Comput Geosci (2007) 11:219–234
DOI 10.1007/s10596-007-9046-x

ORIGINAL PAPER

On unbiased backtransform of lognormal kriging estimates

Jorge Kazuo Yamamoto

Received: 28 November 2006 / Accepted: 26 March 2007 / Published online: 24 April 2007

Abstract Lognormal kriging is an estimation technique that was devised for handling highly skewed data distributions. This technique takes advantage of a logarithmic transformation that reduces the data variance. However, backtransformed lognormal kriging estimates are biased because the nonbias term is totally dependent on a semivariogram model. This paper proposes a new approach for backtransforming lognormal kriging estimates that not only presents none of the problems reported in the literature but also reproduces the sample histogram and, consequently, the sample mean.

Keywords Backtransform · Lognormal kriging · Uncertainty · Interpolation variance · Smoothing effect

1 Introduction

Several natural phenomena follow lognormal distribution, i.e., frequency distributions show positive skewness. Distributions of rare metals such as gold, tungsten, tin, and uranium among others follow lognormal distributions. By definition a lognormal distribution is one whose logarithms of observations follow a normal distribution. Lognormal distributions are characterized by a great quantity of low values and a very small quantity of high values. Given a random variable *Z(x)* presenting a lognormal distribution, then the logarithm of *Z(x)* will present a normal distribution. In fact, the purpose of lognormal transformation is to change the shape of the frequency distribution in which skewed distributions are more or less normalized (pp. 231 and 233 in [6]). This normalization reduces the data variance that improves the calculation of statistics and weighted averages such as ordinary kriging estimates. In geostatistics, a class of estimators based on logarithmic transformation is known as lognormal kriging. Depending on whether the mean is known or not, simple or ordinary lognormal kriging can be used. The main idea of lognormal kriging is to take advantage of transformed data distribution to reduce the influence of few high values (p. 999 in [8]). This is an excellent idea; however, resulting estimates appear as a logarithm of original values, making its interpretation more difficult. Thus, it is necessary to backtransform lognormal kriging estimates, returning them to the original measurement scale. Backtransformation is achieved as the exponential of the kriging estimate plus a nonbias term (pp. 291 and 295 in [4], pp. 216 and 217 in [7], and p. 1001 in [8]):

$$Z^*_{\mathrm{SLK}}(x_\mathrm{o}) = \exp\left(Y^*_{\mathrm{SK}}(x_\mathrm{o}) + \sigma^2_{\mathrm{SK}}/2\right) \tag{1}$$

$$Z^*_{\mathrm{OLK}}(x_\mathrm{o}) = \exp\left(Y^*_{\mathrm{OK}}(x_\mathrm{o}) + \sigma^2_{\mathrm{OK}}/2 - \mu\right) \tag{2}$$

where σ^2_{SK} and σ^2_{OK} are kriging variances, respectively, for simple and ordinary kriging; μ is the Lagrange multiplier; and $\sigma^2_{\mathrm{SK}}/2$ and $\sigma^2_{\mathrm{OK}}/2 - \mu$ are the nonbias terms for simple and ordinary kriging, respectively.

There are two problems in using expression 1 or 2. First, the expected value of backtransformed lognormal kriging estimates is biased; that is, it is not equal to the sample mean (p. 572 in [5]):

$$E\left[\exp\left(Y^*_{\mathrm{OK}}(x_\mathrm{o}) + \sigma^2_{\mathrm{OK}}/2 - \mu\right)\right] \leq E[Z(x)]$$

J. K. Yamamoto (✉)
Department of Environmental and Sedimentary Geology, Institute of Geosciences, University of São Paulo, São Paulo, Brazil
e-mail: jkyamamo@usp.br

These authors have proposed an application of a corrective factor in such a way that:

$$Z^*_{\mathrm{OLK}}(x_o) = k_o \exp\left(Y^*_{\mathrm{OK}}(x_o) + \sigma^2_{\mathrm{OK}}/2 - \mu\right) \quad (3)$$

where k_o is the corrective factor that makes the expected value of backtransformed value equal to the sample mean. According to Journel and Huijbregts ([5], p. 572) the biasedness of expression 1 or 2 is due to its lack of robustness concerning the multivariate lognormal hypothesis.

The second problem regarding expression 1 or 2 derives from the nonbias term that is totally dependent on the semivariogram model. In this sense, several authors have reported this problem, such as: David ([2], p. 119), Chilés and Delfiner ([1], p. 192), and Saito and Goovaerts ([9], p. 4230), among others.

Although the idea behind lognormal kriging is supported by an elegant theory, it results in insufficient backtransformed values. Therefore, the problem is to find a nonbias term that guarantees that the expected value of backtransformed values is equal to the sample mean and that these values do not show loss of local accuracy. This paper proposes a new approach that ensures both of the conditions mentioned above.

2 The new approach

Lognormal kriging is based on a logarithmic transformation of the original random variable *Z(x)*:

$$Y(x) = \mathrm{Ln}Z(x)$$

Sometimes data translation by adding a constant is desirable:

$$Y(x) = \mathrm{Ln}(Z(x) + C)$$

Usually this constant is equal to 1 because it makes all logarithms greater than 0. Depending on the constant that is added for logarithmic transformation, frequency distribution of *Y(x)* is no longer normal. However, the constant reduces the variance of transformed values in such a way that the greater the constant the smaller the variance of logarithms is. Considering the use of a constant added to the variable *Z(x)* before logarithmic transformation, the best alternative is to consider ordinary kriging as an estimation technique instead of simple kriging. In this sense, expression 2 can be rewritten as:

$$Z^*_{\mathrm{OLK}}(x_o) = \exp\left(Y^*_{\mathrm{OK}}(x_o) + \sigma^2_{\mathrm{OK}}/2 - \mu\right) - C \quad (4)$$

The ordinary kriging estimate at unsampled location (x_o) can be computed as follows:

$$y^*_{\mathrm{OK}}(x_o) = \sum_{i=1}^{n} \lambda_i y(x_i)$$

Because weights sum up to 1, the interpolation variance (p. 491 in [10]) can be calculated:

$$S^2_o = \sum_{i=1}^{n} \lambda_i \left(y(x_i) - y^*_{\mathrm{OK}}(x_o)\right)^2$$

Although lognormal kriging works with transformed data, i.e., data presenting reduced variance, resulting estimates will present some smoothing because it is an inherent property of weighted averages. As a consequence, the histogram of logarithms will not be reproduced. The nonbias term in Eq. 1 or 2 are not good enough to restitute the histogram of logarithms and consequently the expected value of backtransformed lognormal kriging estimates will be biased. The main concern then is to restitute the histogram of logarithms. The straightforward solution is the application of the postprocessing algorithm proposed by Yamamoto [11] that guarantees reproduction of the sample histogram and the sample semivariogram without loss of local accuracy. This method is described in detail in Yamamoto [11]. Moreover, in this paper, a new approach for computing a factor that makes the variance of corrected estimates equal to the sample variance is also described.

According to Yamamoto [11], the postprocessing algorithm is based on the following steps:

1. Run a cross-validation process to get for every sample data location the ordinary kriging estimate and the interpolation variance. Because we know the true value for sample data location we can compute the true error as:

   $$\mathrm{TrueError}(x_o) = y^*_{\mathrm{OK}}(x_o) - y(x_o)$$

   Instead of using the true error for compensating the smoothing effect, we may use the true error divided by the interpolation standard deviation:

   $$N_{S_o} = \frac{-\mathrm{TrueError}(x_o)}{S_o}$$

   This new variable was named number of interpolation standard deviation. It is important to note that the minus signal makes the compensation for the smoothing effect. Because low values are overestimated, then a correcting amount must be subtracted from ordinary kriging estimate (N_{S_o} is negative). On the other hand, for high values that are underestimated it is necessary to add a correcting amount to ordinary kriging estimate (in this case N_{S_o} is positive).
2. Run the ordinary kriging procedure for estimating N_{S_o} at all nodes of the regular grid;
3. Now run the ordinary kriging procedure to compute estimates $y^*_{\mathrm{OK}}(x_o)$ at all nodes of the regular grid. In this run, we also obtain the interpolation standard deviation associated with the estimate $y^*_{\mathrm{OK}}(x_o)$;

4. Run the postprocessing procedure for correcting the ordinary kriging estimates as:

$$y^{**}_{OK}(x_o) = y^{*}_{OK}(x_o) + N_{S_o}{*}S_o$$

Sometimes, the corrected estimate falls outside the permissible range of neighbor data used for estimation, that is, $y^{**}_{OK}(x_o) < y\min$ or $y^{**}_{OK}(x_o) > y\max$, where y min is the minimum value and y max the maximum for a given neighbor data. Thus, the correcting amount $N_{S_o}{*}S_o$ is replaced by delta where delta $= y^{*}_{OK}(x_o) - y\min$ if the signal of N_{S_o} is negative; delta $= y\max - y^{*}_{OK}(x_o)$ if the signal of N_{S_o} is positive. Thus, corrected estimates can be computed as:

$$\left\{ y^{**}_{OK}(x_o) = y^{*}_{OK}(x_o) + N_{S_o}\,S_o\text{factor} \right. \tag{5a}$$

$$\left\{ y^{**}_{OK}(x_o) = y^{*}_{OK}(x_o) + \text{delta.factor} \right. \tag{5b}$$

where factor is a constant that makes the variance of corrected estimates equal to the sample variance. The constant value factor is applied to all correcting amounts. This is the main difference between the former algorithm [11] and this one. According to Yamamoto [11], the factor was applied only for corrected estimates falling outside permissible ranges in which the correcting amount was replaced by delta times factor.

These expressions can be rewritten as:

$$y^{**}_{OK}(x_o) = y^{*}_{OK}(x_o) + y^{*}_{NS_o}(x_o)\text{factor} \tag{6}$$

where $y^{*}_{NS_o}(x_o)$ is the correcting amount that replaces either $N_{S_o}S_o$ or delta.

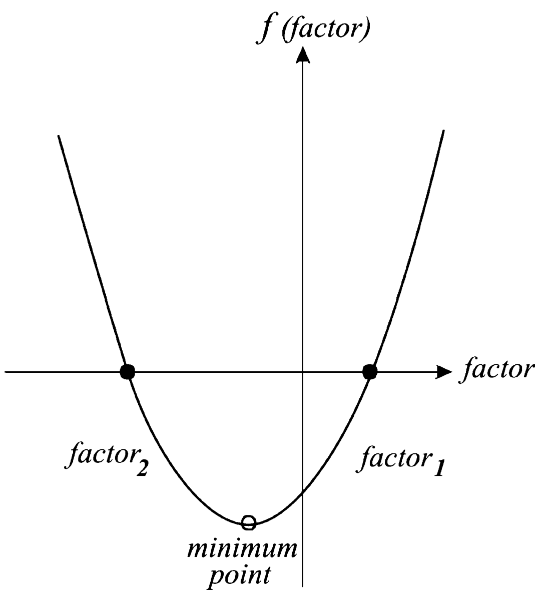

Fig. 1 Quadratic function for finding the optimum factor, which makes the variance of corrected estimates equal to the sample variance

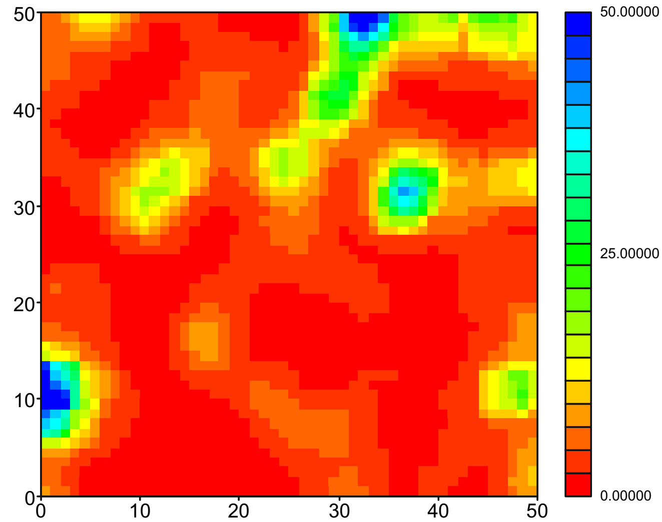

Fig. 2 Image of the exhaustive data set

Now, the problem is to find a factor that makes the variance of corrected estimates equal to the sample variance as follows:

$$\text{var}[Y(x)] = \text{var}\left[Y^{*}_{OK}(x_o) + \text{factor}\,Y^{*}_{NS_o}(x_o)\right]$$

The right side of this expression can be developed as follows:

$$\text{var}[Y(x)] = E\left[\left(Y^{*}_{OK}(x_o) + \text{factor}\,Y^{*}_{NS_o}(x_o)\right)^2\right] - E\left[\left(Y^{*}_{OK}(x_o) + \text{factor}\,Y^{*}_{NS_o}(x_o)\right)\right]^2$$

developing and rearranging we have:

$$\text{var}\left[Y^{*}_{NS_o}(x_o)\right]\text{factor}^2 + 2\text{cov}\left(Y^{*}_{OK}(x_o), Y^{*}_{NS_o}(x_o)\right)\text{factor} + \text{var}\left[Y^{*}_{OK}(x_o)\right] - \text{var}[Y(x)] = 0$$

Therefore, the best factor is a root of a quadratic function. First, let us analyze the quadratic function. The discriminant of this function is:

$$\Delta = 4\text{cov}\left(Y^{*}_{OK}(x_o), Y^{*}_{NS_o}(x_o)\right)^2 - 4\text{var}\left[Y^{*}_{NS_o}(x_o)\right]\left(\text{var}\left[Y^{*}_{OK}(x_o)\right] - \text{var}[Y(x)]\right)$$

Because var$[Y(x)]$ is always greater than var$\left[Y^{*}_{OK}(x_o)\right]$ because of the smoothing effect, the discriminant is always greater than 0; therefore, the quadratic function has two real roots. In addition to this, as var$\left[Y^{*}_{NS_o}(x_o)\right]$ is always greater

Table 1 Summary statistics for the exhaustive data set and samples

Statistics	Exhaustive	Samples					
		1	2	3	4	5	6
No. of data	2500	121	121	121	121	121	121
Mean	4.788	4.646	4.917	4.970	4.665	5.005	4.998
Standard deviation	7.178	6.966	8.247	7.769	7.080	8.037	8.110
Coefficient of variation	1.499	1.499	1.677	1.563	1.518	1.606	1.623
Maximum	72.280	44.503	54.354	53.723	36.109	67.518	52.429
Upper quartile	5.186	5.098	4.315	5.408	4.668	4.967	4.744
Median	2.185	2.070	1.897	2.015	1.763	2.153	2.297
Lower quartile	0.952	0.932	0.959	1.049	0.942	0.930	1.049
Minimum	0.153	0.253	0.282	0.210	0.238	0.238	0.238

than 0, the quadratic function is concave up, having a minimum point:

$$\text{factor}_{\min} = \frac{-\text{cov}\left(Y^*_{\text{OK}}(x_{\text{o}}), Y^*_{NS_{\text{o}}}(x_{\text{o}})\right)}{\text{var}\left[Y^*_{NS_{\text{o}}}(x_{\text{o}})\right]}$$

As the covariance between ordinary kriging estimates and correcting amounts is always positive, the minimum point is always negative. The quadratic function can be seen on a graph as shown in Fig. 1.

The roots of the quadratic function can be calculated as:

$$\text{factor}_{1,2} = \frac{-2\text{cov}\left(Y^*_{\text{OK}}(x_{\text{o}}), Y^*_{NS_{\text{o}}}(x_{\text{o}})\right) \pm \sqrt{\Delta}}{2\text{var}\left[Y^*_{NS_{\text{o}}}(x_{\text{o}})\right]}$$

Because the square root of the discriminant is always greater than $2\text{cov}\left(Y^*_{\text{OK}}(x_{\text{o}}), Y^*_{NS_{\text{o}}}(x_{\text{o}})\right)$, the positive root will be:

$$\text{factor} = \frac{-2\text{cov}\left(Y^*_{\text{OK}}(x_{\text{o}}), Y^*_{NS_{\text{o}}}(x_{\text{o}})\right) + \sqrt{\Delta}}{2\text{var}\left[Y^*_{NS_{\text{o}}}(x_{\text{o}})\right]}$$

This is the optimum factor, which makes possible the variance of corrected estimates be equal to the sample variance.

Expression 6 allows the computation of corrected estimates in the logarithmic domain. Moreover, they reproduce the logarithmic semivariogram and the histogram of logarithms with no loss of local accuracy. Corrected estimates are then backtransformed as follows:

$$z^{**}_{\text{OLK}}(x_{\text{o}}) = \exp\left(y^*_{\text{OK}}(x_{\text{o}}) + y^*_{NS_{\text{o}}}\text{factor}\right) - C \quad (7)$$

where $y^*_{NS_{\text{o}}}$ factor appears as the nonbias term. The expected value of corrected backtransformed estimates is equal to the sample mean because the backtransform is based on corrected lognormal kriging estimates that present the same histogram of logarithms. In other words, when sample data are transformed into logarithms we have a sample histogram of logarithms. Now, if corrected lognormal kriging estimates presenting the same histogram as the sample histogram of logarithms are backtransformed, then we are reproducing histogram and sample statistics.

As a new proposal, backtransformed values according to expression 7 have to be proven to present neither conditional bias nor sensitivity to the nonbias term, such as reported for expression 2.

Moreover, the influence of constant C in expressions 4 and 7 is examined in this paper.

3 Materials and methods

For this study, an exhaustive data set composed of 50×50 values on a regular grid was considered (Fig. 2). This data set was derived from the public domain file named *true.dat* (p. 35 in [3]). Actually, the secondary variable of the

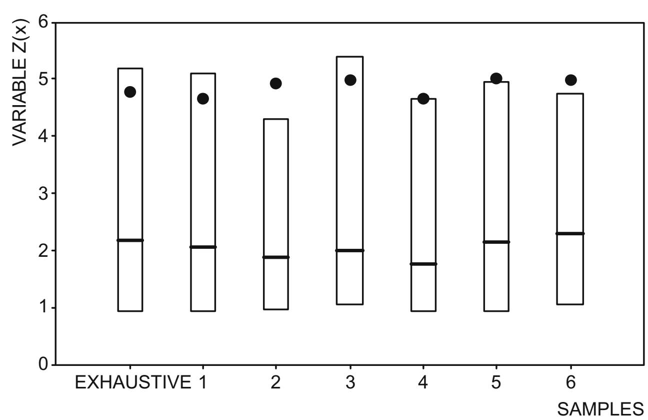

Fig. 3 Interquartile box plots for the exhaustive data and samples. *Black circle* indicates the mean, *thick line* the median

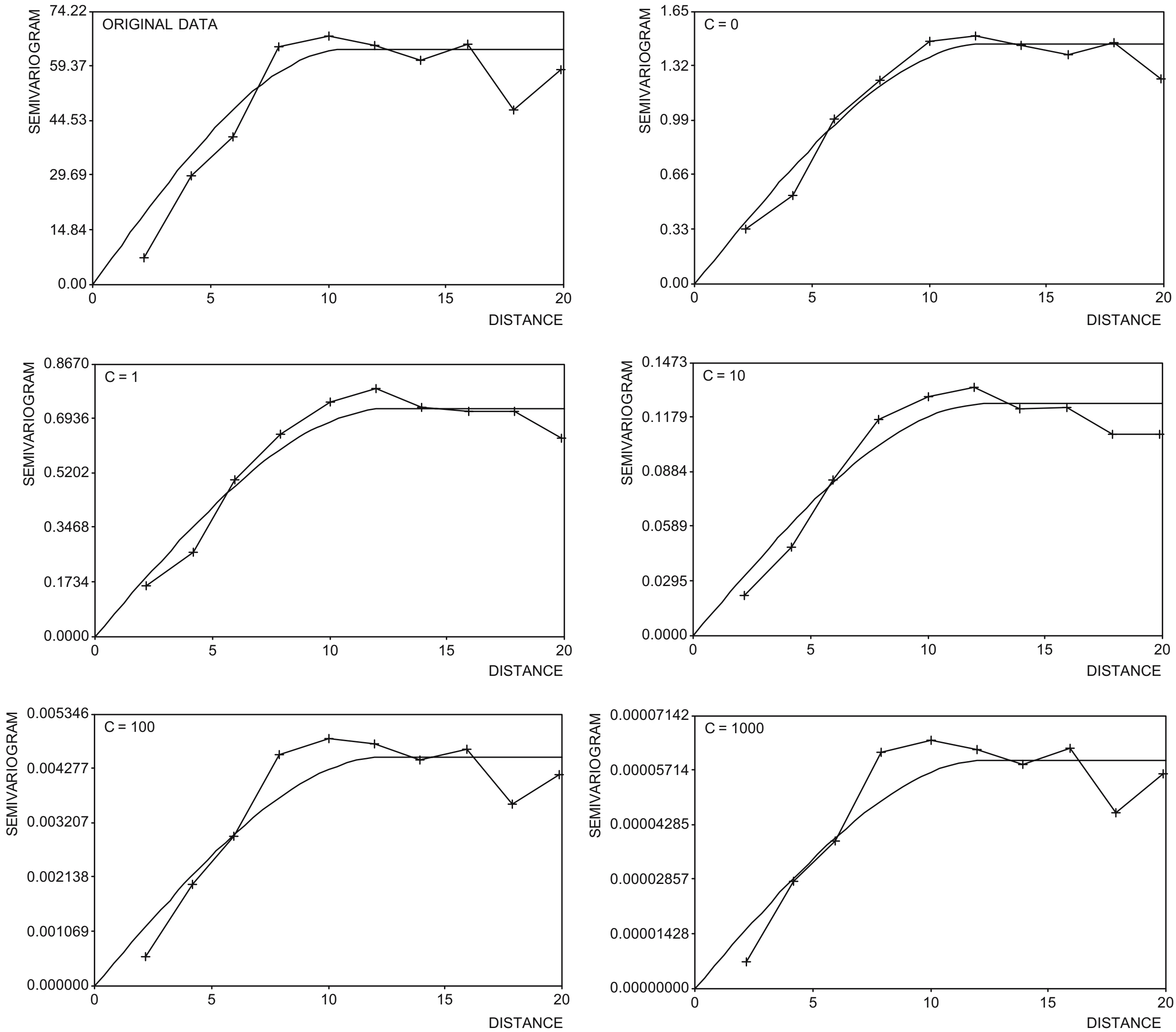

Fig. 4 Semivariogram models for sample 2

original file was mathematically transformed into another variable to enhance the skewness of the frequency distribution, as follows:

$$Z\log(x) = 3.9^{\mathrm{Ln}(Z(x)+1.0)} - 1.0$$

where $Z(x)$ is the secondary variable from true.dat and $Z\log(x)$ is the transformed variable.

From this exhaustive data set, six samples composed of 121 data points were drawn based on stratified random sampling. In fact, only samples presenting a coefficient of variation greater than or equal to 1.5 were retained for this study. The purpose was to test the proposed method working with samples presenting strong skewness. Summary statistics for these samples are presented in Table 1.

Table 2 Number of times that expressions 5a and 5b were used for correcting the smoothing effect of lognormal kriging estimates (for a log-normal constant equal to 0) and for expression 5b the number of cases where corrected estimates were less than y min and greater than y max

Sample	Expression 5a	Expression 5b	$\overset{**}{y}_{\mathrm{OK}}(x_o) < y\min$	$\overset{**}{y}_{\mathrm{OK}}(x_o) > y\max$
1	2,129	371	167	204
2	2,174	326	161	165
3	2,158	342	197	145
4	2,104	396	164	232
5	2,165	335	151	184
6	2,124	376	170	206

Table 3 Mean values for backtransformed values (expressions 4 and 7)

Sample	Expression 4					Expression 7				
	0	1	10	100	1,000	0	1	10	100	1,000
1	4.268	4.183	4.443	4.668	4.709	4.649	4.620	4.618	4.646	4.652
2	4.345	4.257	4.582	4.898	4.970	4.752	4.828	4.903	4.919	4.925
3	4.519	4.410	4.645	4.934	5.002	4.949	4.952	4.935	4.967	4.972
4	4.339	4.273	4.518	4.767	4.810	4.514	4.542	4.620	4.669	4.678
5	4.375	4.329	4.508	4.749	4.804	5.053	5.017	5.004	5.005	5.005
6	4.495	4.409	4.675	4.988	5.060	4.856	4.868	4.946	4.998	5.005

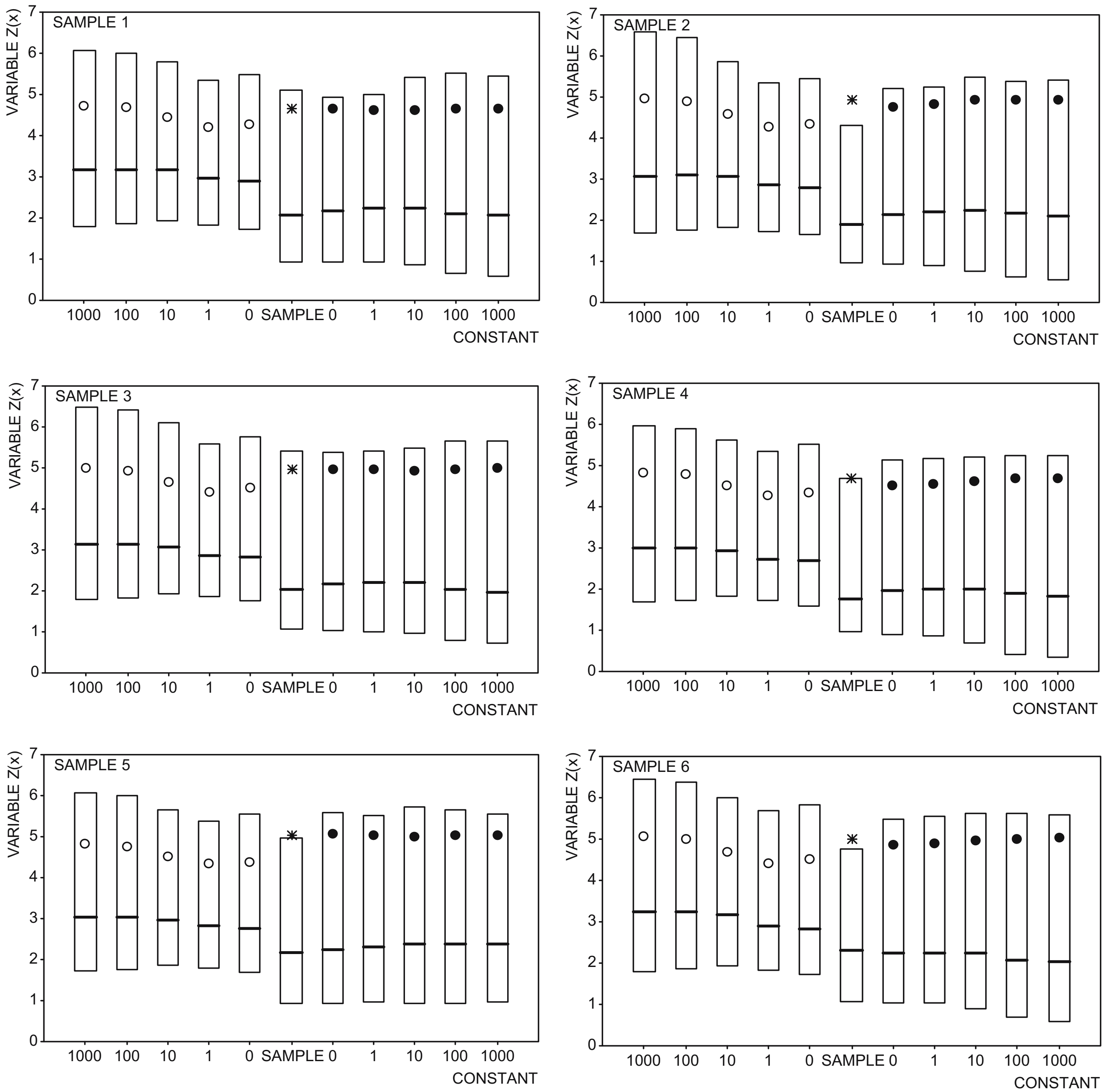

Fig. 5 Four summary statistics displayed as interquartile box plots for backtransformed values. *Empty circle* indicates the mean of backtransformed value after expression 4, *black circle* the mean of backtransformed value after expression 7, *asterisk* the sample mean, *thick line* the median

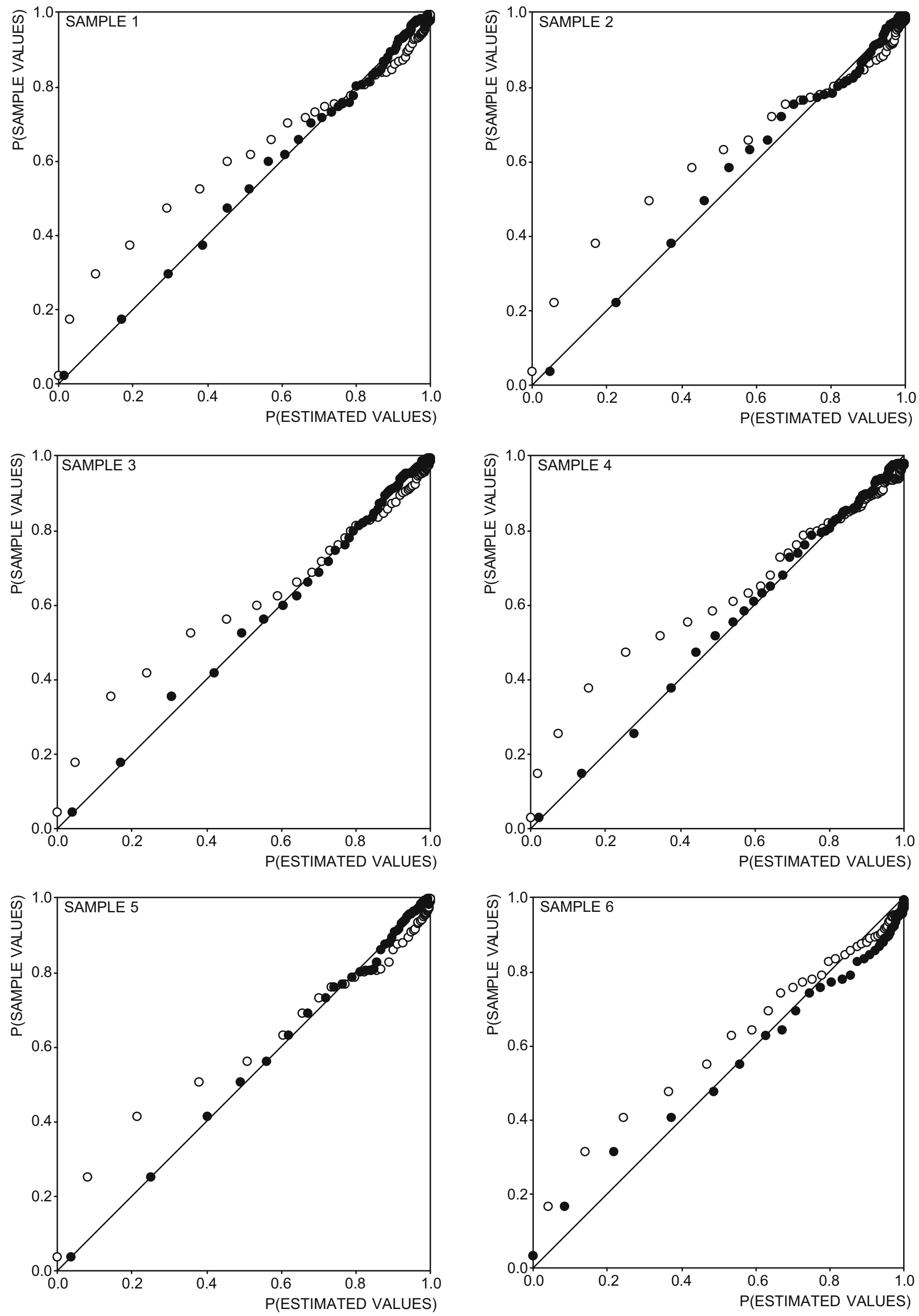

Fig. 6 P–P plots comparing estimated distributions with sample distribution. *Empty circle* indicates the backtransformed values after expression 4, *black circle* the backtransformed value after expression 7

Figure 3 presents interquartile range box plots for the exhaustive data set and samples. This graphic display, that is, a box plot with no whiskers, was chosen because it allows a more precise comparison between samples. Although the mean values are close to the exhaustive mean, the samples are very different from each other due to the sampling error. Moreover, only sample 1 presents skewness close to the exhaustive data, whereas all the others show greater skewness as confirmed by Table 1.

Thus, for each sample, experimental semivariograms were computed for original and transformed data. For logarithmic transformation, the following constant values: 0, 1, 10, 100, and 1,000 have been chosen. Therefore, for each sample, six semivariograms were computed and modeled. For illustration purposes, Fig. 4 presents semivariograms and models for sample 2. This figure clearly demonstrates that the nil constant produces the best semivariogram of logarithms. Increasing the

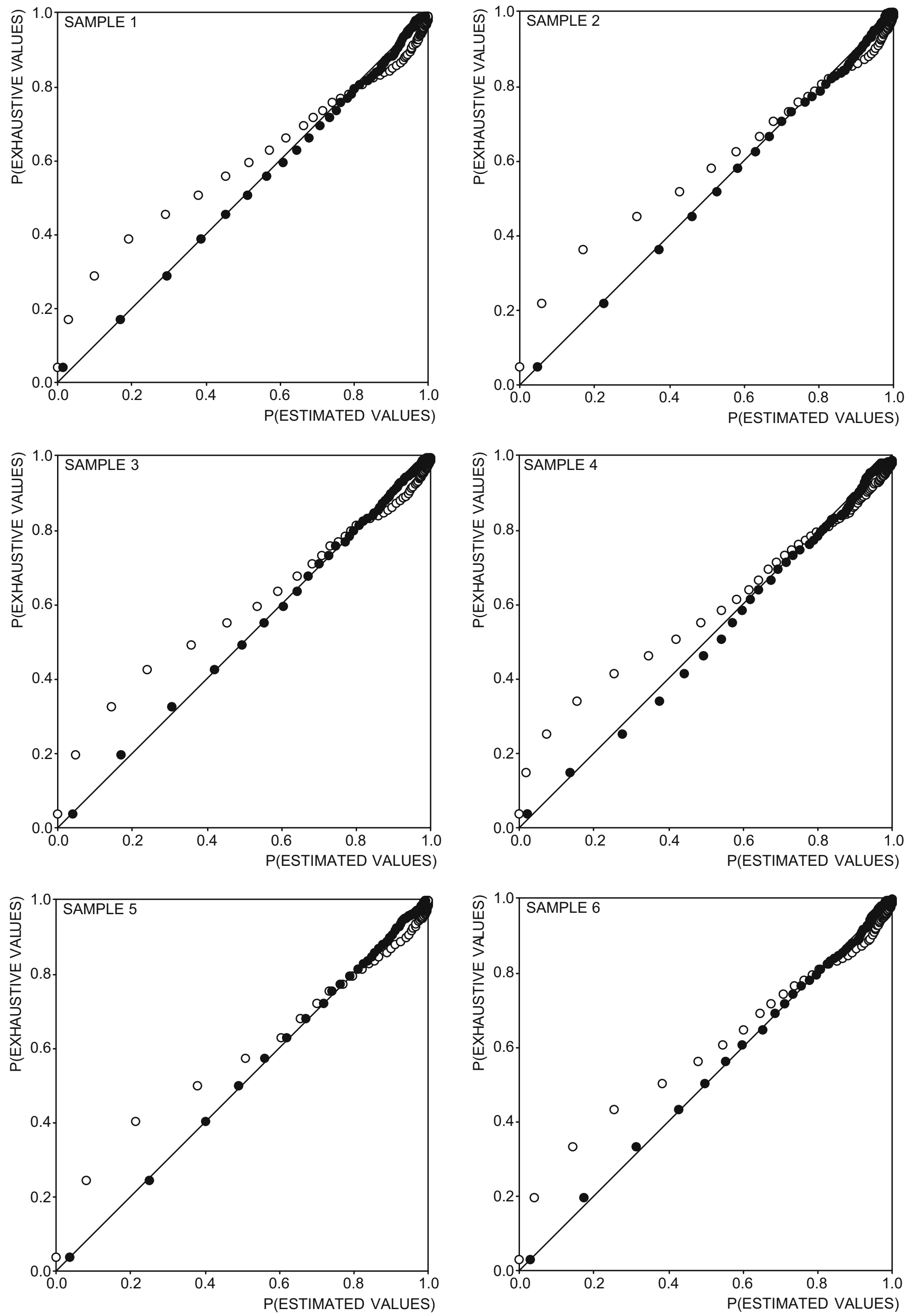

Fig. 7 P–P plots comparing estimated distributions with the exhaustive distribution. *Empty circle* indicates the backtransformed values after expression 4, *black circle* the backtransformed value after expression 7

constant, the semivariogram becomes increasingly similar to the original semivariogram. The difference is that semivariogram function values are scaled down.

A computer program was used to calculate both backtransformed ordinary lognormal kriging estimates (expression 4) and backtransformed corrected estimates (expression 7). For all runs, a unique neighborhood will be considered, i.e., three data points per quadrant. Actually, this program named Crossordkrig2E has been tested over and over and it worked well in all cases (hundreds of examples).

4 Results and discussion

Before presenting all results it would be interesting to see how many times either expression 5a or 5b are used to

Table 4 Correlation coefficients between estimated (expressions 4 and 7) and actual values

Sample	Expression 4					Expression 7				
	0	1	10	100	1,000	0	1	10	100	1,000
1	0.889	0.892	0.892	0.890	0.888	0.903	0.906	0.897	0.875	0.868
2	0.913	0.913	0.917	0.912	0.900	0.913	0.913	0.896	0.871	0.861
3	0.925	0.923	0.919	0.910	0.905	0.927	0.923	0.914	0.892	0.885
4	0.841	0.846	0.848	0.845	0.843	0.893	0.890	0.873	0.851	0.845
5	0.878	0.879	0.881	0.879	0.876	0.904	0.909	0.903	0.885	0.876
6	0.893	0.898	0.903	0.901	0.899	0.920	0.917	0.897	0.866	0.854

examine the effectiveness of the proposed solution. Evidently, the greater the number of times using expression 5a, the better are resulting corrected estimates. Just for illustration purposes we present in Table 2 the number of times that each expression was used for postprocessing lognormal kriging estimates for all samples and using a constant equal to 0.

As we can see the vast majority of cases the corrected estimates were based on expression 5a, proving the effectiveness of the proposed solution. Moreover, Table 2 gives additional information about the number of times that corrected estimates were less than the minimum and greater than the maximum of neighbor data points. According to these data, Eq. 5b is more likely to be used when $y^{**}_{\mathrm{OK}}(x_o) > y\,\mathrm{max}$.

Mean values for backtransformed values according to expressions 4 and 7 are listed in Table 3. Interquartile box plots comparing these mean values are presented in Fig. 5. Mean values for backtransformed values (expression 7) are clearly closer to the sample mean, proving that these backtransformed estimates are not biased. On the other hand, mean values for backtransformed values after expression 4 for constants less than or equal to 10 are always less than the sample mean; that is, they are biased. Moreover, they tend to the sample mean when the constant C increases. In fact, larger constants produce results very close to original data. For instance, experimental semivariogram for original data looks like the semivariogram of logarithms with a constant equal to 1,000. In addition, when the constant is 1,000, the mean values of backtransformed results (expression 4) are very close to sample means. Based on these observations we can conclude that higher constants reduce the nonbias term in Eq. 2, making backtransformed values closer to original values. However, in this case lognormal kriging does not take advantage of transformed values.

Medians of backtransformed values after expression 4 are always greater than sample medians. However, expression 7 provides closer medians to the sample medians.

Estimated distributions can be compared with both sample (Fig. 6) and exhaustive (Fig. 7) distributions. Only P–P plots for the nil constant will be presented, whereas P–P plots for other constants will not be shown to avoid repetitive figures. As mentioned before, the target statistics are always the sample statistics. However, inferences from sample data about population statistics can be made.

Both figures show that backtransformed values, according to expression 7, provide the best results. In this sense, Figs. 6 and 7 prove that the proposed approach reproduces not only the sample histogram but also the exhaustive histogram. Now it is possible to make inferences about exhaustive data from sample results. This is possible because results provided by the new proposal are more precise and accurate than conventional lognormal kriging. Furthermore, correlation coefficients and root mean square (RMS) errors can be computed for quantitative comparisons between estimated and actual values. Tables 4 and 5, respectively, present correlation coefficients and RMS errors.

According to Table 4, correlation coefficients between backtransformed values after expression 7 and actual values

Table 5 RMS errors between estimated (expressions 4 and 7) and actual values

Sample	Expression 4					Expression 7				
	0	1	10	100	1,000	0	1	10	100	1,000
1	4.072	4.224	4.072	3.792	3.727	3.107	3.057	3.186	3.521	3.634
2	3.622	3.773	3.569	3.266	3.212	3.130	3.314	3.614	3.984	4.169
3	3.359	3.576	3.499	3.298	3.270	2.814	2.866	3.016	3.447	3.621
4	4.240	4.328	4.219	4.063	4.035	3.239	3.278	3.518	3.860	3.957
5	3.891	4.106	3.925	3.705	3.663	3.594	3.429	3.491	3.742	3.886
6	3.725	3.828	3.667	3.406	3.358	2.889	2.959	3.383	3.986	4.203

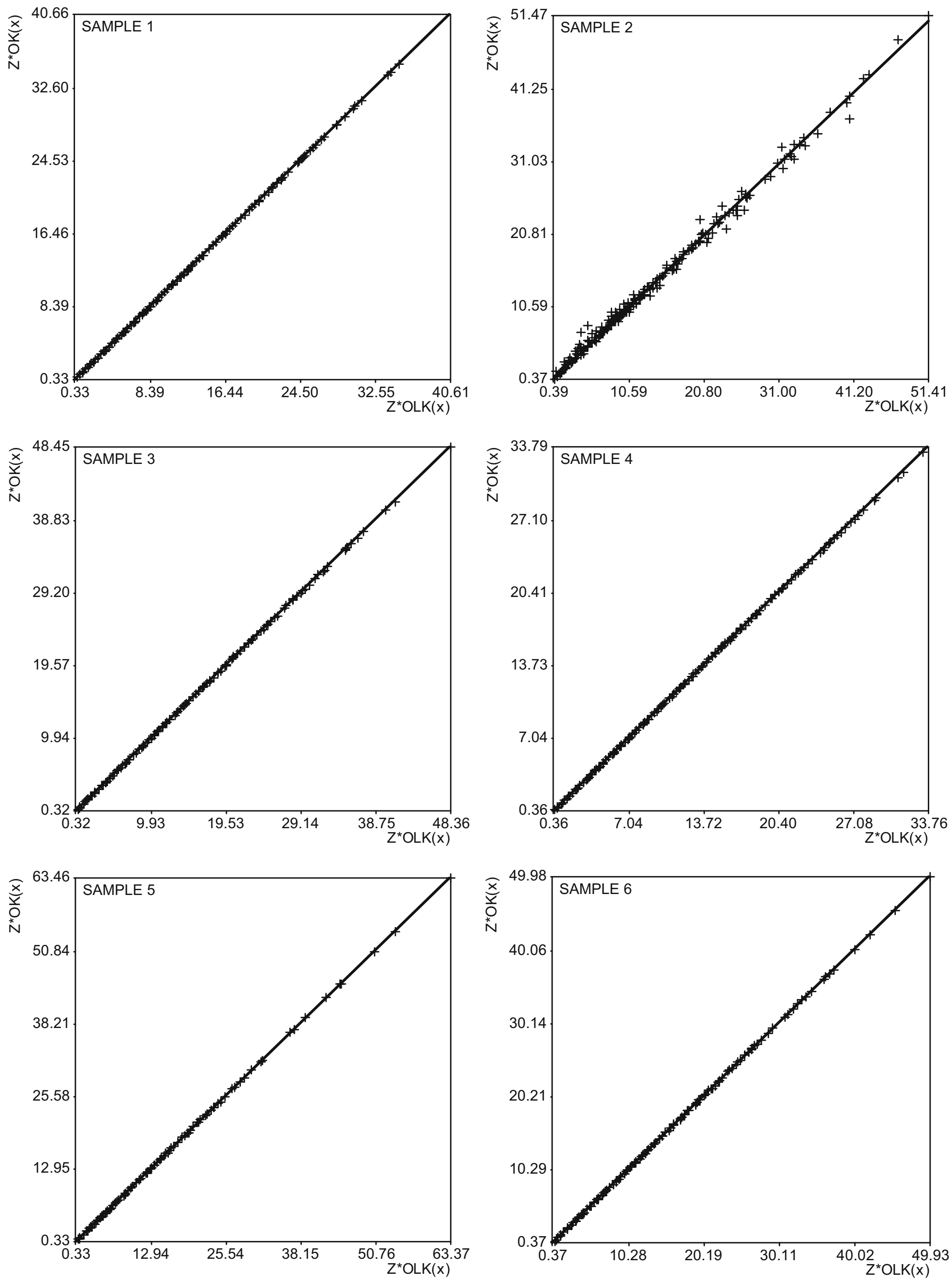

Fig. 8 Scattergrams of ordinary kriging estimates against backtransformed lognormal kriging estimates after expression 4 based on a constant equal to 1,000

are generally greater than those after expression 4. The only difference occurs for sample 2 in which a correlation coefficient equal to 0.917 (expression 4) is slightly superior than 0.913 computed after expression 7. Correlation coefficients measure precision, whereas RMS errors measure accuracy. Table 5 shows that the smaller RMS errors are associated with backtransformed values according to expression 7. Moreover, the smaller errors occur for constants between 0 and 1. Backtransformed values after expression 7 show RMS errors increasing as constant values increase. On the contrary, backtransformed values using expression 4 show RMS errors decreasing as constant values increase. In fact, this happens because as long as the constant increases, the nonbias term tends to 0; therefore, the backtransformed value will approach the ordinary kriging estimate. This reconfirms that the nonbias term for expression 4 is not appropriate. On the other hand, backtransformed values after expression 7 present smaller RMS errors for low constants (between 0 and 1), proving that the nonbias term associated with expression 7 does work. In fact, the proposed backtransform of lognormal kriging estimates takes advantage of the logarithmic transformation.

As mentioned before, higher constants tend to reduce the nonbias term expression 4, making resulting estimates very

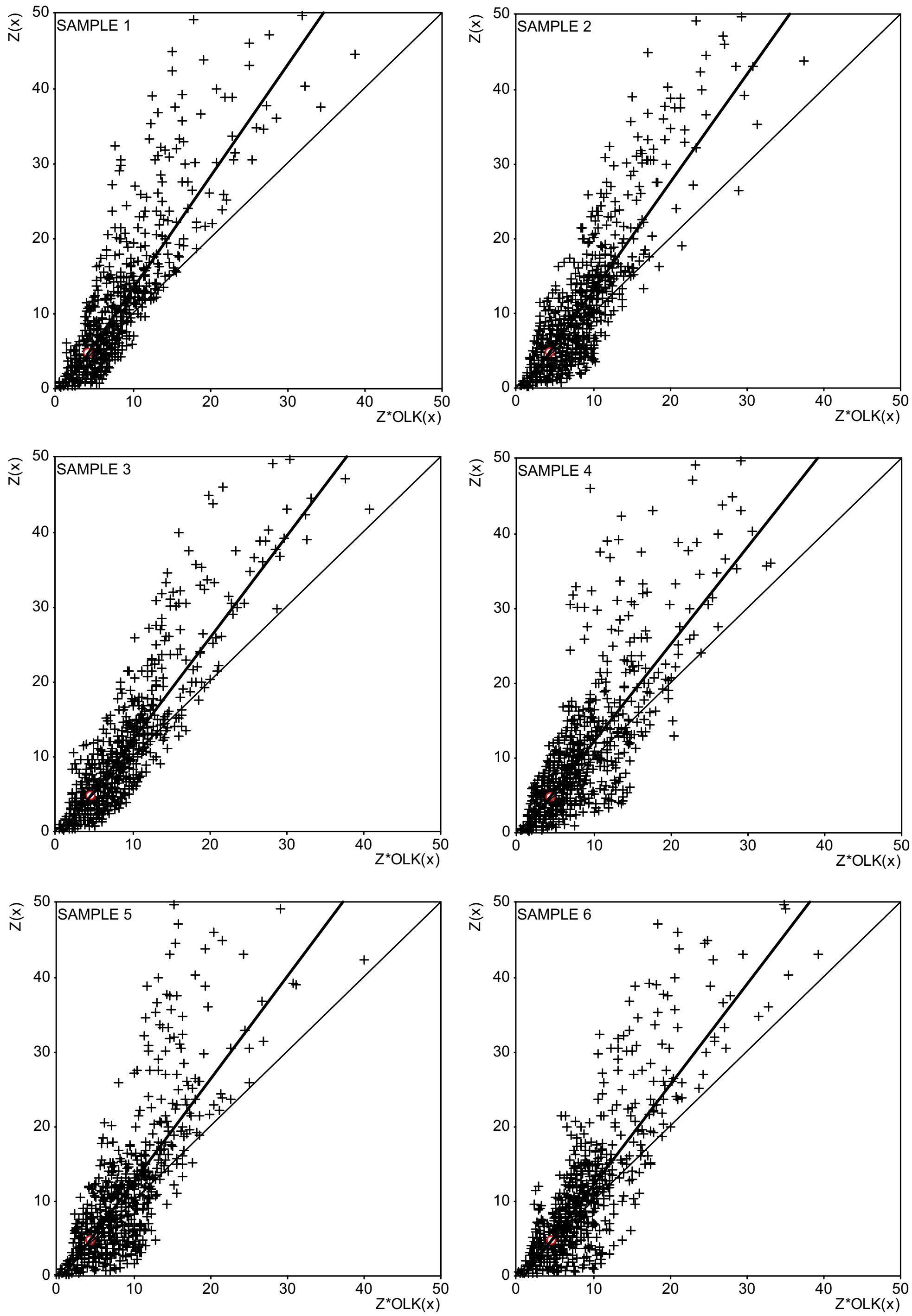

Fig. 9 Scattergram of actual and backtransformed values after expression 4 based on null constant. *Thick line* indicates the minimum square line

close to ordinary kriging estimates. Hence, for illustration purposes, Fig. 8 presents a scattergram between ordinary kriging estimates against backtransformed lognormal kriging estimates after expression 4 based on a constant equal to 1,000.

Figure 8 presents scattergrams showing perfect linear relationships between ordinary kriging estimates and backtransformed (expression 4) lognormal kriging estimates based on logarithmic transformation with a constant equal to 1,000. The nonbias term in Eq. 4 practically vanishes, making the exponential of lognormal kriging estimates equal to ordinary kriging estimates. In this case, the lognormal kriging does not take advantage of the logarithmic transformation; that is, few high values will impact on almost all kriging estimates.

Relationships between estimated and actual values can be visualized as scattergrams. Figures 9 and 10 present, respectively, scattergrams of actual values against back-

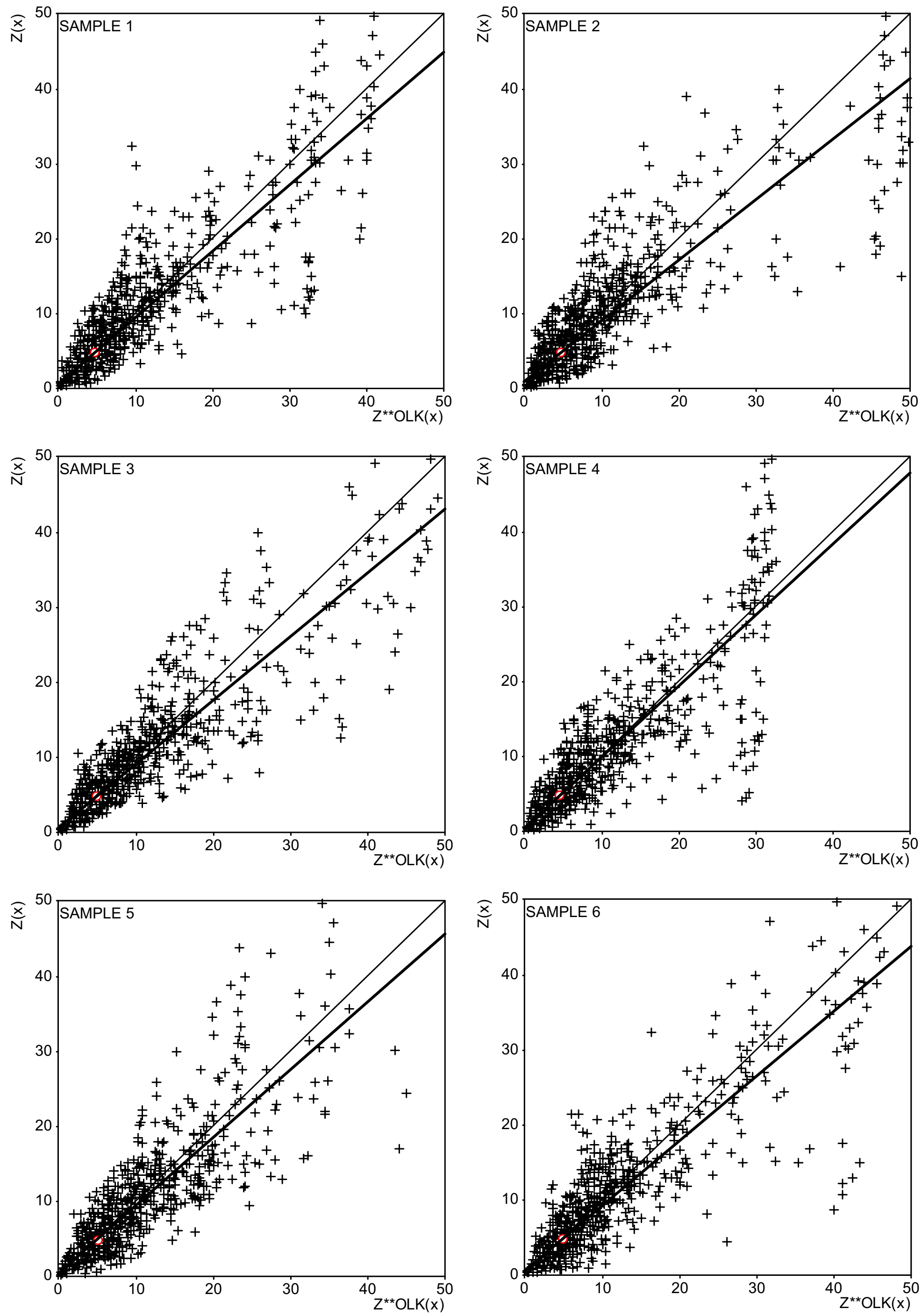

Fig. 10 Scattergram of actual and backtransformed values after expression 7 based on null constant. *Thick line* indicates the minimum square line

transformed estimates after expression 4 and against backtransformed values after expression 7 based on a constant equal to 0. Expression 7 produces backtransformed values closer to actual ones. Note that in Fig. 10, minimum square regression lines are closer to the 45° line. The difference between regression lines in Figs. 9 and 10 is in slopes; slopes in Fig. 9 are always greater than 1, and in Fig. 10, they are always less than 1. This behavior, according to Yamamoto ([11], p. 91), is due to the smoothing effect. In these scattergrams, slopes greater than 1 mean estimates

Table 6 Corrective factor and statistics for backtransformed values after expression 4 multiplied by the corrective factor

Sample	Factor k_o	Mean	SD
1	1.088	4.646	4.387
2	1.132	4.917	5.065
3	1.010	4.970	5.171
4	1.975	4.665	4.718
5	1.144	5.005	5.165
6	1.112	4.998	5.091

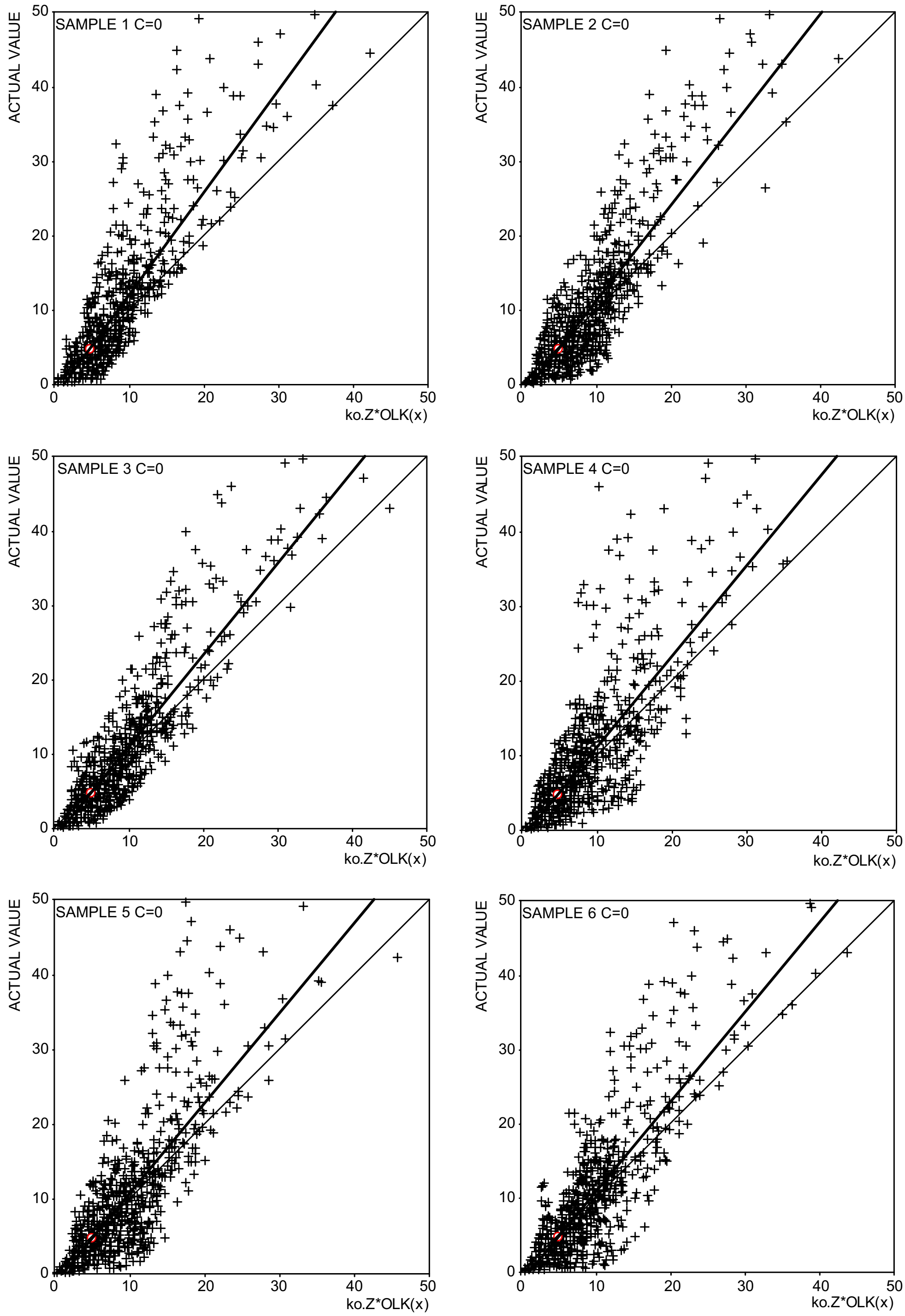

Fig. 11 Scattergram of actual and backtransformed values after expression 4 multiplied by a corrective factor based on null constant. *Thick line* indicates the minimum square line

presenting smoothing effect. On the other hand, estimates with no smoothing effect will present slope less than 1. Therefore, backtransformed lognormal kriging estimates after expression 7 do not present a smoothing effect. In other words, expression 7 brought all characteristics, namely, histogram and semivariogram reproduction, in the logarithmic domain to the original measurement domain. It reconfirms the reliability of the nonbias term of expression 7. In fact, the nonbias term is the amount used for correcting the smoothing effect of lognormal kriging estimates.

Before going further, let us analyze the influence on backtransformed estimates after applying the corrective factor k_o proposed by Journel and Huijbregts ([5], p. 572). When a constant factor is used to multiply a random variable its mean is multiplied by this constant, but its variance will be multiplied by square of constant. Table 6

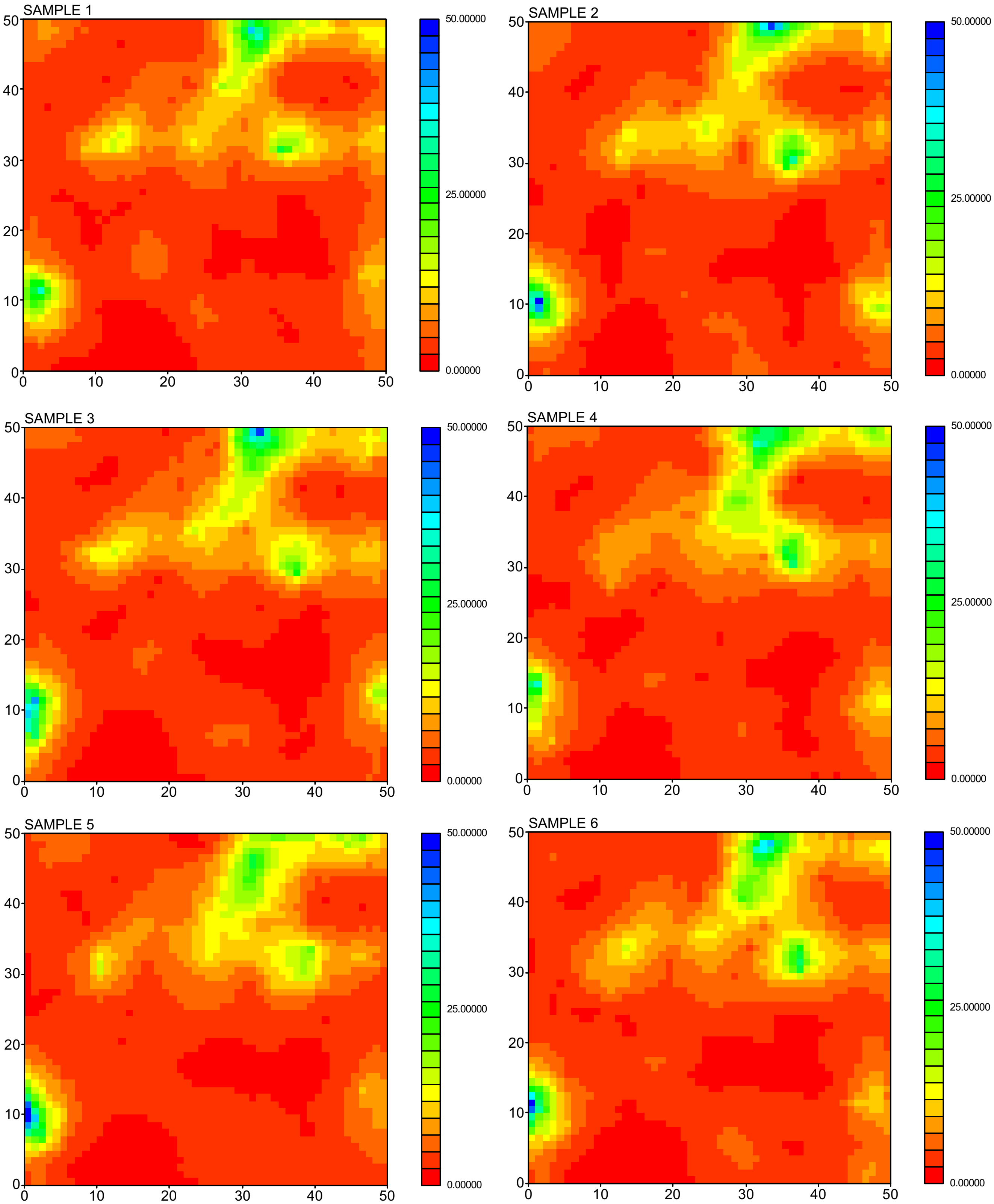

Fig. 12 Resulting images of backtransformed values after expression 4

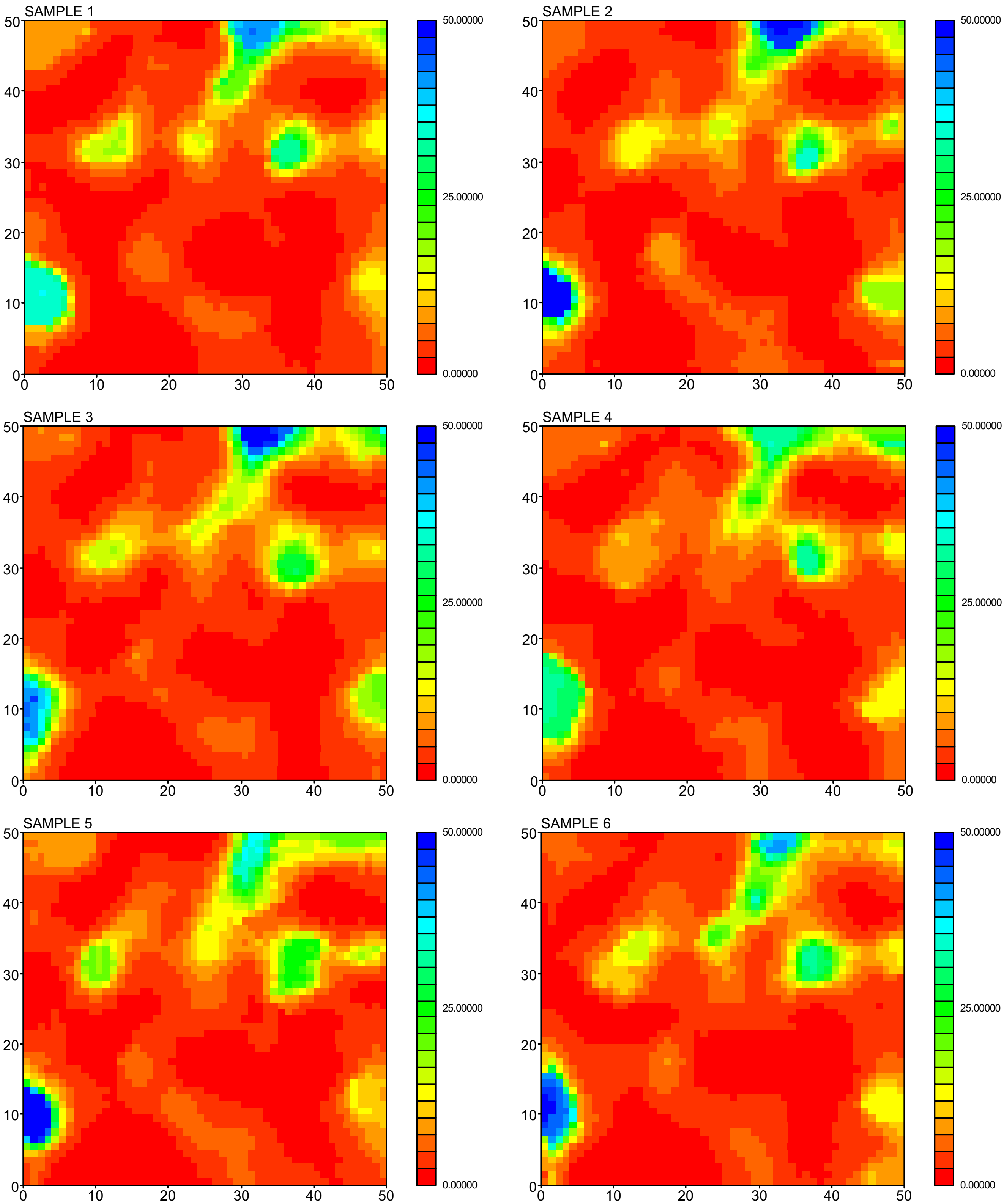

Fig. 13 Resulting images of backtransformed values after expression 7

presents results after applying the corrective factor. Comparing these results with sample statistics, we conclude that the estimated means are equal to the sample means; however, standard deviations are always less than sample standard deviation. Therefore, the corrective factor produces unbiased estimates, but they are still smoothed. Evidently, we can reconfirm that backtransformed estimates after expression 4 and multiplied by a corrective factor are still smooth looking at scattergrams between actual values and corrected estimates (Fig. 11). All slopes of minimum square regression lines are greater than 1, showing that corrected estimates (expression 4) multiplied by a corrective factor are smoothed.

Resulting images after backtransforming lognormal kriging estimates are presented in Figs. 12 and 13. When pairs of figures are compared, it is clear that expression 7 provides images closer to the exhaustive image, reconfirming what has been concluded about effectiveness of the nonbias term used for backtransforming lognormal kriging estimates. Moreover, resulting images from backtransformed lognormal kriging estimates after expression 7 reproduces the pattern as shown by the exhaustive image.

5 Conclusions

This paper proposed an alternative method for backtransforming lognormal kriging estimates that does not present problems reported by several authors. The core of this proposal is based on the nonbias term that is exactly the correcting amount for the smoothing effect of ordinary lognormal kriging estimates. Resulting estimates reproduce the sample histogram and, consequently, the sample mean. Moreover, the nonbias term does not depend directly on a semivariogram model but only on interpolation standard deviation. Comparing backtransformed estimates with exhaustive values, the proposed method always presented higher correlation coefficients and smaller RMS errors, confirming its effectiveness. Regarding the constant C, low values (at most equal to 1) are recommended because high values practically eliminate the advantage of the logarithmic transformation.

Acknowledgements The author wishes to thank *CNPq–Conselho Nacional de Desenvolvimento Científico e Tecnológico* (process 302537/2004-0) for funding his research. The author is grateful to an anonymous reviewer for comments and suggestions that improved this manuscript.

References

1. Chilès, J.P., Delfiner, P.: Geostatistics — Modeling Spatial Uncertainty. Wiley, New York (1999)
2. David, M.: Handbook of Applied Advanced Geostatistical Ore Reserve Estimation. Elsevier, Amsterdam (1988)
3. Deutsch, C.V., Journel, A.G.: GSLIB: Geostatistical Software Library and User's Guide. Oxford University Press, New York (1992)
4. Journel, A.G.: The lognormal approach to predicting local distributions of selective mining unit grades. Math. Geol. **12**, 285–303 (1980)
5. Journel, A.G., Huijbregts, C.J.: Mining Geostatistics. Academic, London (1978)
6. Koch, G.S., Link, R.F.: Statistical Analysis of Geological Data. Dover Publications, New York (1971)
7. Rivoirard, J.: A review of lognormal estimator for *in situ* reserves. Math. Geol. **22**, 213–221 (1990)
8. Roth, C.R.: Is lognormal kriging suitable for local estimation? Math. Geol. **30**, 999–1009 (1998)
9. Saito, H., Goovaerts, P.: Geostatistical interpolation of positively skewed and censored data in a Dioxin-contaminated site. Environ. Sci. Technol. **34**, 4228–4235 (2000)
10. Yamamoto, J.K.: An alternative measure of the reliability of ordinary kriging. Math. Geol. **32**, 489–509 (2000)
11. Yamamoto, J.K.: Correcting the smoothing effect of ordinary kriging estimates. Math. Geol. **37**, 69–94 (2005)