# Kriging Models for Linear Networks and non-Euclidean Distances: Cautions, Solutions, and a Comment on Ladle et al. (2016)

Jay M. Ver Hoef

Marine Mammal Laboratory, NOAA-NMFS Alaska Fisheries Science Center
7600 Sand Point Way NE, Seattle, WA 98115
tel: (907) 456-1995 E-mail: jay.verhoef@noaa.gov

May 9, 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.

2. I give a quick review of kriging and illustrate the problem with several fabricated examples, including locations on a circle, locations on a linear dichotomous network like streams, and locations on a linear trail or road network. I re-examine the linear network distance models from Ladle et al. (2016) and show that they are not guaranteed to have a positive definite covariance matrix.

3. I introduce the reduced rank method, also called predictive process models, fixed-rank kriging, and spatial basis functions, 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 re-analyze the data of Ladle et al. (2016), showing that their fitted models, which used linear network distance in a geostatistical model without any nugget effect, had poor predictive performance compared to a model using Euclidean distance with a nugget effect, and it also had improper coverage for the prediction intervals. The reduced rank approach using linear network distances had the best predictive performance and had proper coverage for the prediction intervals.


---


Key Words: spatial statistics, geostatistics, prediction, reduced-rank methods, predictive process models

# INTRODUCTION

The variety and sophistication of statistical methods in ecology is increasing rapidly (Touchon and McCoy, 2016). Occasionally, this leads to researchers making mistakes when proposing to extend a method without fully realizing that certain foundations and assumptions of that method are violated. There are now several examples in the ecological literature where this has happened when using non-Euclidean distance metrics for autocorrelation models used in kriging, which were developed assuming Euclidean distance. My objective is to help ecologists understand the problem and avoid this mistake. In particular, I comment on the problems with extending kriging to linear networks advocated by Ladle et al. (2016), and reanalyze their data to show a better method for kriging on linear networks.

## 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 has 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 kriging model, like the familiar regression model, can be divided into two parts: 1) the non-stochastic part (also called the fixed effects, which includes covariates and regression parameters) and 2) the stochastic part (the random errors). The ordinary kriging model is,

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

where $Y_i$ is a spatial random variable at location $i$, $i = 1, 2, \ldots, n$, with constant mean $\mu$ (the fixed effect) and random error $\varepsilon_i$. In classical statistics, such as regression, the random errors are assumed to be independent from each other, with a single variance parameter. For kriging, the independence assumption is relaxed, and the spatial distance among locations is used to model autocorrelation among 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 spatial autocorrelation is when sites closer together tend to be more similar than those that are 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 $i$th row and $j$th column of the off-diagonal elements of $\mathbf{R}$ are correlations, from minus one to one, between site $i$ and $j$. Then a covariance matrix $\mathbf{C} = \sigma_{\mathrm{p}}^2\mathbf{R}$ is just a scaled autocorrelation matrix that includes an overall variance, $\sigma_{\mathrm{p}}^2$. In constructing kriging models, practitioners often include a "nugget" effect, which is an independent (uncorrelated) random effect. Constructing a full covariance matrix for a kriging model generally yields

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

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

and $j$, are

$$\begin{aligned}
\rho_{\mathrm{e}}(d_{i,j}) &= \exp(-d_{i,j}/\alpha), \\
\rho_{\mathrm{s}}(d_{i,j}) &= [1 - 1.5(d_{i,j}/\alpha) + 0.5(d_{i,j}/\alpha)^3]\mathcal{I}(d_{i,j} < \alpha), \\
\rho_{\mathrm{g}}(d_{i,j}) &= \exp(-(d_{i,j}/\alpha)^2), \\
\rho_{\mathrm{c}}(d_{i,j}) &= 1/(1 + (d_{i,j}/\alpha)^2), \\
\rho_{\mathrm{h}}(d_{i,j}) &= (\alpha/d_{i,j})\sin(d_{i,j}/\alpha)\mathcal{I}(d_{i,j} > 0) + \mathcal{I}(d_{i,j} = 0),
\end{aligned} \qquad \text{eqn 3}$$

where distances are scaled by $\alpha > 0$, 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, $\sigma_{\mathrm{p}}^2 = 2$, and a nugget effect, $\sigma_0^2 = 1$, are shown in Figure 1a. The exponential model, $\rho_{\mathrm{e}}(d_{i,j})$, is a very popular model, and a special case of the Matern model. It approaches zero autocorrelation asymptotically. The spherical model, $\rho_{\mathrm{s}}(d_{i,j})$, is also very popular, and attains exactly zero autocorrelation at $\alpha$. Both the exponential and spherical models decrease rapidly near the origin, for short distances, whereas the Gaussian model, $\rho_{\mathrm{g}}(d_{i,j})$, decreases more slowly near the origin. This is also a special case of the Matern model, and creates very smooth spatial surfaces. The Cauchy model, $\rho_{\mathrm{c}}(d_{i,j})$ is similar to the Gaussian, but approaches zero autocorrelation very slowly. Finally, The hole effect model, $\rho_{\mathrm{h}}(d_{i,j})$ 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 and Delfiner (1999, p. 80–93).

Kriging is often expressed as variograms and semivariograms. Semivariograms model the variance of the *difference* among variables. If $Z_i$ and $Z_j$ 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,

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

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,

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

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,

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

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, $\mu$ 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 (Cressie, 1985). However, this concept is generalized by restricted maximum likelihood (REML, Patterson and 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, REML eliminates the arbitrary binning of distances for variogram estimation. Although REML was originally derived assuming normality, REML can be viewed as unbiased estimating equations (Heyde, 1994; Cressie and Lahiri, 1996), so normality is not required to estimate covariance parameters. Later, I will use REML for estimation. Also, I focus on covariances, rather than variograms, because their interpretation is more readily understood in the broader context of statistical models.

After covariance parameters are estimated from the data, kriging is the spatial prediction (interpolation) for spatial locations where data were not collected. Kriging provides best linear unbiased predictions (BLUP) in the sense of minimizing the expected squared errors between the data as predictors, and the predictand, subject to unbiasedness (on average). The ordinary kriging prediction equations, in terms of a covariance matrix (Schabenberger and Gotway, 2005, p.33), are

$$\hat{Y}_{n+\ell} = \hat{\mu} + \mathbf{c}'\mathbf{\Sigma}^{-1}(\mathbf{y} - \mathbf{1}\hat{\mu}), \qquad \text{eqn 5}$$

for $M$ predictions with locations indexed by $n+\ell$, $\ell = 1, 2, \ldots, M$. Here, $\mathbf{1}$ is a vector of ones, $\hat{\mu} = (\mathbf{1}'\mathbf{\Sigma}^{-1}\mathbf{y})/(\mathbf{1}'\mathbf{\Sigma}^{-1}\mathbf{1})$, and $\mathbf{c}$ 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 used in $\mathbf{\Sigma}$. The prediction variance (the expected squared errors that were minimized) is given by

$$\mathrm{var}(\hat{Y}_{n+\ell}) = (\sigma_p^2 + \sigma_0^2) - \mathbf{c}'\mathbf{\Sigma}^{-1}\mathbf{c} + \frac{(1 - \mathbf{1}'\mathbf{\Sigma}^{-1}\mathbf{c})^2}{\mathbf{1}'\mathbf{\Sigma}^{-1}\mathbf{1}} \qquad \text{eqn 6}$$

## The Problem

One of the properties shared by all models in eqn 3 is that, when $d_{i,j}$ is Euclidean distance (in 3 dimensions or less), 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: $\sigma_{\mathrm{p}}^2 > 0, \sigma_0^2 > 0$, and $\alpha > 0$. It is important for $\mathbf{\Sigma}$ to be positive definite because many estimators and predictors in statistics are linear functions of the data, kriging being one of them. That is, let $\boldsymbol{\omega}$ be a vector of weights and $\mathbf{y}$ be a vector of random variables with covariance matrix $\mathbf{\Sigma}$. Then an estimator or predictor $\hat{T} = \boldsymbol{\omega}'\mathbf{y}$ will have variance

$$\mathrm{var}(\hat{T}) = \boldsymbol{\omega}'\mathbf{\Sigma}\boldsymbol{\omega}, \qquad \text{eqn 7}$$

which is guaranteed to be positive only if $\mathbf{\Sigma}$ is positive definite. Requiring $\mathbf{\Sigma}$ to be positive definite is the matrix analog of requiring a variance parameter to be positive.

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

$$\mathbf{\Sigma} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}' \qquad \text{eqn 8}$$

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 8 into eqn 7 gives

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

where $\mathbf{v} = \mathbf{Q}'\boldsymbol{\omega}$. Because $v_i^2 \geqslant 0$, $\mathrm{var}(\hat{T})$ is guaranteed to be positive as long as all $\lambda_i$ 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 $11 \times 11$ distance matrix was used with autocorrelation models in eqn 3, and the minimum eigenvalue is plotted in Figure 2b. Notice that as the range parameter $\alpha$ increases, the hole effect, Gaussian, and Cauchy models have a minimum eigenvalue that is less than zero, so for these values of $\alpha$, the matrix is not positive definite, and cannot be a covariance matrix. This example points out a further problem. It appears that the exponential model and spherical model are valid models for all range values; however, this is only true for 11 points that 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 $127 \times 127$ distance matrix with the autocorrelation models in eqn 3 for various $\alpha$ values showed that all models failed to consistently yield 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 $\alpha$ causes 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 and 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 $25 \times 25$ distance matrix with the autocorrelation models in eqn 3 for various $\alpha$ values shows that none of the models are positive definite for all $\alpha$ (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 $\alpha$, and an explanation is provided by Curriero (2006).

Figure 2 demonstrates 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 and Huijbregts, 1978, p. 161). For example, Webster and 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 networks.

## Literature Review

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

For roads and trails, impermissible models have been used by Shiode and Shiode (2011), Selby and Kockelman (2013) and Ladle et al. (2016), who substitute network-based distance for Euclidean distance. However, the exponential is a permissible model for a perfect grid using Manhatton 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 used to replace Euclidean distance in Little et al. (1997), Rathbun (1998), and Jensen et al. (2006), which yields impermissible models. Instead, permissible models based on reduced-rank radial-basis functions were given by Wang and 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 and Gelfand, 1997; Kaluzny et al., 1998), which have been criticized (Banerjee, 2005). Most research has centered on geodetic, or great-circle distance. If geodetic 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. Whether this is a permissible model was examined by Guillot et al. (2014).

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 may lead to covariance functions that do not guarantee positive definite covariance matrices. I will discuss this further after a re-analysis of the data in Ladle et al. (2016).

# REANALYSIS OF LADLE ET AL. (2016)

Prior to a reanalysis of Ladle et al. (2016), I list several specific criticisms of 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. (2016).

## Criticism of Ladle et al. (2016) Analysis

TThese criticism only relate to spatial modeling and kriging used in Ladle et al. (2016). The spherical variogram model was incorrect in used in Ladle et al. (2016). Fig. 2 in Ladle et al. (2016) shows the spherical variogram going up and then back down. The correct spherical models reaches an asymptote and remains constant, as shown in Fig. 1b, and virtually all textbooks on geostatistics (Journel and Huijbregts, 1978; Isaaks and Srivastava, 1989; Cressie, 1993; Goovaerts, 1997; Chiles and Delfiner, 1999; Fortin and Dale, 2005; Webster and Oliver, 2007). Because they fit an incorrect model, none of the results for the spherical variogram model are valid.

To compare variogram fits, Ladle et al. (2016) used AIC based on an assumption that the residuals of a nonlinear least squares were independent and Gaussian. This is not valid. Every point in the empirical variogram is binned, and re-uses the same location many times, both within bins and among bins. This creates a complicated correlation structure that is not independent, even if the spatial data are independent. If the data are normally distributed, then the squared differences, under the best conditions, are chi-squared distributed, and not Gaussian. For a review, see Cressie (1993).

Ladle et al. (2016) fit models without a nugget effect, justifying the decision without examining the data and a prior belief that no nugget was present. Examination of Fig. 2 in Ladle et al. (2016) would lead most spatial statistical modelers to include a nugget effect. Moreover, when variograms are fitted without a nugget effect, they should be checked carefully for fitting and prediction instabilities. It is been well-known that models without nugget can lead to computational instability when inverting the covariance matrix (Diamond and 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 and Simpson, 2005), a small nugget effect can be added to the diagonal (e.g., $1 \times 10^{-}6$ was used inBooker et al. (1999)). 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).

The main objective of this paper, and my prior review, is 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. (2016), using their linear network distance matrix in the models given in eqn 3 showed that none of the models are permissible beyond a certain $\alpha$ 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 $\hat{\alpha} = 7620$ in Ladle et al. (2016) for motorised and $\hat{\alpha} = 14245$ for nonmotorised variables, which yielded positive definite covariance matrices because

$\alpha < 28224$ had all positive eigenvalues (Figure 3a). The (incorrectly) fitted spherical models in Ladle et al. (2016) had estimated range parameters $> 40,000$, which would not yield positive-definite covariance matrices because $\alpha > 15876$ had negative eigenvalues (Figure 3a).

## Review of Non-Euclidean Distance Models

I will review two general approaches 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 and Ver Hoef, 1996), multivariable (cokriging) models (Ver Hoef and 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 and Peterson, 2010), models on the sphere (Gneiting, 2013), and spatio-temporal models (Wikle, 2002). Using the moving average approach requires solving integrals to obtain the autocorrelation function, and while those can be tractable for stream networks when purely dichotomous branching occurs (Ver Hoef et al., 2006), they are not tractable for more general linear networks.

The second approach is a reduced rank idea, also called a dimension reduction (Wikle and Cressie, 1999) and spatial radial basis (Lin and Chen, 2004) method, which 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. (2016). It has been used for shortest path distances in estuaries(Wang and Ranalli, 2007), but it is mostly featured as a method for big data sets (e.g., Wikle and Cressie, 1999; Ruppert et al., 2003; Cressie and 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

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 $\alpha$ is the range parameter for one of the models in eqn 3. For example, $\mathbf{R}_{\mathrm{e},\mathbf{L},\alpha} = \exp(-\mathbf{L}/\alpha)$. 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, $\mathbf{R}_{m,\mathbf{A},\alpha}$ is $239 \times 239$, but we will reduce it to just 120 columns, so $\mathbf{R}^r_{m,\mathbf{A},\alpha}$ is $239 \times 120$.

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 and 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 we denote the knot-to-knot distances as $\mathbf{D}^k$.

Now consider the random effects model,

$$\mathbf{y} = \mathbf{1}\mu + [\mathbf{R}^r_{m,\mathbf{A},\alpha}]\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \qquad \text{eqn 9}$$

where $\boldsymbol{\gamma}$ is a vector of zero-mean random effects, and $\mathrm{var}(\boldsymbol{\varepsilon}) = \sigma_0^2\mathbf{I}$. The model is eqn 9, is just a generalization of eqn 1 in vector notation. It is a mixed model, which are often written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \qquad \text{eqn 10}$$

where $\mathbf{X}$ is a design matrix with covariates, $\boldsymbol{\beta}$ is a vector of regression parameters, and $\mathbf{Z}$ is a random-effects design matrix. In statistical textbooks, $\mathbf{Z}$ in eqn 10 often contains dummy variables (zeros or ones) that indicate some factor level of the random effect. However, $\mathbf{Z}$ can also contain covariates, in which case $\boldsymbol{\gamma}$ would contain random effects for the slope of a line, illustrating that there are no restrictions on the types of values contained in $\mathbf{Z}$. In eqn 9, I have replaced $\mathbf{Z}$ with $\mathbf{R}_{m,\mathbf{A},\alpha}$, and there are no covariates in $\mathbf{X}$, so $\mathbf{X}$ is a vector of ones.

For the linear mixed model, eqn 10, recall that $\text{var}(\mathbf{y}) = \sigma_{\text{p}}^2\mathbf{Z}\mathbf{C}\mathbf{Z}' + \sigma_0^2\mathbf{I}$, where $\mathbf{C}$ is the correlation matrix for $\boldsymbol{\gamma}$ and $\sigma_{\text{p}}^2$ is an overall variance for the random effects. Classically, for mixed models, random effects are assumed independent, so $\mathbf{C} = \mathbf{I}$, and then $\text{var}(\mathbf{y}) = \sigma_{\text{p}}^2\mathbf{Z}\mathbf{Z}' + \sigma_0^2\mathbf{I}$. The innovations for reduced-rank spatial models in eqn 9 occur because: 1) we use correlation models of distance in the random effects design matrix, essentially $\mathbf{Z} = \mathbf{R}^r_{m,\mathbf{A},\alpha}$, and 2) we also allow the random effects $\boldsymbol{\gamma}$ to be spatially autocorrelated using the *inverse* covariance matrix from one of the models in eqn 3. The model in eqn 9 must have a positive definite covariance matrix, so I assume Euclidean distance will be used for the distance among knots. In that case, the most general form of eqn 9 leads to,

$$\boldsymbol{\Sigma} = \sigma_{\gamma}^2\mathbf{R}^r_{m,\mathbf{A},\alpha}[\mathbf{R}_{m,\mathbf{D}^k,\eta}]^{-1}[\mathbf{R}^r_{m,\mathbf{A},\alpha}]' + \sigma_{\varepsilon}^2\mathbf{I} \qquad \text{eqn 11}$$

In fact, each model subscript $m$ in eqn 11 could be different, and $\mathbf{A}$ could be either $\mathbf{D}$ or $\mathbf{L}$, or some other matrix based on any number of distance metrics. Several comments are pertinent for eqn 11.

1. The covariance matrix in eqn 11 is guaranteed to be positive definite because of the quadratic form, similar to the variance of eqn 10, recall it was $\mathbf{ZCZ}' + \sigma^2\mathbf{I}$, which will always be positive definite if $\mathbf{C}$ is positive definite. 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.

2. It might seem strange to model the covariance among the knots as the inverse $[\mathbf{R}_{m,\mathbf{D}^k,\eta}]^{-1}$. Although any positive definite matrix could be used here, and the reasons for the inverse can get involved (Banerjee et al., 2008), some intuition can be gained. 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 $[\mathbf{R}_{m,\mathbf{D},\alpha}]^{-1}$, the inverse of the full covariance matrix, and the inverse cancels one of the full covariance matrices, so in eqn 11, $\sigma^2_{\mathrm{p}}\mathbf{R}_{m,\mathbf{D},\alpha}[\mathbf{R}_{m,\mathbf{D},\alpha}]^{-1}[\mathbf{R}_{m,\mathbf{D},\alpha}]' = \sigma^2_{\mathrm{p}}\mathbf{R}_{m,\mathbf{D},\alpha}$, which is the $n \times n$ symmetric covariance matrix without any reduction in rank. By using the inverse, the formulation in eqn 11 allows us to recover a typical covariance matrix as the knots become equal to the observed locations.

3. 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 eqn 11. Notice that $\mathbf{\Sigma}$ is a $239 \times 239$ 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 and Morrison, 1949; Woodbury, 1950); see an excellent review by Henderson and Searle (1981). In our case, if we choose 120 knots, than the inverse would be for a $120 \times 120$ matrix rather than a $239 \times 239$ matrix.

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

$$\mathbf{\Sigma} = \sigma^2_{\mathrm{p}}\mathbf{R}^r_{\mathrm{e},\mathbf{L},\alpha}[\mathbf{R}_{\mathrm{e},\mathbf{D}^k,\eta}]^{-1}[\mathbf{R}^r_{\mathrm{e},\mathbf{L},\alpha}]' + \sigma^2_0\mathbf{I}. \qquad \text{eqn 12}$$

For this covariance matrix, there are 4 parameters to estimate; $\sigma^2_p$, $\alpha$, $\rho$, and $\sigma^2_0$. In what follows, I fit all reduced rank models using REML.

## Reanalysis of the Ladle et al. (2014) Data

The reanalysis of Ladle et al. (2016) is given in Table 1. The parameter estimates for the two exponential models found in Ladle et al. (2016) for motorised and nonmotorised variables are given in the first row. To evaluate models, I use four criteria, the first being AIC (Akaike, 1973; Burnham and 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).

The rest of the criteria are based on leave-one-out crossvalidation. Let $\mathbf{y}_{-i}$ be the vector of observed data with the $i$th observation removed. Then, using $\mathbf{y}_{-i}$ and the estimated covariance matrix, the $i$th observation is predicted, denoted as $\hat{y}_i$, with eqn 5, and its prediction stardard error, denoted as $se(\hat{y}_i)$, is estimated with eqn 6. The correlation was computed on the set $\{y_i, \hat{y}_i\}$

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 percentage 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. (2016) (model Ladle in Table 1). Note that Ladle et al. (2016) also used correlation between predicted and observed for leave-one-out crossvalidation. Using their model, I do not get exactly the same correlation for the motorised variable as Ladle et al. (2016), where they report 0.472, and I obtained 0.491; however, I obtain exactly the same correlation result for non-motorised (0.639). Of particular interest is the fact that the CI90 for the model in Ladle et al. (2016) covers the true value only 74.5% of the time for the motorised variable, and only 69.9% of the time for the non-motorised variable (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, assuming prediction is better than it really is, which results in estimated prediction errors that are too small.

For all of the rest of the fits, I used REML. The empirical semivariograms in Ladle et al. (2016) clearly show that there should be a nugget effect in the model. I refit the exponential model with linear network distance, but I added a nugget effect and used REML (model LinEN in Table 1) . The nugget effect was estimated to be substantial, being more than 50% of the partial sill (1.45/1.66 for motorised, and 1.19/1.75 for non-motorised). By every cross-validation metric, model linEN did a much better job at prediction than model Ladle (Ladle et al., 2016). The correlation between observed and predicted was higher, the RMSPE was lower, and the 90% prediction interval covered the true value 89.1% of the time, much closer to the nominal 90%. Note that this method is not recommended because linear network distance is not permissible in models designed for Euclidean distance. It merely illustrates that a nugget effect should be included in the models.

Fitting a model with Euclidean distance (model EucEN in Table 1) showed that it performed slightly better than model LinEN for both motorised and non-motorised variables based on AIC, Corr, and RMSPE, and much better than the original Ladle model. The 90% prediction intervals appear to be very accurate, covering the true value 90% of the time in both cases.

The final four models in Table 1 used the reduced rank approach, based on exponential, spherical, Gaussian, and Cauchy autocorrelation models, labeled as RRexp, RRsph, RRgau, and RRcau, respectively, using the covariance matrix shown in eqn 12. The estimated covariance parameters for each of the models are shown in Table 1 for both motorised and non-motorised variables. For the motorised variable, RRcau had the highest Corr value and lowest RMSPE among all models, although RRexp had the lowest AIC. For the non-motorised variable, RRcau had the lowest AIC and RMSPE, and RRsph had the highest correlation. In general, the Ladle model performed worst, with LinEN and EucEN better than Ladle and very similar to each other, but the best models were RRexp, RRsph, RRgau, and RRcau. So not only were the reduced rank models the best performers, they were all completely permissible and computationally faster than the full rank models. There was little actual difference among the reduced rank models in performance.

# DISCUSSION AND CONCLUSIONS

If one is going to promote a statistical method, there are several things that are incumbent on the author. First, the method should be shown to be better than the method it is supposed to replace. In the case of the data in Ladle et al. (2016), there is no benefit to using linear network distance compared to Euclidean distance for models LinEN and EucEN, according to any of the cross-validation statistics (Table 1). While linear network distance may make intuitive sense, if the data exist, there is some obligation to do a comparison. For example, for stream networks, several papers show linear distance models are better than Euclidean distance in a variety of ways (Peterson et al., 2013; Isaak et al., 2014; Rushworth et al., 2015). Secondly, an estimator/predictor is intimately tied to a variance estimate of that estimator/predictor. Statistics is a discipline for modeling uncertainty, and that uncertainty is captured by the standard error estimate. The standard error estimate should appropriately reflect that uncertainty. The model presented by Ladle et al. (2016) did not have proper prediction interval coverage, whose actual coverage was between 70 and 75% for the 90% interval (Table 1). This is easy to check with cross-validation. It is generally advisable to add a nugget effect to geostatistical models and let the data decide how large it should be.

While it is possible fit impermissible models such as Ladle and LinEN (Table 1) and then check the fitted model to ensure that the covariance matrix is positive definite, this practice is discouraged in traditional geostatistics. First, the fitting method itself may be susceptible to irregularities. For example, the hole effect model in Fig. 2 oscillates wildly. An optimization routine that depends on the inverse of the covariance matrix would behave erratically, and it would be hard to constrain any optimization to $\alpha$ (range) values that guaranteed a positive definite covariance matrix. Also, note that models Ladle and LinEN (Table 1) happened to have positive definite covariance matrices for the specific set of locations and estimated $\alpha$ values, resulting in cross-validation predictions that had positive variance estimates. However, when predicting at locations where data were not collected, a larger covariance matrix must be considered. 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

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{o,o} & \boldsymbol{\Sigma}_{o,p} \\ \boldsymbol{\Sigma}'_{o,p} & \boldsymbol{\Sigma}_{p,p} \end{pmatrix}$$

must be positive definite when making predictions at unobserved locations. This can be computationally expensive or impossible to check if there are thousands of prediction locations, as there were in Ladle et al. (2016) (it is computationally expensive to compute eigenvalues). It is much simpler, and safer, to choose permissible models/methods that guarantee positive definite covariance matrices for all spatial configurations and model parameter values.

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. (2016), there was a small benefit, by lowering RMSPE, for several of the linear network distance models (RRexp and RRcau) over Euclidean distance (EucEN) for the motorised variable (Table 1), and a more noticeable advantage for all reduced rank models for the non-motorised variable (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.

## ACKNOWLEDGMENTS

The project received financial support from the National Marine Fisheries Service, NOAA. Aerial surveys were authorized under a Marine Mammal Protection Act General Authorization (LOC No. 14590) issued to the Marine Mammal Laboratory. The findings and conclusions in the paper of the NOAA author(s) do not necessarily represent the views of the reviewers nor the National Marine Fisheries Service, NOAA. Any use of trade, product, or firm names does not imply an endorsement by the U.S. Government.

# DATA AND CODE ACCESSIBILITY

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

Akaike, H. (1973), "Information Theory and an Extension of the Maximum Likelihood Principle," in *Second International Symposium on Information Theory*, eds. Petrov, B. and Csaki, F., Budapest: Akademiai Kiado, pp. 267–281.

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

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

Barry, R. P. and Ver Hoef, J. M. (1996), "Blackbox Kriging: Spatial Prediction without Specifying Variogram Models," *Journal of Agricultural, Biological, and Environmental Statistics*, 1, 297–322.

Bivand, R. S., Pebesma, E. J., and Gomez-Rubio, V. (2008), *Applied Spatial Data Analysis with R*, Springer, NY.

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

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

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

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

Cressie, N. (1985), "Fitting Models by Weighted Least Squares," *Journal of the International Association for Mathematical Geology*, 17, 563–586.

— (1990), "The Origins of Kriging," *Mathematical Geology*, 22, 239–252.

Cressie, N., Frey, J., Harch, B., and Smith, M. (2006), "Spatial Prediction on a River Network," *Journal of Agricultural, Biological, and Environmental Statistics*, 11, 127–150.

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

Cressie, N. and Lahiri, S. N. (1996), "Asymptotics for REML Estimation of Spatial Covariance Parameters," *Journal of Statistical Planning and Inference*, 50, 327–341.

Cressie, N. and Majure, J. J. (1997), "Spatio-temporal statistical modeling of livestock waste in streams," *Journal of Agricultural, Biological, and Environmental Statistics*, 24–47.

Cressie, N. A. C. (1993), *Statistics for Spatial Data, Revised Edition*, New York: John Wiley & Sons.

Curriero, F. C. (2006), "On the use of non-Euclidean distance measures in geostatistics," *Mathematical Geology*, 38, 907–926.

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

Ecker, M. D. and Gelfand, A. E. (1997), "Bayesian variogram modeling for an isotropic spatial process," *Journal of Agricultural, Biological, and Environmental Statistics*, 347–369.

Fortin, M.-J. and Dale, M. R. T. (2005), *Spatial Analysis: A Guide for Ecologists*, Cambridge, UK: Cambridge University Press.

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

Gardner, B., Sullivan, P. J., and Lembo Jr., A. J. (2003), "Predicting Stream Temperatures:

Geostatistical Model Comparison Using Alternative Distance Metrics," *Canadian Journal of Fisheries and Aquatic Sciences*, 60, 344–351.

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

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

Goovaerts, P. (1997), *Geostatistics for Natural Resources Evaluation*, New York, NY: Oxford University Press.

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

Henderson, H. and Searle, S. R. (1981), "On Deriving the Inverse of a Sum of Matrices," *SIAM Review*, 50, 53–60.

Heyde, C. C. (1994), "A Quasi-likelihood Approach to the REML Estimating Equations," *Statistics & Probability Letters*, 21, 381–384.

Higdon, D. (1998), "A Process-convolution Approach to Modelling Temperatures in the North Atlantic Ocean (Disc: P191-192)," *Environmental and Ecological Statistics*, 5, 173–190.

Higdon, D., Swall, J., and Kern, J. (1999), "Non-stationary Spatial Modeling," in *Bayesian Statistics 6 – Proceedings of the Sixth Valencia International Meeting*, eds. Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A., Clarendon Press [Oxford University Press], pp. 761–768.

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

Isaak, D. J., Peterson, E. E., Ver Hoef, J. M., Wenger, S. J., Falke, J. A., Torgersen, C. E., Sowder, C., Steel, E. A., Fortin, M.-J., Jordan, C. E., et al. (2014), “Applications of spatial statistical network models to stream data,” *Wiley Interdisciplinary Reviews: Water*, 1, 277–294.

Isaaks, E. H. and Srivastava, R. M. (1989), *Applied Geostatistics*, New York, NY: Oxford University Press.

Jensen, O. P., Christman, M. C., and Miller, T. J. (2006), “Landscape-based geostatistics: a case study of the distribution of blue crab in Chesapeake Bay,” *Environmetrics*, 17, 605–621.

Journel, A. G. and Huijbregts, C. W. (1978), *Mining Geostatistics*, London, UK: Academic Press.

Kaluzny, S. P., Vega, S. C., Cardoso, T. P., and Shelly, A. A. (1998), “Analyzing Geostatistical Data,” in *S+SpatialStats: Users Manual for Windows and UNIX*, New York, NY: Springer New York, pp. 67–109.

Ladle, A., Avgar, T., Wheatley, M., and Boyce, M. S. (2016), “Predictive modelling of ecological patterns along linear-feature networks,” *Methods in Ecology and Evolution*, 8, 329–338.

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

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

MacQueen, J. B. (1967), “Some Methods for Classification and Analysis of MultiVariate Observations,” in *Proc. of the fifth Berkeley Symposium on Mathematical Statistics and*

*Probability*, eds. Cam, L. M. L. and Neyman, J., University of California Press, vol. 1, pp. 281–297.

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

Matheron, G. (1963), "Principles of Geostatistics," *Economic Geology*, 58, 1246–1266.

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

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

Patterson, H. and Thompson, R. (1974), "Maximum likelihood estimation of components of variance," in *Proceedings of the 8th International Biometric Conference*, Biometric Society, Washington, DC, pp. 197–207.

Patterson, H. D. and Thompson, R. (1971), "Recovery of Inter-block Information When Block Sizes Are Unequal," *Biometrika*, 58, 545–554.

Peterson, E. E., Ver Hoef, J. M., Isaak, D. J., Falke, J. A., Fortin, M.-J., Jordan, C., McNyset, K., Monestiez, P., Ruesch, A. S., Sengupta, A., Som, N., Steel, A., Theobald, D. M., Torgersen, C. E., and Wenger, S. J. (2013), "Stream networks in space: concepts, models, and synthesis," *Ecology Letters*, 16, 707–719.

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

R Core Team (2017), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

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

Ruppert, D., Wand, M. P., and Carroll, R. J. (2003), *Semiparametric Regression*, Campbridge University Press.

Rushworth, A., Peterson, E., Ver Hoef, J., and Bowman, A. (2015), "Validation and comparison of geostatistical and spline models for spatial stream networks," *Environmetrics*, 26, 327–338.

Schabenberger, O. and Gotway, C. A. (2005), *Statistical Methods for Spatial Data Analysis*, Boca Raton, Florida: Chapman Hall/CRC.

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

Sherman, J. and Morrison, W. J. (1949), "Adjustment of an Inverse Matrix Corresponding to Changes in the Elements of a Given Column or a Given Row of the Original Matrix," *Annals of Mathematical Statistics*, 20, 621.

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

Touchon, J. C. and McCoy, M. W. (2016), "The mismatch between current statistical practice and doctoral training in ecology," *Ecosphere*, 7.

Ver Hoef, J. M. and Barry, R. P. (1998), "Constructing and Fitting Models for Cokriging and Multivariable Spatial Prediction," *Journal of Statistical Planning and Inference*, 69, 275–294.

Ver Hoef, J. M., Cressie, N., and Barry, R. P. (2004), "Flexible Spatial Models for Kriging and Cokriging Using Moving Averages and the Fast Fourier Transform (fft)," *Journal of Computational and Graphical Statistics*, 13, 265–282.

Ver Hoef, J. M. and Jansen, J. K. (2015), "Estimating Abundance from Counts in Large Data Sets of Irregularly-Spaced Plots using Spatial Basis Functions," *Journal of Agricultural, Biological, and Environmental Statistics*, 20, 1–27.

Ver Hoef, J. M. and Peterson, E. (2010), "A Moving Average Approach for Spatial Statistical Models of Stream Networks (with discussion)," *Journal of the American Statistical Association*, 105, 6–18.

Ver Hoef, J. M., Peterson, E. E., and Theobald, D. (2006), "Spatial Statistical Models That Use Flow and Stream Distance," *Environmental and Ecological Statistics*, 13, 449–464.

Wang, H. and Ranalli, M. G. (2007), "Low-rank Smoothing Splines on Complicated Domains," *Biometrics*, 63, 209–217.

Webster, R. and Oliver, M. A. (2007), *Geostatistics for Environmental Scientists*, Chichester, England: John Wiley & Sons.

Wikle, C. K. (2002), "A kernel-based spectral model for non-Gaussian spatio-temporal processes," *Statistical Modelling*, 2, 299–314.

Wikle, C. K. and Cressie, N. (1999), "A dimension-reduced approach to space-time Kalman filtering," *Biometrika*, 815–829.

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 *Implementing Reproducible Computational Research*, eds. Stodden, V., Leisch, F., and Peng, R. D., Chapman and Hall/CRC, pp. 3 – 32, iSBN 978-1466561595.

— (2015), *Dynamic Documents with R and knitr*, Boca Raton, Florida: Chapman and Hall/CRC, 2nd ed., iSBN 978-1498716963.

— (2016), *knitr: A General-Purpose Package for Dynamic Report Generation in R*, r package version 1.15.1.

Table 1: Model fits and cross-validations statistics. The top part of the table is for the motorised data found in Ladle et al. (2016), and the lower part for the non-motorised. On the left of the table are parameter estimates using notation from eqn 2, eqn 3, and eqn 12. 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).

| Model | $\sigma_p^2$ | $\alpha$ | $\eta$ | $\sigma_0^2$ | AIC | Corr | RMSPE | CI90 |
|---|---|---|---|---|---|---|---|---|
| | Motorised | | | | | | | |
| Ladle[a] | 4.72 | 7620 | | | | 0.491 | 1.850 | 0.745 |
| LinEN[b] | 1.66 | 14806 | | 1.45 | 968.78 | 0.552 | 1.705 | 0.891 |
| EucEN[b] | 2.05 | 18739 | | 1.45 | 968.19 | 0.555 | 1.698 | 0.900 |
| RRexp[c] | 1.51 | 9983 | 2123 | 1.55 | 967.06 | 0.564 | 1.686 | 0.874 |
| RRsph[c] | 1.35 | 31164 | 7964 | 1.61 | 968.80 | 0.553 | 1.700 | 0.891 |
| RRgau[c] | 1.17 | 15495 | 3954 | 1.62 | 969.16 | 0.549 | 1.706 | 0.891 |
| RRcau[c] | 1.41 | 7632 | 1753 | 1.55 | 967.80 | 0.565 | 1.685 | 0.883 |
| | Non-motorised | | | | | | | |
| Ladle[a] | 5.09 | 14245 | | | | 0.639 | 1.594 | 0.699 |
| LinEN[b] | 1.75 | 18676 | | 1.19 | 899.11 | 0.662 | 1.498 | 0.883 |
| EucEN[b] | 1.73 | 11403 | | 1.18 | 904.71 | 0.665 | 1.492 | 0.900 |
| RRexp[c] | 1.58 | 12545 | 3368 | 1.32 | 899.61 | 0.674 | 1.475 | 0.891 |
| RRsph[c] | 1.44 | 25962 | 9393 | 1.33 | 900.21 | 0.678 | 1.468 | 0.887 |
| RRgau[c] | 1.20 | 10721 | 3586 | 1.35 | 903.76 | 0.671 | 1.481 | 0.883 |
| RRcau[c] | 1.48 | 9768 | 3515 | 1.33 | 899.58 | 0.674 | 1.476 | 0.891 |

[a]Model parameters reported in Ladle et al. (2016)

[b]LinEN, EucEN are classical exponential models with a nugget effect, using linear network distance and Euclidean distance, respectively, fit using REML.

[c]RRexp, RRsph, RRgau, RRcau are the reduced rank models using exponential, spherical, Gaussian, and Cauchy autocorrelation models, respectively, fit using REML.

# FIGURES

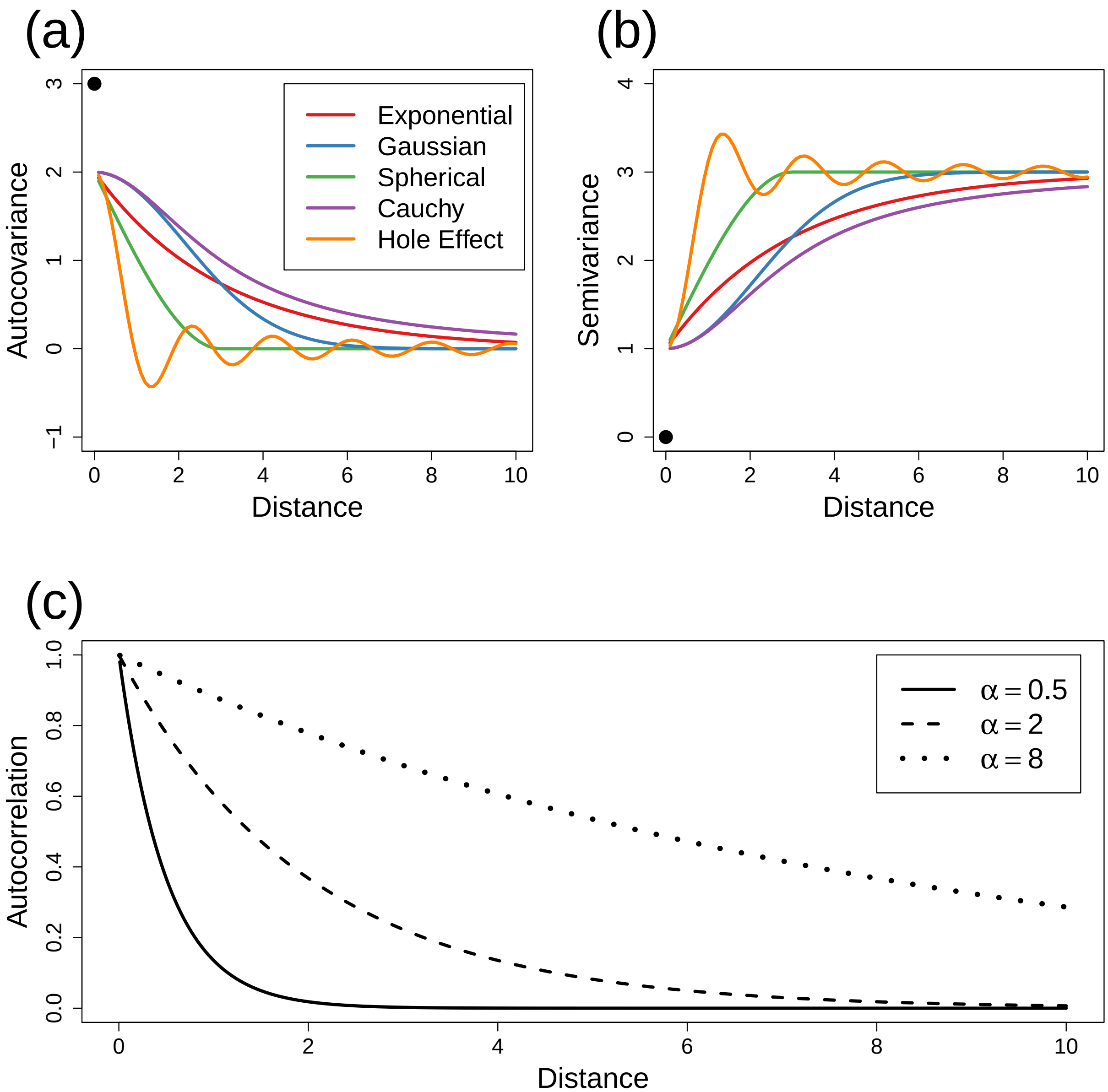


Figure 1: Autocorrelation models. (a) Autocovariance functions for various models, with a partial sill of 2 and a nugget effect of 1. (b) The same models as in (a), except represented as semivariogram models. (c) Effect of the range parameter $\alpha$ on autocorrelation functions, where the exponential model was used as an example.

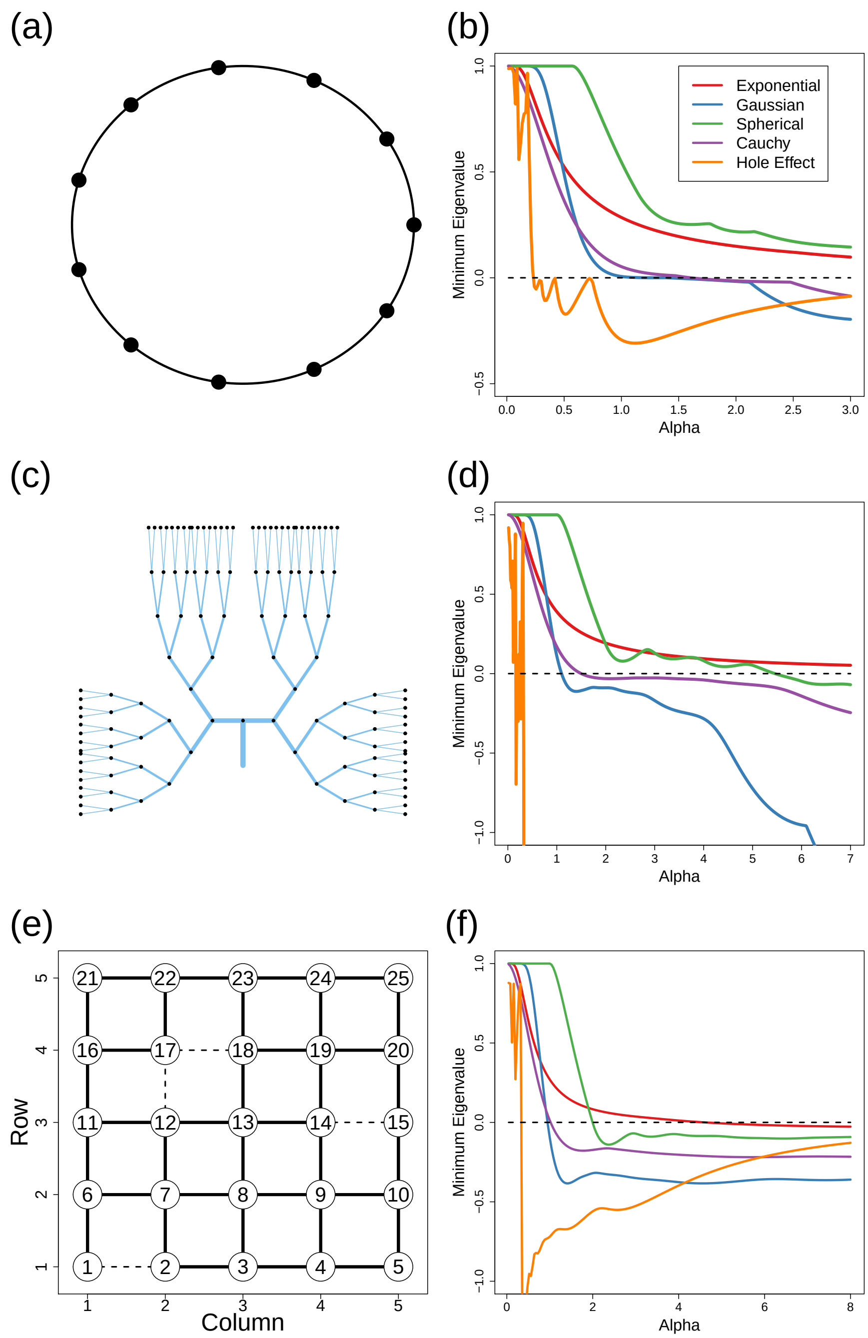


Figure 2: Cautionary examples. (a) 11 spatial locations on a circle are shown with solid circles. (b) Minimum eigenvalue for various autocorrelation models using distances on the circle. (c) A dichotomous branching network (stream) with 127 spatial locations at the node of each branch. (d) Minimum eigenvalue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

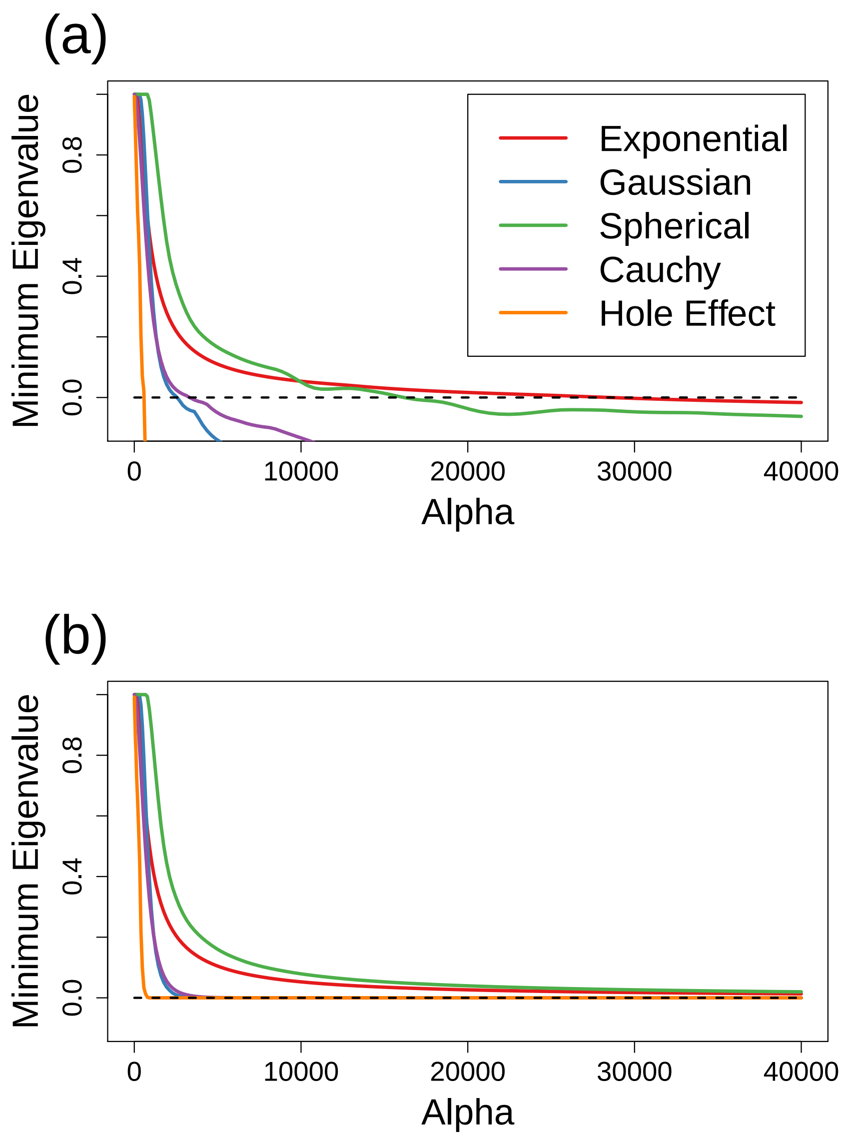


Figure 3: Minimum eigenvalues for various autocorrelation models for Ladle et al. (2016) data set. (a) Using linear distances among cameras. (b) Using Euclidean distances among cameras.

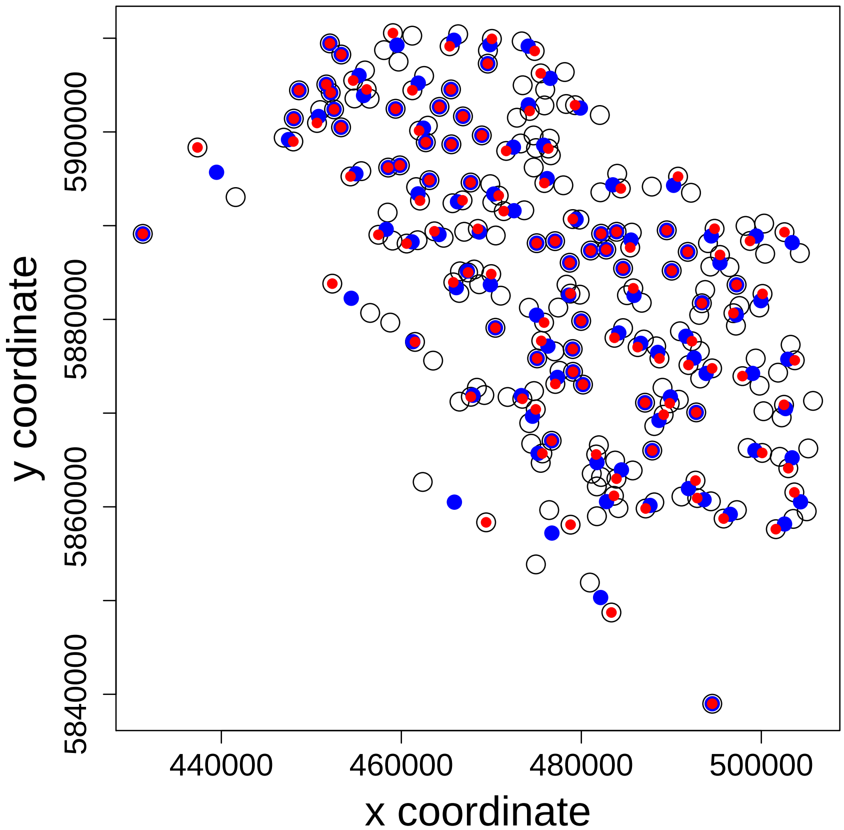


Figure 4: All spatial locations (open circles) and knot locations for reduced rank methods. Initially, k-means on x- and y-coordinates created 120 clusters with center locations given by solid blue circles, and then these were moved to nearest actual locations (solid red circles).