International Journal of Engineering, Mathematical and Physical Sciences
ISSN: 2517-9934
Vol:10, No:2, 2016

Robust Variogram Fitting Using Non-Linear Rank-Based Estimators

Hazem M. Al-Mofleh, John E. Daniels, Joseph W. McKean

Abstract**—**In this paper numerous robust fitting procedures are considered in estimating spatial variograms. In spatial statistics, the conventional variogram fitting procedure (non-linear weighted least squares) suffers from the same outlier problem that has plagued this method from its inception. Even a 3-parameter model, like the variogram, can be adversely affected by a single outlier. This paper uses the Hogg-Type adaptive procedures to select an optimal score function for a rank-based estimator for these non-linear models. Numeric examples and simulation studies will demonstrate the robustness, utility, efficiency, and validity of these estimates.

Keywords**—**Asymptotic relative efficiency, non-linear rank-based, robust, rank estimates, variogram.

I. Introduction

IN geostatistics, a function of the distances and directions separating the sample locations, used to model statistical dependence, is estimated by the variogram. The variogram is then used to build a spatial model that eventually determines the kriging weights and the standard errors of an estimate at a known spatial location with an unknown response.

The theoretical variogram $\gamma(h)$ is often a non-linear function describing the spatial correlation structure of a random field or stochastic process $Z(h)$. There are two stages in creating a variogram. First, the individual variogram points must be determined using the raw data (the estimation step). This is a plot of $\hat{\gamma}(h)$ versus the separation distance h of all pairwise locations, which is often Euclidean. Second, using these variogram points, the parameters associated with the variogram function must be estimated (the fitting step). Geostatistical data often contains outliers. Reference [9] discussed some geostatistical data with 10-15% outliers while [13] showed that this proportion can be as high as 30%.

There are several competing procedures for the estimation step. The original, using the method-of-moments was proposed by [20]:

$$\hat{\gamma}(h) \equiv \frac{1}{|N(h)|}\sum_{N(h)}\left(Z(s_i)-Z(s_j)\right)^2,\ \ s_i, s_j \in \mathbb{R}^d \qquad (1)$$

H. M. Al-Mofleh is with the Department of Mathematics, Tafila Technical University, Tafila, Jordan (phone: 00962788510934; e-mail: almof1hm@cmich.edu).

J.E. Daniels is with the Department of Mathematics at Central Michigan University, Mount Pleasant, MI 48859 USA (e-mail: danie1je@cmich.edu).

J.W. McKean is with the Department of Statistics at Western Michigan University, Kalamazoo, MI 49004 USA (e-mail: joseph.mckean@wmich.edu).

where $N(h) \equiv \left\{(s_i, s_j) : s_i - s_j = h; i, j = 1,\ldots,n\right\}$ and $|N(h)|$ is the number of distinct pairs in $N(h)$.

A robust estimator, proposed by [6] is:

$$\bar{\gamma}(h) \equiv \left\{\frac{1}{|N(h)|}\sum_{N(h)}\left|Z(s_i)-Z(s_j)\right|^{\frac{1}{2}}\right\}^4 \Big/ \left(0.457+\frac{0.494}{|N(h)|}\right) \qquad (2)$$

For the fitting step, we can specify a parametric variogram model, for example, an exponential (3) (there are many more), where $\boldsymbol{\theta} = (\tau, \sigma^2, \phi)'$ denotes the vector of parameters (nugget, sill, and range) respectively.

$$\gamma(h;\boldsymbol{\theta}) = \begin{cases} 0 & \text{if } h = 0 \\ \tau + \sigma^2\left(1-\exp(-h/\phi)\right) & \text{if } h > 0 \end{cases} \qquad (3)$$

Early procedures that relied on an underlying Gaussian assumption like maximum likelihood [19] were found to be biased for small sample sizes. References [24] and [25] developed a procedure that depends on a restricted maximum likelihood. Minimum norm quadratic (MINQ) estimation used by [27] requires linearity in the parameters of the variance-covariance matrix.

First, a discussion of the conventional and proposed fitting steps will be provided in Sections II and III respectively. Section IV will discuss a quasi-block-jackknife method for generating parameter confidence intervals.

In Section V, some numerical examples are given to demonstrate the robustness of the non-linear rank-based procedure. In Section VI, some simulation studies will compare the NLWLS and the robust estimators in terms of efficiency and validity over a family of contaminated normal error distributions. The programs were used to compute the results in Sections V and VI were written in *R* 3.2.2 programming language [26]. These *R* codes are available to the reader from the authors.

II. Non-Linear Weighted Least Squares (NLWLS)

The variogram estimates $\left(\hat{\gamma}(h)\right)$ in both (1) and (2) are correlated with non-constant variances [4]. These are violations of independence and heteroscedasticity in the general assumptions underlying ordinary non-linear least square (NLOLS), so the NLOLS cannot be applied in this

case. Therefore, [4] suggested a non-linear weighted least squares (NLWLS) to fit a variogram.

A. NLWLS to Fit a Variogram

The weights used in NLWLS are suggested by [4] as $\left|N\left(h\right)\right|/\left(\gamma\left(h;\boldsymbol{\theta}\right)\right)^2$, and the NLWLS is approximated by minimizing the norm as:

$$\sum_{i=1}^{n}\left(\left|N\left(h_i\right)\right|/\left(\gamma\left(h_i;\boldsymbol{\theta}\right)\right)^2\right)\left(\hat{\gamma}\left(h_i\right)-\gamma\left(h_i;\boldsymbol{\theta}\right)\right)^2 \qquad (4)$$

where n is the number of lags, $\gamma\left(h_i;\boldsymbol{\theta}\right)$ is the theoretical variogram model whose form is known up to $\boldsymbol{\theta}$, and $\hat{\gamma}(\cdot)$ is the empirical variogram estimated at n lags.

B. Standard Errors of NLWLS Estimators

The standard error of the parameters resulting from NLWLS is similar to its counterpart in linear WLS, except the design matrix X of the linear model is replaced by the $n\times p$ matrix of partial derivatives at $\hat{\boldsymbol{\theta}}$ (the Jacobian matrix at $\hat{\boldsymbol{\theta}}$) $D_{n\times p}\left(\hat{\boldsymbol{\theta}}\right)$, where $D_{i,j}\left(\hat{\boldsymbol{\theta}}\right)=\left[\partial\gamma\left(h_i;\hat{\boldsymbol{\theta}}\right)/\partial\theta_j\right]$, $i=1,2,..n;\ j=1,2,..,p$, where p is a number of parameters and n is a number of observations. So we can approximate the standard error for a NLWLS by

$$s\varepsilon.\left(\hat{\boldsymbol{\theta}}\right)=\left[\operatorname{diag}\left(\hat{\sigma}^2\left(D'\left(\hat{\boldsymbol{\theta}}\right)W\left(\hat{\boldsymbol{\theta}}\right)D\left(\hat{\boldsymbol{\theta}}\right)\right)^{-1}\right)\right]^{1/2} \qquad (5)$$

where; $\hat{\sigma}^2=\sum_{i=1}^{n}w_i\left(\hat{\boldsymbol{\theta}}\right)\left(\hat{\gamma}^*\left(h_i\right)-\gamma\left(h_i;\hat{\boldsymbol{\theta}}\right)\right)^2/\left(n-p\right)$;

$w_i\left(\hat{\boldsymbol{\theta}}\right)=\left|N\left(h_i\right)\right|/\left[\gamma\left(h_i;\hat{\boldsymbol{\theta}}\right)\right]^2$; and W is a $p\times p$ diagonal matrix with entries w_i's in the main diagonal.

III. Rank-Based Estimators for Non-Linear Models

For a vector u in $\mathbb{R}^n$ the pseudo-norm (rank-based norm) $\|.\|_\varphi$ is defined as $\|u\|_\varphi=\sum_{i=1}^{n}a\left(R\left(u_i\right)\right)u_i$, where $R\left(u_i\right)$ denotes rank of u_i among $u_1,u_2,\ldots,u_n$, $a\left(t\right)=\varphi\left(t/\left(n+1\right)\right)$, and φ is a non-decreasing, square-integrable score function defined on the interval $(0,1)$ [10]. Consider the general nonlinear model:

$$Y_i=f_i\left(\boldsymbol{\theta}\right)+\epsilon_i;\qquad i=1,2,\ldots,n \qquad (6)$$

where f_i are known real valued functions defined on a compact space Θ and $\epsilon_1,\epsilon_2,\ldots,\epsilon_n$ are independent and identically distributed random errors with pdf $h\left(t\right)$ and cdf $H\left(t\right)$, were $H\left(t\right)$ is unknown. Let $\boldsymbol{y}=\left(Y_1,\ Y_2,\ \ldots,\ Y_n\right)'$ and $\mathbf{f}\left(\boldsymbol{\theta}\right)=\left(f_1\left(\boldsymbol{\theta}\right),f_2\left(\boldsymbol{\theta}\right),\ldots,f_n\left(\boldsymbol{\theta}\right)\right)'$. Given a norm $\|.\|_2^2$ on n-space, a natural estimator of $\boldsymbol{\theta}$ induced by the norm is a value $\hat{\boldsymbol{\theta}}$ which minimizes the distance between the response vector y and $\mathbf{f}\left(\boldsymbol{\theta}\right)$; i.e.,

$$\hat{\boldsymbol{\theta}}=\operatorname{Argmin}_{\theta\in\Theta}\left\|\boldsymbol{y}-\mathbf{f}\left(\boldsymbol{\theta}\right)\right\|_2^2 \qquad (7)$$

where $\|u\|_2^2=\sum_{i=1}^{n}u_i^2$ is the Euclidean norm.

A. Rank-Based Procedures to Fit a Variogram

For the rank-based procedures, one can simply replace the norm $\|.\|_2^2$ in (7) y the pseudo-norm $\|u\|_\varphi=\sum_{i=1}^{n}a\left(R\left(u_i\right)\right)u_i$. One of the properties of a rank-based fits is that it's robust to outliers in response space (*y*-space) [1], [10].

If the nonlinear model (6) has an intercept, i.e. can be rewritten as:

$$Y_i=1_n\alpha+g_i\left(\boldsymbol{\beta}\right)+\epsilon_i.\qquad i=1,2,\ldots,n \qquad (8)$$

where $\boldsymbol{\beta}=\left(\sigma^2,\phi\right)'$, and 1_n is a vector of n ones. Then the estimator of the parameters, $\boldsymbol{\beta}$, is:

$$\hat{\boldsymbol{\beta}}_\varphi=\operatorname{Argmin}_{\beta\in\Theta}\left\|\boldsymbol{y}-\mathrm{g}\left(\boldsymbol{\beta}\right)\right\|_\varphi \qquad (9)$$

and the intercept α can be estimated by a location estimate based on the residuals $\hat{\boldsymbol{e}}=\boldsymbol{y}-g\left(\hat{\boldsymbol{\beta}}_\varphi\right)$. Reference [10] suggests to use the median of the residuals to estimate α, which denotes by $\hat{\alpha}_s=\operatorname{med}\left\{\hat{\boldsymbol{e}}\right\}$.

The model in (3) can be rewritten as

$$\gamma\left(h;\boldsymbol{\theta}\right)=\begin{cases}0 & if\ \ h=0\\ \tau+g\left(h;\boldsymbol{\beta}\right) & if\ \ h>0\end{cases} \qquad (10)$$

where $\boldsymbol{\theta}=\left(\tau,\boldsymbol{\beta}'\right)'$. Thus, $\gamma\left(h,\boldsymbol{\theta}\right)$ is a nonlinear function with intercept (the nugget τ).

Let $\hat{\boldsymbol{\gamma}}^*$ denote the vector of empirical fitted values $\left(\hat{\gamma}\left(h\right)\right)$ of the variogram, and let $\boldsymbol{g}\left(\boldsymbol{\beta}\right)$ denote the vector of the parametric modeled values $\left(g\left(h;\boldsymbol{\beta}\right)\right)$. Then the estimators of the parameters of the variogram, τ and β, are

$$\hat{\tau}_s=\operatorname{med}\left\{\hat{\boldsymbol{e}}\right\} \qquad (11)$$

$$\hat{\boldsymbol{\beta}}_{\varphi} = \text{Argmin}_{\boldsymbol{\beta}\in\Theta} \left\| \hat{\boldsymbol{\gamma}}^{*} - \boldsymbol{g}\left(\boldsymbol{\beta}\right) \right\|_{\varphi} \tag{12}$$

respectively, and $\hat{\boldsymbol{\theta}}_{\varphi} = \left(\hat{\tau}_s, \hat{\boldsymbol{\beta}}_{\varphi}'\right)'$.

B. Standard Errors of Non-Linear Rank-Based Estimators

The approximate standard error of the rank-based non-linear estimators for $\hat{\boldsymbol{\theta}}_{\varphi} = \left(\hat{\tau}_s, \hat{\boldsymbol{\beta}}_{\varphi}'\right)'$ above with a score φ, is given by

$$s.e.\left(\hat{\boldsymbol{\theta}}_{\varphi}\right) = s.e.\left(\left(\hat{\tau}_s, \hat{\boldsymbol{\beta}}_{\varphi}'\right)'\right) = \text{diag}\left(\hat{V}_{\varphi}\left(\left(\hat{\tau}_s, \hat{\boldsymbol{\beta}}_{\varphi}'\right)'\right)\right)^{1/2} \tag{13}$$

where $\hat{V}_{\varphi}$ is the variance-covariance matrix of the estimator with a given score φ, $\hat{V}_{\varphi}$ defined by

$$\hat{V}_{\varphi} = \begin{bmatrix} n^{-1}\hat{\tau}_s^2 & \mathbf{0}' \\ \mathbf{0} & \hat{\tau}_{\varphi}^2\left(D'\left(\hat{\boldsymbol{\beta}}_{\varphi}\right)D\left(\hat{\boldsymbol{\beta}}_{\varphi}\right)\right)^{-1} \end{bmatrix} \tag{14}$$

where $\hat{\tau}_s$ and $\hat{\tau}_{\varphi}$ are an estimator of the scale parameters τ_s and τ_{φ} respectively, these scales were developed by [18], given by

$$\tau_{\varphi}^{-1} = \int \varphi(u)\varphi_h(u)du \tag{15}$$

$$\tau_s = \left(2h\left(H^{-1}(1/2)\right)\right)^{-1} \tag{16}$$

where

$$\varphi_h(u) = -h'\left(H^{-1}(u)\right)/h\left(H^{-1}(u)\right) \tag{17}$$

and h is a probability density function of the model errors. τ_s and τ_{φ} can be computed by the **Rfit** package [16] on *R* 3.2.2.

C. Optimal Scores

The rank-based analysis aforementioned in Section III-A depends on the selection of a score function $\varphi(\cdot)$, If the form of the underlying error distribution is known, we can obtain an optimal score function that minimizes the variance of the estimator.

From (15), τ_{φ}^{-1} can be rewritten as

$$\begin{aligned}\tau_{\varphi}^{-1} &= \int_0^1 \varphi(u)\varphi_h(u)du \\ &= \left(\int_0^1 \varphi(u)\varphi_h(u)du \Big/ \left(\int_0^1 \varphi_h^2(u)du\right)^{1/2}\right)\left(\int_0^1 \varphi_h^2(u)du\right)^{1/2} \\ &= \rho\left(\sqrt{\int_0^1 \varphi_h^2(u)du}\right) = \rho\sqrt{I(h)}\end{aligned}$$

thus,

$$\tau_{\varphi}^{-1} = \rho\sqrt{I(h)} \tag{18}$$

where ρ is a correlation coefficient and $\sqrt{I(h)}$ is a Fisher Information. Therefore, maximizing $\rho\sqrt{I(h)}$ means minimizing τ_{φ}, but the maximum of $\rho\sqrt{I(h)}$ is satisfied when $\rho = 1$. This is accomplished by taking $\varphi(u) = \varphi_h(u)$. So (15) is the score function which optimizes the rank-based analysis [10], [22].

References [15] and [22] have recommended to fit a model (8) with Wilcoxon score $\varphi_W(u) = \sqrt{12}(u - 1/2)$, then plot the normal Q-Q plot of its residuals $\hat{\boldsymbol{e}}_W$ to determine the error distribution.

We can then choose the optimal score $\varphi(u)$, for example if the error distribution is normal, then the optimal score function simplifies to $\varphi(u) = \Phi^{-1}(u)$, which is the normal scores (where Φ is the cdf of $N(0,1)$).

D. Hogg-Type Adaptive Procedure to Choose the Optimal Score

As we mentioned in Section III *C*, the optimal score function can be chosen by the distribution of the residuals $\hat{\boldsymbol{e}}_W$. Reference [11] proposed an adaptive procedure to choose the optimal score; this procedure is a fully efficient rank-based procedure for testing and estimation, for more details see Chapter 10 of [12].

Let $\hat{\boldsymbol{e}}_W = \left(\hat{e}_1, \hat{e}_2, \ldots, \hat{e}_n\right)'$ be a vector of Wilcoxon residuals of model (8), [12] proposed pair of statistics Q_1 and Q_2 (Q_1 is a measure of skewness and Q_2 is a measure of tail heaviness) as:

$$\begin{aligned} Q_1\left(\hat{\boldsymbol{e}}_W\right) &= \frac{\left(\bar{U}_{0.05} - \bar{M}_{0.5}\right)}{\left(\bar{M}_{0.5} - \bar{L}_{0.05}\right)} \\ Q_2\left(\hat{\boldsymbol{e}}_W\right) &= \frac{\left(\bar{U}_{0.05} - \bar{L}_{0.05}\right)}{\left(\bar{U}_{0.5} - \bar{L}_{0.5}\right)} \end{aligned} \tag{19}$$

where $\bar{U}_\alpha$ is the mean of the upper $\alpha\%$, $\bar{M}_\alpha$ is the mean of the middle $\alpha\%$, and $\bar{L}_\alpha$ is the mean of the lower $\alpha\%$ of the residuals $\hat{\boldsymbol{e}}_W$. Reference [29] suggested a set of cutoffs to select the optimal score function as:

$$\begin{aligned} c_{l_{q1}} &= 0.36+(0.68/n) \\ c_{l_{q2}} &= \begin{cases} 2.17-(3.01/n) & n<25 \\ 2.24-(4.68/n) & n\geq 25 \end{cases} \\ c_{u_{q1}} &= 2.73-(3.72/n) \\ c_{u_{q2}} &= \begin{cases} 2.63-(3.94/n) & n<25 \\ 2.95-(9.37/n) & n\geq 25 \end{cases} \end{aligned} \quad (20)$$

and his scheme was recently discussed by [17]. Depending on these cutoff values, one can select the optimal score function as follows:

- If $Q_1 \leq c_{l_{q1}}$ and $Q_2 \leq c_{l_{q2}}$, then select $\varphi_{\text{bent2}}(u;0.15,0.65,-1,2,0)$
- If $Q_1 \leq c_{l_{q1}}$ and $c_{l_{q2}} < Q_2 \leq c_{u_{q2}}$, then select $\varphi_{\text{bent3}}(u;0.3,-1,2)$
- If $Q_1 \leq c_{l_{q1}}$ and $Q_2 > c_{u_{q2}}$, then select $\varphi_{\text{bent3}}(u;0.5,-1,2)$
- If $c_{l_{q1}} < Q_1 \leq c_{u_{q1}}$ and $Q_2 \leq c_{l_{q2}}$, then select $\varphi_{\text{bent2}}(u;0.25,0.75,-1,1,0)$
- If $c_{l_{q1}} < Q_1 \leq c_{u_{q1}}$ and $c_{l_{q2}} < Q_2 \leq c_{u_{q2}}$, then select $\varphi_W(u)$
- If $c_{l_{q1}} < Q_1 \leq c_{u_{q1}}$ and $Q_2 > c_{u_{q2}}$, then select $\varphi_{\text{bent4}}(u;0.25,0.75,-1,1,0)$
- If $Q_1 > c_{u_{q1}}$ and $Q_2 \leq c_{l_{q2}}$, then select $\varphi_{\text{bent2}}(u;0.35,0.85,-2,1,0)$
- If $Q_1 > c_{u_{q1}}$ and $c_{l_{q2}} < Q_2 \leq c_{u_{q2}}$, then select $\varphi_{\text{bent1}}(u;0.7,-2,1)$
- If $Q_1 > c_{u_{q1}}$ and $Q_2 > c_{u_{q2}}$, then select $\varphi_{\text{bent1}}(u;0.5,-2,1)$

(21)

where $\varphi_W(u)=\sqrt{12}(u-1/2)$ is Wilcoxon score. The remaining score functions are defined as:

$$\varphi_{\text{bent1}}(u;s_1,s_2,s_3)= \begin{cases} ((s_3-s_2)/s_1)u+s_2 & \text{if } 0<u<s_1 \\ s_3 & \text{if } s_1\leq u<1 \end{cases}$$

$$\varphi_{\text{bent2}}(u;s_1,s_2,s_3,s_4,s_5)= \begin{cases} ((s_5-s_3)/s_1)u+s_3 & \text{if } 0<u<s_1 \\ ((s_4-s_5)/(1-s_2))(u-s_2)+s_5 & \text{if } s_1\leq u<1 \text{ and } u>s_2 \\ s_5 & \text{Otherwise} \end{cases}$$

$$\varphi_{\text{bent3}}(u;s_1,s_2,s_3)= \begin{cases} s_2 & \text{if } 0<u<s_1 \\ ((s_3-s_2)/(1-s_2))(u-s_1)+s_2 & \text{if } s_1\leq u<1 \end{cases}$$

$$\varphi_{\text{bent4}}(u;s_1,s_2,s_3,s_4)= \begin{cases} s_4 & \text{if } 0<u<s_2 \\ ((s_4-s_3)/(s_2-s_1))(u-s_1)+s_3 & \text{if } s_2\leq u<1 \end{cases} \quad (22)$$

In this paper $\varphi_W(u)$, $\varphi_{\text{bent1}}(u;0.5,-2,1)$, and $\varphi_{\text{bent4}}(u;0.25,0.75,-1,1,0)$, are denoted WILX, BENT1, and BENT4 respectively. We applied this procedure in Section V-A to choose the optimal scores to fit the variogram model.

IV. Implementation

A. Utility: Simulation Convergence Rate

Reference [3] have used the Newton-Raphson algorithm in their simulation studies to estimate the parameters in (4), but they found around 10% of the simulations fail to converge.

The ***nlminb*** function within **stats** package in *R* 3.2.2 programming language [26] was used in this paper to estimate the parameters in (4) and (12). The ***nlminb*** function uses an unconstrained and box-constrained optimization that depends on PORT routines.

To avoid the initial values problem to estimate the parameters, we used the grid initial values search procedure. We ran this procedure over possible values of the parameters (nugget(τ), sill(σ^2), and rang(ϕ)).

In our simulation study in Section VI, a grid of 6 possible values for the nugget and 20 possible values for the sill and the range were used. We found the grid initial values search procedure with these grid values is adequate, and all the cases converge. This procedure certainly increased the utility of the estimation procedure.

B. Efficiency: Asymptotic Relative Efficiency

In this paper we used another robust nonparametric method to obtain the asymptotic relative efficiency between NLWLS and the rank-based estimator. This method depends on a median absolute deviation (*MAD*). Let $\hat{\boldsymbol{\theta}}$ be the true estimate we used to generate the simulations, and $\hat{\boldsymbol{\theta}}_{k,i}$ be the estimate of the k^{th} model at i^{th} simulation, then, we define *MAD* of the k^{th} model as

$$MAD_k = \text{Median}_i \left| \hat{\boldsymbol{\theta}}_{k,i} - \hat{\boldsymbol{\theta}} \right| \quad (23)$$

Let MAD_1 be the *MAD* of NLWLS model, then the asymptotic relative efficiency (*ARE*) of the k^{th} rank-based estimator with respect to NLWLS estimator is given by

$$ARE = MAD_1 / MAD_k \tag{24}$$

Hence, values of this ratio less than 1 are favorable to the NLWLS while values greater than 1 are favorable to the k^{th} rank-based estimator. We applied this in simulation studies (Section VI) to prove the efficiency of our method to fit the variogram models.

C. Validity: Quasi-Block-Jackknife Method for Constructing Confidence Intervals for Variogram Model Estimates $\boldsymbol{\theta}$

As we mentioned in Section II, $\left(\hat{\gamma}(h)\right)$ in (1) and (2) are correlated, thus the usual standard nonparametric jackknife, and standard nonparametric bootstrap methods perform poorly for confidence intervals unless these correlations are negligible [30]. There are many improvements suggested to develop the bootstrap and jackknife methods for correlated data.

One of these improvements is the quasi-block-jackknife; this method is suggested by [3].

Let $Z = \left\{Z(s_i) : s_i \in \mathbb{R}^d ; i = 1,2,\ldots,n^2\right\}$ be a random field in $n \times n$ equally grid lattice generated using variogram models (3) with true parameters $\boldsymbol{\theta} = (\tau, \sigma^2, \phi)'$. For our case the quasi-block-jackknife method to construct the confidence interval for $\boldsymbol{\theta}$ can be summarized as follows:

i. Calculate the variance-covariance matrix $C(h;\boldsymbol{\theta})$ of the generated data with true parameters $\boldsymbol{\theta} = (\tau, \sigma^2, \phi)'$, where $C(d_{ij};\boldsymbol{\theta})$ is defined as

$$C(d_{ij};\boldsymbol{\theta}) = \begin{cases} \tau + \sigma^2 & \text{if } d_{ij} = 0 \\ \tau + \sigma^2 - \gamma(d_{ij};\boldsymbol{\theta}) & \text{if } d_{ij} > 0 \end{cases} \tag{25}$$

where $\gamma(d_{ij};\boldsymbol{\theta})$ is defined in (10), and $\left[d_{ij}\right]$ are the entries of the $n^2 \times n^2$ distance matrix D.

ii. Calculate a Cholesky decomposition for C as follows: $C = LL'$. Then find the transformation $U = L^{-1}Z$. This U is approximately uncorrelated and normally distributed with mean 0 and variance 1 (i.e. $U \overset{iid}{\sim} N(0,1)$).

iii. Divide the region into B non-overlapping equally size blocks, each block of size l. These blocks are $U^* = \{U_b : b = 1,2,\ldots,B\}$. Note $(l)(B) = n^2$ and each U_b is of size l. Similarly, divide the distance matrix D into B non-overlapping equally size blocks, each block of size l^2, these blocks are $C^* = \{C_b : b = 1,2,\ldots,B\}$. Note $(l^2)(B) = n^4$.

iv. Drop one block b at a time from U^* to get a database $\{U_1^*, U_2^*, \ldots, U_B^*\}$ where $U_j^* = U^* \backslash U_j ; j = 1,2,\ldots,B$. Similarly, drop one block b at a time from C^* to get a database $\{C_1^*, C_2^*, \ldots, C_B^*\}$ where $C_j^* = C^* \backslash C_j ; j = 1,2,\ldots,B$.

v. For each $j = 1,2,\ldots,B$, find the Cholesky decomposition for C_j^* as: $C_j^* = L_j^* L_j^{*\prime}$. Then re-correlate U_j^* by calculating $Z_j^* = L_j^* U_j^*$.

vi. For a certain number of lags *k*, and for each Z_j^*, estimate the empirical variogram to get *h* and $\hat{\gamma}(h)$ from (1) or (2), which we denote $\left\{\left(h_{i,j}, \hat{\gamma}(h_{i,j})\right) : i = 1,\ldots,k ; j = 1,2,\ldots,B\right\}$.

vii. For each variogram estimated in step (vi), fit the variogram model you used to generate the random field above, this would be by (4) for the NLWLS or (12) for the rank-based, let these estimates $\left\{\hat{\boldsymbol{\theta}}_j ; j = 1,2,\ldots,B\right\}$.

viii. Calculate the standard error estimator of $\hat{\boldsymbol{\theta}}$ as:

$$s.e._{jk}(\hat{\boldsymbol{\theta}}) = \left[\left(\frac{B-1}{B}\right)\sum_{j=1}^{B}\left(\hat{\boldsymbol{\theta}}_j - \bar{\hat{\boldsymbol{\theta}}}\right)\left(\hat{\boldsymbol{\theta}}_j - \bar{\hat{\boldsymbol{\theta}}}\right)'\right]^{1/2} \tag{26}$$

where $\bar{\hat{\boldsymbol{\theta}}} = (1/B)\sum_{j=1}^{B}\hat{\boldsymbol{\theta}}_j$.

ix. Finally, construct the $(1-\alpha)\%$ confidence interval for $\boldsymbol{\theta}$ as

$$\hat{\boldsymbol{\theta}} \mp z_{\alpha/2}\, s.e._{jk}(\hat{\boldsymbol{\theta}}) \tag{27}$$

There are some conditions necessary for the quasi-block-jackknife method: the number of blocks *B* should be large enough respect to the number of the lags *k* and the effective range ϕ [3]. Furthermore, the block size l^2 should be greater than the effective range $(l^2 > \phi)$, with large enough block size, so the data from different blocks is approximately uncorrelated.

Reference [2] have defined the integral range $A = \int \rho(h) dh$ as a measure of the strength of spatial correlations over a region *V* (with volume $|V|$), where $\rho(h)$ is the correlation at distance lag *h*.

The closed form of *A* for the exponential model (3) are provided by [7] as: $A_{Exponential} = 2\pi\phi^2 / 9$. Thus, to get valid

confidence intervals for $\boldsymbol{\theta} = (\tau, \sigma^2, \phi)'$, it should be $(|V|/A) > \phi$. This implies $|V| > \phi A$, and $|V| > 2\pi\phi^3/9$ for exponential model.

We used the block-jackknife method in our simulation studies in Section VI to prove the validity of our method to fit the variogram models.

V. Robustness of the Non-Linear Rank-Based Estimator

We will examine some numerical examples to demonstrate the robustness properties of the non-linear rank-based norm.

A. Wheat-Yield Data:

The wheat-yield data [5] were presented by [23] and these data consist of yields on a 20×25 regular lattice ($n = 500$). Fig. 1 shows the locations of these measurements and the relative variation of the yield values.

Reference [21] found in the east-west spectrum. This trend was neither a linear nor periodic [5]. Reference [5] suggested the median-polish procedure to remove this trend. After we ran the median-polish procedure, we replace the yield values by the residuals of this procedure, therefore, stationary can be assumed in the east-west direction (90 degrees).

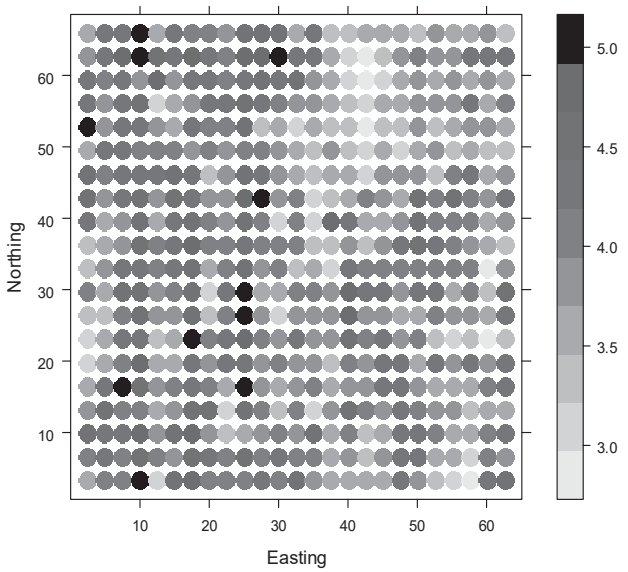

Fig. 1 Sampling Locations of the Wheat-Yield Data (Distances are measured in ft.)

A summary of the data indicates the maximum lagged distance is 86.949 ft. Also, further exploratory data analysis and examination detected an outlier at sits 191 and 208 of grid locations $(40.16, 26.4)$ and $(20.08, 29.7)$ respectively, with 1.1475 and -1.225 median-polish's residuals respectively.

A bubble plot presented in Fig. 2 indicates this outlier and its spatial location. This outlier's contribution to the Matheron estimate can be shown in Fig. 3 is indicated by the points above the dashed line. Here, we just considered that point with distance 20.16717, and $\hat{\gamma}(20.167) = 3.229$ because it is an outlier in *y*- space, and this point is at site 208.

For the estimation step, 13 lags with at least 30 points in each lag were used to ensure that the empirical variogram at each point was well estimated; with maximum lag distance at about half the maximum separation distance [14], [4].

As presented by [5], the data has an apparent exponential variogram model (3). For the fitting step, as we mentioned before, first the Wilcoxon (WLIX) was fitted for model (3) using both the Matheron and Cressie variogram estimators.

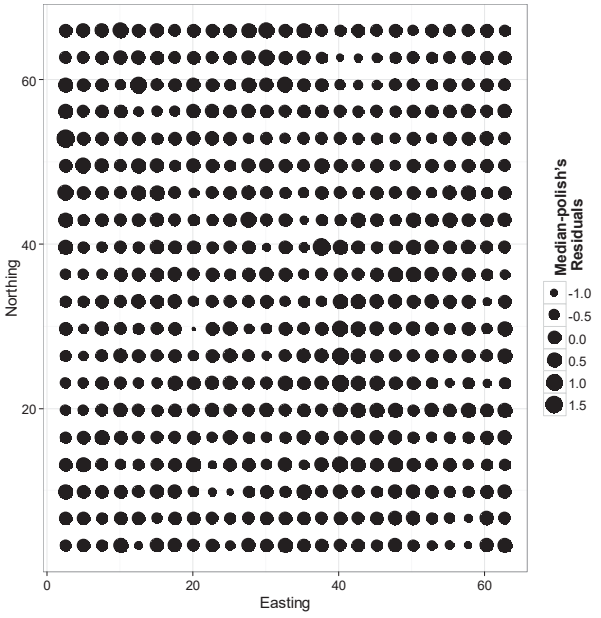

Fig. 2 Bubble Plot of the Wheat-Yield Data (Distances are measured in ft.)

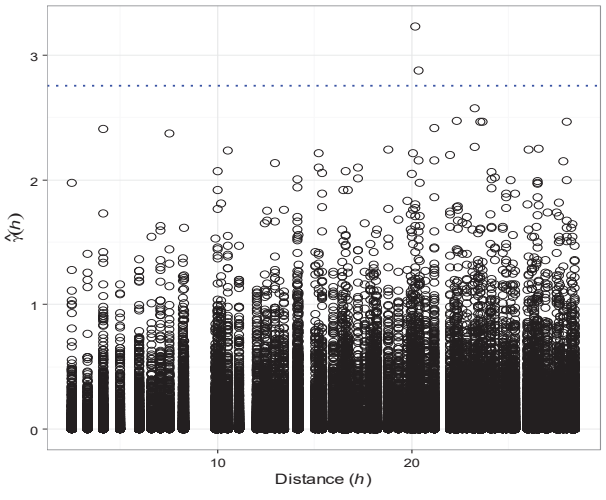

Fig. 3 Variogram Cloud of the Wheat-Yield Data

The normal Q-Q and the density estimate plots of its residuals can be seen in Fig. 4 (a). We can see that the residuals are approximately heavy-tailed symmetric for the Matheron estimate (1) in Fig. 4 (a) and for the Cressie estimate (2) in Fig. 4 (b).

Now to get the optimal scores to fit the exponential model, the Hogg-type adaptive procedures in Section III-D were applied. The BENT4 score function was chosen for both the Matheron and Cressie estimates. The plots of h and $\gamma(h)$, Fig. 5 (a), show the NLWLS fit and rank-based fit for the BENT4 score (with and without outlier and the Matheron estimate), and Fig. 5 (b), shows NLWLS fit and rank-based fits for the BENT4 score (with and without outlier and the Cressie estimate).

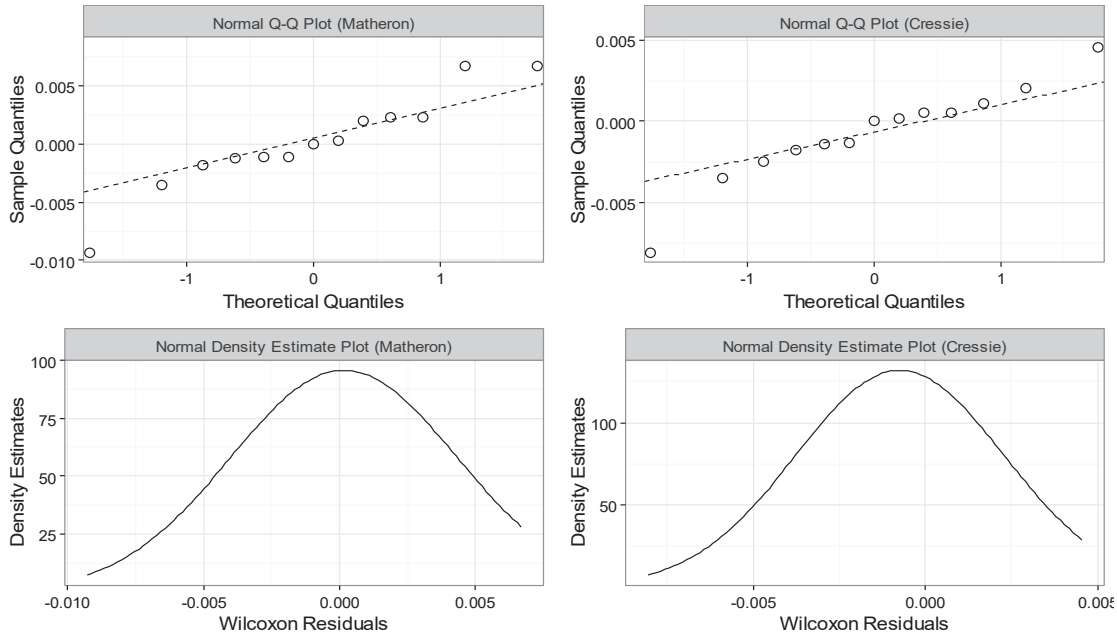

Fig. 4 (a) The Normal Q-Q and Density Estimates Plots Based on the residuals $\hat{e}_W$ of the Wilcoxon Fit of Wheat-Yield Data (With Outlier)

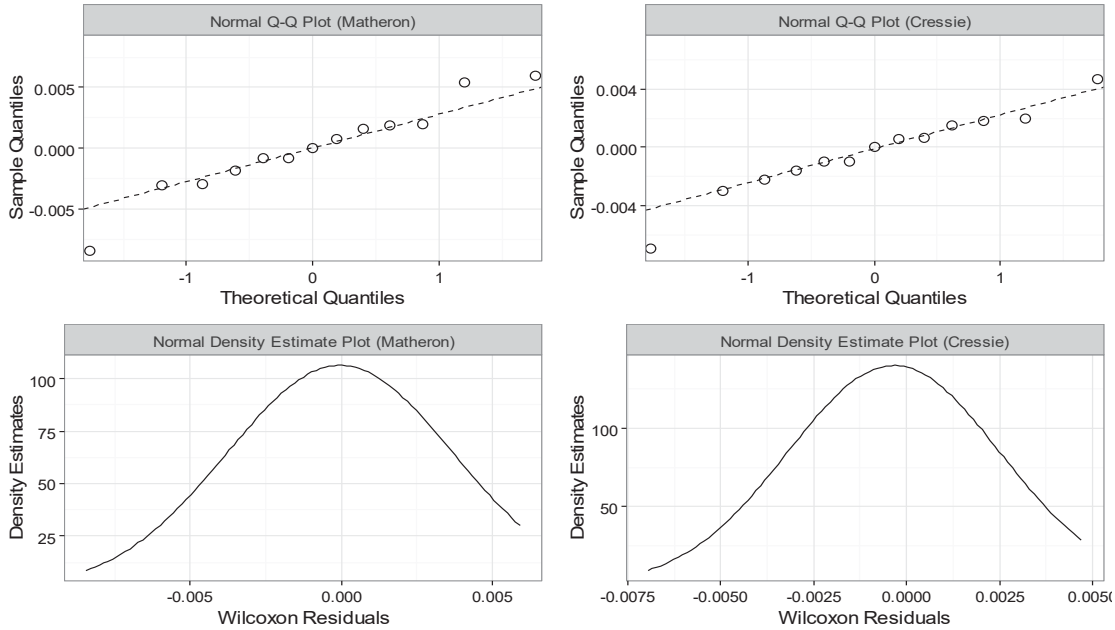

Fig. 4 (b) The Normal Q-Q and Density Estimates Plots Based on the residuals $\hat{e}_W$ of the Wilcoxon Fit of Wheat-Yield Data (Without Outlier)

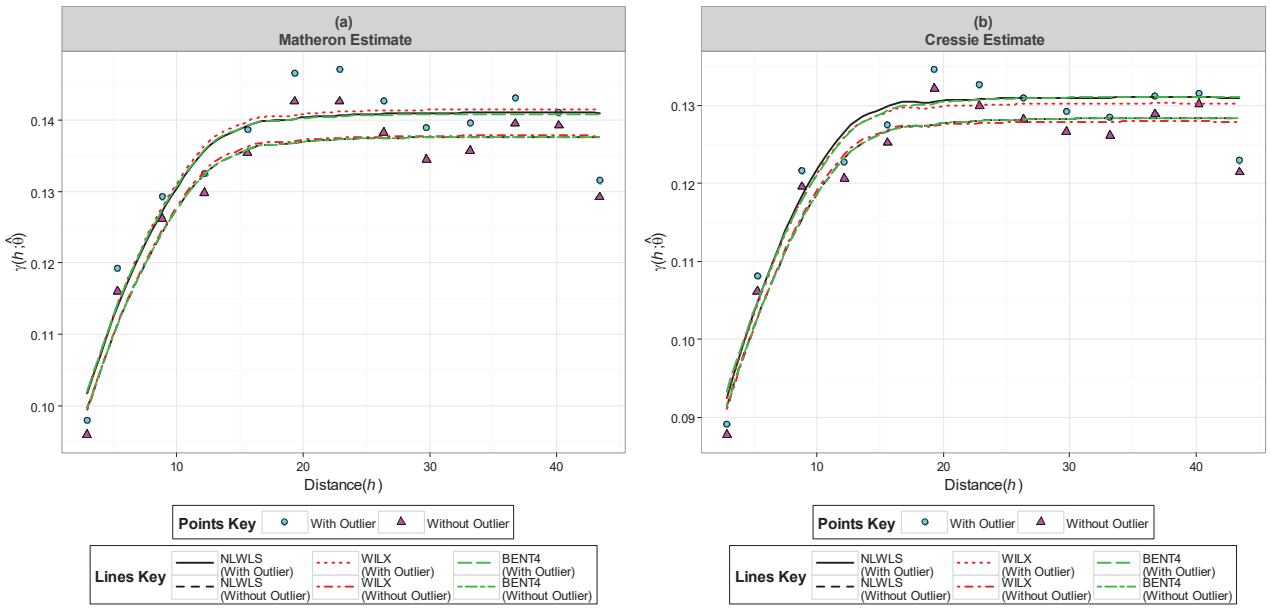

Fig. 5 Plot of the Fits for the Wheat-Data

In Table I, we display the NLWLS, WILX, and BENT4 fits for the Matheron and the Cressie estimates, using model (3).

Along with the estimates, we show the approximate standard errors of the estimates $s.e.\left(\hat{\theta}\right)$ that were discussed in Sections II-B and III-B.

TABLE I
WHEAT-YIELD DATA: EXPONENTIAL VARIOGRAM PARAMETER ESTIMATES AND APPROXIMATE STANDARD ERRORS WITH AND WITHOUT OUTLIER (THE MATHERON AND CRESSIE ESTIMATES)

	Model	Nugget (τ)		Sill (σ^2)		Range (ϕ)	
		With Outlier	Without Outlier	With Outlier	Without Outlier	With Outlier	Without Outlier
Matheron	NLWLS	0.0570	0.0558	0.0844	0.0820	4.2711	4.2543
	Approx.s.e.	0.0299	0.0273	0.0294	0.0269	1.3412	1.2510
	WILX	0.0586	0.0586	0.0822	0.0790	4.2864	4.3980
	Approx.s.e.	0.0201	0.0200	0.0195	0.0194	1.1510	1.2412
	BENT4	0.0593	0.0586	0.0818	0.0790	4.3995	4.3995
	Approx.s.e.	**0.0178**	**0.0140**	**0.0173**	**0.0137**	**1.0684**	**0.8744**
Cressie	NLWLS	0.0449	0.0450	0.0854	0.0830	3.9742	3.9993
	Approx.s.e.	0.0234	0.0219	0.0231	0.0216	0.9399	0.9149
	WILX	0.0525	0.0516	0.0786	0.0768	4.3982	4.3979
	Approx.s.e.	0.0166	0.0144	0.0162	0.0140	1.0381	0.9203
	BENT4	0.0441	0.0516	0.0869	0.0768	3.9862	4.3995
	Approx.s.e.	**0.0134**	**0.0125**	**0.0131**	**0.0121**	**0.6460**	**0.7981**

By comparing the results in Table I, we can see that:

a. For the sill (σ^2), regardless of the outlier, all the robust procedures estimates outperformed (smaller s.e.) the NLWLS estimate with regards to the standard error of the estimates (for both the Matheron and Cressie estimates). The BENT4 outperformed (smaller s.e.) the WILX estimate regardless of outlier (for both the Matheron and Cressie estimates).
b. For the range (ϕ), regardless of outlier, all the robust procedures estimates outperformed the NLWLS estimate with regards to the standard error (smaller s.e.) of the estimates (for the Matheron estimate). The BENT4 outperformed the WILX estimate regardless of outlier (for the Matheron estimate).
c. For the range (ϕ), regardless of outlier, the NLWLS outperformed (the smallest s.e.) the WILX estimate (for the Cressie estimate). The BENT4 outperformed (the smallest s.e.) the NLWLS and WILX estimates regardless of outlier (for the Cressie estimate).

VI. SIMULATION STUDY

We now present the results for some simulations that investigate the utility, validity, and efficiency of the rank-based fits (with the Hogg-type adaptive procedures) compared with NLWLS. All of these simulations were performed on spatially correlated data in $\mathbb{R}^2$ with east-west direction (90 degrees) and 22.5 degrees of tolerance to avoid anisotropy issues.

We used the exponential model (3) of the wheat-yield example and we set the true values of the parameters $\boldsymbol{\theta}$ as follows: $\boldsymbol{\theta} = (\tau, \sigma^2, \phi)' = (0.056, 0.082, 4.375)'$, which is close to the fitted coefficients in Table I.

The ***RFsimulate*** function within the **RandomFields** package [28] using *R* 3.2.2 programming language [26] was used to simulate these models, and for each model we generated 3,000 Gaussian random fields of $n^2 = 1600$ spatially correlated points on a 40×40 equally grid square

lattice within $[0,100]\times[0,100]$. These simulations were obtained by using the direct matrix decomposition method.

We partitioned the equally grid square 40×40 into $B=16$ blocks each of size $l=100$. n^2, l^2, and the B were chosen under the conditions in Section IV-C.

Next, we chose the normal distribution $N(\mu=0,\sigma^2=1^2)$, with proportion levels $\epsilon=0\%$, $\epsilon=5\%$, $\epsilon=10\%$, and $\epsilon=20\%$ for randomly contaminated for each simulated data and equally number of contaminations in each block (i.e. in each block there are $(1/16)\epsilon$ contaminations).

In the estimating step: We used the Cressie estimate, and the number of lags is 34. As we discussed earlier in this paper, and based on the recommendations of [14], we used at least 30 points in each lag and the maximum lag distance is about half the maximum separation distance.

In the fitting step: This step is performed using NLWLS, WILX and the adaptive procedure (AdSch). It should also be mentioned with regards to the importance of three parameters (nugget(τ), sill(σ^2), and range(ϕ)) in respect to the effectiveness on ordinary kriging. The nugget (τ) and sill (σ^2) influence the kriging variances, and the range (ϕ) parameter influences the ordinary kriging weights, so the range (ϕ) is regarded as the most important kriging parameter [8].

We considered the NLWLS, WILX and the adaptive procedure (AdSch) nonlinear fits. For our efficiency study, we investigated the empirical asymptotic relative efficiency (ARE) for the parameters the nugget (τ), sill (σ^2) and the range (ϕ), as we suggested in Section IV-B. Table II displays the results for the *ARE's* for the robust methods relative to NLWLS (in (24)) for each coefficient over the four contamination levels.

TABLE II
EMPIRICAL ARE'S FOR THE ROBUST ESTIMATES OF THE PARAMETERS τ,σ^2, AND ϕ RELATIVE TO NLWLS FOR THE EXPONENTIAL MODEL (THE CRESSIE ESTIMATE)

	θ	Contamination Level			
		$\epsilon=0\%$	$\epsilon=5\%$	$\epsilon=10\%$	$\epsilon=20\%$
NLWLS, WILX	Nugget (τ)	1.201	1.036	0.979	0.994
NLWLS, AdSch		1.166	1.027	0.977	0.993
NLWLS, WILX	Sill	**1.177**	**1.234**	**1.238**	**1.233**
NLWLS, AdSch	(σ^2)	1.144	1.199	1.220	1.231
NLWLS, WILX	Range (ϕ)	**1.077**	**1.126**	**1.127**	**1.181**
NLWLS, AdSch		1.046	1.111	1.103	1.180

Table II shows for the sill (σ^2) and the range (ϕ), the robust estimates are more efficient than the NLWLS estimate for all contamination levels. The WILX estimate shows slightly more efficiency than the AdSch estimate over all contamination levels.

For our validity study, the quasi-block-jackknife method discussed in Section IV were applied to constructed the empirical confidence coefficients of confidence intervals for the parameters nugget (τ), sill (σ^2) and the range (ϕ) as in Section IV.

Table III displays the results for nominal 95% confidence intervals, the method was used to estimate $\boldsymbol{\theta}$ is considered a valid method if its empirical confidence closed to the nominal confidence of 0.95.

TABLE III
EMPIRICAL CONFIDENCE COEFFICIENTS FOR THE NLWLS AND ROBUST ESTIMATES OF THE PARAMETERS τ,σ^2, AND ϕ. THE NOMINAL CONFIDENCE IS 0.95, FOR THE EXPONENTIAL MODEL (THE CRESSIE ESTIMATE)

	θ	Contamination Level			
		$\epsilon=0\%$	$\epsilon=5\%$	$\epsilon=10\%$	$\epsilon=20\%$
NLWLS	Nugget (τ)	0.978	0.953	0.856	0.545
WILX		0.992	0.941	0.785	0.417
AdSch		0.993	0.949	0.824	0.457
NLWLS	Sill (σ^2)	0.982	0.988	0.993	0.996
WILX		0.991	0.993	0.996	0.997
AdSch		0.995	0.996	0.997	0.997
NLWLS	Range (ϕ)	**0.968**	**0.975**	0.980	0.981
WILX		0.977	0.979	0.985	0.984
AdSch		0.981	0.983	0.988	0.988

For the sill (σ^2) and the range (ϕ), the NLWLS, WILX, and the AdSch intervals are conservative. The exceptions: the range (ϕ) are over 0% and 5% contamination levels, the NLWLS intervals are valid (between 0.925 and 0.975).

VII. SUMMARY

Using the results of the numeric and simulation studies, we find the following:

1. With regards to utility (convergence rate), the simulation studies found the initial convergence rate for the Rank estimates was 1% versus 12% for NLWLS. Utilizing the grid initial values search procedure with the ***nlminb*** function, previously discussed in Section IV-A, the convergence rate was later improved to 100% for the Rank procedure and 99.7% for NLWLS.
2. With regards to efficiency (ARE: Asymptotic Relative Efficiency), the simulation studies showed that the ARE for the Rank estimates was superior to NLWLS ranging from 0.46% to 23.8% for the sill and range.
3. With regards to validity, (Empirical Confidence Intervals), the Rank Based procedures provided acceptable results.
4. The choice of estimation procedure does influence the scoring function selected. In general, there is no optimal score function. The adaptive procedure, however, is shown to be most useful in the selection of a score function.

5. The numeric studies demonstrate the overall robustness of the Rank estimates. The standard errors of the rank estimates showed much less sensitivity than NLWLS.
6. Overall, variogram model chosen does influence the results. But, the Rank estimate, in conjunction with the Hogg-Type adaptive procedure, is a robust and superior estimation procedure over NLWLS.
7. Procedures outperform NLWLS regardless of the estimation procedure or level of contamination.

REFERENCES

[1] Abebe A., and McKean J., (2013), Weighted Wilcoxon Estimators in Nonlinear Regression, Australian & New Zealand Journal of Statistics, v. 55, no. 4, pp. 401–420.
[2] Chiles, J. P., and Delfiner, P., (1999). Geostatistics: Modelling Spatial Uncertainty, New York: John Wiley and Sons Ltd.
[3] Clark, R.G., and Allingham, S., (2011). Robust Resampling Confidence Intervals for Empirical Variograms, Mathematical Geosciences, v. 43, no. 2, pp. 243-259.
[4] Cressie N., (1985), Fitting variogram models by weighted least squares. Journal of the International Association for Mathematical Geology, v. 17, no. 5, pp. 563-586.
[5] Cressie N., (1991). Statistics for spatial data, New York: John Wiley and Sons Ltd.
[6] Cressie N., and Hawkins D. (1980), Robust estimation of the variogram, I: Math. Geology, v. 12, no. 2, pp. 115-125.
[7] Garrigues, S., and Allard, D., Baret, F., Weiss, M., (2006), Quantifying spatial heterogeneity at the landscape scale using variogram models. Remote Sensing of Environment, vol. 103, no. 1, pp.81-96.
[8] Genton M., (1998), Highly robust variogram estimation, Mathematical Geology, v. 30, no. 2, pp. 213-221.
[9] Hampel F., (1973), Robust estimation, a condensed partial survey: Zeitschrift fur Wahrschein-lichkeitstheorie und verwandte Gebiete, vol. 27, no. 2, pp. 87-104.
[10] Hettmansperger, T., and McKean J. W., (2011). Robust Nonparametric Statistical Methods, 2nd Ed., New York: Chapman-Hall.
[11] Hogg, R. V., Fisher, D. M., and Randles, R. H. (1975), A two-sample adaptive distribution-free test, Journal of the American Statistical Association, v. 70, pp. 656-661.
[12] Hogg, R. V., McKean, J. W., and Craig, A. T. (2013). Introduction to Mathematical Statistics, 7th Ed., Boston: Pearson.
[13] Huber, P. J., (1996). Robust statistical procedures, 2nd Ed., Philadelphia: Society for Industrial and Applied Mathematics, 56 p.
[14] Journal, A., and Huijbregts, C., (2003). Mining Geostatistics, reprint Ed., New York: Academic Press.
[15] Kloke, J., and McKean, J. W., (2012), Rfit: Rank-based estimation for linear models, R Journal, v. 4, no. 2, pp. 57-64.
[16] Kloke, J., McKean, J., (2015, September 1), Rfit: Rank Estimation for Linear Models. R package version 0.22.0. Retrieved from http://CRAN.R-project.org/package=Rfit.
[17] Kloke, J., McKean, J. W., (2014). Nonparametric statistical methods using R, FL: Chapman-Hall.
[18] Koul, H. L., Sievers, G., and McKean, J. W., (1987), An Estimator of the Scale Parameter forthe Rank Analysis of Linear Models under General Score Functions. Scandinavian Journal of Statistics, v. 14, no. 2, pp. 131-141.
[19] Mardia, K., and Marshall, R., (1984), Maximum likelihood estimation of models for residual covariance in spatial regression, Biometrika. v. 71, no. 1, pp. 135-146.
[20] Matheron, G., (1962), Traite de geostatistique appliquee, Tome I: Memories du Bureau de Recherches Geologiques et Minieres, no. 14, Editions Technip, Paris, 333 p.
[21] McBratney, A. B., Webster, R., and Burgess, T. M., (1981), The design of optimal sampling schemes for local estimation and mapping of regional variables-I: theory and method, Computers and Geosciences, v. 7, pp. 331-334.
[22] McKean, J. W., and Kloke, J., (2014), Efficient and Adaptive Rank-Based Fits for Linear Models with Skewed-Normal Errors. Journal of Statistical Distributions and Applications, v. 1, no. 1, pp. 1-18.
[23] Mercer, W. B. and Hall, A. D. (1911), The experimental error of field trials. Journal of Agricultural. Science 4, pp. 107-132.
[24] Patterson H., and Thompson, R., (1971), Recovery of inter-block information when block sizes are unequal. Biometrika, v. 58, no. 3, pp. 545-54.
[25] Patterson, H., and Thompson, R., (1974), Paper presented at: 8th International Biometrics Conference, Maximum likelihood estimation of components of variance. Constanta, Romania, pp. 197-207.
[26] R Core Team, (2015, September 1), R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from http://www.R-project.org/.
[27] Rao, C., (1979), MINQE Theory and its Relation to ML and MML Estimation of Variance Components, The Indian Journal of Statistics, v. 41, Series B, Pts. 3 and 4, pp. 138-153.
[28] Schlather, M., Malinowski, A., Oesting, M., Boecker, D., Strokorb, K., Engelke, S., Martini, J., Ballani, F., Moreva, O., Menck, PJ., Gross, S., Ober, U., Berreth, C., Burmeister, K., Manitz, J., Morena, O., Ribeiro, P., Singleton, R., Pfaff, B., and R Core Team, (2015, August 1). RandomFields: Simulation and Analysis of Random Fields. R package version 3.1.1. Retrieved from http://CRAN.R-project.org/package=RandomFields (accessed September 1, 2015).
[29] Shomrani, A., (2003), A Comparison of Different Schemes for Selecting and Estimating Score Functions Based on Residuals, unpublished Ph.D. dissertation, Western Michigan University, Kalamazoo, Michigan.
[30] Tang, L., Schucany, W., Woodward W, and Gunst R., (2006), A Parametric Spatial Bootstrap, Technical Report SMU-TR-337, Southern Methodist University, Dallas, Texas.