Uncertainty as the Overlap of Alternate Conditional Distributions

Olena Babak and Clayton V. Deutsch

Centre for Computational Geostatistics
Department of Civil & Environmental Engineering
University of Alberta

An important task in modern geostatistics is the assessment and quantification of resource and reserve uncertainty. This uncertainty is valuable decision support information for many management decisions. Uncertainty at specific locations and uncertainty in the global resource is of interest. There are many different methods to build models of uncertainty including Simple Kriging, Inverse Distance, Simple Cokriging, and so on. Each method leads to different results. We propose a method for combining local uncertainties predicted by different models to obtain a combined measure of uncertainty that, ideally, combines the good features of each alternative. The new estimator is obtained as an overlap of alternate conditional distributions.

Uncertainty as the Overlap of Alternate Conditional Distributions

Let us consider *n* data on spatial random variable $Y(u)$ at locations $u_i, i = 1, \ldots, n$ in the domain of interest *A*. We consider the problem of estimating the value of the variable of interest *Y* at an unsampled location *u** and quantification of the local uncertainty at this estimation location. Consider two different estimation techniques (e.g., Simple Kriging and Simple Cokriging) available to obtain local uncertainty at the estimation location *u**. Figure 2 shows a schematic representation of the results for the local uncertainty at location *u**.

Ideally, we would only have one distribution of uncertainty. In presence of many alternatives, we should simply choose the *best* one. *Best* is defined as the distribution with the most theoretically valid approach, the greatest fidelity with the data, the simplest to apply and so forth. In practice, however, different techniques are *best* in different senses. There may be a need to reconcile distributions from different sources that all have some legitimacy.

In the case considered above, it is known that both techniques provide theoretically valid model of uncertainty and estimate the same variable at the same estimation location; however, the results are different. Because both techniques are theoretically valid, there is a zero probability that variable of interest $Y(u)$ will be equal to some value for which one of the estimators indicates zero probability. In general, the local uncertainty of $Y(u)$ at an unsampled location *u** could be considered to be the minimum or overlap of the local uncertainties obtained by two different estimation techniques (scaled to 1). The uncertainty model obtained as an overlap of the two modeled local uncertainties is often narrow and appears as a reasonable result. The probability is the highest that the value at the unsampled location is in the interval common to both estimators. This estimator will be referred to as a overlap estimator. Figure 1 shows schematic representation of the overlap of the two the local uncertainty models at location *u**.

This approach can easily be extended to the case of several different estimators and models. In particular, Simple Kriging, Ordinary Kriging (with bias correction), Inverse Distance, if secondary information is available then also Simple Cokriging, Bayesian Updating, etc. all can be considered. Then the overlap estimator can take on arbitrary shape, see schematic examples in Figure 2.

Example with Inverse Distance and Simple Kriging

To explain the idea of the overlap uncertainty estimator, we will limit ourselves in this section to considering standard normal spatial random variable and only two techniques for its estimation, that is, Inverse Distance and Simple Kriging. A short description of each of the two techniques follows.

Inverse Distance Interpolator

An Inverse Distance (ID) weighted estimate of the variable of interest *Y* at an unsampled location *u** is a spatially weighted average of the sample values within a search neighborhood (Shepard, 1968; Franke, 1982; Diadato and Ceccarelli, 2005). It is calculated as

$$Y_{ID}^{*}(u) = \sum_{i=1}^{n} \beta_i Y(u_i), \tag{1}$$

where $\beta_i, i = 1, \ldots, n,$ are the weights assigned to each sample. The weights are determined as

$$\beta_i = \frac{\left(\frac{1}{d_i^p}\right)}{\sum_{i=1}^{n}\left(\frac{1}{d_i^p}\right)}, \quad (i = 1, \ldots, n), \tag{2}$$

where d_i are the Euclidian distances between estimation location and sample points, and exponent *p* is the power or distance exponent value.

The mean and variance of the Inverse Distance estimator $Y_{ID}^{*}(u)$ at estimation location u^* under assumption of stationarity are given by

$$E(Y_{ID}^{*}(u)) = m; \tag{3}$$

$$Var(Y_{ID}^{*}(u)) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{1}{d_i^p}\right)\left(\frac{1}{d_j^p}\right)Cov(Y(u_i), Y(u_j))}{\left[\sum_{k=1}^{n}\left(\frac{1}{d_k^p}\right)\right]^2}; \tag{4}$$

where *m* is stationary domain mean (for standard normal random variable *Y* it is zero), d_i are the Euclidian distances between estimation location and sample points, exponent *p* is the power or distance exponent value used in inverse distance interpolation and $Cov(\mathrm{u}_i, \mathrm{u}_j), \quad i, j = 1, \cdots, n,$ denotes data-to-data covariance function calculated though the semivariogram model $2\gamma(h)$.

Simple Kriging

Simple Kriging is a well established linear estimation technique that provides an estimate of the unsampled value *Y*(*u*) as a linear combination of neighboring observations $Y(\mathrm{u}_i), i = 1, \ldots, n$, accounting for the spatial continuity of the variable under study (Deutsch, 2002). The Simple Kriging estimator predicts the value of the variable of interest as

$$Z^{*}_{SK}(\mathrm{u}) = \sum_{i=1}^{n} \lambda_i Y(\mathrm{u}_i) + \left[1 - \sum_{i=1}^{n} \lambda_i\right] m, \tag{5}$$

where m denotes the stationary domain mean (for standard normal random variable *Y* it is zero), $\lambda = (\lambda_1, \ldots, \lambda_n)^T$ denotes the vector of the Simple Kriging weights calculated from the normal system of equations for the estimation location *u**,

$$\sum_{i=1}^{n} \lambda_i Cov(Y(\mathrm{u}_i), Y(\mathrm{u}_j)) = C(Y(u^*), Y(\mathrm{u}_j)), \quad j = 1, \cdots, n, \tag{6}$$

where $Cov(Y(\mathrm{u}_i), Y(\mathrm{u}_j)), \quad i, j = 1, \cdots, n,$ denotes data-to-data covariance function and $Cov(Y(\mathrm{u}^*), Y(\mathrm{u}_j)), \quad j = 1, \cdots, n,$ is data-to-estimation point covariance function calculated though the semivariogram model $2\gamma(h)$.

Simple Kriging is the best linear unbiased estimator, that is, it provides estimates with minimum error variance σ_{SK}^2 in the least square sense given by

$$\sigma_{SK}^2 = \sigma^2 - \sum_{i=1}^{n} \lambda_i Cov(Y(\mathrm{u}^*), Y(\mathrm{u}_j)), \tag{7}$$

where σ^2 is the stationary domain variance.

The mean and variance of the Simple Kriging estimator $Y_{SK}^*(u)$ at estimation location u^* are given by

$$E(Y_{SK}^*(u)) = m; \tag{8}$$

$$Var(Y_{ID}^*(u)) = \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j Cov(Y(u_i), Y(u_j)) = \sigma^2 - \sigma_{SK}^2, \tag{9}$$

where σ_{SK}^2 is Simple Kriging variance defined in (7).

Overlap Uncertainty

Let us consider 10 simulated data in the domain 20 by 20 units shown in Figure 4. The data were generated using zero nugget spherical variogram model with range of correlation equal to the size of the domain. Data distribution is also shown in Figure 3. Let us now determine the local uncertainty at the two estimation locations (8,12) and (9,9) using inverse distance interpolation and simple kriging, then subsequently calculate the overlap uncertainty estimator as the overlap of the two alternate conditional distributions. Figure 4 shows results obtained for estimation location (8,12). Looking at Figure 4 we can clearly note that the local uncertainty predicted by the inverse distance interpolation and simple kriging are quite different. As we combine both techniques to obtain the overlap uncertainty estimator, we observe that the local uncertainty by this approach is significantly narrower than predicted by other of the two estimation approaches. To quantify the change in the local uncertainty at the estimation location (8,12), we will use the interval (P10,P90). For the inverse distance estimator this interval is (-2.55,-0.92); for the simple kriging estimator it is (-1.93,-0.09) and for the overlap uncertainty estimator it is (-2.14,-0.74). So, the change is significant.

Figure 5 shows results obtained for estimation location (9,9). The local uncertainty predicted by the inverse distance interpolation and simple kriging are quite similar As we combine both techniques to obtain the overlap uncertainty, we observe that the local uncertainty by this approach is similar to local uncertainty predicted by two different estimation approaches. To quantify the change in the local uncertainty at the estimation location (9,9), we will use again the interval (P10,P90). For the inverse distance estimator this interval is (-2.66,-1.14); for the simple kriging estimator it is (-2.40,-1.15) and for the overlap uncertainty estimator it is (-2.44,-1.13). Thus, we can see that results in this case of all considered approaches are similar.

Case Study I: Cluster.dat Example

To illustrate the performance of the overlap uncertainty estimator a well known GSLIB (Deutsch and Journel, 1998) data set 'cluster.dat' is selected. The data consists of about 100 data that are sampled on a

random stratified grid and 40 data that are clustered in high valued areas. We discard the clustered data. The 2-D area of interest is 50 by 50 distance units. The distribution of data is approximately lognormal with a mean of 2.5 and a standard deviation of 5.0. The spatial continuity of the data in the normal space is described by isotropic Spherical variogram model with range of correlation 12 and nugget effect of 0.3. Figure 6 shows the location map of the 100 data and their distribution. Figure 6 also shows the same analysis but for the normal score transformed 100 data. The aim of our analysis is to establish local conditional distributions for all 100 data points in crossvalidation using Inverse Distance interpolation approach, Simple Kriging and Overlap Uncertainty method and check results. Specifically, the fairness and accuracy of the local uncertainties is checked as well as direct results of crossvalidation for the errors in estimation. All analysis is conducted in normal space.

Results

Figure 7 shows accuracy plots obtained for Inverse Distance, Simple Kriging and Overlap Uncertainty estimator prediction in crossvalidation of 100 nscore transformed data of Figure 6. From Figure 7 we can clearly see that Overlap Uncertainty estimator is both accurate and precise, thus fair, estimator of uncertainty. Both Simple Kriging and Inverse Distance are accurate, but not precise estimators of uncertainty. Figure 8 shows the crossplots between p10, p50 and p90 of Simple Kriging and Overlap Uncertainty estimator for 100 nscore transformed data. Analogous results for Inverse Distance and Overlap Uncertainty Estimator are shown in Figure 9. Figure 10 shows the p10-p90 local uncertainty intervals for the first 10 data. One can clearly note from Figures 8-10 that taking uncertainty as the overlap of the local conditional distributions from Simple Kriging and Inverse Distance (for example) can result in significant reduction of the probability intervals and, moreover, more fair local uncertainty (see Figure 7). To assess how much narrower are local conditional distributions predicted by Overlap Uncertainty versus Simple Kriging and Inverse Distance, crossplots between the variance of the local conditional distributions (smoothing effect) of the Simple Kriging, Inverse Distance and Overlap Uncertainty estimators for the 100 data of the file cluster.dat are prepared. Figure 11 shows the results. We can observe that the average variance of the local conditional distributions obtained by the Simple Kriging is 0.706, by the Inverse Distance is 0.772 and by the overlap of the local conditional distributions is 0.663. Thus, we see that the average variance of the local conditional distributions is indeed significantly smaller than that of Simple Kriging (more than 6% smaller) and that of Inverse Distance (more than 16% smaller).

Figure 12 shows results of estimates crossvalidation for Inverse Distance, Simple Kriging and Overlap Uncertainty estimator for 100 nscore transformed cluster.dat data. Table below summarizes results shown in Figure 12.

Statistics	Inverse Distance	Simple Kriging	Overlap Uncertainty Estimator
Bias	0.019	0.007	0.011
Correlation	0.481	0.547	0.538

From the table above, we can observe that all three estimators are virtually unbiased, The correlation between true values and estimates are the highest for the Simple Kriging approach, and only slightly lower is for the Overlap Uncertainty method.

Case Study II: Mine650.dat Example

The second case study is based on the data set ‘mine650.dat’. The data consists of 310 data. The 2-D area of interest is 3000 by 5000 distance units. The spatial continuity of the data is described by isotropic Spherical variogram model with range of correlation 1450 and zero nugget effect. Figure 13 shows the locations of the 310 data and their representative (declustered) distribution. Figure 13 also shows the same analysis but for the normal score transformed ‘mine650.dat’ data. The aim of our analysis is the same as before, that is, to establish the fairness and width of local conditional distributions for all data points in crossvalidation using Inverse Distance interpolation approach, Simple Kriging and Overlap Uncertainty method and check results. All analysis is conducted in normal space.

Results

Figure 14 shows accuracy plots obtained for Inverse Distance, Simple Kriging and Overlap Uncertainty estimator prediction in crossvalidation of the nscore transformed 'mine650.dat' data of Figure 13. From Figure 14 we can clearly see that all considered uncertainty estimators are accurate, but not precise estimators of uncertainty. The question is now how wide are local distributions of uncertainty predicted by each of the approaches. Figure 15 shows the crossplots between p10, p50 and p90 of Simple Kriging and Overlap Uncertainty estimator. One can see that local distributions of uncertainty are virtually the same, there are only a few data cases when the local uncertainty distributions obtained based on the Overlap Uncertainty estimator are narrower. The crossplots between p10, p50 and p90 for Inverse Distance and Overlap Uncertainty estimator are shown in Figure 16. Note that there is a huge difference in the width of the local uncertainty distributions. Local uncertainty distributions obtained by the Overlap Uncertainty estimation are significantly narrower. To summarize, we can say that basically only Simple Kriging has an impact on the local conditional distributions obtained by the Overlap Uncertainty approach. This is mainly because the respective means of the local distributions modeled by Simple Kriging and Inverse Distance are very similar, however local distributions obtained in Inverse Distance are much wider than obtained in Simple Kriging.

Conclusion

A flexible approach for combining alternate local conditional distributions to create Overlap Uncertainty estimator was proposed. Both simulated and real case studies (one with 100 data from file cluster.dat and the other one with 310 data from file 'mine650.dat') were considered. It was shown that Overlap Uncertainty estimator can result in significantly narrower intervals for the local uncertainty.

Good results are not guaranteed. One should always check whether the local distributions obtained by the approaches are accurate.

References

Deutsch, C.V., Geostatistical Reservoir Modeling, 2002.

Deutsch, CV and Journel, A.G.: GSLIB: Geostatistical Software Library and Users Guide, Oxford University Press, New York, second edition, 1998.

Diadato, N. and Ceccarelli: Interpolation Processes using Multivariate Geostatistics for Mapping of Climatological Precipitation Mean in the Sannio Mountains (southern Italy), Earth Surface Processes and Landforms, 2005.

Franke, R., Scattered Data Interpolation: Tests of Some Methods, Mathematics of Computation, 1982.

Shepard, D.: A Two-Dimensional Interpolation Function for Irregularly Spaced Data, Proc. 23rd Nat. Conf. ACM, 1968.

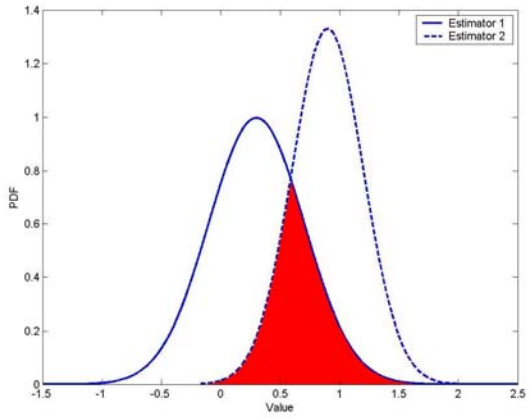

Figure 1: A schematic representation of the example results for two distributions at location *u**. Solid and dashed lines represent local uncertainty obtained by two different approaches; red area is the overlap.

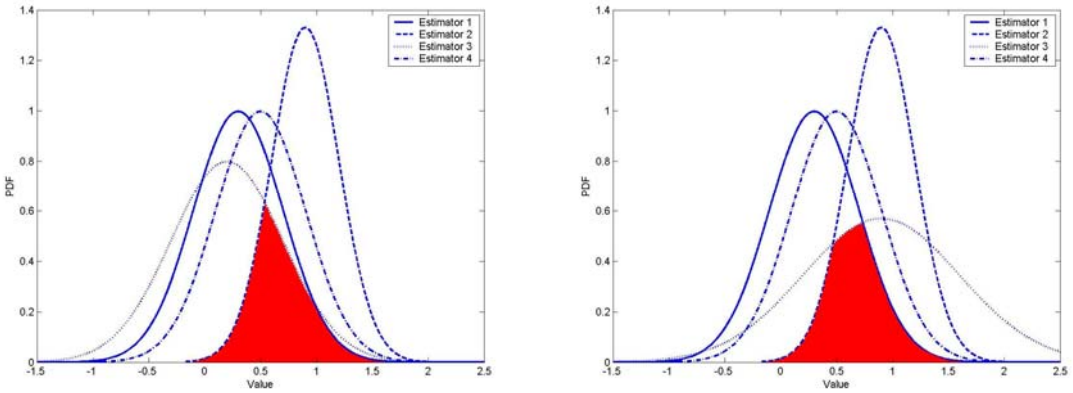

Figure 2: A schematic representation of the two example results for the four local uncertainty at location *u**.Solid dashed, doted and dash-dot lines represent local uncertainty obtained by four different estimation approaches; red area is an overlap.

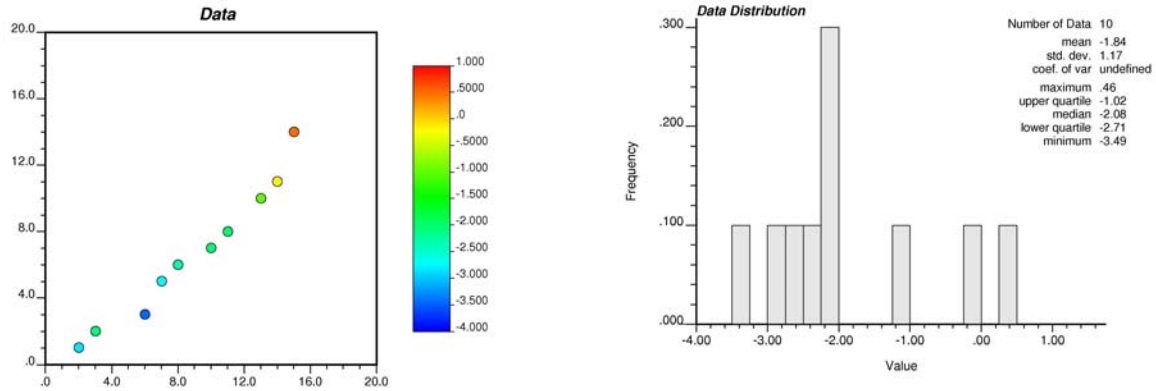

Figure 3: Location map of 10 data (left) and their distribution.

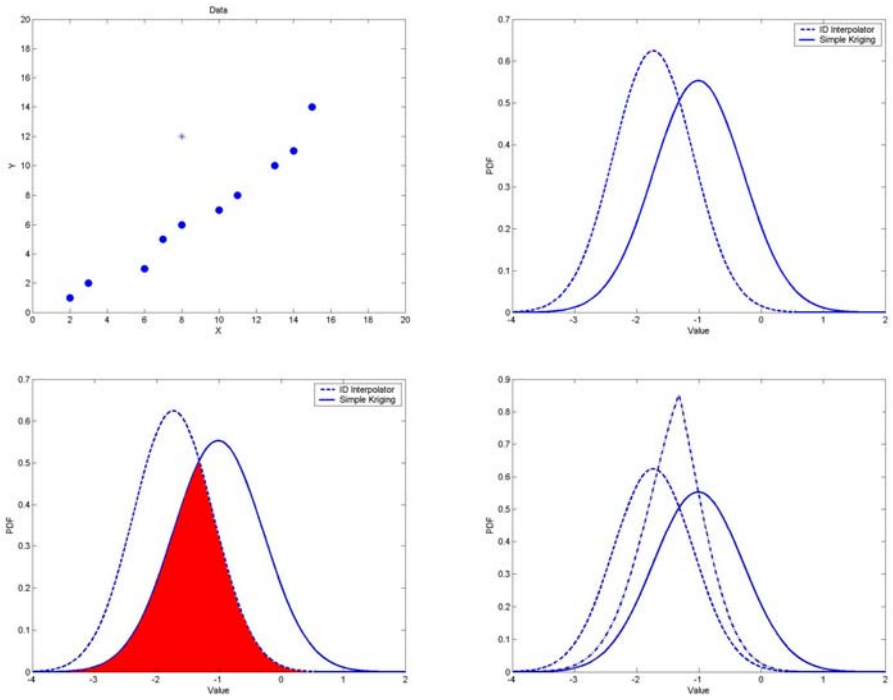

Figure 4: Location map of 10 data (circles) together with estimation location 1 (8,12) (asterisk) (top left), result of the Inverse Distance and Simple Kriging for the local uncertainty at the estimation location 1 (top right), not scaled Overlap Uncertainty estimator together with Inverse Distance and Simple Kriging local uncertainty models (bottom left), and (scaled).

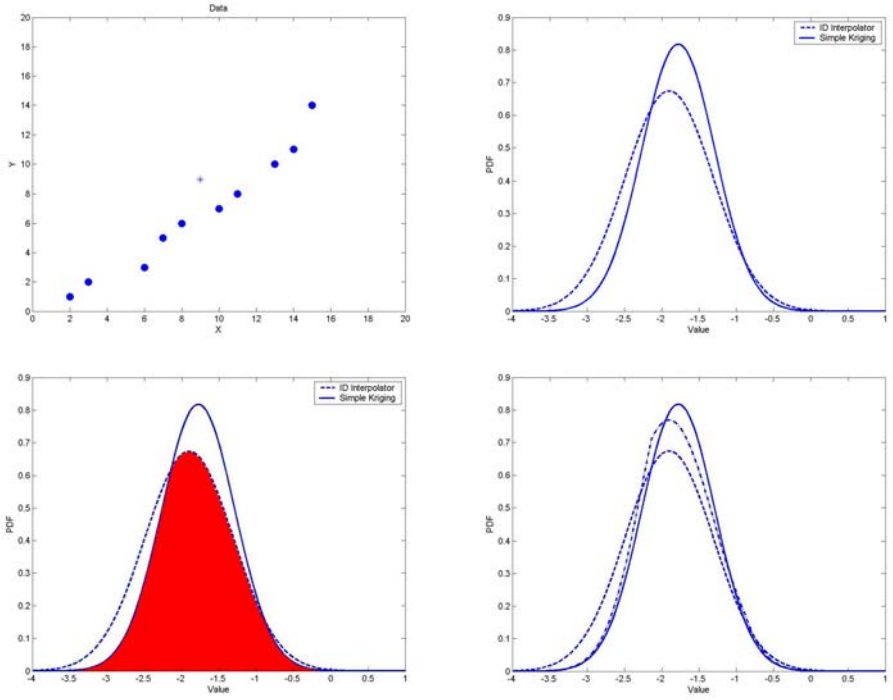

Figure 5: Location map of 10 data (circles) together with estimation location 2 (9,9) (asterisk) (top left), result of the Inverse Distance and Simple Kriging for the local uncertainty.

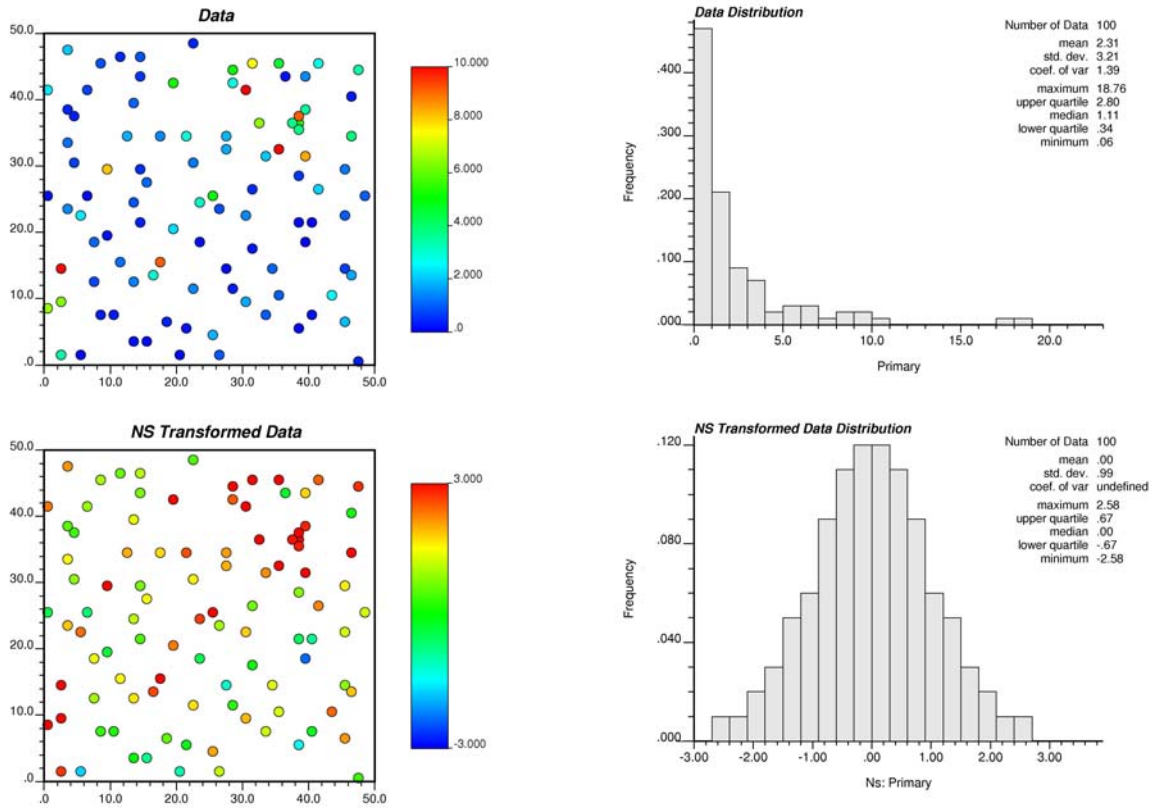

Figure 6: Location map of 100 data from file cluster.dat (top left) and their distribution (top right); Location map of the 100 normal score transformed data from file cluster.dat (bottom left) and their distribution in normal space (bottom right).

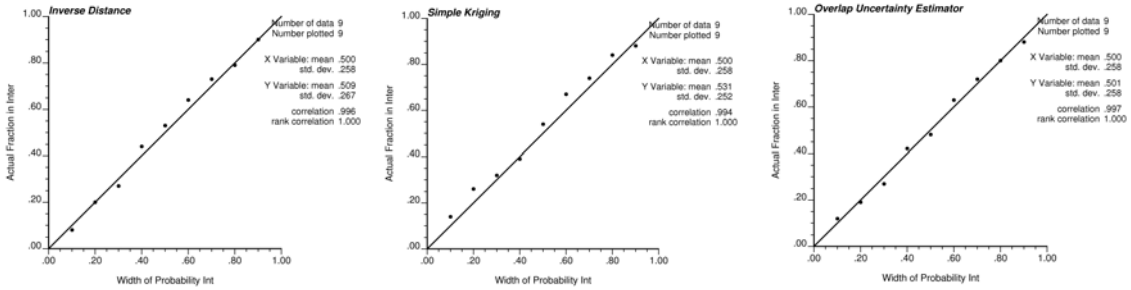

Figure 7: Accuracy plots for Inverse Distance, Simple Kriging and Overlap Uncertainty estimator prediction in crossvalidation of 100 nscore transformed primary data from file cluster.dat.

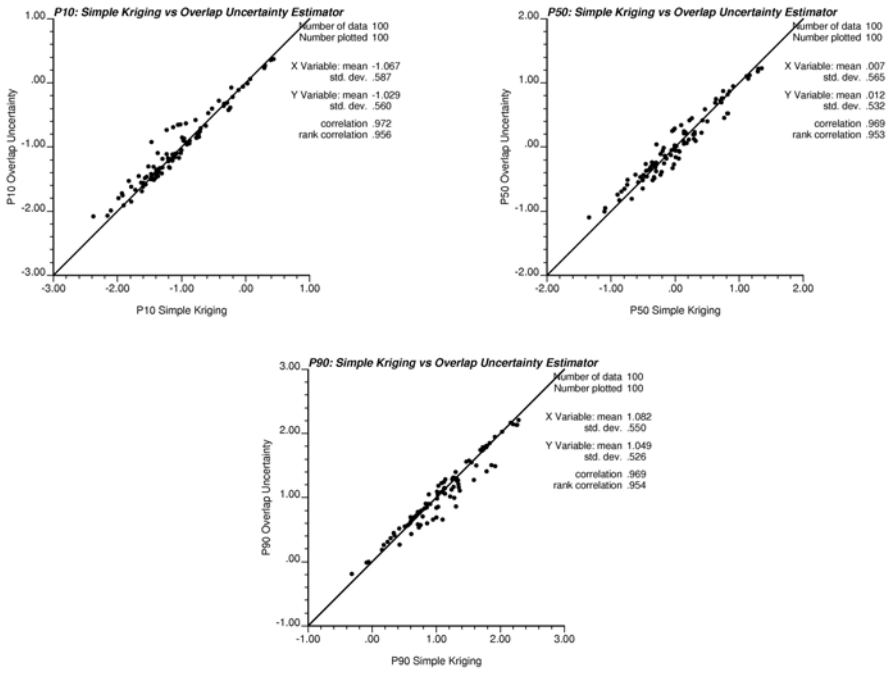

Figure 8: Crossplot between p10 (top left), p50 (top right) and p90 (bottom) of Simple Kriging and Overlap Uncertainty Estimator for 100 nscore transformed primary data from file cluster.dat.

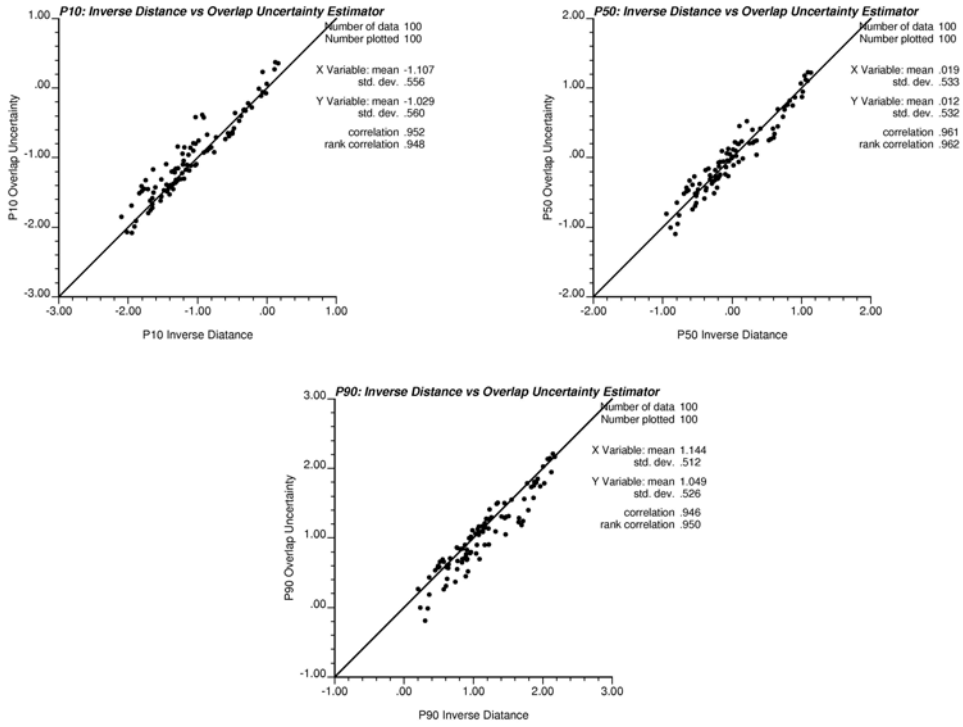

Figure 9: Crossplot between p10 (top left), p50 (top right) and p90 (bottom) of Inverse Distance and Overlap Uncertainty Estimator for 100 nscore transformed primary data from file cluster.dat.

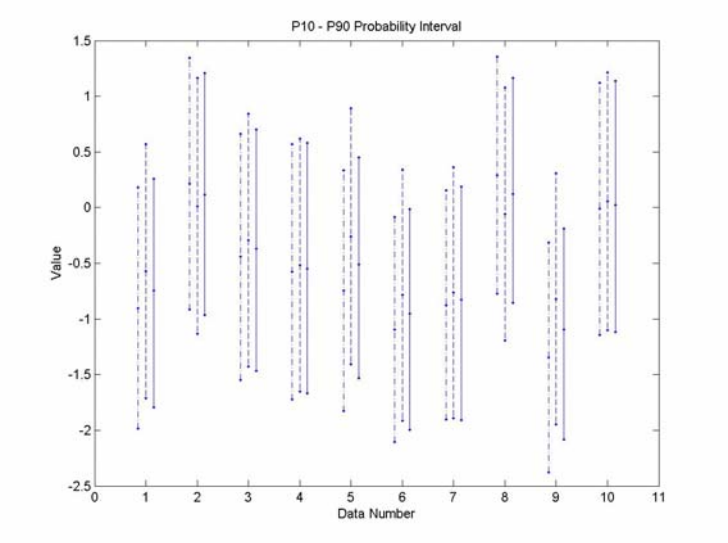

Figure 10: p10-p90 probability intervals obtained for the first 10 data in the cluster.dat data set based on Simple Kriging (dash-dot lines), Inverse Distance interpolation (dashed lines) and Overlap Uncertainty Estimator (solid lines). Medians (p50) for each of the three considered approaches are shown by dots.

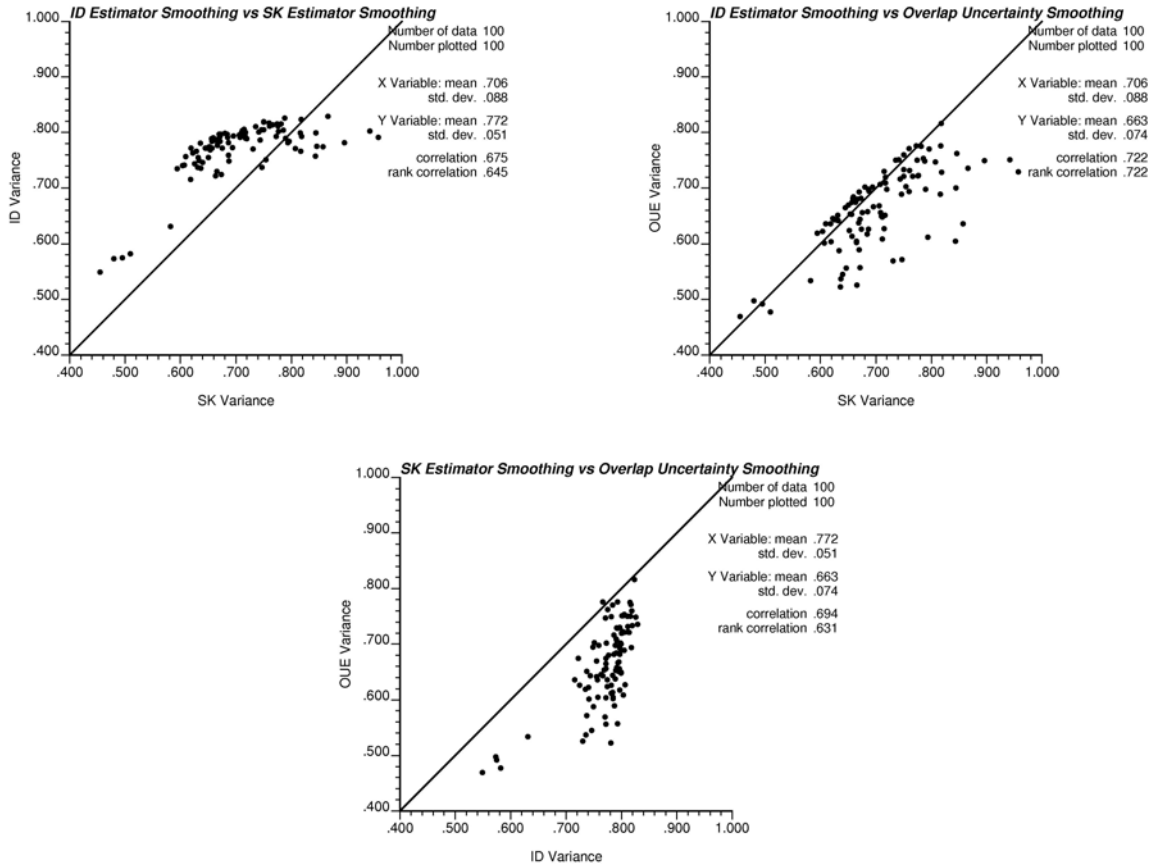

Figure 11: Crossplots of between the variance of the local conditional distributions (smoothing effect) of the Simple Kriging, Inverse Distance and Overlap Uncertainty estimators obtained for the 100 data of the file cluster.dat.

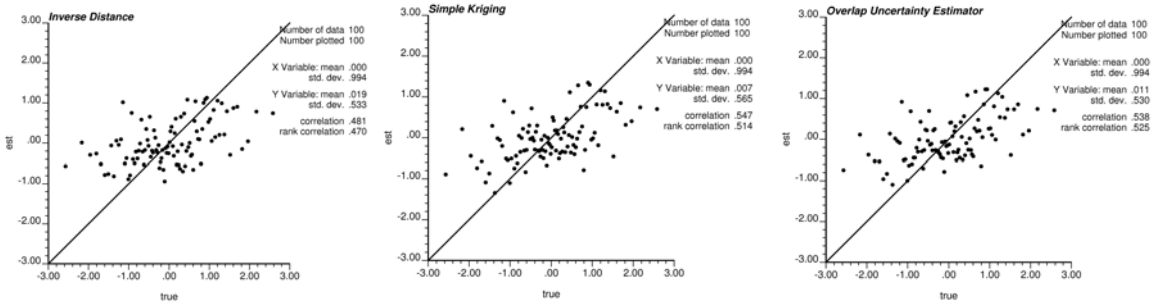

Figure 12: Crossvalidation result for Inverse Distance (top left), Simple Kriging (top right) and Overlap Uncertainty estimator (bottom) for 100 nscore transformed primary data from file cluster.dat.

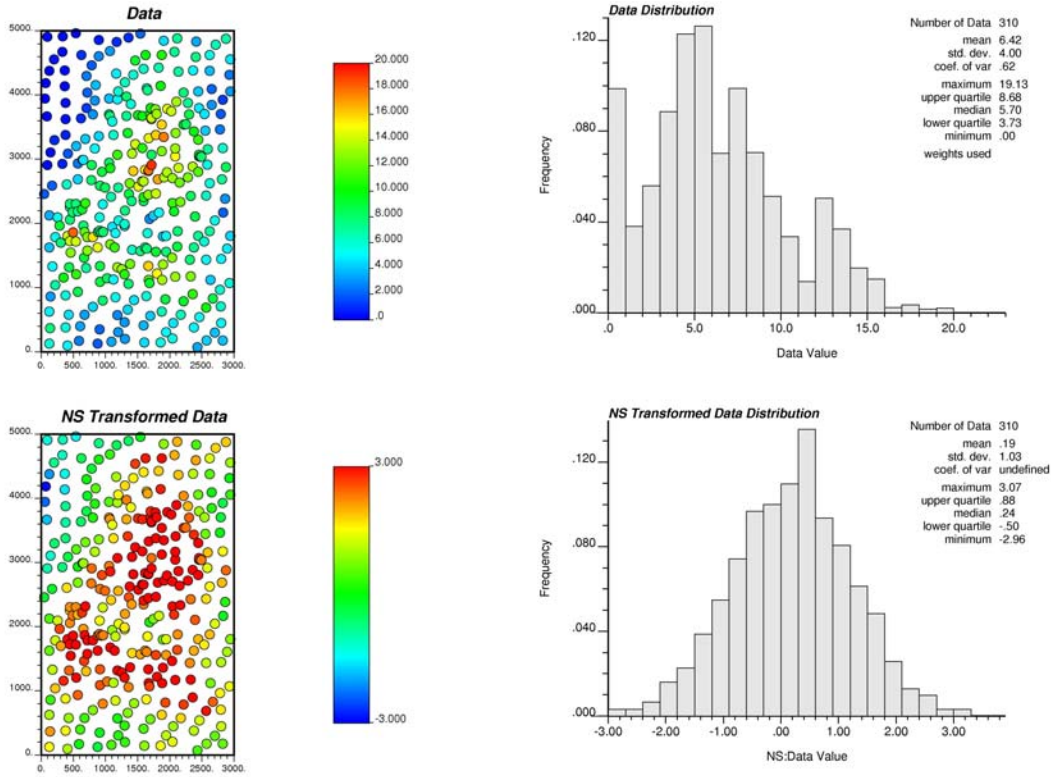

Figure 13: Location map of the data from file 'Mine650.dat' (top left) and their representative distribution (top right); Location map of the normal score transformed data (bottom left) and their distribution in normal space (bottom right).

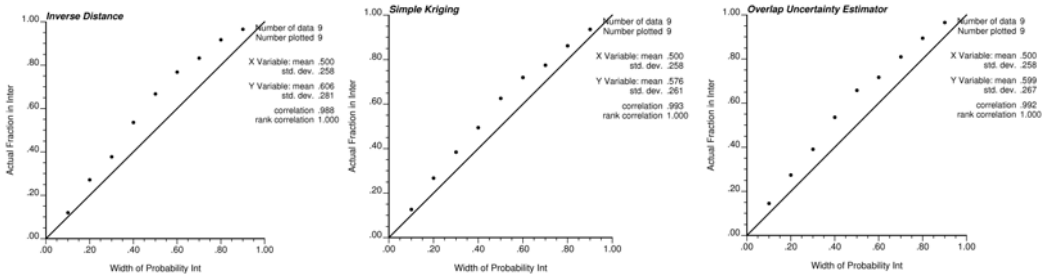

Figure 14: Accuracy plots for Inverse Distance (top left), Simple Kriging (top right) and Overlap Uncertainty estimator (bottom) prediction in crossvalidation of nscore transformed data from file 'Mine650.dat'.

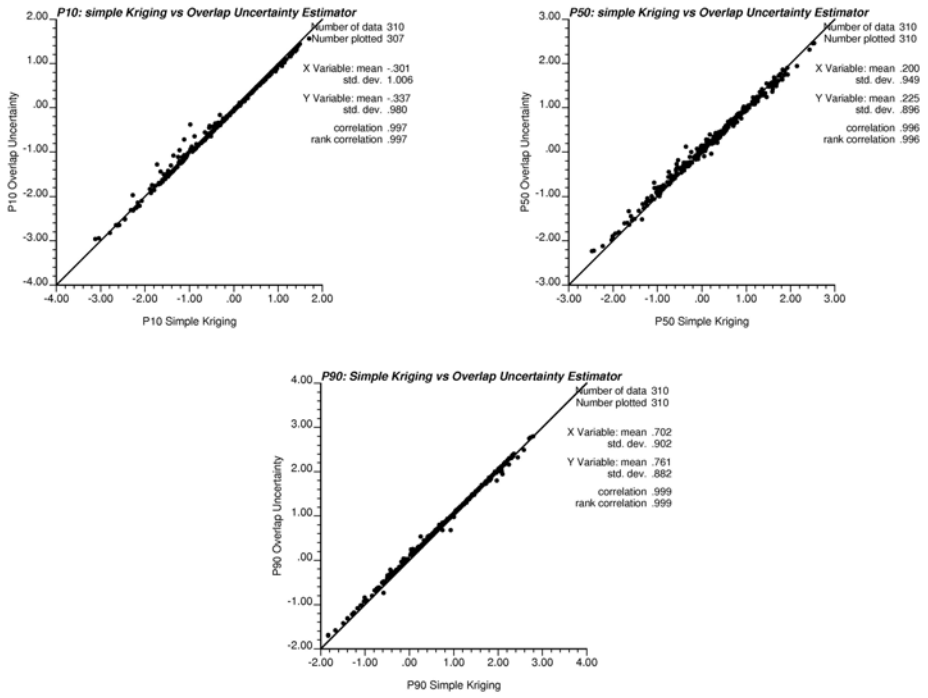

Figure 15: Crossplot between p10 (top left), p50 (top right) and p90 (bottom) of Simple Kriging and Overlap Uncertainty Estimator for nscore transformed data from file 'Mine650.dat'.

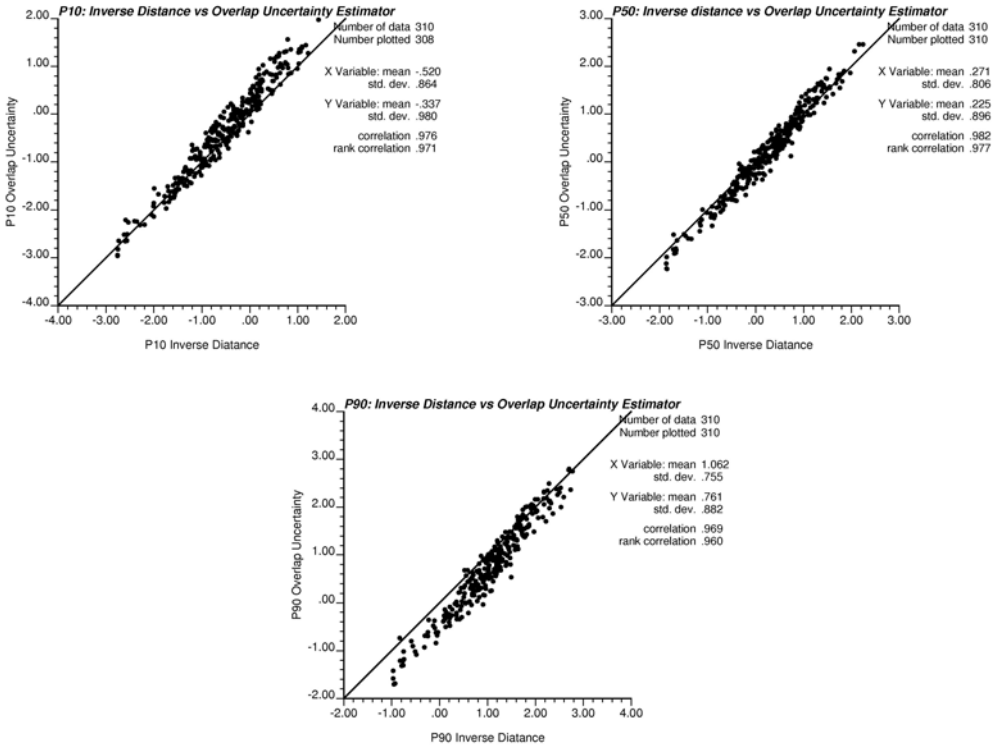

Figure 16: Crossplot between p10 (top left), p50 (top right) and p90 (bottom) of Inverse Distance and Overlap Uncertainty Estimator for nscore transformed data from file 'Mine650.dat'.