World Academy of Science, Engineering and Technology
International Journal of Mathematical and Computational Sciences
Vol:10, No:9, 2016

Variogram Fitting Based on the Wilcoxon Norm

Hazem Al-Mofleh, John Daniels, Joseph McKean

***Abstract*—**Within geostatistics research, effective estimation of the variogram points has been examined, particularly in developing robust alternatives. The parametric fit of these variogram points which eventually defines the kriging weights, however, has not received the same attention from a robust perspective. This paper proposes the use of the non-linear Wilcoxon norm over weighted non-linear least squares as a robust variogram fitting alternative. First, we introduce the concept of variogram estimation and fitting. Then, as an alternative to non-linear weighted least squares, we discuss the non-linear Wilcoxon estimator. Next, the robustness properties of the non-linear Wilcoxon are demonstrated using a contaminated spatial data set. Finally, under simulated conditions, increasing levels of contaminated spatial processes have their variograms points estimated and fit. In the fitting of these variogram points, both non-linear Weighted Least Squares and non-linear Wilcoxon fits are examined for efficiency. At all levels of contamination (including 0%), using a robust estimation and robust fitting procedure, the non-weighted Wilcoxon outperforms weighted Least Squares.

***Keywords*—**Non-Linear Wilcoxon, robust estimation, Variogram estimation.

I. Introduction

A. Estimation Step

VARIOGRAM construction is a two stage process. First, the variogram points must be estimated from the raw data. Second, the parameters associated with the variogram must also be estimated (variogram fit). To avoid confusion in this paper, the first step will be called the estimation step and the second step will be called the fitting step. Both of these are vital steps in the spatial prediction process since this process determines the kriging weights. Since most data contain contamination, due to either the nature of the random process which produced the data or human error in measuring the data, the use of procedures such as the one developed in this paper become one of the ways of improving inference. Reference [1] discussed that measured data can contain 10-15% outliers while [2] showed that this proportion can be as high as 30%. The estimation step has received a great deal of attention and there are many competing procedures. The first, proposed by [3] and referred to as the *Matheron* estimator in this paper, is based on the method-of-moments

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

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

H. M. Al-Mofleh is with the Department of Mathematics, Tafila Technical University, Tafila, 66110 Jordan (phone: 00962788510934, e-mail: almof1hm@cmich.edu).

J. E. Daniels is with the Department of Mathematics, Central Michigan University, Mount Pleasant,MI, 48858 USA (e-mail: danie1je@cmich.edu).

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

A robust estimator, proposed by [4] and referred to as the *Cressie* estimator in this paper, is based on fourth roots of the squared differences.

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

where $N(h) \equiv \{(s_i, s_j) : s_i - s_j = h; i, j = 1, \ldots, n\}$ and $|N(h)|$ is the number of distinct pairs in $N(h)$. Further robust procedures for the estimation step have been proposed by [5], [6].

B. Fitting Step

Once the variogram points have been estimated, the fitting step must include a theoretical model in which to approximately describe the spatial continuity of the data. Certain models (i.e., mathematical functions) that are known to be positive definite already exist and are commonly used. One example, a spherical variogram model, is defined as:

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

where $\boldsymbol{\theta} = (\tau, \sigma^2, \phi)'$.

In considering the evolution of the fitting step, after the theoretical model is chosen, early procedures that relied on an underlying Gaussian assumption like maximum likelihood [7] were found to be biased with small sample sizes. Further research resulted in restricted maximum likelihood first developed by [8], [9]. Minimum norm quadratic (MINQ) estimation used by [10] requires linearity in the parameters of the variance-covariance matrix. Both ordinary, generalized, and weighted least squares fitting are highly appealing, are presently in use, but suffer from the same outlier issues that have always surrounded mean based procedures. To further this research, we present a robust fitting step based on the non-linear Wilcoxon estimator.

II. Non-Linear Weighted Least Squares

The variogram estimates $(\hat{\gamma}(h))$ in both (1) and (2) are correlated with non-constant variances [11]. These are violations of the independence and heteroscedasticity general assumptions of ordinary non-linear least square, hence this method cannot be applied in this context. Reference [11] suggested a non-linear weighted least squares (NLWLS) procedure to fit a variogram. The weights used in NLWLS suggested by [11] are $|N(h)|/\left(\gamma(h;\boldsymbol{\theta})\right)^2$. The NLWLS estimates are obtained by minimizing the norm

$$\sum_{j=1}^{n} \frac{|N(h_j)|}{(\gamma(h_j,\boldsymbol{\theta}))^2} (\hat{\gamma}(h_j) - \gamma(h_j;\boldsymbol{\theta}))^2 \quad (4)$$

where n is the number of lags, $\gamma(h_j,\boldsymbol{\theta})$ 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.

A. 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 $\hat{\boldsymbol{\theta}}$) $D_{n\times p}(\hat{\boldsymbol{\theta}})$, where $D_{(i,j)}(\hat{\boldsymbol{\theta}}) = [\partial\gamma_i(\hat{\boldsymbol{\theta}})/\partial\theta_j]$, $i = 1,2,\ldots,p; j = 1,2,..,n$, 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

$$s.e._{NLWLS}(\hat{\boldsymbol{\theta}}) = \left[diag\left(\hat{\sigma}^2 \left(D'(\hat{\boldsymbol{\theta}}) W(\hat{\boldsymbol{\theta}}) D(\hat{\boldsymbol{\theta}}) \right)^{-1} \right) \right]^{1/2} \quad (5)$$

where $\hat{\sigma}^2 = \sum_{i=1}^{n} w_i(\hat{\boldsymbol{\theta}}) \left(\hat{\gamma}^*(h_i) - \gamma(h_i;\hat{\boldsymbol{\theta}}) \right)^2/(n-p)$, $w_i(\hat{\boldsymbol{\theta}}) = |N(h_i)|/(\gamma(h_i,\boldsymbol{\theta}))^2$, and W is a $p \times p$ diagonal matrix with entries w_i's in the main diagonal.

III. Robust Wilcoxon Estimator

The robust Wilcoxon estimator of the variogram is based on the Wilcoxon nonlinear estimator which we now briefly describe. It is analogous to the Least Square nonlinear estimator because the Euclidean norm is replaced by another norm. We describe it in terms of a general nonlinear model and then proceed to discuss its application to the problem at hand.

Consider the nonlinear model given by

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

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 $h(t)$ and cdf $H(t)$, were $H(t)$ is unknown.

Let $\boldsymbol{y} = (Y_1, Y_2, \ldots, Y_n)'$ and $\mathbf{f}(\boldsymbol{\theta}) = (f_1(\boldsymbol{\theta}), f_2(\boldsymbol{\theta}), \ldots, f_n(\boldsymbol{\theta}))'$. Given a norm $\|.\|$ 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 $\mathbf{y}$ and $\mathbf{f}(\boldsymbol{\theta})$; i.e., $\hat{\boldsymbol{\theta}} = \text{Argmin}_{\boldsymbol{\theta}\in\Theta} \|\boldsymbol{y} - \mathbf{f}(\boldsymbol{\theta})\|$. If the norm is the Euclidean norm then $\hat{\boldsymbol{\theta}}$ is the LS estimate.

For a vector $\boldsymbol{v}$ in $\mathbb{R}$, the Wilcoxon norm is given by

$$\|\boldsymbol{v}\|_W = \sum_{i=1}^{n} \varphi_W \left[\frac{R(v_i)}{n+1} \right] v_i \quad (7)$$

where the score function φ_W is defined by $\varphi_W(v_i) = \sqrt{12}(v_i - (1/2))$, and $R(v_i)$ denotes the rank of v_i among $v_1, v_2, \ldots, v_n$; see Chapters 2 and 3 of [12] or see [13]. Hence, the Wilcoxon estimator of the parameter $\boldsymbol{\theta}$ in (7) is,

$$\hat{\boldsymbol{\theta}}_W = \text{Argmin}_{\boldsymbol{\theta}\in\Theta} \|\boldsymbol{y} - \mathbf{f}(\boldsymbol{\theta})\|_W \quad (8)$$

The properties of the Wilcoxon nonlinear estimator were developed by [14]. The usual Gauss-Newton algorithm which is used to compute the WLS nonlinear estimator can also be used for computation of $\hat{\boldsymbol{\theta}}_W$. Although not entirely applicable in this context, due to spatial correlation, it has been shown that if in fact the random errors are independent and have a normal distribution then the Asymptotic Relative Efficiency (ARE: ratio of LS variance to Wilcoxon variance) is 0.955; that is, at the normal the Wilcoxon estimator is only 4.5% less efficient than the LS estimator.

For error distributions with thicker tails than the normal distribution the Wilcoxon estimator is generally more efficient ($ARE > 1$) than the LS estimator. The influence function of the LS estimator gives further evidence that the LS estimator is not robust. Based on the influence function of the Wilcoxon estimator, obtained by [14], the Wilcoxon estimator is robust to outliers in the response space. Further, although this paper uses the unweighted Wilcoxon versus WLS, weighted versions of the Wilcoxon estimator offer protection for cases where the tangent plane to the surface of $\mathbf{f}(\boldsymbol{\theta})$ at $\boldsymbol{\theta}_0$ can be unbounded; see [14] for discussion. Further research will continue to explore use of the weighted Wilcoxon estimator with spatially correlated data.

For our application, once the raw data have been obtained, we select a procedure for the variogram estimation step. In our applications and simulations, we considered either the *Matheron* estimator as given in (1) or the *Cressie* estimator as given in (2). From this estimation step, suppose a variogram function has been specified; for example, a spherical variogram model provided in (3).

A. Wilcoxon Procedure to Fit a Variogram

Let $\boldsymbol{\theta}$ denote the vector of parameters and let $\gamma(h,\boldsymbol{\theta})$ denote the model of the variogram. For the spherical variogram $\boldsymbol{\theta} = (\tau, \sigma^2, \phi)'$, and $\gamma(h,\boldsymbol{\theta})$ is defined in (3). Denote the empirical estimate of the variogram by $\hat{\gamma}^*(h)$, where $N(h) \equiv \{(s_i, s_j) : s_i - s_j = h; i,j = 1,\ldots,n\}$ and $|N(h)|$ is the number of distinct pairs in $N(h)$. We have used the superscript * to avoid confusion with the parametric model. Let $\hat{\boldsymbol{\gamma}}^*$ denote the vector of empirical fitted values $(\hat{\gamma}(h))$ of the variogram and let $\boldsymbol{\gamma}(\boldsymbol{\theta})$ denote the vector of the parametric modeled values $(\gamma(h,\boldsymbol{\theta}))$. Then our Wilcoxon estimator of the parameters of the variogram, $\boldsymbol{\theta}$, is

$$\hat{\boldsymbol{\theta}}_W = \text{Argmin}_{\boldsymbol{\theta}\in\Theta} \|\hat{\boldsymbol{\gamma}}^* - \boldsymbol{\gamma}(\boldsymbol{\theta})\|_W \quad (9)$$

For the example and simulations discussed, we computed $\hat{\boldsymbol{\theta}}_W$ by the Gauss-Newton algorithm; see [14].

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

$$Y_i = \mathbf{1}_n \tau + g_i(\boldsymbol{\beta}) + \epsilon_i; \qquad i = 1,2,\ldots,n \quad (10)$$

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

$$\hat{\boldsymbol{\beta}}_W = \text{Argmin}_{\boldsymbol{\beta}\in\Theta}\|\boldsymbol{y} - \mathbf{g}(\boldsymbol{\beta})\|_W \tag{11}$$

and the intercept τ can be estimated by a location estimate based on the residuals $\hat{\boldsymbol{e}} = \boldsymbol{y} - \mathbf{g}(\hat{\boldsymbol{\beta}}_W)$. Reference [15] suggest to use the median of the residuals to estimate τ, which denotes by $\hat{\tau}_s = \text{med}(\hat{\boldsymbol{e}})$.

The models in (3) can be rewritten as

$$\gamma(h;\boldsymbol{\theta}) = \begin{cases} 0 & \text{if } h = 0 \\ \tau + g(h;\boldsymbol{\beta}) & \text{if } h > 0 \end{cases} \tag{12}$$

where $\boldsymbol{\theta} = (\tau, \boldsymbol{\beta}')'$ and $g(h;\boldsymbol{\beta})$ is defined as

$$g(h;\boldsymbol{\beta}) = \begin{cases} \sigma^2\left(\frac{3}{2}\left(\frac{h}{\phi}\right) - \frac{1}{2}\left(\frac{h}{\phi}\right)^3\right) & \text{if } 0 < h \leq \phi \\ \sigma^2 & \text{if } h > \phi \end{cases} \tag{13}$$

Thus, $\gamma(h,\boldsymbol{\theta})$ is a nonlinear function with intercept (the nugget τ).

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

$$\hat{\tau}_s = \text{med}(\hat{\boldsymbol{e}}) \tag{14}$$

$$\hat{\boldsymbol{\beta}}_W = \text{Argmin}_{\boldsymbol{\beta}\in\Theta}\|\hat{\boldsymbol{\gamma}}^* - \boldsymbol{g}(\boldsymbol{\beta})\|_W \tag{15}$$

respectively, and $\hat{\boldsymbol{\theta}}_W = (\hat{\tau}, \hat{\boldsymbol{\beta}}_W')'$.

B. Standard Errors of Wilcoxon Estimates

The approximate standard error of the Wilcoxon estimators for $\hat{\boldsymbol{\theta}}_W = (\hat{\tau}, \hat{\boldsymbol{\beta}}_W')'$ above, is given by

$$s.e._{\textit{Wilcoxon}}(\hat{\boldsymbol{\theta}}_W) = \left[\text{diag}\left(\widehat{V}_W\left((\hat{\tau}_s, \hat{\boldsymbol{\beta}}_W')'\right)\right)\right]^{1/2} \tag{16}$$

where $\widehat{V}_W$ is the variance-covariance matrix of the Wilcoxon estimator, and $\widehat{V}_W$ defined by

$$\widehat{V}_W = \begin{bmatrix} n^{-1}\hat{\tau}_s^2 & \mathbf{0}' \\ \mathbf{0} & \hat{\tau}_W^2\left(D'(\hat{\boldsymbol{\beta}}_W)D(\hat{\boldsymbol{\beta}}_W)\right)^{-1} \end{bmatrix} \tag{17}$$

where $\hat{\tau}_s$ and $\hat{\tau}_W$ are estimators of the scale parameters τ_s and τ_W respectively. Reference [15] proposed an estimator of τ_ϕ. Estimation of τ_s is discussed in [12]. These scales are given by

$$\tau_{\varphi_W}^{-1} = \int \varphi_W(u)\varphi_h(u)du \tag{18}$$

$$\tau_s = \left(2h(H^{-1}(1/2))\right)^{-1} \tag{19}$$

where $\varphi_W(u) = \sqrt{12}(u - (1/2))$ and $\varphi_h(u) = -h'\left(H^{-1}(u)\right)/h\left(H^{-1}(u)\right)$, and h is a probability density function of the model errors. The **Rfit** package [16] on R 3.2.2 computes these estimates.

IV. Implementation

A. Utility: Simulation Convergence Rate

Reference [17] have used the Newton-Raphson algorithm in their simulation studies to estimate the parameters, but they found around 10% of the simulations fail to converge. The *nlminb* function within **stats** package in R 3.2.2 programming language [18] was used in this paper to estimate the parameters in (4) and (9). 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 range(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 (ARE) 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 k^{th} model at the i^{th} simulation. Then, we define MAD of the k^{th} model as

$$MAD_k = \text{Median}_i|\hat{\boldsymbol{\theta}}_{k,i} - \hat{\boldsymbol{\theta}}| \tag{20}$$

Let MAD_1 and MAD_2 be the median absolute deviations (MAD's) of the NLWLS and Wilcoxon models, respectively, then the asymptotic relative efficiency (ARE) of the Wilcoxon estimator with respect to NLWLS estimators is given by

$$ARE = MA\ D_1/\ MAD_2 \tag{21}$$

Hence, values of this ratio less than 1 are favorable to the NLWLS while values greater than 1 are favorable to the Wilcoxon 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 Parameters

As we mentioned in Section II, $(\hat{\gamma}(h_j))$ 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 [19]. 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 [17]. Let $Z = Z(s_i) : s_i \in \mathbb{R}^d; i = 1, 2, \ldots, n^2$ be a random field in $n \times n$ equally grid lattice generated using the variogram model (3) with true parameters $\boldsymbol{\theta} = (\tau, \sigma^2, \phi)'$. For our case, [20] have developed

the quasi-block-jackknife method to construct the confidence interval for $\boldsymbol{\theta}$, this method can be summarized as:

1) 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

$$C(d_{ij};\boldsymbol{\theta}) = \begin{cases} \tau + \sigma^2 & \text{if } d_{ij} = 0 \\ \tau + \sigma^2 - \gamma(d_{ij};\boldsymbol{\theta}) & \text{if } d_{ij} > 0 \end{cases} \quad (22)$$

where $\gamma(\cdot;\boldsymbol{\theta})$ is defined as in (12), and (d_{ij}) are the entries of the $n^2 \times n^2$ distance matrix D.
2) 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. $U \overset{iid}{\sim} N(0,1)$).
3) 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 $(l^2)(B) = n^2$.
4) 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 $C_j^* = C^* \backslash C_j; j = 1, 2, \ldots, B$.
5) For each $j = 1, 2, \ldots, B$, find the Cholesky decomposition for C_j^* as follows: $C_j^* = L_j^* L_j^{*\prime}$. Then re-correlate U_j^* by calculating $Z_j^* = L_j^* U_j^*$.
6) For a certain number of lags k, and for each Z_j^*, estimate the empirical variogram to get h and $\hat{\gamma}(h)$ from (3), which we denote $\{(h_{i,j}, \hat{\gamma}(h_{i,j})) : i = 1, 2, \ldots, k; j = 1, 2, \ldots, B\}$.
7) For each variogram estimated in step (6), fit the variogram model you used to generate the random field Z, this would be (4) for the NLWLS or (9) for the Wilcoxon, let these estimates $\{\hat{\theta}_j; j = 1, 2, \ldots, B\}$.
8) Calculate the standard error estimator of $\hat{\theta}$:

$$s.e._{Jack}(\hat{\boldsymbol{\theta}}) = \left[\left(\frac{B-1}{B} \right) \sum_{j=1}^{B} \left(\hat{\boldsymbol{\theta}}_j - \bar{\hat{\boldsymbol{\theta}}} \right) \left(\hat{\boldsymbol{\theta}}_j - \bar{\hat{\boldsymbol{\theta}}} \right)' \right]^{1/2} \quad (23)$$

where $\bar{\hat{\boldsymbol{\theta}}} = \sum_{j=1}^{B} \hat{\boldsymbol{\theta}}_j$
9) Finally, construct the $(1-\alpha)\%$ confidence interval for $\boldsymbol{\theta}$ as:

$$\hat{\boldsymbol{\theta}} \pm z_{\alpha/2}\, s.e._{Jack}(\hat{\boldsymbol{\theta}}) \quad (24)$$

There are some necessary conditions 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 [17]. 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 [21] has 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 forms of A for the spherical model (3) was provided by [22] as: $A_{Spherical} = \pi\phi^2/5$. 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 $|V| > \pi\phi^3/5$. 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 Wilcoxon Estimator

We now consider a numeric example in order to demonstrate the robustness properties of the Wilcoxon estimator. We will analyze the Jura, Pb dataset. The Jura data were collected by the Swiss Federal Institute of Technology at Lausanne. For more details see [23], [24]. Data were recorded at $n = 359$ scattered locations with the concentration of seven heavy metals (Cd, Cu, Pb, Co, Cr, Ni, and Zn) in the topsoil measured at each location. Here we will focus on the results as Pb. The data are highly right skewed, [23] have suggested a $\log_{10}$ transformation to reduce this skewness and stabilize its variance. Fig. 1 shows the locations of these measurements and the relative variation in the $\log_{10}(Pb)$ values.

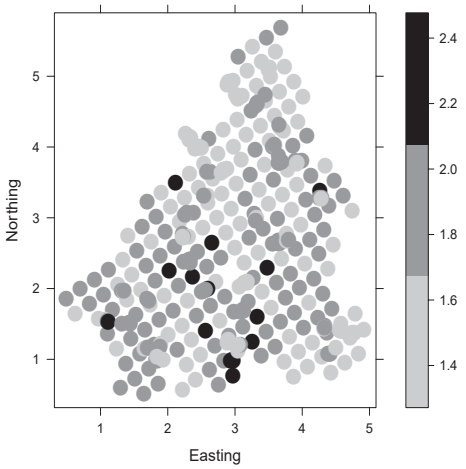

Fig. 1 Sampling locations of the Jura Pb data (Distances measured in km)

Exploratory data analysis indicates that within these reoriented $n = 359$ observations there is no trend in the east-west direction (90 degrees). A summary of the data indicates the maximum separation distance is 5.6199 km. Also, an outlier examination deducted an outlier at site 30 of grid locations $(3.482, 2.295)$ with 2.4771 as $\log_{10}(Pb)$ value. A bubble plot presented in Fig. 2, which indicates this outlier and its spatial location, and the variogram cloud ($Matheron$ estimate) is shown in Fig. 3. In this case, the outliers influence can be considered by any point above the dashed line at distances 1.5816 and 1.5263, and $\gamma(1.5816) = 0.7269$ and $\gamma(1.5263) = 0.7191$ respectively which are at site 30.

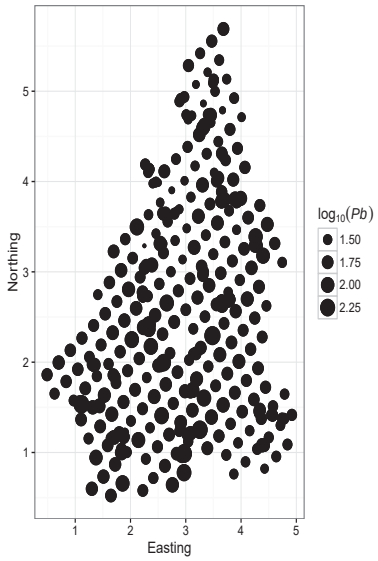

Fig. 2 Bubbleplot of the Jura Pb data (Distances are measured in km)

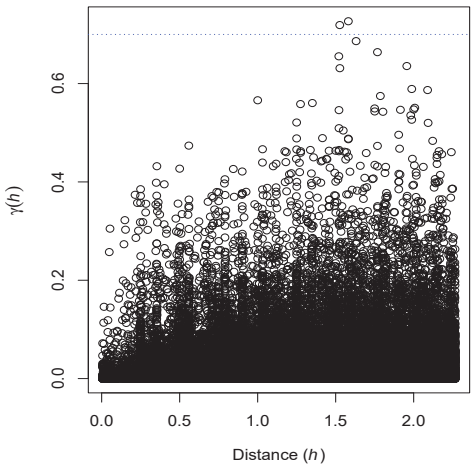

Fig. 3 Variogram cloud of the Jura Pb data

For the estimation step the number of lags used were 11 with at least 30 points in each, for both the *Matheron* and *Cressie* variogram estimates. As presented by [23], the data has an apparent spherical variogram (3).

A plot of $\gamma(h)$ versus h, shown in Fig. 4, shows the NLWLS fit and Wilcoxon fit (with and without outlier) using the *Cressie* estimate.

In Table I, we display the NLWLS and WILX fits (with the outlier) for the *Matheron* estimate to the Jura Pb data. Along with these estimates, we show the approximate standard errors of the estimates $s.e.(\hat{\boldsymbol{\theta}})$ as discussed in Sections II-A and III-B. In the bottom half of Table I (without outlier), location $(3.482, 2.295)$ was removed and the variogram was re-estimated and refit under the same spherical model conditions (3) without any discernible effect in the apparent lack of trend in the east-west direction (90 degrees).

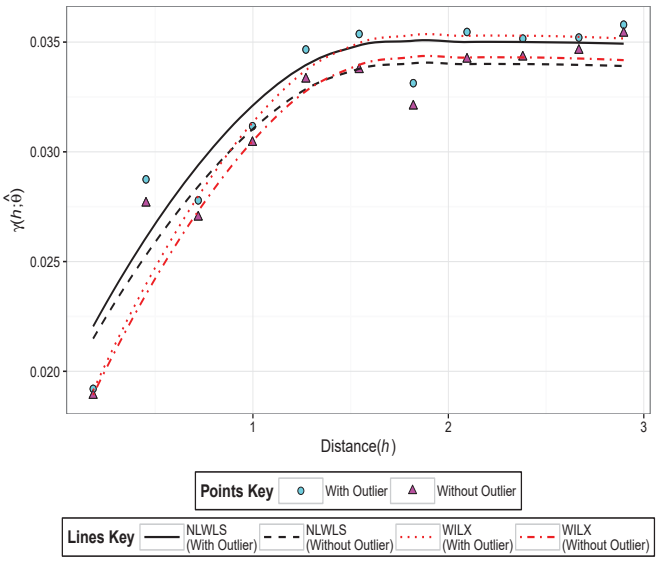

Fig. 4 NLWLS and Wilcoxon fits with and without outlier (*Cressie* Estimate)

TABLE I
JURA PB DATA: SPHERICAL VARIOGRAM PARAMETER ESTIMATES AND APPROXIMATE STANDARD ERRORS WITH AND WITHOUT OUTLIERS (*Matheron* ESTIMATES)

With Outliers			
Model	Nugget(τ)	Sill(σ^2)	Range(ϕ)
NLWLS	0.0247	0.0144	1.7410
Approx. *s.e.*	0.0017	0.0017	0.2219
WILX	0.0224	0.0165	1.7014
Approx. *s.e.*	**0.0013**	**0.0013**	**0.1665**
Without Outliers			
Model	Nugget(τ)	Sill(σ^2)	Range(ϕ)
NLWLS	0.0244	0.0129	1.9872
Approx. *s.e.*	0.0013	0.0013	0.2332
WILX	0.0221	0.0151	1.8084
Approx. *s.e.*	**0.0010**	**0.0011**	**0.1599**

In a similar fashion, Table II, we display the NLWLS and WILX fits for the *Cressie* estimate to the Jura Pb data. Along with the estimates, we show the approximate standard errors of the estimates $s.e.(\hat{\boldsymbol{\theta}})$ as discussed in Sections II-A and III-B. In the the bottom half of Table I (without outlier), location (3.482, 2.295) was removed and the variogram was re-estimated and refit under the same spherical model conditions (3) without any discernible effect in the apparent lack of trend in the east-west direction (90 degrees).

By examining the results in Tables I and II, we can see that:

1) Using the non-robust *Matheron* estimates, with regards to the standard error of the estimates, the WILX procedure outperformed (smallest *s.e.* in **bold**) the NLWLS estimate with the outlier and smaller standard errors without the outlier.
2) Using the robust *Cressie* estimates, with regards to the standard error of the estimates, the WILX outperformed (smallest *s.e.* in **bold**) the NLWLS estimate with or without the outlier.

TABLE II
JURA PB DATA: SPHERICAL VARIOGRAM PARAMETER ESTIMATES AND APPROXIMATE STANDARD ERRORS WITH AND WITHOUT OUTLIERS (*Cressie* ESTIMATE)

With Outliers			
Model	Nugget(τ)	Sill(σ^2)	Range(ϕ)
NLWLS	0.0195	0.0154	1.5906
Approx. *s.e.*	0.0023	0.0023	0.2581
WILX	0.0161	0.0191	1.6451
Approx. *s.e.*	**0.0012**	**0.0012**	**0.1275**
Without Outliers			
Model	Nugget(τ)	Sill(σ^2)	Range(ϕ)
NLWLS	0.0191	0.0148	1.6258
Approx. *s.e.*	0.0021	0.0021	0.2538
WILX	0.0161	0.0182	1.6538
Approx. *s.e.*	**0.0013**	**0.0013**	**0.1477**

VI. SIMULATION STUDIES

We now present the results for some simulations that investigate the efficiency of the WILX procedure compared to 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 again used the spherical model (3) of the Jura Pb example, and we set the true values of the parameters $\boldsymbol{\theta}$ as follows: $\boldsymbol{\theta} = (\tau, \sigma^2, \phi) = (0.022, 0.016, 1.700)$, which is close to the fitted coefficients in Table I. The $RFsimulate$ function within the **RandomFields** package [25] using R 3.2.2 programming language 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 $[0, 6] \times [0, 6]$ for this spherical model. These simulations were obtained by using the direct matrix decomposition method. Then, we partitioned the equally grid square 40×40 into $B = 16$ blocks each of size $l = 100$. Next, we chose the standered normal distribution $N(\mu = 0, \sigma = 1)$, with contamination levels $\epsilon = 0\%, \epsilon = 5\%, \epsilon = 10\%$, and $\epsilon = 20\%$. There were an equal number of contaminations per h block (i.e. in each block there are $(\epsilon/16)$ contaminations).

In the estimating step, we used the *Cressie* estimate, and the number of lags was chosen to be 34. As we discussed earlier in this paper, and based on the recommendations of [26], we used at least 30 points in each lag and the maximum lag distance is about half the maximum separation distance.

The fitting step was performed using NLWLS and WILX. It should also be mentioned with regards to the importance of three parameters nugget (τ), sill (σ^2), and range (ϕ) that with respect to the effectiveness on ordinary kriging, the nugget (τ) and sill (σ^2) influence the kriging variances while the range (ϕ) parameter influences the ordinary kriging weights. So in this simulation, the range (ϕ) is regarded as the most important kriging parameter [6]. To evaluate the efficiency of these estimates, 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 III displays the results of the $AREs$ for the *Cressie* estimates comparing NLWLS to WILX over the four contamination levels. Again, $AREs > 1$ are favorable to WILX.

TABLE III
EMPIRICAL $AREs$ FOR THE WILCOXON ESTIMATES OF THE PARAMETERS τ, σ^2, AND ϕ RELATIVE TO NLWLS (THE *Cressie* ESTIMATE)

		Contamination Level			
	θ	$\epsilon = 0\%$	$\epsilon = 5\%$	$\epsilon = 10\%$	$\epsilon = 20\%$
NLWLS,WILX	Nugget (τ)	1.107	1.006	1.000	0.975
NLWLS,WILX	Sill (σ^2)	1.033	1.030	1.018	1.320
NLWLS,WILX	Range (ϕ)	1.059	1.071	1.071	1.320

Table III indicates that at all levels of contamination, including **no** contamination, the WILX estimates are more efficient than the NLWLS estimates. The only exception being the nugget at 20% contamination.

For our validity study, the quasi-block-jackknife method discussed in section IV-C were applied to constructed the empirical confidence coefficients of confidence intervals for the parameters nugget (τ), sill (σ^2) and the range (ϕ). Table IV 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 IV
EMPIRICAL CONFIDENCE COEFFICIENT FOR THE NLWLS AND WILCOXON ESTIMATES OF THE PARAMETERS τ, σ^2, AND ϕ: THE NOMINAL CONFIDENCE IS 0.95 (THE *Cressie* ESTIMATE)

		Contamination Level			
	θ	$\epsilon = 0\%$	$\epsilon = 5\%$	$\epsilon = 10\%$	$\epsilon = 20\%$
NLWLS	Nugget	0.992	0.196	0.105	0.423
WILX	(τ)	0.994	0.123	0.014	0.001
NLWLS	Sill	0.975	0.999	0.999	0.991
WILX	(σ^2)	0.974	0.999	0.999	0.998
NLWLS	Range	0.975	0.990	0.997	0.996
WILX	(ϕ)	0.978	0.990	0.998	0.997

Table IV shows for the sill (σ^2) and range (ϕ), the NLWLS and WILX intervals are conservative.

VII. SUMMARY

In this paper, we have presented a robust alternative for fitting the empirical variogram. The geometry of the non-linear unweighted Wilcoxon is similar to that of the least squares estimator in that another norm (Wilcoxon) is substituted for the least square (Euclidean) norm. Further research is necessary to develop and evaluate a weighted non-linear Wilcoxon procedure. In the analysis of the Jura Pb data, it was shown that when the outlier was eliminated, weighted least squares suggests a significant change in the variogram model. The unweighted non-linear Wilcoxon, in conjunction with the robust (*Cressie*) variogram estimator, is not affected in this way. In several simulation studies, we showed that the unweighted Wilcoxon estimators became more efficient than the weighted Least Squares estimators with ranging from 1.8% to 32% for the sill and range. This paper demonstrates that when contaminated spatial data are an issue, as they often are, robust procedures for both the estimation and fitting stages of variogram creation are essential for proper modeling.

VIII. Conclusion

From this research, we have concluded the following:

1) Both a robust variogram estimator and robust fit are necessary to ensure the estimates have the lowest standard errors (highest efficiencies).
2) The robust variogram estimation, along with the robust fit, are more efficient than NLWLS even at 0% contamination. This is due in part to the overall instability of fitting a 3-parameter non-linear model.
3) Since un-weighted WILX has outperformed NLWLS, we would expect weighted WILX to show even better results.

References

[1] F. R. Hampel, "Robust estimation: A condensed partial survey," *Zeitschrift fr Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 27, no. 2, pp. 87–104, 1998.
[2] P. J. Huber, *Robust Statistical Procedures*, 2nd ed. Philadelphia: Society for Industrial and Applied Mathematics, 1996.
[3] G. Matheron, *Trait de gostatistique applique*, ser. Mémoires du Bureau de Recherches Géologiques et Minières. Éditions Technip, 1963, no. v. 2.
[4] N. Cressie and D. M. Hawkins, "Robust estimation of the variogram: I," *Mathematical Geology*, vol. 12, no. 2, pp. 115–125, 1980.
[5] N. Cressie, *Statistics for Spatial Data*. New York: John Wiley and Sons, Ltd, 1991.
[6] M. G. Genton, "Highly robust variogram estimation," *Mathematical Geology*, vol. 30, no. 2, pp. 213–221, 1998.
[7] K. V. Mardia and R. J. Marshall, "Maximum likelihood estimation of models for residual covariance in spatial regression," *Biometrika*, vol. 71, no. 1, pp. 135–146, 1984.
[8] H. Patterson and R. Thompson, "Recovery of inter-block information when block sizes are unequal," *Biometrika*, vol. 58, no. 3, pp. 545–554, 1971.
[9] ——, "Maximum likelihood estimation of components of variance." Constanta, Romania: $8th$ International Biometrics Conference, 1974, pp. 197–207.
[10] C. R. Rao, "Minqe theory and its relation to ml and mml estimation of variance components," *Sankhy: The Indian Journal of Statistics, Series B (1960-2002)*, vol. 41, no. 3/4, pp. 138–153, 1979.
[11] N. Cressie, "Fitting variogram models by weighted least squares," *Journal of the International Association for Mathematical Geology*, vol. 17, no. 5, pp. 563–586, 1985.
[12] T. P. Hettmansperger and J. W. McKean, *Robust Nonparametric Statistical Methods*, 2nd ed. New York: Chapman-Hall, 2011.
[13] J. W. McKean and J. D. Kloke, "Efficient and adaptive rank-based fits for linear models with skew-normal errors," *Journal of Statistical Distributions and Applications*, vol. 1, no. 1, pp. 1–18, 2014.
[14] A. Abebe and J. W. McKean, "Highly efficient nonlinear regression based on the wilcoxon norm," in *D. Umbach (ed.), Festschrift in Honor of Mir Masoom Ali on the Occasion of his Retirement*, Ball State University, Muncie, IN, 2007, pp. 340–357.
[15] H. L. Koul, G. L. Sievers, and J. W. McKean, "An estimator of the scale parameter for the rank analysis of linear models under general score functions," *Scandinavian Journal of Statistics*, vol. 14, no. 2, pp. 131–141, 1987.
[16] J. D. Kloke and J. W. McKean, "Rfit: Rank-based estimation for linear models," *The R Journal*, vol. 4, no. 2, pp. 57–64, 2012.
[17] R. G. Clark and S. Allingham, "Robust resampling confidence intervals for empirical variograms," *Mathematical Geosciences*, vol. 43, no. 2, pp. 243–259, 2011.
[18] R Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2015, r package version 3.2.2. (Online). Available: https://www.R-project.org/
[19] L. Tanga, W. R. Schucany, W. A. Woodwardb, and R. F. Gunstb, "A parametric spatial bootstrap," Southern Methodist University, Dallas, Texas, Tech. Rep. SMU-TR-337, 2006.
[20] H. M. Al-Mofleh, J. E. Daniels, and J. W. McKean, "Robust variogram fitting using non-linear rank-based estimators," *International Journal of Mathematical, Computational, Physical, Electrical and Computer Engineering*, vol. 10, no. 2, pp. 68 – 77, 2016. (Online). Available: http://waset.org/Publications?p=110
[21] J.-P. Chiles and P. Delfiner, *Geostatistics: Modelling Spatial Uncertainty*. New York: John Wiley and Sons, Ltd, 1999.
[22] S. Garrigues, D. Allard, F. Baret, and M. Weiss, "Quantifying spatial heterogeneity at the landscape scale using variogram models," *Remote Sensing of Environment*, vol. 103, no. 1, pp. 81–96, 2006.
[23] O. Atteia, J.-P. Dubois, and R. Webster, "Geostatistical analysis of soil contamination in the swiss jura," *Environmental Pollution*, vol. 86, no. 3, pp. 315–327, 1994.
[24] P. Goovaerts, *Geostatistics for Natural Resources Evaluation*. New York: Oxford University Press, 1997.
[25] M. Schlather, A. Malinowski, P. J. Menck, M. Oesting, and K. Strokorb, "Analysis, simulation and prediction of multivariate random fields with package RandomFields," *Journal of Statistical Software*, vol. 63, no. 8, pp. 1–25, 2015. (Online). Available: http://www.jstatsoft.org/v63/i08/
[26] A. G. Journel and C. J. Huijbregts, *Mining Geostatistics*, reprint ed. New York: The Blackburn Press, 2003.