



An asymptotically unbiased weighted least squares estimation criterion for parametric variograms of second order stationary geostatistical processes

DOI:
[10.1080/03610918.2018.1508698](https://doi.org/10.1080/03610918.2018.1508698)

Document Version
Accepted author manuscript

[Link to publication record in Manchester Research Explorer](#)

Citation for published version (APA):
Boshnakov, G., Das, S., & Subba Rao, T. (2020). An asymptotically unbiased weighted least squares estimation criterion for parametric variograms of second order stationary geostatistical processes. *Communications in Statistics: Simulation and Computation*, 49(7), 1839-1854. <https://doi.org/10.1080/03610918.2018.1508698>

Published in:
Communications in Statistics: Simulation and Computation

Citing this paper
Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights
Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy
If you believe that this document breaches copyright please refer to the University of Manchester's Takedown Procedures [<http://man.ac.uk/04Y6Bo>] or contact uml.scholarlycommunications@manchester.ac.uk providing relevant details, so we can investigate your claim.



An asymptotically unbiased weighted least squares estimation criterion for parametric variograms of second order stationary geostatistical processes

Sourav Das¹, Tata Subba Rao² and Georgi N. Boshnakov²

¹Epicentre, University of Melbourne

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 Cressie (1985) remains a popular choice for variogram estimation. Simplicity, 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, Matérn class, wave variogram, cross-validation kriging.

1 Introduction

The 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. A major advantage is that it does not require any distributional assumptions on the original process. Also, the use of a diagonal weight matrix instead of a full variance-covariance matrix, substantially reduces the computational burden, compared to generalized least squares or Gaussian maximum likelihood estimation. This approach also preserves the heteroscedasticity of the variogram estimator. Such properties make the method popular among practitioners. In fact, for these reasons the same criterion remains popular among scientists working with spatio-temporal data under assumptions of symmetry within the model based paradigm (see for example Cressie & Huang 1999, Gneiting 2002 and Gneiting et al. 2007). But it also has well documented limitations (see for example Curriero & Lele 1999, Subba Rao & Terdik 2017). Before discussing them we introduce the method formally.

Let $\{Z(\mathbf{s}_1), Z(\mathbf{s}_2), \dots, Z(\mathbf{s}_n)\}'$ be a sample from a second order stationary spatial process $Z(\mathbf{s})$ observed at n fixed locations. Let $2\gamma(\mathbf{h}, \boldsymbol{\theta})$ be the theoretical variogram for this process, defined as

$$2\gamma(\mathbf{h}, \boldsymbol{\theta}) = \text{Var}\{Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s})\}, \quad (1)$$

to which we want to fit the classical variogram estimator, defined by Matheron (Matheron 1963),

$$2\hat{\gamma}(\mathbf{h}) = \frac{1}{|N(\mathbf{h})|} \sum_{N(\mathbf{h})} \{Z(\mathbf{s}_i) - Z(\mathbf{s}_j)\}^2, \quad (2)$$

where $N(\mathbf{h})$ is the set of distinct pairs $(\mathbf{s}_i, \mathbf{s}_j)$, such that $N(\mathbf{h}) = \{(\mathbf{s}_i, \mathbf{s}_j) : \mathbf{s}_i - \mathbf{s}_j = \mathbf{h}\}$, $|N(\mathbf{h})|$ is the cardinality of $N(\mathbf{h})$, $\boldsymbol{\theta}_{p \times 1} \in \Theta \subset \mathbb{R}^p$ is the vector of unknown parameters and Θ is a compact space. In generalized least squares (GLS) estimation, parameters are chosen subject to minimization of the following weighted sum of squares.

$$Q(\boldsymbol{\theta}) = \mathbf{g}'(\boldsymbol{\theta})V(\boldsymbol{\theta})\mathbf{g}(\boldsymbol{\theta}), \quad (3)$$

where $\mathbf{g}(\boldsymbol{\theta}) = \{2\hat{\gamma}(\mathbf{h}_1) - 2\gamma(\mathbf{h}_1, \boldsymbol{\theta}), \dots, 2\hat{\gamma}(\mathbf{h}_k) - 2\gamma(\mathbf{h}_k, \boldsymbol{\theta})\}'$. $2\hat{\gamma}(\mathbf{h}_1), 2\hat{\gamma}(\mathbf{h}_2), \dots, 2\hat{\gamma}(\mathbf{h}_k)$ are the sample variogram estimators, see equation (2), calculated at k different lag distances $\{\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_k\}$. $V(\boldsymbol{\theta})$ is a weight matrix, which is usually taken to be the inverse of the dispersion matrix, $E[\mathbf{g}(\boldsymbol{\theta})\mathbf{g}'(\boldsymbol{\theta})]$, of the sample variogram. 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 $V^{-1}(\boldsymbol{\theta})$ introduces a huge computational cost. For the practitioner the challenge is to propose an alternative weight matrix that maintains a balance between computational cost and efficiency loss.

If the random process $Z(\mathbf{s})$ is Gaussian, then $Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s}) \sim N(0, 2\gamma(\mathbf{h}, \boldsymbol{\theta}))$. 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, Ch 2) are weak, that is

$$\text{Cov} [\{Z(\mathbf{s}_i + \mathbf{h}_l) - Z(\mathbf{s}_i)\}^2, \{Z(\mathbf{s}_j + \mathbf{h}_m) - Z(\mathbf{s}_j)\}^2] \simeq 0, \quad (4)$$

for all $i, j = 1, 2, \dots, n$ and $l, m = 1, 2, \dots, k$, then one may ignore the off-diagonal elements of the variance-covariance matrix to get

$$\text{Var}[2\hat{\gamma}(\mathbf{h}_i)] \simeq 2(2\gamma(\mathbf{h}_i, \boldsymbol{\theta}))^2 / |N(\mathbf{h}_i)|. \quad (5)$$

Based on these approximations, Cressie (1985) proposed the weighted least squares (WLS) criterion (the weighting matrix with diagonal elements) for parametric variogram estimation

$$Q_{1n}(\boldsymbol{\theta}) = \sum_{i=1}^k [2\hat{\gamma}(\mathbf{h}_i) - 2\gamma(\mathbf{h}_i, \boldsymbol{\theta})]^2 v_{1i}(\boldsymbol{\theta}), \quad (6)$$

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

$v_{1i}(\boldsymbol{\theta})$ denotes the weight function. The method is simple yet efficient, provided that the assumptions on Gaussianity and covariance are valid. But apart from the Gaussianity assumption (that is relaxed later in the paper) the assumption on the covariance is quite strong for many real 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. Further, the dependence of the weight function $v_{1i}(\boldsymbol{\theta})$ on the parameters $\boldsymbol{\theta}$ leads to biased generalized estimating equations (see e.g Diggle & Ribeiro 2007, p. 109) corresponding to $Q_{1n}(\boldsymbol{\theta})$.

In this paper we propose two alternative weighted least squares criteria that result in unbiased generalized estimating equations and reduce the asymptotic variance of parameters, while retaining the diagonal nature of the weight matrix. In Section 2 we present the proposals. In Section 3 we

compare the asymptotic variance of variogram estimates of one of the proposed criteria with $Q_{1n}(\boldsymbol{\theta})$. Simulation studies are given in Section 4. We simulate from both Gaussian and non-Gaussian processes. In Section 5 the methods are applied to a spatial rainfall data and we demonstrate the advantage of using a least squares based variogram estimation method compared to model based maximum likelihood estimation. Conclusions are given in Section 6.

2 Modified weighted least squares

We consider two alternative weight functions, $v_{2i}(\boldsymbol{\theta})$ and $v_{3i}(\boldsymbol{\theta})$, and illustrate that even within the framework of WLS we can improve the efficiency of the parameter estimates. The alternative criteria we propose are

$$\begin{aligned} Q_{2n}(\boldsymbol{\theta}) &= \sum_{i=1}^k [2\hat{\gamma}(\mathbf{h}_i) - 2\gamma(\mathbf{h}_i, \boldsymbol{\theta})]^2 v_{2i}(\boldsymbol{\theta}) \text{ and} \\ Q_{3n}(\boldsymbol{\theta}) &= \sum_{i=1}^k [\ln 2\hat{\gamma}(\mathbf{h}_i) - \ln 2\gamma(\mathbf{h}_i, \boldsymbol{\theta})]^2 v_{3i}(\boldsymbol{\theta}) \end{aligned} \quad (7)$$

where the weight functions are

$$\begin{aligned} v_{2i}(\boldsymbol{\theta}) &= |N(\mathbf{h}_i)| / \left[\frac{1}{|N(\mathbf{h}_i)|} \sum_{N(\mathbf{h}_i)} \{Z(\mathbf{s}_l) - Z(\mathbf{s}_m)\}^2 - 2\hat{\gamma}(\mathbf{h}_i) \right]^2 \\ v_{3i}(\boldsymbol{\theta}) &= \frac{|N(\mathbf{h}_i)|}{2} \end{aligned}$$

Note: Since $v_{2i}(\boldsymbol{\theta})$ and $v_{3i}(\boldsymbol{\theta})$ do not depend on $\boldsymbol{\theta}$, henceforth we denote them by v_{2i} and v_{3i} , respectively. In constructing v_{2i} we replace the unknown parametric function $v_{1i}(\boldsymbol{\theta})$ in equation (6) by its sample counterpart.

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

v_{3i} , on the other hand, is based on the theory of variance stabilization. It is well known that when the variance of a random variable is proportional to its population mean then a logarithm of the random variable has a constant mean. For the sample variogram estimator in equation (5) we have,

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

which leads to

$$2 \log\{\hat{\gamma}(\mathbf{h}_i)\} \propto \frac{2}{|N(\mathbf{h}_i)|}.$$

Thus, we use $\frac{|N(\mathbf{h}_i)|}{2}$ as the weight function, v_{3i} , in $Q_{3n}(\boldsymbol{\theta})$.

It can be easily checked that the parameter free weight function v_{3i} also makes the estimating equations asymptotically unbiased, thus eliminating a common criticism against the WLS method for variogram fitting (see for example Curriero & Lele 1999). In Section 3, using a general result obtained by Lahiri et al. (2002), we show that the asymptotic variance of estimates obtained by minimizing Q_{3n} is smaller than that of estimates obtained using $Q_{1n}(\boldsymbol{\theta})$. In this paper we use their result as a tool to derive expressions of variances for the proposed estimators and omit details of their statement and the mixed increasing domain sampling scheme. Interested readers may follow the original paper Lahiri et al. (2002).

3 Comparison of the efficiencies of estimators

Lahiri et al. (2002, Section 3) have established asymptotic normality for estimators of parametric variograms, obtained using least squares, under “mixed increasing domain asymptotics”.

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 et al. (2002). Let $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_3$ denote the true minima corresponding to $Q_{1n}(\boldsymbol{\theta})$ and $Q_{3n}(\boldsymbol{\theta})$, respectively.

We denote by $\nabla_l \mathbf{g}_1(\boldsymbol{\theta})$ the vector of first derivatives of $\mathbf{g}_1(\boldsymbol{\theta})$ with respect to θ_l and by $D_1(\boldsymbol{\theta}_0)$ the corresponding matrix of all first derivatives. The index 1 specifies that these quantities are related to the first criterion, $Q_{1n}(\boldsymbol{\theta})$. We also define

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

Similarly to above, the vector of first derivatives of $\mathbf{g}_3(\boldsymbol{\theta})$, with respect to θ_l is denoted by $\nabla_l \mathbf{g}_3(\boldsymbol{\theta})$, while $D_3(\boldsymbol{\theta}_0)$ denotes the first derivative matrix. We assume that assumptions C.1–C.7 of Lahiri et al. (2002) are true. Now using the approximation due to Cressie (1985) (see Section 1 above) and Theorem 3.2 in Lahiri et al. (2002) we obtain that for the first criterion, $Q_{1n}(\boldsymbol{\theta})$, we have

$$\begin{aligned} \mathbf{g}(\boldsymbol{\theta}) &\simeq \text{MVN}(\mathbf{0}, \Sigma_1(\boldsymbol{\theta}_0)) \text{ where} \\ \Sigma_1(\boldsymbol{\theta}_0) &= \text{diag} (2(2\gamma(\mathbf{h}_i, \boldsymbol{\theta}))^2 / |N(\mathbf{h}_i)|)_{i=1,2,\dots,K}. \end{aligned} \quad (8)$$

Now $\ln\{\gamma(\mathbf{h}, \boldsymbol{\theta})\}$ is a differentiable function of $\gamma(\mathbf{h}, \boldsymbol{\theta})$. Thus application of the delta method (see for example Brockwell & Davis 1991) to equation (8) establishes that

$$\begin{aligned} \mathbf{g}_3(\boldsymbol{\theta}) &\simeq \text{MVN}(\mathbf{0}, \Sigma_3(\boldsymbol{\theta}_0)), \text{ where} \\ \Sigma_3(\boldsymbol{\theta}_0) &= \text{diag} (2/|N(\mathbf{h}_i)|)_{i=1,2,\dots,K} \end{aligned} \quad (9)$$

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 $V_3(\boldsymbol{\theta}) = \Sigma_3^{-1}(\boldsymbol{\theta})$ be the inverses of the corresponding covariance matrices. Then using equation (9), Theorem 3.2 and Corollary 3.2 from Lahiri et al. (2002) we have

$$\begin{aligned} \hat{\boldsymbol{\theta}}_{1n} &\approx \text{MVN}(\boldsymbol{\theta}_0, \Sigma_{v1}(\boldsymbol{\theta}_0)) \\ \Sigma_{v1}(\boldsymbol{\theta}_0) &= B_1(\boldsymbol{\theta}_0) D_1(\boldsymbol{\theta}_0)' V_1(\boldsymbol{\theta}_0) \Sigma_1(\boldsymbol{\theta}_0) V_1(\boldsymbol{\theta}_0) D_1(\boldsymbol{\theta}_0) B_1(\boldsymbol{\theta}_0) \\ B_1(\boldsymbol{\theta}_0) &= \{D_1(\boldsymbol{\theta}_0)' V_1(\boldsymbol{\theta}_0) D_1(\boldsymbol{\theta}_0)\}^{-1} \\ \hat{\boldsymbol{\theta}}_{3n} &\approx \text{MVN}(\boldsymbol{\theta}_0, \Sigma_{v3}(\boldsymbol{\theta}_0)) \\ \Sigma_{v3}(\boldsymbol{\theta}_0) &= B_3(\boldsymbol{\theta}_0) D_3(\boldsymbol{\theta}_0)' V_3(\boldsymbol{\theta}_0) \Sigma_3(\boldsymbol{\theta}_0) V_3(\boldsymbol{\theta}_0) D_3(\boldsymbol{\theta}_0) B_3(\boldsymbol{\theta}_0) \\ B_3(\boldsymbol{\theta}_0) &= \{(D_3(\boldsymbol{\theta}_0)' V_3(\boldsymbol{\theta}_0) D_3(\boldsymbol{\theta}_0))\}^{-1} \end{aligned} \quad (10)$$

Using the expressions in equation (10) following simple algebra one can easily deduce that

$$\text{Var}(\hat{\boldsymbol{\theta}}_1) \rightarrow D_1^{-1}(\boldsymbol{\theta}_0) \Sigma_1(\boldsymbol{\theta}_0) D_1^{-1'}(\boldsymbol{\theta}_0) \quad (11)$$

$$\text{Var}(\hat{\boldsymbol{\theta}}_3) \rightarrow D_3^{-1}(\boldsymbol{\theta}_0) \Sigma_3(\boldsymbol{\theta}_0) D_3^{-1'}(\boldsymbol{\theta}_0) \quad (12)$$

To compare the relative efficiencies of the estimates, let us consider the difference between i^{th} elements of the diagonal terms of $\Sigma_1(\boldsymbol{\theta})$ and $\Sigma_3(\boldsymbol{\theta})$. Note that,

$$\frac{2\{2\gamma(\mathbf{h}_i, \boldsymbol{\theta}_0)\}^2}{|N(\mathbf{h}_i)|} - 2/|N(\mathbf{h}_i)| > 0 \quad \text{if } \text{Var}(Z(\mathbf{s}_l) - Z(\mathbf{s}_m)) > 1^1, \text{ for all } l, m \in N(\mathbf{h}_i). \quad (13)$$

Then

$$\begin{aligned}
\Sigma_{v_1} - \Sigma_{v_3} &= D_1(\boldsymbol{\theta}_0)^{-1} \Sigma_1(\boldsymbol{\theta}_0) D_1'(\boldsymbol{\theta}_0)^{-1} - D_3(\boldsymbol{\theta}_0)^{-1} \Sigma_3(\boldsymbol{\theta}_0) D_3'(\boldsymbol{\theta}_0)^{-1} \\
&> D_3(\boldsymbol{\theta}_0)^{-1} \Sigma_1(\boldsymbol{\theta}_0) D_3'(\boldsymbol{\theta}_0)^{-1} - D_3(\boldsymbol{\theta}_0)^{-1} \Sigma_3(\boldsymbol{\theta}_0) D_3'(\boldsymbol{\theta}_0)^{-1} \\
&= D_3(\boldsymbol{\theta}_0)^{-1} [\Sigma_1(\boldsymbol{\theta}_0) - \Sigma_3(\boldsymbol{\theta}_0)] D_3'(\boldsymbol{\theta}_0)^{-1} \\
&\geq 0,
\end{aligned} \tag{14}$$

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 equation (13) $[\Sigma_1(\boldsymbol{\theta}_0) - \Sigma_3(\boldsymbol{\theta}_0)] > \mathbf{0}$. 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.

4 Simulation

We now present empirical comparison of parameter estimators of variogram parameters using the weight functions $v_{1i}(\boldsymbol{\theta})$, v_{2i} and v_{3i} , 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 and v_3 respectively. The results presented in this section are based on simulated Gaussian spatial processes with the well known Matérn class variogram functions (see Stein 1999). The Matérn class variogram function is defined as

$$\gamma(\mathbf{h}; \nu, \phi, \sigma) = \sigma^2 - \sigma^2 \frac{1}{\Gamma(\nu) 2^{(\nu-1)}} \left(\frac{\|\mathbf{h}\|}{\phi} \right)^\nu K_\nu \left(\frac{\|\mathbf{h}\|}{\phi} \right), \quad \nu > 0, \phi > 0, \tag{15}$$

where $K_\nu(\cdot)$ is the modified Bessel function of the second kind. Here, σ^2 is the variance of the process. The parameter ν , also called the order of the Matérn class, is a shape parameter which characterizes the smoothness of the random process $Z(\mathbf{s})$ (see Stein 1999, Ch 2 and 3). In this parametrization, the scale parameter ϕ is inversely proportional to the practical range of the variogram function and $1/\phi$ is the rate at which the variogram "saturates" to σ^2 (the sill) as $\|\mathbf{h}\| \rightarrow \infty$. Further details on the empirical and theoretical properties of the Matérn class functions can be found in Matérn (1986), Stein (1999) and Diggle et al. (1998). Gaussian spatial processes are generated with Exponential ($\nu = 0.5$), Matérn 1 ($\nu = 1$), Matérn 1.5 ($\nu = 1.5$) and Gaussian ($\nu \rightarrow \infty$). The processes are sampled at 100 fixed locations on a two dimensional plane using Cholesky decomposition method (see e.g. Hadley 1961, Ch 7 and Cressie 1993, Ch 2). 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. We choose the process variance σ^2 to be 20 ($\sigma = 4.47$) for each spatial process. While the range parameters ϕ are chosen based on practical range (see for example Schabenberger & Gotway 2005) of 100 units. The

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.

chosen values are 16.69041 for exponential variogram, 10.49 for Matérn 1, 8.37 for Matérn 1.5 and 28.89 for the Gaussian variogram.

The computations were done with the statistical system R (R Development Core Team 2016). We used package `geoR` (Ribeiro Jr & Diggle 2009) in the estimation of spatial variograms and the routine `nls` from package `nlme` (Pinheiro et al. 2012) for optimization.

4.1 Results and discussion

Table 1 gives the estimates of σ from 200 independent simulations, along with their MSEs, estimated using WLS method and the three weight functions. The corresponding results for ϕ are given in Table 2. We observe that for all chosen Matérn class variograms, the MSEs of parameter estimates

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

Variogram	$v_1(\boldsymbol{\theta})$			v_2			v_3		
	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
Matérn 1	4.46	0.74	-0.02	4.32	0.54	-0.16	4.36	0.49	-0.12
Matérn 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

obtained using both v_2 and v_3 are smaller than the MSE of the estimates obtained using $v_1(\boldsymbol{\theta})$. The estimates obtained using v_3 have the smallest MSE for all variograms considered here. Table 2 shows

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

Variogram	$v_1(\boldsymbol{\theta})$			v_2			v_3		
	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
Matérn 1	11.94	50.00	1.45	10.95	20.43	0.46	10.64	13.19	0.15
Matérn 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

that, for exponential variogram, estimators of ϕ have much higher MSEs for all weight functions considered. But the MSE obtained using v_3 is always smaller than $v_1(\boldsymbol{\theta})$. Further, estimators obtained using v_3 have the smallest MSEs, followed by estimates obtained using v_2 .

In Appendix B we present two more simulation studies comparing parameter estimates from Gaussian, gamma and Laplace spatial processes on uniform and non-uniform locations. We also provide some guidance on choosing starting values for optimization. Further details are given by Das (2011).

5 Analysis of Swiss Rainfall data

We now apply WLS estimation for analysis of Swiss Rainfall data (see Gregoire 1997) earlier considered by Diggle et al. (1998, Ch. 5). The data corresponds to the record of rainfall measured on 8th May 1986 at 467 fixed locations across Switzerland. Locations (Cartesian coordinates) are measured in meters and rainfall in 1/10th of a millimetre. To estimate the variogram structure we use the data recorded at all 467 locations. Figure 1 gives the plot of the rainfall. The plot does not

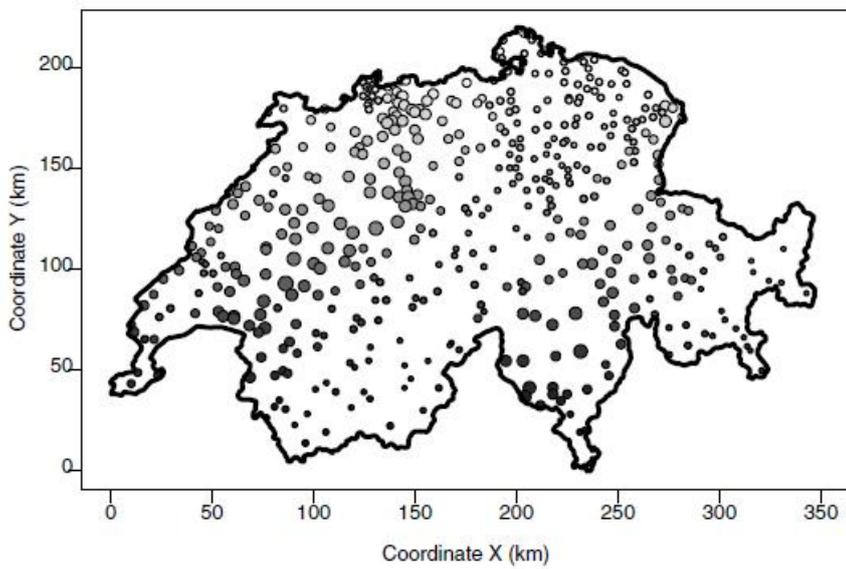


Figure 1: Swiss Rainfall Plot. The figure shows the rainfall received at a particular geographical location in Switzerland. The radii of the circles are proportional to the amount of rainfall. *Figure reproduced from Diggle et al. (1998)*

indicate the presence of any particular spatially varying trend surface. Besides, Figure 1 does not indicate directional clustering of observations. Diggle et al. (1998, Ch. 5) have previously made the same observation about this data. Thus we assume that the rainfall data is stationary and isotropic. Diggle et al. (1998) recommended a transformation of the rainfall data to the logarithmic scale and selected a Matérn class variogram function based on Gaussian Maximum Likelihood estimation. But Das (2011) has discussed the limitations of this selection both in estimation and prediction. The mean square prediction errors are particularly high if one uses Gaussian assumption on the logarithm transformed data.

The sample variogram estimate from equation (2) and the weighted least squares variogram estimation method also illustrate that a Matérn class variogram is not suitable for the rainfall data. We suggest that the periodically saturating wave variogram is better suited for the rainfall data. The Wave variogram is defined as:

$$\gamma(\mathbf{h}) = \sigma^2 - \sigma^2 \left(\frac{\phi}{\|\mathbf{h}\|} \right) \sin \left(\frac{\|\mathbf{h}\|}{\phi} \right), \quad \sigma > 0, \phi > 0. \quad (16)$$

The parameters σ and ϕ have similar interpretation to the corresponding variance and range parameters of the Matérn class functions. The estimates of σ and ϕ and their standard errors for the Swiss rainfall data are given in Table 3. We observe that the estimated standard errors are minimum if one uses the criterion $Q_3(\theta)$ with the proposed logarithmic weight function v_{3i} .

Table 3: Estimates and Standard Errors of Variogram Parameters

Variogram function	Statistic	$v_1(\theta)$		v_2		v_3	
		$\hat{\sigma}$	$\hat{\phi}$	$\hat{\sigma}$	$\hat{\phi}$	$\hat{\sigma}$	$\hat{\phi}$
Exponential	estimate	118.729	34.072	117.767	33.74	118.754	35.298
	stderr	(2.094)	(4.189)	(1.996)	(4.157)	(2.163)	(4.183)
Matérn 1	estimate	117.336	19.482	116.638	19.49	117.072	19.525
	stderr	(1.48)	(1.347)	(1.441)	(1.574)	(1.473)	(1.368)
Matérn 1.5	estimate	116.56	14.154	116.309	14.928	116.42	14.3682
	stderr	(1.4692)	(0.8946)	(1.355)	(1.08)	(1.4361)	(0.9235)
Matérn 2	estimate	115.711	11.1104	116.158	12.5408	116.09	11.758
	stderr	(1.7342)	(0.6432)	(1.3344)	(0.8679)	(1.469)	(0.749)
Wave	estimate	103.7291	17.6437	102.7444	17.6592	103.1531	17.6269
	stderr	(1.0933)	(0.2126)	(1.1206)	(0.2248)	(1.0312)	(0.2472)

We observe that the estimates of σ are much smaller (around 14 percent) for the wave variogram than the Matérn class variogram functions considered. This indicates that for the Swiss rainfall data the Matérn class variogram functions overestimate the variance of the random process due to their non-periodic nature.

Further, for both σ and ϕ the standard errors of the estimates are substantially lower for the wave variogram, compared to the other variogram functions considered. For σ the minimum standard error of 1.03 is obtained using the weight function $v_3(\theta)$ while the minimum standard error of estimate of ϕ is 0.21 for $v_1(\theta)$. But in view of the correlation of the estimates of σ and ϕ , we calculate a measure to study the joint variability. We define it as the *Joint Standard Error* (JSE) (see equation (17) in Appendix A). The JSE's corresponding to the estimates in Table 3 are given in Table 4. We denote the JSE corresponding to the weight function $v_i(\theta)$ by $JSE_i(\theta)$; ($i = 1, 2, 3$). We see that the JSE

Table 4: Joint Standard Error for $\hat{\sigma}$ and $\hat{\phi}$

	$JSE_1(\boldsymbol{\theta})$	$JSE_2(\boldsymbol{\theta})$	$JSE_3(\boldsymbol{\theta})$
Exponential	5.88	5.73	5.95
Matérn 1	2.51	2.66	2.52
Matérn 1.5	2.07	2.11	2.06
Matérn 2	2.14	1.91	1.94
Wave	1.17	1.19	1.03

of the wave variogram is also smaller than the JSE of Matérn class variogram functions. Within the wave variogram, JSE of estimates of σ and ϕ is minimum (1.03) for v_3 compared to 1.17 for $v_1(\boldsymbol{\theta})$.

We now consider the goodness of the fit aspects of the estimates. Figures 2 and 3 show the plots of the fitted variograms corresponding to the respective weight functions along with the sample variogram, for comparison.

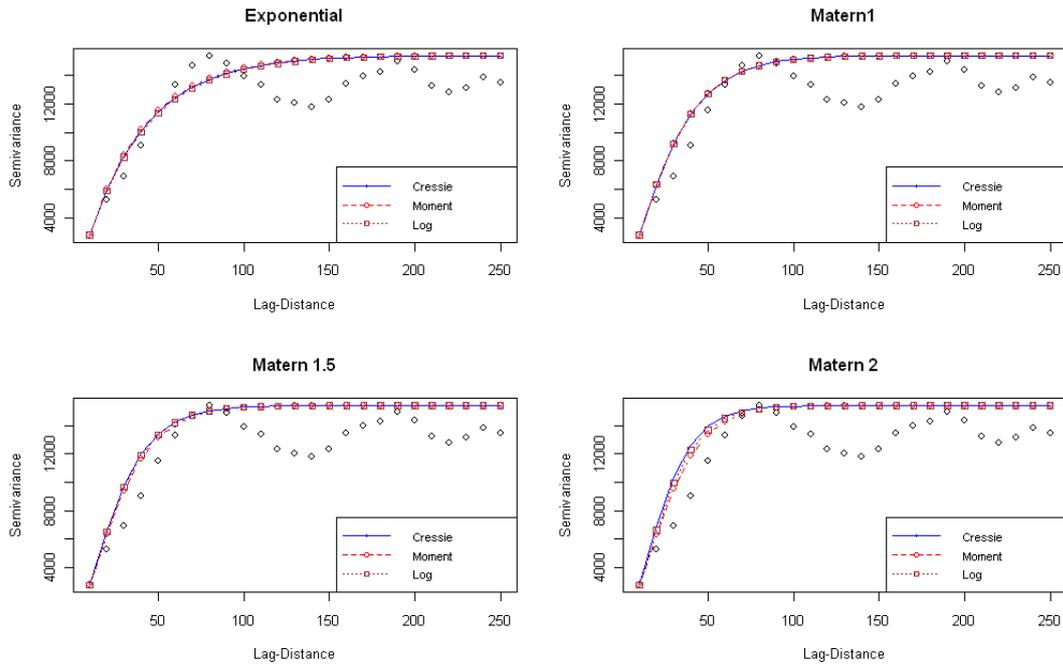


Figure 2: Variogram Fits for Rainfall Data — exponential, Matérn 1, Matérn 1.5 and Matérn 2 variogram functions. The plot for each variogram function shows the fitted variograms corresponding to the three weights, $v_1(\boldsymbol{\theta})$, v_2 and v_3 , superimposed over the sample variogram.

A comparison of Figure 2 with Figure 3 reveals that there has been a substantial improvement in the fit of the variogram curve by using wave variogram as the parametric function. Table 5 summarizes the above observations using a descriptive goodness of fit measure (equation (18) in Appendix A).

Note that the higher order Matérn functions overestimate the theoretical variogram by deviating

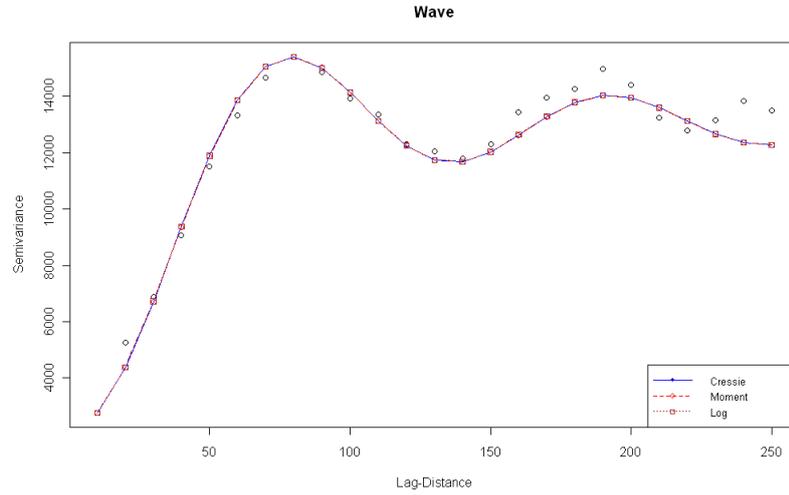


Figure 3: Variogram Fit for Rainfall Data — Wave Variogram. The three curves corresponding to the weight functions $v_1(\boldsymbol{\theta})$, v_2 and v_3 are superimposed over the sample variogram.

Table 5: Goodness-of-Fit

	Exponential	Matérn 1	Matérn 1.5	Matérn 2	Wave
$v_1(\boldsymbol{\theta})$	6382.617	5465.87	5334.34	5555.24	2752.401
v_2	6288.232	5472.62	5294.38	5238.38	2726.082
v_3	6406.145	5457.32	5307.59	5335.07	2698.278

towards the higher sample variogram values thereby spuriously reducing the goodness-of-fit measure. Table 6 compares the predictive capabilities of the models, using a measure based on cross-validation kriging residuals (see equation (19) in Appendix A and Cressie 1993, p. 102). The poor prediction performance of the overfitted variograms can be seen clearly. Indeed, we note that the higher order

Table 6: Kriging Residual Sum of Squares - Original data Wave Variogram

	Exponential	Matérn 1	Matérn 1.5	Matérn 2	Wave
$v_1(\boldsymbol{\theta})$	0.570166	2.61485	8.98134	20.0277	0.17036
$v_2(\boldsymbol{\theta})$	0.575658	2.6305	9.00066	19.9505	0.157217
$v_3(\boldsymbol{\theta})$	0.570873	2.62075	8.99209	19.9623	0.156936

variogram functions for which the fitted variogram curves shift towards the higher sample variogram values, result in higher residual sums of squares. Thus, overestimation of the variogram adversely effects the prediction of the random process in kriging. The above analysis using the weighted least squares variogram estimation technique (also called curve fitting) makes it clear that the Matérn class of variograms cannot properly explain the underlying sample variogram of the Swiss Rainfall data.

The residual prediction sums-of-squares are smaller for the wave variogram, compared to all other variogram functions. Based on the above analyses we believe that for the Swiss Rainfall data and for other spatial processes which exhibit periodic sample variogram, the wave variogram is appropriate.

The minimum value of the residual sum of squares is 0.1569 observed for v_3 as the weight function. The residual sum of squares of prediction, obtained for the wave variogram using the weight function $v_1(\boldsymbol{\theta})$ is higher than those obtained using the proposed weights v_2 and v_3 . The above tables provide clear empirical evidence that estimation using the proposed weight functions v_2 and v_3 results in better prediction results, when we use weighted least squares for estimation.

6 Discussion

In this paper we have proposed two parameter-free weight matrices, v_{i2} and v_{3i} , for use in weighted least squares criteria for parametric spatial variogram estimation (Cressie 1985). We observe that the weight function v_{3i} improves asymptotic efficiency of parameter estimates. v_{3i} also renders the estimating equation asymptotically unbiased, which is often held as a limitation against weighted least squares method of estimation. We corroborate our observations using extensive simulation under different spatial orientation of locations, parametric covariances and random fields. We note that one can similarly supplement the weighted least squares estimating criterion while analysing spatio-temporal data (Cressie & Huang 1999, Eq. 9). In the past couple of decades there have been significant improvements in the maximum likelihood and quasi-likelihood methods and their applications to estimation of parameters of spatial processes (see e.g Varin et al. 2011). But in this paper, through simulations and real data analysis, we show that the distribution free least squares method and the associated techniques remain insightful tools for an applied scientist.

7 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 also like to thank an anonymous referee and the Associate Editor for several helpful suggestions.

References

- Brockwell, P. & Davis, R. (1991), *Time series: theory and methods*, Springer.
- 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 & Sons, New York.
- Cressie, N. & Huang, H. (1999), ‘Classes of Nonseparable, Spatio-Temporal Stationary Covariance Functions.’, *J. Am. Statist. Ass.* **94**(448).
- Curriero, F. C. & Lele, S. (1999), ‘A composite likelihood approach to semivariogram estimation’, *Journal of Agricultural, biological and Environmental statistics* **4**(1), 9–28.
- Das, S. (2011), Statistical Estimation of Variogram and Covariance Parameters of Spatial and Spatio-Temporal Random Processes, PhD thesis, School of Mathematics, University of Manchester.
- Diggle, P. J. & Ribeiro, P. J. (2007), *Model-based geostatistics*, Springer-Verlag New York.
- Diggle, P. J., Tawn, J. & Moyeed, R. (1998), ‘Model-based geostatistics’, *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **47**(3), 299–350.
- Gneiting, T. (2002), ‘Nonseparable, stationary covariance functions for space-time data’, *J. Am. Statist. Ass.* **97**(458), 590–600.
- Gneiting, T., Genton, M. G. & Guttorp, P. (2007), ‘Geostatistical space-time models, stationarity, separability, and full symmetry’, *Monographs On Statistics and Applied Probability* **107**, 151–175.
- Gregoire, D. (1997), *Spatial Interpolation Comparison*.
URL: https://wiki.52north.org/AI_GEOSTATS/WebHome
- Hadley, G. (1961), *Linear Algebra Reading*, Addison-Wesley Publishing Company, Boston.
- Lahiri, S., Lee, Y. & 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), 65–85.
- Matérn, B. (1986), *Spatial variation*, Springer-Verlag Berlin.
- Matheron, G. (1963), ‘Principles of geostatistics’, *Economic geology* **58**(8), 1246.
- Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D. & R Core Team (2012), *nlme: Linear and Nonlinear Mixed Effects Models*, R Foundation for Statistical Computing, Vienna, Austria. R package version 3.1-104.
URL: <https://CRAN.R-project.org/package=nlme>
- R Development Core Team (2016), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
URL: <https://www.R-project.org/>

Ribeiro Jr, P. J. & Diggle, P. J. (2009), *geoR: Analysis of Geostatistical Data*.

URL: <https://CRAN.R-project.org/package=geoR>

Schabenberger, O. & Gotway, C. (2005), *Statistical methods for spatial data analysis*, CRC Press, Florida.

Soetaert, K. (2016), *Nonlinear Root Finding, Equilibrium and Steady-State Analysis of Ordinary Differential Equations*. R package version ≥ 2.01 .

Stein, M. (1999), *Interpolation of Spatial Data: some theory for kriging*, Springer Verlag, New York.

Subba Rao, T. & Terdik, G. (2017), ‘A space-time covariance function for spatio-temporal random processes and spatio-temporal prediction’, *J. Time Ser. Anal.* **38**(6), 936–959.

Varin, C., Reid, N. & Firth, D. (2011), ‘An overview of composite likelihood methods’, *Statistica Sinica* **21**(1), 5–42.

A Errors of fitted models

Joint Standard Error

$$\text{Joint Standard Error (JSE)} = \sqrt{\{\text{se}(\hat{\sigma})\}^2 + \{\text{se}(\hat{\phi})\}^2 + 2 \text{Cov}\{\hat{\sigma}, \hat{\phi}\}} \quad (17)$$

Goodness-of-fit measure The goodness of fit measure is expressed as the average of the sum of squared deviations of the fitted and empirical variograms for k chosen lags.

$$\left[(1/k) \sum_{j=1}^k \{(\gamma(\mathbf{h}_j) - \hat{\gamma}(\mathbf{h}_j))\}^2 \right]^{1/2}. \quad (18)$$

Cross Validation Kriging Once we have decided on a suitable theoretical variogram — using the methods described above — we use it to predict the random process at various locations, omitting the observation at j^{th} location and using the rest of the available locations to predict the value at this location, using kriging. We denote such a predictor by $\hat{Z}((s)_{-j})$ and the associated standard error of prediction by $\sigma_{-j}((s)_j)$. The sum of the squares of residuals, a descriptive measure of the prediction error is given by

$$\left[(1/n) \sum_{i=1}^n \left\{ \left(Z(\mathbf{s}_j) - \hat{Z}(\mathbf{s}_{-j}) \right) / \sigma_{-j}((s)_j) \right\}^2 \right]^{1/2}. \quad (19)$$

B Simulations

We present a more comprehensive simulation study to illustrate the benefits of the proposed weight function. Each simulation is based on 100 sampled locations. We simulate random fields from Gaussian, gamma and double exponential random fields with exponential, Matérn 1, Matérn 1.5 and Gaussian covariance functions. The sill or process variance is fixed at $\sigma^2 = 1$, in each case. The range parameters ϕ (see equation (15)) were then chosen based on a pre-specified practical range (see for example Schabenberger & Gotway 2005, Ch 4). We opt for two different sampling schemes: uniform locations and non-uniform locations.

B.1 Uniformly oriented locations

In the first case we uniformly sample 100 locations on a 2-dimensional square planar area of length 100 units. The maximum pairwise distance was 129.97 units. Figure 4 shows the orientation of selected points. The original range parameters are $\phi_1 = 16.69$, $\phi_2 = 10.49$, $\phi_3 = 8.37$ and $\phi_4 = 28.89$, for

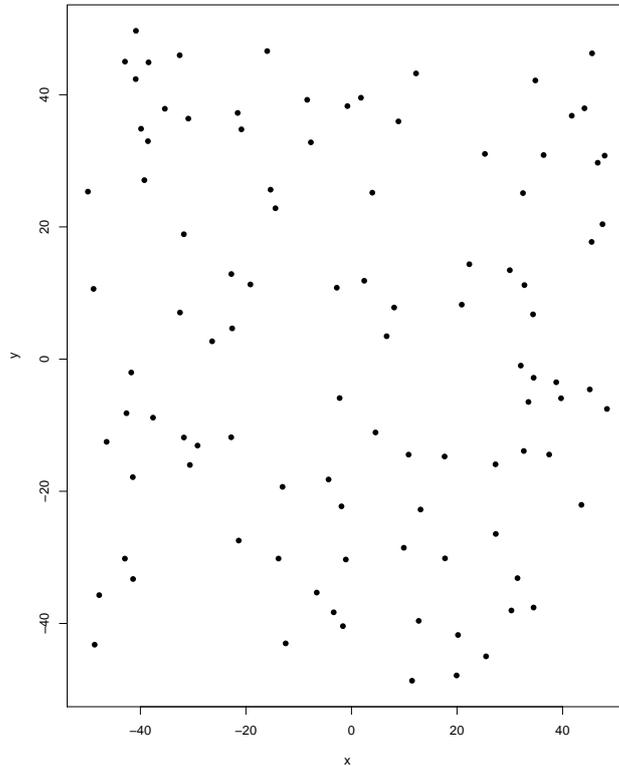


Figure 4: The plot shows the spatial orientation of 100 continuously and uniformly selected locations on a 2 dimensional square of length 100 units.

exponential, Matérn 1, Matérn 1.5 and Gaussian covariance functions, respectively. For parameter estimation the overall distances were partitioned into 20 distinct bins. Tables 7–9 give descriptive statistics for estimated values of σ . The corresponding estimates for ϕ are given in Tables 10–12. We can see that for each simulated model, the bias and the mean squared errors (MSE) for the estimates of both parameters, σ and ϕ , are lower for the proposed weights v_2 and v_3 . In fact, the log-transformed weight, v_{3i} (see equation (8)), leads to minimum bias and MSE, in each case. For non-Gaussian processes, the proposed weight functions have comparatively much smaller bias and MSE.

Table 7: Sampling properties of $\hat{\sigma}$ - Gaussian process

Gaussian	$v_{1\theta}$			v_2			v_3		
	Mean	Bias	MSE	Mean	Bias	MSE	Mean	Bias	MSE
Exponential	1.031	0.031	0.043	0.996	-0.004	0.033	1.013	0.013	0.035
Matérn 1	1.013	0.013	0.031	0.980	-0.020	0.026	0.998	0.013	0.031
Matérn-1.5	1.011	0.011	0.033	0.976	-0.024	0.027	0.995	-0.005	0.027
Gaussian	1.021	0.021	0.119	0.971	-0.029	0.077	0.987	-0.013	0.075

Table 8: Sampling properties of $\hat{\sigma}$ - gamma process

gamma	$v_{1\theta}$			v_2			v_3		
	Mean	Bias	MSE	Mean	Bias	MSE	Mean	Bias	MSE
Exponential	1.052	0.052	0.121	1.008	0.008	0.073	1.017	0.017	0.062
Matérn 1	1.011	0.011	0.054	0.977	-0.023	0.044	0.991	-0.009	0.043
Matérn 1.5	1.005	0.005	0.055	0.971	-0.029	0.045	0.985	-0.015	0.044
Gaussian	0.986	-0.014	0.113	0.954	-0.046	0.100	0.968	-0.032	0.099

Table 9: Sampling properties of $\hat{\sigma}$ - double exponential process

Laplace	$v_{1\theta}$			v_2			v_3		
	Mean	Bias	MSE	Mean	Bias	MSE	Mean	Bias	MSE
Exponential	1.053	0.053	0.088	1.010	0.010	0.058	1.026	0.026	0.057
Matérn 1	1.017	0.017	0.045	0.982	-0.018	0.039	1.000	0.000	0.041
Matérn 1.5	1.010	0.010	0.044	0.974	-0.026	0.039	0.994	-0.006	0.041
Gaussian	0.991	-0.009	0.093	0.947	-0.053	0.088	0.970	-0.030	0.088

Table 10: Sampling properties of $\hat{\phi}$ - Gaussian process

Gaussian	$v_{1\theta}$			v_2			v_3		
	Mean	Bias	MSE	Mean	Bias	MSE	Mean	Bias	MSE
Exponential	20.960	4.270	308.800	18.656	1.965	124.555	18.991	2.300	124.932
Matérn 1	11.446	0.956	29.218	10.684	0.194	12.341	10.822	0.332	11.722
Matérn 1.5	8.944	0.574	16.458	8.375	0.005	5.527	8.479	0.109	5.108
Gaussian	30.496	1.607	275.951	28.448	-0.440	89.946	28.473	-0.415	64.833

Table 11: Sampling properties of $\hat{\phi}$ - gamma process

gamma	$v_{1\theta}$			v_2			v_3		
	Mean	Bias	MSE	Mean	Bias	MSE	Mean	Bias	MSE
Exponential	25.504	8.814	1332.314	21.655	4.965	459.826	20.802	4.112	304.001
Matérn 1	12.115	1.625	39.769	11.374	0.884	23.770	11.202	0.712	16.964
Matérn 1.5	9.385	1.015	17.346	8.871	0.501	10.405	8.746	0.376	7.215
Gaussian	29.794	0.906	98.200	28.736	-0.152	71.479	28.880	-0.008	58.757

Table 12: Sampling properties of $\hat{\phi}$ - double exponential process

Laplace	$v_{1\theta}$			v_2			v_3		
	Mean	Bias	MSE	Mean	Bias	MSE	Mean	Bias	MSE
Exponential	22.950	6.260	982.851	19.987	3.297	283.176	19.565	2.875	234.570
Matérn 1	11.456	0.966	26.530	10.912	0.422	18.432	10.750	0.260	13.698
Matérn 1.5	8.885	0.515	10.663	8.504	0.134	7.672	8.415	0.045	5.616
Gaussian	29.094	0.206	83.775	27.600	-1.288	66.738	27.860	-1.028	49.984

B.2 Non-uniform locations

The second set of simulations are based on 467 non-uniform locations of the Swiss rainfall data (see Section 4). The orientation of the sampled locations is given in Figure 5. A common issue in simulation is that if there are multiple locations in close proximity, the spatial covariance matrix, $\Sigma(\theta)$, becomes singular. To address this we sampled 100 locations, uniformly, from the set of all locations (467) that are at least 5 KM apart. Of the 467 coordinates 253 locations satisfy this criterion. The maximum pairwise distance for the selected locations is 324 KM. As earlier, we fixed the practical range to choose the range-parameters (ϕ). The original values for range-parameters are $\phi_1 = 50.07$, $\phi_2 = 37.51$, $\phi_3 = 31.62$ and $\phi_4 = 86.67$, respectively for the exponential, Matérn 1, Matérn 1.5 and Gaussian covariance functions. Tables 13–15 give estimates for σ while the estimates for ϕ are given in Tables 16–18. We note that for all simulated models the MSE and bias are minimum when we use the proposed weight, v_{3i} . v_{2i} also has lower bias and MSE for most models. These observations are consistent across all three distributions: Gaussian, gamma and double exponential.

Table 13: Sampling properties of $\hat{\sigma}$ — Gaussian process, non-uniform locations

Gaussian	$v_{1\theta}$			v_2			v_3		
	Mean	Bias	MSE	Mean	Bias	MSE	Mean	Bias	MSE
Exponential	1.075	0.075	0.141	1.062	0.062	0.132	1.051	0.051	0.100
Matérn 1	1.020	0.020	0.080	1.008	0.008	0.074	1.006	0.020	0.080
Matérn 1.5	1.007	0.007	0.081	0.996	-0.004	0.076	0.996	-0.004	0.072
Gaussian	1.013	0.013	0.255	0.988	-0.012	0.165	0.981	-0.019	0.156

Table 14: Sampling properties of $\hat{\sigma}$ — gamma process, non-uniform locations

gamma	$v_{1\theta}$			v_2			v_3		
	Mean	Bias	MSE	Mean	Bias	MSE	Mean	Bias	MSE
Exponential	1.030	0.030	0.133	1.023	0.023	0.191	0.998	-0.002	0.066
Matérn 1	0.986	-0.014	0.087	0.971	-0.029	0.069	0.964	-0.036	0.056
Matérn 1.5	0.974	-0.026	0.098	0.961	-0.039	0.076	0.953	-0.047	0.061
Gaussian	0.959	-0.041	0.193	0.935	-0.065	0.149	0.931	-0.069	0.123

A well known challenge for the above optimization problem is the choice of starting values. We have found this to be the case irrespective of the estimating equation, sampling scheme, distribution or the covariance function. In this paper we have not addressed the issue formally but we suggest the following. Note that in real life we usually observe only one sample from an assumed generating

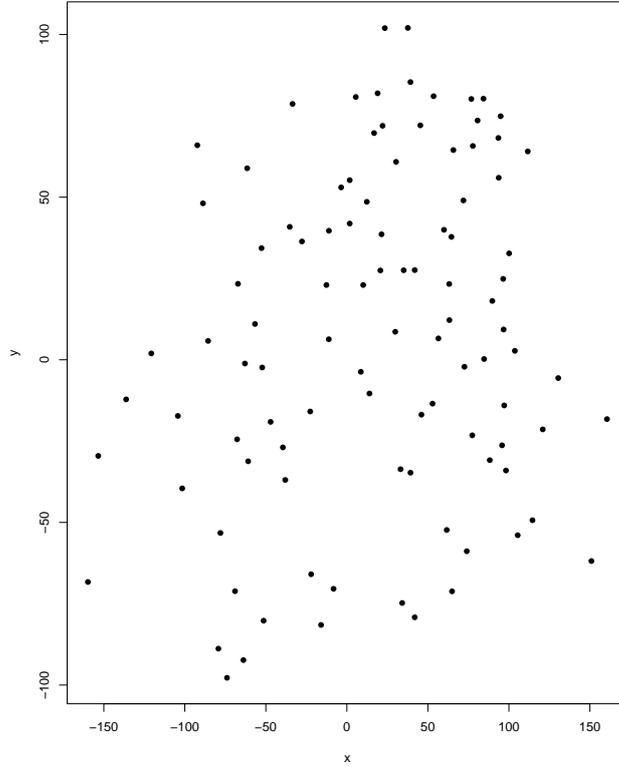


Figure 5: The plot shows the spatial orientation of 100 continuously sampled locations from the Swiss rainfall data.

Table 15: Sampling properties of $\hat{\sigma}$ — double exponential process, non-uniform locations

Laplace	$v_{1\theta}$			v_2			v_3		
	Mean	Bias	MSE	Mean	Bias	MSE	Mean	Bias	MSE
Exponential	1.079	0.079	0.330	1.047	0.047	0.145	1.036	0.036	0.112
Matérn 1	1.013	0.013	0.094	0.999	-0.001	0.082	0.999	-0.001	0.082
Matérn 1.5	1.003	0.003	0.100	0.992	-0.008	0.092	0.986	-0.014	0.081
Gaussian	0.987	-0.013	0.207	0.969	-0.031	0.180	0.965	-0.035	0.173

Table 16: Sampling properties of $\hat{\phi}$ — Gaussian process, non-uniform locations

Gaussian	$v_{1\theta}$			v_2			v_3		
	Mean	Bias	MSE	Mean	Bias	MSE	Mean	Bias	MSE
Exponential	74.644	24.573	11331.160	75.053	24.982	10818.584	65.372	15.301	4254.896
Matérn 1	39.613	2.099	473.905	40.623	3.109	423.993	38.641	1.127	279.242
Matérn 1.5	32.130	0.510	203.942	33.049	1.429	188.019	31.747	0.127	132.886
Gaussian	88.195	1.530	3400.264	86.622	-0.042	1264.717	84.625	-2.039	862.913

Table 17: Sampling properties of $\hat{\phi}$ — Gamma process, non-uniform locations

gamma	$v_{1\theta}$			v_2			v_3		
	Mean	Bias	MSE	Mean	Bias	MSE	Mean	Bias	MSE
Exponential	68.699	18.628	14527.333	75.284	25.213	47124.438	56.170	6.099	2122.752
Matérn-1	38.322	0.808	583.819	38.959	1.445	429.025	36.320	-1.194	210.359
Matérn-1.5	31.447	-0.173	293.567	32.057	0.438	210.188	30.277	-1.343	106.239
Gaussian	88.867	2.203	2976.916	87.576	0.912	1617.319	85.713	-0.951	1003.752

Table 18: Sampling properties of $\hat{\phi}$ — double exponential process, non-uniform locations

Laplace	$v_{1\theta}$			v_2			v_3		
	Mean	Bias	MSE	Mean	Bias	MSE	Mean	Bias	MSE
Exponential	95.459	45.388	178869.839	73.969	23.897	18897.861	64.638	14.566	8133.804
Matérn 1	39.563	2.049	513.110	40.408	2.894	411.747	40.408	2.894	411.747
Matérn 1.5	32.220	0.600	233.414	33.122	1.502	218.007	31.354	-0.266	125.350
Gaussian	88.866	2.202	1661.175	88.242	1.578	1160.106	86.589	-0.075	902.222

model. Given this data, estimate the sample variogram using equation (2). Choose the maximum of the sample variogram to be an estimate of the marginal variance (sill), $\hat{\sigma}_{st}$. Find the minimum distance, u^* , at which the estimated sample variogram attains 95% of $\hat{\sigma}_{st}$. Use $(u^*, \hat{\sigma}_{st})$ as inputs in a non-linear equation solver to estimate a starting value for the range parameter, α_{st} based on a given parametric covariance function. We have used the R package `rootSolve` (Soetaert 2016). However, more research is required to test the consistency of this approach.