A Recursive Stepwise Transformation Scheme for Limited Multivariate Data

Behrang Koushavand and Clayton V. Deutsch

Abstract

There are very well established multivariate Geostatistical techniques, but most of these methods require many data, a strong assumption of multivariate Gaussian distribution or can be applied to a limited number of variables. This paper presents a new method based on a modification to the stepwise transformation. A recursive scheme is applied. The bivariate stepwise transformation is performed recursively between first and the rest of variables to generate a new set of variables which are uncorrelated with first variable. Then the first variable is removed and the second variable is used with the rest of variables. The process is performed until all variables are uncorrelated with each other. The advantage of the method is it can be done with few data samples and, therefore, with many variables.

Introduction

The stepwise conditional transformation (Leuangthong, 2003) is a well-established robust transformation technique for multivariate data. It is used to transfer multivariate correlated data to a Gaussian space that none of the variables are correlated. At this stage all very well-known Gaussian simulation techniques such as Sequential Gaussian Simulation (SGS) can be employed to each of the transferred variables separately. The back transformation process is also very straightforward algorithm. Figure 1 shows the schematic illustrations of stepwise transformation for two variables. Leuangthong (2003) concluded that transformed variables have not significant correlation. For more than two variables, the algorithm can be applied too. Back transformation process honors the distributions and correlation of variables at original unit. Full stepwise transformation needs many data samples to identify all conditional distributions. The conditional distributions will not be representative with fewer samples. Leuangthong (2003) suggested supplementary algorithms to generate new pseudo samples based on a multivariate distribution. The algorithm presented here use only bivariate stepwise transformation in a recursive scheme. Therefore it can be done with few samples and many variables.

Methodology

The recursive stepwise transformation can be applied for a multivariate distribution with n dimension. Instead of doing one step-wise transformation using all n variables, maximum $\frac{(n-1)(n)}{2}$ times bivariate stepwise transformations are performed recursively.

The steps of the recursive stepwise transformation for n variables are as follow:

1. Perform a Normal score transformation to first variable : $Y_1^{(1)}$
2. Use bivariate stepwise transformation to transfer next variable using transformed first variable: $Y_2^{(1)},...,Y_n^{(1)}$
3. Repeat step 2 until all variables have been transferred to normal scores. At this step, first variable is uncorrelated with other n-1 variables.
4. Remove first variable and choose transferred second variables: $Y_2^{(1)}$
5. Do bivariate stepwise transformation using second variable and rest of n-2 variables: $Y_3^{(2)},...,Y_n^{(2)}$
6. Repeat steps 4 to 5 n-1 times.

Table 1 show the conditional transformation table for a five variable case. Not that for this particular case, one normal score transformation and 10 stepwise transformations should be done. At the end, there will be 5 none correlated variables: $Y_1^{(1)}$, $Y_2^{(1)}$, $Y_3^{(2)}$, $Y_4^{(3)}$ and $Y_5^{(4)}$. Lower index shows the variable number and upper index shows the number of transformation that has been applied for that variables. For example $Y_4^{(3)}$ shows that for variable 4, three times stepwise transformation has been applied.

Case study

The same steps have been taken for Nickel data set. There are more than 29000 samples. Each variable was transformed to normal scores independently. The scatter plot of normal score values of variables are showed in Figure 2. The algorithm was applied to this data set with 7 variables. Figure 3 shows the scatter plot of transferred variables. Scatter plot of samples and simulated values between variables are plotted in Figure 4 and Figure 5 respectively.

Conclusion

A new method based on stepwise conditional transformation has been presented to transfer high dimensional correlated variable to non-correlated Gaussian variables. There is no need for co-simulation of transformed variables. All univariate simulation methods can be used for each of transformed variables independently. Back transformed values reproduced the histogram, variogram and bivariate structures on scatter plot. Because only two variables are used in stepwise transformation, a few hundred of samples are enough for transformation of large number of correlated variables.

References

[1] Deutsch, C. V. and Journel, A. G. (1998). GSLIB : geostatistical software library and user's guide. in *Applied geostatistics series*, vol. 2. New York: Oxford University Press, pp. 369.

[2] Goovaerts, P. (1997). *Geostatistics for natural resources evaluation.* Oxford University Press, New York, Pages 483.

[3] Leuangthong, O. (2003). Stepwise Conditional Transformation for Multivariate Geostatistical Simulation.Thesis, University of Alberta, Edmonton, Pages 187.

Table 1. The transformation table of recursive stepwise for five variables.

z_1	z_2	z_3	z_4	z_5
$Y_1^{(1)} = G^{-1}\left[F_1(z_1)\right]$	z_2	z_3	z_4	z_5
	$Y_2^{(1)} = G^{-1}\left[F_{2\|1}(z_2 \| z_1)\right]$	$Y_3^{(1)} = G^{-1}\left[F_{3\|1}(z_3 \| z_1)\right]$	$Y_4^{(1)} = G^{-1}\left[F_{4\|1}(z_4 \| z_1)\right]$	$Y_5^{(1)} = G^{-1}\left[F_{5\|1}(z_5 \| z_1)\right]$
		$Y_3^{(2)} = G^{-1}\left[F_{3\|2}\left(Y_3^{(1)} \| Y_2^{(1)}\right)\right]$	$Y_4^{(2)} = G^{-1}\left[F_{4\|2}\left(Y_4^{(1)} \| Y_2^{(1)}\right)\right]$	$Y_5^{(2)} = G^{-1}\left[F_{5\|2}\left(Y_5^{(1)} \| Y_2^{(1)}\right)\right]$
			$Y_4^{(3)} = G^{-1}\left[F_{4\|3}\left(Y_4^{(2)} \| Y_3^{(2)}\right)\right]$	$Y_5^{(3)} = G^{-1}\left[F_{5\|3}\left(Y_5^{(2)} \| Y_3^{(2)}\right)\right]$
				$Y_5^{(4)} = G^{-1}\left[F_{5\|4}\left(Y_5^{(3)} \| Y_4^{(3)}\right)\right]$

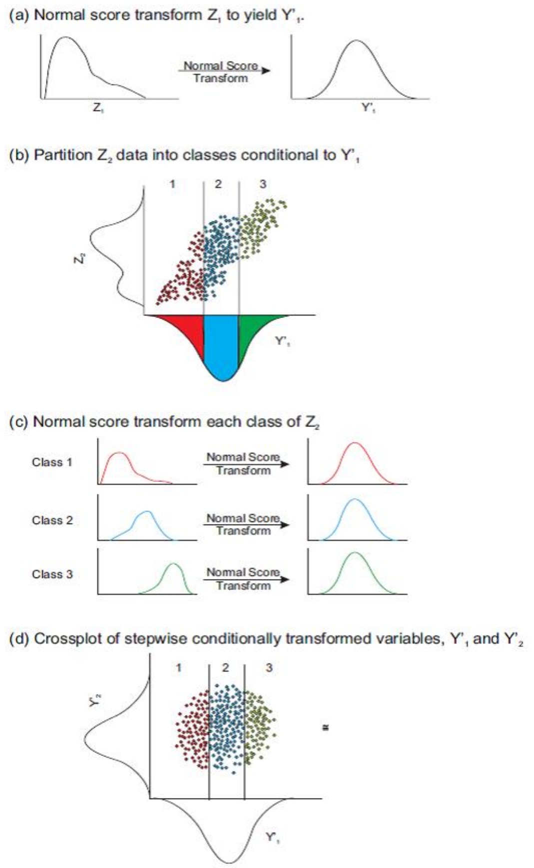
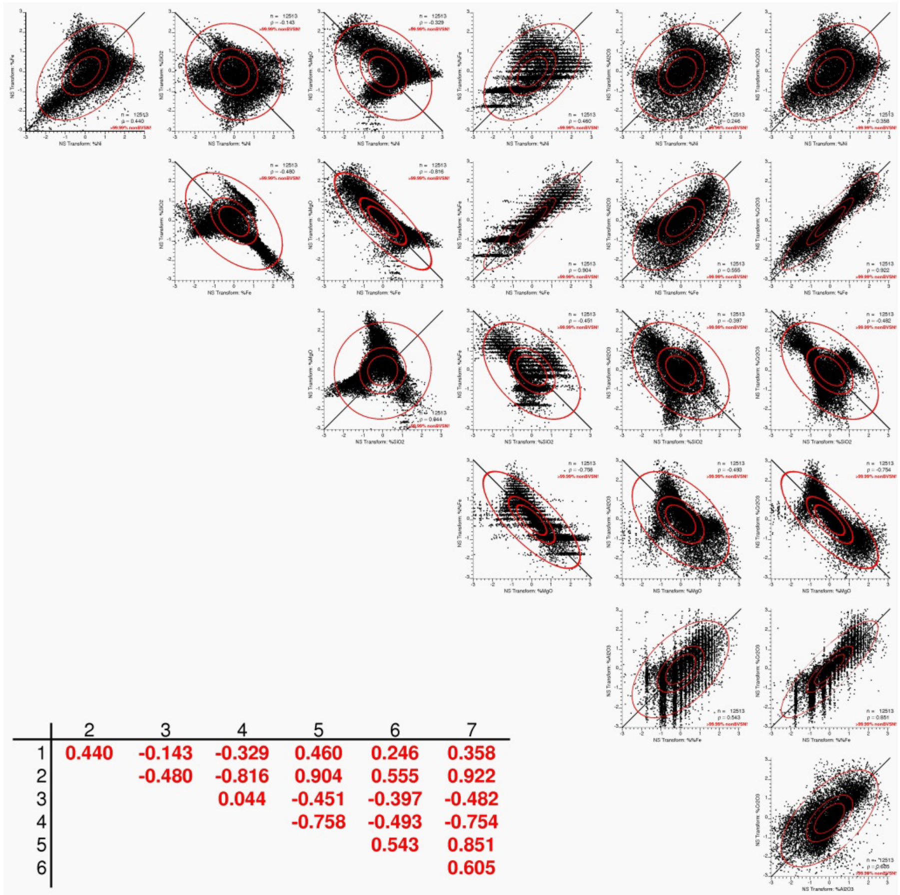
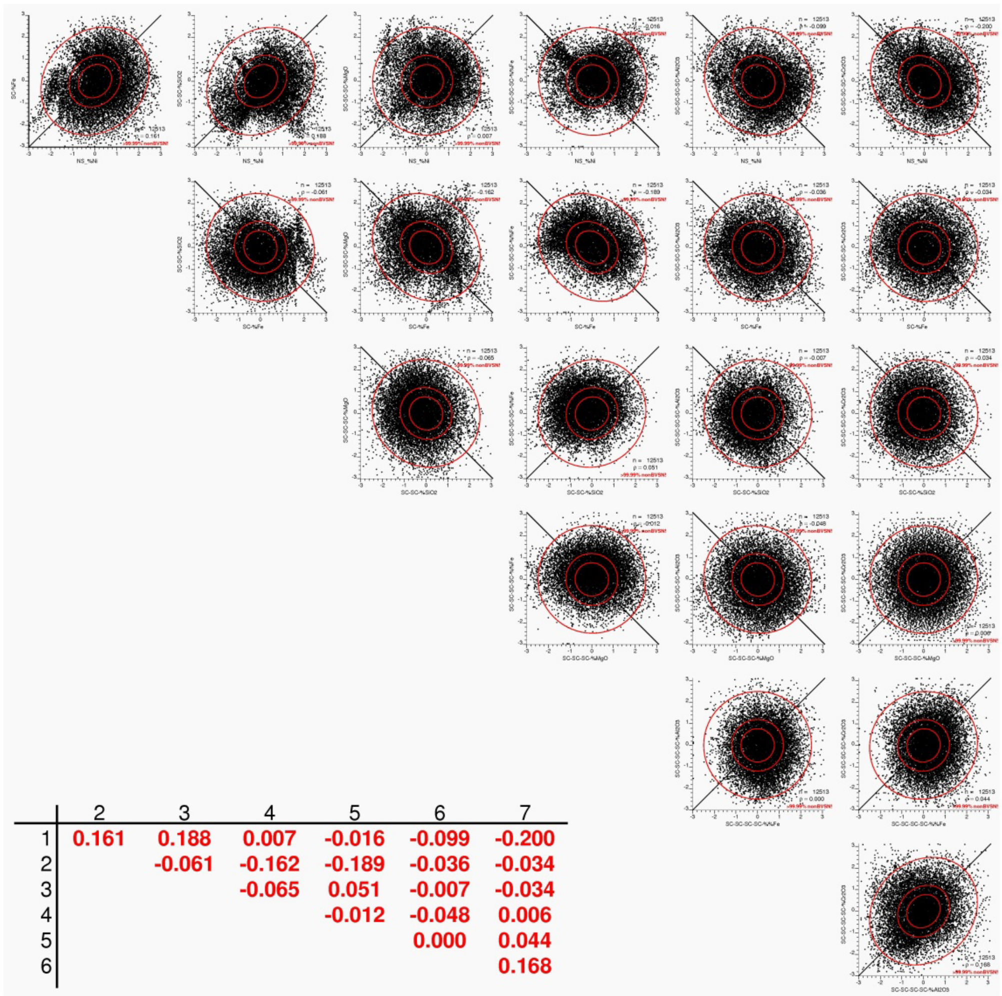


Figure 1. Schematic illustration of stepwise conditional transformation of two variables, Z1 and Z2, with Z1 as the primary variable: (a) normal score transform the primary variable, Z1; (b) partition Z2 data based on classes of Y ’1; (c) perform normal score transform of each class of Z2; (d) cross plot stepwise conditional variables to show bivariate Gaussian distribution with approximately zero correlation. (Leuangthong, 2003) .



	2	3	4	5	6	7
1	0.440	-0.143	-0.329	0.460	0.246	0.358
2		-0.480	-0.816	0.904	0.555	0.922
3			0.044	-0.451	-0.397	-0.482
4				-0.758	-0.493	-0.754
5					0.543	0.851
6						0.605

Figure 2. Scatterplots for 7 variables transferred to normal scores independently.



	2	3	4	5	6	7
1	0.161	0.188	0.007	-0.016	-0.099	-0.200
2		-0.061	-0.162	-0.189	-0.036	-0.034
3			-0.065	0.051	-0.007	-0.034
4				-0.012	-0.048	0.006
5					0.000	0.044
6						0.168

Figure 3. Scatterplots of transferred variables using the recursive step wise transformation.

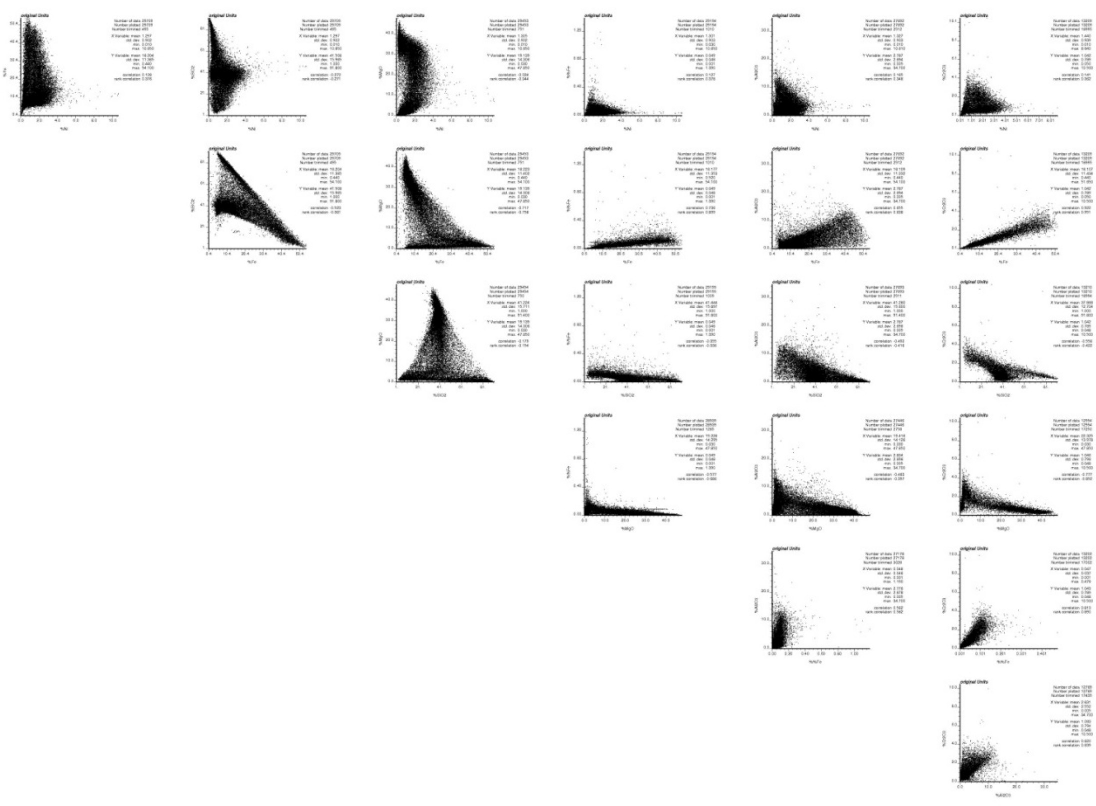


Figure 4. scatterplots of original variables

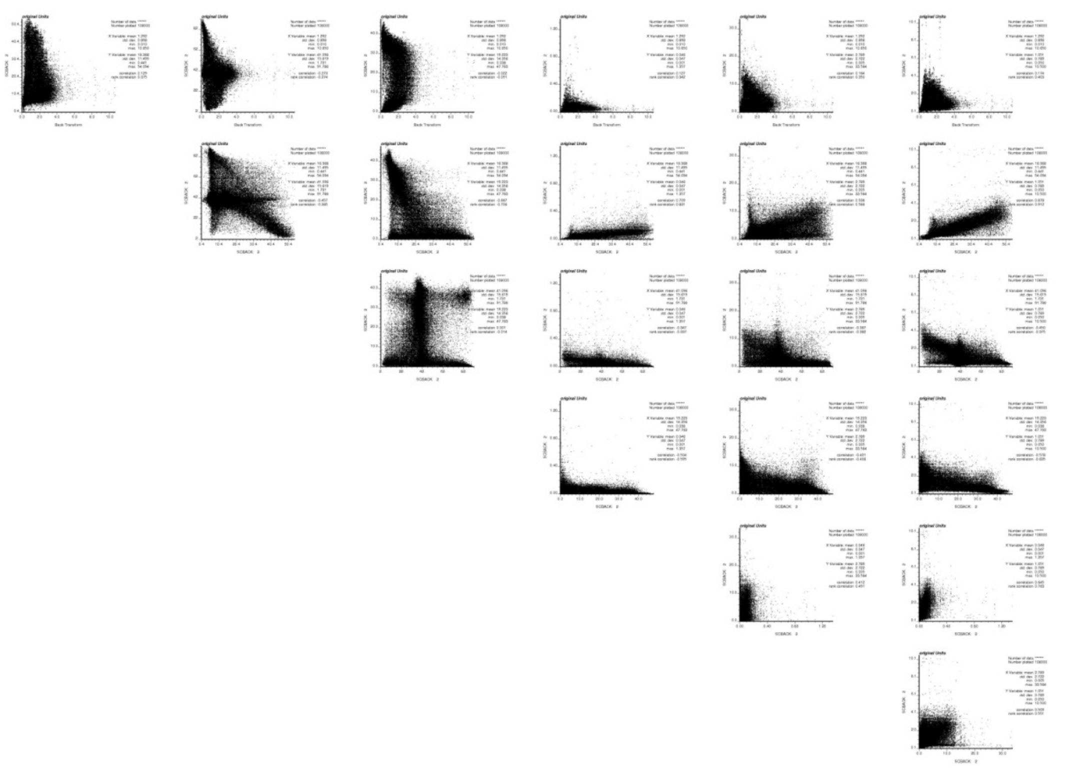


Figure 5. Scatterplots of simulated values.