講演要旨

Correcting the Smoothing Effect of Ordinary Kriging Estimates

Jorge Kazuo Yamamoto*

*Department of Environmental and Sedimentary Geology, Institute of Geosciences, University of
Sao Paulo, Rua do Lago, 562, CEP 05.508-900, Sao Paulo, SP, Brazil. E-mail: jkyamamo@usp.br.

Key words: Smoothing effect, Global accuracy, Interpolation variance, Ordinary kriging

1. INTRODUCTION

Ordinary kriging has been established as a reliable method of data interpolation because it is an exact estimator and provides estimates with minimum error variance in the least square sense. These characteristics guarantee local accuracy for kriging estimates. However, ordinary kriging estimates present a serious drawback well known by geostatisticians as the smoothing effect in which small values are usually overestimated and large values underestimated. In this paper, we present an approach for correcting the smoothing effect of ordinary kriging estimates, which was derived partially from that one suggested by Olea and Pawlowsky (1996).

2. MEASURING UNCERTAINTY ASSOCIATED WITH ORDINARY KRIGING ESTIMATES

The ordinary kriging estimate ($Z^*(x_o)$) results from the following weighted average of neighboring data values:

$$Z_{OK}^*(x_o) = \sum_{i=1}^n \lambda_i Z(x_i) \quad (1)$$

where $\{\lambda_i, i = 1, n\}$ are the kriging weights. A reliable measurement of uncertainty can be derived from ordinary kriging weights, which was named as interpolation variance (Yamamoto, 2000) as follows:

$$S_o^2 = \sum_{i=1}^n \lambda_i [Z(x_i) - Z^*(x_o)]^2 \quad (2)$$

Through cross-validation it is possible to obtain for each sampling point the true error:

$$TrueError = [Z^*(x_o) - Z(x_o)] \quad (3)$$

where $Z(x_o)$ is the actual value of the random variable. Note that the true error can be positive if the estimate value is greater than actual value and negative if vice-versa.

3. SMOOTHING EFFECT

Estimates based on weighted average formula such as ordinary kriging (1) present a reduced variability, which is often referred to as

the smoothing effect, according to Isaaks and Srivastava (1989, p. 420). The smoothing effect is due to a deficit of variance of the kriging estimator and for simple kriging (SK) is precisely equal to the kriging variance as shown by Journel, Kyriakidis, and Mao (2000, p. 791):

$$Var\{Z(x)\} - Var\{Z_{SK}^*(x)\} = \sigma_{SK}^2(x).$$

Similarly, the deficit of variance for ordinary kriging (OK) can be derived from the following expression (Yamamoto, 2000, p. 507):

$$Var\{Z(x)\} - Var\{Z_{OK}^*(x)\} = \sigma_{OK}^2(x) + 2\mu \geq 0,$$

or in terms of the interpolation variance as:

$$Var\{Z(x)\} - Var\{Z_{OK}^*(x)\} = E\{S_o^2\} \geq 0$$

4. CORRECTION FOR SMOOTHING EFFECT

The proposed procedure for correcting the smoothing effect is based on a four-step approach:

a) Run the cross-validation procedure in order to achieve for each data point the interpolation standard deviation (square root of expression (2)) and the true error (3). These variables are then transformed into another variable named number of interpolation standard deviations as follows:

$$N_{S_o} = \frac{-TrueError}{S_o} \quad \text{with } S_o \neq 0 \quad (4).$$

Note that N_{S_o} is either positive or negative. The minus sign makes

compensation of the smoothing effect. If the true error is positive it means that the estimated value is greater than the actual value and, therefore, an error quantity must be subtracted. On the other hand, a negative true error means that the estimated value is less than the actual value and so an error quantity must be added.

b) Run the ordinary kriging procedure using the number of standard deviations (N_{S_o}) as variable in order to have it computed in all nodes

of the grid. It is important to note that this estimation is made using a minimum neighborhood, in order to avoid smoothing effect over N_{S_o} .

c) Run the ordinary kriging procedure using the original variable $Z(x_i)$.

d) The estimated value $Z_{OK}^*(x_o)$ is then corrected:

$$Z_{OK}^{**}(x_o) = Z_{OK}^* + N_{S_o}(x_o) * S_o \quad (5),$$

as a post-processing step. If the absolute value of $N_{S_o}(x_o)$ is greater than a limit (e.g. 3.0) then $N_{S_o}(x_o)$ is firstly corrected and after that expression (5) is applied:

$$N_{S_o} = \text{sgn}(N_{S_o}) * \text{random} * \text{limit},$$

where $\text{sgn}()$ means the sign of N_{S_o} . Actually, when N_{S_o} is too large it means that so is the uncertainty. Thus, in these cases a random N_{S_o} is generated within a predefined limit.

5. CASE STUDY

For a case study let us consider a synthetic 2-D data set comprising 48 x 42 values on a regular grid. Hereafter this data set will be used as the exhaustive data. The sample data set was drawn from the exhaustive data set using stratified random sampling scheme totalizing 224 points. Figure 1 shows histograms and semivariograms for the exhaustive data set and for sample data.

The processing parameters were as follows: cross-validation was processed using 8 nearest neighbor points, the N_{S_o} values were estimated in all nodes of the grid using only 4 nearest neighbor points, and for ordinary kriging the neighborhood was chosen to be composed of 4 and 20 points, in order to study the smoothing effect. A limit equal to ± 3 interpolation standard deviations was established. Examining the obtained results, it is possible to verify that the algorithm removes the smoothing effect without a loss of local accuracy, that is an important issue. Figure 2B and 3A show clearly that both the histogram and the semivariogram of original data are reproduced. On the other hand, when the searching area is enlarged the smoothing effect is increased (Figures 2C and 2D). In this case, the algorithm removes only partially the smoothing effect, because the ordinary kriging estimate is so smoothed that it is almost impossible to remove fully. Finally, in Figure 4 we have scattergrams between reference values versus corrected estimates. Both scattergrams show regression lines with slopes lesser than 1. As long as smoothing effect increases, the slope of regression line gets further from 1. Even after

removing the smoothing effect, there is a residual conditional bias (Fig. 4A). However, it is important to realize that the correction for the smoothing effect improved the local accuracy resulting always in smaller RMS errors.

6. CONCLUSIONS

A simple method for correcting the smoothing effect of ordinary kriging estimates is presented. Corrected estimates present local and global accuracies, which is an important issue in modern geostatistics. This is a first approach for the solution of the problem of smoothing effect of ordinary kriging estimates.

REFERENCES

- Isaaks, E.H., and Srivastava, R.M. (1989) An introduction to applied geostatistics: Oxford University Press, New York, 561 p.
- Journel, A., Kyriakidis, P.C., and Mao, S. (2000) Correcting the smoothing effect of estimators: a spectral postprocessor. *Math. Geology*, vol. 32, no. 7, pp. 787-813.
- Olea, R., and Pawlowsky, V. (1996) Compensating for estimation smoothing in kriging: *Math. Geology*, vol. 28, no. 4, pp. 407-417.
- Yamamoto, J. K. (2000) An alternative measure of the reliability of ordinary kriging estimates: *Math. Geology*, vol. 32, no. 4, pp. 489-509.

Figure 1: Histograms for the exhaustive data set (A) and for sample data set (B); semivariograms for the exhaustive data set (full symbols) and for the sample data set (empty symbols). Legend: square = N135° and circle = N45°. Thick lines represent the semivariogram model for the exhaustive data set.

Figure 2: Histograms for ordinary kriging estimates (A) and corrected estimates (B) using 4 nearest points and the same using 20 nearest points (respectively, C and D).

Figure 3: Semivariograms for ordinary kriging estimates (empty symbols) and for corrected estimates (full symbols) using 4 nearest points (A) and using 20 nearest points (B).

Figure 4: Scattergrams between reference values versus corrected estimates using 4 nearest points (A) and using 20 nearest points (B). Black regression lines refer to corrected estimates and red ones to OK estimates.