when fitting with REML.

389 The rest of the criteria are based on leave-one-out crossvalidation. Let \mathbf{y}_{-i} be the vector of

390 observed data with the i th observation removed. Then, using \mathbf{y}_{-i} and the estimated covariance

391 matrix, with the i th row and column removed, the i th observation is predicted, denoted as \hat{Y}_i ,

392 with eqn 6, and its prediction standard error, denoted as $se(\hat{Y}_i)$, is estimated with (the square

393 root of) eqn 7. The correlation was computed on the set of pairs $\{(y_i, \hat{Y}_i); i = 1, \dots, n\}$ for all i

394 and reported as Corr in Table 1. Root-mean-squared prediction error (RMSPE, Table 1) was

395 computed as the square root of the mean of $(y_i - \hat{Y}_i)^2$ for all i . The coverage of the 90%

396 prediction interval (CI90, Table 1) was the proportion of times that the interval

397 $[\hat{Y}_i - 1.645se(\hat{Y}_i), \hat{Y}_i + 1.645se(\hat{Y}_i)]$ contained the true value y_i for all i .

398 First, I consider the fitted exponential model reported in Ladle et al. (2017) (the first row in

399 Table 1). The fitted model, which did not have a nugget effect, along with the empirical

400 semivariogram, are shown as the dashed line for the exponential model in Figure 5. Of particular
401 interest is the fact that the CI90 for the model in Ladle et al. (2017) covers the true value only
402 69.9% of the time (Table 1). This is due to the lack of a nugget effect. The covariance matrix is
403 forcing high autocorrelation among sites that are close together, and hence the prediction
404 variance assumes prediction is better than it really is, which results in estimated prediction
405 standard errors that are too small. When semivariograms are fitted without a nugget effect, they
406 should be checked carefully for fitting and prediction instabilities. Models without nugget can
407 lead to computational instability when inverting the covariance matrix (Diamond & Armstrong,
408 1984; Posa, 1989; O'Dowd, 1991; Ababou et al., 1994). If the modeler insists on excluding the
409 nugget effect (as often occurs when using kriging to approximate deterministic computer models,
410 e.g. Martin & Simpson, 2005), a small nugget effect can be added to the diagonal (e.g. 1×10^{-6}
411 was used in Booker et al. (1999)) to improve computational stability. Problems can occur due to
412 model type (Gaussian autocorrelation is the worst) and the arrangement of the spatial locations,
413 when “near duplicate” locations can cause apparently singular matrices for computational
414 purposes (Bivand et al., 2008, p. 220).

415 I fit all other models in eqn 3, both with and without a nugget effect, where linear network
416 distance was used in place of Euclidean distance. These form rows 2-10 in Table 1). REML was
417 not used to fit these models because REML depends on the inverse of the covariance matrix,
418 which was unstable for these models because their covariance matrices were not positive definite.
419 For models without a nugget, CWLS, which adds weight to empirical semivariogram values with
420 smaller distances, provided poor fits due to the lack of congruence between the model being
421 forced to zero at the origin, and the empirical semivariogram values. Thus, all models without a
422 nugget effect were fitted by WLS, and all models with a nugget effect were fitted with
423 CWLS(Table 1, Figure 5a). The results show that, other than the exponential model, all fitted

424 models without a nugget effect had negative eigenvalues and, when using cross-validation,
425 produced substantial numbers of negative values for prediction standard errors when using eqn 7
426 (31 for spherical, 97 for Gaussian, 121 for Cauchy, and 125 for Hole-effect). Adding a nugget
427 effect helped, but only exponential, spherical, and Cauchy models had positive-definite covariance
428 matrices. However, CI90 for all three models were well below the nominal 90% level. Of
429 particular interest is the hole-effect model with a nugget effect. It would appear to have the best
430 fit visually (Figure 5), yet even when a nugget effect is included, it produced a cross-validation
431 prediction with a negative prediction standard error (Table 1).

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

438 Four models in Table 1 used the reduced-rank approach, based on exponential, spherical,
439 Gaussian, and Cauchy autocorrelation models in eqn 3 as used in eqn 14 (the hole-effect model
440 always performed poorly, so was eliminated). The estimated covariance parameters for each of the
441 models are shown in Table 1. Note that all reduced-rank models outperformed all other models in
442 terms of RMSPE, and they also had lower AIC than their Euclidean distance counterparts. CI90
443 for the reduced-rank models was always above 88%, so very close to the nominal 90%. Not only
444 were the reduced-rank models the best performers, they were all completely permissible and
445 computationally faster than the Euclidean distance models. There was little actual difference
446 among the reduced-rank models in performance.

447 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 distance and linear network distance for the data in Ladle et al. (2017). In this case, it will be very difficult to see a large advantage in linear network distance models over a Euclidean distance models, or vice versa, which is what we see in 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, were

⁴⁶⁹ the reduced rank construction is added, rather than replacing Euclidean distance, so

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

⁴⁷⁰ where the random effect \mathbf{Z} has a Euclidean distance covariance matrix. For example, I fit a model

⁴⁷¹ that has a covariance matrix

$$\boldsymbol{\Sigma} = \sigma_{\text{Euc}}^2 \mathbf{R}_{s,\mathbf{D},\phi} + \sigma_p^2 \mathbf{R}_{s,\mathbf{L},\alpha}^r [\mathbf{R}_{s,\mathbf{D}^k,\eta}]^{-1} [\mathbf{R}_{s,\mathbf{L},\alpha}^r]' + \sigma_0^2 \mathbf{I}.$$

⁴⁷² where $\mathbf{R}_{s,\mathbf{D},\phi}$ is an autocorrelation matrix based a spherical model with full Euclidean distance

⁴⁷³ matrix \mathbf{D} among all sites, range parameter ϕ , and σ_{Euc}^2 is the Euclidean distance variance

⁴⁷⁴ component. The fitted model parameters are shown as the last two rows in Table 1, with the first

⁴⁷⁵ row the linear network distance component, and the last row the Euclidean distance component.

⁴⁷⁶ Combining both linear network distance and Euclidean distance provided the best predictions

⁴⁷⁷ overall, with the lowest RMSPE and good CI90. According to AIC = 903.35, the variance

⁴⁷⁸ component model does not warrant estimating the two extra parameters because AIC was lower

⁴⁷⁹ for exponential, spherical, and Cauchy reduce-rank-only models, however cross-validation

⁴⁸⁰ summaries indicated otherwise. A variance component approach, combining covariance models

⁴⁸¹ based on linear networks, with those based on Euclidean distance, was also recommended for

⁴⁸² stream network models (Ver Hoef & Peterson, 2010), and is an intuitively appealing idea that

⁴⁸³ puts both components in the model and lets the data decide on their relative contributions.

484 **DISCUSSION AND CONCLUSIONS**

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

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

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

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

510 ACKNOWLEDGMENTS

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

515 DATA AND CODE ACCESSIBILITY

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

522 REFERENCES

- 523 Ababou, R., A. C. Bagtzoglou and E. F. Wood, 1994. On the condition number of covariance
524 matrices in kriging, estimation, and simulation of random fields. Mathematical Geology
525 **26**(1):99–133.
- 526 Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In

- 527 B. Petrov & F. Csaki, editors, Second International Symposium on Information Theory, pages
528 267–281. Akademiai Kiado, Budapest.
- 529 Bakka, H., J. Vanhatalo, J. Illian, D. Simpson and H. Rue, 2016. Accounting for physical barriers
530 in species distribution modeling with non-stationary spatial random effects. arXiv preprint
531 arXiv:1608.03787 .
- 532 Banerjee, S., 2005. On geodetic distance computations in spatial modeling. *Biometrics*
533 **61**(2):617–625.
- 534 Banerjee, S., A. E. Gelfand, A. O. Finley and H. Sang, 2008. Gaussian predictive process models
535 for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical
536 Methodology)* **70**(4):825–848.
- 537 Barry, R. P. and J. M. Ver Hoef, 1996. Blackbox kriging: Spatial prediction without specifying
538 variogram models. *Journal of Agricultural, Biological, and Environmental Statistics* **1**:297–322.
- 539 Bivand, R. S., E. J. Pebesma and V. Gomez-Rubio, 2008. Applied Spatial Data Analysis with R.
540 Springer, NY. URL <http://www.asdar-book.org/>.
- 541 Booker, A. J., J. Dennis Jr, P. D. Frank, D. B. Serafini, V. Torczon and M. W. Trosset, 1999. A
542 rigorous framework for optimization of expensive functions by surrogates. *Structural
543 Optimization* **17**(1):1–13.
- 544 Bradburd, G. S., P. L. Ralph and G. M. Coop, 2013. Disentangling the effects of geographic and
545 ecological isolation on genetic differentiation. *Evolution* **67**(11):3258–3273.
- 546 Burnham, K. P. and D. R. Anderson, 2002. Model Selection and Multimodel Inference: A
547 Practical Information-Theoretic Approach. Springer-Verlag Inc, New York.

- 548 Chiles, J.-P. and P. Delfiner, 1999. Geostatistics: Modeling Spatial Uncertainty. John Wiley &
549 Sons, New York.
- 550 Conn, P. B., D. S. Johnson, J. M. Ver Hoef, M. B. Hooten, J. M. London and P. L. Boveng, 2015.
551 Using spatiotemporal statistical models to estimate animal abundance and infer ecological
552 dynamics from survey counts. Ecological Monographs **85**(2):235–252.
- 553 Cressie, N., 1985. Fitting models by weighted least squares. Journal of the International
554 Association for Mathematical Geology **17**:563–586.
- 555 Cressie, N., 1990. The origins of kriging. Mathematical Geology **22**:239–252.
- 556 Cressie, N., J. Frey, B. Harch and M. Smith, 2006. Spatial prediction on a river network. Journal
557 of Agricultural, Biological, and Environmental Statistics **11**(2):127–150.
- 558 Cressie, N. and G. Johannesson, 2008. Fixed rank kriging for very large spatial data sets. Journal
559 of the Royal Statistical Society, Series B **70**(1):209–226.
- 560 Cressie, N. and S. N. Lahiri, 1996. Asymptotics for REML estimation of spatial covariance
561 parameters. Journal of Statistical Planning and Inference **50**:327–341.
- 562 Cressie, N. and J. J. Majure, 1997. Spatio-temporal statistical modeling of livestock waste in
563 streams. Journal of Agricultural, Biological, and Environmental Statistics pages 24–47.
- 564 Cressie, N. A. C., 1993. Statistics for Spatial Data, Revised Edition. John Wiley & Sons, New
565 York.
- 566 Curriero, F. C., 2006. On the use of non-euclidean distance measures in geostatistics.
567 Mathematical Geology **38**(8):907–926.

- 568 Dale, M. and M.-J. Fortin, 2010. From graphs to spatial graphs. *Annual Review of Ecology,*
569 *Evolution, and Systematics* **41**(1):21.
- 570 Diamond, P. and M. Armstrong, 1984. Robustness of variograms and conditioning of kriging
571 matrices. *Mathematical Geology* **16**(8):809–822.
- 572 Ecker, M. D. and A. E. Gelfand, 1997. Bayesian variogram modeling for an isotropic spatial
573 process. *Journal of Agricultural, Biological, and Environmental Statistics* pages 347–369.
- 574 Gandin, L. S., 1963. *Objective Analysis of Meteorological Fields*, volume 242.
575 Gidrometeorologicheskoe Izdatel'stvo (GIMIZ), Leningrad, (translated by Israel Program for
576 Scientific Translations Jerusalem, 1965).
- 577 Gardner, B., P. J. Sullivan and A. J. Lembo Jr., 2003. Predicting stream temperatures:
578 geostatistical model comparison using alternative distance metrics. *Canadian Journal of
579 Fisheries and Aquatic Sciences* **60**:344–351.
- 580 Gelfand, A. E., S. Banerjee and A. O. Finley, 2012. Spatial design for knot selection in
581 knot-based dimension reduction models. *Spatio-Temporal Design: Advances in Efficient Data
582 Acquisition* pages 142–169.
- 583 Gneiting, T., 2013. Strictly and non-strictly positive definite functions on spheres. *Bernoulli*
584 **19**(4):1327–1349.
- 585 Goovaerts, P., 1997. *Geostatistics for Natural Resources Evaluation*. Oxford University Press,
586 New York, NY.
- 587 Guillot, G., R. L. Schilling, E. Porcu and M. Bevilacqua, 2014. Validity of covariance models for
588 the analysis of geographical variation. *Methods in Ecology and Evolution* **5**(4):329–335.

- 589 Hefley, T. J., K. M. Broms, B. M. Brost, F. E. Buderman, S. L. Kay, H. R. Scharf, J. R. Tipton,
590 P. J. Williams and M. B. Hooten, 2016. The basis function approach for modeling
591 autocorrelation in ecological data. *Ecology* page Wiley Online Library.
- 592 Henderson, H. and S. R. Searle, 1981. On deriving the inverse of a sum of matrices. *SIAM Review*
593 **50**:53–60.
- 594 Heyde, C. C., 1994. A quasi-likelihood approach to the REML estimating equations. *Statistics &*
595 *Probability Letters* **21**:381–384.
- 596 Higdon, D., 1998. A process-convolution approach to modelling temperatures in the North
597 Atlantic Ocean (Disc: P191-192). *Environmental and Ecological Statistics* **5**:173–190.
- 598 Higdon, D., J. Swall and J. Kern, 1999. Non-stationary spatial modeling. In J. M. Bernardo,
599 J. O. Berger, A. P. Dawid & A. Smith, editors, *Bayesian Statistics 6 – Proceedings of the Sixth*
600 *Valencia International Meeting*, pages 761–768. Clarendon Press [Oxford University Press].
- 601 Hoeting, J. A., R. A. Davis, A. A. Merton and S. E. Thompson, 2006. Model selection for
602 geostatistical models. *Ecological Applications* **16**(1):87–98.
- 603 Jensen, O. P., M. C. Christman and T. J. Miller, 2006. Landscape-based geostatistics: a case
604 study of the distribution of blue crab in chesapeake bay. *Environmetrics* **17**(6):605–621.
- 605 Journel, A. G. and C. W. Huijbregts, 1978. *Mining Geostatistics*. Academic Press, London, UK.
- 606 Kaluzny, S. P., S. C. Vega, T. P. Cardoso and A. A. Shelly, 1998. Analyzing geostatistical data.
607 In *S+SpatialStats: Users Manual for Windows and UNIX*, pages 67–109. Springer New York,
608 New York, NY.

- 609 Ladle, A., T. Avgar, M. Wheatley and M. S. Boyce, 2017. Predictive modelling of ecological
610 patterns along linear-feature networks. *Methods in Ecology and Evolution* **8**(3):329–338.
- 611 Lin, G.-F. and L.-H. Chen, 2004. A spatial interpolation method based on radial basis function
612 networks incorporating a semivariogram model. *Journal of Hydrology* **288**(3):288–298.
- 613 Lindgren, F., H. Rue and J. Lindström, 2011. An explicit link between Gaussian fields and
614 Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal
615 of the Royal Statistical Society: Series B (Statistical Methodology)* **73**(4):423–498.
- 616 Little, L. S., D. Edwards and D. E. Porter, 1997. Kriging in estuaries: as the crow flies, or as the
617 fish swims? *Journal of Experimental Marine Biology and Ecology* **213**(1):1–11.
- 618 MacQueen, J. B., 1967. Some methods for classification and analysis of multivariate observations.
619 In L. M. L. Cam & J. Neyman, editors, *Proc. of the fifth Berkeley Symposium on Mathematical
620 Statistics and Probability*, volume 1, pages 281–297. University of California Press.
- 621 Martin, J. D. and T. W. Simpson, 2005. Use of kriging models to approximate deterministic
622 computer models. *AIAA journal* **43**(4):853–863.
- 623 Matheron, G., 1963. Principles of geostatistics. *Economic Geology* **58**:1246–1266.
- 624 McIntyre, J. and R. P. Barry, 2017. A lattice-based smoother for regions with irregular
625 boundaries and holes. *Journal of Computational and Graphical Statistics*, **early
626 on-line**:<http://dx.doi.org/10.1080/10618600.2017.1375935>.
- 627 Miller, D. L. and S. N. Wood, 2014. Finite area smoothing with generalized distance splines.
628 *Environmental and ecological statistics* **21**(4):715–731.

- 629 O'Dowd, R., 1991. Conditioning of coefficient matrices of ordinary kriging. Mathematical
630 Geology **23**(5):721–739.
- 631 Okabe, A. and K. Sugihara, 2012. Spatial Analysis Along Networks: Statistical and
632 Computational Methods. John Wiley & Sons.
- 633 Patterson, H. and R. Thompson, 1974. Maximum likelihood estimation of components of
634 variance. In Proceedings of the 8th International Biometric Conference, pages 197–207.
635 Biometric Society, Washington, DC.
- 636 Patterson, H. D. and R. Thompson, 1971. Recovery of inter-block information when block sizes
637 are unequal. *Biometrika* **58**:545–554.
- 638 Peterson, E. E., J. M. Ver Hoef, D. J. Isaak, J. A. Falke, M.-J. Fortin, C. Jordan, K. McNyset,
639 P. Monestiez, A. S. Ruesch, A. Sengupta et al., 2013. Stream networks in space: concepts,
640 models, and synthesis. *Ecology Letters* **16**:707–719.
- 641 Posa, D., 1989. Conditioning of the stationary kriging matrices for some well-known covariance
642 models. *Mathematical Geology* **21**(7):755–765.
- 643 R Core Team, 2017. R: A Language and Environment for Statistical Computing. R Foundation
644 for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org>.
- 645 Ramsay, T., 2002. Spline smoothing over difficult regions. *Journal of the Royal Statistical
646 Society: Series B (Statistical Methodology)* **64**(2):307–319.
- 647 Rathbun, S. L., 1998. Spatial modelling in irregularly shaped regions: kriging estuaries.
648 *Environmetrics* **9**(2):109–129.

- 649 Rue, H., S. Martino and N. Chopin, 2009. Approximate bayesian inference for latent gaussian
650 models by using integrated nested laplace approximations. Journal of the Royal Statistical
651 Society: Series B (Statistical Methodology) **71**(2):319–392. URL
652 <http://dx.doi.org/10.1111/j.1467-9868.2008.00700.x>.
- 653 Ruppert, D., M. P. Wand and R. J. Carroll, 2003. Semiparametric Regression. Cambridge
654 University Press.
- 655 Sangalli, L. M., J. O. Ramsay and T. O. Ramsay, 2013. Spatial spline regression models. Journal
656 of the Royal Statistical Society: Series B (Statistical Methodology) **75**(4):681–703.
- 657 Schabenberger, O. and C. A. Gotway, 2005. Statistical Methods for Spatial Data Analysis.
658 Chapman Hall/CRC, Boca Raton, Florida.
- 659 Selby, B. and K. M. Kockelman, 2013. Spatial prediction of traffic levels in unmeasured locations:
660 applications of universal kriging and geographically weighted regression. Journal of Transport
661 Geography **29**:24–32.
- 662 Sherman, J. and W. J. Morrison, 1949. Adjustment of an inverse matrix corresponding to changes
663 in the elements of a given column or a given row of the original matrix. Annals of
664 Mathematical Statistics **20**:621.
- 665 Shiode, N. and S. Shiode, 2011. Street-level spatial interpolation using network-based IDW and
666 ordinary kriging. Transactions in GIS **15**(4):457–477.
- 667 Ver Hoef, J. M. and R. P. Barry, 1998. Constructing and fitting models for cokriging and
668 multivariable spatial prediction. Journal of Statistical Planning and Inference **69**:275–294.
- 669 Ver Hoef, J. M., N. Cressie and R. P. Barry, 2004. Flexible spatial models for kriging and

- 670 cokriging using moving averages and the fast Fourier transform (FFT). *Journal of*
671 *Computational and Graphical Statistics* **13**(2):265–282.
- 672 Ver Hoef, J. M. and J. K. Jansen, 2015. Estimating abundance from counts in large data sets of
673 irregularly-spaced plots using spatial basis functions. *Journal of Agricultural, Biological, and*
674 *Environmental Statistics* **20**:1–27.
- 675 Ver Hoef, J. M. and E. Peterson, 2010. A moving average approach for spatial statistical models
676 of stream networks (with discussion). *Journal of the American Statistical Association* **105**:6–18.
- 677 Ver Hoef, J. M., E. E. Peterson and D. Theobald, 2006. Spatial statistical models that use flow
678 and stream distance. *Environmental and Ecological Statistics* **13**(1):449–464.
- 679 Wang, H. and M. G. Ranalli, 2007. Low-rank smoothing splines on complicated domains.
680 *Biometrics* **63**:209–217.
- 681 Webster, R. and M. A. Oliver, 2007. *Geostatistics for Environmental Scientists*. John Wiley &
682 Sons, Chichester, England.
- 683 Wikle, C. K., 2002. A kernel-based spectral model for non-gaussian spatio-temporal processes.
684 *Statistical Modelling* **2**(4):299–314.
- 685 Wikle, C. K. and N. Cressie, 1999. A dimension-reduced approach to space-time kalman filtering.
686 *Biometrika* pages 815–829.
- 687 Wood, S. N., M. V. Bravington and S. L. Hedley, 2008. Soap film smoothing. *Journal of the*
688 *Royal Statistical Society: Series B (Statistical Methodology)* **70**(5):931–955. URL
689 <http://dx.doi.org/10.1111/j.1467-9868.2008.00665.x>.

- 690 Woodbury, M. A., 1950. Inverting modified matrices. Memorandum Report 42, Statistical
691 Research Group, Princeton N.J.
- 692 Xie, Y., 2014. knitr: a comprehensive tool for reproducible research in R. In V. Stodden,
693 F. Leisch & R. D. Peng, editors, Implementing Reproducible Computational Research, pages 3
694 – 32. Chapman and Hall/CRC. URL
695 <http://www.crcpress.com/product/isbn/9781466561595>. ISBN 978-1466561595.
- 696 Xie, Y., 2015. Dynamic Documents with R and knitr. 2nd edition. Chapman and Hall/CRC,
697 Boca Raton, Florida. URL <http://yihui.name/knitr/>. ISBN 978-1498716963.
- 698 Xie, Y., 2016. knitr: A General-Purpose Package for Dynamic Report Generation in R. URL
699 <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. (2017). Models are given in eqn 3, and Y in the RR column indicates the reduced-rank version. The distance matrix used (Lin or linear, Euc for Euclidean) had 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 15. A blank indicates it was not part of the model. The column heading PD has a Y if the fitted covariance matrix is 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	14245			Y	0	0.639	1.594	0.699	
Exp		Lin	CWLS	4.9	28558		1.1	Y	0	0.672	1.483	0.866	
Sph		Lin	WLS	4.8	36393				31				
Sph		Lin	CWLS	3.6	43664		1.2	Y	0	0.659	1.507	0.858	
Gau		Lin	WLS	4.7	15710				97				
Gau		Lin	CWLS	3.2	22311		1.8		0	0.603	1.692	0.782	
Cau		Lin	WLS	5.1	12099				121				
Cau		Lin	CWLS	4.1	21812		1.7	Y	0	0.613	1.593	0.828	
Hol		Lin	WLS	4.2	7856				125				
Hol		Lin	CWLS	2.5	8912		1.8		1				
Exp		Euc	CWLS	4.6	15947		1.0	Y	0	0.664	1.496	0.883	
Sph		Euc	CWLS	3.6	29959		1.3	Y	0	0.665	1.492	0.883	
Gau		Euc	CWLS	3.1	15091		1.8	Y	0	0.640	1.537	0.866	
Cau		Euc	CWLS	3.9	14125		1.7	Y	0	0.654	1.512	0.879	
Hol		Euc	CWLS	2.5	6246		1.9	Y	0	0.617	1.573	0.866	
Exp		Euc	REML	3.0	11406		1.4	Y	0	906.71	0.665	1.492	0.900
Sph		Euc	REML	3.3	27928		1.5	Y	0	905.23	0.668	1.488	0.887
Gau		Euc	REML	2.2	8965		1.8	Y	0	907.02	0.663	1.496	0.891
Cau		Euc	REML	2.7	9496		1.8	Y	0	906.52	0.661	1.499	0.900
Hol		Euc	REML	2.0	5737		2.3	Y	0	918.30	0.621	1.567	0.912
Exp	Y	Lin	REML	1.6	12545	3368	1.3	Y	0	901.61	0.674	1.475	0.891
Sph	Y	Lin	REML	1.4	25962	9393	1.3	Y	0	902.21	0.678	1.468	0.887
Gau	Y	Lin	REML	1.2	10721	3586	1.3	Y	0	905.76	0.671	1.481	0.883
Cau	Y	Lin	REML	1.5	9768	3515	1.3	Y	0	901.58	0.674	1.476	0.891
Sph	Y	Lin	REML	0.8	16739	10078	1.4	Y	0	903.35	0.686	1.453	0.895
Sph	N	Euc		2.0	27595								

700 **FIGURES**

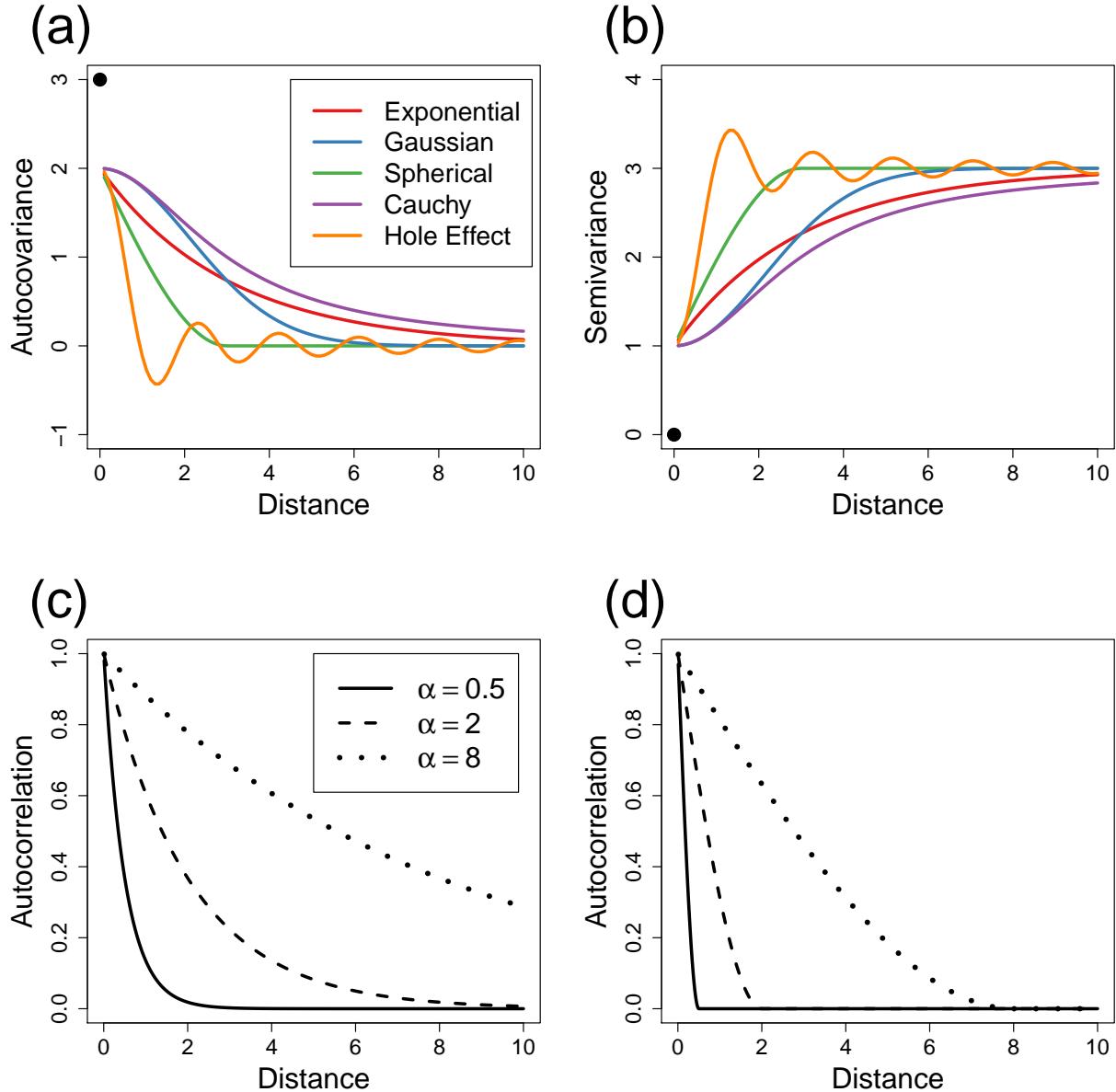


Figure 1: Autocorrelation models. (a) Autocovariance functions for various models, with a partial sill of 2 and a nugget effect of 1. (b) The same models as in (a), except represented as semivariogram models. Effect of the range parameter α on the (c) exponential model, and (d) spherical model. Note 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.

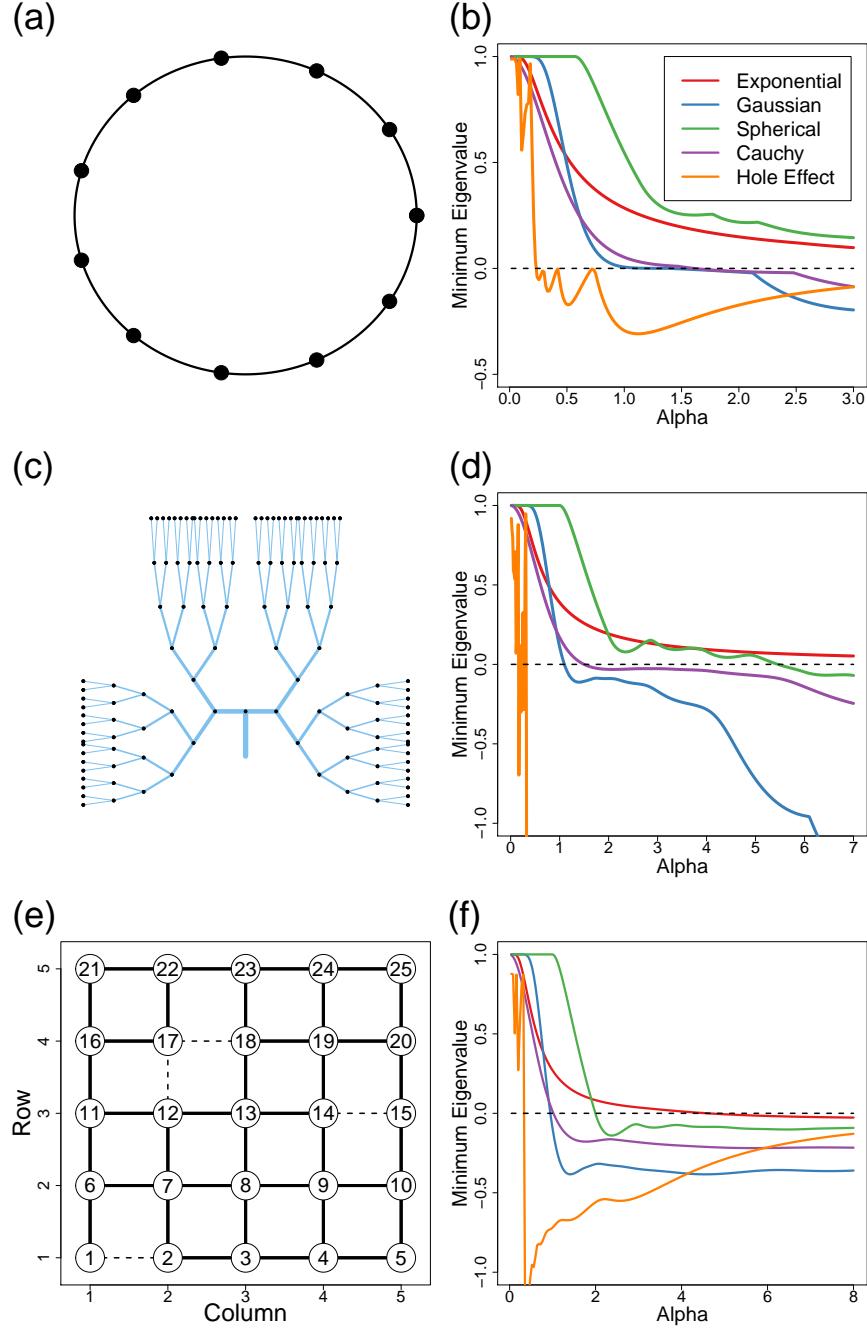


Figure 2: Cautionary examples. (a) 11 spatial locations on a circle are shown with solid circles. (b) Minimum eigenvalue for various autocorrelation models using distances on the circle. (c) A dichotomous branching network (stream) with 127 spatial locations at the node of each branch. (d) Minimum eigenvalue 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 for various autocorrelation models using shortest path distances along the irregular lattice.

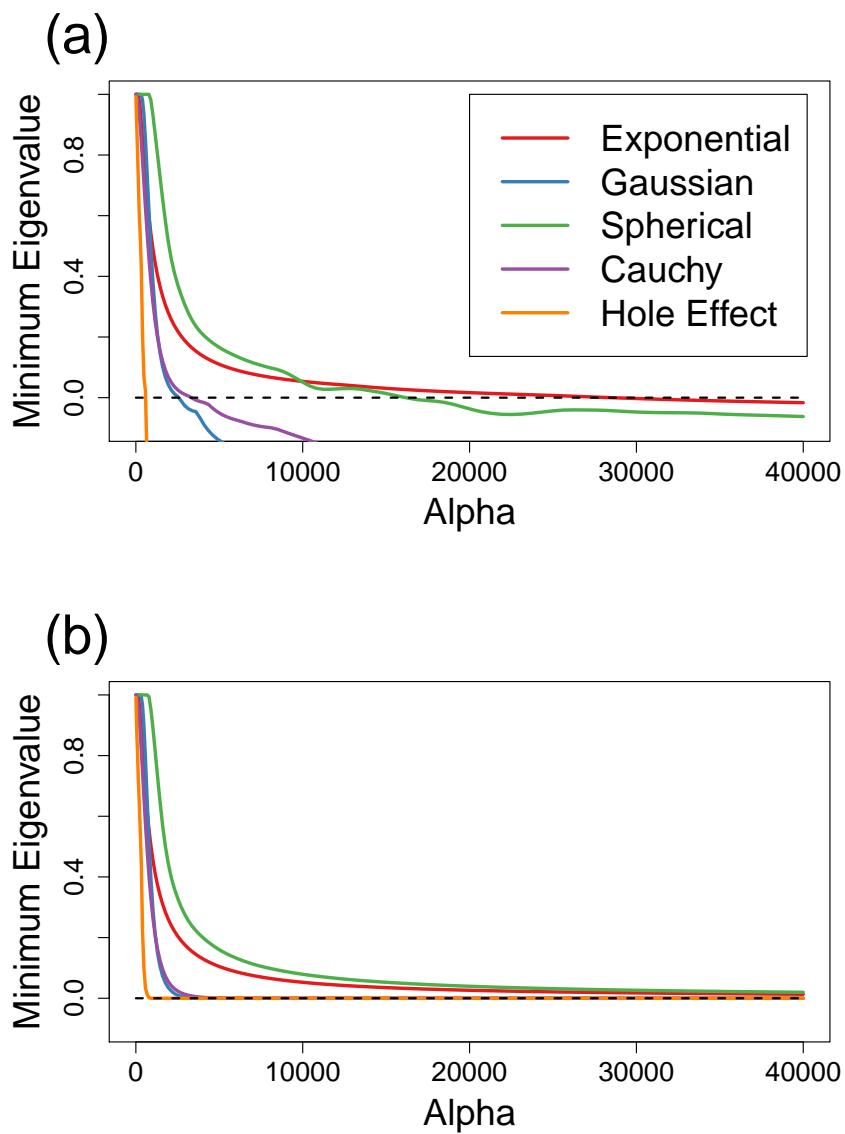


Figure 3: Minimum eigenvalues for various autocorrelation models for Ladle et al. (2016) data set.
 (a) Using linear distances among cameras. (b) Using Euclidean distances among cameras.

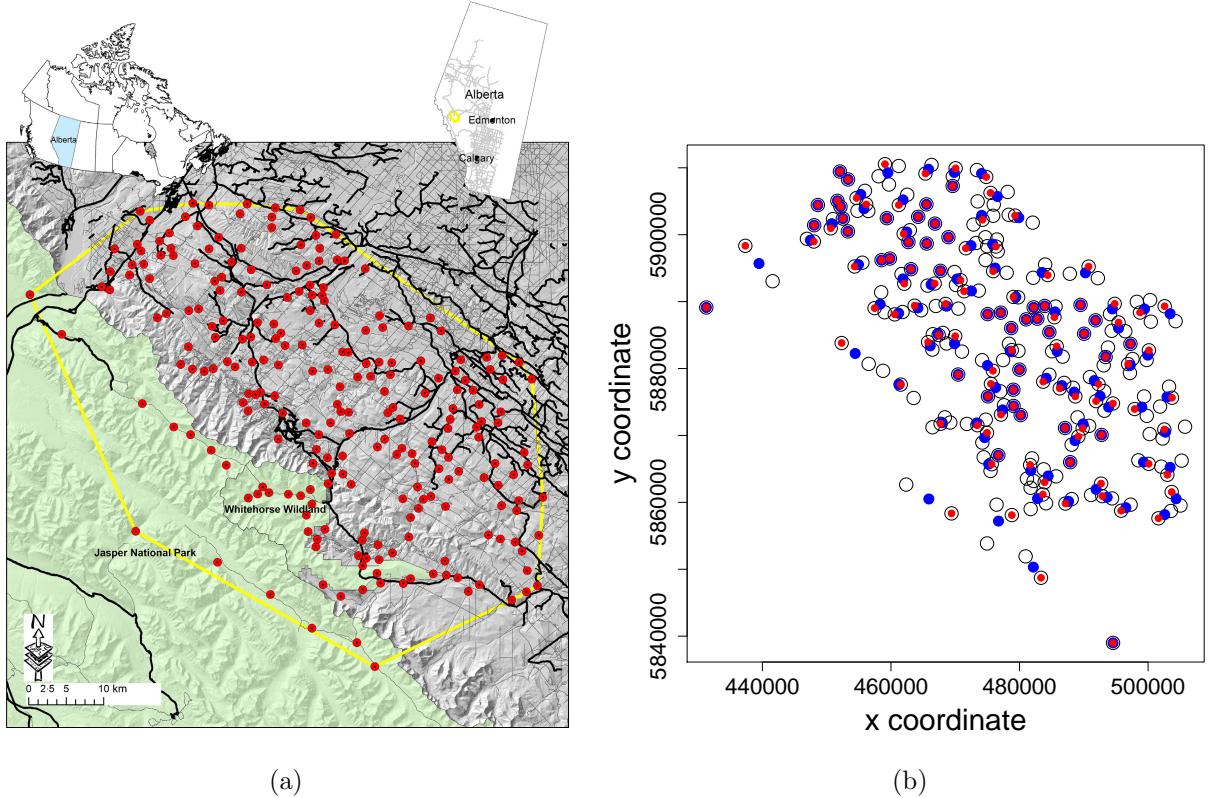


Figure 4: (a) Study area, [will seek permission to reprint], from Ladle et al. (2017). Roads are shown as black lines, trails as gray lines, and camera locations as solid red circles. (b) All spatial locations (open circles) and knot locations for reduced-rank methods. Initially, k-means on x- and y-coordinates 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), (a), and the on-line data, (b), especially along the western and southern borders. All analyses in this paper used the on-line data.

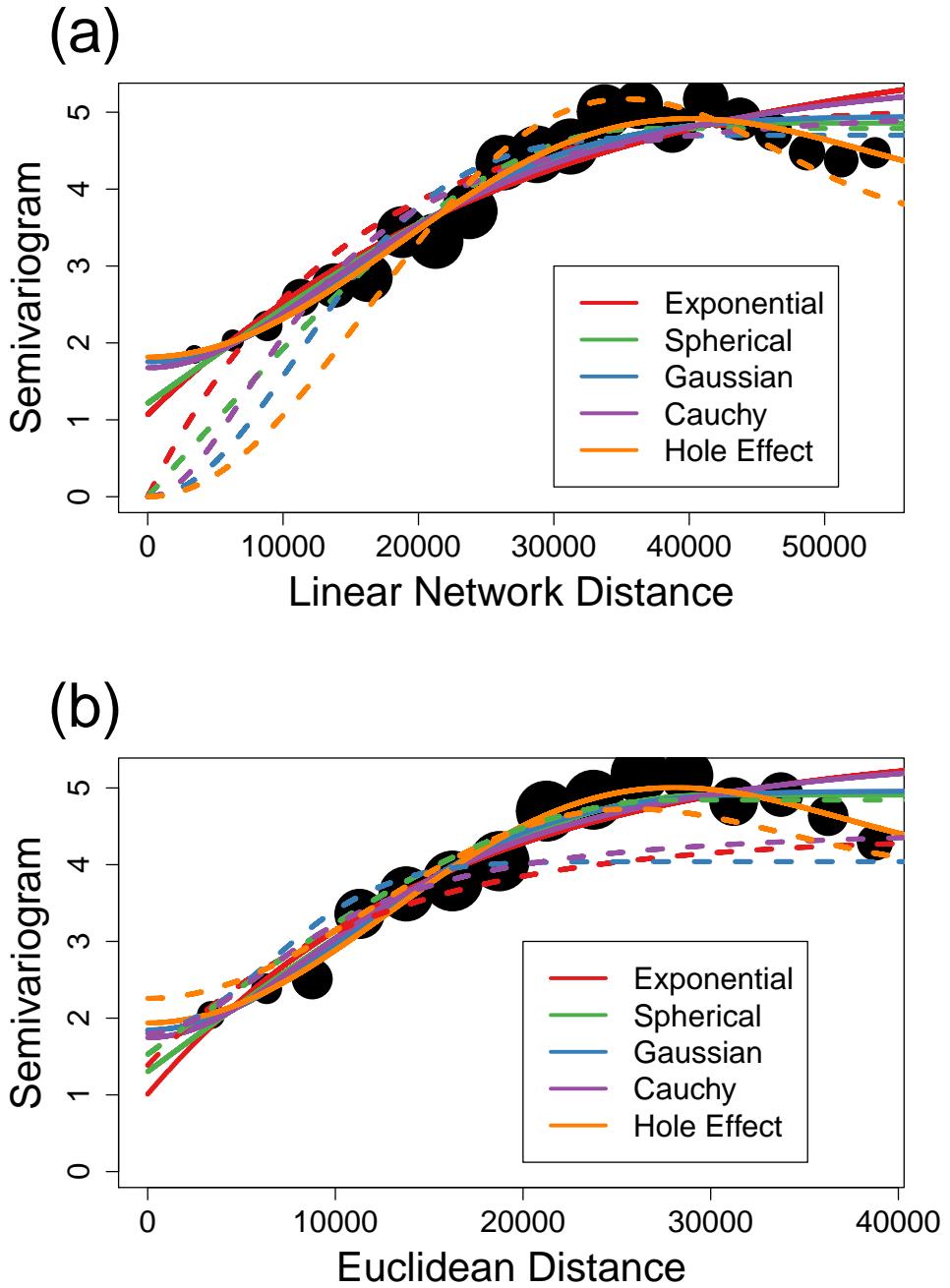


Figure 5: Empirical semivariogram with various fits. The solid black circles are empirical semivariogram values in distance classes, with size proportional to number of pairs of points in each distance class. (a) Linear network distances, where the dashed lines are fitted models without a nugget effect 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)

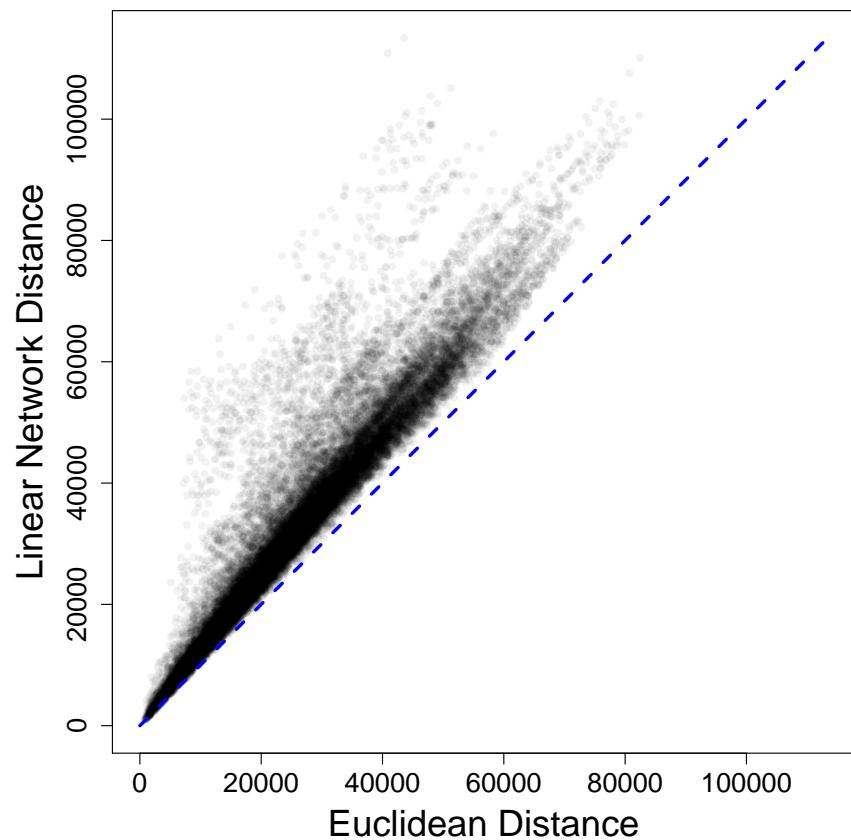


Figure 6: Scatter plot of Euclidean distance versus linear network distance for real data example. The points are semitransparent to reveal a strong correlation between distance metrics.

701 **SUPPLEMENTAL MATERIAL**

702 **Estimation Methods**

703 I use two methods to fit theoretical semivariograms eqn 3 to empirical semivariograms eqn 5. The
 704 first is simple weighted least squares. To show the dependence of the theoretical semivariogram on
 705 parameters, write any of the models, eqn 3, in semivariogram form with a nugget effect, $\gamma(h_k|\boldsymbol{\theta}) =$
 706 $\sigma_0^2 + \sigma_p^2(1 - \rho_m(h_k|\alpha))$, where $\boldsymbol{\theta} = (\sigma_p^2, \sigma_0^2, \alpha)$. Then the weighted least squares estimator of $\boldsymbol{\theta}$ is,

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

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

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

708 REML does not use an empirical semivariogram. Rather, let \mathbf{y} be a vector of observed data of
 709 length n , \mathbf{X} a fixed effects design matrix with n rows and p linearly independent columns, $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$
 710 an $n \times n$ covariance matrix, in the same order as the data, that depends on distances between
 711 observations, and a set of parameters, as given in eqn 2. Note that I show the dependence of $\boldsymbol{\Sigma}$ on
 712 $\boldsymbol{\theta}$ with a subscript. Then REML estimates are given by

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

713 where

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

⁷¹⁴ is the generalized least squares estimator of β . Note that for our case, $\mathbf{X} = \mathbf{1}$, where $\mathbf{1}$ is a vector
⁷¹⁵ of all 1s, and $\beta = \mu$, a scalar.

⁷¹⁶ Simple Example on Negative Variances from Improper Covariance Matrices

⁷¹⁷ For a very simple, worked example in R on how a covariance matrix that is not positive definite
⁷¹⁸ can lead to negative variances, consider the 4 locations in a linear network shown in Figure S1.

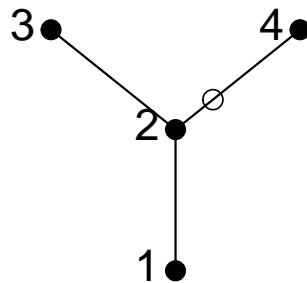


Figure S1: A simple 4-location network, where each location is given by a solid circle numbered from 1 to 4, along with a prediction location, shown by the open circle.

⁷¹⁹ Let the linear distance between each connected location be 1 unit, so the distance matrix among
⁷²⁰ 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))
```

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

⁷²¹ I will use the Gaussian autocorrelation model, eqn 3, with $\sigma_p^2 = 1$, $\alpha = 3$, and a small nugget effect,
⁷²² $\sigma_0 = 0.01$.

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

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

⁷²³ The spectral decomposition, $\Sigma = \mathbf{Q}\Lambda\mathbf{Q}'$ (eqn 10) is

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

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

⁷²⁴ The eigenvectors, $\mathbf{v}_i; i = 1, \dots, 4$, in $\mathbf{Q} = [\mathbf{v}_1 | \mathbf{v}_2 | \mathbf{v}_3 | \mathbf{v}_4]$ are orthonormal, which means that $\mathbf{v}_i' \mathbf{v}_j = 0$

⁷²⁵ if $i \neq j$, but $\mathbf{v}_i' \mathbf{v}_i = 1$.

```
Q[, 1] %*% Q[, 4]
```

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

```
Q[, 4] %*% Q[, 4]
```

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

⁷²⁶ Now, consider 4 random variables, $\mathbf{Y} = \{Y_1, Y_2, Y_3, Y_4\}$. The linear combination $\mathbf{v}_4' \mathbf{Y} = -0.321Y_1 + 0.831Y_2 - 0.321Y_3 - 0.321Y_4$ is a perfectly valid construction, and must have a positive variance. However, if \mathbf{Y} has covariance matrix Σ in eqn S.3, then $\text{var}(\mathbf{v}_4' \mathbf{Y}) = \mathbf{v}_4' \Sigma \mathbf{v}_4 = -0.026$, which is the 4th eigenvalue,

```

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

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

```

730 which is not a valid variance, so Σ in eqn S.3 is not a valid covariance matrix.

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

```

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

## [1] 0.8287989 0.9900498 0.8287989 0.9470111

```

736 Using eqn 7, the prediction variance of the location with the open circle, using data from the
 737 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

```

738 which is negative, so we see that the larger matrix, where Σ is appended with covariances that
 739 include the prediction location, eqn 9, is not a valid covariance matrix.