Estimation in a Finite Domain: Fixing the String Effect

Olena Babak and Clayton V. Deutsch

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

A new method for estimation in a finite domain is proposed. This method is referred to as Finite Domain Kriging. The method combines the features of both Kriging and Inverse Distance estimation methods. Finite Domain Kriging was tested using two real data sets, one from mining and the other one from petroleum. In both cases, the proposed method was shown to outperform Simple and Ordinary Kriging.

Introduction

The boundary points in a string of data receive more weight than other data in the string. To explain and analyze this phenomenon a comprehensive study of the influence of each and all data in a string on the estimation of finite domain was conducted by Babak (2006). The weights are theoretically valid for a stationary and ergodic random function. We believe, however, that this weighting of the boundary data could lead to biased estimation in a finite domain, especially if the data exhibits trends with boundary/border effects.

In this paper, we propose a new method for kriging in a finite domain. We will refer to this method as the Finite Domain Kriging. The Finite Domain Kriging will correct the structure of the data influence in the string on all locations in the finite domain. The correct structure will be imposed based on the ordering of weights in kriging using the distances from the location where we are estimating to all data in the string. The proposed method will be tested using several small examples and will be applied to two real data sets from mining and petroleum. Finally, the advantages and improvements in estimation obtained by using Finite Domain Simple Kriging or Finite Domain Ordinary Kriging in comparison to SK and OK, respectively, will be discussed.

Proposal: Formulation of the Finite Domain Kriging

In order to correct the structure of the influence of end samples in a string when estimating a finite domain, we propose to constrain the weights in kriging. Specifically, we propose to order the weights assigned by each location of interest to the string of data in the following way: the closest data in the string to the location of interest will receive the largest weight; the second closest data to the location of interest will receive the second weight and so on. In that way, the data in the string which is located furthest from the estimate location is assigned a weight based on a certain prescribed structure.

To summarize the proposed approach, we note that we do not change the estimation problem. That is, we still work with Kriging. Specifically, we still choose the estimate for the location of interest which minimizes the estimation variance. However, we solve a constraint optimization problem, where a certain structure is prescribed to the assigned weights.

It is worth noting that the Finite Domain Kriging mixes characterististics of both Kriging and Inverse Distance estimation methods. It was already mentioned that the Finite Domain Kriging, SK and OK minimize the estimation variance. However, in Inverse Distance approach, the closest data (not the boundary data) will receive the largest weight in estimation.

Now let us formalize the Finite Domain estimation problem. We consider both Finite Domain Simple Kriging and Finite Domain Ordinary Kriging estimation problems.

Specifically, by the Finite Domain Simple Kriging (FDSK), the value at the location of interest is found as

$$X_{FDSK}* = \sum_{i=1}^{n} \lambda_{FDSK,i} X_i + \left[1 - \sum_{i=1}^{n} \lambda_{FDSK,i}\right] m, \tag{1}$$

where X_i denotes the i^{th} value of the variable of interest in the string, $X_{FDSK}*$ denotes the Finite Domain Simple Kriging estimate at the location of interest, *m* denotes the population mean, and the FDSK weights $\lambda_{FDSK,i}$ are found by solving the following constraint optimization problem

$$\underset{\lambda}{\text{minimize}} \ \sigma_{est}^2 = \sigma^2 - 2\sum_{i=1}^{n} \lambda_{FDSK,i} Cov(X_i, X^*) + \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_{FDSK,i}\lambda_{FDSK,j} Cov(X_i, X_j) \tag{2}$$

subject to

$$\lambda_{FDSK,i} > \lambda_{FDSK,j}, \text{ if } d_i < d_j, \text{ for each } \ i, j = 1, \ldots, n, \tag{3}$$

where σ_{est}^2 denotes the estimation variance, σ^2 denotes the variance of the data; $Cov(X_i, X_j)$ and $Cov(X_i, X^*)$ denote the covariance between data in the string and between data in the string and the location of interest, respectively, and d_i denotes the distance from the location where we are estimating to the i^{th} data point in the string, $i, j = 1, \ldots, n.$

By the Finite Domain Ordinary Kriging (FDOK) the value at the location of interest is found as

$$X_{FDOK}* = \sum_{i=1}^{n} \lambda_{FDOK,i} X_i, \tag{4}$$

where X_i denotes the i^{th} value of the variable of interest in the string, $X_{FDOK}*$ denotes the Finite Domain Ordinary Kriging estimate at the location of interest, *m* denotes the population mean, and the FDSK weights $\lambda_{FDOK,i}$ are found by solving the following constraint optimization problem

$$\underset{\lambda}{\text{minimize}} \ \sigma_{est}^2 = \sigma^2 - 2\sum_{i=1}^{n} \lambda_{FDOK,i} Cov(X_i, X^*) + \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_{FDOK,i}\lambda_{FDOK,j} Cov(X_i, X_j) \tag{5}$$

subject to

$$\lambda_{FDOK,i} > \lambda_{FDOK,j}, \text{ if } d_i < d_j, \ i, j = 1, \ldots, n, \tag{6}$$

$$\sum_{i=1}^{n} \lambda_{FDOK,i} = 1, \tag{7}$$

where σ_{est}^2 denotes the estimation variance, σ^2 denotes the variance of the data; $Cov(X_i, X_j)$ and $Cov(X_i, X^*)$ denote the covariance between data in the string and between data in the string and location of interest, respectively, and d_i denotes the distance from the location where we are estimating to the i^{th} data in the string, $i, j = 1, \ldots, n.$

Implementation

The Finite Domain Simple Kriging and Finite Domain Ordinary Kriging problems formulated by Equations (1)-(3) and (4)-(7), respectively, can be solved by using a non-linear constraint optimization. However, in this report, we propose to solve these problems differently. That is, we will reformulate both Finite Domain Simple Kriging and Finite Domain Ordinary Kriging in Equations (1)-(3) and (4)-(7) in such a way that simple nonlinear unconstrained optimization can be applied for solving them.

Let us start with Finite Domain Simple Kriging. The minimization problem stated in (2)-(3) is equivalent to the following

$$\underset{\mu}{\text{minimize}} \ \sigma_{est}^2 = \sigma^2 - 2\sum_{i=1}^{n} \lambda_{FDSK,i} Cov(X_i, X^*) + \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_{FDSK,i}\lambda_{FDSK,j} Cov(X_i, X_j) \tag{8}$$

where σ_{est}^2 denotes the estimation variance, σ^2 denotes the variance of the data; the FDSK weights $\lambda_{FDSK,i}$ are given by

$$\lambda_{FDSK,i} = [n - k_i + 1]\text{- th largest element in vector } \mu\,, \tag{9}$$

where μ is the new parameter vector of size n by 1 with respect to which minimization is performed, $k = [k_1 \ k_2 \ \ldots k_n]$ with k_i denoting the position of the distance from the location where we are estimating to the i^{th} data in the string, d_i, is a vector of all distances $d = [d_1 \ d_2 \ \ldots d_n]$ sorted in ascending order; and $Cov(X_i, X_j)$ and $Cov(X_i, X^*)$ denote the covariance between data in the string and between data in the string and location of interest, respectively, $i, j = 1, \ldots, n.$

Note that despite the minimization problem given by (8)-(9) is equivalent (it is equivalent because for each vector μ there is unique a vector) to problem (2)-(3), it can be solved much more easily since we don't need to bother about the constraints. Further, Finite Domain Simple Kriging problem in form (1), (8)-(9) will be considered.

Similarly, we can rewrite the minimization problem stated in Equations (5)-(7) for the Finite Domain Ordinary Kriging as follows

$$\underset{\mu}{\text{minimize}} \ \sigma_{est}^2 = \sigma^2 - 2\sum_{i=1}^{n} \lambda_{FDOK,i} Cov(X_i, X^*) + \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_{FDOK,i}\lambda_{FDOK,j} Cov(X_i, X_j) \tag{10}$$

where σ^2_{est} denotes the estimation variance, σ^2 denotes the variance of the data; the FDOK weights $\lambda_{FDOK,i}$ are given by

$$\lambda_i = \frac{[n - k_i + 1]\text{ - th largest element in vector } \mu}{\text{sum of all element in } \mu}, \tag{11}$$

where μ is a new parameter vector of size n by 1 with respect to which minimization is performed, $k = [k_1\ k_2\ \ldots\ k_n]$ with k_i denoting the position of the distance from the location where we are estimating to the i^{th} data in a string, d_i, is a vector of all distances $d = [d_1\ d_2\ \ldots\ d_n]$ sorted in ascending order; and $Cov(X_i, X_j)$ and $Cov(X_i, X^*)$ denote the covariance between data in the string and between data in the string and location of interest, respectively, $i, j = 1, \ldots, n.$ Further, Finite Domain Ordinary Kriging problem in form (4), (10)-(11) will be considered.

Program

For solution of the Finite Domain Simple Kriging in Equations (8)-(9) and Finite Domain Ordinary Kriging in Equations (10)-(11), the optimization subroutine MINF1 from the Scientific Subroutine Library II (SSL II) was used. This subroutine is designed to perform minimization of a function with several variables using revised quasi-Newton method based on function values only. For convenience this subroutine was incorporated to kt3d program. The new program is called kt3d_up. This program uses the same parameter file as kt3d, see Deutsch and Journel (1998) for reference. The result of running kt3d_up program, one obtains the file containing either 4 columns (grid option) with the 'usual' kriging estimate (either OK or SK depending on the chosen option) and its variance in columns 1 and 2 and, respectively, then the Finite Domain Kriging estimate and its variance in columns 3 and 4, respectively. Alternatively, the output may consist of 10 columns (if cross validation or jackknife option is chosen) with the same 7 columns as would be obtained using kt3d program and Finite Domain Kriging estimate, its estimation variance and error in columns 8, 9 and 10.

The following table compares the time required for program kt3d and kt3d_up to perform estimation for three small estimation exercises:

	Kt3d	Kt3d_up
Exercise 1	0.26245740E+03	0.60797422E+03
Exercise 2	0.20897048E+03	0.48253385E+03
Exercise 3	0.29446342E+03	0.48515762E+03

In general, we note that when estimating a finite domain using program kt3d_up, which performs both 'usual' Kriging (either Simple Kriging or Ordinary Kriging) and the corresponding type of Finite Domain Kriging, we spend about twice as much time compared to estimating a finite domain using program kt3d, which performs 'usual' Kriging. The time required for program kt3d_up to complete estimation depends, of course, on the number of steps required for function (8) in the Finite Domain Ordinary Kriging case or (10) in the Finite Domain Simple Kriging case to achieve the minimum at each location.

Small Examples

To illustrate how optimal weights are found in Finite Domain Kriging approach several small exercises were performed. The aim of the exercises was the estimation of four locations at (1, 7), (1.8, 7), (2.8, 7) and (3.8,7) based on the string of 7 data located at (1,0), (2,0), (3,0), (4,0), (5,0), (6,0) and (7,0), respectively, using Finite Domain Kriging, SK and OK. Results of the estimation are shown in Figure 1 for Finite Domain Ordinary Kriging and Ordinary Kriging and, in Figure 2 for Finite Domain Simple Kriging and Simple Kriging. The structure of optimal weights obtained by Finite Domain Kriging approaches is, of course, almost always different from the ones obtained based on both Simple and Ordinary Kriging. The only exception is when estimating locations which are located at the shortest distance to one of the two boundary data in the string (see Figure 2), then Finite Domain Simple Kriging results in the same estimate as SK.

Practical Applications

For comparison of the Finite Domain Simple Kriging and Finite Domain Ordinary Kriging with SK and OK, respectively, two real data sets were considered. One data set was chosen from a petroleum reservoir (data set 1) and one from a mineral deposit (data set 2). Both data sets contain the information from several vertical wells. Locations of these wells in the *XY* plane are shown in Figure 3 for both data sets. Figure 4 shows the histograms of the variables of interest for the petroleum and the mining data set. Figures 5 and 6 show the experimental variograms and their theoretical fits for the variable of interest in data set 1 and data set 2, respectively.

Figure 7 shows the first and the middle slice in the XY plane of the 3D model for the variable of interest obtained based on the petroleum data set (data set 1) using OK and Finite Domain Ordinary Kriging. Figure 8 shows analogous results for the variable of interest but obtained based on the SK and Finite Domain Simple Kriging, respectively. For mining data set (data set 2) results are shown in Figures 9 and 10.

In order to assess the improvement of Finite Domain Kriging estimation of the variable of interest over the ‘usual’ Kriging estimation, both cross validation and jackknife were applied. The number of data used in both validation techniques was set to be from 10 to 20 for data set 1 and data set 2. For jackknife validation only wells with at least two observations were used. As a measure of improvement over Simple and Ordinary Kriging, respectively, the following statistics were used

$$\text{Improvement over SK} = \frac{(\text{Sum of abs values of residuals of SK - Sum of abs values of residuals of FDSK})}{\text{Sum of abs values of residuals of SK}} \cdot 100\%$$

$$\text{Improvement over OK} = \frac{(\text{Sum of abs values of residuals of OK - Sum of abs values of residuals of FDOK})}{\text{Sum of abs values of residuals of OK}} \cdot 100\% \tag{12}$$

Note that the above statistics can take on both positive and negative values. The positive values, of course, correspond to the fact that Finite Domain Kriging indeed performed better than ‘usual’ kriging. On the other hand, negative values imply that Finite Domain Kriging performed worse than ‘usual’ kriging.

The results of the crossvalidation for data set 1 are shown in Figures 11 and 12 and summarized in Table 1 using several statistics including the improvement measures given by equation (12). The results of the jackknife for this data set are given in Figures 13 and 14 and summarized in

Table 2. Figures 13 and 14 show the sign and magnitude as well as histogram of the improvement of Finite Domain Kriging estimation of the variable of interest over the ‘usual’ kriging estimation.

The analogous results for cross validation of the mining data set (data set 2) are shown in Figures 15, 16 and summarized in Table 3. The results of the jackknife for this data set are given in Figures 17 and 18 and summarized in Table 4.

Conclusions

A new approach for kriging in a finite domain using strings of data is proposed. This approach, referred to as Finite Domain Kriging (FDK), combines characteristics of both Inverse Distance and Kriging. Similar to an Inverse Distance estimate, the estimate obtained by FDK weights conditioning data based on the distance from the location of interest. Similar to a kriged estimate, the estimate obtained using FDK is chosen to be the one which minimizes the estimation variance (but subject to distance constraints). With respect to the constraint on the sum of weights given to conditioning data, two flavors of Finite Domain Kriging are considered. That is, Finite Domain Simple Kriging, for which the sum of the weights can take on any value, and Finite Domain Ordinary Kriging, for which the sum of the weights is constrained to be one.

The proposed approach to estimation of a finite domain using strings of data was tested using two real data sets, one data set from mining and one data set from petroleum. The cross validation results for both data sets reveal an improvement over both SK and OK. Specifically, it was observed for data set 1 that both Finite Domain Ordinary Kriging and Finite Domain Simple Kriging perform about 1.7% better than OK and SK, respectively. For data set 2, improvement of Finite Domain approaches was even higher, it was over 2%. Note that improvement was measured based on the sum of absolute values of error distributions obtained in cross validation.

With respect to jackknife for petroleum data set, we observed that average improvement of Finite Domain Ordinary Kriging (FDOK) over OK is close to 5%. The average improvement of Finite Domain Simple Kriging (FDSK) over SK for this data set is more than 3%. For data set 2 the improvement was not as significant, it was about 1.5% for FDOK over OK and only about 0.6% for FDSK over SK. In general, note, however, that in both cases method developed in this report performed better than both Simple and Ordinary Kriging with respect to both cross validation and jack knife.

References

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

Babak, O.: The Problem of Kriging when Estimating in a Finite Domain, CCG Report 8, 2005-2006.

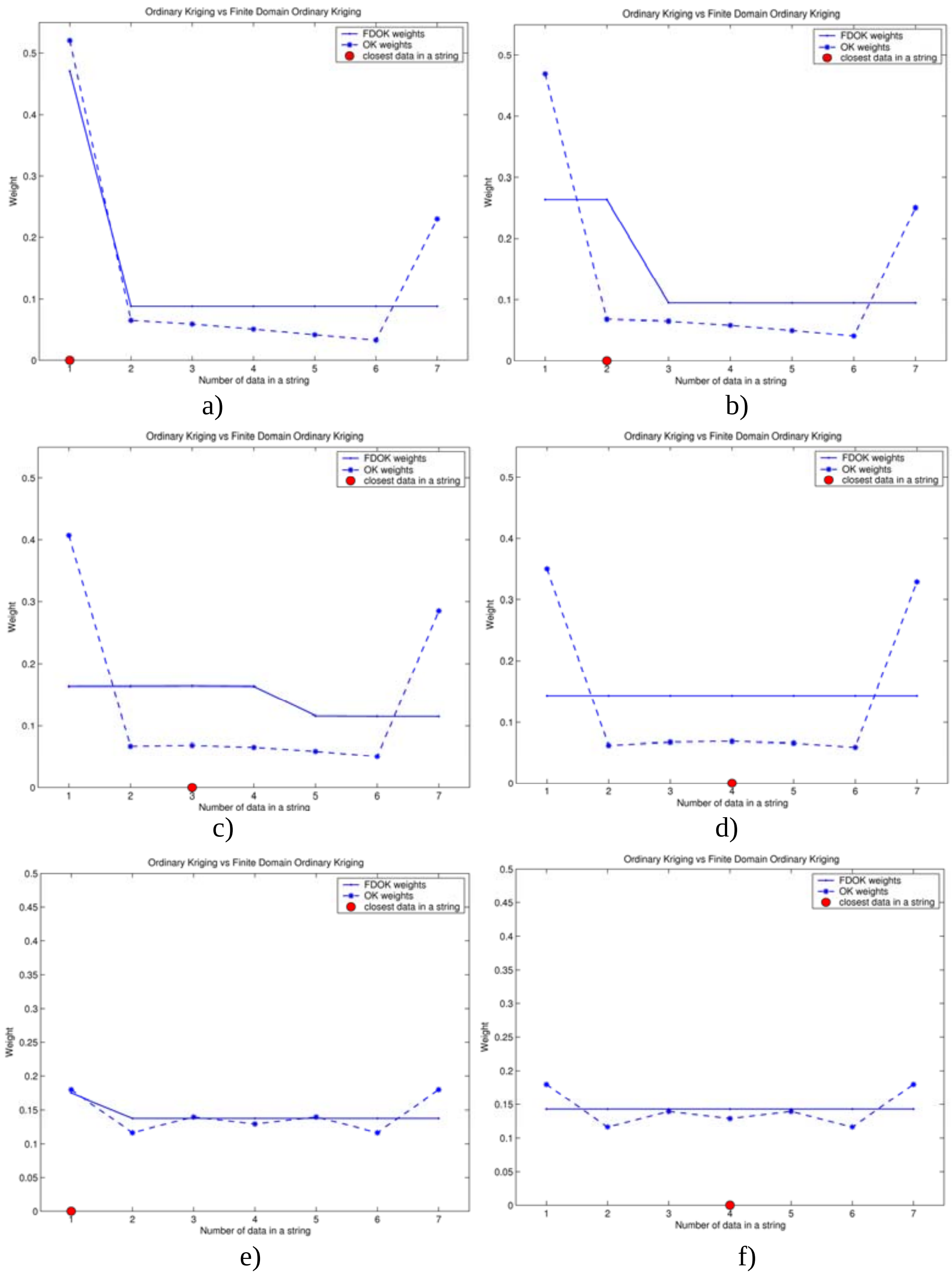


Figure 1: Structure of the Finite Domain Ordinary Kriging (solid line) and Ordinary Kriging (dashed line) weights when estimation location: a) (1,7); b) (1.8, 7); c) (2.8, 7) and d) (3.8,7) using Spherical variogram model with the range of correlation 20; e) (1,7) and f) (3.8,7) using Spherical variogram model with the range of correlation based on the string of 7 data located at (1,0), (2,0), (3,0), (4,0), (5,0), (6,0) and (7,0), respectively.

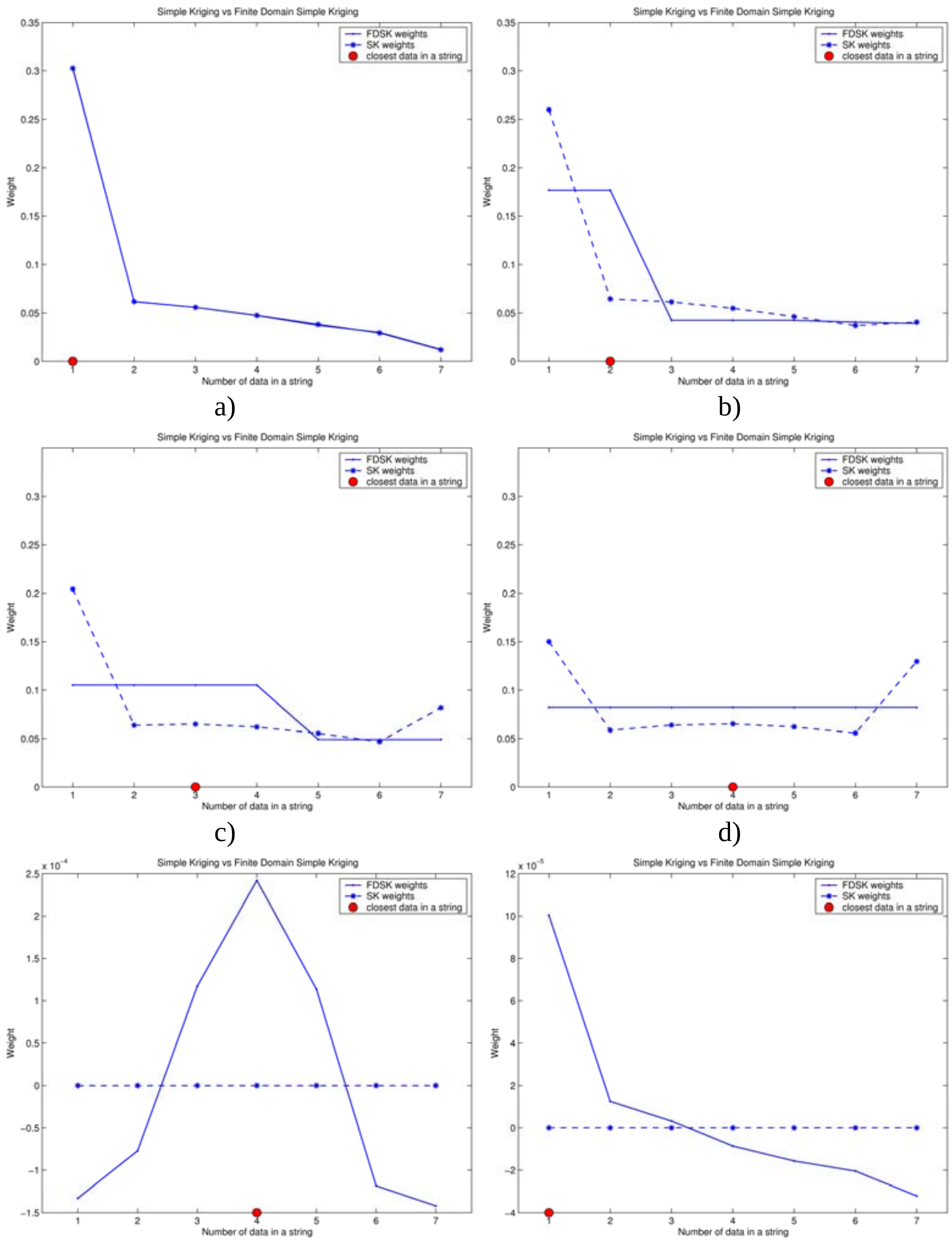
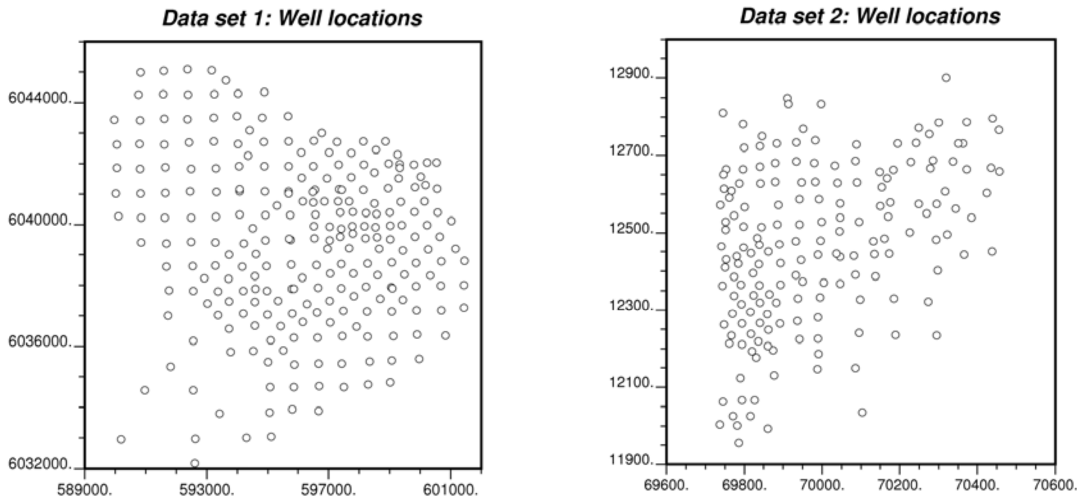
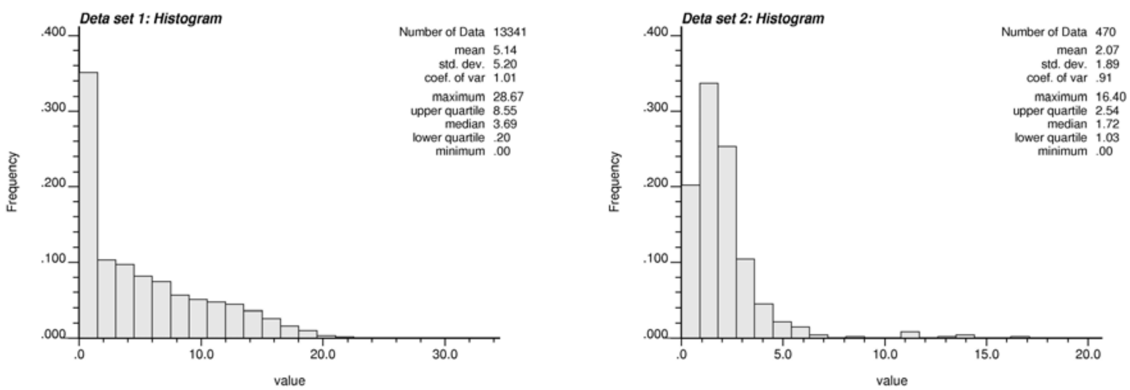


Figure 2: Structure of the Finite Domain Simple Kriging (solid line) and Simple Kriging (dashed line) weights when estimation location: a) (1,7); b) (1.8, 7); c) (2.8, 7) and d) (3.8,7) using Spherical variogram model with the range of correlation 20; e) (1,7) and f) (3.8,7) using Spherical variogram model with the range of correlation based on the string of 7 data located at (1,0), (2,0), (3,0), (4,0), (5,0), (6,0) and (7,0), respectively.



a) b)

Figure 3: Locations of the wells for a) data set 1 and b) data set 2.



a) b)

Figure 4: Histogram of the variable of interest in a) data set 1 and b) data set 2.

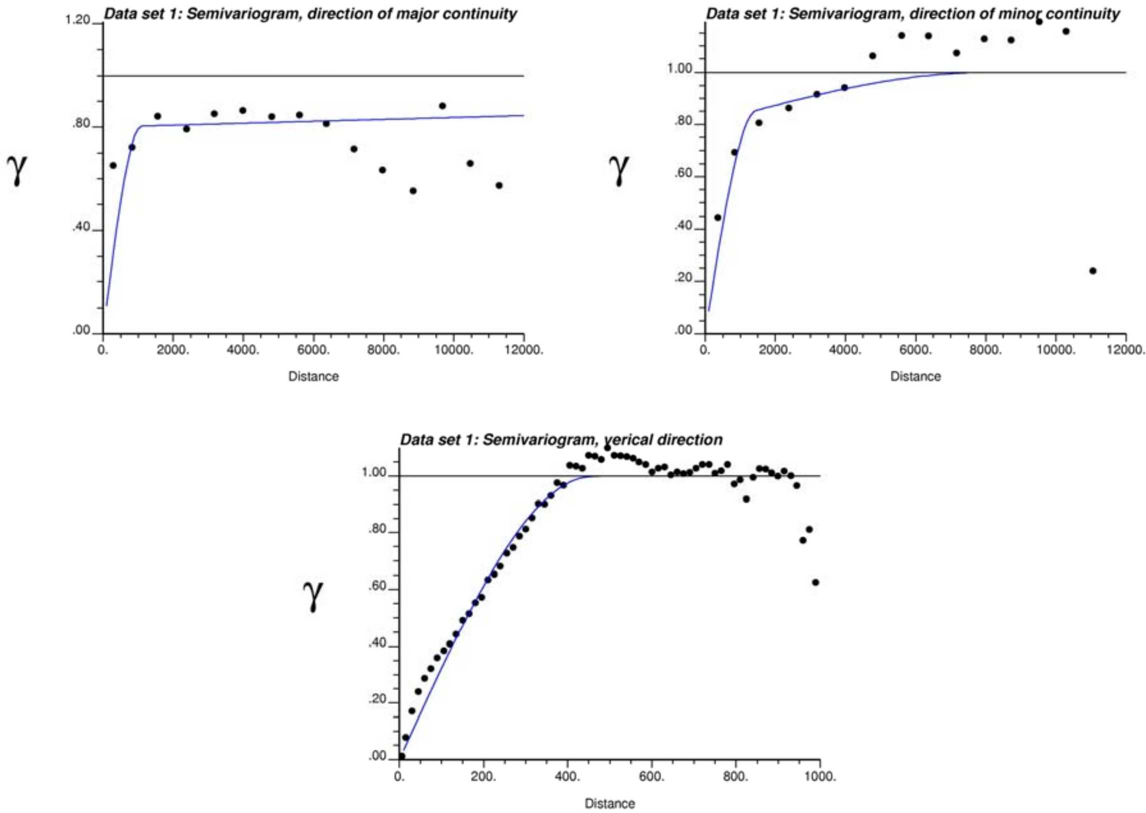


Figure 5: Experimental variogram and its theoretical fit in the three directions of major continuity for the variable of interest in the data set 1.

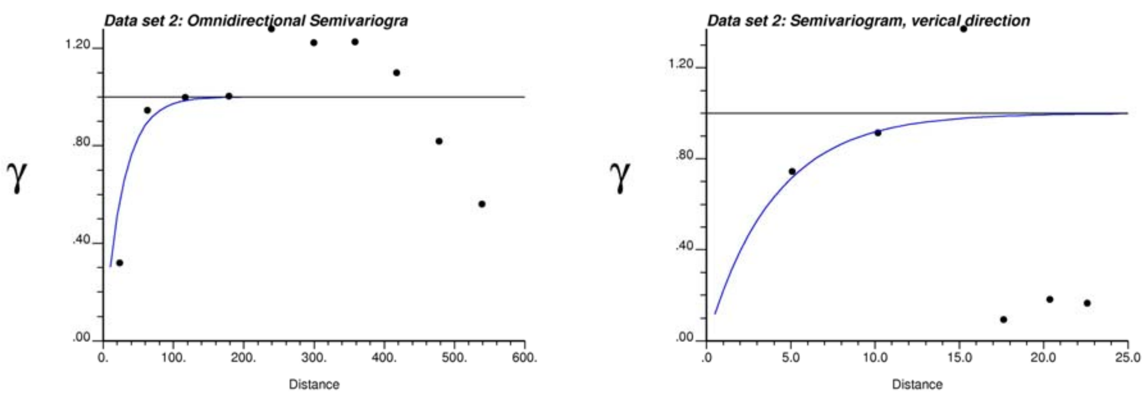


Figure 6: Experimental variogram and its theoretical fit for the variable of interest in the data set 2.

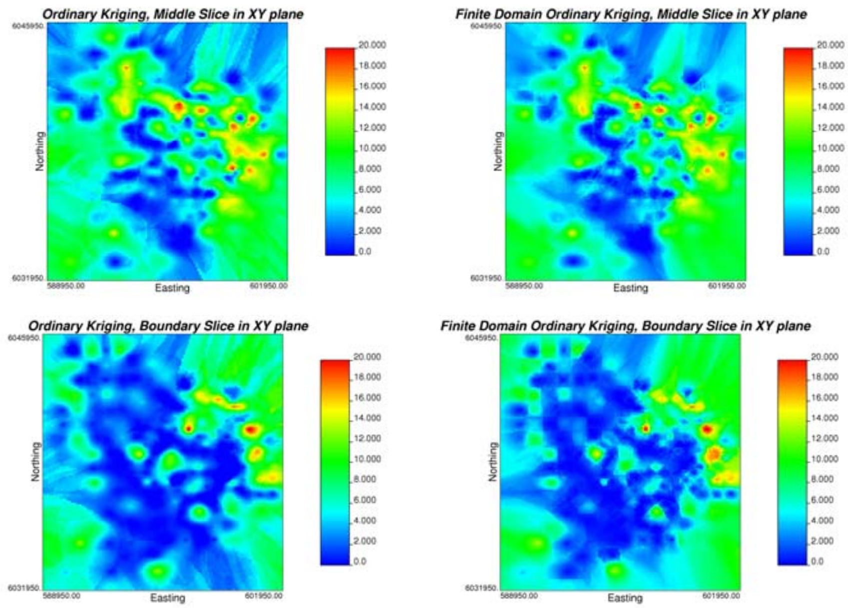


Figure 7: The first (top) and the middle (bottom) slice in the XY plane of the 3D model for the variable of interest obtained based on the petroleum data set (data set 1) using Ordinary Kriging (first column) and Finite Domain Ordinary Kriging (second column).

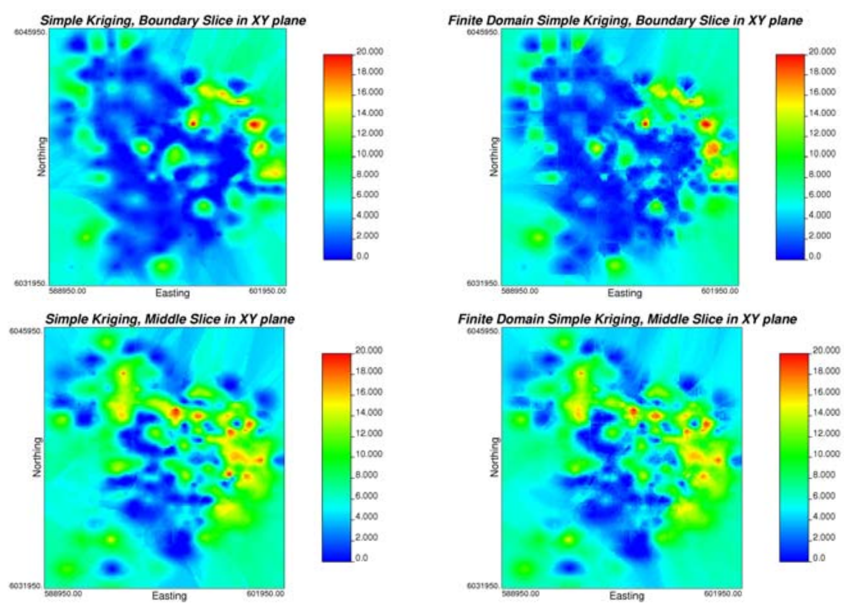


Figure 8: The first (top) and the middle (bottom) slice in the XY plane of the 3D model for the variable of interest obtained based on the petroleum data set (data set 1) using Ordinary Kriging (first column) and Finite Domain Ordinary Kriging (second column).

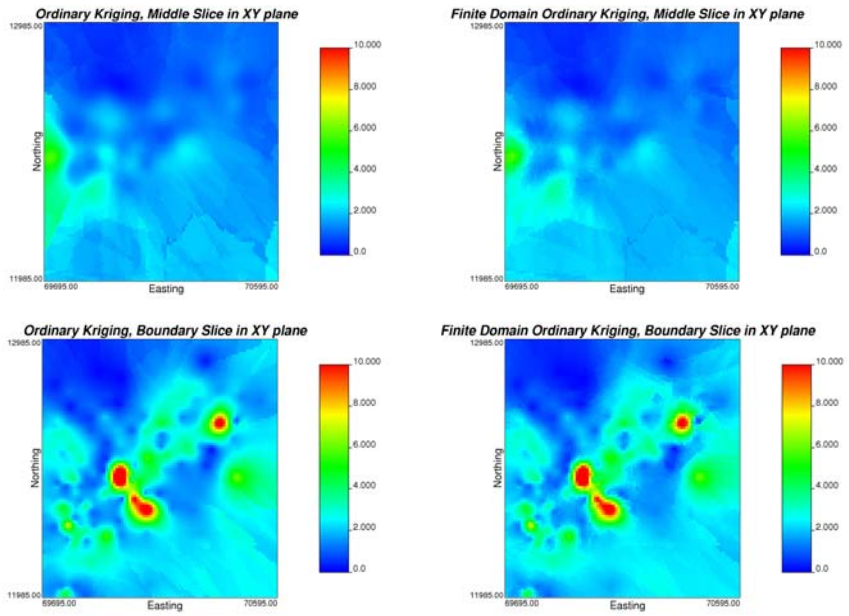


Figure 9: The first (top) and the middle (bottom) slice in the XY plane of the 3D model for the variable of interest obtained based on the mining data set (data set 2) using Ordinary Kriging (first column) and Finite Domain Ordinary Kriging (second column).

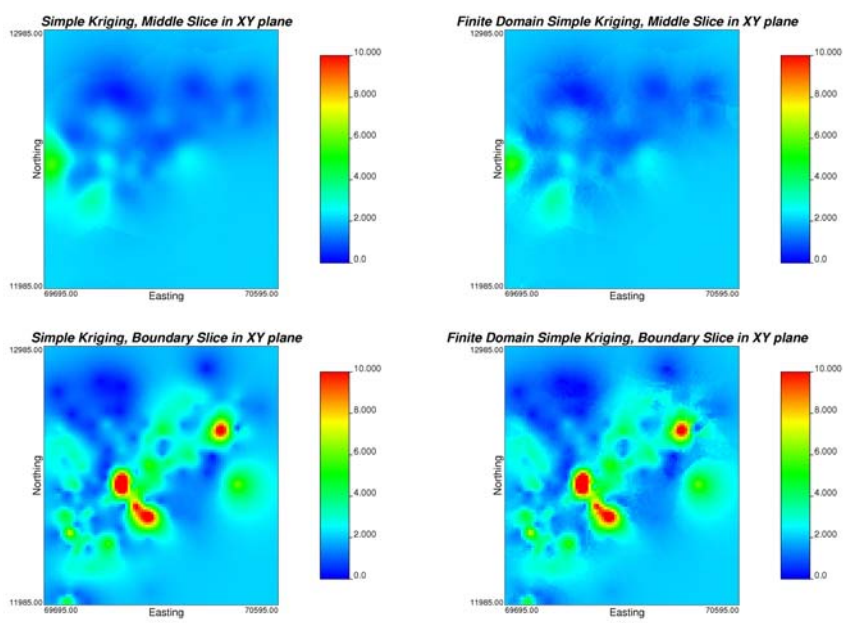


Figure 10: The first (top) and the middle (bottom) slice in the XY plane of the 3D model for the variable of interest obtained based on the mining data set (data set 2) using Ordinary Kriging (first column) and Finite Domain Ordinary Kriging (second column).

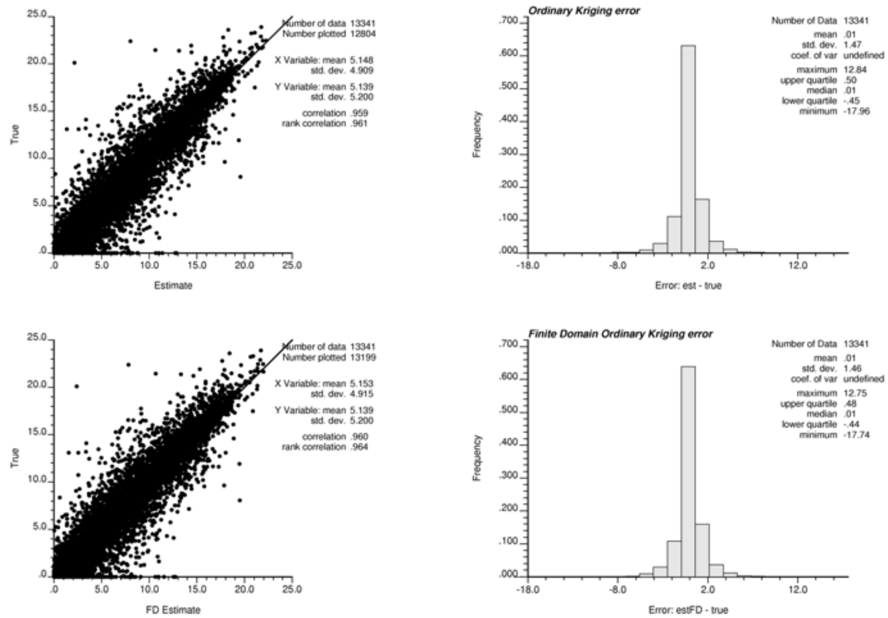


Figure 11: Result of cross validation for Ordinary Kriging (top) and Finite Domain Ordinary Kriging (bottom) for data set 1: crossplot between true value of the variable of interest and its estimate (first column) and histogram of errors (second column).

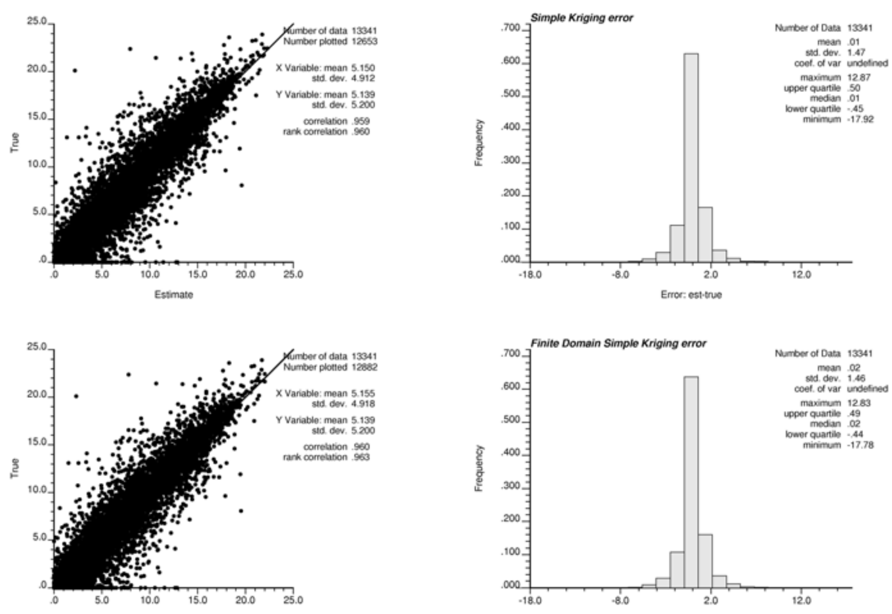


Figure 12: Result of cross validation for Simple Kriging (top) and Finite Domain Simple Kriging (bottom) for data set 1: crossplot between true value of the variable of interest and its estimate (first column) and histogram of errors (second column).

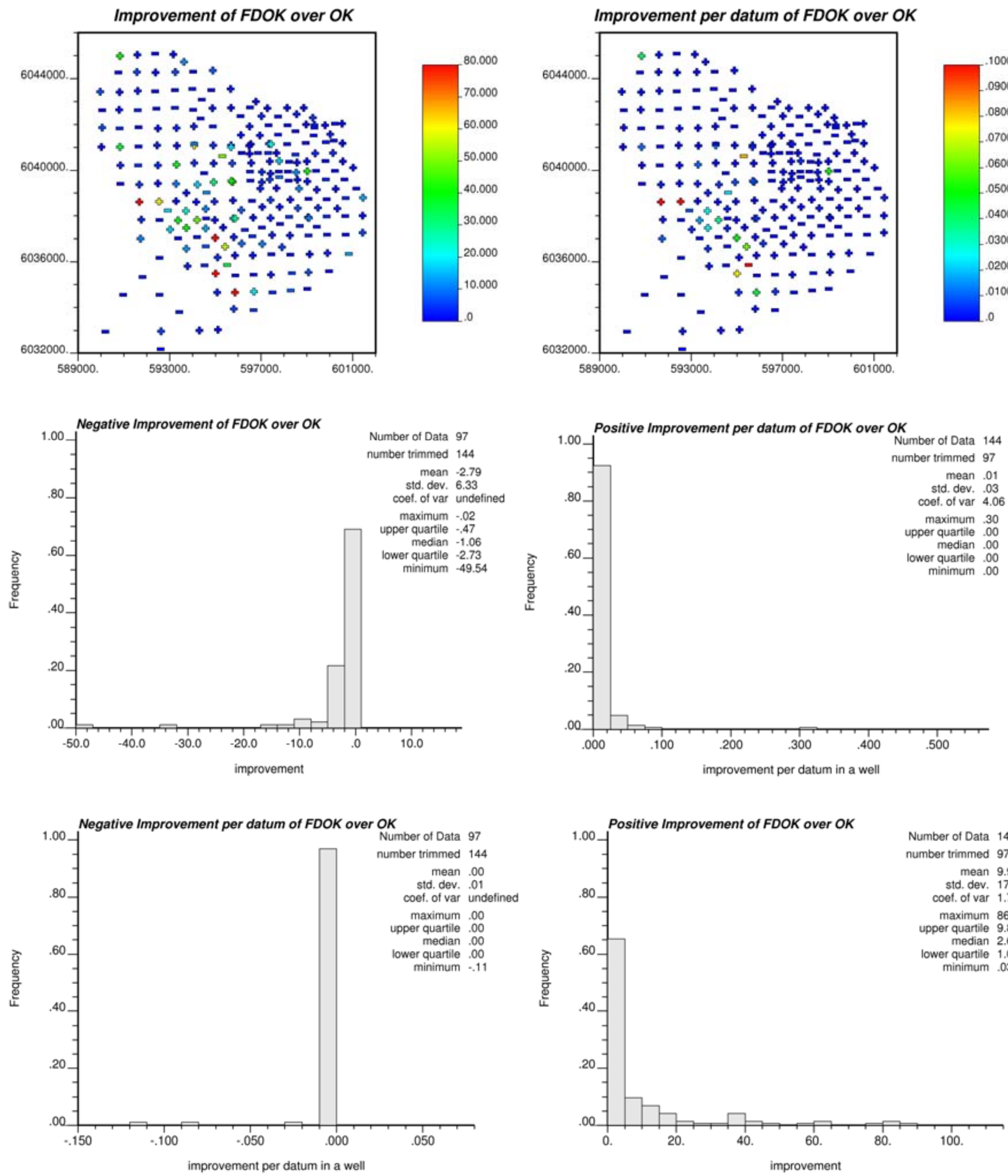


Figure 13: Result of jackknife shown as improvement of Finite Domain Ordinary Kriging over Ordinary Kriging for data set 1: improvement map and improvement per datum map (top), histograms of positive and negative improvement (middle) and histograms of positive and negative improvement (bottom).

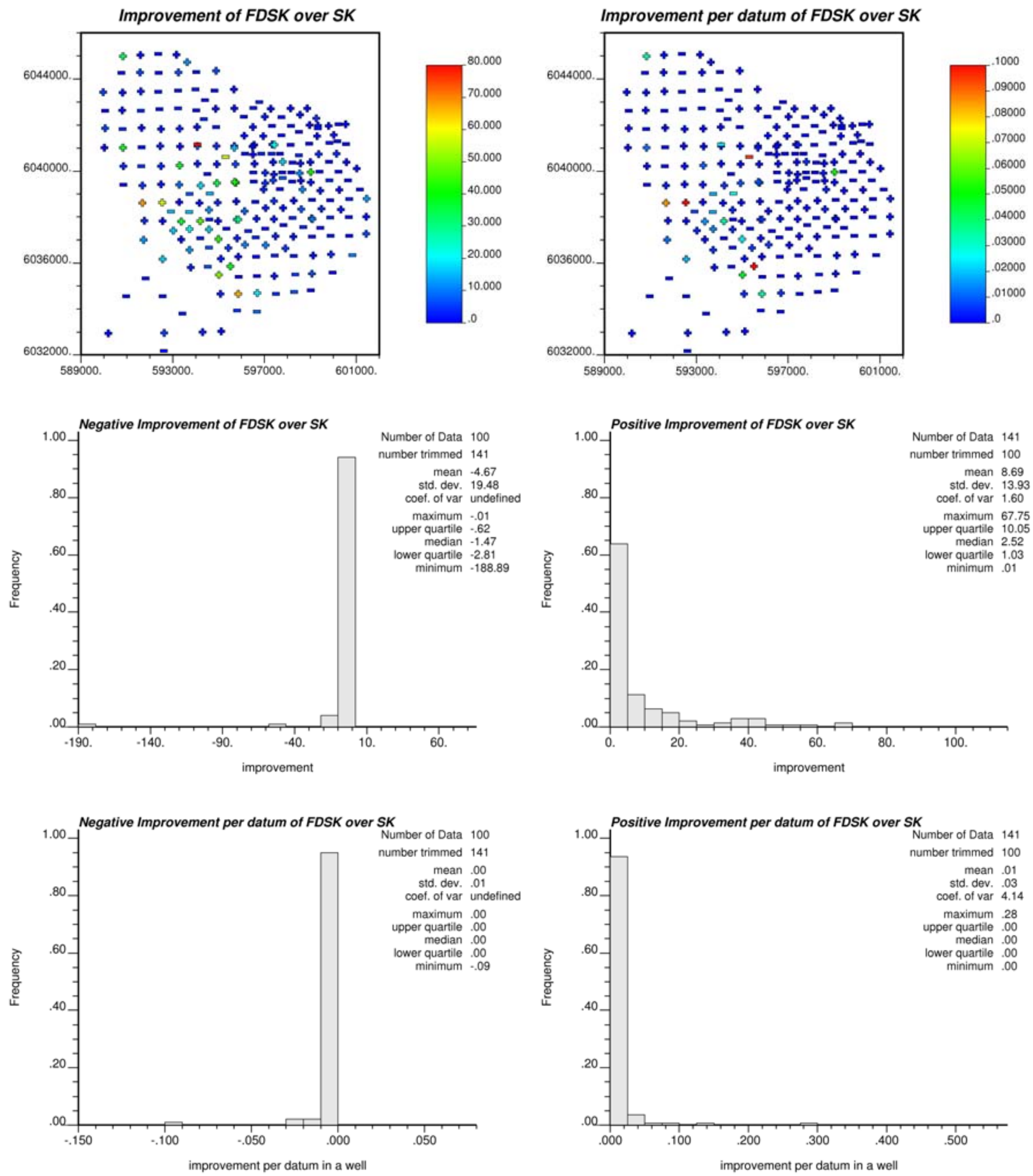


Figure 14: Result of jackknife shown as improvement of Finite Domain Simple Kriging over Simple Kriging for data set 1: improvement map and improvement per datum map (top), histograms of positive and negative improvement (middle) and histograms of positive and negative improvement (bottom).

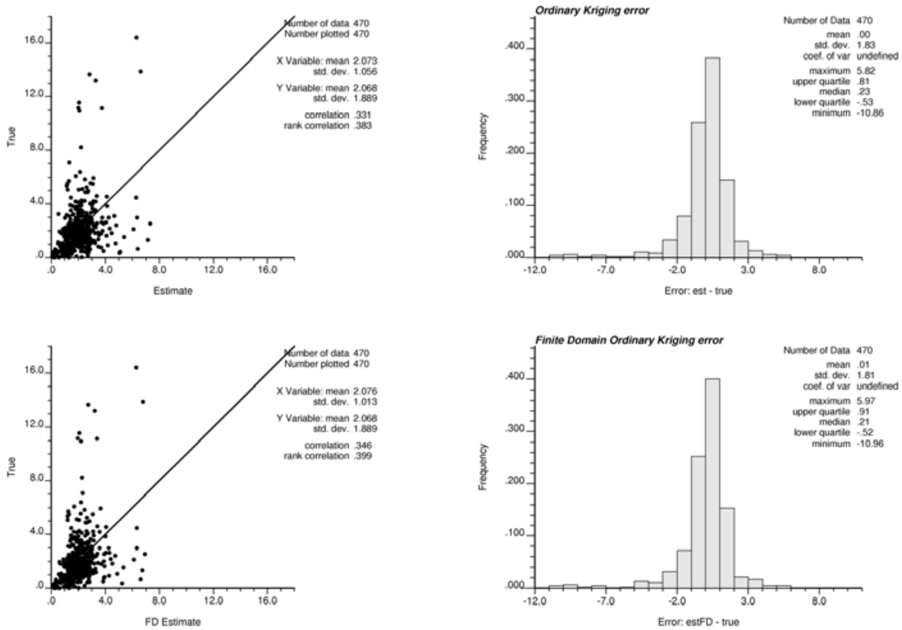


Figure 15: Result of cross validation for Ordinary Kriging (top) and Finite Domain Ordinary Kriging (bottom) for data set 2: crossplot between true value of the variable of interest and its estimate (first column) and histogram of errors (second column).

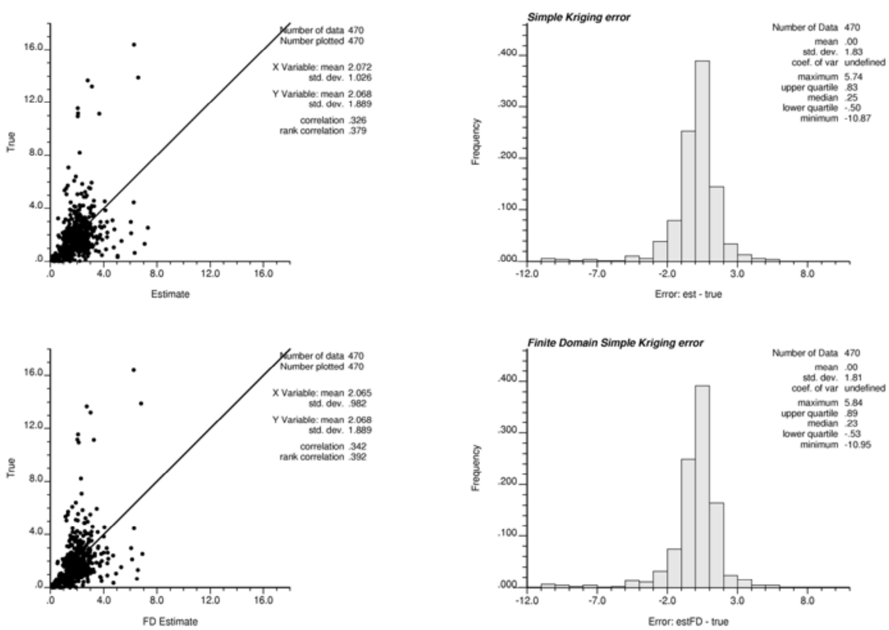


Figure 16: Result of cross validation for Simple Kriging (top) and Finite Domain Simple Kriging (bottom) for data set 2: crossplot between true value of the variable of interest and its estimate (first column) and histogram of errors (second column).

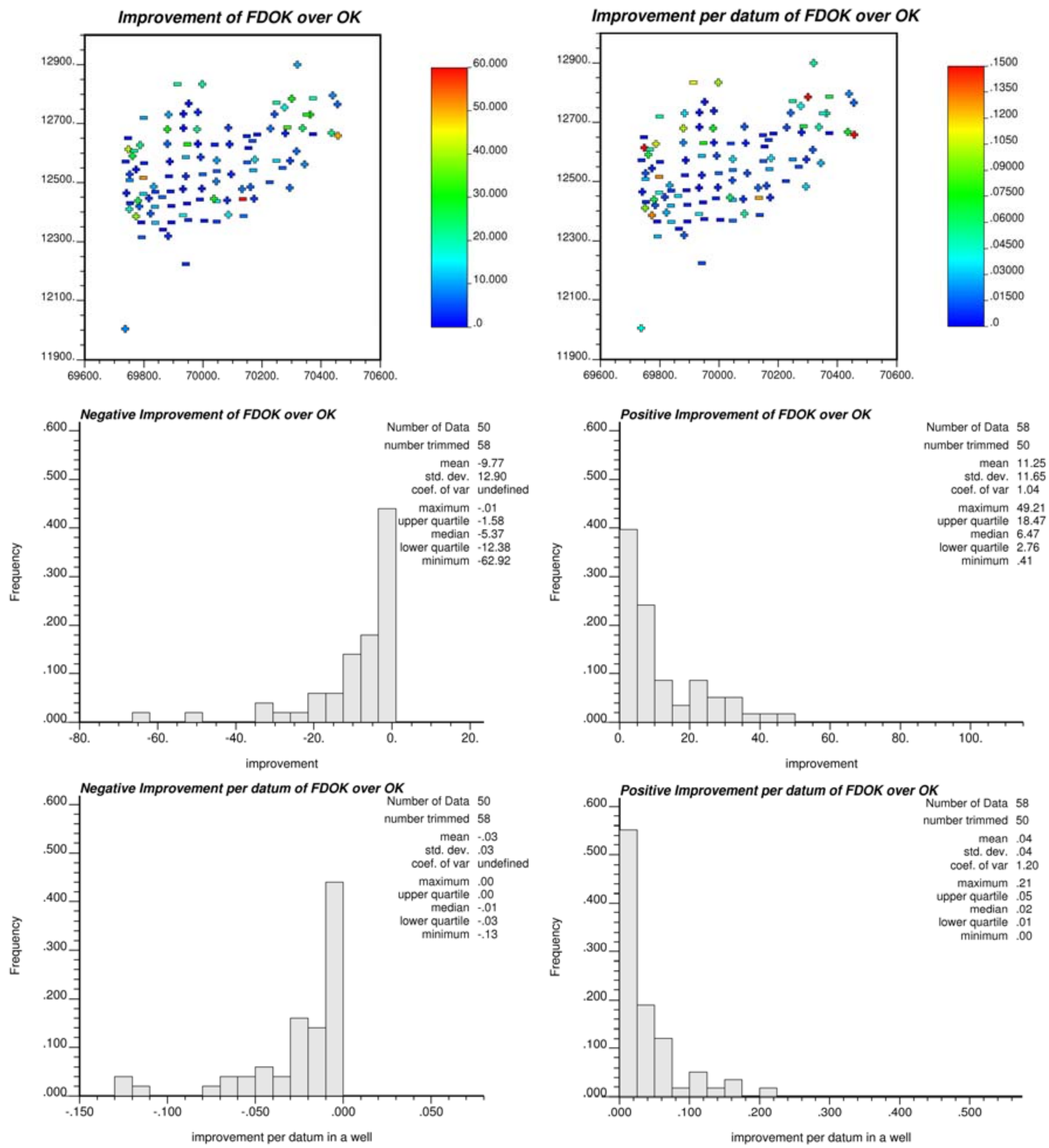


Figure 17: Result of jackknife shown as improvement of Finite Domain Ordinary Kriging over Ordinary Kriging for data set 2: improvement map and improvement per datum map (top), histograms of positive and negative improvement (middle) and histograms of positive and negative improvement (bottom).

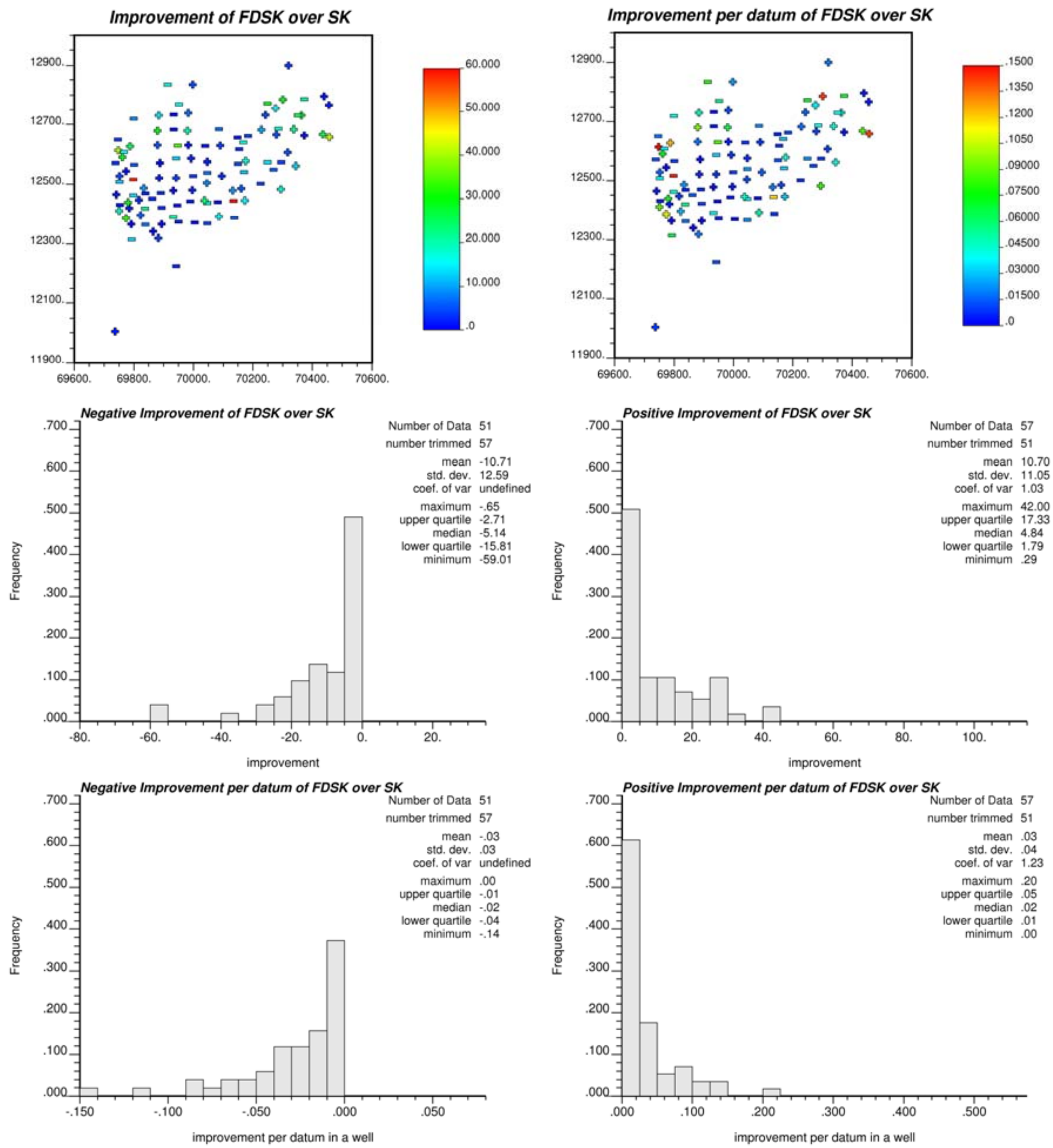


Figure 18: Result of jackknife shown as improvement of Finite Domain Simple Kriging over Simple Kriging for data set 2: improvement map and improvement per datum map (top), histograms of positive and negative improvement (middle) and histograms of positive and negative improvement (bottom).

Table 1: Result of cross validation for all 13341 data of the data set 1.

Type of Kriging	Min error	Max error	Average error	Std of error	Correlation between true and estimate	Sum abs error	Improvement %	Impr. Per data, %
OK	-17.9570	12.8420	0.0092	1.4738	0.9591	11681.86	1.6932	0.0001
FDOK	-17.7430	12.7530	0.0133	1.4649	0.9596	11484.07		
SK	-17.9170	12.8670	0.0110	1.4723	0.9592	11699.12	1.6656	0.0001
FDSK	-17.7790	12.8280	0.0161	1.4636	0.9597	11504.26		

Table 2: Result of jackknife for 241 wells of the data set 1.

Types of Kriging	Min improv., %	Min improv. per datum, %	Max improv., %	Max improv. per datum, %	Average improv., %	Average improv. per datum, %	Number of positive improv.	Number of negative improv.	Abs relative improv., times
FDOK vs OK	-49.543	-0.1111	86.1440	0.3034	4.8411	0.0031	144	97	5.3133
FDSK vs SK	-188.89	-0.0936	67.7523	0.2840	3.1502	0.0030	141	100	2.6268

Table 3: Result of cross validation for all 470 data of the data set 2.

Type of Kriging	Min error	Max error	Average error	Std of error	Correlation between true and estimate	Sum abs error	Improvement %	Impr. Per data, %
OK	-10.8630	5.8210	0.0048	1.8333	0.3315	530.4240	2.3012	0.0049
FDOK	-10.9600	5.9650	0.0078	1.8088	0.3458	518.2180		
SK	-10.8740	5.7390	0.0041	1.8318	0.3264	529.9470	2.0313	0.0043
FDSK	-10.9460	5.8420	-0.0025	1.8072	0.3415	519.1820		

Table 4: Result of jackknife for 108 wells of the data set 2.

Types of Kriging	Min improv., %	Min improv. per datum, %	Max improv., %	Max improv. per datum, %	Average improv., %	Average improv. per datum, %	Number of positive improv.	Number of negative improv.	Abs relative improv., times
FDOK vs OK	-62.920	-0.1293	49.2140	0.2094	1.5194	0.0082	58	50	1.3360
FDSK vs SK	-59.009	-0.1449	42.0032	0.2021	0.5862	0.0051	57	51	1.1159