

## Research Article

## Open Access

B. Schaffrin, T.-S. Bae\*, and Y. Felus

# Optimal biased Kriging: Homeogram tapering and applications to geoid undulations in Korea

<https://doi.org/10.1515/jogs-2018-0016>

Received April 8, 2018; accepted November 27, 2018

**Abstract:** This article studies the *Optimal Biased Kriging (OBK)* approach which is an alternative geostatistical method that gives up the unbiasedness condition of Ordinary Kriging (OK) to gain an improved Mean Squared Prediction Error (MSPE). The system of equations for the optimal linear biased Kriging predictor is derived and its MSPE is compared with that of Ordinary Kriging. A major impediment in implementing this system of equations and performing Kriging interpolation with massive datasets is the inversion of the spatial coherency matrix. This problem is investigated and a novel method, called “homeogram tapering”, which exploits spatial sorting techniques to create sparse matrices for efficient matrix inversion, is described. Finally, as an application, results from experiments performed on a geoid undulation dataset from Korea are presented. A precise geoid is usually the indispensable basis for meaningful hydrological studies over wide areas. These experiments use the theory presented here along with a relatively new spatial coherency measure, called the *homeogram*, also known as the non-centered covariance function.

**Keywords:** geoid undulations, homeogram, homeogram tapering, spatial data sorting, spatial statistics

## 1 Introduction

Geostatistical interpolation methods, for instance various types of Kriging, are often used to generate a predicted surface or map of attribute values at unsampled sites from discrete measurements made at point locations within the same area or region. Many criteria have been proposed

to evaluate the quality of an interpolation method; accuracy, degree of biasedness, and computational complexity are among these criteria. Methods which give up the unbiasedness to obtain an improved relative accuracy as expressed by the Mean Squared Prediction Error MSPE, are popular in various statistical and mathematical applications. These methods, often termed “shrinkage estimators”, were developed to solve problems that contain near-linear dependence (multicollinearity) among the predictor variables (e.g. Montgomery and Friedman, 1993; Hoerl and Kennard, 1970), to treat cases with non-normal data, and to obtain an improved MSPE (Schaffrin 1985, 2000a, 2008). However, it has not been until fairly recently, through the work of Schaffrin (1993) and Gotway and Cressie (1993) that shrinkage methods have been extended to the geostatistical analysis.

Gotway and Cressie (1993) constructed a large class of predictors with a risk (defined as the mean of the total sum of all squared prediction errors - MSPE) that is uniformly smaller than classical Ordinary Kriging (also known as the Best homogeneously Linear Unbiased Prediction - homBLUP). These various predictors were generated using different estimators for this mean, integrated in the Simple Kriging equations.

Schaffrin (1993) investigated homogeneous-isotropic processes on the sphere; these processes cannot be both Gaussian and ergodic. Thus, when considering a global phenomenon such as (incremental) geoid undulations, magnetic or gravity fields, etc., one cannot use standard geostatistical methods. For example, the covariance function estimate will always be biased on a sphere leading to a biased predictor. To solve this problem, Schaffrin (1993) proposed the Optimal Biased Kriging predictor (OBK, also termed homBLIP - Best homogeneously LInear Prediction) as a biased alternative to Ordinary Kriging (OK) with a (slightly) reduced mean squared error. The following section further develops the OBK predictor formulas (with a full proof in the Appendix). In particular,

(i) the different factors that affect this method are analyzed,

(ii) the computational aspects of this technique are investigated, and

**B. Schaffrin:** School of Earth Sciences, Geodetic Science Division, The Ohio State University, Columbus, OH 43210, USA

\***Corresponding Author: T.-S. Bae:** Dept. Geoinformation Engineering, Sejong University, Seoul, Korea, E-mail: baezae@sejong.ac.kr

**Y. Felus:** Survey of Israel, Tel-Aviv, Israel

(iii) a dataset of geoid undulations from Korea is used to test this method.

This paper is organized as follows: Section 2 presents the OBK concepts, its formulas, and the non-centered covariance function (the homeogram) as measure of spatial coherency. Section 3 describes a novel method to efficiently invert the spatial coherency matrix. Experiments with the OBK in interpolating a geoid undulation dataset from Korea are then provided in Section 4. Section 5 concludes our findings and provides a brief outlook on further research issues.

## 2 The Optimal Biased Kriging

For clarity of discussion, the definitions of the mathematical model for the spatial process under investigation will be reviewed. In a vector form it reads:

$$\begin{aligned} \mathbf{y} &= \mathbf{x} + \mathbf{e}, & \text{and} & \left[ \begin{array}{c} \mathbf{e} \\ \mathbf{e}_0 \end{array} \right] \sim \left( \left[ \begin{array}{c} \mathbf{0} \\ \mathbf{0} \end{array} \right], \left[ \begin{array}{cc} \sigma_e^2 \cdot \mathbf{I}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_x \end{array} \right] \right) \end{aligned} \quad (1)$$

where  $\mathbf{y} := [y(s_1), \dots, y(s_n)]^T$  is the  $n \times 1$  vector of observed process values (data) at sites  $s_1, \dots, s_n$ ;

$\mathbf{x} := [x(s_1), \dots, x(s_n)]^T$  is the  $n \times 1$  random effects vector of the actual process at sites  $s_1, \dots, s_n$ ;

$\mathbf{e} := [e_1, \dots, e_n]^T$  is the  $n \times 1$  vector of random observation errors;

$\mathbf{e}_0 := [e_{01}, \dots, e_{0n}]^T$  is the  $n \times 1$  vector of random process mis-centering errors;

$\beta$  is the process mean value (usually unknown);

$\mathbf{C}_x$  is the  $n \times n$  dispersion matrix of  $\mathbf{x}$  with the elements  $C_x(s_i, s_j)$  derived from the respective covariance function;  $\sigma_e^2$  is the variance component for the observational noise (assumed to be otherwise uncorrelated); and

$\boldsymbol{\tau} := [1, 1, \dots, 1]^T$  is the  $n \times 1$  "summation vector".

Note that the process  $x(s')$  is a scalar-valued process at an arbitrary site denoted by  $s'$ . This process is predicted through Optimal Biased Kriging (or homBLIP) using the non-centralized covariance function, also called the *homeogram*, which is defined as:

$$\eta_x(s, s+h) = E\{x(s) \cdot x(s+h)\} \quad (2)$$

where  $\eta_x(s, s+h)$  describes the spatial coherency between points  $s$  and  $s' = s+h$ ;  $E$  is the expectation operator, and  $h$  is the distance difference or lag vector. The homeogram was already recommended by Jeannée and de Fouquet (2000) and Journel (1988), as well as Schaffrin (1985), for cases where the data have an unknown or positively skewed distribution. The following derivation shows that the homeogram can provide a simplified system for OBK.

The homeogram vector between the unknown point at location  $s'$  and the sample data can be expressed as:

$$\boldsymbol{\eta}(s') = [\eta_x(s', s_1), \dots, \eta_x(s', s_n)]^T, \quad (3)$$

and the homeogram matrix between the sample data points as:

$$\mathbf{H} = [\eta_x(s_{ij})] + \sigma_e^2 \mathbf{I}_n \quad \text{for } i, j \in \{1, \dots, n\} \quad (4)$$

The Optimal Biased Kriging prediction using the homeogram function is given by (see Appendix for a complete derivation):

$$\tilde{\tilde{x}}(s') = \boldsymbol{\eta}(s')^T \cdot \mathbf{H}^{-1} \cdot \mathbf{y} \quad (5)$$

with

$$MSPE\{\tilde{\tilde{x}}(s')\} = \eta_x(s', s') - \boldsymbol{\eta}(s')^T \cdot \mathbf{H}^{-1} \cdot \boldsymbol{\eta}(s'). \quad (6)$$

In order to see the difference between the OBK and OK equations, the OK solution will be expressed in terms of the homeogram as given in Schaffrin (2001), where the general equivalence between OK and least-squares collocation (with constant trend) has also been established under mild conditions. Denote:

$$\mathbf{K} := (\sigma_e^2 \mathbf{I}_n + \mathbf{C}_x) = \mathbf{H} - \boldsymbol{\tau} \cdot \beta^2 \cdot \boldsymbol{\tau}^T$$

as the  $n \times n$  covariance matrix of the observed process  $y(s_i)$  in the sample points, and

$$\boldsymbol{\kappa}(s') := [C_x(s', s_1), \dots, C_x(s', s_n)]^T = \boldsymbol{\eta}(s') - \boldsymbol{\tau} \cdot \beta^2 \quad (7a)$$

as the corresponding  $n \times 1$  vector of covariances between the point at the new location  $s'$  and the sample points, with  $\beta = E\{x(s')\} = E\{x(s_i = s' + h_i)\}$  as the spatial process mean. Then the Ordinary Kriging prediction, following Cressie (1993, p. 123), may be given by:

$$\begin{aligned} \tilde{x}(s') &= \hat{\beta} + \boldsymbol{\kappa}(s')^T \cdot \mathbf{K}^{-1} \cdot (\mathbf{y} - \boldsymbol{\tau} \cdot \hat{\beta}) \\ &= \tilde{\tilde{x}}(s') + \bar{v}(s') \cdot (\boldsymbol{\tau}^T \cdot \mathbf{H}^{-1} \cdot \mathbf{y}) \end{aligned} \quad (8a)$$

with

$$\begin{aligned} \hat{\beta} &= (\boldsymbol{\tau}^T \cdot \mathbf{K}^{-1} \cdot \boldsymbol{\tau})^{-1} \cdot (\boldsymbol{\tau}^T \cdot \mathbf{K}^{-1} \cdot \mathbf{y}) \\ &= (\boldsymbol{\tau}^T \cdot \mathbf{H}^{-1} \cdot \boldsymbol{\tau})^{-1} \cdot (\boldsymbol{\tau}^T \cdot \mathbf{H}^{-1} \cdot \mathbf{y}) \end{aligned} \quad (8b)$$

and

$$\begin{aligned} MSPE\{\tilde{x}(s')\} &= \sigma_x^2 - \boldsymbol{\kappa}(s')^T \cdot \mathbf{K}^{-1} \cdot \boldsymbol{\kappa}(s') \\ &= \frac{[1 - \boldsymbol{\kappa}(s')^T \cdot \mathbf{K}^{-1} \cdot \boldsymbol{\tau}]^2}{(\boldsymbol{\tau}^T \cdot \mathbf{K}^{-1} \cdot \boldsymbol{\tau})} = \end{aligned} \quad (9a)$$

$$= MSPE\{\tilde{\tilde{x}}(s')\} + [\bar{v}(s')]^2 (\boldsymbol{\tau}^T \cdot \mathbf{H}^{-1} \cdot \boldsymbol{\tau}) \geq MSPE\{\tilde{\tilde{x}}(s')\}, \quad (9b)$$

where  $\bar{v}(s')$  may be taken from the system

$$\begin{bmatrix} \mathbf{H} & -\boldsymbol{\tau} \\ -\boldsymbol{\tau}^T & 0 \end{bmatrix} \begin{bmatrix} \bar{\chi}(s') \\ \bar{v}(s') \end{bmatrix} = \begin{bmatrix} \boldsymbol{\eta}(s') \\ -1 \end{bmatrix} \quad (10a)$$

or computed directly via

$$\bar{v}(s') = (\boldsymbol{\tau}^T \cdot \mathbf{H}^{-1} \cdot \boldsymbol{\tau})^{-1} [1 - \boldsymbol{\tau}^T \cdot \mathbf{H}^{-1} \cdot \boldsymbol{\eta}(s')]. \quad (10b)$$

Obviously,

$$MSPE\{\tilde{\chi}(s')\} = \sigma_x^2 - \boldsymbol{\kappa}(s')^T \cdot \mathbf{K}^{-1} \cdot \boldsymbol{\kappa}(s') + \frac{[1 - \boldsymbol{\kappa}(s')^T \cdot \mathbf{K}^{-1} \cdot \boldsymbol{\tau}]^2}{(\beta^{-2} + \boldsymbol{\tau}^T \cdot \mathbf{K}^{-1} \cdot \boldsymbol{\tau})} \quad (11)$$

holds when using the covariance function.

It follows from Eq. (11), that if the number of elements  $n$  is large, then  $\boldsymbol{\tau}^T \mathbf{K}^{-1} \boldsymbol{\tau}$  in the denominator of the equation is very large relative to  $\beta^{-2}$ , and the improvement may be negligible. On the other hand, if the mean is very small (for example after applying a trend removing process like "median polish") there can be a noticeable improvement, and the MSPE will be closer to that of Simple Kriging (SK):

$$MSPE\{\tilde{\chi}_{SK}(s')\} = \sigma_x^2 - \boldsymbol{\kappa}(s')^T \cdot \mathbf{K}^{-1} \cdot \boldsymbol{\kappa}(s') \quad (12)$$

From the discussion above and Eqs. (5) and (6) it is clear that Optimal Biased Kriging should be computed using the homeogram matrix, thereby avoiding the unnecessary problem of computing any estimator of  $\beta$  which is not even needed for OK in Eq. (8a). Implementation of Eqs. (5) and (6) requires an inversion of the  $n \times n$  homeogram matrix ( $n$  being the number of given data points). This complex operation could be a major obstacle in trying to apply this method to large datasets, consisting of more than a few hundred points. In the next section, a method to alleviate the problem of working with large spatial coherency matrices will be presented; we may call it "homeogram tapering".

### 3 Inverting the spatial coherency matrix using spatial sorting; especially homeogram tapering

A major obstacle of spatial statistics methods is the high computational complexity associated with the inversion of the large spatial coherency matrices; cf. Eqs. (5)–(12). This problem has led to many investigations of improved algorithms and of the use of high performance computers. Li (1996) reported on ways of using parallel processing for spatial statistics algorithms. Goovaerts (1997, pp. 178–179) suggested taking only a subset of the given data

points instead of the entire dataset. Gundlich et al. (2003) employed sampling-based Monte Carlo methods for the efficient computation of large covariance matrices.

Furrer et al. (2006) as well as Barry and Pace (1997) exploit the sparseness of the covariance matrix  $\mathbf{K}$  to efficiently solve the Kriging equations. If an observation displays stochastic dependence with its nearest  $m - 1$  neighbors, only  $m$  non-zero entries exist per row of the spatial covariance matrix  $\mathbf{K}$ . Sparsity enhances computational efficiency. For example, multiplying an  $n \times n$  matrix  $\mathbf{A}_1$  by an  $W2LOK$  matrix  $\mathbf{A}_2$  requires  $O(n_2 \cdot n^2)$  operations using dense matrices while, with the equivalent sparse matrices, multiplication computation requires  $O(n_2 \cdot n \cdot m)$  operations. The enhancement is even more significant for matrix inversion. This operation can be built upon the Cholesky factorization or, more generally, the LU factorization (lower times upper triangular matrix) which, using Gaussian elimination, requires  $O(n^3)$  operations for dense matrices. However, if the matrix is sparse and has a band structure, with lower bandwidth  $p$  and upper width  $q$ , the operation requires time; see Golub and van Loan (1996, pp. 150-151). Unfortunately, the problem of constructing a band-limited sparse matrix is known to be NP-complete; see Skiena (1997, pp. 202-203). Thus, any solution to this problem relies on so-called "brute-force" methods or ad-hoc heuristics.

Here, a novel approach is presented that takes advantage of spatial sorting to create a stable and nearly band-limited sparse covariance, resp. homeogram matrix, thus making the computation more efficient. The Morton ordering, also known as the "Peano Key" method, was chosen as the spatial sorting technique. This method has the lowest mean absolute-difference measures and the overall minimum value of the Moran statistics. It is also monotonic and quadrant-recursive; cf. Abel and Mark (1990). The Morton ordering technique rearranges the data by using both the X-coordinate and the Y-coordinate and by interleaving the X-, Y-coordinate numbers. Higher resolution in the spatial sorting may be achieved by applying the method using the binary representation of the coordinates (see Fig. 1).

Morton Ordering can be performed using a Radix sorting which has a worst case of  $O(n \cdot \log n)$  time but, with a good implementation scheme, can run in an  $O(n)$  average time (Skiena, 1997). Morton ordering will not provide a perfectly band-limited covariance matrix, due to the discontinuities in the spatial sorting (Fig. 2), and a slight improvement using for example one or two iterations of the *Cuthill-McKee ordering* algorithm may be needed; cf. Meurant (1999), for instance.

In addition to creating a nearly band-limited matrix, spatial sorting can be used to detect observations that

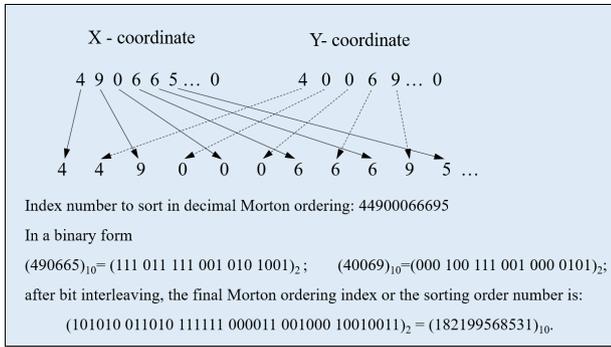


Figure 1: Example for the Morton Order sorting algorithm.

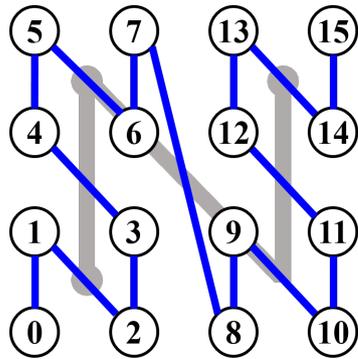


Figure 2: Morton Ordering, spatial sorting as implemented on grid data. This method is applied here to irregularly sampled data. Note the jump between points 7 and 8.

are very close to each other (relative to the resolution), or nearly duplicated. These observations should be removed since the equation system will become unstable as was shown by Davis and Morris (1997). Naturally, elimination of observations will negatively affect reliability as described by Schaffrin (2000b).

So far, we have discussed the covariance function and methods to turn it into a band-limited stable matrix. Unlike the covariance function, the *homeogram function* will not converge to zero at infinity if the data exhibit a trend. To create a sparse matrix in this case, the homeogram matrix can be decomposed via a rank-1 modification as:

$$\mathbf{H} = \bar{\mathbf{H}} + (\boldsymbol{\tau} \cdot \alpha_h \cdot \boldsymbol{\tau}^T) \quad (13)$$

where is a constant that represents the *plateau value* of the homeogram such that  $\bar{\mathbf{H}}$  becomes sparse while remaining positive-definite, with zeros at values beyond the range, i.e. at the plateau. Consequently, the inverse of  $\mathbf{H}$  can be computed using the well known formula:

$$\mathbf{H}^{-1} = \bar{\mathbf{H}}^{-1} - \bar{\mathbf{H}}^{-1} \cdot \boldsymbol{\tau} \cdot (\alpha_h^{-1} + \boldsymbol{\tau}^T \bar{\mathbf{H}}^{-1} \boldsymbol{\tau})^{-1} \boldsymbol{\tau}^T \bar{\mathbf{H}}^{-1} \quad (14)$$

The adapted homeogram matrix and the covariance matrix  $\mathbf{K}$  will contain  $nm$  non-zero elements out of possi-

ble elements which is an  $m/n$  proportion of non-zero elements. These two matrices can be converted into nearly sparse matrices using Morton Ordering.

We shall call the above technique “homeogram tapering”. Other tapering methods, mainly applied to the covariance function, have been discussed by Furrer et al. (2006), Sang et al. (2011), Sang and Huang (2012), as well as Vetter et al. (2014). We also refer to the literature review by Ozanne et al. (2014) which covers a lot of ground on “covariance tapering”, but nothing on “homeogram tapering”.

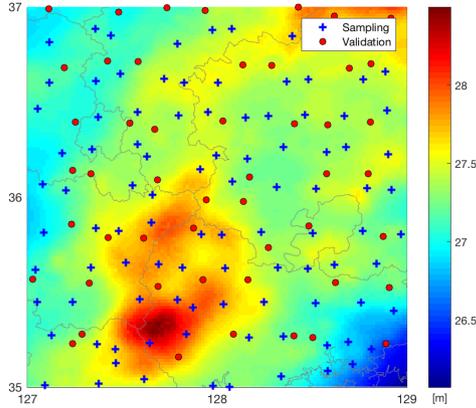
## 4 Experimental results

### 4.1 The data: Geoid undulations from Korea

For the numerical test of the Optimal Biased Kriging based on the spatial coherency functions, we selected the hybrid geoid of Korea which was generated from airborne and terrestrial gravimetric data along with the GPS/leveling observations. The test area is  $2^\circ$  ( $2^\circ$  with the spatial resolution of about 1.8 km, which includes various conditions ranging from benign to mountainous rough terrain. The Continuously Operating Reference Stations (CORS) of Korea provide both ellipsoidal and orthometric heights, thus the CORS within the test area are grouped into two sets; a total of 100 sampling points was used to generate the empirical values of spatial coherence, and another 50 validation points were chosen to compare the estimated geoid after Kriging (Fig. 3). The GPS/leveling data were compared with the geoid model at sampled points, resulting in a mean value of  $-0.85$  cm, and the root-mean-squared error (RMSE) to be  $\pm 2.9$  cm. Based on the statistics at sampling points, the variance of the random observation error was reasonably set to  $100 \text{ cm}^2$ . In order to better estimate the spatial coherency functions, the linear trend in the geoid model was removed by fitting it to a planar model with the geodetic latitude and longitude as parameter lines.

### 4.2 Fitting the homeogram and the covariance function

Two spatial coherency functions were estimated in this study, namely homeogram and covariance function. Since the homeogram is the non-centralized covariance function, the empirical homeogram is shifted by the square of a mean value. Therefore, we present the fitting results of the homeogram to the analytical models. According to Cressie



**Figure 3:** The geoid and the location of CORS for data sampling and validation. The geoid shows the residuals after removing the linear trend by fitting to a planar model (Bae, 2016).

**Table 1:** The quality of fit for the spatial coherency functions.

|                     | Exponential             | Wave                    |
|---------------------|-------------------------|-------------------------|
| Homeogram           | $1.7889 \times 10^{-7}$ | $6.3970 \times 10^{-7}$ |
| Covariance function | $2.4630 \times 10^{-7}$ | $4.7590 \times 10^{-7}$ |

(1993), the exponential model, also known as ‘Gaussian model’, is given by

$$\eta_Y(h) = \begin{cases} c_0 \cdot e^{-a^2|h|^2} + \mu^2 & |h| \neq 0, \\ \sigma_Y^2 + \mu^2 = c_0 + \sigma_e^2 & |h| = 0, \end{cases} \quad (15)$$

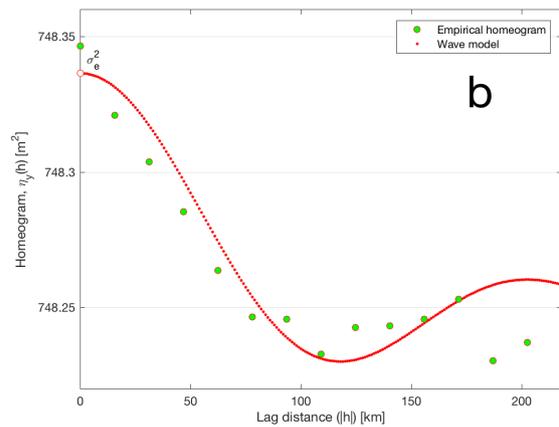
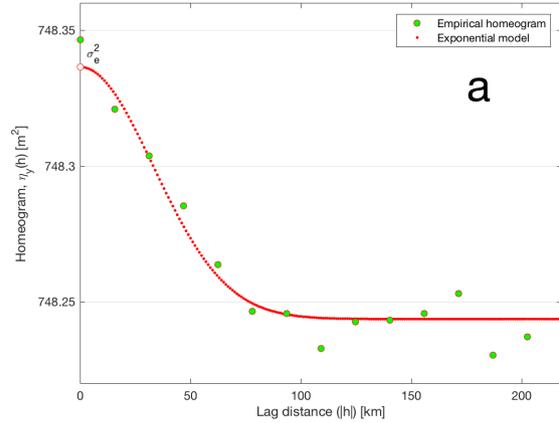
and the wave model can be represented as

$$\eta_Y(h) = \begin{cases} c_0 - c_h \left[ 1 - \frac{a_h}{|h|} \sin\left(\frac{|h|}{a_h}\right) \right] & |h| \neq 0, \\ \sigma_Y^2 + \mu^2 & |h| = 0, \end{cases} \quad (16)$$

The empirical homeogram was fitted using least-squares adjustment with the inverse distance weighting scheme (Fig. 4), resulting in the quality-of-fit as taken from the residuals and given in Table 1.

### 4.3 Comparing Ordinary Kriging (OK) with Optimal Biased Kriging (OBK): Smallest MSPE vs. best Kriging results

The MSPE of OBK was compared with that of OK for two spatial coherency functions. The choice of the empirical spatial coherency function seems to favor the homeogram over the covariance function, though the difference between the OBK and the OK is minor in terms of the expected MSPE (Figure 5). The validation points in the latter part (No. 43-49) show a disordered estimation with a



**Figure 4:** Fitting an empirical homeogram to the analytical models. (a) Exponential model and (b) wave model.

big number of MSPE. This can be attributed to the fact that these stations are located in a highly rough terrain (southern part of the test area).

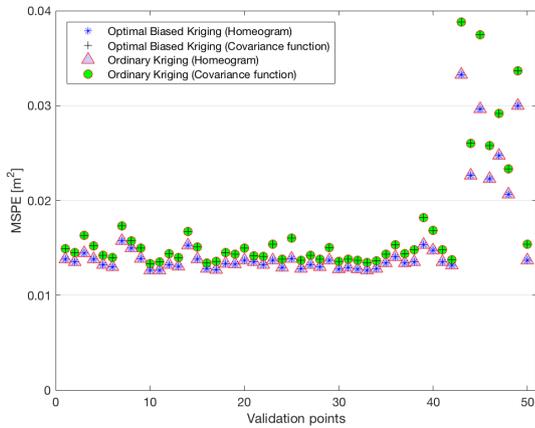
As can be seen in Fig. 6, the type of Kriging predictor does not appear to matter much. However, at the validation points in rough condition, the homeogram provides relatively reliable Kriging results compared to those of the covariance function.

The performance using the empirical covariance function compares well with the use of the empirical homeogram. For both homeogram and covariance function, the empirical spatial coherency was better fitted to the exponential model than the wave model; thus the average value of the Kriging results shows a smaller number for the exponential model.

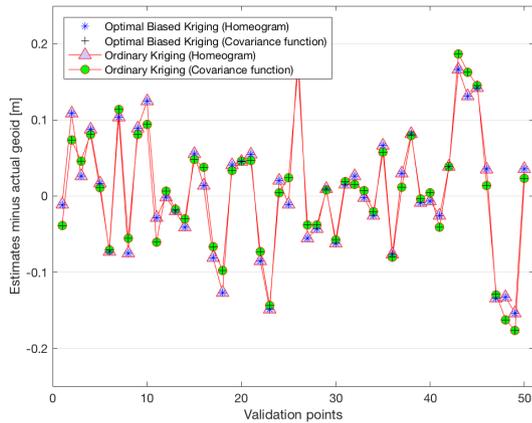
Both Optimal Biased Kriging and Ordinary Kriging lead to essentially the same prediction results, whereas the formal MSPE of OBK appears always smaller than that of OK. In theory, the choice of the spatial coherency function (homeogram, covariance function) should be irrelevant.

**Table 2:** Statistics of the Kriging results. The average Kriging results represent the difference between the interpolated value and the geoid model; the mean squared deviation was calculated based on the Kriging results.

| Predictor                   | Exponential | Homeogram |             | Covariance function |         |
|-----------------------------|-------------|-----------|-------------|---------------------|---------|
|                             |             | Wave      | Exponential | Wave                | Wave    |
| Avg. Kriging results [cm]   | OBK         | 0.6528    | 1.2792      | 0.5779              | 1.0189  |
|                             | OK          | 0.6536    | 1.2797      | 0.5788              | 1.0195  |
| Mean squared deviation [cm] | OBK         | 8.1174    | 10.7980     | 8.0586              | 10.7599 |
|                             | OK          | 8.1175    | 10.7980     | 8.0588              | 10.7599 |



**Figure 5:** The MSPE of Optimal Biased Kriging and Ordinary Kriging at 50 validation points for the spatial coherency of homeogram and covariance function.



**Figure 6:** The resulting estimates of the Kriging predictors for Optimal Biased Kriging and Ordinary Kriging.

However, since these functions have to be determined empirically, their choice matters indeed. It is interesting that *the better fit of the empirical functions does not always yield better Kriging results*. The wave model seems to be particularly affected by this phenomenon. Oftentimes, the for-

mal MSPE does not represent the actual predictive quality. More systematic studies are still necessary in this direction.

## 5 Discussion and outlook

Our case study had three unique goals as mentioned in the Introduction.

The results of the Optimal Biased Kriging prediction experiments were quite unexpected as the differences between the Mean Squared Prediction Error from Ordinary Kriging and the Optimal Biased Kriging were *relatively minute*. Comparing Eqs. (6) and (9a-b), this is rather surprising at first glance. However, as was mentioned before, with more points in the dataset,  $\tau^T \mathbf{K}^{-1} \tau$  becomes a relatively large element in the denominator of the MSPE Eq. (9a) and therefore the improvement through the addition of  $\beta^{-2}$  will be subdued. These results complement the study by Gotway and Cressie (1993), who computed the relative risks of the James-Stein estimator and the Best Linear Uniformly Unbiased Estimate (BLUUE) as a function of the norm of the bias vector  $\tau\beta$ , i.e.  $\beta\sqrt{n}$ , on simulated data. The greatest reduction in risk i.e. occurs for  $\beta$  near zero, and this reduced relative risk was just 99.89% of the original. As  $\beta$  increases, the improvement in the relative risk becomes *more and more negligible*.

Moreover, in contrast to what is expected from biased Kriging, namely that the results are allowed to be biased, the results of this study exhibit almost no bias in both the detrended and the original data with a trend. The key issue in the understanding of this phenomenon is the *homeogram*. Unlike the covariance function, the homeogram carries information about the mean through the plateau parameter. The use of the homeogram function as part of the Kriging equation system is a relatively new topic. Further research has been done to investigate the characteristics of the homeogram function and will be published elsewhere; nevertheless it appears that OBK has already

shown its potential to become a serious competitor to Ordinary Kriging in many practical situations.

Measurements of CPU time, during the experiments, have shown that Kriging computation using sparse matrix techniques is very efficient and increases the speed of the interpolation algorithm by almost threefold. Sorting the data with the Morton Order technique creates a band-limited matrix, which is not yet optimal due to the spatial jump (see Fig. 2). However, given the nearly ordered matrix, an additional improvement by using, for example, the *Cuthill-McKee ordering* (Meurant, 1999) is much faster and the overall computational speed is improved by an average of 15%, compared with the same process without sorting; further details on the empirical CPU timing experiments are given in Felus (2001). These results support the analytical complexity analysis of the algorithm as described in a previous section. Spatial statistics can greatly benefit by using GIS methods; for example, spatial coherency function computation, an time process, can be significantly enhanced by using spatial sorting to classify data point into distance lags, thus reducing the algorithm time. Further research is needed toward exploiting this potential.

**Acknowledgement:** This research was supported by a grant (18RDRP-B076564-05) from Regional Development Research Program funded by Ministry of Land, Infrastructure and Transport of Korean government.

## A Appendix. Optimal Biased Kriging (OBK)

Starting from the mathematical model of Eq. (1) and using the standard assumption of no correlation between the measurement error vector  $\mathbf{e}$ , the spatial process random variation vector  $\mathbf{e}_0$ , and the spatial process  $\mathbf{x}$  itself, namely:  $C\{\mathbf{e}, \mathbf{e}_0\} = \mathbf{0} \Rightarrow C\{\mathbf{x}, \mathbf{e}\} = \mathbf{0}$ , the following conditions should be satisfied for the OBK prediction  $\tilde{x}(s')$  of  $x(s')$ :

(1) Homogeneously linear:

$$\tilde{x}(s') = \mathbf{L} \cdot \mathbf{y} \quad (\text{A.1})$$

where  $\mathbf{L}$  is an unknown row vector of size  $1 \times n$ .

(2) Best or minimum MSPE (i.e., minimum mean squared prediction error):

$$\begin{aligned} MSPE\{\tilde{x}(s')\} &= E\{[\tilde{x}(s') - x(s')]^2\} \\ &= D\{\tilde{x}(s') - x(s')\} + (E\{\tilde{x}(s') - x(s')\})^2 \\ &= D\{\mathbf{L} \cdot \mathbf{y} - x(s')\} + (E\{\mathbf{L} \cdot \mathbf{y} - x(s')\})^2 \\ &= D\{\mathbf{L} \cdot (\mathbf{x} + \mathbf{e}) - x(s')\} + \beta^2 \cdot (\mathbf{L} \cdot \boldsymbol{\tau} - 1)^2 \\ &= \mathbf{L} \cdot (\sigma_e^2 \mathbf{I}_n) \cdot \mathbf{L}^T + \mathbf{L} \cdot \mathbf{C}_x \cdot \mathbf{L}^T - \boldsymbol{\kappa}(s')^T \cdot \mathbf{L}^T \\ &\quad - \mathbf{L} \cdot \boldsymbol{\kappa}(s') + \mathbf{C}_x(s', s') + \beta^2 \cdot (\mathbf{L} \cdot \boldsymbol{\tau} - 1)^2 \\ &= \Phi(\mathbf{L}^T) \end{aligned} \quad (\text{A.2})$$

where  $D$  denotes the dispersion (or variance-covariance) matrix. Here, we assumed that the covariance function is isotropic and homogeneous/stationary, which means that

$$\begin{aligned} C(\mathbf{x}, x(s')) &= C(x(s'), \mathbf{x})^T = [C(x(s_i), x(s'))] = \boldsymbol{\kappa}(s'), \\ & i \in \{1, \dots, n\}; \end{aligned}$$

$$C(x(s'), x(s')) = C_x(s', s') = C_x(0) = \sigma_x^2;$$

and from Eq. (1):

$$C(\mathbf{e}, \mathbf{e}) = D\{\mathbf{e}\} = \sigma_e^2 \mathbf{I}_n;$$

$$C(\mathbf{x}, \mathbf{x}) = D\{\mathbf{x}\} = \mathbf{C}_x.$$

Now,  $\Phi(\mathbf{L}^T)$  is the target function that needs to be minimized. Euler-Lagrange necessary conditions (first partial derivatives set to zero) are used for this minimization:

$$\begin{aligned} \frac{\partial \Phi(\mathbf{L}^T)}{2 \cdot \partial \mathbf{L}^T} &= [\sigma_e^2 \cdot \mathbf{I}_n + (\mathbf{C}_x + \boldsymbol{\tau} \cdot \beta^2 \cdot \boldsymbol{\tau}^T)] \cdot \mathbf{L}^T \\ &\quad - [\boldsymbol{\kappa}(s') + \boldsymbol{\tau} \cdot \beta^2] = \mathbf{0} \end{aligned} \quad (\text{A.3})$$

which, after rearranging Eq. (A.3), results in:

$$\mathbf{L}^T = (\sigma_e^2 \mathbf{I}_n + \mathbf{C}_x + \boldsymbol{\tau} \cdot \beta^2 \cdot \boldsymbol{\tau}^T)^{-1} \cdot (\boldsymbol{\kappa}(s') + \boldsymbol{\tau} \cdot \beta^2), \quad (\text{A.4})$$

and consequently in:

$$\tilde{x}(s') = \mathbf{L} \cdot \mathbf{y} = (\boldsymbol{\kappa}(s') + \boldsymbol{\tau} \cdot \beta^2)^T \cdot (\sigma_e^2 \mathbf{I}_n + \mathbf{C}_x + \boldsymbol{\tau} \cdot \beta^2 \cdot \boldsymbol{\tau}^T)^{-1} \cdot \mathbf{y}. \quad (\text{A.5})$$

Using the definition of the homeogram and the notations from Eqs. (7a) and (7b), one can directly obtain:

$$\tilde{x}(s') = \eta(s')^T \cdot \mathbf{H}^{-1} \cdot \mathbf{y} \quad (\text{A.6})$$

with  $\mathbf{L} = \eta(s')^T \cdot \mathbf{H}^{-1}$ .

It is easy to check that the sufficient condition is satisfied and the matrix of second derivatives is positive-definite:

$$\frac{\partial^2 \Phi(\mathbf{L}^T)}{2 \cdot \partial \mathbf{L} \cdot \partial \mathbf{L}^T} = \sigma_e^2 \mathbf{I}_n + \mathbf{C}_x + \boldsymbol{\tau} \cdot \beta^2 \cdot \boldsymbol{\tau}^T = \mathbf{H} > \mathbf{0}. \quad (\text{A.7})$$

The MSPE of this prediction is computed from Eq. (A.2) as:

$$\begin{aligned} MSPE\{\tilde{x}(s')\} &= \mathbf{L} \cdot (\sigma_e^2 \mathbf{I}_n + \mathbf{C}_x + \boldsymbol{\tau} \cdot \beta^2 \cdot \boldsymbol{\tau}^T) \cdot \mathbf{L}^T \\ &\quad - \mathbf{L} \cdot [\boldsymbol{\kappa}(s') + \boldsymbol{\tau} \cdot \beta^2] - [\boldsymbol{\kappa}(s')^T + \beta^2 \cdot \boldsymbol{\tau}^2] \cdot \mathbf{L}^T + \sigma_x^2 + \beta^2 \\ &= \boldsymbol{\eta}(s')^T \cdot \mathbf{H}^{-1} \cdot (\sigma_e^2 \mathbf{I}_n + \mathbf{C}_x + \boldsymbol{\tau} \cdot \beta^2 \cdot \boldsymbol{\tau}^T) \cdot \mathbf{H}^{-1} \cdot \boldsymbol{\eta}(s') \\ &\quad + \{\sigma_x^2 + \beta^2\} - \boldsymbol{\eta}(s')^T \cdot \mathbf{H}^{-1} \cdot [\boldsymbol{\kappa}(s') + \boldsymbol{\tau} \cdot \beta^2] \\ &\quad - [\boldsymbol{\kappa}(s')^T + \beta^2 \boldsymbol{\tau}^T] \cdot \mathbf{H} \cdot \boldsymbol{\eta}(s') \\ &= \boldsymbol{\eta}(s')^T \cdot \mathbf{H}^{-1} \cdot \boldsymbol{\eta}(s') + \eta_x(s', s') - \boldsymbol{\eta}(s')^T \cdot \mathbf{H}^{-1} \cdot \boldsymbol{\eta}(s') \\ &\quad - \boldsymbol{\eta}(s')^T \mathbf{H}^{-1} \cdot \boldsymbol{\eta}(s'), \end{aligned}$$

and finally as:

$$MSPE\{\tilde{x}(x')\} = \eta_x(s', s') - \boldsymbol{\eta}(s')^T \cdot \mathbf{H}^{-1} \cdot \boldsymbol{\eta}(s') \quad (\text{A.8})$$

## References

- Abel, J.D., Mark, M.D., 1990. A comparative analysis of some two-dimensional orderings. *International Journal of Geographical Information Systems* 4 (1), 21-31.
- Bae, T.S., 2016. Estimation of spatial coherency functions for Kriging of spatial data (in Korean). *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography* 34 (1), 91-98.
- Barry, P.R., Pace, R.K., 1997. Kriging with large data sets using sparse matrix techniques. *Communications in Statistics: Simulation and Computation* 26 (2), 619-629.
- Cressie, N.A.C., 1993. *Statistics for Spatial Data*, second ed. Wiley, New York.
- Davis, J.G., Morris, D.M., 1997. Six factors which affect the condition number of matrices associated with Kriging. *Mathematical Geology* 29 (5), 669-683.
- Felus, Y., 2001. New methods for spatial statistics in geographic information systems. *PhD dissertation*, Dept. of Civil & Environ. Engineering and Geodetic Sci., The Ohio State University, Columbus, Ohio, 176p.
- Furrer, R., Genton, M.G., Nychka, D., 2006. Covariance tapering for interpolation of large spatial datasets. *Journal of Computational and Graphical Statistics* 15 (3), 502-523.
- Golub, H.G., van Loan, F.C., 1996. *Matrix Computations*, third ed. The Johns Hopkins University Press, Baltimore, MD.
- Goovaerts, P., 1997. *Geostatistics for Natural Resources Evaluation*. Oxford University Press, New York.
- Gotway, C.A., Cressie, N.A.C., 1993. Improved multivariate prediction under a general linear model. *Journal of Multivariate Analysis* 45 (1), 56-72.
- Gundlich, B., Koch, K.R., Kusche, J., 2003. Gibbs sampler for computing and propagating large covariance matrices. *J. Geodesy* 77 (9), 514-528.
- Hoerl, A.E., Kennard, R.W., 1970. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* 12 (1), 55-67.
- Jeannée, N., de Fouquet, C., 2000. Characterization of soil pollutions from former coal processing sites. In: Kleingeld, W.J., Krige, J.D. (Eds.), *Geostatistics Congress 2000: Vol. 2*. Cape Town, South Africa, pp. 526-537.
- Journal, A.G., 1988. New distance measures: The route toward truly non-Gaussian geostatistics. *Mathematical Geology* 20 (4), 459-475.
- Li, B., 1996. Implementing spatial statistics on parallel computers. In: Arlinghaus, S.L., Griffith, D.A. (Eds.), *Practical Handbook of Spatial Statistics*. CRC Press, Boca Raton/FL, USA, pp. 107-149.
- Meurant, G., 1999. *Computer Solution of Large Linear Systems*. Elsevier, Amsterdam, The Netherlands.
- Montgomery, D.C., Friedman, D.J., 1993. Prediction using regression models with multicollinear predictor variables. *IIE Transactions* 25 (3), 73-85.
- Ozanne, M., Schneider, G., White, S., Yin, M., Craigmile, P., Herbei, R., Notz, W., 2014. A covariance tapering literature review. The Ohio State University, Columbus/OH, USA, 38p.
- Pace, R.K., Barry, P.R., 1997. Quick computation of spatial autoregressive estimators. *Geographical Analysis* 29 (3), 232-247.
- Sang, H., Huang, J.Z., 2012. A full-scale approximation of covariance functions for large spatial data sets. *J. Roy. Statist. Soc. B-74* (1), 111-132.
- Sang, H., Jun, M., Huang, J.Z., 2011. Covariance approximation for large multivariate spatial data sets with an application to multiple climate model errors. *Ann. Appl. Statist.* 5 (4), 2519-2548.
- Schaffrin, B., 1985. The geodetic datum with stochastic prior information (in German). Publication of the German Geodetic Commission C-313, Munich, Germany.
- Schaffrin, B., 1993. Biased Kriging on the sphere? In: Soares, A. (Ed.), *Geostatistics Troia 92*, Vol. I. Kluwer, Dordrecht, The Netherlands, pp. 121-131.
- Schaffrin, B., 2000a. Minimum mean square error adjustment, Part 1: The empirical BLE and the repro-BLE for direct observations. *Journal of the Geodetic Society of Japan* 46 (1), 21-30.
- Schaffrin, B., 2000b. On the reliability of data obtained by Kriging. In: Jan Beek, K., Molenaar, M. (Eds.), *International Archives of Photogrammetry and Remote Sensing*, Vol. XXXIII, Part B4. Amsterdam, The Netherlands, pp. 893-900.
- Schaffrin, B., 2001. Equivalent systems for various forms of Kriging, including least-squares collocation. *Zeitschrift für Vermessungswesen* 126 (2), 87-93.
- Schaffrin, B., 2008. Minimum mean squared error (MSE) adjustment and the optimal Tykhonov-Phillips regularization parameter via reproducing best invariant quadratic uniformly unbiased estimates (repro-BIQUE). *J. Geodesy* 82 (2), 113-121.
- Skiena, S.S., 1997. *The Algorithm Design Manual*. Springer-Verlag, New York.
- Vetter, P., Schmid, W., Schwarze, R., 2014. Efficient approximation of the spatial covariance function for large data sets - Analysis of atmospheric CO<sub>2</sub> concentrations. *J. Environ. Statist.* 6 (3), 1-36.