Properties and Applications of the Interpolation Variance Associated with Ordinary Kriging Estimates

Jorge Kazuo Yamamoto and Marcelo Monteiro da Rocha

Department of Environmental and Sedimentary Geology, Institute of Geosciences, University of Sao Paulo, Brazil.

Abstract. This paper shows some properties and applications of the interpolation variance. The interpolation variance is a reliable alternative to the kriging variance. The interpolation standard deviation presents correlation with estimates and also spatial correlation following the spatial pattern of the original data. Moreover, it can also be used for computing the global estimation variance associated with the average value of ordinary kriging estimates. Based on this uncertainty measure we have developed numerical procedures for correcting the smoothing effect of ordinary kriging estimates and backtransforming lognormal kriging estimates. Corrected or backtransformed estimates do not present any bias when compared to classical solutions. Properties and applications of the interpolation variance are illustrated on three data sets presenting different frequency distributions: lognormal, normal and negative skewness.

Keywords: kriging variance, interpolation variance, smoothing effect, ordinary kriging

1. Introduction

The kriging variance was proposed as an uncertainty measure associated with ordinary kriging estimates. However, it gives just the spatial configuration of neighbor data used to estimate an unsampled location (Journel and Rossi1). It means if the same configuration is used for distinct locations, then the kriging variance will be exactly the same. Actually, the kriging variance is known to be homoscedastic, because it does not depend on data values and just on semivariogram model. One of the most common applications of the kriging variance was ore reserve classification. Fortunately, ore reserve classification proposals based on the kriging variance have been abandoned mainly after Armstrong2.

An alternative measure was proposed by Yamamoto3 and named as interpolation variance. This uncertainty measure uses both the kriging weights and data values and, therefore, it is heteroscedastic. In this paper we want to show how it does work, its properties and some applications.

2. Uncertainty measures associated with ordinary kriging estimates

The ordinary kriging estimator is given as:

\[ Z_{\text{OK}}^* (x_o) = \sum_{i=1}^{n} \lambda_i Z(x_i) \] (1)

The kriging variance is computed as:

\[ \sigma_{\text{OK}}^2 = \sum_{i=1}^{n} \lambda_i \gamma(x_i - x_o) + \mu \] (2)

Where \( \gamma(x_i - x_o) \) is the semivariance for a distance \( (x_i - x_o) \) as measured on the semivariogram model and \( \mu \) is the Lagrange multiplier.

The interpolation variance is simply:

---

* Corresponding author. Tel.: +5511-3091-4203; fax: +5511-3091-4207.
E-mail address: jkyamamo@usp.br.
The ordinary kriging weights must be all positive otherwise the interpolation variance can result negative that is unacceptable. Thus, it is necessary to post-process ordinary kriging weights in order to constrain them to be all positive. For details, please see Yamamoto\textsuperscript{3} (p.493-495).

This simple expression can be used for both point and block kriging (see proof in Yamamoto\textsuperscript{3}, p.507-509). Actually, we want to show how reliable is this expression as a measure of uncertainty associated with ordinary kriging estimates.

3. Computation of the global estimation variance

The global estimation variance of a deposit is equal to the sum of variance within blocks and variance between blocks, in which the variance within blocks is simply the average of interpolation variances computed for each block (Yamamoto\textsuperscript{4}):

\[ S_D^2 = \frac{1}{N} \sum_{i=1}^{N} \lambda_i \left[ Z(x_i) - Z_{OK}^* (x_o) \right]^2 \tag{3} \]

Note that the variance between blocks is less than the sample variance due to the smoothing effect of ordinary kriging estimates. This expression can be rewritten as:

\[ Var[Z(x)] = E[S_D^2] + Var[Z_{OK}^* (x)] \]

This expression shows that the sample variance is equal to the mean of interpolation variances plus the variance of ordinary kriging estimates. The mean of interpolation variance can be interpreted as the variance smoothing of the kriging estimator (Yamamoto\textsuperscript{5}, p.72). Therefore, the global estimation variance of a deposit is none other than the sample variance. As a matter of fact they are very close but not exactly equal because a bias between mean of kriging estimates and the sample mean.

The variance within blocks corrects the global variance of the smoothing effect. The post-processing algorithm proposed by Yamamoto\textsuperscript{5} makes the distribution of an error derived from the interpolation variance for each ordinary kriging estimate. Developing equation (4) we get the same formula proposed by Crozel and David\textsuperscript{6}:

\[ S_D^2 = \sum_{\alpha=1}^{N} \lambda_{\alpha} \left[ Z(x_{\alpha}) - Z_{\alpha}^* \right]^2 \]

Where \[ \lambda_{\alpha} = \frac{1}{N} \sum_{i=1}^{N} \lambda_i \] is the global weight attached to the \[ \alpha_{th} \] sample obtained from successive block krigings, according to Crozel and David\textsuperscript{6}. Details of this development can be found in Yamamoto\textsuperscript{4}.

4. The smoothing effect of ordinary kriging

The most important application of the interpolation variance concerns correcting the smoothing effect of ordinary kriging estimates (Yamamoto\textsuperscript{5}). The unknown actual value is equal to the ordinary kriging estimate plus an error term:

\[ Z(x_o) = Z_{OK}^* (x_o) + \gamma(x_o) \]

The error term can be either simulated or estimated. Therefore, we must compute the error term to be added to every ordinary kriging estimate without loss of local accuracy. In simulation the error term is chosen randomly (Monte Carlo). The solution proposed by Yamamoto\textsuperscript{5} is based on post-processing ordinary kriging estimates. The most important information comes from running the cross-validation on sample data. In the cross-validation procedure we have for each data point, the estimated and actual values, as well as the interpolation standard deviation. Combining these values we get another variable which was named number of interpolation standard deviations, as follows:

\[ TrueError(x_o) = z_{OK}^* (x_o) - z(x_o) \]
This new variable is estimated at every grid node using the same parameters for estimation of the variable under study. It means that after ordinary kriging we have for each node, the ordinary kriging estimate, the interpolation standard deviation and the number of interpolation standard deviations. The error term is simply the number of interpolation standard deviations times the interpolation standard deviation. Adding this term to the ordinary kriging estimate we have the corrected estimate as:

\[ z^{**}_{OK} (x_o) = z^{*}_{OK} (x_o) + N_{S_e} \sigma_o \]

This is the basic idea behind post-processing ordinary kriging estimates. Further details of this algorithm can be found in Yamamoto\(^5\) and\(^7\).

5. Materials and methods

Three synthetic data sets will be considered to show some properties and applications of the interpolation variance. Exhaustive data sets were generated in computer presenting different frequency distributions: lognormal (positive skewness), normal and a distribution with negative skewness. The exhaustive data sets are composed of 2500 values distributed on 50 x 50 grid nodes. From these exhaustive data sets 121 data points were drawn based on stratified random sampling method. Therefore, these samples represent 4.84% of the exhaustive information. Hereafter, these samples will be named as Zlog, Zgauss and Zinv. Summary statistics are presented in Table 1.

Exhaustive and sample data sets are materials to be considered in this paper. First, unsampled locations will be estimated based on ordinary kriging technique. Kriging and interpolation variances will be computed and compared with resulting estimates. Ordinary kriging estimates will be post-processed for correcting the smoothing effect according to Yamamoto\(^5\). Moreover, this algorithm will be also applied for back-transforming lognormal kriging estimates (Yamamoto\(^7\)).

Table 1: Summary statistics for samples Zlog, Zgauss and Zinv.

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Zlog</th>
<th>Zgauss</th>
<th>Zinv</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nb. of data</td>
<td>121</td>
<td>121</td>
<td>121</td>
</tr>
<tr>
<td>Mean</td>
<td>1.599</td>
<td>15.275</td>
<td>52.723</td>
</tr>
<tr>
<td>Variance</td>
<td>3.932</td>
<td>19.369</td>
<td>3.932</td>
</tr>
<tr>
<td>Std dev.</td>
<td>1.983</td>
<td>4.401</td>
<td>2.024</td>
</tr>
<tr>
<td>Coeff. Var.</td>
<td>1.240</td>
<td>0.288</td>
<td>0.038</td>
</tr>
</tbody>
</table>

6. Results and discussion

Results of ordinary kriging and corrected estimates and uncertainties for samples Zlog, Zgauss and Zinv are presented in Figures 1, 2 and 3.

For lognormal data (Figure 1) it is clear the positive linear relationship between ordinary kriging estimates and interpolation standard deviation. However, there is no relationship with kriging standard deviation, because it does not depend on data values. This positive correlation is known as proportional effect (Isaaks and Srivastava\(^8\)).

For normal data (Figure 2) estimates and associated errors do not present any linear relationship. This is what happens with normal data, in which there is no relationship between estimates and errors. The kriging standard deviation image shows some repetitive patterns because it is low around data points and high when it gets further.

Figure 3 shows a very interesting relationship. There is a negative linear relationship between ordinary kriging estimates and interpolation variances. That is, when estimates are high the interpolation variances are low and vice versa. Besides, errors keep higher continuity along 135° direction.

This is an important property of the interpolation standard deviation because it gives a connection between the data distribution and behavior of errors and estimates. For normal data (no skewness), errors and estimates do not present any correlation. Usually the correlation coefficient for normal data is around zero. Exceptionally in this case the correlation coefficient was equal to -0.241.
Figure 1: Scattergrams between kriging standard deviation (A) and between interpolation standard deviation and estimates (B); Images for kriging standard deviation (C) and for interpolation standard deviation (D) for lognormal data (Zlog).

Figure 2: Scattergrams between kriging standard deviation (A) and between interpolation standard deviation and estimates (B); Images for kriging standard deviation (C) and for interpolation standard deviation (D) for normal data (Zgauss).
Now, let us examine some global statistics derived from ordinary kriging and corrected estimates (Table 2).

Table 2: Global estimation variance for ordinary kriging and variance for corrected estimates.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Ordinary Kriging</th>
<th>Corrected OK</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$E[Z_{OK}^*]$</td>
<td>$E[Z_{**}^*]$</td>
</tr>
<tr>
<td>Zlog</td>
<td>1.571</td>
<td>1.599</td>
</tr>
<tr>
<td>Zgauss</td>
<td>15.226</td>
<td>15.275</td>
</tr>
<tr>
<td>Zinv</td>
<td>52.686</td>
<td>52.723</td>
</tr>
</tbody>
</table>

Comparing means and variances with sample means and sample variances we verify that ordinary kriging means and global estimation variances are very close to the sample mean and sample variance. However, the same statistics for corrected estimates are much better and the variance is much closer than the global estimation variance. In other words, the variance of corrected estimates is equal to the global estimation variance. Besides, it reconfirms the effectiveness of the post-processing algorithm for correcting the smoothing effect that distributes the smoothing variance for all corrected estimates.

Another property of the interpolation variance is that it presents some spatial correlation since it depends on data values. Thus, we can compute semivariograms for kriging and interpolation standard deviations as shown in Figure 4 to 6.

As we can see on Figures 4 to 6 the kriging standard deviation does not reproduce the spatial correlation of original data, whereas the interpolation standard deviation does. The range for these semivariograms is always less than the original ones because of the repetitive patterns at short distances.

Now, let us explore more interesting application that comes for correcting the smoothing effect of ordinary kriging estimates. This correction comes at no extra cost that is without loss of local accuracy as happens with geostatistical simulations. Actually, corrected estimates reproduce both histogram and semivariogram while keeping local accuracy.
Corrected estimates always present better correlation with actual data (Table 3). Moreover, corrected estimates for normal and negative skewness data do not present conditional bias as ordinary kriging does.

Table 3: Correlation coefficients between actual and estimated values.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Ordinary kriging</th>
<th>Backtransformed lognormal kriging</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Smoothed</td>
<td>Corrected</td>
</tr>
<tr>
<td>Zlog</td>
<td>0.894</td>
<td>0.915</td>
</tr>
<tr>
<td>Zgauss</td>
<td>0.919</td>
<td>0.941</td>
</tr>
<tr>
<td>Zinv</td>
<td>0.941</td>
<td>0.962</td>
</tr>
</tbody>
</table>

For the lognormal data, there is a conditional bias. Thus, in order to reduce the bias, we ran lognormal kriging based on Journel\(^9\). Then, lognormal kriging estimates are backtransformed to the original scale of measurement using either the classical formula (Journel\(^9\)) or that one proposed by Yamamoto\(^7\). Backtransformation call for a non-bias term to be added to lognormal kriging estimate before taking its inverse logarithm. Journel\(^9\) proposed to use half of the kriging variance plus the Lagrange multiplier as the non-bias term. A completely different approach was proposed by Yamamoto\(^7\), in which the non-bias term is simply an amount that corrects for the smoothing effect of lognormal kriging estimates.
lognormal kriging estimates after Yamamoto\textsuperscript{7} provided much better correlation with actual data and with a very low conditional bias (Figure 7).

![Figure 7: Scattergrams between actual and backtransformed lognormal kriging estimates: (A) after Journel\textsuperscript{9} and (B) after Yamamoto\textsuperscript{7} for lognormal data.](image)

7. Conclusions

The interpolation variance or its square root the interpolation standard deviation is a reliable measure of uncertainty associated with ordinary kriging estimates. This idea can be extended to any other weighted average method, such as inverse of distance. In this paper we have shown an important connection between sample distribution and correlation of errors and estimates on scattergrams. For distributions presenting positive skewness the linear correlation of errors and estimates will also be positive. On the other hand, when the distribution presents negative skewness there will be a negative linear relationship between errors and estimates. For normal data, there is no linear correlation between errors and estimates.

It was also shown the calculation of the global estimation variance from interpolation variances computed for individual blocks. However, for corrected ordinary kriging estimates the variance smoothing vanishes and the global estimation variance can be computed directly from corrected estimates.

Another property is the spatial correlation of the interpolation standard deviation following the sample data. None of mentioned properties are shared with the kriging standard deviation.

The most important application of the interpolation standard deviation comes with correcting the smoothing effect of ordinary kriging estimates and backtransforming lognormal kriging estimates. For all cases studied corrected estimates were superior to the traditional approach. It shows how reliable is the interpolation standard deviation as an uncertainty measure associated with ordinary kriging estimates. For lognormal data we have tested just one sample presenting a coefficient of variation equal to 1.24. Evidently, this kind of data deserves a more careful study considering samples with higher coefficient of variations. However, this is a subject for a further research.

8. Acknowledgements

I am very grateful to the National Council for Scientific and Technological Development – CNPq (Process 303505/2007-9) for funding this research.

9. References


