Statistics and Its Interface Volume 5 (2012) 159–168

Variogram estimation in the presence of trend

Nikolay Bliznyuk, Raymond J. Carroll, Marc G. Genton*, and Yuedong Wang

Estimation of covariance function parameters of the error process in the presence of an unknown smooth trend is an important problem because solving it allows one to estimate the trend nonparametrically using a smoother corrected for dependence in the errors. Our work is motivated by spatial statistics but is applicable to other contexts where the dimension of the index set can exceed one. We obtain an estimator of the covariance function parameters by regressing squared differences of the response on their expectations, which equal the variogram plus an offset term induced by the trend. Existing estimators that ignore the trend produce bias in the estimates of the variogram parameters, which our procedure corrects for. Our estimator can be justified asymptotically under the increasing domain framework. Simulation studies suggest that our estimator compares favorably with those in the current literature while making less restrictive assumptions. We use our method to estimate the variogram parameters of the short-range spatial process in a U.S. precipitation data set.

Keywords and phrases: Bias, Covariance function, Nonlinear regression, Nonparametric regression, Spatio-temporal dependence, Time series.

1. INTRODUCTION

This paper is devoted to estimation of covariance function parameters from data containing an unknown smooth trend. This important problem arises in nonparametric regression models

$$y(x_i) = g(x_i) + \epsilon(x_i), \tag{1}$$

where $g(\cdot)$ is the unknown smooth regression function, $\epsilon(\cdot)$ is the error process with a covariance function $C(\cdot, \cdot; \phi)$ parameterized by a vector ϕ and $y(x_i)$ is the observation at the design point x_i for $i = 1, \dots, n$. We are interested in the models where the indices x_i belong to a subset of the d-dimensional Euclidean space. Estimation of ϕ subsequently allows one to select an appropriate amount of smoothing in order to estimate $g(\cdot)$ using a local smoother corrected for dependence in ϵ [4, 11]. Failure to correct the smoother for dependence in ϵ results in undersmoothed estimates [11] when the errors are positively correlated.

In the current literature surveyed in [11], two classes of estimators can be identified: (i) those based on the differences of observations $y(x_i)$, e.g., [7], and (ii) those based on variogram fitting using residuals from a preliminary nonparametric model fit with the amount of smoothing chosen as in the case with independent errors, e.g., [4]. In the first class, the decision to use a particular difference between the responses $y(x_i)$ and $y(x_j)$ is based on the distance between x_i and x_j. However, we are not aware of any difference-based estimators when the dimension of x_i exceeds 1. As for the second class, the use of residuals from a preliminary fit that assumes independent errors may be improper: following [11], the residuals may not have the same mean and dependence structure as the errors $\epsilon(x_i)$ when the errors are positively correlated because the local smoother not corrected for dependence in errors is prone to undersmoothing.

In this paper, we develop an estimator of ϕ based on fitting a regression model to the squared differences of observations, $s_{ij} = \{y(x_i) - y(x_j)\}^2$. To obtain a model for s_{ij}, represent

$$E(s_{ij}) = E[\{\epsilon(x_i) - \epsilon(x_j)\}^2] + \{g(x_i) - g(x_j)\}^2. \tag{2}$$

The term $E[\{\epsilon(x_i) - \epsilon(x_j)\}^2]$ is known as the variogram of the process ϵ. When $g(\cdot)$ is constant, an estimator of ϕ obtained by regressing the squared differences s_{ij} on the corresponding theoretical values of the variogram is known as a variogram cloud estimator [10]. Whenever $g(\cdot)$ is nonconstant, the second term in the sum (2) creates a bias, which our estimator will correct for. In contrast with the existing difference-based estimators, we do not assume that the indices x_i are one-dimensional. Unlike in the work of Tong and Wang [15], we do not assume independence of the errors $\epsilon(x_i)$, but when the errors are independent and homoscedastic, Tong and Wang's variance estimator arises as a special case of ours. Our work is motivated by — but is not limited to — spatial statistics, where $g(\cdot)$ represents an unknown long-range smooth spatial trend and ϵ captures the short-range variability of the process y. Beyond smoothness of $g(\cdot)$, we make standard assumptions of the expanding domain asymptotic framework [9], which allow one to consistently estimate the variogram in the simpler models where the signal is observed without the trend.

This paper is organized as follows. In Section 2, we develop our bias-corrected variogram cloud estimator, show its connections with a classical binned estimator, and provide

*Corresponding author.

justifications. In Section 3, we report results of simulation experiments to evaluate the finite-sample performance of our estimator and, whenever possible, to compare it with existing estimators. Subsequently, we use our procedure to estimate the variogram parameters of the short-range spatial dependence process in a U.S. precipitation data set. Technical and supplementary details are provided in the Appendix.

2. BIAS-CORRECTED VARIOGRAM ESTIMATOR

2.1 Variogram cloud estimator in one dimension

In this section, we develop a bias-corrected estimator of the variogram parameters that is based on fitting a least squares model to the squared differences of pairs of observations $y(x_i)$ in (1).

Consider a nonparametric regression model (1) with the design points $\{x_1, \ldots, x_n\}$ in $[0,1]$. We require that the first four moments of $\epsilon(x_i)$ are finite and that $g(\cdot)$ is continuously differentiable. Since our goal is estimation of the variogram and its parameters, we operate under the expanding domain asymptotic framework [3]. For simplicity, one can assume, as in [7], a stationary variogram model

$$E[\{y(x_i) - y(x_j)\}^2] = \gamma_n(x_i, x_j; \phi) = \gamma\{n(x_i - x_j); \phi\}.$$

This representation is obtained upon rescaling the one-dimensional (expanding) domain from $[0,n]$ to $[0,1]$. For example, $\gamma_n(x_i, x_j; \phi) = 2\sigma^2\{1 - \exp(-\lambda n|x_i - x_j|)\}$ for the exponential variogram model, where $\phi = (\sigma^2, \lambda)$. In the case of an equally-spaced design with $x_i = i/n$, this is the variogram of an $AR(1)$ process on $[0,n]$, $\gamma_n(x_i, x_j; \phi) = 2\sigma^2\{1 - \rho^{|i-j|}\}$, where $\rho = \exp(-\lambda)$ is the lag-one correlation.

Replacing the second term in (2) with a Taylor series expansion of $g(x_i)$ about x_j, we obtain

$$\begin{aligned}(3)\quad E(s_{ij}) &= \gamma_n(x_i, x_j; \phi) + [g(x_j) + g'(x_j)(x_i - x_j) \\ &\quad + O\{(x_i - x_j)^2\} - g(x_j)]^2 \\ &= \gamma_n(x_i, x_j; \phi) + \{g'(x_j)\}^2(x_i - x_j)^2 \\ &\quad + O\{(x_i - x_j)^3\}.\end{aligned}$$

Because of the stationarity, we can redefine $\gamma_n(x_i, x_j; \phi) = \gamma_n(a_{ij}; \phi)$, where $a_{ij} = |x_i - x_j|$.

In the case of a general space-filling design in $[0,1]$ (see [13]), we propose to estimate ϕ by nonlinear least squares in the model

$$(4)\qquad s_{ij} = \gamma_n(a_{ij}; \phi) + \beta a_{ij}^2 + e_{ij}$$

under the assumption that $E(e_{ij}) = 0$. The estimator is obtained as a global minimizer of the ordinary nonlinear least squares criterion $h_O(\cdot)$ with respect to (β, ϕ), where

$$(5)\qquad h_O(\phi, \beta) = \sum_{(i,j)\in\mathcal{I}} \{s_{ij} - \gamma_n(a_{ij}; \phi) - \beta a_{ij}^2\}^2$$

for (i,j) in the adjacency list $\mathcal{I} = \{i < j : a_{ij} \le (m/n)\}$ for some "radius" m that does not depend on n. Practical selection of bandwidth in one- and multidimensional index settings is discussed in the Appendix A.2.

In the following subsections we show that, even though the coefficients of a_{ij}^2 in the Taylor series expansion in (3) depend on the x_j's and are unknown, the regression model (4) indeed effectively corrects for the variogram cloud bias $\{g'(x_j)\}^2(x_i - x_j)^2$ due to not knowing $g(\cdot)$. Subsequently, we generalize the procedure from one-dimensional to general d-dimensional indexed random fields.

2.2 Binning to aggregate the variogram cloud biases

To make the exposition more transparent, we now restrict attention to the setting of equally spaced design in one dimension, where $x_i = i/n$.

For $k = 1, \ldots, m$, let $S_k = (n-k)^{-1}\sum_{i=1}^{n-k} s_{i,i+k}$, be the classical binned variogram estimator. A development similar to that used to obtain (3) yields

$$\begin{aligned}(6)\quad S_k &= (n-k)^{-1}\sum_{i=1}^{n-k}\{\epsilon(x_{i+k}) - \epsilon(x_i)\}^2 \\ &\quad + a_k^2(n-k)^{-1}\sum_{i=1}^{n-k}\{g'(x_{i+k})\}^2 \\ &\quad + 2a_k(n-k)^{-1}\sum_{i=1}^{n-k} g'(x_{i+k})\{\epsilon(x_{i+k}) - \epsilon(x_i)\} \\ &\quad + O_p(a_k^3),\end{aligned}$$

where $a_k = a_{i,i+k} = k/n$.

Define $J = \int_0^1 \{g'(x)\}^2 dx$. Since

$$\begin{aligned}(7)\quad E(S_k) &= \gamma_n(a_k; \phi) + a_k^2\{J + O(n^{-1})\} + O(a_k^3) \\ &= \gamma_n(a_k; \phi) + Ja_k^2 + O(a_k^3),\end{aligned}$$

the dominant term of bias in the binned variogram depends on the variogram cloud biases only through J.

2.3 Justification of bias-corrected binned variogram estimator

Estimation of the variogram parameters by minimizing $\sum_{k=1}^m \{S_k - \gamma_n(a_k; \phi) - \beta a_k^2\}^2$ is equivalent to the following two-step procedure:

1. Holding ϕ fixed, regress $Z_k(\phi) = S_k - \gamma_n(a_k; \phi)$ on a_k^2 to estimate $\widehat{\beta}(\phi)$.
2. Minimize in ϕ the sum of squared "residuals" $\sum_{k=1}^m \{Z_k(\phi) - \widehat{\beta}(\phi)a_k^2\}^2$. Equivalently, minimize the criterion $\|(I_m - H)Z(\phi)\|_2^2$, where I_m is the identity matrix of size m, $H = bb^{\mathrm{T}}/b^{\mathrm{T}}b$ is the "hat" matrix, $b = (1, 2^2, \ldots, m^2)^{\mathrm{T}}$, $Z(\phi) = \{Z_1(\phi), \ldots, Z_m(\phi)\}^{\mathrm{T}}$ and $\|\cdot\|_2$ is the Euclidean vector norm.

Consistency of $\widehat{\phi} = \arg\min_\phi \|(I_m - H)Z(\phi)\|_2^2$ follows by

Proposition 1. *Assume*

A1: $V = (V_1, \dots, V_m)^{\mathrm{T}}$ *converges in probability to* $v(\phi^*) = \{\gamma(1;\phi^*), \dots, \gamma(m;\phi^*)\}^{\mathrm{T}}$, *where* ϕ^* *is the true* ϕ *and* $V_k = (n-k)^{-1}\sum_{i=1}^{n-k}\{\epsilon(x_{i+k}) - \epsilon(x_i)\}^2$ *[9].*

A2: $\|(I_m - H)\{v(\phi_0) - v(\phi_1)\}\|_2^2 = 0$ *if and only if* $\phi_0 = \phi_1$, *for all valid* ϕ_0, ϕ_1.

Then $\widehat{\phi}$ *converges to* ϕ^* *in probability.*

The proof is outlined in the Appendix A.1.

2.4 Equivalence of variogram cloud and binned variogram estimation

Consider two nonlinear regression models,

$$S_k = \gamma_n(a_k;\phi) + \beta a_k^2 + e_k, \tag{8}$$

$$s_{i,i+k} = \gamma_n(a_k;\phi) + \beta a_k^2 + e_{i,i+k}, \tag{9}$$

along with the two corresponding objective functions

$$h_W(\phi,\beta) = \sum_{k=1}^{m} w_k\{S_k - \gamma_n(a_k;\phi) - \beta a_k^2\}^2, \tag{10}$$

$$h_O(\phi,\beta) = \sum_{k=1}^{m}\sum_{i=1}^{n-k}\{s_{i,i+k} - \gamma_n(a_k;\phi) - \beta a_k^2\}^2, \tag{11}$$

where $w_k = n - k$. It can be seen that the estimating equations obtained by differentiating either h_W or h_O with respect to (β, ϕ) are the same, hence the estimation methods that minimize the weighted binned variogram criterion (10) or the unweighted variogram cloud criterion (11) are equivalent. If m does not depend on n as assumed in this work, $w_k = n + O(1)$ and estimation by minimizing h_W is asymptotically equivalent to minimization of $\|(I_m - H)Z(\phi)\|_2^2$.

2.5 Extension to multidimensional index settings

We now extend the above one-dimensional procedure to random fields on the d-dimensional hypercube $[0,1]^d$. In order to simplify exposition, assume the process is observed on the lattice $\mathcal{X} = \{0.5, 1.5, \dots, n^{1/d} - 0.5\}^d / n^{1/d}$.

Repeating the decomposition in (3),

$$\begin{aligned} E(s_{ij}) &= \gamma_n(x_i - x_j;\phi) + \{g(x_j) + (x_i - x_j)^{\mathrm{T}}\nabla g(x_j) \\ &\quad + O(\|x_i - x_j\|_2^2) - g(x_j)\}^2 \\ &= \gamma_n(x_i - x_j;\phi) + \|x_i - x_j\|_2^2\left[\{\nabla g(x_j)\}^{\mathrm{T}}u_{ij}\right]^2 \\ &\quad + O(\|x_i - x_j\|_2^3), \end{aligned} \tag{12}$$

where $u_{ij} = (x_i - x_j)/\|x_i - x_j\|_2$, ∇ is the gradient operator. Notice that the bias depends not only on the gradient, but also on the direction u_{ij}.

To account for the dependence of the bias on direction, we propose the following generalization of the one-dimensional binned estimator:

1. Choose a set of directions $\mathcal{H}$, for example, if $d = 2$, $\{(0,1);(1,0);(1,1);(1,-1)\}/\sqrt{n}$.
2. For each $h \in \mathcal{H}$:
   (a) For $k = 1, \dots, m$, compute the classical (directional) variogram estimators
   $$S_k(h) = |\mathcal{I}(hk)|^{-1}\sum_{x_i \in \mathcal{I}(hk)}\{y(x_i + hk) - y(x_i)\}^2,$$
   where $\mathcal{I}(hk) = \{x_i \in \mathcal{X} : x_i + hk \in \mathcal{X}\}$.
   (b) Define a directional analogue of the column vector $Z(\phi)$ of Section 2.3 as $Z(\phi;h) = \{Z_1(\phi;h), Z_2(\phi;h), \dots, Z_m(\phi;h)\}^T$, where $Z_k(\phi;h) = S_k(h) - \gamma_n(kh;\phi)$ for $k = 1, 2, \dots, m$.
3. Minimize $\sum_{h\in\mathcal{H}}\|(I_m - H)Z(\phi;h)\|_2^2$, where $H = bb^{\mathrm{T}}/b^{\mathrm{T}}b$ and $b = (1, 2^2, \dots, m^2)^{\mathrm{T}}$.

Notice that, because the same number of equally-spaced "steps" were taken in each direction h, the projection matrix H is the same for every direction. More generally, the criterion for minimization is obtained by projecting $Z(\phi;h)$ onto the null space of the squared lags. As in Sections 2.1 and 2.4, one could consider a procedure based on the variogram cloud points that lie along a given direction (at least, approximately), rather than the binned variogram.

Asymptotic justification of the procedure in the multidimensional index case is similar to that in the one-dimensional case of Section 2.3, and is omitted here.

3. NUMERICAL EXPERIMENTS

3.1 Experiments with one-dimensional designs

It can be seen that the estimator of Tong and Wang in [15] is a special case of our variogram estimator. The model of [15], which appears in (4) with a constant variogram (as a function of lag), would be misspecified when there is dependence in the errors $\epsilon(x_i)$. The primary goal of this section is to examine the impact of this model misspecification. It is also of interest to compare our estimator with that in Section 2.2 of [7], where Hall and Van Keilegom [7] assume that ϵ is an order-p autoregressive error process in one dimension. Their procedure, which does not account for the trend explicitly, is based upon estimation of (i) σ^2 as an average of classical semivariogram estimators of $\gamma(m;\phi)$ for "large" lags $m = m_1, \dots, m_2$, where $m_1/\log(n) \to \infty$ and $m_2 = O(n^{1/2})$, and of (ii) $\gamma(k;\phi)$ for small values of $k = 1, \dots, p$. The parameters of an $AR(p)$ process are subsequently recovered via Yule-Walker equations, which, more generally, can be achieved by least squares. Of course, in the spatial context of most interest to us, the estimator of [7] is not applicable.

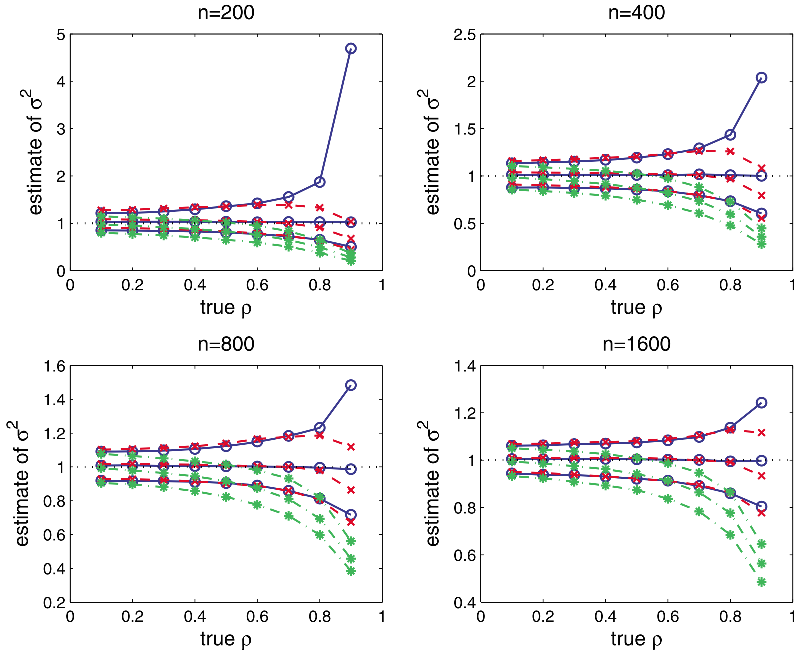

Figure 1. Experiment of Section 3.1: plots of the 5th, 50th and 95th sample percentiles of estimators of σ^2: our variogram estimator (–o–), the Hall and Van Keilegom estimator of [7] (--x--) and the Tong and Wang estimator of [15] (–·*–·*). The true σ^2 is 1 ($\cdots$). The trend is $g(x) = 10 + 12.5x^3(1 - 0.5x)^3$.*

We conducted simulations under the model of Section 3 of [7]. We used $g(x) = 10 + 12.5x^3(1 - 0.5x)^3$. The errors $\epsilon(x_i)$ follow a Gaussian $AR(1)$ process with the variogram model $\gamma_n(x_i, x_j; \phi) = 2\sigma^2\{1 - \exp(-\lambda n|x_i - x_j|)\}$, where $\phi = (\sigma^2, \lambda)$. We simulated data for $x_i = (i - 0.5)/n$, $i = 1, 2, \ldots, n$ for $n = 200, 400, 800,$ 1,600, $\rho = \exp(-\lambda) = 0.1, 0.2, \ldots, 0.9$ and $\operatorname{var}\{\epsilon(x_i)|\epsilon(x_{i-1})\} = 1/16$. For each experimental setting, 500 data sets were generated. The bandwidth parameters are $m = \sqrt{n}$ for the estimators of Tong and Wang in [15] and of ours, and $m_1 = n^{0.4}$ and $m_2 = \sqrt{n}$ for the estimator of Hall and Van Keilegom as in the simulations in [7]. We used 3 random starting points in optimization; the solution with the lowest value of h_O was retained.

In Figure 1, we plot the 5th, 50th and 95th normalized sample percentiles of σ^2 for the three estimators; the estimates of $\operatorname{var}\{\epsilon(x_i)\}$ were divided by the true variance $(1/16)/(1 - \rho^2)$. It is seen that for data with strong dependence, the estimated σ^2 is heavily underestimated by the Tong and Wang estimator in [15]. When dependence is high, our estimator appears more variable but less biased than that of [7]. In Figure 2, we plot the corresponding sample percentiles of estimators of ρ. The plots for the two estimators are very similar to each other, with our estimator being slightly more variable.

The estimator of Hall and Van Keilegom from [7] effectively ignores the trend, which does not impact the large-sample performance of their procedure under the assumptions they make. Although in the above experiment their estimator is well-behaved and the trend is indeed negligible, it is of great practical interest whether ignoring the trend compromises the estimation procedure when the variability in the trend is substantial. We fixed $n = 2{,}000$ and increased the amount of signal from the trend by using $g(x) = c\{10 + 12.5x^3(1 - 0.5x)^3\}$, where $c = 2, 4, 8, 16$, while keeping the remaining experimental conditions as above. In Figures 3–4, we plot the percentiles of the estimators of σ^2 and ρ for the procedure of [7] and for ours. The behavior of our estimator barely changes as c increases, while that of the estimator of [7] deteriorates beyond what is acceptable.

3.2 Experiments with two-dimensional designs

The goal of this section is to assess the performance of our estimator and the impact of ignoring the trend when estimating the variogram using a two-dimensional test problem from [4].

The set of design points is $\mathcal{X} = \{0.5, 1.5, \ldots, \sqrt{n} - 0.5\}^2/\sqrt{n}$ and the number of design points, n, is a square of an integer. We use the exponential variogram model $\gamma_n(x_i, x_j; \phi) = 2\sigma^2\{1 - \exp(-\lambda\sqrt{n}\|x_i - x_j\|_2)\}$, where $\phi = (\sigma^2, \lambda)$. To account for potentially different magnitudes of $g(x)$ and $\epsilon(x)$, the variance of the noise, σ^2, is equal to the second central moment of $g(\cdot)$ times a multiplicative constant τ. In the experiments of this section, the values of τ are $0.5, 1.0, 2.0$ and the values of $\rho = \exp(-\lambda)$ are $0.25, 0.50, 0.75$. For each experimental setting, 500 data sets were generated.

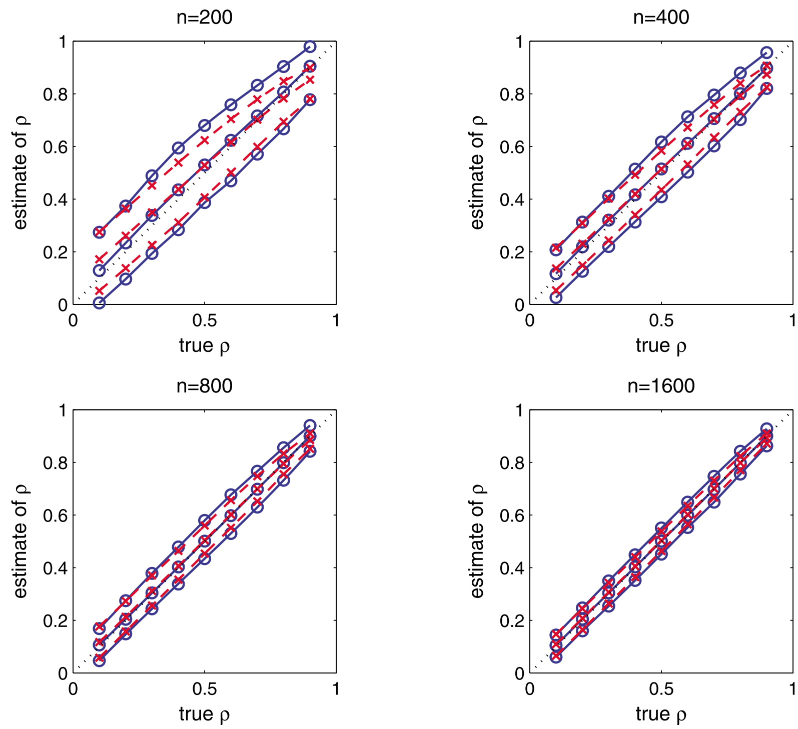

Figure 2. Experiment of Section 3.1: plots of the 5th, 50th and 95th sample percentiles of estimators of ρ: our variogram estimator (-o-) and the estimator of [7] (--x--). The trend is $g(x) = 10 + 12.5x^3(1 - 0.5x)^3$.

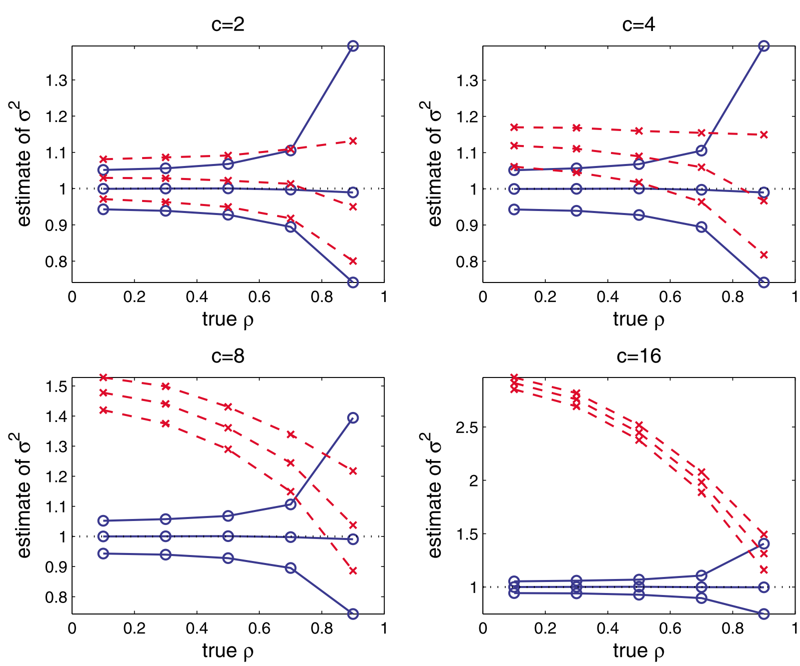

Figure 3. Experiment of Section 3.1: plots of the 5th, 50th and 95th sample percentiles of estimators of σ^2: our variogram estimator (-o-) and the estimator of [7] (--x--). The trend is $g(x) = c\{10 + 12.5x^3(1 - 0.5x)^3\}$, $n = 2,000$. The true σ^2 is 1 ($\cdots$).

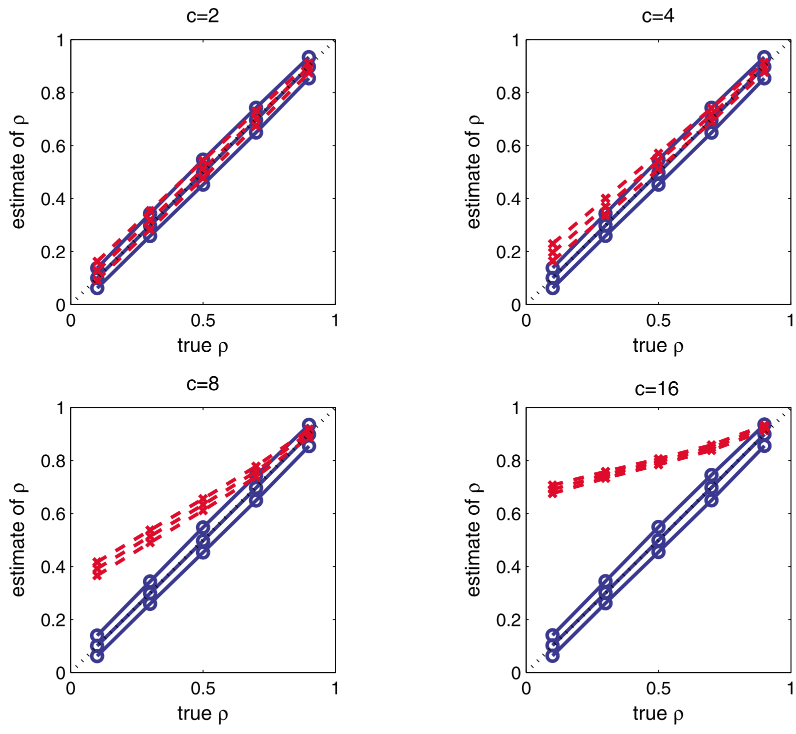

Figure 4. Experiment of Section 3.1: plots of the $5th$, $50th$ *and* $95th$ *sample percentiles of estimators of* ρ*: our variogram estimator (*`-o-`*) and the estimator of [7] (*`--x--`*). The trend is* $g(x) = c\{10 + 12.5x^3(1 - 0.5x)^3\}$, $n = 2,000$.

To estimate the variogram parameters, we use the procedure of Section 2.5 with $m = 5$ and $\mathcal{H} = \{(0,1); (1,0); (1,1); (-1,1)\}/\sqrt{n}$, although a greater number of directions and lags could be considered. Since a stationary variogram is symmetric, i.e., $\gamma(h;\phi) = \gamma(-h;\phi)$, the use of directions from the first two quadrants is sufficient.

In Tables 1–2, we summarize the results of the experiment with the function $g(\cdot)$ that was used by [4], $g(x_1, x_2) = \sin(2\pi x_1) + 4(x_2 - 0.5)^2$. The focus of their work is local polynomial regression adjusted for dependence in the error process, not estimation of the variogram parameters by itself. Given that estimates of the variogram parameters based on the residuals from a preliminary nonparametric fit to $g(\cdot)$ — under the independence assumption of errors — can be misleading [11], estimation of the variogram parameters that bypasses estimation of $g(\cdot)$, such as our present work, is of particular importance.

We estimateed the variogram parameters by minimizing the criterion $\sum_{h\in\mathcal{H}} \|(I_m - H)Z(\phi; h)\|_2^2$ from Section 2.5 and compared them to the estimates from minimization of the criterion $\sum_{h\in\mathcal{H}} \|Z(\phi; h)\|_2^2$ that ignores the trend. Tables 1 and 2 juxtapose the respective estimates of τ and ρ by the two procedures. The fact that our procedure uniformly outperforms the one that does not correct for the trend strongly suggests the importance of bias correction in variogram estimation and provides further support for our methodology.

4. CASE STUDY: U.S. PRECIPITATION DATA

In this section, we apply our procedure to estimate parameters of the short-range spatial variability in a large spatial data set of U.S. precipitation in April 1948. The data set is available as part of the `spam` package for `R` and has been used in numerous papers; see Section 4 of [5] for discussion and references. The data consist of monthly averages of *anomaly*, defined as standardized square root of precipitation readings, recorded by $n = 5,909$ weather stations that are scattered irregularly over the continental U.S. We converted the geographical coordinates (latitude and longitude) into rectangular coordinates (in miles along the horizontal and vertical axes) through a map projection [2], after which we rounded the rectangular coordinates of stations to the nearest mile.

Furrer et al. [5] state that the anomaly field is close to being second-order stationary. They represent the field stochastically as a sum of two independent processes, one with a rapidly decaying exponential covariance function and another with a slowly decaying one. For a data set of this size or larger, estimation of covariance function parameters in the presence of long-range dependence, e.g., by a likelihood-based method, can become computationally prohibitive due to evaluation and factorization of an $n\times n$ covariance matrix. On the other hand, if the long-range dependence is modeled

Table 1. Sample percentiles of estimates of τ in the experiment of Section 3.2 with $g(x_1, x_2) = \sin(2\pi x_1) + 4(x_2 - 0.5^2)$ when directional correction for the variogram bias is used. The bias-corrected estimator (left) is the minimizer of $\sum_{h\in\mathcal{H}} \|(I_m - H)Z(\phi; h)\|_2^2$; the uncorrected estimator (right) is the minimizer of $\sum_{h\in\mathcal{H}} \|Z(\phi; h)\|_2^2$. For each set of values of τ, ρ and n, 500 data sets were generated

	true		percentile of estimate of τ (corrected)			percentile of estimate of τ (uncorrected)		
τ	ρ	n	50th	5th	95th	50th	5th	95th
0.5	0.25	400	0.595	0.502	0.735	1.547	1.111	2.546
0.5	0.25	900	0.533	0.477	0.598	0.851	0.752	1.008
0.5	0.25	1,600	0.511	0.471	0.552	0.678	0.620	0.740
0.5	0.5	400	0.720	0.498	1.924	2.689	1.182	30.566
0.5	0.5	900	0.567	0.461	0.735	1.094	0.839	1.762
0.5	0.5	1,600	0.526	0.448	0.605	0.790	0.652	0.980
0.5	0.75	400	3.956	0.479	25.751	67.544	2.182	210.201
0.5	0.75	900	0.907	0.443	10.156	5.154	1.146	129.679
0.5	0.75	1,600	0.596	0.374	1.429	1.496	0.802	7.349
1	0.25	400	1.090	0.912	1.311	1.627	1.308	2.130
1	0.25	900	1.036	0.926	1.164	1.270	1.148	1.447
1	0.25	1,600	1.012	0.931	1.092	1.152	1.062	1.248
1	0.5	400	1.166	0.861	2.143	2.017	1.300	3.959
1	0.5	900	1.077	0.873	1.354	1.428	1.181	1.901
1	0.5	1,600	1.027	0.878	1.179	1.233	1.050	1.451
1	0.75	400	2.531	0.717	35.367	7.260	1.407	296.729
1	0.75	900	1.311	0.740	11.082	2.246	1.225	11.575
1	0.75	1,600	1.077	0.699	2.271	1.598	1.042	3.168
2	0.25	400	2.084	1.741	2.473	2.489	2.154	3.046
2	0.25	900	2.047	1.826	2.295	2.249	2.036	2.506
2	0.25	1,600	2.009	1.848	2.171	2.140	1.974	2.321
2	0.5	400	2.134	1.607	3.274	2.746	2.037	4.344
2	0.5	900	2.096	1.706	2.590	2.374	1.992	3.007
2	0.5	1,600	2.027	1.742	2.328	2.204	1.925	2.540
2	0.75	400	2.766	1.219	63.219	4.005	1.744	71.464
2	0.75	900	2.271	1.395	16.040	2.865	1.860	6.321
2	0.75	1,600	2.054	1.371	3.989	2.492	1.746	4.164

as being due to an unknown long-range smooth trend, our method of Section 2 will estimate the parameters of the short-range process at a much lower cost. Consequently, one can obtain predictions of the overall anomaly process upon estimating $g(\cdot)$ by a smoothing method that corrects for dependence in errors [4, 11].

It is notable that our method does not assume that the whole process is stationary in the mean, which allows one to fit a wider range of models under less stringent assumptions.

To estimate parameters of the exponential covariance function, parameterized as

$$\mathrm{cov}\{\epsilon(x_i), \epsilon(x_j)\} = \sigma^2 \exp(-\lambda\|x_i - x_j\|_2),$$

we rescaled the distance (in miles) so that a unit distance equals 100 miles. The smallest rectangle that contains the spatial domain has side lengths $(28.78, 17.96)$ after the transformation. In our parameterization, the parameter values obtained by Furrer et al. in [5] for σ^2 and λ in the short-range process are 0.28 and 2.46, respectively.

To estimate the variogram parameters, it is important to use a set of directions and lags that "cover" the effective range of the short-range process. For the choice of directions $\mathcal{H} = \{(0,1), (1,0), (1,1), (-1,1), (2,1), (-2,1)\}/100$ and lags $k = 1, 2, \ldots, 70$, the estimates for σ^2 and λ are around 0.25 and 2.09, respectively. Using a different set of directions and lags to cover the same spatial range produced similar estimates, which is consistent with the representation in [5]. Considerably increasing the number of lags results in misleading parameter estimates because the long-range effect (in the representation of [5]) is getting partially captured by our short-range variogram model.

5. DISCUSSION AND EXTENSIONS

We have presented a nonlinear least squares estimator of the variogram parameters in nonparametric regression models with dependent errors. While making fewer assumptions, our procedure combines merits of the methods developed in [15] and [7], but does not suffer from their draw-

Table 2. Sample percentiles of estimates of ρ in the experiment of Section 3.2 with $g(x_1,x_2) = \sin(2\pi x_1) + 4(x_2 - 0.5^2)$ when directional correction for the variogram bias is used. The bias-corrected estimator (left) is the minimizer of $\sum_{h\in\mathcal{H}} \|(I_m - H)Z(\phi;h)\|_2^2$; the uncorrected estimator (right) is the minimizer of $\sum_{h\in\mathcal{H}} \|Z(\phi;h)\|_2^2$. For each set of values of τ, ρ and n, 500 data sets were generated

			percentile of estimate of τ (corrected)			percentile of estimate of τ (uncorrected)		
τ	true ρ	n	50th	5th	95th	50th	5th	95th
0.5	0.25	400	0.371	0.249	0.507	0.815	0.718	0.901
0.5	0.25	900	0.298	0.232	0.369	0.639	0.566	0.730
0.5	0.25	1,600	0.268	0.219	0.315	0.520	0.455	0.567
0.5	0.5	400	0.644	0.476	0.858	0.914	0.798	0.993
0.5	0.5	900	0.556	0.461	0.658	0.797	0.719	0.879
0.5	0.5	1,600	0.525	0.445	0.584	0.717	0.641	0.771
0.5	0.75	400	0.963	0.727	0.993	0.997	0.937	0.999
0.5	0.75	900	0.852	0.716	0.985	0.976	0.890	0.999
0.5	0.75	1,600	0.787	0.670	0.905	0.918	0.843	0.982
1	0.25	400	0.316	0.200	0.427	0.609	0.479	0.723
1	0.25	900	0.277	0.213	0.345	0.462	0.391	0.554
1	0.25	1,600	0.259	0.210	0.306	0.387	0.330	0.436
1	0.5	400	0.576	0.423	0.759	0.777	0.630	0.896
1	0.5	900	0.535	0.433	0.631	0.675	0.586	0.772
1	0.5	1,600	0.516	0.435	0.573	0.620	0.545	0.678
1	0.75	400	0.893	0.640	0.991	0.962	0.804	0.999
1	0.75	900	0.803	0.665	0.976	0.890	0.790	0.980
1	0.75	1,600	0.768	0.653	0.880	0.845	0.768	0.921
2	0.25	400	0.285	0.175	0.391	0.447	0.321	0.573
2	0.25	900	0.265	0.198	0.333	0.358	0.293	0.440
2	0.25	1,600	0.254	0.206	0.302	0.318	0.267	0.368
2	0.5	400	0.533	0.396	0.702	0.660	0.515	0.797
2	0.5	900	0.522	0.420	0.618	0.594	0.511	0.698
2	0.5	1,600	0.509	0.429	0.569	0.564	0.486	0.624
2	0.75	400	0.825	0.604	0.991	0.879	0.705	0.993
2	0.75	900	0.774	0.647	0.964	0.826	0.724	0.927
2	0.75	1,600	0.760	0.647	0.867	0.801	0.718	0.880

backs. In particular, tractable random fields include temporal, spatial, spatio-temporal, as well as general random fields with d-dimensional indices, including multivariate random fields [1]. Under standard regularity conditions of increasing domain asymptotics, the proposed estimator is consistent, and possesses attractive finite sample properties, which was studied through simulations. Proof of asymptotic normality along the lines of [9] is possible, but was not pursued since such results are not indicative of finite-sample operational characteristics of estimators.

Although we justified our estimator under lattice designs, the methodology extends easily to scattered data. In particular, it is applicable to spatio-temporal data when site locations are scattered and, possibly, sparse, but each site produces long readings over time — a scenario common in environmental statistics. In fact, this happens in the full precipitation data set, a subset of which we considered in Section 4: monthly averages are computed for readings from a network of stations operating, roughly, over the past 100 years [8]. Other famous spatio-temporal datasets that are spatially sparse but temporally dense include the Irish wind data [6] and Canadian weather data [12]. In these settings, one would form spatio-temporal sample variograms for all spatio-temporal lags of interest, but for each choice of a spatial lag, the binned variograms would be projected onto the null space of the vector of corresponding squared temporal lags. That is, there is a separate nuisance (i.e., variogram bias) parameter for each spatial lag.

To improve its efficiency, our estimator based on unweighted nonlinear least squares can be extended to weighted or to generalized nonlinear least squares. Since the projection matrix $(I_m - H) = QQ^{\mathrm{T}}$, where Q is an $m \times (m-1)$ matrix whose columns form an orthonormal basis for the null space of b^{T}, one can rewrite the criterion of Section 2.5 as

$$\sum_{h\in\mathcal{H}} \|(I_m - H)Z(\phi;h)\|_2^2 = \sum_{h\in\mathcal{H}} \|Q^{\mathrm{T}} Z(\phi;h)\|_2^2.$$

Replacing the standard Euclidean vector norm $\|\cdot\|_2$ with a generalized Euclidean norm defined for a vector v as

$\|v\|_A = \sqrt{v^{\mathrm{T}} A v}$ for some positive definite matrix A, one obtains $\sum_{h \in \mathcal{H}} \|Q^{\mathrm{T}} Z(\phi; h)\|^2_{W^{-1}(\phi;h)}$, where $W(\phi; h)$ is the large-sample covariance matrix of $Q^{\mathrm{T}} Z(\phi; h)$ in the case of generalized least squares, or a diagonal matrix of variances of $Q^{\mathrm{T}} Z(\phi; h)$ in the case of weighted least squares. In either case, the analytical expression for $W(\phi; h)$ can be obtained using the results in [9]. Use of weighted or generalized least squares in the variogram cloud criterion in (5) is not recommended even in the mean-stationary models, as the estimation is inconsistent without correct specification of the second moment (or the covariances) of s_{ij} [10].

When the goal is estimation of the parameters of the short-range spatial process, our procedure is a computationally more efficient alternative to a likelihood-based method that represents the response as a sum of long- and short-range processes, which can be computationally intractable due to the cost of matrix factorizations when the sample sizes measure in tens of thousands as in the case of modern spatial data sets. For a recent review on geostatistics for large data sets, see [14].

APPENDIX A

A.1 Outline of the proof of Proposition 1

Using (6) and (7), rewrite

$$S_k = V_k + a_k^2 J + W_k + O_p(a_k^3), \tag{13}$$

where $W_k = 2a_k(n-k)^{-1} \sum_{i=1}^{n-k} g'(x_{i+k})\{\epsilon(x_{i+k}) - \epsilon(x_i)\} = o_p(a_k)$. Define $S = (S_1, \ldots, S_m)^{\mathrm{T}}$ and $W = (W_1, \ldots, W_m)^{\mathrm{T}}$, so that $S = V + bJ + W + O_p\{(m/n)^3\}$.

Examine

$$\begin{aligned}
&(I_m - H)\{S - v(\phi)\} \\
&\quad = (I_m - H)[V + bJ - v(\phi) + W + O_p\{(m/n)^3\}] \\
&\quad = (I_m - H)\{V - v(\phi)\} + (I_m - H)[W + O_p\{(m/n)^3\}].
\end{aligned}$$

The second term converges in probability (i.p.) to 0, while the first converges i.p. to $(I_m - H)\{v(\phi^*) - v(\phi)\}$, by assumption (A1). Therefore, $\|(I_m - H)\{S - v(\phi^*)\}\|_2^2$ converges i.p. to 0. Since

$$\begin{aligned}
0 &\le \|(I_m - H)\{S - v(\widehat{\phi})\}\|_2^2 = \min_{\phi} \|(I_m - H)\{S - v(\widehat{\phi})\}\|_2^2 \\
&\le \|(I_m - H)\{S - v(\phi^*)\}\|_2^2,
\end{aligned}$$

$\|(I_m - H)\{S - v(\widehat{\phi})\}\|_2^2$ converges i.p. to $\|(I_m - H)\{v(\phi^*) - v(\widehat{\phi})\}\|_2^2 = 0$. By assumption (A2), $\widehat{\phi}$ converges i.p. to ϕ^*.

A.2 Selection of bandwidth parameters

Selection of the number of lags m and the set of directions $\mathcal{H}$ for variogram estimation, which we jointly refer to as bandwidth parameters, is a question of great practical importance. However, even in the simpler case when ϵ is a white noise process, the choice is far from obvious. In [15], the authors were able to derive an expression for the dominant terms of the mean squared error of their estimator, which was used to argue that an optimal bandwidth is $m = O(n^{1/2})$. However, their numerical experiments suggest that a choice $m = n^{1/3}$ is more appropriate in practice, particularly when g is "wiggly". The choice becomes even more difficult in our context because of the potential weak indentifiability between g and ϵ when g is highly variable but ϵ has slowly-decaying dependence.

Our procedure requires that ϕ is identifiable under our criterion, namely that, for all valid ϕ_0, ϕ_1, m and $\mathcal{H}$ be chosen so that

$$\sum_{h \in \mathcal{H}} \|(I_m - H)\{v(\phi_0) - v(\phi_1)\}\|_2^2 = 0$$

if and only if $\phi_0 = \phi_1$, where $v(\phi_i)$ is the vector of theoretical variogram values used in $\sum_{h \in \mathcal{H}} \|(I_m - H)Z(\phi; h)\|_2^2$. This requirement can be easily checked once a parametric variogram family is selected.

In practice, we recommend that the choice of m and $\mathcal{H}$ be tailored to each individual application based on a two-step pilot study as follows:

Step 1: Given a set of design points $\{x_1, \ldots, x_n\}$ and a variogram family, simulate a stochastic process ϵ for a set of plausible values of ϕ and estimate ϕ using a set of plausible choices of $\mathcal{H}$ and m. This is an analogue of least squares estimation of ϕ from the empirical variogram, which is widely used in spatial statistics. Additionally, this step helps one to determine if the variogram parameters can be estimated well (if at all) for a given design and a choice of the variogram family.

Step 2: For promising choices of $\mathcal{H}$ and m, repeat the simulation exercise incorporating a trend g suggested by an area of application, in order to narrow down the set of $\mathcal{H}$ and m. Since estimation of ϕ is typically very fast for given values of m and $\mathcal{H}$ unless m is large, one need not find "the best" pair, but rather a range, of bandwidth parameters giving answers consistent with each other.

In case the estimates of ϕ from a real data set are dramatically different from the values considered in the pilot study, the pilot study should be repeated.

In our simulation experiments, a choice of m of 5 or 10 worked well in all scenarios, although we used $m = n^{1/2}$ to give fair comparison with the procedure of [15]. This might explain why our estimator was slightly more variable than that of [7] in the first one-dimensional index experiment.

ACKNOWLEDGEMENTS

Bliznyuk's research was supported by a postdoctoral training grant from the National Cancer Institute (CA090301). Carroll's research was supported by a grant from the National Cancer Institute (R37-CA057030) and

Genton's research was partially supported by the National Science Foundation (DMS-1007504). Carroll and Genton were also supported by Award Number KUS-CI-016-04, made by King Abdullah University of Science and Technology (KAUST).

Received 1 March 2011

REFERENCES

[1] Apanasovich, T. V. and Genton, M. G. (2010). Cross-covariance functions for multivariate random fields based on latent dimensions, *Biometrika* **97**, 15–30. MR2594414

[2] Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2004). *Hierarchical Modeling and Analysis for Spatial Data*. Boca Raton: Chapman and Hall-CRC.

[3] Cressie, N. (1993). *Statistics for Spatial Data*. Chichester: John Wiley & Sons. MR1239641

[4] Francisco-Fernandez, M. and Opsomer, J. D. (2005). Smoothing parameter selection methods for nonparametric regression with spatially correlated errors, *The Canadian Journal of Statistics* **33**, 279–295. MR2193032

[5] Furrer, R., Genton, M. G., and Nychka, D. (2006). Covariance tapering for interpolation of large spatial datasets, *Journal of Computational and Graphical Statistics* **15**, 502–523. MR2291261

[6] Gneiting, T., Genton, M. G., and Guttorp, P. (2007). Geostatistical space-time models, stationarity, separability and full symmetry, in *Statistics of Spatio-Temporal Systems*, B. Finkenstaedt, L. Held, V. Isham (eds), Chapman & Hall/CRC Press, Monographs in Statistics and Applied Probability, pp. 151–175.

[7] Hall, P. and Van Keilegom, I. (2003). Using difference-based methods for inference in nonparametric regression with time series errors, *Journal of the Royal Statistical Society, Series B* **65**, 443–456. MR1983757

[8] Johns, C., Nychka, D., Kittel, T., and Daly, C. (2003). Infilling sparse records of spatial fields, *Journal of the American Statistical Association* **98**, 796–806. MR2055488

[9] Lahiri, S. N., Lee, Y., and Cressie, N. (2002). On asymptotic distribution and asymptotic effciency of least squares estimators of spatial variogram parameters, *Journal of Statistical Planning and Inference* **103**, 65–85. MR1896984

[10] Müller, W. G. (1999). Least-squares fitting from the variogram cloud, *Statistics & Probability Letters* **43**, 93–98. MR1707256

[11] Opsomer, J., Wang, Y., and Yang, Y. (2001). Nonparametric regression with correlated errors, *Statistical Science* **16**, 134–153 MR1861070

[12] Ramsey, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*, 2nd ed., New York: Springer–Verlag. MR2168993

[13] Santner, T. J., Williams, B. J., and Notz, W. (2003). *The Design and Analysis of Computer Experiments*. New York: Springer–Verlag. MR2160708

[14] Sun, Y., Li, B., and Genton, M. G. (2012). Geostatistics for large datasets, in *Space-Time Processes and Challenges Related to Environmental Problems*, E. Porcu, J. M. Montero, M. Schlather (eds), Chapter 3, Springer Lecture Notes in Statistics, Vol. 207, *to appear*.

[15] Tong, T. and Wang, Y. (2005). Estimating residual variance in nonparametric regression using least squares, *Biometrika* **92**, 821–830. MR2234188

Nikolay Bliznyuk
Department of Statistics
University of Florida
406 McCarty C
Gainesville, Florida 32611-0339
USA
E-mail address: nbliznyuk@stat.ufl.edu

Raymond J. Carroll
Department of Statistics
Texas A&M University
College Station, Texas 77843-3143
USA
E-mail address: carroll@stat.tamu.edu

Marc G. Genton
Department of Statistics
Texas A&M University
College Station, Texas 77843-3143
USA
E-mail address: genton@stat.tamu.edu

Yuedong Wang
Department of Statistics and Applied Probability
University of California – Santa Barbara
Santa Barbara, California 93106-3110
USA
E-mail address: yuedong@pstat.ucsb.edu