ARTICLE IN PRESS

Computational Statistics and Data Analysis ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

A new class of semiparametric semivariogram 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]

[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

ARTICLE INFO

Article history:
Received 14 September 2010
Received in revised form 21 July 2011
Accepted 15 October 2011
Available online xxxx

Keywords:
Unsupervised brain imaging
Nonparametric
Bessel basis
Isotropic
Node space
Regular lattice
Negative definiteness

ABSTRACT

Several authors have proposed nonparametric semivariogram estimators. Shapiro and Botha (1991) did so by application of Bochner's theorem and Cherry et al. (1996) further investigated this technique where it performed favorably against parametric estimators even when data were generated under the parametric model. While the former makes allowances for a prescribed nugget and the latter outlines a possible approach, neither of these demonstrate nugget estimation in practice, which is essential to spatial modeling and proper statistical inference. We propose a modified form of this method, which admits practical 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 unsupervised semiparametric method is demonstrated via application and simulation, where it is shown to be comparable with correctly specified parametric models while outperforming misspecified ones. We conclude with remarks about selecting the appropriate basis and node space definition.

1. Introduction

Let

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

be a sample from a k-dimensional spatial process with

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

where $\mu(\cdot)$ is the mean function and $\delta(\cdot)$ is the spatial error process. Provided that $\mathrm{E}\left[\delta\left(\mathbf{s}\right)\right] = 0$, $\mathrm{Var}\left[\delta\left(\mathbf{s}\right)\right] = \sigma^2 < \infty$, and $\mathrm{Cov}\left[\delta\left(\mathbf{s}_i\right), \delta\left(\mathbf{s}_j\right)\right] = C\left(h_{ij}\right)$, where $C(\cdot)$ is a positive definite covariance function (i.e., $\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j C(h_{ij}) \geq 0, \forall\, a_i, h_{ij}$, and n), and $h_{ij} = \left\|\mathbf{s}_i - \mathbf{s}_j\right\|$, the spatial error process is said to be isotropic. That is, the variance of the process is finite, and the covariance between any two spatial locations only depends on the distance between them.

Commonly, spatial modeling uses *semivariance*,

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

* Corresponding author. Tel.: +1 501 450 5660; fax: +1 501 450 5662.
E-mail address: patrickc@uca.edu (P.S. Carmack).

0167-9473/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2011.10.017

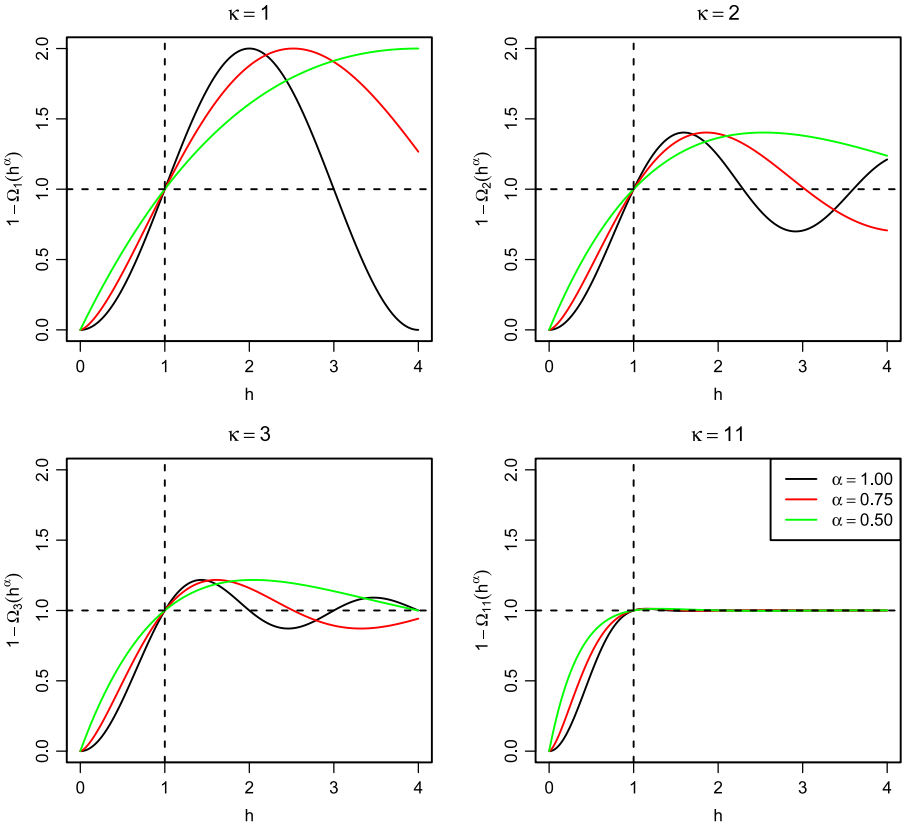

Fig. 1. Plots of semivariogram basis elements for various values of κ and α. The dashed horizontal line shows the unit sill, while the dashed vertical line indicates the first lag. All of the basis functions pass through the unit sill for the first time at the first lag. As κ increases, the oscillatory behavior about the unit sill quickly dampens beyond the first lag. The influence of α toward the origin is comparable in all four plots.

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 and Huijbregts (1978), Isaaks and 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, possibly from multiple sources 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 usually attributed to measurement error. At large distances, the semivariance between locations usually becomes constant and 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 $\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \gamma\left(h_{ij}\right) \leq 0, \forall\, a_i, h_{ij}$, 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 and Botha (1991), Sampson and Guttorp (1992), Lele (1995) and Genton and Gorsich (2002). We shall focus our efforts on presenting and improving upon Shapiro and Botha's approach, which was generalized by Fernández-Casal et al. (2003). Invoking Bochner's Theorem, $C(\cdot)$ can be represented as a spectral integral

$$C(h) = \int_0^\infty \Omega_\kappa(ht)\, dF(t),$$

where $\Omega_\kappa(x) = \left(\frac{2}{x}\right)^{(\kappa-2)/2} \Gamma\left(\frac{\kappa}{2}\right) J_{(\kappa-2)/2}(x)$, $\Gamma(\cdot)$ is the gamma function, $J_\nu(\cdot)$ is the Bessel function of the first kind of order $\nu = (\kappa-2)/2$, and $F(\cdot)$ is a 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, κ may take on non-integer values and, if set to any value higher than the dimension of the spatial process (i.e., $\kappa \geq k$), will still yield a valid covariance function for the dimension in question. Finally, as $\kappa \to \infty$, $\Omega_\kappa\left(\sqrt{2\kappa}h\right) \to \exp\left(-h^2\right)$, which corresponds to a Gaussian spatial error process and will be of importance in modifying the method. Some plots of $1 - \Omega_\kappa(\cdot)$ for selected values of κ are shown in Fig. 1.

Please cite this article in press as: Carmack, P.S., et al., A new class of semiparametric semivariogram and nugget estimators. Computational Statistics and Data Analysis (2011), doi:10.1016/j.csda.2011.10.017

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 and 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 for which nonnegative least squares (Lawson and Hanson, 1974) can be employed to obtain a nonparametric semivariogram estimate by solving

$$\begin{aligned}\hat{\gamma}(h) &= C(0) - C(h)\\ &= \int_0^\infty [\Omega_\kappa(0) - \Omega_\kappa(ht)]\, dF(t)\\ &= \sum_{i=1}^{m} [1 - \Omega_\kappa(ht_i)]\, p_i,\end{aligned}$$

for jumps $p_i \geq 0$ by minimizing

$$\sum_{j=1}^{\ell} \left[\hat{\gamma}(h_j) - \tilde{\gamma}(h_j)\right]^2$$

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. The nodes, t_i, and their number, m, are user defined. 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 others, $1 - \Omega_\kappa(\cdot)$ results in a rich semivariogram basis that compares favorably with parametric semivariogram estimators. In their article, Cherry et al. (1996) outlined a potential iterative nugget estimation method, which we found to yield degenerate estimates in practice. 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 certain class of parametric isotropic semivariogram models can be represented as:

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

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 spatial error process, $\alpha = 1/2$ is an exponential spatial error process, and $\alpha = 1$ corresponds to a Gaussian spatial error process.

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

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

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

$$\begin{aligned}\gamma_\alpha(h) &= \int_0^\infty [\Omega_\kappa(0) - \Omega_\kappa(h^\alpha t)]\, dF(t)\\ &= \sum_{i=1}^{m} [1 - \Omega_\kappa(h^\alpha t_i)]\, p_i.\end{aligned}$$

We have dispensed with the $\sqrt{2\kappa}$ 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 spatial error processes; 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. 1.

Please cite this article in press as: Carmack, P.S., et al., A new class of semiparametric semivariogram and nugget estimators. Computational Statistics and Data Analysis (2011), doi:10.1016/j.csda.2011.10.017

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 4.00–16.16 resulting in a total of 200 nodes. They further noted that saturating this space only increased computation time and had little impact on the final fit. We propose that in addition to the change in the argument to $\Omega_\kappa(\cdot)$, the definition of the node space is crucial for obtaining nugget estimates and stable sill estimates.

Our initial interest in nonparametric semivariogram estimation was to have a method that could operate in unsupervised brain-imaging applications. In our experience implementing the 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 are the instances where the sill estimate, $\sum_{i=1}^{m} p_i$, far exceeds the maximum value in the empirical semivariogram, as addressed in Cherry (1997). 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 $1 - \Omega_\kappa(\cdot)$ basis reveals why both phenomena occur. The first root of $\Omega_\kappa(\cdot)$ corresponds to the distance at which $1 - \Omega_\kappa(\cdot)$ will first cross 1, after which that basis element will proceed to oscillate about 1. Thus, very high frequency nodes start oscillating about 1 before the first lag in the empirical semivariogram and become 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 about 1 until well beyond the hull of the empirical semivariogram. The locations of these crossings coincide with the roots of the Bessel function of the first kind of order $\nu = \frac{\kappa-2}{2}$, for which numeric approximations are given in Ball (2000).

Using the first root of the Bessel function of order ν, t'_κ, define the node space as follows:

$$t_i = \frac{t'_\kappa}{h_i^\alpha}, \quad i = 2, \ldots, \ell,$$

where h_i is the ith ordered distance in the empirical semivariogram (i.e., $h_i < h_j$ for $i < j$). t_1 is defined to be the node at infinity (i.e., $t_1 = \lim_{h\to 0^+} t'_\kappa / h^\alpha = \infty$; $\lim_{t_1\to\infty} 1 - \Omega_\kappa(h^\alpha t_1) = 1, \forall h \geq 0$) 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 obtain and start oscillating about 1 before the second unique distance in the empirical semivariogram since $ht_i < t'_\kappa$, which implies $1 - \Omega_\kappa(ht_i) \neq 1$ 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 crossing within the hull of the data since $1 - \Omega_\kappa\left(h_i^\alpha t_i\right) = 1, i = 2, \ldots, \ell$, 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 κ

As Cherry et al. (1996) noted, as κ 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 $\kappa = 3$, or the sinc function, over $\kappa = 1$ (cosine) or $\kappa = 2$ (J_0), even for lower dimension problems. We take their argument further and suggest that even higher values of κ desirable on the same smoothness grounds (Fig. 1).

The Bessel function of the first kind has an infinite number of roots whose spacing converges to π as $h \to \infty$ (Watson, 1958). This means that while each basis element passes through 1 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 $\kappa = 1$, which is a form of cosine interpolation. Even the sinc function exhibits this behavior given its slowly decaying cyclic nature. For values of $\kappa \geq 11$, $\Omega_\kappa(\cdot)$ 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 $\kappa \to \infty$, which would result in using members of the exponential family as basis elements. The problem here is that as $\kappa \to \infty$, $t'_\kappa \to \infty$. Thus, we would forfeit our ability to precisely control where a given basis element passes through 1. In addition, $\Omega_\kappa(\sqrt{2\kappa}h^\alpha)$ converges fairly rapidly to $\exp(-h^{2\alpha})$. Thus, we use $\kappa = 11$ for the purpose of simulation and application, even though other values of κ are certainly valid for the proposed method.

3.3. Estimating α

As Fig. 1 shows, κ primarily affects the behavior of the basis elements after they pass through 1, while α primarily impacts their behavior between the first lag and the origin. Given the new basis and node space definitions, a natural question arises

Please cite this article in press as: Carmack, P.S., et al., A new class of semiparametric semivariogram and nugget estimators. Computational Statistics and Data Analysis (2011), doi:10.1016/j.csda.2011.10.017

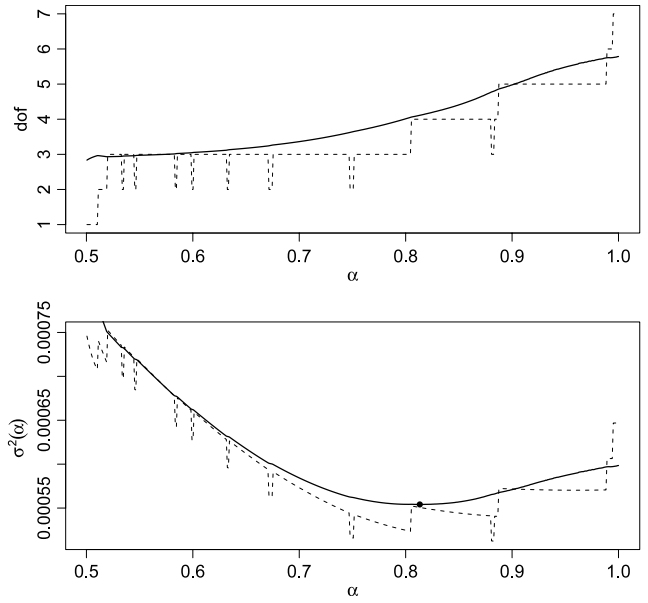

Fig. 2. In the upper plot, the dashed line shows the number of nonzero NNLS parameters as a function of α, while the solid curve shows the degrees of freedom as estimated by ridge regression. The lower plot shows $\sigma^2(\cdot)$ as a function of α using the number of nonzero NNLS parameters as degrees 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 the minimization of 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.

concerning how to choose the tuning parameter α. Model selection criteria, such as Akaike's information criterion (Akaike, 1973), Bayesian information criterion (Schwarz, 1978), or generalized cross validation (Craven and Wahba, 1979) can be used to estimate α. Unfortunately, we found all three methods tended to underestimate α, which may be due to the heavy correlation inherent in the empirical semivariogram. Thus, we propose the following optimization criterion:

$$\sigma^2(\alpha) = \sum_{i=1}^{\ell} \frac{\left[1 - \tilde{\gamma}(h_i)/\hat{\gamma}_\alpha(h_i)\right]^2}{\ell - df\left(\hat{\gamma}_\alpha\right)},$$

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

$$\mathbf{p}(\alpha) = \underset{\mathbf{p}\geq\mathbf{0}}{\operatorname{argmin}} \{\mathbf{A}_\alpha\mathbf{p} - \tilde{\gamma}\}^T \{\mathbf{A}_\alpha\mathbf{p} - \tilde{\gamma}\},$$

where $(\mathbf{A}_\alpha)_{ij} = 1 - \Omega_\kappa\left(h_i^\alpha t_j\right)$, $i = 1, \ldots, \ell$, $j = 1, \ldots, \ell$, and $\tilde{\gamma}^T = \left[\tilde{\gamma}(h_1), \ldots, \tilde{\gamma}(h_\ell)\right]$ is the vector of empirical semivariogram values.

NNLS is similar to shrinkage methods such as ridge regression (Hoerl and Kennard, 1988) and Lasso. Zou et al. (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, Zou et al. (2004) 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 $\sigma^2(\cdot)$ curve extremely difficult (Fig. 2).

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

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

where

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

and

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

is the ridge regression smoother matrix.

Please cite this article in press as: Carmack, P.S., et al., A new class of semiparametric semivariogram and nugget estimators. Computational Statistics and Data Analysis (2011), doi:10.1016/j.csda.2011.10.017

This approximation of degrees of freedom for the NNLS fit is a smooth function of α, making minimization of the $\sigma^2(\cdot)$ curve numerically stable (Fig. 2). Thus, α is estimated as

$$\hat{\alpha} = \underset{0\le\alpha\le 1}{\operatorname{argmin}} \sum_{i=1}^{\ell} \frac{\left[1-\tilde{\gamma}(h_i)/\hat{\gamma}_\alpha(h_i)\right]^2}{\ell - \operatorname{tr}\left(S_{\lambda_\alpha}\right)}, \tag{1}$$

yielding the final fit given by

$$\hat{\gamma}_{\hat{\alpha}}(h) = \sum_{i=1}^{\ell}\left[1-\Omega_\kappa\left(h^{\hat{\alpha}}t_i\right)\right] p\left(\hat{\alpha}\right)_i.$$

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 $\boldsymbol{\Theta}^T = [\theta_1, \theta_2, \theta_3]$ to generate 20×20 two-dimensional realizations. The fields are generated using the root covariance matrix method where the eigen decomposition of $\boldsymbol{\Sigma} = \mathbf{VDV}^T$, $(\boldsymbol{\Sigma})_{ij} = \theta_1 + \theta_2 - \gamma\left(h_{ij}\right)$, is used to obtain $\boldsymbol{\Sigma}^{\frac{1}{2}} = \mathbf{VD}^{\frac{1}{2}}\mathbf{V}^T$, 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,

$$\text{WISE}\left(\hat{\gamma}\right) = \int_0^{h_\ell} \frac{1}{\gamma(h)^2}\left[\gamma(h) - \hat{\gamma}(h)\right]^2 dh = \int_0^{h_\ell}\left[1-\frac{\hat{\gamma}(h)}{\gamma(h)}\right]^2 dh, \tag{2}$$

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. Extensive simulation has shown that fixing $\alpha = 0.00, 0.575, 0.750$, and 1.000 for white noise, exponential, spherical, and Gaussian spatial error processes, respectively, to be nearly optimal at estimating the true parametric semivariogram in terms of WISE for a wide variety of choices for $\boldsymbol{\Theta}$.

For each random realization, the proposed semiparametric method, using both the estimated α and 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 $\boldsymbol{\Theta}$ 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% of the sill, and range parameters (θ_3) were set to either medium or long. Medium range was defined as 1 for the exponential spatial error process, $\sqrt{3}$ for the Gaussian spatial error process, and 3 for the spherical spatial error process, giving each the same approximate range of 3. The long range was set at 2 for the exponential spatial error process, $2\sqrt{3}$ for the Gaussian spatial error process, and 6 for the spherical spatial error process resulting in an approximate 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 1000 times each. The maximum lag included in the empirical semivariogram was $3\sqrt{5}$, which is approximately 1/4 the maximum distance. We ran 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. 3–5 show boxplots for the nugget estimates and WISE for the semiparametric and parametric fits. An examination of the estimated nuggets using the fixed value of α as indicated earlier 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 spatial error process 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 with which to

Please cite this article in press as: Carmack, P.S., et al., A new class of semiparametric semivariogram and nugget estimators. Computational Statistics and Data Analysis (2011), doi:10.1016/j.csda.2011.10.017

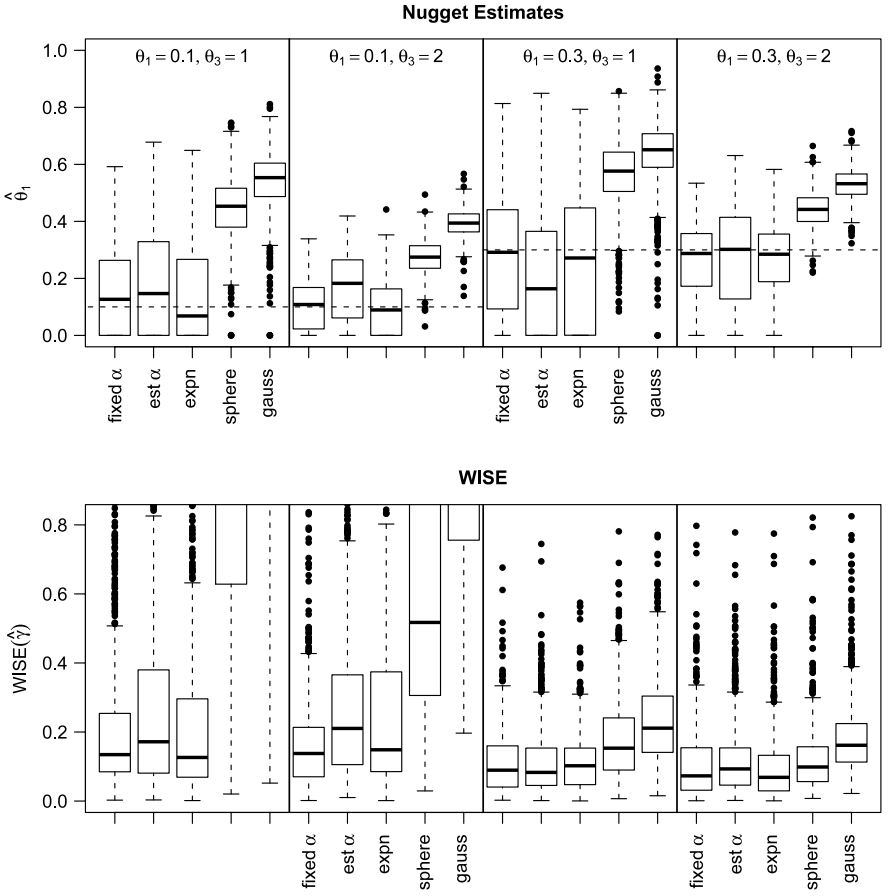

Fig. 3. Nugget estimates and WISE for the 2D exponential spatial field simulations. The upper portion of the figure consists of five side-by-side boxplots of the nugget estimates produced by the new semiparametric method using $\alpha = 0.575$, an estimated value from Eq. (1), and from the three parametric fits for 1000 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 Eq. (2).

identify 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. Application

Our research team routinely employs spatial modeling (kriging) in brain imaging for the purpose of statistical inference and data driven smoothing (Spence et al., 2007), both of which depend heavily on nugget estimation. Even for a modestly sized study, it is impractical to manually inspect and adjust semivariogram fits for every region of interest (ROI) on a subject-by-subject basis. Hence, we have a motivation to use a nonparametric semivariogram estimator that could function in an unsupervised fashion, while providing nugget estimates.

The example given here deals with functional magnetic resonance imaging (fMRI), where the subject lies in a magnetic scanner while images are acquired during a performance task. The magnet records changes in blood oxygen level dependent (BOLD) signals at locations across the brain to see which areas are involved in performing the task. The temporal signal-to-noise ratio in fMRI experiments tends to be very low, which makes spatial modeling's ability to borrow information in space to strengthen these inherently weak signals a natural application. More information about fMRI methods can be found in Lindquist (2008).

In this specific experiment, which lasted 304 s, the subject was shown nonsense words at random times on an overhead view screen and asked to repeat them silently. The time between scans, or time to repeat (TR), was 2 s, which yielded a total of 152 scans. We then extracted the BOLD signal at 1557 spatial locations (voxels) in the superior temporal gyrus, which is a brain structure thought be to associated with the experiment. Next, we modeled the hemodynamic response function (HRF) at each location in space under the standard linear convolution invariance assumption in conjunction with a 13 parameter finite impulse response (FIR). Since TR = 2 s, the FIR covers 26 s after each stimulus. This step removes the temporal aspect of the data leaving a 13 parameter FIR vector estimate at each location in space.

Please cite this article in press as: Carmack, P.S., et al., A new class of semiparametric semivariogram and nugget estimators. Computational Statistics and Data Analysis (2011), doi:10.1016/j.csda.2011.10.017

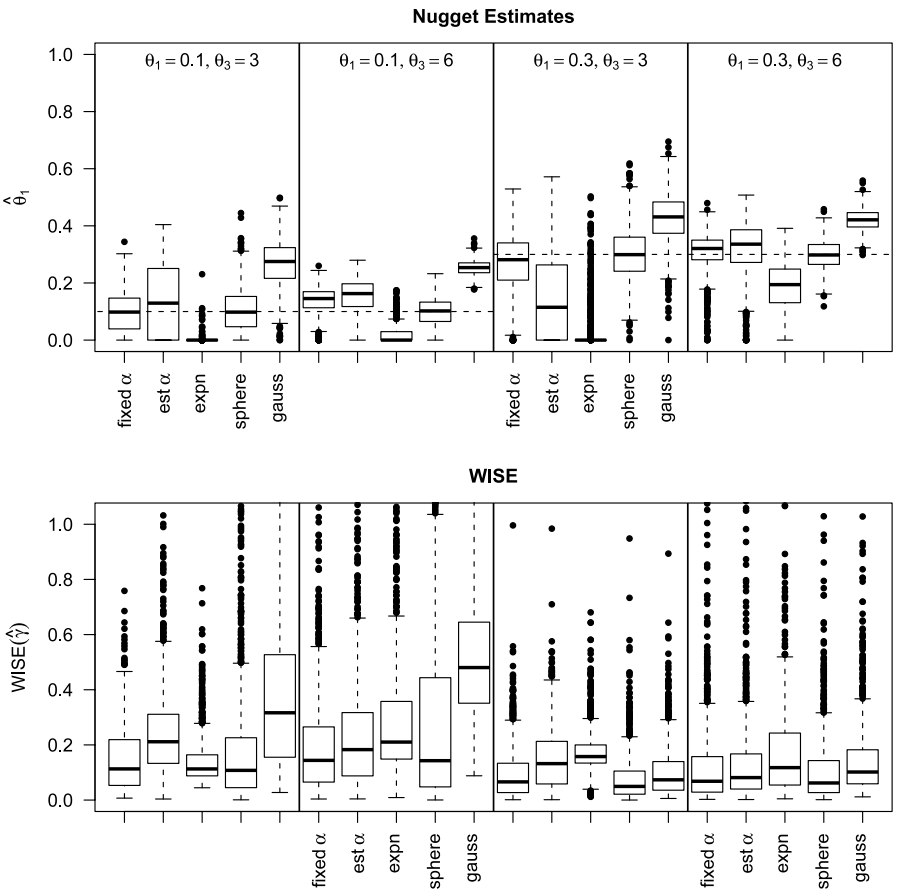

Fig. 4. Nugget estimates and WISE for the 2D spherical spatial field simulations. The upper portion of the figure consists of five side-by-side boxplots of the nugget estimates produced by the new semiparametric method using $\alpha = 0.750$, an estimated value from Eq. (1), and from the three parametric fits for 1000 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 Eq. (2).

We then computed the empirical semivariogram of the third FIR parameter (Fig. 6), which is generally the peak of the HRF 6 s after a stimulus for this particular paradigm and subject. For demonstration purposes, we fit Gaussian (not shown), spherical, and exponential parametric semivariograms. As the figure shows, the exponential semivariogram fit (green) poorly fits the empirical semivariogram and yields a nugget estimate that is 0% of its estimated sill. The spherical semivariogram fit (red) is better with a nugget estimate that is 20% of its estimated sill, but appears to overestimate the range. The Gaussian semivariogram fit is not shown to avoid overcrowding the figure, but it tends to overshoot the early lags albeit to a lesser degree than the exponential. It has the largest nugget estimate at 36% of its estimated sill. Finally, the fit from our proposed method is shown in black with $\hat{\alpha} = 0.63$, which uses four non-zero jumps including the nugget estimated at 9% of its estimated sill. The semiparametric fit is arguably better in the body of the empirical semivariogram and unsurprisingly yields a nugget estimate between that of the exponential and spherical fits in light of the discussion of using fixed values of α in the simulation section. In addition, all the parametric fits required manually set starting parameter values in a non-linear optimization algorithm, while the new method operated in an entirely automated fashion yielding a fit in under 2 s on a 2006 Mac Pro including the time required to estimate α.

6. Discussion

Nonparametric semivariograms offer an effective alternative to choosing a parametric form for spatial modeling. 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. Cherry et al. (1996) discussed an approach to obtain nugget estimates, but our attempts to implement their suggestion resulted in degenerate estimates. Hence, we have proposed several key modifications to improve and extend the method. First, a more flexible basis was introduced by replacing the argument of $\Omega_\kappa(\cdot)$ 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 unsupervised semiparametric semivariogram and nugget estimation possible.

Please cite this article in press as: Carmack, P.S., et al., A new class of semiparametric semivariogram and nugget estimators. Computational Statistics and Data Analysis (2011), doi:10.1016/j.csda.2011.10.017

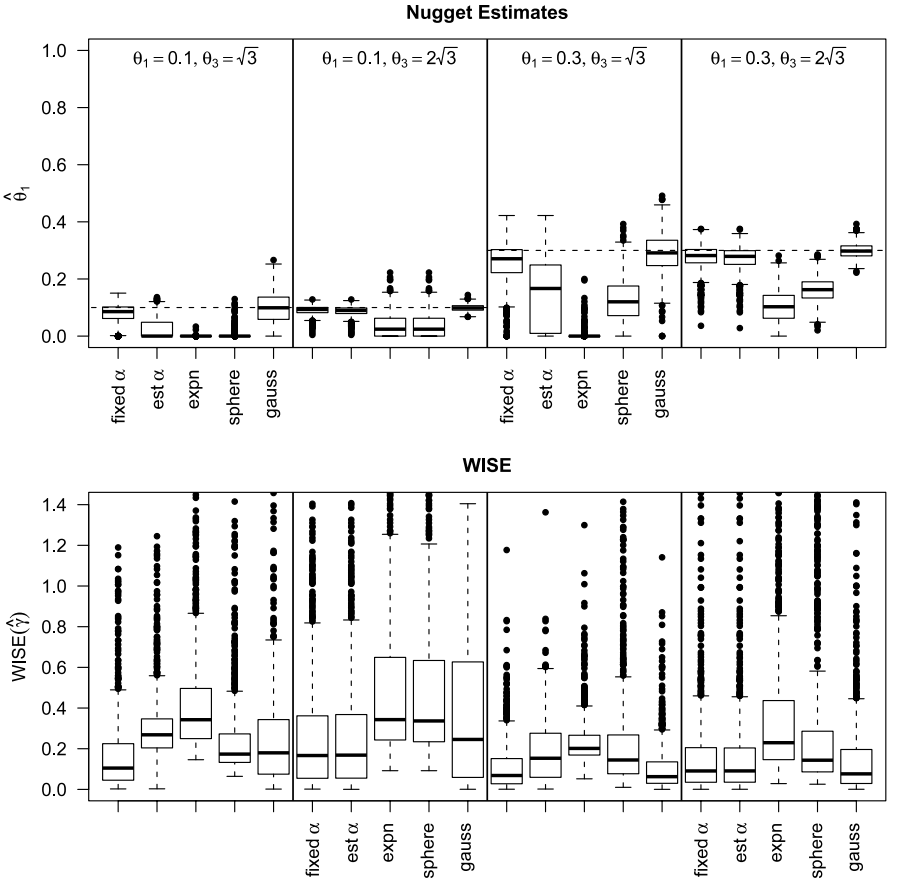

Fig. 5. Nugget estimates and WISE for the 2D Gaussian spatial field simulations. The upper portion of the figure consists of five side-by-side boxplots of the nugget estimates produced by the new semiparametric method using $\alpha = 1.000$, an estimated value from Eq. (1), and from the three parametric fits for 1000 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 Eq. (2).

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 fMRI application demonstrates that the method can be reliably used in place of parametric semivariograms in unsupervised brain imaging where it is impractical to manually inspect every semivariogram fit in such massive datasets.

The rank of $\mathbf{A}_\alpha^T\mathbf{A}_\alpha$ is 1 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 errors 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 semiparametric 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 $t_i^{(\kappa)}$ to be the ith root of the Bessel function of order $\nu = \frac{\kappa-2}{2}$ and redefines h^α to be $(h/h_\ell)^\alpha$ so that all the distances in the empirical semivariogram fall in the unit interval. This has the advantage that $\Omega_\kappa(\cdot)$ forms an orthogonal basis for covariograms with respect to the inner-product weighting function $w_\kappa(h) = h^{\kappa-1}$ since $\int_0^1 \Omega_\kappa(ht_i^{(k)})\Omega_\kappa(ht_j^{(k)})w_\kappa(h)\,dh = 0$ for $i \neq j$.

Using this fact, it can be shown that $1-\Omega_\kappa(\cdot)$ forms a quasi-orthogonal basis for semivariograms with respect to $w_\kappa(\cdot)$ for large κ. 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 $t_1^{(\kappa)}(h_\ell/h_1)^\alpha$ can be used in the fitting process since the corresponding basis element would pass through 1 before the first distance in the empirical semivariogram. This restriction is compounded by the fact that the first root grows larger as κ 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

Please cite this article in press as: Carmack, P.S., et al., A new class of semiparametric semivariogram and nugget estimators. Computational Statistics and Data Analysis (2011), doi:10.1016/j.csda.2011.10.017

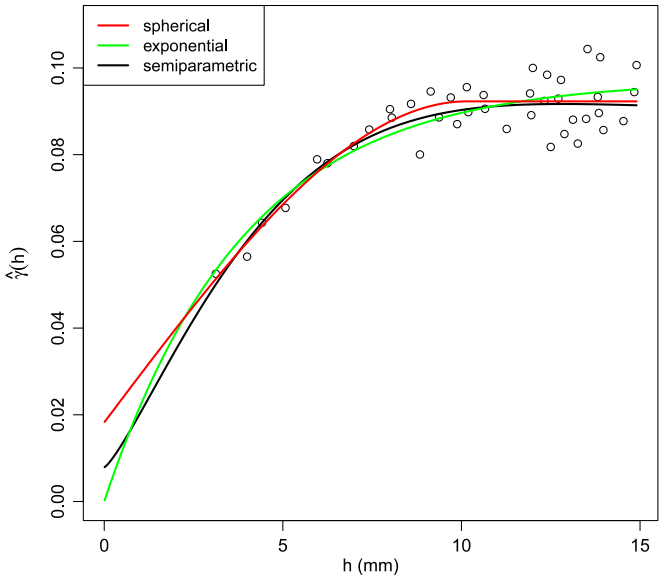

Fig. 6. Plot of three semivariograms fits to an empirical semivariogram shown as hollow circles. The empirical semivariogram is based on the third (peak) parameter of the hemodynamic response function (HRF) estimated using a finite impulse response (FIR) fit to event related fMRI data in the superior temporal gyrus consisting of 1557 voxel locations collected at 152 time points with a time to repeat (TR) of 2 s. The nuggets estimated by the spherical, exponential, and semiparametric fits are 20%, 0%, and 9% of their estimated sills, respectively. For the semiparametric fit, $\hat{\alpha} = 0.63$, which is between an exponential and spherical fit.

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 $\int_0^{h_\ell} [1 - \Omega_\kappa (h^{\alpha_u} u_i)]\, dh = \int_0^{h_\ell} [1 - \Omega_\kappa (h t_i)]\, dh$. 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 model selection 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.

Acknowledgments

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

References

Akaike, H., 1973. Information theory and an extension of maximum likelihood principle. In: Petrov, B., Csàki, F. (Eds.), 2nd International Symposium on Information Theory. Akadémia Kiadó, Budapest, pp. 267–281.
Ball, J., 2000. Automatic computation of zeros of Bessel functions and other special functions. Journal of Scientific Computing 21, 1458–1464.
Cherry, S., 1997. Non-parametric estimation of the sill in geostatistics. Environmetrics 8, 13–27.
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. John Wiley and Sons, New York, revised ed.
Fernández-Casal, R., González-Manteiga, W., Febrero-Bande, M., 2003. Flexible spatio-temporal stationary variogram models. Statistics and Computing 13 (2), 127–136.
Genton, M., Gorsich, D., 2002. Nonparametric variogram estimation with Fourier–Bessel matrices. Computational Statistics and Data Analysis 41 (1), 47–57.
Hoerl, A., Kennard, R., 1988. Ridge Regression. In: Encyclopedia of Statistical Sciences, vol. 8. Wiley, New York, pp. 129–136.
Isaaks, E., Srivastava, R., 1989. An Introduction to Applied Geostatistics. Oxford University Press, Oxford.
Journel, A., Huijbregts, C., 1978. Mining Geostatistics. Academic Press, New York.
Kythe, P., Puri, P., 2002. Computational Methods for Linear Integral Equations. Birkhäuser, Boston.
Lawson, C., Hanson, R., 1974. Solving Least Squares Problems. Prentice-Hall, Englewood Cliffs, New Jersey.
Lele, S., 1995. Inner product matrices, kriging, and nonparametric estimation of the variogram. Mathematical Geology 27 (5), 673–692.
Lindquist, M., 2008. The statistical analysis of fmri data. Statistical Science 23 (4), 439–464.

Please cite this article in press as: Carmack, P.S., et al., A new class of semiparametric semivariogram and nugget estimators. Computational Statistics and Data Analysis (2011), doi:10.1016/j.csda.2011.10.017

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.
Spence, J., Carmack, P., Gunst, R., Schucany, W., Woodward, W., Haley, R., 2007. Accounting for spatial dependence in the analysis of spect brain imaging data. Journal of the American Statistical Association 102 (478), 464–473.
Watson, G., 1958. Theory of Bessel Functions, second ed. Cambridge University Press, New York.
Zou, H., Hastie, T., Tibshirani, R., 2004. On the degrees of freedom of the lasso. Annals of Statistics 35 (5), 2173–2192.

Please cite this article in press as: Carmack, P.S., et al., A new class of semiparametric semivariogram and nugget estimators. Computational Statistics and Data Analysis (2011), doi:10.1016/j.csda.2011.10.017