SURFACE RECONSTRUCTION VIA TOTAL LEAST-SQUARES ADJUSTMENT OF THE SEMI-VARIOGRAM

Burkhard Schaffrin [a, *], Sibel Uzun [a]

[a] School of Earth Sciences, The Ohio State University, Columbus, OH 43210-1398, (schaffrin.1,uzun.1)@osu.edu,

Commission III, WG III/2

KEY WORDS: Surface reconstruction, Kriging, Semi-variogram, TLS

ABSTRACT:

One well established technique to construct a surface that best fits to an observed scattered point cloud is based on the Kriging methodology that uses semi-variograms. It somewhat resembles least-squares collocation which, however, uses the covariance function to define spatial coherency. This approach requires the assumption of ergodicity for the underlying spatial process that ultimately determines the surface in need for reconstruction by pointwise spatial prediction and, thereby, avoids the restrictions that come with a parametric surface description. An essential part of the Kriging approach, though, is the estimation of the empirical semi-variogram which is usually found by employing a weighted least-squares technique to best fit a number of representative values, derived from the data set, that describe the average loss of similarity between surface heights over growing distances. As this semi-variogram regularly turns out to have a steep slope near the origin - where it matters most -, a better idea seems to be seeking a best fit on the basis of the Total Least-Squares (TLS) principle. This approach does guarantee that any measures of misfit are taken perpendicular to the adjusted curve rather then in the vertical direction.

In the present contribution, an attempt will be made to quantify the improvement, due to the TLS adjustment, over the traditional weighted least-squares fit. An exemplary set of aeromagnetic data from West Antarctica will serve as a realistic application case for this novel approach to surface reconstruction.

INTRODUCTION

In most datasets from the earth sciences we can recognize *spatial dependence* (resp. correlation) in the sense that two data points are more likely to have similar values the closer they are to each other. This behavior is usually expressed mathematically in form of a *spatial coherency function* such as the semi-variogram, the covariance function, or the homeogram. In this contribution we shall restrict ourselves to the *semi-variogram* which is the traditional tool in geostatistical analysis; see, e.g., *G.Matheron (1971)*, *A.G.Journel and Ch.J.Huijbregts (1978)*, *R.Dutter (1985)*, *N.Cressie (1991)*, *P.K.Kitanidis (1997)*, or *J.P.Chilès and P.Delfiner (1999)*.

Proper spatial dependence analysis is a vital step towards any geostatistical approach to surface reconstruction since it provides initial information about the likely surface behavior in space and, moreover, a mathematical description of the quantities needed for Kriging interpolation. In general, semi-variograms measure the spatial dissimilarity of the surface value at different locations, including the data points. In a planar area, the semi-variogram would only depend on the distance vector between the two points involved if the so-called "*intrinsic hypothesis*" holds true which is actually a similar, but slightly weaker property than "*second order stationarity*"; cf. *A.G.Journel and Ch.J.Huijbregts (1978, pp. 32-34)* for a short discussion. If the spatial process that describes the surface, in addition, follows the requirement of "*isotropy*" (i.e., rotational invariance) the semi-variogram will only depend on one scalar parameter, namely the distance itself. After some prepocessing, including trend removal and/or "median polish" according to *N.Cressie (1991, pp. 46-48)*, the resulting incremental process may oftentimes be assumed to be, at least, second order stationary (if not isotropic, too).

In order to estimate the semi-variogram from the collected data, which typically represent observations of only one realization of the spatial process, we also need to assume the property of "*ergodicity*" that allows the full description of the process by analyzing any one of its realizations. Note that ergodic processes will necessarily turn out to be stationary; thus, non-stationary processes cannot be handled by using methods that require ergodicity for their application.

As a result of this diffuse situation, no universal technique to derive the semi-variogram from a given dataset has been firmly established. Those commonly recommended start from the "*classical*" estimator of *G.Matheron (1963)* that is based on the "method of moments". Since these "empirical" function values will (generally) not obey the characteristics of a semi-variogram, namely to be "conditionally positive-definite", a certain model from an *admissible* class must be fitted in a second step, *with special consideration of the behavior near the origin*.

The quality of fit will usually be judged in terms of a discrete "measure of deviation", for instance the (standard) l^2-norm as applied by *M.David (1977)* among others, taken along the ordinates. *N.Cressie (1985)* proposed a related (suitably) weighted least-squares criterion which contrasts to the

* Burkhard Schaffrin, School of Earth Sciences, The Ohio State University, Columbus, OH 43210-1398, Schaffrin.1@osu.edu

MInimum **N**orm **Qu**adratic **E**stimation (MINQUE) by *M.L.Stein (1987)*, or the **B**est (**I**nvariant) **Q**uadratic **U**nbiased **E**stimation (BIQUE) by *P.K.Kitanidis (1985)*, resp. *R.J.Marshall and K.V.Mardia (1985)*. In a large Monte Carlo study, *D.L.Zimmerman and M.B.Zimmerman (1991)* eventually concluded that, with respect to a number of well-established criteria, the **W**eighted **L**east-**S**quares (WLS) approach yields sometimes the best, but never poor results.

In our view, WLS could be improved indeed by modifying the distance criterion to not only reflect the deviations in the ordinates, *but also in the abscissas* which result from an averaging process, too. In fact, the deviation will be measured perpendicular to the semi-variogram graph when using the criterion of ***T**otal **L**east-**S**quares (TLS)* as it was first discussed by *G.H.Golub and C.F.van Loan (1980)* in a different context. By applying this modified criterion in two stages, Felus and Schaffrin (2005) could already see a better approximation of the "classical" estimator near the origin while generally preserving the "good nature" of the WLS by *N.Cressie (1985)*.

In the following, we shall first introduce the notation and provide a brief overview of the mathematical setting before introducing the Total Least-Squares approach in one stage and its adaptation to the spatial context, namely the fitting of a semi-variogram model based on the "empirical" values. A MATLAB program has been developed for this purpose and tested on an aeromagnetic dataset from Antarctica, using two different models for their semi-variogram. We conclude with a discussion of the merits that are specific to either the WLS, or the TLS procedure in view of our experience so far.

1. Spatial Process and Their Prediction: A Review

The semi-variogram offers one way to summarize *spatial coherency* among the process values within its domain of definition. More formally, let us introduce the real-valued stochastic process *x(s)*, that describes the surface, by

$$x: S \to \mathbb{R} \quad \text{for} \quad s \in S \subset \mathbb{R}^d \tag{1.1}$$

where *d* denotes the Euclidean dimension. For the elements of the domain *S*, let the process values *x(s)* be random numbers with *constant expectation* (or "mean")

$$E\{x(s)\} = \beta \quad \text{for all} \quad s \in S\,, \tag{1.2}$$

such that the *(auto)-covariance*

$$C\{x(s), x(s+h)\} = E\{x(s)\cdot x(s+h)\} - \beta^2 =: C_x(h) \tag{1.3}$$

only depends on the distance vector *h*. In this case we call the process *x(s)* "*second order stationary*"; moreover, the *variogram*

$$\begin{aligned} D\{x(s+h)-x(s)\} &= E\left\{\left[x(s+h)-x(s)\right]^2\right\} = \\ &= 2\left[C_x(0)-C_x(h)\right] =: 2\gamma_x(h) \end{aligned} \tag{1.4}$$

will also depend on the distance vector *h* only. Here *D* denotes "dispersion" (or "variance"). With (1.2) and (1.4) the process *x(s)* fulfills the "*intrinsic property*".
In addition, we would assume *x(s)* to be "*ergodic*" which allows us to replace the expectation operator by a spatial integral over a realization of the process, here also denoted by *x(s)*. Hence, the mean from (1.2) can be rewritten

$$E\{x(s)\} = \frac{1}{\|S\|}\int_S x(s)\,ds = \beta \tag{1.5}$$

and the semi-variogram from (1.4) as

$$\gamma_x(h) = \frac{1}{2\|S_h\|}\int_{S_h}\left[x(s+h)-x(s)\right]^2 ds \tag{1.6}$$

where S_h is that portion of the domain *S* that guarantees *s*+*h* to be still in *S* for a given *h* if and only if $s \in S_h$.

We emphasize the fact that "ergodicity" implies "second order stationarity" and consequently the "intrinsic property". If either of these latter characteristics does not hold true, the use of any identities such as (1.5) or (1.6) is *not advised*, due to the lack of "ergodicity".

In this study, we shall restrict ourselves to the dimension *d*=2, i.e. to true surfaces defined over a *planar* domain, and shall assume "isotropy" as additional property. In this case, both covariance function and semi-variogram will only depend on the distance

$$\|h\| := \sqrt{(X_1-X_2)^2+(Y_1-Y_2)^2} \tag{1.7}$$

between two locations $s:(X_1,Y_1)$ and $s+h:(X_2,Y_2)$, namely:

$$C\{x(s)\,x(s+h)\} = C_x(\|h\|) \tag{1.8}$$

$$D\{x(s+h)-x(s)\}/2 = \gamma_x(\|h\|) \tag{1.9}$$

If we, furthermore, consider the "*homeogram*"

$$E\{x(s)\cdot x(s+h)\} =: \eta_x(\|h\|) = C_x(\|h\|) + \beta^2 \tag{1.10}$$

we may readily establish the identities

$$\gamma_x(\|h\|) = C_x(0) - C_x(\|h\|) = \eta_x(0) - \eta_x(\|h\|), \tag{1.11}$$

provided the afore-mentioned quantities exist. Obviously, covariance function and homeogram need to be "*positive-definite*" functions whereas the variogram may only be "*conditionally positive-definite*"; see *A.G.Journel and Ch.J.Huijbregts (1978, p.35)* for more details.

In order to evaluate formulas such as (1.5) and (1.6), a realization of the process *x(s)* ought to be available on the entire domain *S*. In practice, this will be an unrealistic requirement; instead, the process will be "sampled" (i.e., observed) at a number of (mostly) discretely distributed locations, say $s_j\,(j=1,\cdots,n)$, leading to a "digitized surface" with the $n\times 1$ vector

$$\boldsymbol{y} = \left[y_1,\cdots,y_n\right]^T := \left[x(s_1)+e_1,\cdots,x(s_n)+e_n\right]^T \tag{1.12}$$

where e_j ($j=1,...,n$) denotes the additive random observation noise associated with the respective sample point S_j. As usual, we assume

$$E\{e_j\}=0\,,\qquad D\{e_j\}=\sigma_e^2\,,\qquad \text{for all } j\in\{1,\cdots,n\}\,, \tag{1.13a}$$

$$C\{e_j,e_k\}=0\,,\qquad \text{if } j\neq k\,, \tag{1.13b}$$

$$C\{e_j,x(s)\}=0\,,\qquad \text{for all if } j\in\{1,\cdots,n\} \text{ and } s\in S\,, \tag{1.13c}$$

and can thus conclude

$$E\{y\}=\underset{n\times 1}{\boldsymbol{\tau}}\cdot\beta\,,\quad D\{y\}=:\underset{n\times n}{\boldsymbol{K}}=:\underset{n\times n}{\boldsymbol{C}}+\sigma_e^2\boldsymbol{I}_{\boldsymbol{n}}\,. \tag{1.14a}$$

Here $\boldsymbol{\tau}:=[1,\cdots,1]^T$ denotes the $n\times 1$ "*summation vector*", and the typical element in the $n\times n$ covariance matrix $\boldsymbol{K}$ is defined by

$$K_{jk}:=C\{y_j,y_k\}=C_x(\|s_j-s_k\|)+\sigma_e^2\cdot\delta_{jk} \tag{1.14b}$$

with δ_{jk} as "Kronecker delta". Moreover, we may define the $n\times 1$ vector

$$\begin{aligned}\boldsymbol{\kappa}(s)&:=\left[C\{x(s),y_1\},\cdots,C\{x(s),y_n\}\right]^T=\\&=\left[C\{x(s),x(s_1)\},\cdots,C\{x(s),x(s_n)\}\right]^T=\\&=\left[C_x(\|s_1-s\|),\cdots,C_x(\|s_n-s\|)\right]^T\end{aligned} \tag{1.14c}$$

in order to obtain the *Simple Kriging (SK)* prediction

$$\bar{x}(s)=\beta+\boldsymbol{\kappa}(s)^T\cdot\boldsymbol{K}^{-1}\cdot(\boldsymbol{y}-\boldsymbol{\tau}\beta) \tag{1.15a}$$

with its Mean Squared Prediction Error

$$\boldsymbol{MSPE}\{\bar{x}(s)\}=\sigma_x^2-\boldsymbol{\kappa}(s)^T\cdot\boldsymbol{K}^{-1}\cdot\boldsymbol{\kappa}(s)\,,\quad \sigma_x^2:=C_x(0), \tag{1.15b}$$

in accordance with *G.Matheron (1971)*.

Obviously, the use of (1.15a) requires the process mean β to be *known* which, most of the time, is unrealistic to assume. In those situations, the *Ordinary Kriging (OK)* prediction is preferred, namely

$$\tilde{x}(s)=\hat{\beta}+\boldsymbol{\kappa}(s)^T\cdot\boldsymbol{K}^{-1}\cdot(\boldsymbol{y}-\boldsymbol{\tau}\cdot\hat{\beta}) \tag{1.16a}$$

with the *estimated* process mean

$$\hat{\beta}=(\boldsymbol{\tau}^T\boldsymbol{K}^{-1}\boldsymbol{y})/(\boldsymbol{\tau}^T\boldsymbol{K}^{-1}\boldsymbol{\tau}), \tag{1.16b}$$

and (slightly) deteriorated Mean Squared Prediction Error

$$\begin{aligned}&\boldsymbol{MSPE}\{\tilde{x}(s)\}-\boldsymbol{MSPE}\{\bar{x}(s)\}\\&=[1-\boldsymbol{\tau}^T\boldsymbol{K}^{-1}\boldsymbol{\kappa}(s)]^2/(\boldsymbol{\tau}^T\boldsymbol{K}^{-1}\boldsymbol{\tau})\geq 0;\end{aligned} \tag{1.16c}$$

see, e.g., *N.Cressie (1991, p.123)*.

Both Kriging variants, SK as well as OK, can also be represented in terms of the *semi-variogram* (or the homeogram for that matter) instead of the covariance function; see, e.g., *B.Schaffrin (2001)*. In case of the latter (OK), we define the slightly altered $n\times n$ matrix Γ by its typical element

$$\begin{aligned}\Gamma_{jk}&:=D\{y_j-y_k\}/2=\gamma_x(\|x_j-x_k\|)+\sigma_e^2\cdot(1-\delta_{jk})\\&=(\sigma_x^2+\sigma_e^2)-K_{jk}\end{aligned} \tag{1.17a}$$

and the (similarly altered) $n\times 1$ vector, at an arbitrary location $s\in S$, by

$$\begin{aligned}\boldsymbol{\gamma}(s)&:=\left[\gamma_x(\|s_1-s\|),\cdots,\gamma_x(\|s_n-s\|)\right]^T+\sigma_e^2\cdot\boldsymbol{\tau}\\&=(\sigma_x^2+\sigma_e^2)\cdot\boldsymbol{\tau}-\boldsymbol{\kappa}(s)\end{aligned} \quad \text{if } s\neq s_j\,, \tag{1.17b}$$

$$\begin{aligned}\boldsymbol{\gamma}^*(s_j)&:=\boldsymbol{\gamma}(s_j)+\sigma_e^2\cdot\boldsymbol{\eta}_j\\&=(\sigma_x^2+\sigma_e^2)\cdot\boldsymbol{\tau}-\boldsymbol{\kappa}(s_j)\end{aligned} \quad \text{if } j\in\{1,\cdots n\}\,; \tag{1.17c}$$

here $\boldsymbol{\eta}_j:=[0,\cdots,1,\cdots,0]^T$ denotes the j^{th} unit vector of size $n\times 1$. Due to these identities and the fact that (1.16a-c) can be derived from the extended system

$$\begin{bmatrix}\boldsymbol{K} & -\boldsymbol{\tau}\\ -\boldsymbol{\tau}^T & 0\end{bmatrix}\begin{bmatrix}\hat{\boldsymbol{\chi}}(s)\\ \hat{v}(s)\end{bmatrix}=\begin{bmatrix}\boldsymbol{\kappa}(s)\\ -1\end{bmatrix} \tag{1.18}$$

$$\hat{v}(s)=\left[1-\boldsymbol{\tau}^T\boldsymbol{K}^{-1}\boldsymbol{\kappa}(s)\right]/(\boldsymbol{\tau}^T\boldsymbol{K}^{-1}\boldsymbol{\tau}) \tag{1.19a}$$

$$\hat{\boldsymbol{\chi}}(s)=\boldsymbol{K}^{-1}\boldsymbol{\kappa}(s)+\boldsymbol{K}^{-1}\boldsymbol{\tau}\left[1-\boldsymbol{\tau}^T\boldsymbol{K}^{-1}\boldsymbol{\kappa}(s)\right]/(\boldsymbol{\tau}^T\boldsymbol{K}^{-1}\boldsymbol{\tau}) \tag{1.19b}$$

$$\tilde{x}(s)=\hat{\boldsymbol{\chi}}(s)^T\cdot\boldsymbol{y} \tag{1.19c}$$

$$\boldsymbol{MSPE}\{\tilde{x}(s)\}=\sigma_x^2-\hat{\boldsymbol{\chi}}(s)^T\cdot\boldsymbol{\kappa}(s)+\hat{v}(s) \tag{1.19d}$$

we are able to establish the equivalency of (1.18) with the equations

$$\begin{aligned}&(\sigma_x^2+\sigma_e^2)\cdot\boldsymbol{\tau}-\boldsymbol{\gamma}(s)=\boldsymbol{\kappa}(s)=\boldsymbol{K}\cdot\hat{\boldsymbol{\chi}}(s)-\boldsymbol{\tau}\cdot\hat{v}(s)=\\&=(\sigma_x^2+\sigma_e^2)\boldsymbol{\tau}\cdot\left[\boldsymbol{\tau}^T\hat{\boldsymbol{\chi}}(s)\right]-\boldsymbol{\Gamma}\cdot\hat{\boldsymbol{\chi}}(s)-\boldsymbol{\tau}\cdot\hat{v}(s)=\\&=(\sigma_x^2+\sigma_e^2)\boldsymbol{\tau}-\boldsymbol{\Gamma}\cdot\hat{\boldsymbol{\chi}}(s)-\boldsymbol{\tau}\cdot\hat{v}(s),\end{aligned} \tag{1.20a}$$

$$\boldsymbol{\tau}^T\cdot\hat{\boldsymbol{\chi}}(s)=1\,, \tag{1.20b}$$

if $s\neq s_j$ $(j=1,\cdots,n)$ or, alternatively, with the system

$$\begin{bmatrix}\boldsymbol{\Gamma} & \boldsymbol{\tau}\\ \boldsymbol{\tau}^T & 0\end{bmatrix}\begin{bmatrix}\hat{\boldsymbol{\chi}}(s)\\ \hat{v}(s)\end{bmatrix}=\begin{bmatrix}\boldsymbol{\gamma}(s)\\ 1\end{bmatrix}\,, \tag{1.21}$$

which firmly indicates the importance of a *well determined semi-variogram* for the purpose of OK prediction. In the system (1.21), on the right hand side, the vector $\gamma(s)$ is obviously to be replaced by $\boldsymbol{\gamma}^*(s_j)$ following (1.17c), should s coincide with a sample point s_j; see also the respective remarks by *N.Cressie (1991, pp.128-130)*. The system (1.21) now yields the solution

$$\hat{v}(s) = -\left[1 - \boldsymbol{\tau}^T \boldsymbol{\Gamma}^{-1} \boldsymbol{\gamma}(s)\right] / (\boldsymbol{\tau}^T \boldsymbol{\Gamma}^{-1} \boldsymbol{\tau}), \tag{1.22a}$$

$$\hat{\boldsymbol{\chi}}(s) = \boldsymbol{\Gamma}^{-1} \boldsymbol{\gamma}(s) + \boldsymbol{\Gamma}^{-1} \boldsymbol{\tau} \left[1 - \boldsymbol{\tau}^T \boldsymbol{\Gamma}^{-1} \boldsymbol{\gamma}(s)\right] / (\boldsymbol{\tau}^T \boldsymbol{\Gamma}^{-1} \boldsymbol{\tau}), \tag{1.22b}$$

$$\tilde{x}(s) = \hat{\boldsymbol{\chi}}(s)^T \cdot \boldsymbol{y}, \tag{1.22c}$$

$$\boldsymbol{MSPE}\left\{\tilde{x}(s)\right\} = \hat{\boldsymbol{\chi}}(s)^T \cdot \boldsymbol{\gamma}(s) + \hat{v}(s) - \sigma_e^2. \tag{1.22d}$$

$$\begin{bmatrix} \boldsymbol{\Gamma} & \boldsymbol{\tau} \\ \boldsymbol{\tau}^T & (\sigma_x^2 + \sigma_e^2)^{-1} \end{bmatrix} \begin{bmatrix} \bar{\boldsymbol{\chi}}(s) \\ \bar{v}(s) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\gamma}(s) \\ 1 \end{bmatrix} \tag{1.23}$$

if $s \neq s_j$ $(j = 1, \cdots, n)$, via

$$\bar{v}(s) = (\sigma_x^2 + \sigma_e^2)\left[1 - \boldsymbol{\tau}^T \bar{\boldsymbol{\chi}}(s)\right] \tag{1.24a}$$

$$\begin{aligned} \bar{\boldsymbol{\chi}}(s) &= \left[\boldsymbol{\Gamma} - (\sigma_x^2 + \sigma_e^2) \cdot \boldsymbol{\tau}\boldsymbol{\tau}^T\right]^{-1} \left[\boldsymbol{\gamma}(s) - (\sigma_x^2 + \sigma_e^2) \cdot \boldsymbol{\tau}\right] \\ &= \boldsymbol{K}^{-1} \cdot \boldsymbol{\kappa}(s) \end{aligned} \tag{1.24b}$$

$$\bar{x}(s) = \beta + \bar{\boldsymbol{\chi}}(s)^T (\boldsymbol{y} - \boldsymbol{\tau} \cdot \beta), \tag{1.24c}$$

$$\boldsymbol{MSPE}\left\{\bar{x}(s)\right\} = \bar{\boldsymbol{\chi}}(s)^T \cdot \boldsymbol{\gamma}(s) + \bar{v}(s) - \sigma_e^2. \tag{1.24d}$$

Again, on the right hand side of (1.23) as well as in (1.24d), the vector $\gamma(s)$ ought to be replaced by $\gamma^*(s_j)$ it one of the sample points s_j is chosen for the S*imple Kriging prediction using the semi-variogram*. (Note that such a possibility was negated by *J.P.Chilès and P.Delfiner (1999, p.170)*.) Of course, the value for $\sigma_x^2 + \sigma_e^2$ can be taken as the "sill" of the semi-variogram while σ_e^2 represents the "nugget effect", at least in the absence of microscale variation.

2. Empirical Semi-Variogram Values and Suitable Models to be Fitted to

Since we had assumed "ergodicity" and "isotropy" for our spatial process, *empirical values* for the semi-variogram can be obtained from a *discrete* version of (1.6) when applied to the sample data, namely

$$\hat{\gamma}_y\left(\left\|\hat{h}^i\right\|\right) := \sum_{j=1}^{n} \sum_{l=1}^{N(H_j^i)} \left[y(s_j) - y(s_j + h_l^i)\right]^2 / \left[2N(H^i)\right] \tag{2.1a}$$

where $s_j + h_l^i$ denotes any other sample point within a neighborhood H_j^i of s_j, defined by $\|h\|_{i-1} < \|h\| \leq \|h\|_i$ with chosen lower and upper bounds, which has $N(H_j^i)$ elements with respect to this sample point s_j. The *total number* of squared differences is, therefore,

$$N(H^i) := \sum_{j=1}^{n} N(H_j^i) \tag{2.1b}$$

and the central value for this neighborhood would be the *average lag* in all H_j, namely

$$\left\|\hat{h}^i\right\| := \sum_{j=1}^{n} \sum_{l=1}^{N(H_j^i)} \left\|h_l^i\right\| / N(H^i). \tag{2.1c}$$

Except for the origin, the function values of $\gamma_y\left(\|h\|\right)$ are expected to be shifted away from the semi-variogram $\gamma_x\left(\|h\|\right)$ by the noise variance:

$$\gamma_y\left(\|h\|\right) = \gamma_x\left(\|h\|\right) + \sigma_e^2 \qquad \text{if} \quad h \neq 0 \tag{2.2}$$

which was already reflected in the formulas (1.17a-c). The formulas (2.1a-c) represent the "classical" estimate by *G.Matheron (1963)*, based on the method of moments; see, e.g., *N.Cressie (1991, pp. 69-70)* for more details, and *J.P.Chilès and P.Delfiner (1999, pp. 34-57)* for possible modifications.

As a result, we obtain an empirical value $\hat{\gamma}_y\left(\left\|\hat{h}^i\right\|\right)$ for each neighborhood (or interval) H^i $(i = 1, \cdots, I)$ that approximates the respective semi-variogram $\gamma_y\left(\|h\|\right)$ at $\|h\| = \left\|\hat{h}^i\right\|$. These discrete values must be fitted to a continuous *semi-variogram model* that necessarily provides a (conditionally) positive-definite semi-variogram as theoretically required. *N.Cressie (1991, pp. 61-63)* presents six valid isotropic standard models while *J.-P.Chilès and P.Delfiner (1999, pp. 80-93)* discuss even more complex models. In this contribution, we shall restrict ourselves to the following two isotropic standard models:

Model I:"Spherical Semi-Variogram"

$$\gamma_y\left(\|h\|\right) := \begin{cases} 0 & \text{if } \|h\| = 0 \\ \sigma_e^2 + \sigma_x^2\left[(3\|h\|/r) - (\|h\|/r)^3\right]/2 & \text{if } 0 < \|h\| \leq r \\ \sigma_e^2 + \sigma_x^2 & \text{if } \|h\| \geq r \end{cases} \tag{2.3}$$

Model II:"Exponential Semi-Variogram"

$$\gamma_y\left(\|h\|\right) := \begin{cases} 0 & \text{if } \|h\| = 0 \\ \sigma_e^2 + \sigma_x^2\left[1 - \exp(-\|h\|/r)\right] & \text{if } \|h\| > 0 \end{cases} \tag{2.4}$$

Both models depend on *three parameters* that have to be derived from the empirical semi-variogram values by a certain estimation principle. They are

- σ_e^2, the noise variance (or "nugget effect"),
- σ_x^2, the process variance (or $\sigma_e^2 + \sigma_x^2$ as "sill"),
- r, the impact interval (or "range"),

but do not automatically form a linear relation. Hence, we *linearize first* by using approximate values $\overset{\circ}{r}$ and $\overset{\circ}{\sigma}{}_x^2$ that are straight-forwardly derived from an analysis of the empirical semi-variogram values at the far end (i.e., for large lags $\left\|\hat{h}^i\right\|$, along with $\overset{\circ}{\sigma}{}_e^2 := 0$, and obtain:

Linearized Model I:

$$\delta\gamma_y(\|h\|) := \gamma_y(\|h\|) - \overset{\circ}{\sigma}{}_x^2\left[3\|h\|/\overset{\circ}{r} - (\|h\|/\overset{\circ}{r})^3\right]/2 =$$
$$= \begin{cases} \sigma_e^2 + \left[(3\|h\|/\overset{\circ}{r} - (\|h\|/\overset{\circ}{r})^3)/2\right]\cdot\delta\sigma_x^2 - 3(\sigma_x^2/\overset{\circ}{r})\left[(\|h\|/\overset{\circ}{r} - (\|h\|/\overset{\circ}{r})^3)/2\right]\cdot\delta r & \text{if } 0 < \|h\| \le \overset{\circ}{r} \\ \sigma_e^2 + \delta\sigma_x^2 & \text{if } \|h\| \ge \overset{\circ}{r} \end{cases} \quad (2.5)$$

Linearized Model II:

$$\delta\gamma_y(\|h\|) := \gamma_y(\|h\|) - \overset{\circ}{\sigma}{}_x^2\left[1 - \exp\left(-\|h\|/\overset{\circ}{r}\right)\right] =$$
$$= \sigma_e^2 + \left[1 - \exp\left(-\|h\|/\overset{\circ}{r}\right)\right]\cdot\delta\sigma_x^2 - (\overset{\circ}{\sigma}{}_x^2/\overset{\circ}{r})\cdot(\|h\|/\overset{\circ}{r})\cdot\exp(-\|h\|/\overset{\circ}{r})\cdot\delta r \quad \text{if } \|h\| > 0 \quad (2.6)$$

From these linearized models, we can readily obtain estimates for σ_e^2, $\delta\sigma_x^2$ and δr by applying a suitable principle such as "*Weighted Least-Squares (WLS)*" or "*Total Least-Squares (TLS)*". In any case, the procedure has to be repeated with improved approximate values until convergence is reached.

3. Fitting Principle: Weighted Least-Squares vs. Total Least-Squares

Here we discuss methods how the empirical values $\hat{\gamma}_y(\|\hat{h}^i\|)$ from (2.1a-c) for $i = 1,\cdots,I$ can be "best" fitted to one of the admissible semi-variogram models such as (2.3)-(2.4), for instance, or (2.5)-(2.6) after linearization. Obviously, *Weighted Least-Squares (WLS)* according to *N.Cressie (1985)* comes to mind in which estimates for the parameters $\Xi := \left[\sigma_e^2, \sigma_x^2, r\right]^T$ are found from the *objective function*

$$\left[\hat{\boldsymbol{\gamma}}_y(\|\hat{h}^i\|) - \boldsymbol{\gamma}_y(\|\hat{h}^i\|,\Xi)\right]^T \cdot \boldsymbol{W}(\Xi) \cdot \left[\hat{\boldsymbol{\gamma}}_y(\|\hat{h}^i\|) - \boldsymbol{\gamma}_y(\|\hat{h}^i\|,\Xi)\right] = \min_{\Xi}. \quad (3.1)$$

where the brackets denote an $I\times 1$ vector of *increments* between empirical values and the chosen semi-variogram model, and $\boldsymbol{W}(\Xi)$ is an $I\times I$ weight matrix that is still to be defined. Note that this weight matrix itself may generally depend on the parameters in Ξ.

As a reasonable choice, we would try to find the weight matrix $\boldsymbol{W}(\Xi)$ by approximating the inverse covariance matrix of the empirical values, namely

$$\boldsymbol{W}(\Xi) \approx (D\left\{\left[\hat{\gamma}_y(\|\hat{h}^i\|)\right]\right\})^{-1}. \quad (3.2)$$

Since this may prove a very tedious endeavor in itself, let us restrict ourselves to the *diagonal elements* only, which leads to the definition

$$\boldsymbol{W}(\Xi) := Diag\left[(D\left\{\left[\hat{\gamma}_y(\|\hat{h}^i\|)\right]\right\})^{-1}\right] = Diag\left[w_i(\Xi)\right] \quad (3.3)$$

Under *quasi-normal* assumptions, the variance of each empirical value $\hat{\gamma}_y(\|\hat{h}^i\|)$ can be readily obtained by exploiting the fact that twice this value would represent an estimated variance. So, we first get

$$D\left\{2\hat{\gamma}_y(\|\hat{h}^i\|)\right\} = 2\left[\gamma_y(\|\hat{h}^i\|,\Xi)\right]^2 / N(H^i) = 4D\left\{\hat{\gamma}_y\left(\|\hat{h}^i\|\right)\right\} \quad (3.4a)$$

in agreement with *N.Cressie (1991, p.96)* and, furthermore,

$$w_i(\Xi) := (D\left\{\hat{\gamma}_y(\|\hat{h}^i\|)\right\})^{-1} = 2N(H^i)/\left[\gamma_y(\|\hat{h}^i\|,\Xi)\right]^2 \quad (3.4b)$$

for the weights $(i = 1,\cdots,I)$. Consequently, up to the factor 2, the objective function (3.1) is reduced to

$$\sum_{i=1}^{I} N(H^i)\left[\hat{\gamma}_y(\|\hat{h}^i\|)/\gamma_y(\|\hat{h}^i\|,\Xi) - 1\right]^2 = \min_{\Xi}. \quad (3.5)$$

which can be solved by well-established techniques.

We may, however, go one step further and replace the weight matrix in (3.3) by its *empirical counterpart*

$$\hat{\boldsymbol{W}} := Diag\left[(\hat{D}\left\{\hat{\gamma}_y(\|\hat{h}^i\|)\right\})^{-1}\right] = Diag\left[\hat{w}_i\right] \quad (3.6a)$$

with

$$\hat{w}_i := (\hat{D}\left\{\hat{\gamma}_y(\|\hat{h}^i\|)\right\})^{-1} = 2N(H^i)/\left[\hat{\gamma}_y(\|\hat{h}^i\|)\right]^2 \quad (3.6b)$$

for $i = 1,\cdots,I$ which now no longer depends on Ξ. As a result, we arrive at the *new objective function*

$$\sum_{i=1}^{I} N(H^i)\left[1 - \gamma_y(\|\hat{h}^i\|,\Xi)/\hat{\gamma}_y(\|\hat{h}^i\|)\right]^2 \quad (3.7)$$

which, after linearization , reads:

$$\sum_{i=1}^{I} N(H^i)\left[1 - \gamma_y(\|\hat{h}^i\|,\overset{\circ}{\Xi})/\hat{\gamma}_y(\|\hat{h}^i\|) - \boldsymbol{A}_\gamma^i\boldsymbol{\xi}\right]^2 = \min_{\xi}. \quad (3.8)$$

Here, $\boldsymbol{\xi} := \left[\sigma_e^2, \delta\sigma_x^2, \delta r\right]^T$ denotes the 3×1 vector of (incremental) parameters; $\boldsymbol{A}_\gamma$ is a $I\times 3$ matrix with coefficients taken from either of the linearized models (2.5)-(2.6) and then divided by the respective empirical value $\hat{\gamma}_y(\|\hat{h}^i\|)$, with $\boldsymbol{A}_\gamma^i$ as the i$^{\text{th}}$ row of $\boldsymbol{A}_\gamma$ $(i = 1,\cdots,I)$. We may further introduce the $I\times 1$ vector

$$\boldsymbol{y}_\gamma := \left[1 - \gamma_y(\|\hat{h}^i\|,\overset{\circ}{\Xi})/\hat{\gamma}_y(\|\hat{h}^i\|)\right]_{i=1,\cdots,I} \quad (3.9)$$

and the additional weight matrix of size, $I\times I$

$$\boldsymbol{P}_\gamma := Diag\left[N(H^i)\right], \quad (3.10)$$

to give (3.8) the *vectorial form*

$$(\boldsymbol{y}_\gamma - \boldsymbol{A}_\gamma\cdot\boldsymbol{\xi})^T\boldsymbol{P}_\gamma(\boldsymbol{y}_\gamma - \boldsymbol{A}_\gamma\cdot\boldsymbol{\xi}) = \min_{\boldsymbol{\xi}}. \quad (3.11)$$

In our view, (3.11) offers the *most practical* approach to WLS, but still suffers from the fact that the fit, although properly weighted, occurs *only in the direction of the ordinate*. This may not be critical at the far end of the semi-variogram, but it becomes a more serious issue near the origin where we may see a substantial slope. In such a case, the target function (3.11) does not provide an estimated semi-variogram that is "*nearest*" to the empirical values in the *geometric sense*, namely measured along perpendicular projections onto the graph of the semi-variogram.

Such an estimate can, however, be obtained by the principle of "*Total Least-Squares (TLS)*" as defined in the following, using *equal weighting* for both ordinate and abscissa. Note that, in the future, we shall try to allow different weights for the abscissa values $\|\hat{h}^i\|$ in accordance with the averaging (2.1c). For now, let us introduce

$$\bar{\boldsymbol{y}}_\gamma := \boldsymbol{P}_\gamma^{1/2}\boldsymbol{y}_\gamma, \quad \bar{\boldsymbol{A}}_\gamma := \boldsymbol{P}_\gamma^{1/2}\boldsymbol{A}_\gamma, \quad \boldsymbol{P}_\gamma := \boldsymbol{P}_\gamma^{1/2}\cdot\boldsymbol{P}_\gamma^{1/2} \tag{3.12}$$

to form the *Linear Model*

$$(\bar{\boldsymbol{A}}_\gamma - \bar{\boldsymbol{E}}_\gamma)\boldsymbol{\xi} - (\bar{\boldsymbol{y}}_\gamma - \bar{\boldsymbol{e}}_\gamma) = 0 \tag{3.13a}$$

$$E\left\{\left[\bar{E}_\gamma, \bar{e}_\gamma\right]\right\} = 0, \quad D\left\{vec\left[\bar{E}_\gamma, \bar{e}_\gamma\right]\right\} = \Sigma_o \otimes I_I \tag{3.13b}$$

where $\Sigma_o := \sigma_o^2 \boldsymbol{I}_4$ has size 4×4 (with unknown σ_o^2), $\boldsymbol{I}_I$ is the $I\times I$ identity matrix, and $\otimes$ denotes the "*Kronecker-Zehfuss product*" of matrices, defined by

$$\boldsymbol{M}\otimes\boldsymbol{N} := \left[m_{ij}\cdot\boldsymbol{N}\right] \quad \text{for} \quad \boldsymbol{M} = \left[m_{ij}\right]; \tag{3.14}$$

the "vec" operator stacks one column of a matrix under the other, moving from left to right.

It is easy to see that WLS is based on the minimization of $\bar{\boldsymbol{e}}_\gamma^T\bar{\boldsymbol{e}}_\gamma$ under the constraint $\bar{\boldsymbol{E}}_\gamma = \boldsymbol{0}$. In contrast, the equally weighted "*Total Least-Squares*" principle starts from the objective function

$$\bar{\boldsymbol{e}}_\gamma^T\bar{\boldsymbol{e}}_\gamma + (vec\bar{\boldsymbol{E}}_\gamma)^T vec\bar{\boldsymbol{E}}_\gamma = \min_\xi. \tag{3.15}$$

The solution to the TLS problem was originally provided by *G.H.Golub and C.F.van Loan (1980)*; see also *G.H.Golub and C.F.van Loan (1983, pp.420-425)* for a brief overview, or *S.van Huffel and J.Vandewalle (1991)* for many details.

In essence, the identity (3.13a) of size $I\times 1$ is equivalently transformed into one of size 4×1 via

$$\begin{bmatrix}\bar{\boldsymbol{A}}_\gamma^T - \bar{\boldsymbol{E}}_\gamma^T\\ \bar{\boldsymbol{y}}_\gamma^T - \bar{\boldsymbol{e}}_\gamma^T\end{bmatrix}\left[\bar{\boldsymbol{A}}_\gamma - \bar{\boldsymbol{E}}_\gamma, \bar{\boldsymbol{y}}_\gamma - \bar{\boldsymbol{e}}_\gamma\right]\begin{bmatrix}\boldsymbol{\xi}\\ -1\end{bmatrix} = \left(\begin{bmatrix}\bar{\boldsymbol{A}}_\gamma^T\bar{\boldsymbol{A}}_\gamma & \bar{\boldsymbol{A}}_\gamma^T\bar{\boldsymbol{y}}_\gamma\\ \bar{\boldsymbol{y}}_\gamma^T\bar{\boldsymbol{A}}_\gamma & \bar{\boldsymbol{y}}_\gamma^T\bar{\boldsymbol{y}}_\gamma\end{bmatrix} - \boldsymbol{\Delta}\right)\begin{bmatrix}\boldsymbol{\xi}\\ -1\end{bmatrix} =$$

$$= \begin{bmatrix}\boldsymbol{A}_\gamma^T\boldsymbol{P}_\gamma\boldsymbol{A}_\gamma & \boldsymbol{A}_\gamma^T\boldsymbol{P}_\gamma\boldsymbol{y}_\gamma\\ \boldsymbol{y}_\gamma^T\boldsymbol{P}_\gamma\boldsymbol{A}_\gamma & \boldsymbol{y}_\gamma^T\boldsymbol{P}_\gamma\boldsymbol{y}_\gamma\end{bmatrix}\begin{bmatrix}\boldsymbol{\xi}\\ -1\end{bmatrix} - \boldsymbol{\Delta}\cdot\begin{bmatrix}\boldsymbol{\xi}\\ -1\end{bmatrix} = \boldsymbol{0}$$

where the 4×4 matrix $\boldsymbol{\Delta} = \boldsymbol{\Delta}(\bar{\boldsymbol{E}}_\gamma, \bar{\boldsymbol{e}}_\gamma)$ is still unknown, but will be determined such that it has *minimum Frobenius norm*

$$\|\boldsymbol{\Delta}\| = \left[tr(\boldsymbol{\Delta}\boldsymbol{\Delta}^T)\right]^{1/2} = \min. \tag{3.17}$$

among all matrices $\boldsymbol{\Delta}$ that make the difference

$$\begin{bmatrix}\boldsymbol{A}_\gamma^T\boldsymbol{P}_\gamma\boldsymbol{A}_\gamma & \boldsymbol{A}_\gamma^T\boldsymbol{P}_\gamma\boldsymbol{y}_\gamma\\ \boldsymbol{y}_\gamma^T\boldsymbol{P}_\gamma\boldsymbol{A}_\gamma & \boldsymbol{y}_\gamma^T\boldsymbol{P}_\gamma\boldsymbol{y}_\gamma\end{bmatrix} - \boldsymbol{\Delta} = \left[\boldsymbol{U}, \boldsymbol{u}\right]\begin{bmatrix}\boldsymbol{\Lambda} & \boldsymbol{0}\\ \boldsymbol{0} & \lambda_o\end{bmatrix}\begin{bmatrix}\boldsymbol{U}^T\\ \boldsymbol{u}^T\end{bmatrix} - \boldsymbol{\Delta} \tag{3.18}$$

singular. On the right hand side, we have applied the *Singular Value Decomposition* with the provision that all the three eigenvalues in $\boldsymbol{\Lambda}$ are *larger than* the smallest eigenvalue λ_o. In this case, we first obtain

$$\boldsymbol{\Delta}_{\min} = \lambda_o\cdot(\boldsymbol{u}\boldsymbol{u}^T) = \boldsymbol{\Delta}(\left[\tilde{\bar{\boldsymbol{E}}}_\gamma, \tilde{\bar{\boldsymbol{e}}}_\gamma\right]), \tag{3.19a}$$

and consequently

$$(\boldsymbol{U}\cdot\boldsymbol{\Lambda}\cdot\boldsymbol{U}^T)\cdot\begin{bmatrix}\hat{\boldsymbol{\xi}}\\ -1\end{bmatrix} = \boldsymbol{0} \tag{3.19b}$$

from which we learn the proportionality of $\left[\hat{\boldsymbol{\xi}}^T, -1\right]$ to $\boldsymbol{u}^T = \left[u_1, u_2, u_3, u_4\right]$, leading to

$$\hat{\boldsymbol{\xi}}^T = -\frac{1}{u_4}\left[u_1, u_2, u_3\right]. \tag{3.20}$$

This solution is *unique* if rk $\boldsymbol{A}_\gamma$ =3 and λ_o has a *multiplicity* of 1, whereas the *residual matrix* $\left[\tilde{\bar{\boldsymbol{E}}}_\gamma, \tilde{\bar{\boldsymbol{e}}}_\gamma\right]$ may not immediately be isolated from (3.19a). Recognizing, however, that the *orthogonality relations*

$$\begin{bmatrix}\bar{\boldsymbol{A}}_\gamma^T - \tilde{\bar{\boldsymbol{E}}}_\gamma^T\\ \bar{\boldsymbol{y}}_\gamma^T - \tilde{\bar{\boldsymbol{e}}}_\gamma^T\end{bmatrix}\cdot\left[\tilde{\bar{\boldsymbol{E}}}_\gamma, \tilde{\bar{\boldsymbol{e}}}_\gamma\right] = 0 \tag{3.21a}$$

and thus the identities

$$\boldsymbol{\Delta}_{\min} = \begin{bmatrix}\tilde{\bar{\boldsymbol{E}}}_\gamma^T\\ \tilde{\bar{\boldsymbol{e}}}_\gamma^T\end{bmatrix}\cdot\left[\tilde{\bar{\boldsymbol{E}}}_\gamma, \tilde{\bar{\boldsymbol{e}}}_\gamma\right] = \begin{bmatrix}\bar{\boldsymbol{A}}_\gamma^T\\ \bar{\boldsymbol{y}}_\gamma^T\end{bmatrix}\cdot\left[\tilde{\bar{\boldsymbol{E}}}_\gamma, \tilde{\bar{\boldsymbol{e}}}_\gamma\right] \tag{3.21b}$$

hold true, we may get it from the *Singular Value Decomposition* of

$$\begin{aligned}\left[\bar{\boldsymbol{A}}, \bar{\boldsymbol{y}}_\gamma\right] &= \left[\boldsymbol{V}_1, \mathbf{v}, \boldsymbol{V}_2\right]\begin{bmatrix}\boldsymbol{\Lambda}^{1/2} & \boldsymbol{0}\\ \boldsymbol{0} & \lambda_o^{1/2}\\ \boldsymbol{0} & \boldsymbol{0}\end{bmatrix}\begin{bmatrix}\boldsymbol{U}^T\\ \boldsymbol{u}^T\end{bmatrix} =\\ &= \boldsymbol{V}_1\cdot\boldsymbol{\Lambda}^{1/2}\cdot\boldsymbol{U}^T + \lambda_o^{1/2}\cdot(\mathbf{v}\boldsymbol{u}^T)\\ &= \boldsymbol{V}_1\cdot\boldsymbol{\Lambda}^{1/2}\cdot\boldsymbol{U}^T + \left[\tilde{\bar{\boldsymbol{E}}}_\gamma, \tilde{\bar{\boldsymbol{e}}}_\gamma\right]\end{aligned} \tag{3.22a}$$

$$\left[\tilde{\bar{\boldsymbol{E}}}_\gamma, \tilde{\bar{\boldsymbol{e}}}_\gamma\right] = \lambda_o^{1/2}\cdot(\mathbf{v}\boldsymbol{u}^T). \tag{3.22b}$$

If we now re-insert $\Delta_{\min}$ from (3.19a) back into (3.18), formula (3.19b) will give us the relations

$$(\boldsymbol{N}_\gamma - \lambda_o \boldsymbol{I}_3)\hat{\boldsymbol{\xi}} = \mathbf{c}_\gamma \quad for \quad \left[\boldsymbol{N}_\gamma, \boldsymbol{c}_\gamma\right] := \boldsymbol{A}_\gamma^T \boldsymbol{P}_\gamma \left[\boldsymbol{A}_\gamma, \boldsymbol{y}_\gamma\right] \tag{3.23}$$

which may be considered as the "*normal equations*" for TLS, and

$$\begin{aligned} \lambda_o &= \mathbf{y}_\gamma^{\mathsf{T}} \boldsymbol{P}_\gamma \boldsymbol{y}_\gamma - \boldsymbol{c}_\gamma^T \hat{\boldsymbol{\xi}} = \\ &= tr\boldsymbol{\Delta}_{\min} = \tilde{\boldsymbol{e}}_\gamma^T \tilde{\boldsymbol{e}}_\gamma + (vec\tilde{\boldsymbol{E}}_\gamma)^T vec\tilde{\boldsymbol{E}}_\gamma \end{aligned} \tag{3.24}$$

as minimized "*sum of squared residuals*". Obviously, λ_o ought to be smaller than any of the three eigenvalues of N_γ for (3.23) to provide a *unique* solution $\hat{\boldsymbol{\xi}}$ of type (equally weighted) TLS. How to procede in the "*degenerate case*" where N_γ has at least one eigenvalue that is equal to (or even smaller than) λ_o, has been described by *S.van Huffel and J.Vandewalle (1991,chapter 3)*. Also, suggestions to generalize the weighting pattern for the TLS principle have been presented before, but will not be considered in the present contribution. The interested reader is referred to *Schaffrin and Wieser (2008)* for an overview over recently made progress.

4. Case Study: Aeromagnetic Data From West Antarctica and The Goodness-of-Fit

The data that have been used to test the procedure, described in the previous section, stem from a data set of aeromagnetic data that have been collected in West Antarctica. The raw data were first preprocessed by using standard techniques as described by Felus (2002) in detail, particularly to ensure that no dominant trend remains in the data of which only a small portion, namely 1322 points, was actually employed. In essence, the surface that is being constructed via Kriging represents only the fine structure of the overall surface that is defined by the magnetic field values at every point of the sampled domain.

For these data, 30 neighbourhoods (intervals) were introduced with the representative distances between pairs of points reaching from ≈ 500 m to almost 15 km. The empirical semivariogram values spread from just below 4000 to over 11000. After these thirty values had been obtained, a Total Least-Squares fit was computed, using either of the two linearized models, defined in (2.5) and (2.6). The results came out in form of the following estimated parameters:

	"nugget effect" $\hat{\sigma}_e^2$	"sill", $\hat{\sigma}_e^2 + \hat{\sigma}_x^2$	"range", $\hat{r} = \overset{\circ}{r} + \hat{\delta}r$
Spherical model	(still to be specified)		
Exponential model	3830.3	9939.2	5000.8

In order to obtain precision measures for these nonlinear (!) estimates, major error propagations would have to be performed which is left as a future task.

Instead, let us look into the issue how well the respective models fit the empirical semivariogram values, and whether the Total Least-Squares adjustment turns out to be superior to the Weighted Least-Squares adjustment indeed. For this comparison, the sum of weighted squared residuals is being considered, using formula (3.24) in case of TLS, and formula (3.11) for weighted LS. The results are seen in the following table for the range zone only, thereby reflecting on the actual redundancy in these models.

	Spherical Model	Exponential Model
TLS	(still to be specified)	
WLS	4.7258	6.5357

As expected, the TLS procedure provides a clearly better fit than the weighted LS adjustment. The improvement, however, that originates from a model change may still turn out much larger. This obviously means that a poor model choice can normally not be counterbalanced by a smart estimation technique. For the aeromagnetic data analyzed here, the spherical semivariogram model appears to be a suitable choice, to which the TLS may contribute roughly another 15% improvement in fit when compared with the weighted LS adjustment.

4. Conclusions and Outlook

In this contribution a novel approach to semivariogram fitting has been presented, in support of surface reconstruction by Kriging. It is based on the Total Least-Squares (TLS) principle that promises a better representation of the semivariogram function near the origin where it matters most. The fitting algorithm is a bit more complex than in the case of Weighted Least-Squares (WLS) adjustment since the TLS approach leads to nonlinear estimates. This is the reason why, at this stage, no analytical precision measures could be presented. This remains one of the main tasks for the future.

However, the goodness-of-fit was determined for an aeromagnetic dataset where two semivariogram models have been compared, the spherical and the exponential model. In this case, the spherical model provided a much better fit both when using TLS or WLS, respectively. For both models, the TLS fit turned out superior to the WLS fit by 5-15% overall, including mainly the range zone.

In the future, it is planned to introduce a suitable weighting scheme for TLS as well as a modified one for WLS. First steps into this general direction have been undertaken successfully by Schaffrin and Wieser (2008).

Acknowledgment

The authors are exteremely grateful to Prof. Günter Schmitt who served as their host at the Geodetic Institute, University of Karlsruhe (Germany), in the first half of 2007 when most of this paper was written. The assistance by Prof. Yaron Felus in providing the dataset and some preliminary computations is also very much acknowledged.

References

Chilès, J.-P., and P. Delfiner (1999): *Geostatistics. Modelling Spatial Uncertainty*, Wiley: New York etc., 695 pages.

Cressie, N. (1985): Fitting variogram models by weighted least squares, *Math. Geology* **17**(1985), 563-586.

Cressie, N. (1991): *Statistics for Spatial Data* (2nd Edition 1993), Wiley: New York etc., 900 pages.
David, M. (1977): *Geostatistical Ore Reserve Estimation*, Elsevier: Amsterdam/NL.

Dutter, R. (1985): *Geostatistics. An Introduction* (in German), Teubner: Stutgart, 159 pages.

Felus, Y.A. (2002): New methods for spatial statistics in Georaphical Information Systems, Ph.D dissertation, Dept. of Civil and Envir. Engrg. And Geodetic Science, The Ohio State University, Columbus/OH, U.S.A.

Felus, Y.A., and B. Schaffrin (2005): A Total Least-Squares approach in two stages for semivariogram modeling of aeromagnetic data; in: *GIS and Spatial Analysis* (Q. Cheng/ G. Bonham-Carter, eds.), China University of Geosciences Printing House: Beijing 2005, pp.215-220.

Golub, G.H. and C.F. van Loan (1980): An analysis of the Total Least-Squares problem, *SIAM J. Numer. Anal.* **17**(1980), 883-893.

Golub, G.H. and C.F. van Loan (1983): *Matrix Computations*, North Oxford Academic: Oxford/U.K., 476 pages.

Journel, A.G., and Ch. J. Huijbregts (1978): *Mining Geostatistics,* Academic Press: London etc., 600 pages

Kitanidis, P.K. (1985): Minimum-variance unbiased quadratic estimation of covariances of regionalized variables, *Math. Geology* **17**(1985), 195-208.

Kitanidis, P.K. (1997): *Introduction to Geostatistics. Applications in Hydrogeology*, Cambridge University Press: Cambridge/U.K., 249 pages.

Marshall, R. J., and K. V. Mardia (1985): Minimum norm quadratic estimation of components of spatial covariance, *Math. Geology* **17**(1985), 517-525.

Matheron, G. (1963): Principles of geostatistics, *Economic Geology* **58**(1963), 1246-1266.

Matheron, G. (1971): The Theory of Regionalized Variables and its Applications, Cahiers du Centre de Morphologie Mathematiques No.5, Fontainebleau/France.

Schaffrin, B. (2001): Equivalent systems for various forms of Kriging, including least-squares collocation, *Z. für Vermessungwesen* **126** (2001), 87-94.

Schaffrin, B., and A. Wieser (2008): On Weighted Total Least-Squares adjustment for linear regression, *J. of Geodesy* **81** (2008), to appear.

Stein, M.L. (1987): Minimum norm quadratic estimation of spatial variograms, *J. of the Amer. Statist. Assoc*. **82**(1987), 765-772.

van Huffel, S., and J. Vandewalle (1991): *The Total Least-Squares Problem. Computational Aspects and Analysis,* SIAM: Phildelphia/PA, 300 pages.

Zimmermann, D.L., and M.B. Zimmerman (1991): A Monte Carlo comparison of spatial semivariogram estimators and corresponding Ordinary Kriging predictors, *Technometrics* **33**(1991), 77-91.