DEPARTMENT OF MATHEMATICS
TECHNICAL REPORT

A New Class of Semiparametric Semivariogram and Nugget Estimators

Patrick S. Carmack

October 2009

No. 2009 – 1

UNIVERSITY OF CENTRAL ARKANSAS
Conway, AR 72035

A New Class of Semiparametric Semvariogram and Nugget Estimators

Patrick S. Carmack*[a]
Jeffrey S. Spence[b]
William R. Schucany[c]
Richard F. Gunst[c]
Qihua Lin[b]
Robert W. Haley[b]

Revised October 21, 2009

[a] *Department of Mathematics, University of Central Arkansas, 201 Donaghey Avenue, Conway, AR 72035-5001, USA*

[b] *Department of Internal Medicine, Epidemiology Division, University of Texas Southwestern Medical Center at Dallas, 5323 Harry Hines Boulevard, Dallas, TX 75390-8874, USA*

[c] *Department of Statistical Science, Southern Methodist University, P.O. Box 750332, Dallas, TX 75275-0332, USA*

* Corresponding author. Department of Mathematics, University of Central Arkansas, 201 Donaghey Avenue, Conway, AR 72035-5001. Fax +1–501–450–5662
*E-mail:*patrickc@uca.edu (P.S. Carmack).

Abstract

Several authors have proposed nonparametric semivariogram estimators. Shapiro & Botha (1991) did so by application of Bochner's theorem and Cherry, Banfield & Quimby (1996) further investigated this technique where it performed favorably against parametric estimators even when data were generated under the parametric model. While this approach is sound, it lacks nugget estimation which is essential to spatial modeling and proper statistical inference. We propose a modified form of this method, which admits nugget estimation and broadens the basis. This is achieved by a simple change to the basis and an appropriate restriction of the node space as dictated by the first root of the Bessel function of the first kind of order ν. The efficacy of this new method is demonstrated via simulation. We conclude with remarks about selecting the appropriate basis and node space definition.

Key Words: Bessel basis, isotropic, node space, regular lattice, negative definiteness

1 Introduction

Let

$$\mathbf{r} \equiv \left\{ r\left(\mathbf{s}_i\right) : 1 \leq i \leq n, \mathbf{s}_i \in D \subset \mathbb{R}^k \right\} \tag{1}$$

be a sample from k-dimensional spatial process with

$$r\left(\mathbf{s}_i\right) = \mu\left(\mathbf{s}_i\right) + \delta\left(\mathbf{s}_i\right), \tag{2}$$

where $\mu\left(\cdot\right)$ is the mean function and $\delta\left(\cdot\right)$ is the spatial error process. Provided that three conditions hold,

$$\begin{aligned} \mathrm{E}\left[\delta\left(\mathbf{s}\right)\right] &= 0, && (3) \\ \mathrm{Var}\left[\delta\left(\mathbf{s}\right)\right] &= \sigma^2 < \infty, \text{and} && (4) \\ \mathrm{Cov}\left[\delta\left(\mathbf{s}_i\right), \delta\left(\mathbf{s}_j\right)\right] &= C\left(h_{ij}\right), && (5) \end{aligned}$$

where $C\left(\cdot\right)$ is a positive definite covariance function (i.e., $\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j C\left(h_{ij}\right) \geq 0, \forall\, a_i$, h_{ij}, and n), and $h_{ij} = \|\mathbf{s}_i - \mathbf{s}_j\|$, the spatial error process is said to be istropic. In other words, the variance of the process is finite, and the covariance between any two spatial location only depends on the distance between them.

Commonly, spatial modeling uses *semivariance*,

$$\gamma\left(\mathbf{s}_i, \mathbf{s}_j\right) = \frac{\mathrm{Var}\left[\delta\left(\mathbf{s}_i\right) - \delta\left(\mathbf{s}_j\right)\right]}{2}, \tag{6}$$

between spatial locations instead of covariance since this represents a richer class of models. A more rigorous treatment of spatial modeling can be found in Journel & Huijbregts (1978), Isaaks & Srivastava (1989), and Cressie (1993). When the spatial error process is isotropic, semivariance can be expressed in terms of covariance simply as $\gamma(h) = C(0) - C(h)$. Under these conditions, the *nugget* can be defined as $\lim_{h\to 0^+} \gamma(h) = \lim_{h\to 0^+} C(0) - C(h)$. Hence, the nugget represents the semivariance between locations close in space. If $C(\cdot)$ is right continuous at the origin, the nugget will be identically zero meaning the spatial error process is smooth. On the other hand, a non-zero limit implies a rough spatial error process due to irreducible source(s) of variation. This could be due to a combination of measurement error and small scale variation. In practice, the attribution to these two sources is generally unknown, so the nugget is commonly attributed to measurement error. At large distances, the semivariance between locations is known as the *sill*.

Typically, a parametric function, which ensures conditional negative definiteness, is fit to the empirical semivariogram using a non-linear algorithm. A function is said to be conditionally negative definite if $\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \gamma(h_{ij}) \leq 0, \forall\, a_i, h_{ij}$, and n such that $\sum_{i=1}^{n} a_i = 0$. It can be shown that for any positive definite function $C(\cdot)$, $\gamma(h) = C(0) - C(h)$ is necessarily conditionally negative definite. Several nonparametric semivariogram fitting procedures guaranteeing this property have been put forth in Shapiro & Botha (1991), Sampson & Guttorp (1992), and Lele (1995). We shall focus our efforts on presenting and improving upon Shapiro and Botha's approach. As posited in their paper, $C(\cdot)$ can be represented as a spectral integral

$$C(h) = \int_0^\infty \Omega_k(ht)\, dF(t) \tag{7}$$

invoking Bochner's Theorem,

$$\Omega_k(x) = \left(\frac{2}{x}\right)^{(k-2)/2} \Gamma\left(\frac{k}{2}\right) J_{(k-2)/2}(x), \tag{8}$$

where $\Gamma(\cdot)$ is the gamma function, $J_\nu(\cdot)$ is the Bessel function of the first kind of order ν, and $F(\cdot)$ is an non-decreasing bounded function for $t \geq 0$. It is worth noting that $\Omega_1(x) = \cos(x)$, $\Omega_2(x) = J_0(x)$, and $\Omega_3(x) = \sin(x)/x$, which appear as interpolators with certain optimal properties in various fields and applications. Also, k may take on non-integer values, and, if set to any value higher than the dimension of the spatial process, will still yield a valid covariance function for the dimension in question. Finally, as $k \to \infty$, $\Omega_k\left(\sqrt{2k}h\right) \to \exp\left(-h^2\right)$, which corresponds to a Gaussian spatial error process and will be of importance in modifying the method. Some plots of $\Omega_k(\cdot)$ for selected values of k are shown in Fig. 1.

Solving for the unknown $F(\cdot)$ involves a Fredholm equation of the first kind, which can be found in linear integral equation texts such as Kythe & Puri (2002). Solutions to such

equations are generally unstable, but assuming that $F(\cdot)$ is a step function transforms the integral into a finite sum where nonnegative least squares(Lawson & Hanson, 1974) can be employed to obtain a nonparametric semivariogram estimate by solving

$$\begin{aligned}\hat{\gamma}(h) &= C(0) - C(h) && (9)\\ &= \int_0^\infty (\Omega_k(0) - \Omega_k(ht))\, dF(t) && (10)\\ &= \sum_{i=1}^{m} (1 - \Omega_k(ht_i))\, p_i, && (11)\end{aligned}$$

for jumps $p_i \geq 0$, user defined nodes t_i (thus, m is implicitly user selected) by minimizing

$$\sum_{j=1}^{l} \left(\hat{\gamma}(h_j) - \tilde{\gamma}(h_j)\right)^2 \tag{12}$$

with respect to $\mathbf{p}^{\mathrm{T}} = [p_1, \ldots, p_m]$, where $\tilde{\gamma}(h) = \sum_{\|\mathbf{s}_i - \mathbf{s}_j\| = h} \frac{(\delta(\mathbf{s}_i) - \delta(\mathbf{s}_j))^2}{2N(h)}$ is the usual empirical semivariogram estimate at distance h with $N(h)$ pairs satisfying $\|\mathbf{s}_i - \mathbf{s}_j\| = h$, and l is the number of unique distances in the empirical semivariogram. This formulation readily yields a sill estimate as $\lim_{h\to\infty} \hat{\gamma}(h) = \sum_{i=1}^{m} p_i$.

As pointed out by Cherry et al. (1996), $1 - \Omega_k(\cdot)$ results in a rich semivariogram basis that compares favorably with parametric semivariogram estimators. In their article, the authors also noted a frustration with not being able to obtain estimates for the nugget since $\lim_{h\to 0^+} \hat{\gamma}(h) = 0$. As will be shown, this problem is intimately connected with the selection of the nodes; however, even careful node selection fails to completely address nugget estimation. We propose key modifications to the approach, which will admit nugget estimates and broaden this class of nonparametric semivariogram estimators.

2 Theoretical Motivation

A class of parametric isotropic semivariogram models can be represented as:

$$\gamma_\alpha(h) = \theta_1 + \theta_2 \left\{ 1 - \exp\left[-\left(\frac{h}{\theta_3}\right)^{2\alpha} \right] \right\}, \tag{13}$$

where $\gamma_\alpha(\cdot)$ is the semivariance function, h is the euclidean distance between two points in space, θ_1 is the *nugget*, θ_2 is the *partial sill*, θ_3 is the *range* parameter, and $\alpha \in [0, 1]$. When $\alpha = 0$, this results in a white noise model $(\theta_1 + \theta_2)$, $\alpha = 1/2$ is the exponential, and

$\alpha = 1$ corresponds to the Gaussian.

Recalling that $\Omega_k\left(\sqrt{2k}h\right) \to \exp\left(-h^2\right)$, one can generalize this to $\Omega_k\left(\sqrt{2k}h^\alpha\right) \to \exp\left(-h^{2\alpha}\right)$ as $k \to \infty$. Thus, this class of parametric semivariograms is the limiting case of a more general class of seimivariograms given by:

$$\gamma_\alpha(h) = \theta_1 + \theta_2\left\{1 - \Omega_k\left[\sqrt{2k}\left(\frac{h}{\theta_3}\right)^\alpha\right]\right\}. \tag{14}$$

In the spectral integral representation, this would correspond to $dF(\cdot)$ being defined as $dF\left(\sqrt{2k}/\theta_3^\alpha\right) = \theta_2$, $dF(\infty) = \theta_1$, and zero otherwise. Returning to the more general step function definition for $F(\cdot)$ in the spectral integral,

$$\begin{aligned} \gamma_\alpha(h) &= \int_0^\infty \left(\Omega_k(0) - \Omega_k\left(h^\alpha t\right)\right) dF(t) && (15) \\ &= \sum_{i=1}^m \left(1 - \Omega_k\left(h^\alpha t_i\right)\right) p_i. && (16) \end{aligned}$$

We have dispensed with the $\sqrt{2k}$ since that can be absorbed into the nodes, t_i. Thus, the only modification that has been introduced is substituting h^α for h in the original formulation. This would seem unimportant given the original basis does an excellent job fitting semivariograms from a variety of nugget-free parametrically generated data, including non-Gaussian ones; however, it is a key ingredient when considering nugget estimation. This subtle change controls how the basis behaves between the origin and the first observable empirical semivariogram value as shown in Fig. 2.

3 Methodology

3.1 Node Space Definition

Cherry et al. (1996) used an arbitrary node space spanning from 0.04 to 16.16 with 0.04 spacing for the nodes between 0.04 and 4.00 and then 0.16 spacing for nodes above 4.00 to 16.16 resulting in a total of 200 nodes. They further noted that saturating this only increased computation time and seemed to have little impact on the final fit. We propose that in addition to the change in the argument to $\Omega_k(\cdot)$, the definition of the node space is crucial for obtaining nugget estimates and stable sill estimates.

In our experience implementing their method, non-degenerate nugget estimates were not possible, as acknowledged in their paper with some possible approaches outlined in

their closing remarks. Another difficulty using this method not mentioned in their paper are the instances where the sill estimate, $\sum_{i=1}^{m} p_i$, far exceeds the maximum value in the empirical semivariogram. Our investigations found these always corresponded to very low spatial frequency nodes which caused the sill of the nonparametric semivariogram estimate to occur well beyond the maximum distance included in the fitting process.

An inspection of the $\Omega_k(\cdot)$ basis reveals why both phenomena occur. The zeros of this function correspond to where a given basis element will start to oscillate about its respective jump, p_i (Fig. 3). Thus, very high frequency nodes will start oscillating about their respective jumps before the first value in the empirical semivariogram and will be highly aliased with the nugget. With this insight, the nugget can be thought of as the jump associated with the node at infinity. Similarly, extremely low frequency nodes will not start oscillating until well beyond the hull of the empirical semivariogram (Fig. 3). The locations of these jump crossings coincide with the roots of the Bessel function of the first kind of order $\nu = \frac{k-2}{2}$, for which bounds for the first root are given in Watson (1958) by

$$\frac{\sqrt{(k-2)(k+4)}}{2} < t'_k < \sqrt{\frac{k(k+8)}{3}}, \tag{17}$$

provided $k > 10$, where t'_k denotes the first root of the Bessel function. Bounds for lower values of k exist, but we recommend using higher values of k to avoid excessively oscillatory basis elements as demonstrated by the lower order curves in Fig. 1.

Using a numerical root finding algorithm such as `uniroot.all` in the library `rootSolve` in `R` and the bounds supplied above, we use t'_k to define our node space as follows:

$$t_i = \frac{t'_k}{h_i^{\alpha}}, i = 2, \ldots, l, \tag{18}$$

where h_i is the ith distance in the empirical semivariogram. t_1 is defined to be the node at infinity (i.e., $\Omega_k(h^{\alpha} t_1) = 0$) with its respective jump, p_1, being the nugget estimate. This definition of the node space has several advantages over the previous approach. With the exception of t_1, none of the basis elements start oscillating about their respective jumps before the second unique distance in the empirical semivariogram since $ht_i < t'_k$, which implies $\Omega_k(ht_i) \neq 0$ for $h < h_2$. This eliminates high frequency nodes that are highly aliased with the nugget. If the first unique distance were included, the corresponding basis element would essentially be a constant within the hull of the data and confounded with any nugget estimate. Second, all of the nodes achieve at least one jump crossing within the hull of the data since $\Omega_k(h_i^{\alpha} t_i) = 0$, $i = 2, \ldots, l$, eliminating extremely low frequency nodes and potentially unstable sill estimates. This definition is also independent of the scale of the particular distances being employed. Finally, the placement and number of the nodes are

dictated by the unique distances in the empirical semivariogram, generally reducing computational overhead since most practical applications have fewer than 200 unique distances.

3.2 Selecting k

As Cherry et al. (1996) noted, as k increases, so does the smoothness of the basis elements. From an informal point-of-view, this makes sense given that satisfying conditionally negative definiteness imposes more and more constraints as the dimension of the spatial process increases (see Schoenberg (1938) for a more formal argument). Thus, they favored using $k = 3$, or the sinc function, over $k = 1$ (cosine) or $k = 2$ (J_0), even for lower dimension problems. We take their argument further and suggest that $k > 10$ is desirable on the same smoothness grounds plus the added benefit of sharper bounds when numerically solving for the first root of the Bessel function.

The Bessel function of the first kind has an infinite number of roots whose spacing converges to π. This means that while each basis element passes through its corresponding jump within the hull of the data using the node space defined above, it may do so several times, which could result in fits with wiggly behavior between the empirical semivariogram data points. This phenomenon is plainly evident for $k = 1$, which is a form of cosine interpolation. Even the sinc function exhibits this behavior given its slowly decaying cyclic nature. For values of $k > 10$, $\Omega_k(\cdot)$ dampens rapidly beyond its first root (Fig. 1), which greatly reduces this behavior. One might wonder if this argument should be taken to its extreme and let $k \to \infty$, which would result in using members of the exponential family as basis elements. The problem here is that as $k \to \infty$, $t'_k \to \infty$. Thus, we would forfeit our ability to precisely control when a given basis element passes through its respective jump. In addition, $\Omega_k\left(\sqrt{2k}h^{\alpha}\right)$ converges fairly rapidly to $\exp\left(-h^{2\alpha}\right)$. Thus, we will use $k = 11$ for the purpose of simulation, even though other values of k are certainly valid for the proposed method with some minor modification to the bounds for finding t'_k needed for $k \leq 10$.

3.3 Estimating α

Given the new basis and node space definitions, a natural question arises concerning how to choose α. Goodness-of-fit criteria such as Akaike's information criterion (Akaike, 1973), Bayesian information criterion (Schwarz, 1978), or generalized cross validation (Craven & Wahba, 1979) can be used to estimate α. Unfortunately, we found all three methods tended to underestimate α, so we propose the following optimization criteria:

$$\sigma^2(\alpha) = \sum_{i=1}^{l} \frac{\left(1 - \tilde{\gamma}(h_i)/\hat{\gamma}_\alpha(h_i)\right)^2}{l - \mathrm{df}\left(\hat{\gamma}_\alpha\right)}, \tag{19}$$

where $\mathrm{df}\left(\hat{\gamma}_\alpha\right)$ is the degrees of freedom of the fit, $\hat{\gamma}_\alpha\left(\cdot\right)$, obtained using nonnegative least squares (NNLS).

NNLS is similar to shrinkage methods such as ridge regression (Hoerl & Kennard, 1988) and Lasso. Zou, Hastie & Tibshirani (2004) showed that the expected number of nonzero parameters in Lasso is the degrees of freedom in the framework of Stein's unbiased risk estimation (SURE). Hence, they used the number of nonzero parameters for a particular sample as the degrees of freedom for Lasso. While this is an unbiased estimate, the degrees of freedom are now a stochastic integer quantity. In the present context, taking the number of positive parameters as the degrees of freedom for the NNLS fits is unsatisfactory since it is not a smooth function of α with jumps back and forth between consecutive integers a common occurrence. This phenomenon makes reliably minimizing the $\sigma^2\left(\cdot\right)$ curve extremely difficult (Fig 4).

Hence, we approximate the degrees of freedom of the NNLS fit using the degrees of freedom of the closest ridge regression fit. That is,

$$\mathrm{df}\left(\hat{\gamma}_\alpha\right) \approx \mathrm{tr}\left(S_{\lambda_\alpha}\right), \tag{20}$$

where

$$\lambda_\alpha = \underset{\lambda \geq 0}{\mathrm{argmin}} \left\{\mathbf{A}\left(\mathbf{A}^{\mathrm{T}}\mathbf{A} + \lambda\mathbf{I}\right)^{-1}\mathbf{A}^{\mathrm{T}}\tilde{\boldsymbol{\gamma}} - \mathbf{A}\mathbf{p}\right\}^{\mathrm{T}} \left\{\mathbf{A}\left(\mathbf{A}^{\mathrm{T}}\mathbf{A} + \lambda\mathbf{I}\right)^{-1}\mathbf{A}^{\mathrm{T}}\tilde{\boldsymbol{\gamma}} - \mathbf{A}\mathbf{p}\right\}, \tag{21}$$

and

$$S_{\lambda_\alpha} = \mathbf{A}\left(\mathbf{A}^{\mathrm{T}}\mathbf{A} + \lambda_\alpha\mathbf{I}\right)^{-1}\mathbf{A}^{\mathrm{T}} \tag{22}$$

is the ridge regression smoother matrix, $\left(\mathbf{A}\right)_{ij} = 1 - \Omega_k\left(h_i^\alpha t_j\right)$, $\tilde{\boldsymbol{\gamma}}^{\mathrm{T}} = \left[\tilde{\gamma}\left(h_1\right), \ldots, \tilde{\gamma}\left(h_l\right)\right]$ is the vector of empirical semivariogram values, and $\mathbf{p} = \underset{\mathbf{p}\geq\mathbf{0}}{\mathrm{argmin}}\left(\mathbf{A}\mathbf{p} - \tilde{\boldsymbol{\gamma}}\right)^{\mathrm{T}}\left(\mathbf{A}\mathbf{p} - \tilde{\boldsymbol{\gamma}}\right)$ is the NNLS solution vector. This approximation of degrees of freedom for the NNLS fit is a smooth function of α, making minimization of the $\sigma^2\left(\cdot\right)$ curve numerically stable (Fig. 4). Thus, α is estimated as

$$\hat{\alpha} = \underset{0\leq\alpha\leq 1}{\mathrm{argmin}} \sum_{i=1}^{l} \frac{\left(1 - \tilde{\gamma}\left(h_i\right)/\hat{\gamma}_\alpha\left(h_i\right)\right)^2}{l - \mathrm{tr}\left(S_{\lambda_\alpha}\right)}. \tag{23}$$

4 Simulations

This section will demonstrate the efficacy of the new method via simulation. Three commonly used parametric semivariograms – the exponential, spherical, and Gaussian – are chosen using various parameter combinations of $\boldsymbol{\Theta}^{\mathrm{T}} = \left[\theta_1, \theta_2, \theta_3\right]$ to generate 20×20 two-dimensional realizations. The fields are generated using the root covariance matrix method

where the eigen decomposition of $\mathbf{\Sigma} = \mathbf{V}\mathbf{D}\mathbf{V}^{\mathrm{T}}$, $(\mathbf{\Sigma})_{ij} = \theta_1 + \theta_2 - \gamma(h_{ij})$, is used to obtain $\mathbf{\Sigma}^{\frac{1}{2}} = \mathbf{V}\mathbf{D}^{\frac{1}{2}}\mathbf{V}^{\mathrm{T}}$, which is then applied to a random vector of appropriate length drawn from a $N(0,1)$ to generate a realization with the desired semivariance structure.

Our ultimate goal is to compare the performance of the new technique and these traditional parametric models in terms of nugget estimation and weighted integrated squared error (WISE). WISE is defined in the spirit of weighted least squares (Cressie, 1985) so that lower lags receive more weight than later ones:

$$\mathrm{WISE}(\hat{\gamma}) = \int_0^{h_l} \frac{1}{\gamma(h)^2} \left(\gamma(h) - \hat{\gamma}(h)\right)^2 dh = \int_0^{h_l} \left(1 - \frac{\hat{\gamma}(h)}{\gamma(h)}\right)^2 dh, \tag{24}$$

where $\gamma(\cdot)$ is the true semivariogram, and $\hat{\gamma}(\cdot)$ is the estimated semivariogram using either the new method or one of the parametric models. This measure is also known as integrated squared relative error. Along with listing the forms of the parametric semivariograms, Table 1 also has a column for α. Extensive simulation has shown these fixed values of α to be the best at estimating the true parametric semivariogram in terms of WISE for a wide variety of choices for $\mathbf{\Theta}$.

For each random realization, the proposed nonparametric method both estimating α and using the corresponding fixed value of α, and the three parametric models were fit using weighted least squares. The parametric fits used the Nelder-Mead optimization algorithm in the `R` function `optim`. In contrast to popular alternatives like the Broyden-Fletcher-Goldfarb-Shannon (BFGS) optimization algorithm, we were able to obtain convergence for every fit using the generally slower Nelder-Mead algorithm. The initial value of $\mathbf{\Theta}$ for optimization was given as the minimizer of WISE between the model being fit and the true model. The nugget estimate and WISE were recorded for each fit.

Nuggets (θ_1) were set at either 10% or 30%, and ranges (θ_3) were set to either medium or long. Medium range was defined as 1 for the exponential, $\sqrt{3}$ for the Gaussian, and 3 for the spherical, giving each the same effective range of 3. The long range was set at 2 for the exponential, $2\sqrt{3}$ for the Gaussian, and 6 for the spherical resulting in an effective range of 6 for all three parametric semivariograms. The partial sills were set at $\theta_2 = 1 - \theta_1$ to achieve a unit sill. Each of the four combinations of nugget and range for each parametric semivariogram was run 1,000 times each. The maximum lag included in the empirical semivariogram was $3\sqrt{5}$, which is approximately 1/4 the maximum distance. We have run many more combinations of nuggets, ranges, parametric functions such as the rational quadratic, and three dimensional fields, but choose these particular ones to conserve space while still presenting a wide variety of semivariograms.

Figs. 5 - 7 show boxplots for the nugget estimates and WISE for the nonparametric and parametric fits. An examination of the estimated nuggets using the fixed value of α

as indicated in Table 1 shows this approach is generally competitive with, and sometimes better than, the fits from the parametric form generating the field. This is not terribly surprising since fixing α is tantamount to knowing the true form of the model. A similar pattern emerges examining the WISE boxplots where using a fixed value of α is again competitive with, or better than, the fits using the correct parametric function. The misspecified parametric models did not compare favorably to using a fixed α in terms of nugget estimation with the same observation generally applying to WISE.

Turning to the results when estimating α, the nugget estimates do not perform as well as using a fixed α, but they are better than the misspecified models and are arguably competitive with the true parametric function. The degraded performance comes as no surprise since the method is trying to determine how to approach the origin in a entirely data driven way. Estimating α for the Gaussian runs tends to underestimate the nugget, which makes sense given α has an upper bound of 1 and the fixed $\alpha = 1$ nugget estimates also exhibit negative bias. Also, the nugget estimates seem to exhibit a negative bias when the range is medium and the nugget is high across all three parametric forms. Overall, the nugget estimates perform better at the longer range where the method has more lags to lock onto the underlying parametric form. In terms of WISE, estimating α does not do as well as using a fixed value, but is again competitive with or better than fitting the true and misspecified models, respectively.

5 Discussion

Nonparametric semivariograms offer a powerful alternative to choosing a parametric form for spatial modeling. While others have laid the groundbreaking work in terms of applying Bochner's theorem and demonstrating the efficacy of nonparametric semivariograms in a nugget-free setting, we have proposed several key modifications to improve and extend the method. First, a more flexible basis was introduced by replacing the argument of $\Omega_k(\cdot)$ by h^α. Then, careful consideration was given to the definition of the node space to make nugget estimation feasible and to ensure stable sill estimation. Finally, a method for estimating α was set forth making automated nonparametric semivariogram and nugget estimation possible.

The simulations using either a fixed or estimated α demonstrate that the new method is competitive with fits from the true parametric form, while outperforming the misspecified models. The new method, especially when estimating α, admittedly exhibits biased nugget estimates, but much less so than the misspecified models, which speaks to its robustness.

The rank of $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ is one when $\alpha = 0$ and increases as $\alpha \to 1$, which translates into

more usable basis elements as α increases. This is likely the reason Cherry et al. (1996) noted that only three or four basis elements are typically used despite having a saturated node space. Hence, the sum of squared error is a generally decreasing function of α, which necessitates estimating the degrees of freedom for model selection. The new method also needs medium ranges to obtain reliable nugget estimates, which is why we have included the option of using a fixed value for α. While using a fixed value still allows for nonparametric modeling for short ranged semivariograms, the decision is again in the hands of the modeler instead of being data driven. The solution to both these problems may ultimately be solved via the node space, as we discuss in the next paragraphs.

The definition of the node space set forth in this paper is not unique, and we experimented with several different approaches. One of the more promising ones defines $t_i^{(k)}$ to be the ith root of the Bessel function of order $\nu = \frac{k-2}{2}$ and redefines h^α to be $(h/h_l)^\alpha$ so that all the distances in the empirical semivariogram fall in the unit interval. This has the advantage that $\Omega_k(\cdot)$ forms an orthogonal basis for covariograms with respect to the inner-product weighting function $w_k(h) = h^{k-1}$ since $\int_0^1 \Omega_k\left(ht_i^{(k)}\right)\Omega_k\left(ht_j^{(k)}\right) w_k(h)\,dh = 0$ for $i \neq j$.

Using this fact, it can be shown that $1 - \Omega_k(\cdot)$ forms a quasi-orthogonal basis for semivariograms with respect to $w_k(\cdot)$ for large k. Such a basis could potentially take advantage of generalized fourier series theory, but there are two problems with this approach. First, no root larger than $t_1^{(k)}(h_l/h_1)^\alpha$ can be used in the fitting process since it would pass through its jump before the first distance in the empirical semivariogram. This restriction is compounded by the fact that the first root grows larger as k increases. Thus, quasi-orthogonality comes at the price of a sparse node space, which severely impacts the flexibility of the basis. Even if quasi-orthogonality is discarded, empirical semivariograms covering a short range of distances will suffer from node sparsity. For ones covering a large range of distances, we were able to obtain good fits using this technique and will continue to pursue this avenue of research.

A second approach involves modifying the node space so that the basis elements are equivalent in a certain sense. A general sketch of the technique is to define the t_i nodes for $\alpha = 1$ as proposed in this paper. For $\alpha_u < 1$, each u_i node for that space is defined so that $\int_0^{h_l}(1 - \Omega_k(h^{\alpha_u}u_i))\,dh = \int_0^{h_l}(1 - \Omega_k(ht_i))\,dh$. Some of the u_i nodes will have to be discarded since they will not obtain their respective jumps within the hull of the data. The corresponding t_i nodes will also have to be removed to keep the two node spaces on parity. Taken to the extreme of $\alpha_u = 0$, only the node corresponding to the nugget will be left in both spaces. We have experimented with restricting α_u to a lower bound of 0.5, say, and then applying goodness-of-fit criteria with some success. This remains an active area of research where a different definition of node equivalence may eventually obviate the

need to estimate degrees of freedom.

Acknowledgements

This study was supported by the VA IDIQ contract number VA549-P-0027 awarded and administered by the Dallas, TX VA Medical Center. The content of this paper does not necessarily reflect the position or the policy of the U.S. government, and no official endorsement should be inferred.

References

Akaike, H. (1973). Information theory and an extension of maximum likelihood principle. In B. Petrov & F. Csàki (Eds.), 2nd *International Symposium on Information Theory* (pp. 267–281). Budapest: Akadémia Kiadó.

Cherry, S., Banfield, J., & Quimby, W. (1996). An evaluation of a nonparametric method of estimating semivariograms of isotropic spatial processes. *Journal of Applied Statistics*, *23*(4), 435–449.

Craven, P. & Wahba, G. (1979). Smoothing noisy data with spline functions. *Numerical Mathematics*, *31*, 377–403.

Cressie, N. (1985). Fitting variogram models by weighted least squares. *Mathematical Geology*, *17*(5), 563–586.

Cressie, N. (1993). *Statistics for Spatial Data* (revised ed.). New York: John Wiley and Sons.

Hoerl, A. & Kennard, R. (1988). Ridge regression. In *Encyclopedia of Statistical Sciences*, volume 8 (pp. 129–136). New York: Wiley.

Isaaks, E. & Srivastava, R. (1989). *An Introduction to Applied Geostatistics*. Oxford: Oxford University Press.

Journel, A. & Huijbregts, C. (1978). *Mining Geostatistics*. New York: Academic Press.

Kythe, P. & Puri, P. (2002). *Computational Methods for Linear Integral Equations*. Boston: Birkhäuser.

Lawson, C. & Hanson, R. (1974). *Solving Least Squares Problems*. Englewood Cliffs, New Jersey: Prentice-Hall.

Lele, S. (1995). Inner product matrices, kriging, and nonparametric estimation of the variogram. *Mathematical Geology*, *27*(5), 673–692.

Sampson, P. & Guttorp, P. (1992). Nonparametric estimation of nonstationary spatial covariance structure. *Journal of the American Statistical Association*, *87*(417), 108–119.

Schoenberg, I. (1938). Metric spaces and completely monotone functions. *Annals of Mathematics*, *39*(4), 811–841.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*(2), 461–464.

Shapiro, A. & Botha, J. (1991). Variogram fitting with a general class of conditionally nonnegative definite functions. *Computational Statistics and Data Analysis*, *11*(1), 87–96.

Watson, G. (1958). *Theory of Bessel Functions* (2nd ed.). New York: Cambridge University Press.

Zou, H., Hastie, T., & Tibshirani, R. (2004). On the "degrees of freedom" of the lasso. *Annals of Statistics*, *35*(5), 2173–2192.

Tables

Table 1: Some common parametric semivariograms and corresponding empirically determined values of α. The nugget is θ_1 in all the models. The sill is $\theta_1 + \theta_2$.

Model Name	$\gamma(h)$	α
white noise	$\theta_1 + \theta_2$	0.000
exponential	$\theta_1 + \theta_2 \{1 - \exp(-h/\theta_3)\}$	0.575
spherical	$\theta_1 + \theta_2 \left\{ \frac{3}{2}\frac{h}{\theta_3} - \frac{1}{2}\left(\frac{h}{\theta_3}\right)^3 \right\}$	0.750
Gaussian	$\theta_1 + \theta_2 \left\{ 1 - \exp\left[-(h/\theta_3)^2 \right] \right\}$	1.000

Figures

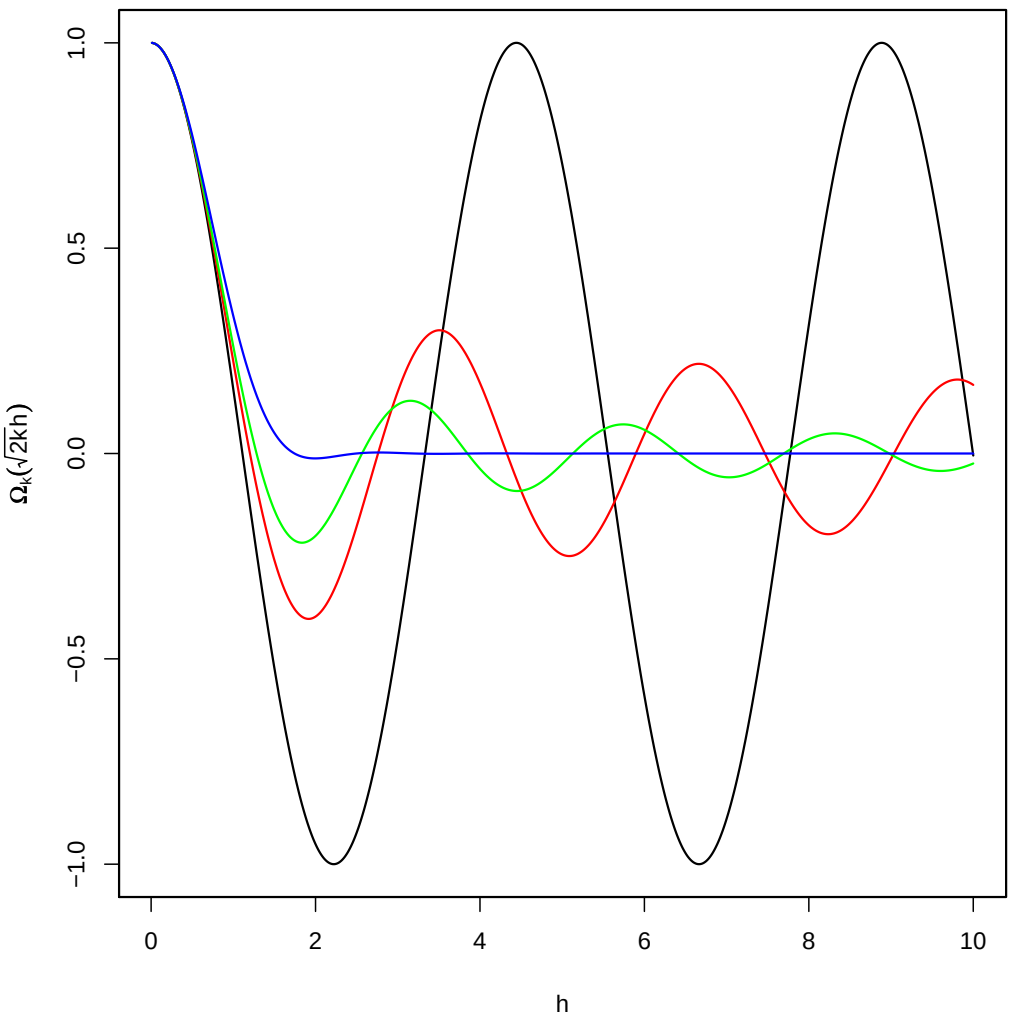

Figure 1: Plots of $\Omega_k(\cdot)$ where $\Omega_1\left(\sqrt{2}h\right) = \cos\left(\sqrt{2}h\right)$ is shown in black, $\Omega_2(2h) = J_0(2h)$ is shown in red, $\Omega_3\left(\sqrt{6}h\right) = \sin\left(\sqrt{6}h\right)/\sqrt{6}h$ is shown in green, and $\Omega_{11}\left(\sqrt{22}h\right)$ is shown in blue. Note how the cyclic behavior dampens more rapidly as k increases.

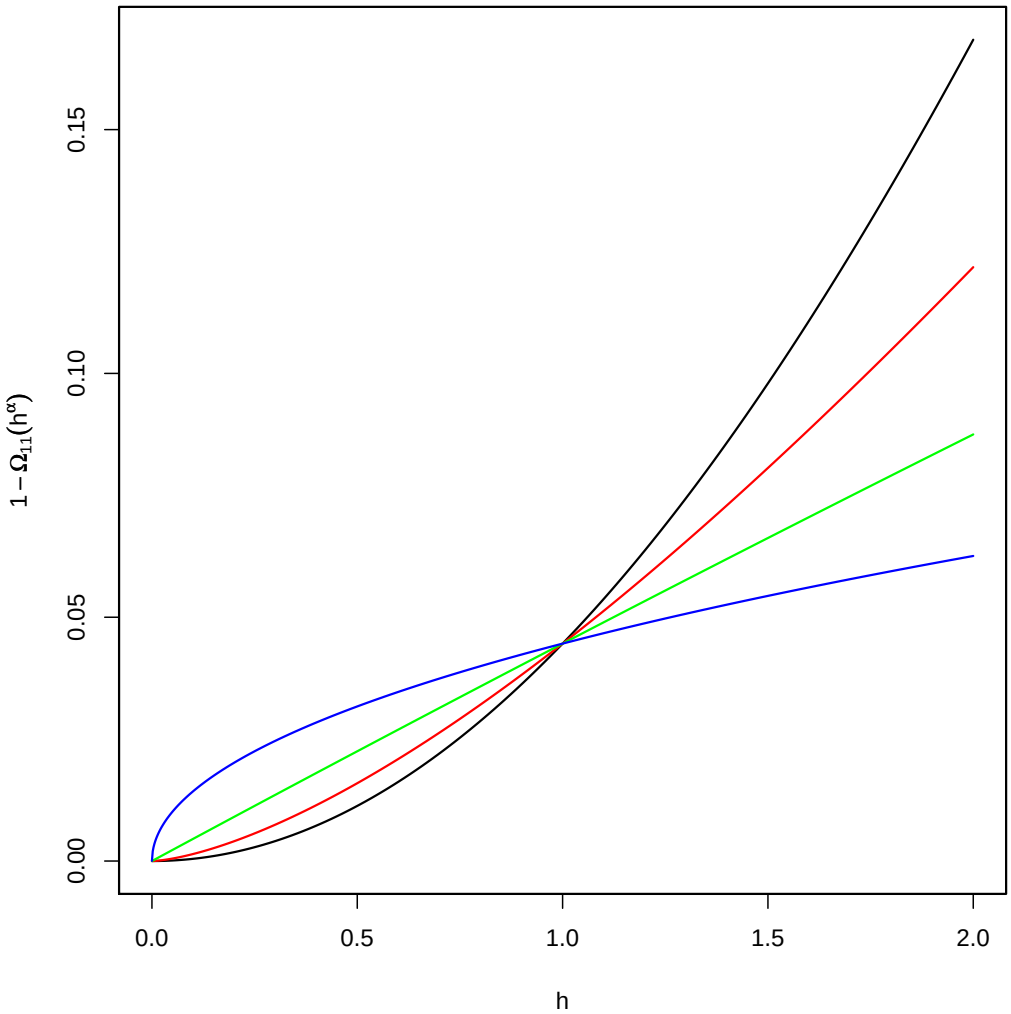

Figure 2: Plots of $1-\Omega_{11}\left(h^{\alpha}\right)$, where $1-\Omega_{11}\left(h^{1.00}\right)$ is shown in black, $1-\Omega_{11}\left(h^{0.75}\right)$ is shown in red, $1-\Omega_{11}\left(h^{0.5}\right)$ is shown in green, and $1-\Omega_{11}\left(h^{0.25}\right)$ is shown in blue. Each approaches the origin in an extremely different manner, which is critical for accurate nugget estimates.

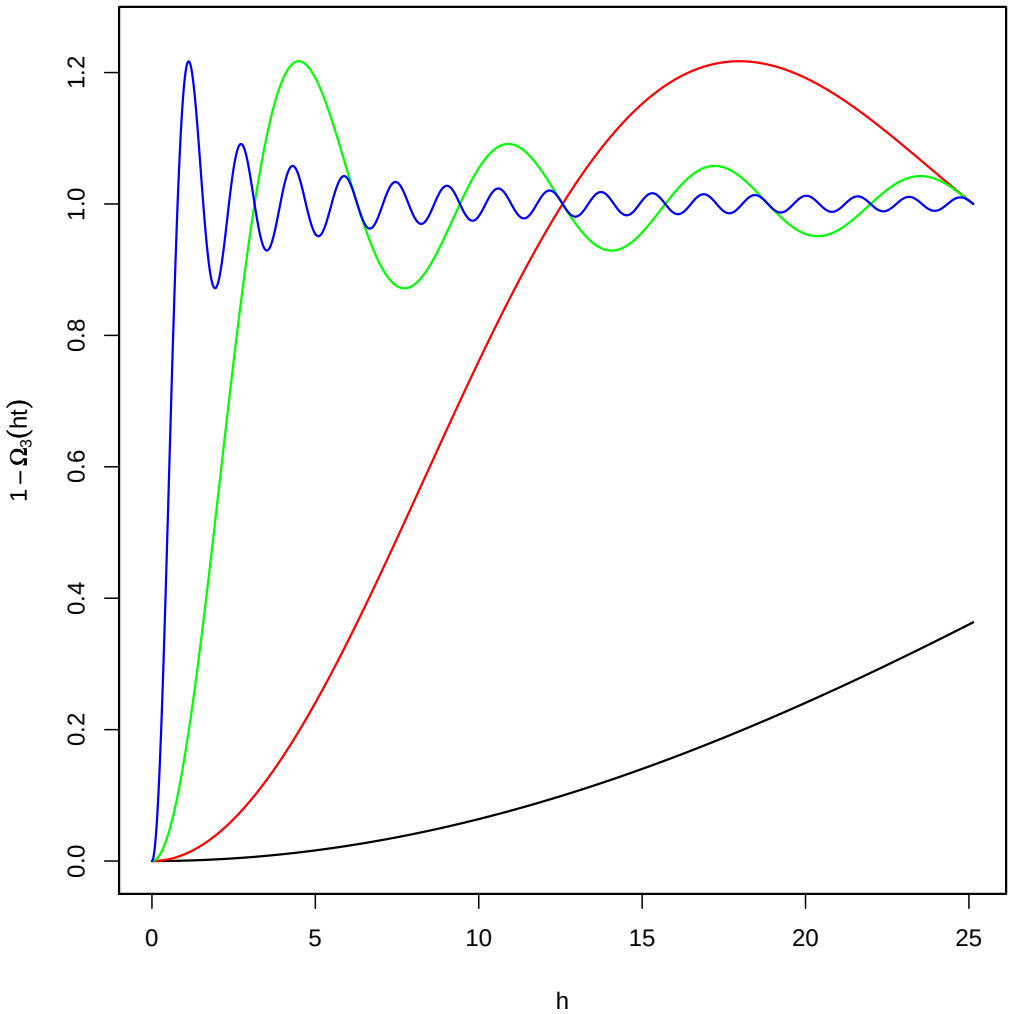

Figure 3: Plots of $1-\Omega_3(ht)$ where $1-\Omega_3(h/16)$ is shown in black, $1-\Omega_3(h/4)$ is shown in red, $1-\Omega_3(h)$ is shown in green, and $1-\Omega_3(4h)$ is shown in blue. $k=3$ is selected to emphasize the oscillatory behavior of the basis elements about their respective jumps, which are set to 1 for all the functions in this example.

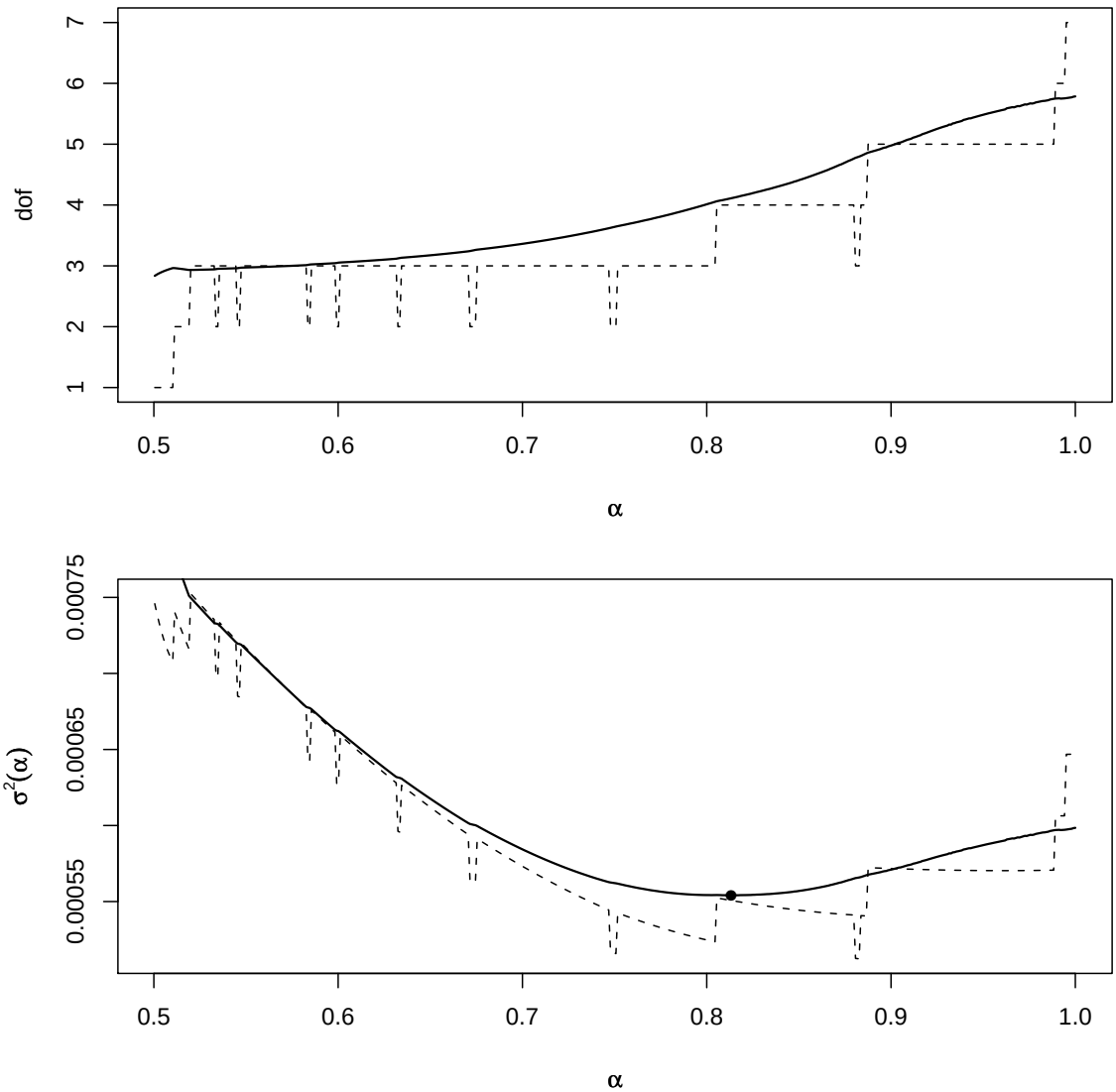

Figure 4: In the upper plot, the dashed line shows the number of nonzero NNLS parameters as a function of α, while the solid curve shows the degrees of freedom as estimated by ridge regression. The lower plot shows $\sigma^2(\cdot)$ as a function of α using the number of nonzero NNLS parameters as degrees of freedom as the dashed line, and using the ridge regression estimated degrees of freedom as the solid curve. The dot indicates the minimum of the solid curve. The erratic behavior in the upper dashed curve makes reliably minimizing the lower dashed curve difficult. Both plots are based on the same sample generated by a spherical semivariogram with 10% nugget, and a range of 6 from a 20×20 realization.

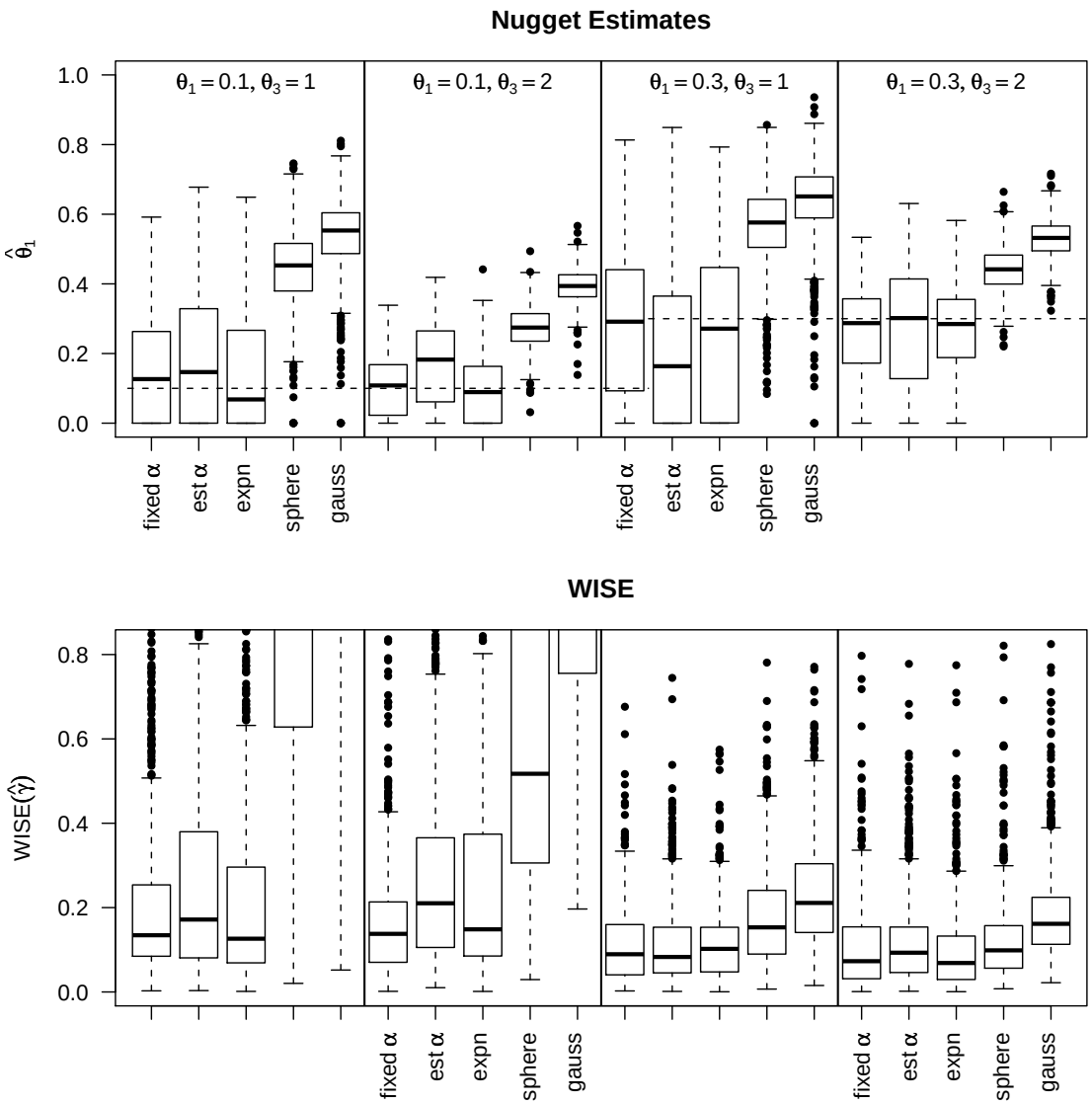

Figure 5: Nugget estimates and WISE for the 2D exponential simulations. The upper portion of the figure consists of five side-by-side boxplots of the nugget estimates produced by the new nonparametric method using a fixed α from Table 1, an estimated value from Equation (23), and from the three parametric fits for 1,000 20×20 two-dimensional realizations. The true nugget is indicated as a dashed horizontal line across the boxplots with the true values for θ_1 and θ_3 indicated at the top of each panel. The lower portion of the figure is the weighted integrated squared error (WISE) defined in Equation (24).

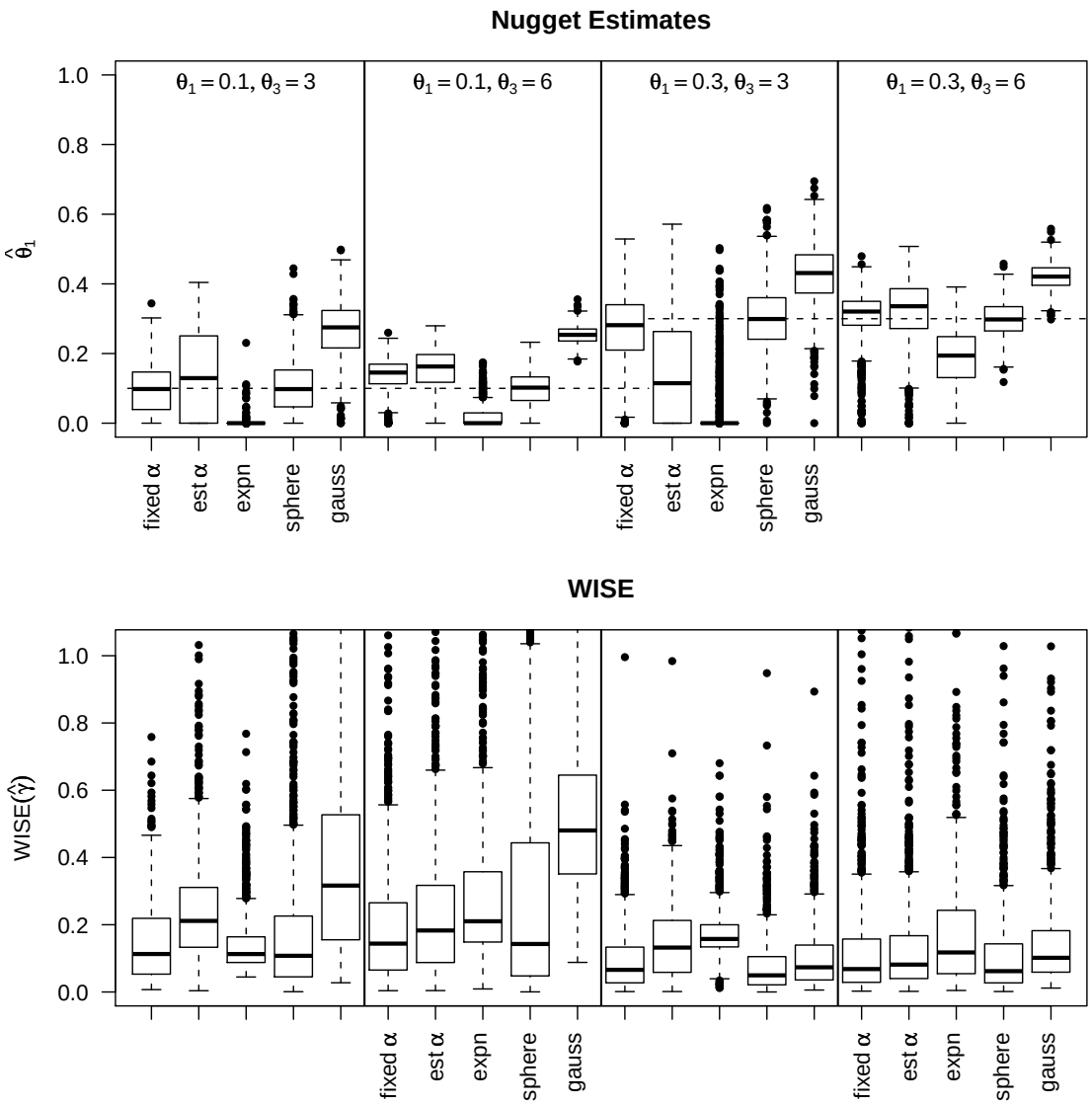

Figure 6: Nugget estimates and WISE for the 2D spherical simulations. The upper portion of the figure consists of five side-by-side boxplots of the nugget estimates produced by the new nonparametric method using a fixed α from Table 1, an estimated value from Equation (23), and from the three parametric fits for 1,000 20×20 two-dimensional realizations. The true nugget is indicated as a dashed horizontal line across the boxplots with the true values for θ_1 and θ_3 indicated at the top of each panel. The lower portion of the figure is the weighted integrated squared error (WISE) defined in Equation (24).

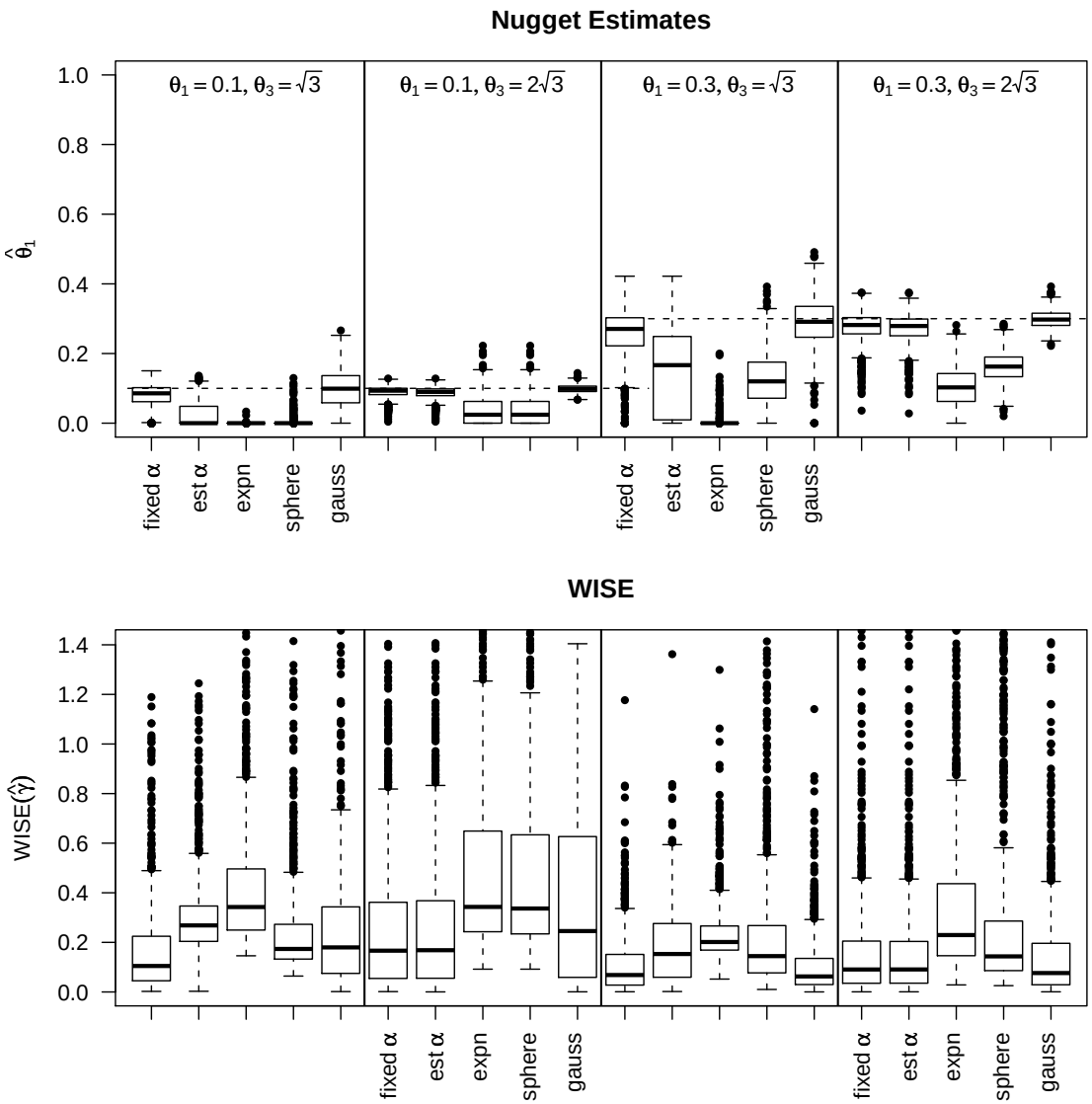

Figure 7: Nugget estimates and WISE for the 2D Gaussian simulations. The upper portion of the figure consists of five side-by-side boxplots of the nugget estimates produced by the new nonparametric method using a fixed α from Table 1, an estimated value from Equation (23), and from the three parametric fits for 1,000 20×20 two-dimensional realizations. The true nugget is indicated as a dashed horizontal line across the boxplots with the true values for θ_1 and θ_3 indicated at the top of each panel. The lower portion of the figure is the weighted integrated squared error (WISE) defined in Equation (24).