

On the estimation of parameters of variograms of spatial stationary isotropic random processes

Sourav Das¹, Tata Subba Rao² and Georgi Boshnakov²

¹National University of Singapore

²School of Mathematics, University of Manchester(UK)

Corresponding Author: Sourav Das, e-mail: sourav2.das@googlemail.com

Abstract

In many fields of science dealing with geostatistical data, the weighted least squares proposed by Noel Cressie remains a popular choice for variogram estimation. The main advantages of Simplicity (yet rigour), ease of implementation and non-parametric nature are its principle advantages. It also avoids the heavy computational burden of Generalized least squares. But that comes at the cost of loss of information due to the use of a diagonal weight matrix. Besides, the parameter dependent weight matrix makes the estimating equations biased. In this paper we propose two alternative weight matrices which do not depend on the parameters. We show that one of the weight matrices gives parameter estimates with lower asymptotic variance and also has asymptotically unbiased estimating equations. The observations are validated using simulation and real data.

Keywords: variance stabilisation, Matern class, wave variogram, cross-validation kriging.

1 Introduction

Weighted least squares method of variogram estimation was proposed by Cressie (1985) and has remained one of the frequently used methods for variogram estimation for geostatistical processes in applications. It has the following principal advantages- a) as a particular case of a least squares method, it does not require any distributional assumption on the original process b) it substantially reduces the computational burden involved in the inversion of a full variance covariance matrix of the variogram estimator and c) it retains the heteroscedasticity of the variogram estimator. These advantages make the method easy to interpret and implement. But it also has well documented limitations (see for example Curriero and Lele(1999)). Before discussing them let us introduce the method formally.

Let $\{Z(\mathbf{s}_1), Z(\mathbf{s}_2), \dots, Z(\mathbf{s}_n)\}'$ be the sample of size n from the second order stationary process $\{Z(\mathbf{s})\}$ observed at n fixed locations. Let $2\gamma(\mathbf{h}, \boldsymbol{\theta})$ be a valid theoretical variogram to which we want to fit the classical variogram estimator for this process, where $\{\boldsymbol{\theta}_{p \times 1} \in \Theta \subset \mathbb{R}^p\}$ is the vector of unknown parameters and Θ is an open parameter space. According to the generalized least squares (GLS) fitting procedure, parameters are chosen subject to minimization of a weighted sum of squares.

$$Q(\boldsymbol{\theta}) = \mathbf{g}'(\boldsymbol{\theta})V(\boldsymbol{\theta})\mathbf{g}(\boldsymbol{\theta}), \quad (1)$$

where $\mathbf{g}(\boldsymbol{\theta}) = \{2\hat{\gamma}(\mathbf{h}_1) - 2\gamma(\mathbf{h}_1, \boldsymbol{\theta}), \dots, 2\hat{\gamma}(\mathbf{h}_k) - 2\gamma(\mathbf{h}_k, \boldsymbol{\theta})\}'$ and $V(\boldsymbol{\theta})$ is the weight matrix, which is usually taken as the inverse of the dispersion matrix of the sample variogram, $E[\mathbf{g}(\boldsymbol{\theta})\mathbf{g}'(\boldsymbol{\theta})]$. Thus the method gives more weight to sample variograms with small variance and less weight to sample variograms with large variance (usually as the lag distance increases). But a full variance covariance

matrix $V(\boldsymbol{\theta})$ (assuming it can be calculated!) introduces a huge computational cost. The challenge is then to propose a weight matrix that on one hand reduces the computational burden but on the other reduces the loss in efficiency.

Cressie (1985) proposed that if the random process $Z(\mathbf{s})$ is Gaussian with a constant mean, then $Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s}) \sim N(0, 2\gamma(\mathbf{h}))$. This implies that $\{Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s})\}^2/2\gamma(\mathbf{h}, \boldsymbol{\theta}) \sim \chi_1^2$, and $\text{Var}[\{Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s})\}^2] = 2(2\gamma(\mathbf{h}, \boldsymbol{\theta}))^2$. Now, if the "between-lag" covariances of the variogram clouds (see Cressie(1993)) are weak, that is

$$\text{Cov}[\{Z(\mathbf{s}_i + \mathbf{h}_l) - Z(\mathbf{s}_i)\}^2, \{Z(\mathbf{s}_j + \mathbf{h}_m) - Z(\mathbf{s}_j)\}^2] \simeq 0,$$

for all $i, j = 1, 2, \dots, n$ and $l, m = 1, 2, \dots, k$ the above implies that, $\text{Var}[2\gamma(\hat{\mathbf{h}}_i)] \simeq 2(2\gamma(\mathbf{h}_i, \boldsymbol{\theta}))^2/|N(\mathbf{h}_i)|$. This lead Cressie to propose the following weighted least squares (WLS) criterion for parametric variogram estimation

$$Q_{1n}(\boldsymbol{\theta}) = \sum_{i=1}^k [2\hat{\gamma}(\mathbf{h}_i) - 2\gamma(\mathbf{h}_i, \boldsymbol{\theta})]^2 v_{1i}(\boldsymbol{\theta}), \quad (2)$$

where $v_{1i}(\boldsymbol{\theta}) = \frac{|N(\mathbf{h}_i)|}{2(2\gamma(\mathbf{h}_i, \boldsymbol{\theta}))^2}, \quad i = 1, \dots, k.$

Here k denotes the number of distinct lag distances for which the sample and theoretical variograms are computed. And $v_{1i}(\boldsymbol{\theta})$ is the weight function. The method is simple yet powerful provided that the assumptions on Gaussianity and covariance are valid. But apart from the Gaussianity assumption (which Cressie relaxes later) the assumption on the covariance is quite strong for many geostatistical processes. Consequently this weight function, despite being an improvement over the ordinary least squares, leaves scope for further improvement on the efficiencies of parameter estimators. Also, the fact that the weight function depends on the parameters $\boldsymbol{\theta}$ leads to biased generalized estimating equations (see e.g Diggle(2007, p. 109)) for $Q_{1n}(\boldsymbol{\theta})$.

Here we consider two alternative weight functions, $v_{2i}(\boldsymbol{\theta})$ and $v_{3i}(\boldsymbol{\theta})$ as improvements in these regards and illustrate that even within the framework of WLS we can improve the efficiency of the estimated parameters. The alternative criteria we propose are

$$Q_{2n}(\boldsymbol{\theta}) = \sum_{i=1}^k [2\hat{\gamma}(\mathbf{h}_i) - 2\gamma(\mathbf{h}_i, \boldsymbol{\theta})]^2 v_{2i}(\boldsymbol{\theta}) \quad \text{and} \quad Q_{3n}(\boldsymbol{\theta}) = \sum_{i=1}^k [\ln 2\hat{\gamma}(\mathbf{h}_i) - \ln 2\gamma(\mathbf{h}_i, \boldsymbol{\theta})]^2 v_{3i}(\boldsymbol{\theta}) \quad (3)$$

where the weight functions are

$$v_{2i}(\boldsymbol{\theta}) = |N(\mathbf{h}_i)| / \left[\frac{1}{|N(\mathbf{h}_i)|} \sum_{N(\mathbf{h}_i)} [\{Z(\mathbf{s}_l) - Z(\mathbf{s}_m)\}^2 - 2\hat{\gamma}(\mathbf{h}_i)]^2 \right]$$

$$v_{3i}(\boldsymbol{\theta}) = \frac{|N(\mathbf{h}_i)|}{2} \quad (4)$$

In constructing $v_{2i}(\boldsymbol{\theta})$ we replace the unknown parametric function $v_{1i}(\boldsymbol{\theta})$ (see (2)) by its sample counterpart.

Thus $v_{2i}(\boldsymbol{\theta})$ is the inverse of the sample estimator of $\text{Var}[2\gamma(\hat{\mathbf{h}}_i)]$. It doesn't depend on $\boldsymbol{\theta}$. Later we present a comparative analysis of the standard errors and mean squared errors of estimates of $\boldsymbol{\theta}$, obtained from WLS using $v_{2i}(\boldsymbol{\theta})$ and $v_{1i}(\boldsymbol{\theta})$.

$v_{3i}(\boldsymbol{\theta})$ on the other hand is based on the theory of variance stabilization. It is well known that if the variance of a random variable $2\gamma(\hat{\mathbf{h}}_i)$ is proportional to its population mean that is,

$$\text{Var}[2\gamma(\hat{\mathbf{h}}_i)] \propto (2\gamma(\mathbf{h}_i, \boldsymbol{\theta}))^2,$$

the logarithmic transformation of $2\hat{\gamma}(\mathbf{h}_i)$ has a variance proportional to $\frac{2}{|N(\mathbf{h}_i)|}$. Thus, we use $\frac{|N(\mathbf{h}_i)|}{2}$ as the weight function, $v_{3i}(\boldsymbol{\theta})$, in $Q_{3n}(\boldsymbol{\theta})$.

It can be checked that parameter free weight function $v_{3i}(\boldsymbol{\theta})$ also makes the estimating equations asymptotically unbiased, which is a common criticism against the WLS method for variogram fitting (see for example Curriero and Lele(1999)). Also we show later in Section 2, using a general result obtained by Lahiri, Lee and Cressie (2002), that the main advantage of using $v_{3i}(\boldsymbol{\theta})$ is, we obtain estimators with smaller asymptotic variance.

In Section 2 we compare the asymptotic variances of the parameter estimators obtained from $Q_{1n}(\boldsymbol{\theta})$ with that of $Q_{3n}(\boldsymbol{\theta})$. In Section 3 we present the simulation results.

2 Comparison of the efficiencies of estimators

Lahiri, Lee and Cressie (2002, sec. 3) have established asymptotic normality for estimators of variogram parameters obtained using least squares. Since $\ln(\gamma(\mathbf{h}, \boldsymbol{\theta}))$ is differentiable function of $\gamma(\mathbf{h}, \boldsymbol{\theta})$, an application of Theorem 3.2 in Lahiri, Lee and Cressie(2002) and delta method establishes the asymptotic normality of $\ln\{2\hat{\gamma}(\mathbf{h}_i)\}$. We now evaluate the asymptotic variances of estimators of $\boldsymbol{\theta}$ obtained by minimizing $Q_{1n}(\boldsymbol{\theta})$ and $Q_{3n}(\boldsymbol{\theta})$, using the results of Lahiri, Lee and Cressie (2002). We show that estimators of $\boldsymbol{\theta}$ obtained by minimizing $Q_{3n}(\boldsymbol{\theta})$ have smaller asymptotic variance. Let

$\boldsymbol{\theta}_1$: unique minimizer of the first objective function

$\boldsymbol{\theta}_3$: unique minimizer of the third objective function Let $\nabla_l \mathbf{g}(\boldsymbol{\theta}_0)$ be the vector of first derivatives of $\mathbf{g}(\boldsymbol{\theta}_0)$ and $D(\boldsymbol{\theta}_0)$ the matrix of first derivatives. In this section we use the suffix 1 ($\nabla_l \mathbf{g}_1(\boldsymbol{\theta}_0)$, $D_1(\boldsymbol{\theta}_0)$) to indicate the corresponding matrices due to the first criterion $Q_{1n}(\boldsymbol{\theta})$. We also define,

$$\mathbf{g}_3(\boldsymbol{\theta}) = (\ln 2\hat{\gamma}(\mathbf{h}_1) - \ln 2\gamma(\mathbf{h}_1, \boldsymbol{\theta}), \ln 2\hat{\gamma}(\mathbf{h}_2) - \ln 2\gamma(\mathbf{h}_2, \boldsymbol{\theta}), \dots, \ln 2\hat{\gamma}(\mathbf{h}_k) - \ln 2\gamma(\mathbf{h}_k, \boldsymbol{\theta}))'_{(k \times 1)}$$

The vector of the first derivatives of $\mathbf{g}_3(\boldsymbol{\theta})$, with respect to θ_l is denoted as $\nabla_l \mathbf{g}_3(\boldsymbol{\theta})$, while $D_3(\boldsymbol{\theta}_0)$ denotes the first derivative matrix. Now using the approximation due to Cressie((1985)) and Theorem 3.2 in Lahiri, Lee and Cressie (2002) we note that for the first criterion $Q_{1n}(\boldsymbol{\theta})$, we have $\mathbf{g}(\boldsymbol{\theta}) \simeq \text{MVN}(\mathbf{0}, \Sigma_1(\boldsymbol{\theta}_0))$ where $\Sigma_1(\boldsymbol{\theta}_0) = \text{diag}(2(2\gamma(\mathbf{h}_i, \boldsymbol{\theta}))^2/|N(\mathbf{h}_i)|)_{i=1,2,\dots,K}$. Similarly using Theorem 3.2 in Lahiri, Lee and Cressie (2002) and delta method for $Q_{3n}(\boldsymbol{\theta})$, we have $\mathbf{g}_3(\boldsymbol{\theta}) \simeq \text{MVN}(\mathbf{0}, \Sigma_3(\boldsymbol{\theta}_0))$, where

$$\Sigma_3(\boldsymbol{\theta}_0) = \text{diag}(2/|N(\mathbf{h}_i)|)_{i=1,2,\dots,K} \quad (5)$$

Let $\hat{\boldsymbol{\theta}}_{1n}$ and $\hat{\boldsymbol{\theta}}_{3n}$ denote the parameter estimators of $\boldsymbol{\theta}$ obtained by minimizing $Q_1(\boldsymbol{\theta})$ and $Q_3(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$, respectively. Also let $V_1(\boldsymbol{\theta}) = [\Sigma_1^{-1}(\boldsymbol{\theta})]$ and $V_3(\boldsymbol{\theta}) = [\Sigma_3^{-1}(\boldsymbol{\theta})]$, the inverses corresponding to the covariance matrices of the sample variogram . Then using equation (5), Theorem 3.2 and Corollary 3.2 from Lahiri, Lee and Cressie (2002) we have

$$\begin{aligned} \hat{\boldsymbol{\theta}}_{1n} &\approx \text{MVN}(\boldsymbol{\theta}_0, \Sigma_{v_1}(\boldsymbol{\theta}_0)) \\ \Sigma_{v_1}(\boldsymbol{\theta}_0) &= B_1(\boldsymbol{\theta}_0)D_1(\boldsymbol{\theta}_0)'V_1(\boldsymbol{\theta}_0)\Sigma_1(\boldsymbol{\theta}_0)V_1(\boldsymbol{\theta}_0)D_1(\boldsymbol{\theta}_0)B_1(\boldsymbol{\theta}_0) \\ B_1(\boldsymbol{\theta}_0) &= [(D_1(\boldsymbol{\theta}_0)'V_1(\boldsymbol{\theta}_0)D_1(\boldsymbol{\theta}_0))]^{-1} \\ \hat{\boldsymbol{\theta}}_{3n} &\approx \text{MVN}(\boldsymbol{\theta}_0, \Sigma_{v_3}(\boldsymbol{\theta}_0)) \\ \Sigma_{v_3}(\boldsymbol{\theta}_0) &= B_3(\boldsymbol{\theta}_0)D_3(\boldsymbol{\theta}_0)'V_3(\boldsymbol{\theta}_0)\Sigma_3(\boldsymbol{\theta}_0)V_3(\boldsymbol{\theta}_0)D_3(\boldsymbol{\theta}_0)B_3(\boldsymbol{\theta}_0) \\ B_3(\boldsymbol{\theta}_0) &= [(D_3(\boldsymbol{\theta}_0)'V_3(\boldsymbol{\theta}_0)D_3(\boldsymbol{\theta}_0))]^{-1} \end{aligned} \quad (6)$$

From these expressions it's not hard to deduce that

$$\begin{aligned}\text{Var}[\hat{\boldsymbol{\theta}}_1] &\rightarrow D_1(\boldsymbol{\theta}_0)^{-1}\Sigma_1(\boldsymbol{\theta}_0)D_1'(\boldsymbol{\theta}_0)^{-1} \\ \text{Var}[\hat{\boldsymbol{\theta}}_3] &\rightarrow D_3(\boldsymbol{\theta}_0)^{-1}\Sigma_3(\boldsymbol{\theta}_0)D_3'(\boldsymbol{\theta}_0)^{-1}\end{aligned}$$

First, let us consider the difference between the i^{th} diagonal terms of $\Sigma_1(\boldsymbol{\theta})$ and $\Sigma_3(\boldsymbol{\theta})$. Note that,

$$\frac{2(2\gamma(\mathbf{h}_i, \boldsymbol{\theta}_0))^2}{|N(\mathbf{h}_i)|} - 2/|N(\mathbf{h}_i)| > 0 \quad \text{if } \text{Var}(Z(\mathbf{s}_l) - Z(\mathbf{s}_m)) > 1^1, \text{ for all } l, m \in N(\mathbf{h}_i), \quad (7)$$

Then

$$\begin{aligned}\Sigma_{v_1} - \Sigma_{v_3} &= D_1(\boldsymbol{\theta}_0)^{-1}\Sigma_1(\boldsymbol{\theta}_0)D_1'(\boldsymbol{\theta}_0)^{-1} - D_3(\boldsymbol{\theta}_0)^{-1}\Sigma_3(\boldsymbol{\theta}_0)D_3'(\boldsymbol{\theta}_0)^{-1} \\ &> D_3(\boldsymbol{\theta}_0)^{-1}\Sigma_1(\boldsymbol{\theta}_0)D_3'(\boldsymbol{\theta}_0)^{-1} - D_3(\boldsymbol{\theta}_0)^{-1}\Sigma_3(\boldsymbol{\theta}_0)D_3'(\boldsymbol{\theta}_0)^{-1} \\ &= D_3(\boldsymbol{\theta}_0)^{-1}[\Sigma_1(\boldsymbol{\theta}_0) - \Sigma_3(\boldsymbol{\theta}_0)]D_3'(\boldsymbol{\theta}_0)^{-1} \\ &\geq 0,\end{aligned} \quad (8)$$

since $[\Sigma_1(\boldsymbol{\theta}_0) - \Sigma_3(\boldsymbol{\theta}_0)] = \text{diag}\left(\frac{2(2\gamma(\mathbf{h}_i, \boldsymbol{\theta}_0))^2}{|N(\mathbf{h}_i)|} - 2/|N(\mathbf{h}_i)|\right)$ and thus by (7) $[\Sigma_1(\boldsymbol{\theta}_0) - \Sigma_3(\boldsymbol{\theta}_0)] > \mathbf{0}$. Thus we observe that for large samples, drawn under a mixed increasing domain sampling scheme, the asymptotic variance of the parameter estimators obtained by minimizing $Q_3(\boldsymbol{\theta})$ is smaller than those obtained using $Q_1(\boldsymbol{\theta})$.

Note: Here we have not provided theoretical comparison of the performance of estimators obtained using $Q_1(\boldsymbol{\theta})$ with those obtained using $Q_2(\boldsymbol{\theta})$. But in the next two sections we present empirical evidence that the mean squared errors of the parameter estimators of $\boldsymbol{\theta}$ obtained by minimizing $Q_2(\boldsymbol{\theta})$ are smaller than those obtained using $Q_1(\boldsymbol{\theta})$ for the majority of the examples considered.

3 Discussion of simulation results

In this section we compare empirically, the performance of parameter estimators of variogram parameters using the weight functions $v_{1i}(\boldsymbol{\theta}), v_{2i}(\boldsymbol{\theta})$ and $v_{3i}(\boldsymbol{\theta})$, for various second order stationary spatial processes.

Henceforth, we will refer to the three different weight functions vis-a-vis criteria as $v_1(\boldsymbol{\theta}), v_2(\boldsymbol{\theta})$ and $v_3(\boldsymbol{\theta})$ respectively. Here we compare the performances of the parameter estimators for the proposed weights $v_2(\boldsymbol{\theta})$ and $v_3(\boldsymbol{\theta})$ with $v_1(\boldsymbol{\theta})$ for a Gaussian spatial process with Matern class variogram functions (see Stein(1999)). For a more comprehensive simulation under different distributional assumptions and real data analysis see Das (2011). The Matern class is defined as

$$\begin{aligned}\gamma(\mathbf{h}; \nu, \phi, \sigma) &= \sigma^2 - \sigma^2 \frac{1}{\Gamma(\nu) 2^{(\nu-1)}} \left(\frac{\|\mathbf{h}\|}{\phi}\right)^\nu K_\nu\left(\frac{\|\mathbf{h}\|}{\phi}\right) \quad \nu > 0, \phi > 0, \\ K_\nu(\cdot) &\text{ is the modified Bessel function of the second kind.}\end{aligned} \quad (9)$$

Here, σ^2 is the variance of the process. The parameter ν , also called the order of the Matern class, is a shape parameter which characterizes the smoothness of the random process $Z(\mathbf{s})$. The scale

¹This is not a very restrictive assumption since, for geostatistical data emerging from mining and many weather attributes of meteorology $\text{Var}(Z(\mathbf{s}_l) - Z(\mathbf{s}_m)) > 1$ is a common observation. Note also that in the neighbourhood of $\mathbf{h} \rightarrow 0$ we expect $2\gamma(\mathbf{h}_i, \boldsymbol{\theta}_0) \rightarrow 0$. But for purposes of parameter estimation of variogram functions for common geostatistical processes, the minimum lag distance at which the variogram function is computed is large enough such that the assumption is satisfied.

parameter ϕ is proportional to the practical range of the variogram function and $1/\phi$ is the rate at which the covariance function decays to zero as $\|\mathbf{h}\| \rightarrow \infty$. Further details on the empirical and theoretical properties of the Matern class functions can be found in Matern (1986), Stein (1999) and Diggle, Tawn and Moyeed(1998). For each of the above mentioned variogram functions a Gaussian spatial process is generated and sampled at 100 fixed locations on a two dimensional plane using the commonly used Cholesky decomposition method (see e.g Hadley(1961), Cressie (1993) and Schabenberger and Gotway (2005)). The locations are kept fixed for all simulations so that the distance matrix and the isotropic covariance matrix is fixed. The maximum euclidean distance among all pairs, for the simulated locations, was approximately 130 units. The computations were done with the statistical system **R** ((2012)).

3.1 Results and discussion

Given below are the tables of parameter estimates and their mean squared errors (MSE). In Table 1 we present the estimators of the parameter σ for 200 independent simulations, along with their MSEs obtained using the three weight functions. The corresponding results for the parameter ϕ are given in Table 2. The process variance σ^2 are chosen to be 20. While the range parameters ϕ are chosen based on practical range (see for example Schabenberger and Gotway (2005)) of 100 units. The chosen values are 16.69041 for exponential variogram, 10.49 for Matern 1, 8.37 for Matern 1.5 and 28.89 for the Gaussian variogram.

Table 1: Sampling Properties of $\hat{\sigma}$ (Gaussian Random Process)

Variogram	$v_1(\boldsymbol{\theta})$			$v_2(\boldsymbol{\theta})$			$v_3(\boldsymbol{\theta})$		
	Mean	MSE	Bias	Mean	MSE	Bias	Mean	MSE	Bias
Exponential	4.55	1.24	0.08	4.42	0.86	-0.05	4.44	0.70	-0.03
Matern 1	4.46	0.74	-0.02	4.32	0.54	-0.16	4.36	0.49	-0.12
Matern 1.5	4.44	0.75	-0.04	4.29	0.54	-0.18	4.33	0.50	-0.14
Gaussian	4.39	1.70	-0.08	4.24	1.58	-0.23	4.30	1.53	-0.17

We observe that for all Matern class variograms chosen, the MSEs of the estimators of the parameters obtained using both $v_2(\boldsymbol{\theta})$ and $v_3(\boldsymbol{\theta})$ are smaller than the MSE of the estimators obtained using $v_1(\boldsymbol{\theta})$. The estimates obtained using $v_3(\boldsymbol{\theta})$ have the smallest MSE for all variograms considered here.

Table 2: Sampling properties of estimates of $\hat{\phi}$ (Gaussian Random Process)

Variogram	$v_1(\boldsymbol{\theta})$			$v_2(\boldsymbol{\theta})$			$v_3(\boldsymbol{\theta})$		
	Mean	MSE	Bias	Mean	MSE	Bias	Mean	MSE	Bias
Exponential	22.03	616.00	5.34	20.12	281.93	3.43	18.91	167.12	2.21
Matern 1	11.94	50.00	1.45	10.95	20.43	0.46	10.64	13.19	0.15
Matern 1.5	9.36	26.11	0.99	8.53	9.41	0.16	8.32	5.77	-0.05
Gaussian	29.71	119.94	0.82	28.91	110.63	0.02	28.63	84.47	-0.26

We observe from Table 2 that among the distributions considered here, for the exponential variogram estimators of ϕ have much higher MSEs for all weight functions considered. But the MSE obtained using $v_3(\boldsymbol{\theta})$ is always smaller than $v_1(\boldsymbol{\theta})$. For the Gaussian random process, for all variogram functions, estimators obtained using $v_3(\boldsymbol{\theta})$ have the smallest MSEs followed by estimates obtained using $v_2(\boldsymbol{\theta})$. For all variograms considered in this paper we observe that the MSEs of the

estimators of ϕ obtained using the proposed weight functions $v_2(\boldsymbol{\theta})$ and $v_3(\boldsymbol{\theta})$ are much lower than those obtained using $v_1(\boldsymbol{\theta})$. We observe that by using $v_3(\boldsymbol{\theta})$ we obtain estimators with smallest mean squared errors for all the variograms and distributions considered here.

4 Acknowledgements

The work presented in this article was pursued at the University of Manchester, with sponsorship from British Council under the UKIERI research grant and the first author received additional funds from Department of Science and Technology, as a Research Associate at the C.R.Rao AIMSCS, for preparing the article. The authors would like to thank Dr. Peter Neal of University of Manchester, Prof. Peter Diggle of University of Lancaster and Dr. Suhasini Subbarao of Texas AM University for many useful suggestions that helped to improve the article.

References

- Cressie, N. (1985). “Fitting variogram models by weighted least squares”. *Mathematical geology* 17.5, pp. 563–586.
- Cressie, N.A.C. (1993). *Statistics for spatial data*. John Wiley & Sons, New York.
- Curriero, F. C. and Lele, S. (1999). “A composite likelihood approach to semivariogram estimation”. *Journal of Agricultural, biological, and Environmental statistics*, pp. 9–28.
- Das, S. (2011). “Statistical Estimation of Variogram and Covariance Parameters of Spatial and Spatio-Temporal Random Processes”. PhD thesis. School of Mathematics.
- Diggle, Peter J and Ribeiro, Paulo J (2007). *Model-based geostatistics*. Springer.
- Diggle, P.J., Tawn, J.A., and Moyeed, R.A. (1998). *Model-based geostatistics*. Vol. 47. 3. Wiley, pp. 299–350.
- Hadley, G. (1961). *Linear Algebra (Reading)*.
- Lahiri, S.N., Lee, Y., and Cressie, N. (2002). “On asymptotic distribution and asymptotic efficiency of least squares estimators of spatial variogram parameters”. *Journal of Statistical Planning and Inference* 103.1-2, pp. 65–85.
- Matérn, B. (1986). *Spatial variation*. Springer-Verlag Berlin.
- R Development Core Team (2012). *R: A Language and Environment for Statistical Computing*. ISBN 3-900051-07-0. R Foundation for Statistical Computing. Vienna, Austria. URL: <http://www.R-project.org/>.
- Schabenberger, O. and Gotway, C.A. (2005). *Statistical methods for spatial data analysis*. CRC Press. ISBN: 1584883227.
- Stein, M.L. (1999). *Interpolation of Spatial Data: some theory for kriging*. Springer Verlag.