Kriging, Stationarity and Optimal Estimation: Measures and Suggestions

Jared L. Deutsch and Clayton V. Deutsch

Kriging searches are routinely limited to optimize the smoothing for volume variance relations, decrease model reliance on stationarity and decrease the computational time required. The optimal kriging search depends on the type of estimate being made, from visualization to interim to final decisions. There are a number of kriging measures to help quantify the impact of a limited search on estimates including Krige's kriging efficiency and the slope of regression. In this paper, a new statistically based measure of kriging efficiency which compares the kriging variance to the global simple kriging variance is proposed. A number of cross validation case studies demonstrate that the lowest mean squared error for estimates can be obtained for kriging with large numbers of search data. This also increases the theoretical slope of regression and kriging efficiency.

Definition and Origin

Consider estimating a random variable Z for block volume V with estimate Z_V^*. If we are estimating at this location with kriging, then there are a number of statistics that can be calculated for this estimate, including the kriging variance, kriging efficiency and slope of regression. Kriging efficiency was first introduced by Krige (1997a) as a measure of the efficiency of block estimates. The goal of this definition is to express the kriging variance *KV* (estimation variance) normalized by the variance of the true blocks *BV* as a percentage. A high efficiency means that the kriging variance is low and the variance of the block estimates is approximately equal to the variance of the true block values. A low efficiency implies a high kriging variance. The kriging variance varies from block to block, so the kriging efficiency will vary as well. The original definition of Krige's kriging efficiency (KE_{DK}), expressed as a percentage is:

$$KE_{DK}(\%) = \frac{BV - KV}{BV} \tag{1}$$

The kriging variance is σ^2_K; calculation of the kriging variance will depend on the type of kriging used. The block variance is the variance of the true block values. Current geostatistical notation for the true block variance is:

$$BV = \sigma^2 = \bar{C}(V,V) \tag{2}$$

The limiting cases for the kriging efficiency are given in terms of the dispersion variance *DV*. The dispersion variance referenced is the variance of the block estimates:

$$DV = \sigma^2_{Z_V^*} \tag{3}$$

For the duration of this note, standard geostatistical notation will be used. The limiting cases described by Krige follow (Eq. 4-7). For perfect valuations:

$$\sigma^2_K = 0,\ \sigma^2_{Z_V^*} = \sigma^2 \text{ and } KE_{DK} = \frac{\sigma^2 - 0}{\sigma^2} = 100\% \tag{4}$$

Where all blocks are valued at the global mean (global estimate of all blocks is only estimate made):

$$\sigma^2_{Z_V^*} = 0,\ \sigma^2_K = \sigma^2 \text{ and } KE_{DK} = \frac{\sigma^2 - \sigma^2}{\sigma^2} = 0\% \tag{5}$$

With *no conditional bias* for imperfect valuations:

$$\sigma^2_{Z_V^*} = \sigma^2 - \sigma^2_K \text{ and } KE_{DK} = \frac{\sigma^2 - \sigma^2_K}{\sigma^2} = \frac{\sigma^2_{Z_V^*}}{\sigma^2} \tag{6}$$

With *a conditional bias* for imperfect valuations:

$$\sigma^2_{Z_V^*} > \sigma^2 - \sigma^2_K \text{ and } KE_{DK} = \frac{\sigma^2 - \sigma^2_K}{\sigma^2} < \frac{\sigma^2_{Z_V^*}}{\sigma^2} \tag{7}$$

Krige also notes that the efficiency can be negative if the kriging variance is greater than the true block variance. When the estimation variance exceeds the block variance, Krige dubs this a kriging anomaly and states that valuing the block with the mean would be more efficient. Figure 1 is a schematic of the block and dispersion variance referenced by Krige.

Historical and Current Application

In Krige's paper introducing kriging efficiency (1997a), a short study which using both the slope of regression and kriging efficiency was conducted. A thorough review of the slope of regression and relation to conditional bias has been completed (Deutsch, 2007); the basics will be reviewed here. The slope of regression is an approximation of the conditional expectation of the true block values given the estimated values:

$$E\{Z_V \mid Z_V^* = z\} \cong a + bz \tag{8}$$

For a bivariate Gaussian relationship between the truth and estimate, the slope can be calculated for a general linear estimator:

$$b = \frac{Cov\left\{Z_V, Z_V^*\right\}}{\sigma^2_{Z_V^*}} = \frac{\sum_{i=1}^{n} \lambda_i \bar{C}(v_i, V)}{\sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j C(v_i, v_j)} \tag{9}$$

A theoretically conditionally unbiased estimator, such as simple kriging, will have a slope of regression of 1. In practice, ordinary kriging and other estimators will have a slope of regression less than 1. This implies that high grade estimates are higher than the true grades and low grade areas are undervalued due to the smoothing effect of kriging. Expressions for the slope of regression in terms of the kriging weights are available for simple and ordinary kriging (Deutsch, 2007).

One criterion for a valid kriged estimate given by Krige (1997a) is that the slope of regression be greater than 95% for conditional unbiasedness. Using ordinary block kriging, Krige found that for slopes exceeding 95% the efficiencies were generally greater than 40% and often greater than 60%. Other observations were that the efficiency improved when the variogram range increased or when simple kriging was used. These observations are expected; increased continuity in the orebody will lead to a lower kriging variance and therefore higher kriging efficiency. Similarly, simple kriging will have a lower kriging variance and higher kriging efficiency than ordinary kriging except in the case where the ordinary kriging solution degenerates to simple kriging.

The improvement in the kriging efficiency and in the slope of regression with increasing number of data points accessed was demonstrated by Krige in a small case study (1997b). When a very low number of points were used for ordinary kriging (less than approximately 15 in Krige's study), the kriging efficiency was negative and Krige indicated that the ordinary kriging solution for these blocks was anomalous and the mean should have been used instead. The efficiency and slope of regression monotonically increased with the number of data points accessed when greater than an arbitrarily low (1 or 2) number of data points were used in the solution.

Based on these observations by Krige and the form of Eq. 1, it is the authors' opinion that the kriging efficiency as defined by Krige is useful in that it rescales the kriging variance as a percentage, but it does provide any information that could not be gained by looking at the kriging variance. Expressing the kriging efficiency as a percentage would be useful for identifying blocks where the kriging variance exceeds the data variance. In addition, Krige's efficiency is not a true statistical measure of efficiency since it does not compare with a theoretically best estimator.

The criterion of the slope exceeding a set value for valid kriging parameters is still used in kriging neighbourhood analysis (KNA) (Boyle, 2010; Rivoirard, 1987; Vann et al., 2003). Vann et al. propose four criteria for evaluating the kriging neighbourhood. These criteria are to examine the 1) slope of regression, 2) weight of the mean for simple kriging, 3) distribution of kriging weights and 4) kriging variance. Equivalently, the kriging efficiency could be examined in place of the kriging variance. A map of the kriging variance, or kriging efficiency, is used to assess the estimate quality given the data density and geometry (ie: due to drillhole clustering).

In the case study by Boyle (2010) on KNA using the Jura data set containing information on heavy metal concentrations in soil for an area of Switzerland, little value was found in optimizing the search pattern based on the theoretical slope of regression. This slope was found to have little correlation to the actual slope of regression. The inclusion of data points far away from the sample location may be violating the assumption of stationarity. Given this, Boyle concluded that understanding the area geology was much more important than optimizing the search strategy through KNA to avoid over extending the search area and placing too much reliance on the assumption of stationarity.

Statistical Efficiency

Efficiency is a measure of the relative amount of effort to accomplish a task. If a different process can accomplish the same task with less effort, then it is more efficient. In classical statistics, the efficiency of a statistical quantity is defined differently depending on the property. The three most commonly considered are the efficiencies of 1) an estimator, 2) an experimental design matrix and 3) a hypothesis test. For an unbiased estimator, the efficiency is normally quantified as the variance of the estimator normalized by the minimum possible estimation variance. The minimum possible variance is determined by the sample size; the minimum possible variance will change depending on if we have 10, 100 or 1 million samples. This minimum possible variance is given by the Cramér-Rao bound which states that the minimum possible variance of the estimator is the inverse of the Fisher information matrix (Rao, 1945). The most efficient estimator (if one exists) will have an efficiency of 1. Less efficient estimators will have efficiencies in [0,1).

For example, when estimating the slope of a line passing through a series of points with independent errors, the least-squares approach is the most statistically efficient method and will therefore give the minimum possible variance.

Krige's definition of kriging efficiency differs from the classical definition in that the kriging efficiency is a measure of the variance in the true value that is not represented in the kriged estimate. In addition, the absolute minimum variance would be the global simple kriging variance. If the definition of efficiency proposed by Krige were to be reworked as a statistical measure of efficiency; it should incorporate the global simple kriging variance which is the absolute minimum kriging variance. If the kriging variance is equal to the global simple kriging variance, then the estimator is efficient and will have an efficiency of 1. If the kriging variance is larger than the global simple kriging variance, then the efficiency will be less than 1. This could also be expressed as a percent. The efficiency will vary depending on the block being estimated and data available. Note that although global simple kriging is the lowest variance *linear* estimator; there may exist lower variance *non-linear* estimators. Restricting the definition of efficiency to linear, unbiased estimators:

$$KE_{JD/CD}(\mathbf{u}) = \frac{GSKV(\mathbf{u})}{KV(\mathbf{u})} = \frac{\sigma^2_{GSK}(\mathbf{u})}{\sigma^2_K(\mathbf{u})} \tag{10}$$

Although global simple kriging is a statistically efficient unbiased linear estimator, this also forces the model to rely heavily on a strong stationarity assumption, which is unrealistic in many cases. For example, issues with an increase in the mean squared error and decrease in the actual slope of regression were encountered by Boyle (2010) when a very large search area was used. The reason for using ordinary kriging is to have a local estimate of the mean, at the expense of introducing some conditional bias by limiting the search, and therefore reduce reliance on the assumption of stationarity.

Constraining the estimator will introduce a conditional bias. Biased estimators are sometimes used in statistics because it is possible that an estimator with a small bias will have a smaller mean squared error (the mean squared error is equal to the variance plus the square of the bias). To compare efficiencies, the ratio of efficiencies is taken. This gives a measure of how much more efficient one estimator is compared to a second estimator.

For example, kriging with a more restrictive search radius will decrease the relative efficiency compared to kriging with a large search radius. The ratio of efficiencies would give a measure of how much variance is being incurred to decrease reliance on the stationarity assumption. Ultimately, knowledge of how much the decision of stationarity could be relied upon would be necessary to determine an acceptable efficiency level (McLennan, 2007).

Case Studies

A series of case studies were examined to determine the effect of different numbers of data in kriging searches on the slope of regression, Krige's efficiency and the kriging efficiency definition introduced in this paper. The first case study considers a low grade copper porphyry deposit (lgp.dat). A histogram of the 10596 copper data (average grade 0.25%), location map and variograms are shown below (Figure 2). The vertical variogram is well behaved, albeit with a significant nugget effect of 20%.

Kriging in cross validation mode considering anywhere from 5 to 100 data was performed with simple and ordinary kriging. Note that when kriging in cross validation mode, data coming from the same drillhole as the location being estimated are excluded from the estimate. The mean squared error, Krige's efficiency, theoretical slope of regression and the kriging efficiency introduced in this paper are plotted in Figure 3.

Immediately, it can be seen that increasing the number of search data for both ordinary and simple kriging results in a lower mean squared error. In kriging practice, ordinary kriging is often preferred over simple kriging because it estimates a local mean rather than relying on a global mean as simple kriging does. This can lead to lower mean squared error (MSE) estimates. The crossover, or point at which the use of ordinary kriging results in a lower MSE compared to simple kriging is observed at approximately 37 search data in Figure 3. When only a small number of search data are considered, simple kriging will be better because of the additional information of a global mean that ordinary kriging does not consider. The MSE continues to decrease for both ordinary and simple kriging, even with very large numbers of data (>60).

The shape of Krige's efficiency, the theoretical slope of regression and kriging efficiency are all similar with the exception of the simple kriging slope. Recall that for simple kriging, the slope of regression will always be 1. A key aspect of this plot is that as the number of search data increase, the mean kriging efficiency (using the measure introduced in this paper) for both ordinary and simple kriging asymptotically approaches 1. This is because the estimator variances are approaching that of the theoretically best linear estimator: global simple kriging. This contrasts with what is observed for Krige's efficiency which plateaus around a value of 0.3. This is because the kriging variance is not being compared to the best possible case but instead to the block variance.

The second case study considered was an oil sands training data set used in mine design projects at the University of Alberta. A histogram, location map and variograms of bitumen content are shown in Figure 4. As before, the number of search data considered was varied between 5 and 100 search data. The kriged MSE and measures as a function of the number of search data are plotted in Figure 5.

A similar set of conclusions can be drawn from oilsands.dat data as from the lgp.dat data. Increase the number of search data continues to decrease the MSE; ordinary kriging improves over simple kriging at approximately 25 search data. In this case, the differences between ordinary and simple kriging results are small for large number of search data. This is observed in the calculated efficiencies and MSE.

A final case study, using a zinc training data set, sem01.dat, was conducted using the same procedure (Figure 6). Very similar results are observed in this case. Here, simple and ordinary kriging yield very similar results for greater than 20 data. Modest improvements in the MSE are observed when increasing numbers of data are used.

These case studies all indicate that using more search data is beneficial in terms of reducing the mean squared error. This decrease in the MSE is reflected in the increase theoretical slope of regression and kriging efficiency. The number of search data should be increased: sufficient computing power is generally available to perform this task. Increasing the number of search data will often lead to lower MSE results. This could be confirmed by a simple cross validation script like the one used to generate these figures. The script used and test cases for oilsands.dat and lgp.dat accompanies this paper.

Framework for Optimizing a Kriging Search Strategy

Ultimately, the kriging search strategy depends on the estimate type. Estimate types can be broadly classified as 1) visualization and geological understanding, 2) interim estimates or 3) final estimates. When the goal is visualization and gaining a level of geological understanding, a very smooth map is desired since this is easier to understand. For this purpose global simple kriging (to prevent any search artifacts) could be considered. Interim estimates are used for long term planning. The goal is to anticipate the information effect for the future and consider volume variance relations. For interim estimates a restricted search to increase the variance of the estimates could be considered. This increases understanding of what variability can be expected in the future.

Final estimates come down to the decision of ore or waste. These decisions must be made with no conditional bias and should have a minimum mean squared error. From the results of the case studies presented, the practitioner should consider the use of ordinary kriging with a large number of search data. A series of cross validation case studies could be done with the data set to confirm that this was the best approach. To facilitate these studies, the new version of `kt3d`, `kt3dn`, described in Paper 403 in this report (Deutsch and Deutsch, 2012) includes an auto search optimization feature which will iteratively test a range of search data. The kriged results can then be post processed to determine a reasonable number of search data for the estimate.

Conclusion

There are a number of useful kriging measures including Krige's efficiency, the slope of regression and the statistically based kriging efficiency proposed in this paper. The proposed kriging efficiency considers the estimation variance compared to the global simple kriging variance, which is the theoretically minimum possible estimation variance. In the case studies considered, the lowest mean squared error in cross validation was normally for ordinary kriging with a very large number of search data. In this case, the kriging efficiency was close to 1. Ultimately the decision of how many search data to use and search strategy depends on the type of estimate being made.

References

Boyle, C., 2010. Kriging neighbourhood analysis by slope of regression and weight of mean; evaluation with the Jura data set. Mining Technology, 119(2): 49-58.

Deutsch, C.V., 2007. The slope of regression for kriging estimators. Centre for Computational Geostatistics, 9:311.

Deutsch, J.L. and Deutsch, C.V., 2012. A New Version of kt3d with Test Cases. Centre for Computational Geostatistics, 14:403.

Krige, D.G., 1997a. A practical analysis of the effects of spatial structure and of data available and accessed, on conditional biases in ordinary kriging. Geostatistics Wollongong '96, Vols 1 and 2, 8(1-2): 799-810.

Krige, D.G., 1997b. Block kriging and the fallacy of endeavouring to reduce or eliminate smoothing, Proceedings of the Regional APCOM, Moscow.

McLennan, J.A., 2007. The decision of stationarity, University of Alberta, Edmonton, AB, 167 pp.

Rao, R., 1945. Information and the accuracy attainable in the estimation of statistical parameters. Bull. Calcutta Math. Soc., 37: 81-91.

Rivoirard, J., 1987. Two Key Parameters When Choosing the Kriging Neighborhood. Mathematical Geology, 19(8): 851-856.

Vann, J., Jackson, S. and Bertoli, O., 2003. Quantitative kriging neighbourhood analysis for the mining geologist - A description of the method with worked case examples. 5th International Mining Geology Conference, 2003(8): 215-223.

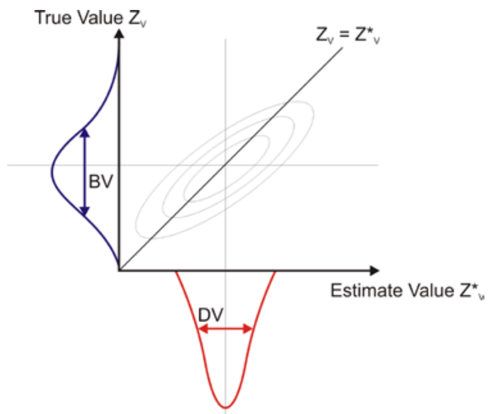

Figure 1: A schematic illustration showing Krige's dispersion variance and block variance in the context of a cross validation plot.

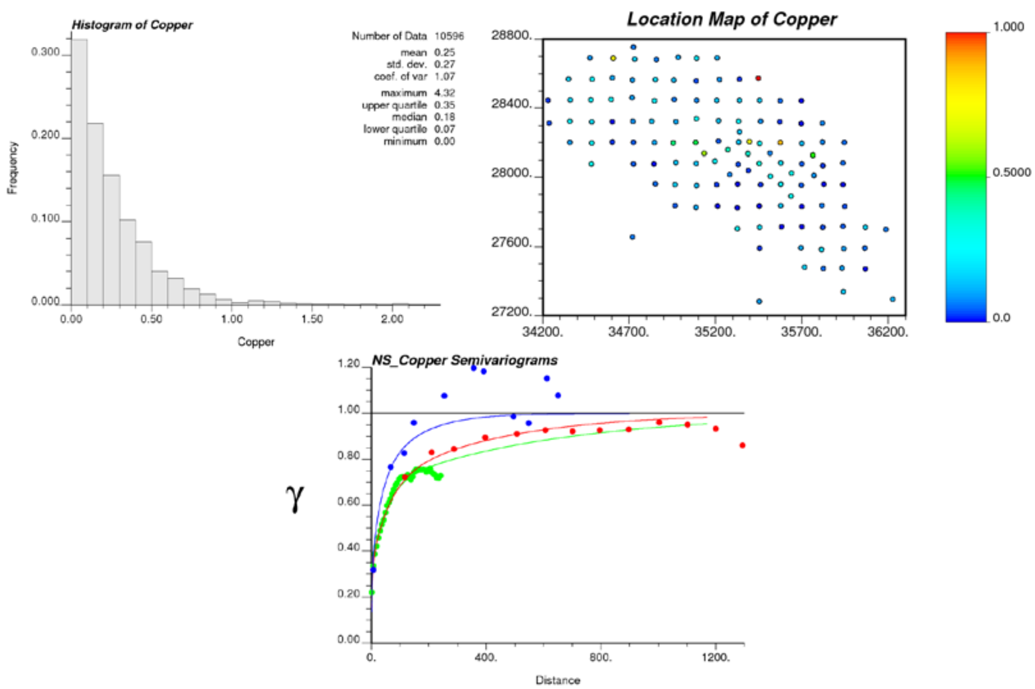

Figure 2: Histogram, location map and variograms (red = 110° horizontal, blue = 20° horizontal, green = vertical) of lgp.dat copper data.

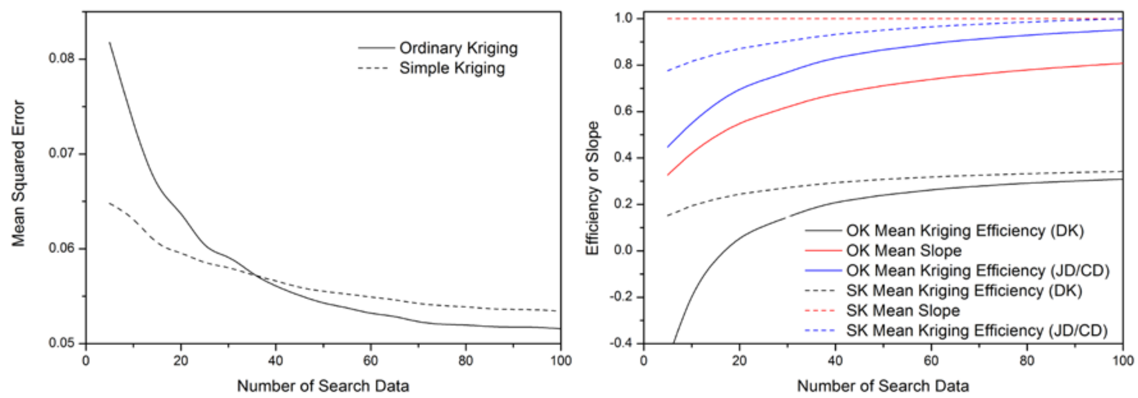

Figure 3: Effect of number of search data on simple and ordinary kriging results with lgp.dat copper data.

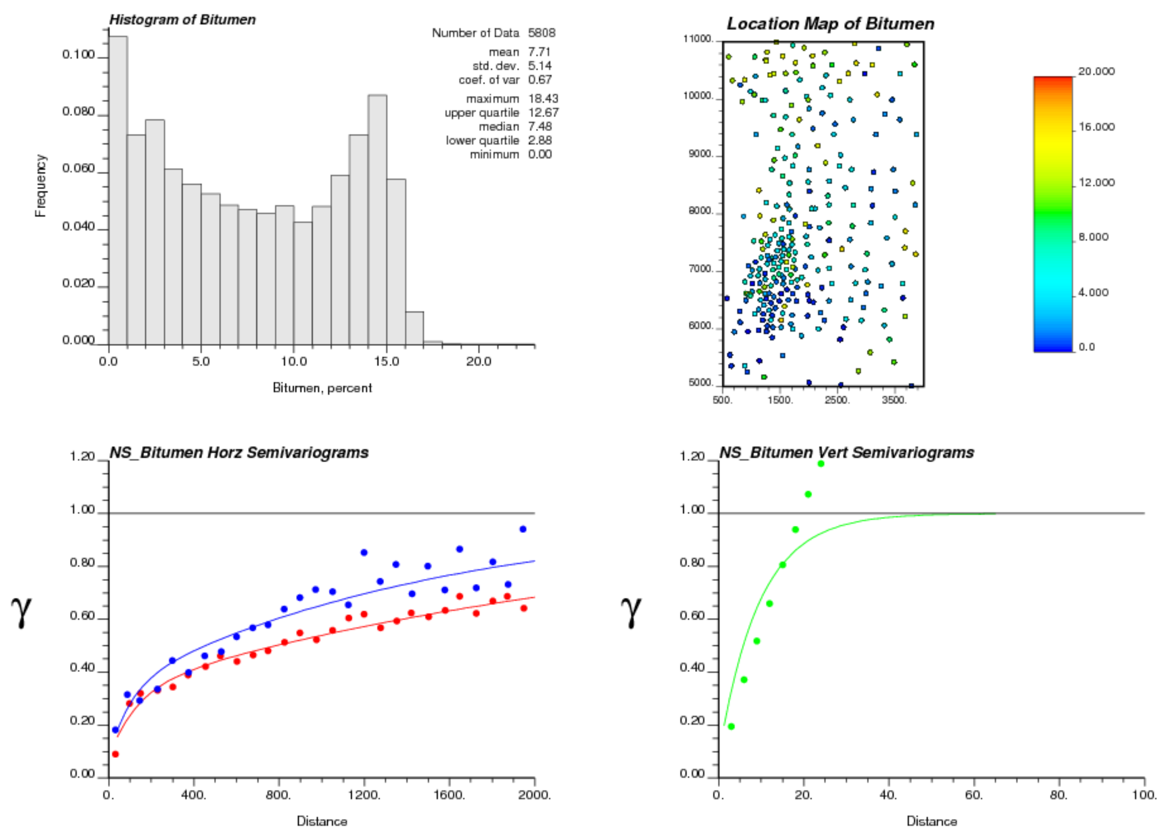

Figure 4: The histogram, location map and variograms of oilsands.dat bitumen data.

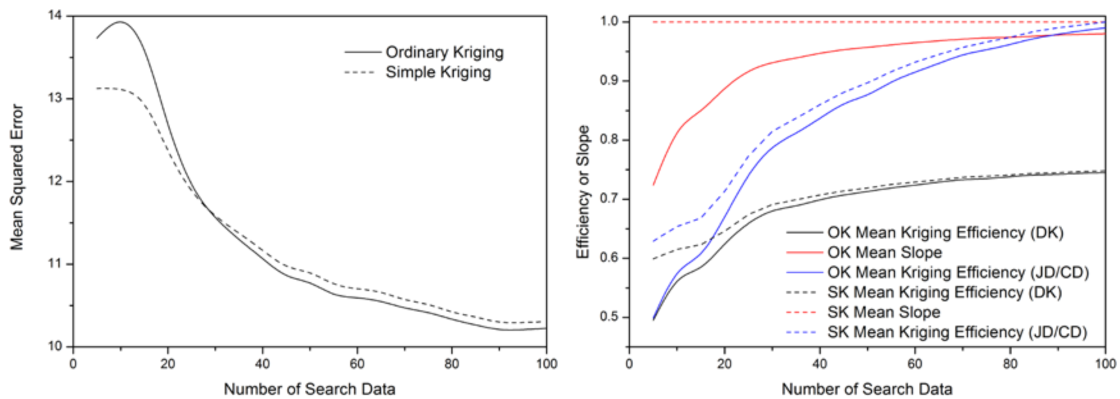

Figure 5: Average kriging results for different numbers of search data with oilsands.dat bitumen data.

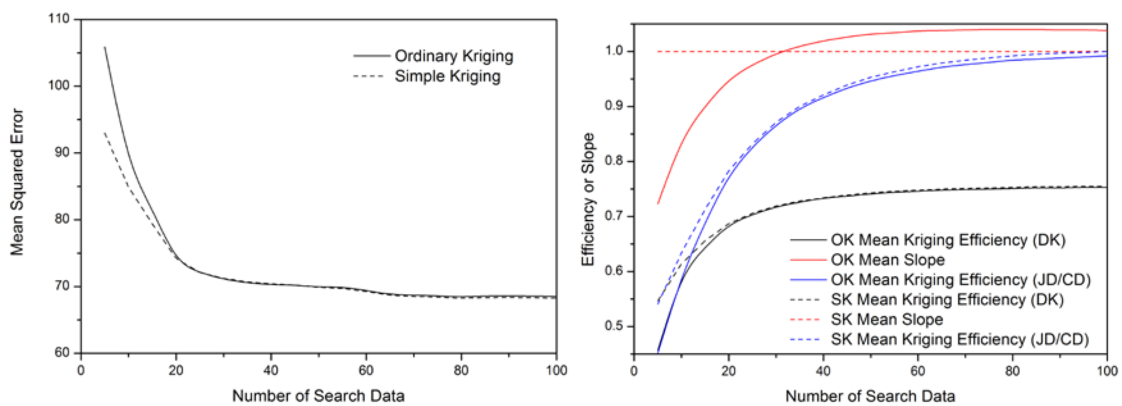

Figure 6: Mean squared error, kriging efficiency and slope of regression for simple and ordinary kriging with sem01.dat zinc data.