Stepwise Conditional Transformation for Simplified Cosimulation of Reservoir Properties

Oy Leuangthong and Clayton V. Deutsch, University of Alberta
(oy@ualberta.ca, cdeutsch@civil.ualberta.ca)

Abstract

The cosimulation of multiple reservoir properties remains a significant outstanding problem in reservoir characterization. Gaussian methods are commonly used because of their simplicity; however, they assume that all distributions between model variables are multivariate Gaussian. Then, a model of coregionalization such as the linear model of coregionalization or a Markov-type "collocated" model must also be assumed. We present the Stepwise Conditional Transformation procedure, which transforms multiple variables to be multivariate Gaussian and independent. This greatly simplifies cosimulation; the variables are transformed, independent Gaussian simulation is performed, and back- transformation imparts the correct multivariate correlation to the simulated values.

*The stepwise-conditional transformation procedure was introduced by Rosenblatt in 1952. The method was not quickly adopted because of data and computer requirements. Modern reservoir characterization removes these limitations. We present the methodology, two programs (*stepcon *and* backstep*) for transformation and backtransformation, and several petroleum related examples. The simplified simulation or porosity and permeability is illustrated.*

Introduction

Petroleum reservoir characterization is increasingly being accomplished by employing geostatistical tools. Large scale lithofacies modeling is typically followed by smaller scale modeling of more uncertain features, such as porosity and permeability. Different simulation methods can be used to develop a suitable numerical reservoir model of heterogeneity and for uncertainty assessment. These include sequential indicator simulation, p-field simulation, simulated annealing, and the more commonly used sequential Gaussian simulation (SGSIM).

The use of Gaussian techniques, such as sequential Gaussian simulation, to simulate regionalized variables is dependent on the characteristics of a Gaussian variable. In the presence of two or more variables, the conventional procedure is to normal score transform each variable one at a time. This ensures that each variable is univariate normal; however, the multivariate distributions (two or more at a time) are not necessarily multivariate Gaussian. Yet, all Gaussian simulation algorithms assume this to be the case.

Once transformed, the model variables are cosimulated in order to preserve the correlation between them. Cosimulation requires a model of coregionalization. The variogram and cross variogram modeling must be modeled in a mathematically consistent manner, that is, through the linear model of coregionalization or a Markov model suitable for collocated
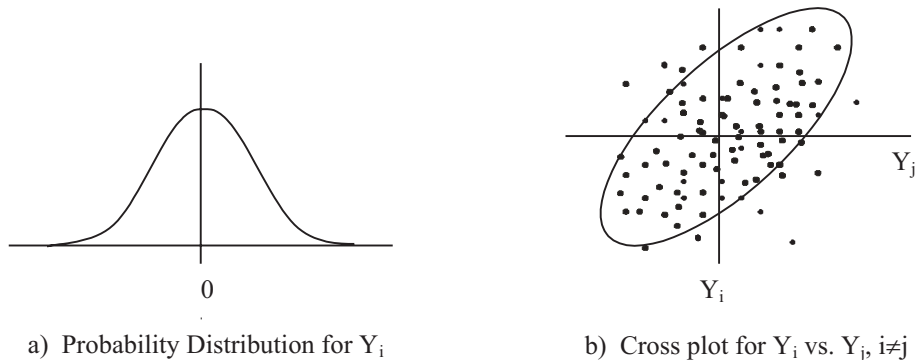
a) Probability Distribution for Y_i b) Cross plot for Y_i vs. Y_j, i≠j

Figure 1: Gaussian distributions: (a) univariate case, illustrating normal distribution of variable Y_i; (b) bivariate Gaussianity, showing correlation between two normal variables Y_i and Y_{i+1}.

cokriging. Alternatively, one might consider the application of a more sophisticated transformation that removes correlation between the model variables hence eliminating the need for cosimulation.

This paper presents the stepwise-conditional transformation to produce variables that conform to multiGaussian behaviour. Several petroleum- related examples are shown to compare the results to conventional normal transformation.

Multivariate Gaussianity

Gaussian-based geostatistical tools require the data variables to follow a Gaussian distribution. Earth science data, however, rarely are Gaussian distributed; therefore, the original N data variables ($Z_i, i = 1, \ldots N$) must be transformed to Gaussian variables ($Y_i, i = 1, \ldots N$). The merit of any transformation technique is dependent on how well the transformed variables, Y_i's follow the multivariate Gaussian distribution of the following orders:

1. First order: univariate case where each $Y_i, i = 1, \ldots N$ should have a Gaussian density function with zero mean and unit variance and characteristic bell shape, see Figure 1.a.

2. Second order: bivariate scenario where all cross plots between Y_i and Y_j should show a bivariate Gaussian probability distribution with elliptical probability contours along a line through the origin, see Figure 1.b.

3. Third and higher orders: A distribution of order k between k- variables in k-dimensions should follow a multivariate Gaussian distribution with probability contours following a hyperellipsoid in k-dimensional space.

With no loss of generality, we will consider the techniques based on how closely the distributions resemble the first and second order Gaussian distribution. The technique we propose, however, works for any number of variables N provided sufficient data is available.

The requirement for multivariate Gaussianity extends to multivariate spatial distributions, for example, the bivariate distribution of two variables separated by lag distance

2

vector $\mathbf{h}$: $Y_i(\mathbf{u})$ and $Y_j(\mathbf{u} + \mathbf{h})$. The transformation method proposed here does *not* ensure the Gaussianity of such multivariate spatial distributions. The stepwise-conditional transformation procedure ensures that "collocated" variables are strictly Gaussian and independent.

Normal Transformation

The normal or Gaussian distribution[1] is characterized by its mean m and variance σ^2. The probability density function (pdf) that produces its characteristic bell shape is:

$$g(y) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left[-\frac{1}{2}\left(\frac{z - m}{\sigma}^2\right)\right] \qquad (1)$$

A standard normal distribution refers to a mean of zero and a unit variance. The corresponding pdf is:

$$g(y) = \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{z^2}{2}\right) \qquad (2)$$

The normal distribution is the limit distribution for the Central Limit Theorem. The sum of two or more normal distributions is also a normal distribution. In fact, the sum of independent random variables following any distribution tends toward a normal distribution. This result implies great simplicity for simulation and is the main reason why Gaussian approaches are commonly used. However, most earth sciences variables are not normally distributed. In order to apply the Gaussian approach, the variables must first be transformed to normal space.

   The three basic steps in the common normal score transformation process (also known as the "graphical" or "quantile" transformation) are described below and illustrated in Figure 2:

1. The original data are ranked in ascending order.

2. The sample cumulative distribution function of the original data variable, Z, must be calculated. The transformation from one distribution to another is accomplished based on the cumulative probability associated to each data value.

3. For each sample data, $z(\mathbf{u}_j)$, the corresponding cumulative probability is identified. Once determined, the normal score value, $y(\mathbf{u}_j)$, corresponding to each probability is found:

$$y(\mathbf{u}_j) = G^{-1}[F(z(\mathbf{u}_j))] = G^{-1}(p_j) \qquad (3)$$

   where $y(\mathbf{u}_j)$ is the normal transform of $z(\mathbf{u}_j)$, $G^{-1}(\cdot)$ is the inverse of the cumulative Gaussian distribution, and $F(z(\mathbf{u}_j))$ is the data-derived cumulative distribution function value of data $z(\mathbf{u}_j)$.

---

[1]The words "normal" and "Gaussian" are used interchangeably; they refer to exactly the same probability distribution.
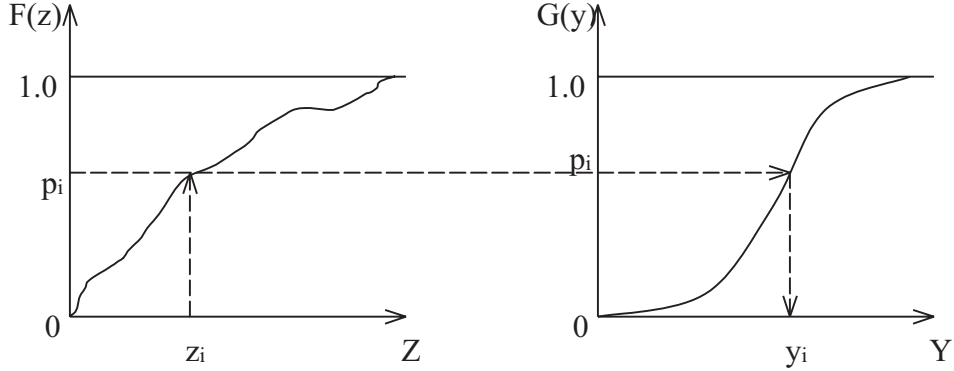
Figure 2: Transforming original data $(z(\mathbf{u}_i))$ to a standard normal value $(y(\mathbf{u}_i))$.

Alternative transformation procedures such as fitting Hermite polynomials could be considered. This normal score transformation procedure is simple and effective.

Once transformation is complete, the transformed variables $(y(\mathbf{u}_j))$ are simulated. The results of simulation must then be back transformed. Figure 3 illustrates the basic steps in simulating normal score data.

Stepwise Conditional Transformation

This technique, first introduced by Rosenblatt in 1952, bears resemblance to the normal score transformation technique. In the univariate case, the stepwise-conditional technique is identical to the normal score transform, that is, the variable is transformed using the Gaussian distribution.

In a bivariate case, the normal transformation of the second variable is conditional to the probability class of the first variable. Correspondingly, for k-variate problems, the kth variable is conditionally transformed based on the $(k-1)$ first variables.

$$\begin{aligned}
Y_1 &= Prob(Z_1 \leq z_1) \\
Y_2 &= Prob(Z_2 \leq z_2 \mid Z_1 = z_1) \\
Y_3 &= Prob(Z_3 \leq z_3 \mid Z_2 = z_2, Z_1 = z_1)
\end{aligned}$$

Figure 4 shows the steps to accomplish this conditional transformation. Once the data are binned based on their conditional probabilities, each group of data is normally transformed. Simulation is then performed on the normal score values of the Y_i's and back transformation is performed. For example, Z_1 can be determined from Y_1 with the correct conditional distribution; from Z_1 and the simulated value of Y_2, Z_2 can be calculated; and so forth.

As mentioned above, the multivariate spatial relationship of the original model variables is not transformed (i.e. no modification of bivariate spatial distributions $Y(\mathbf{u})$ and $Y(\mathbf{u}+\mathbf{h})$, or trivariate distributions $Y(\mathbf{u})$, $Y(\mathbf{u}+\mathbf{h}_1)$ and $Y(\mathbf{u}+\mathbf{h}_2)$, ...).

There is no correlation between the transformed variables. For example, each class of Y_2 data is independently transformed to a normal distribution removing any correlation
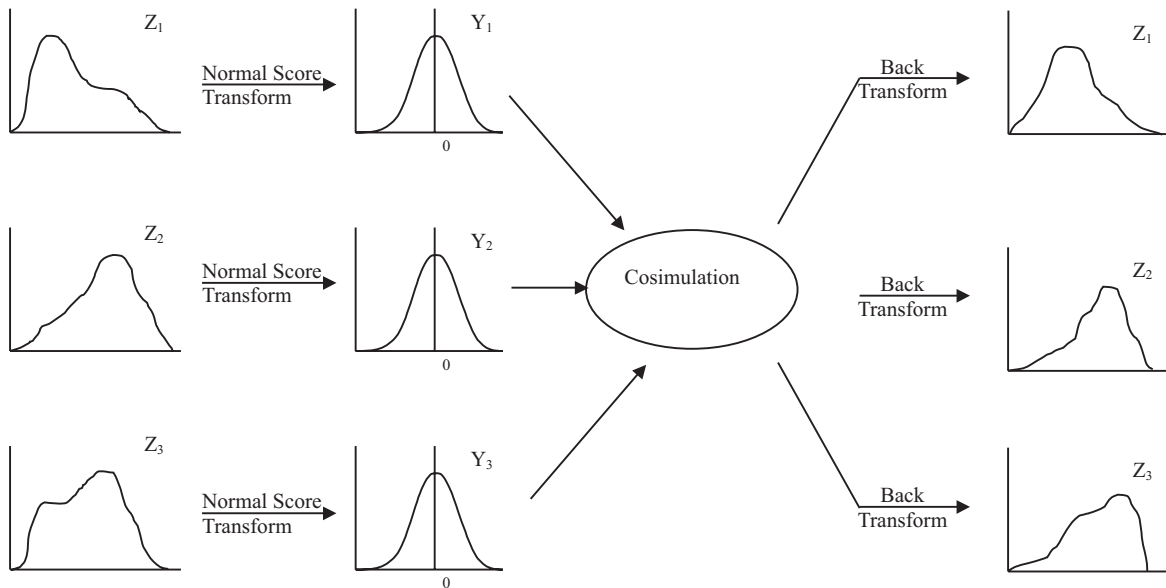
4

Figure 3: Schematic of application of a normal transformation and simulation of three variables.

between Y_2 and Y_1. Consequently, the simulation of a multivariate problem does not require cosimulation due to the independence of the transformed variables. This is the primary motivation for transforming multiple variables in a stepwise conditional fashion.
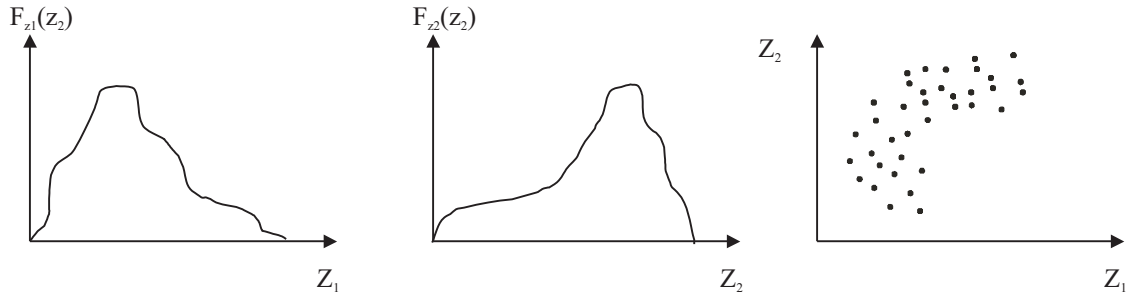
Limitations of Stepwise Conditional Transformation

The variables $Y_k, k = 2, \ldots, N$ are combinations of multiple "real" variables. These values cannot be backtransformed by an inverse Gaussian transformation. Each variable must be back transformed in the same order in which they were originally transformed.

The main limitation of the stepwise conditional transformation lies in the need for sufficient data. In order to classify data and transform each class, there must be sufficient data to identify a conditional distribution. Sparse data leads to eratic and nonrepresentative conditional distributions. There is no general rule; however, we estimate the need for 10^N to 20^N data, where N is the number of variables.

The data are partitioned into classes corresponding to equal probability intervals. Depending on the dataset, the conditional transformation of secondary variables may result in artifacts in the transformed data. For example, two identical secondary data values should have the same transformed values (Y_2). However, if the corresponding primary variables belong to different probability intervals, then transformation may produce significant differences in the secondary variable Y_2.
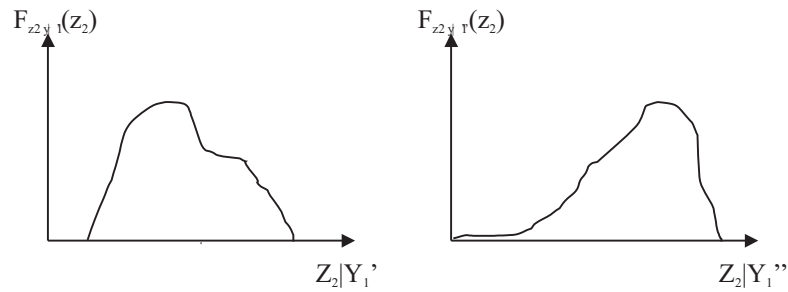
5

a) Raw Data

b) Normal Transform of Z_1 to Y_1
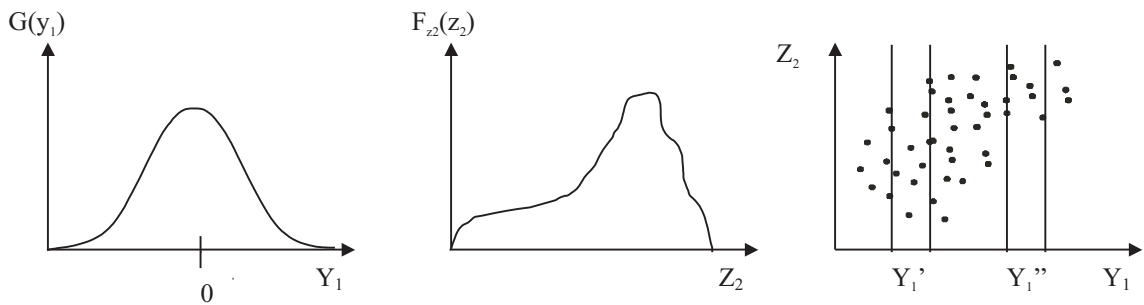
c) Determine Conditional Distributions $F_{z2|y_1'}(z_2)$ and $F_{z2|y_1''}(z_2)$

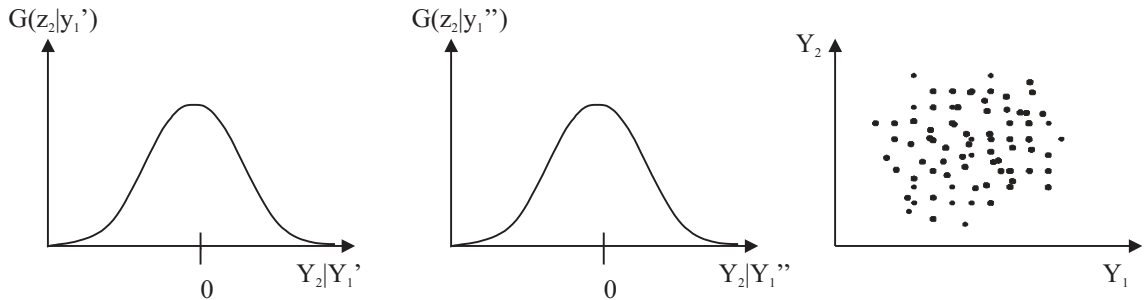d) Normal Transform $Z_2|Y_1$, using corresponding conditional distribution

Figure 4: Processes involved in Stepwise Conditional Transformation, (a) Distribution of data variables and the scatterplot of Z_2 and Z_1; (b) Transform variable Z_1 into normal scored data, Y_1, and categorize Z_2 data conditional to probability classes of Y_1 ; (c) Determine conditional distributions for each Z_2 class conditional to Y_1 data; and (d) Normally transform each $Z_2|Y_1$ class independent from each other, and the scatterplot between one $Z_2|Y_i$ class and $Z_2|Y_{i+1}$ will show no correlation between the normally transformed classes.

Application

Programs `stepcon` and `backstep` were developed to perform the stepwise conditional transformation and the back transformation, respectively. Several petroleum related examples are shown below in Figures 5 and 6. For each sample data set, cross plots are shown for the original data, conventional normal score transformation, and stepwise conditional transformation. In all cases, porosity is the primary variable of transformation, and the logarithm of permeability is conditionally transformed.

The "two-well" data used in the GSLIB training course (see CCG Report 1) is considered as a specific example to illustrate (1) the conventional cosimulation of two model variables (see Figure 7), and (2) simulation using the stepwise conditionally transformed variables (see Figure 8). In both cases, the cross plots of the simulated results are shown after back transformation has been performed.

The original data shows a correlation coefficient of 0.586. The results of the stepwise conditional transformation and subsequent independent simulation of the porosity and the log(permeability) exhibit a correlation of 0.531; while, the results of applying a typical normal score transform and performing cosimulation show a correlation coefficient of 0.572. A visual inspection of the cross plot shows the conditional transformation method reproduces the correlation and non- Gaussian features between the model variables.

Figure 9 shows the distributions for both porosity and log(permeability) in each of the three scenarios: original data, a cosimulated realization, and an independently simulated realization after conditional transformation.

Although the cross plot for the cosimulated results reproduces the correlation between the model variables, the cross plot of the normal score values of the cosimulated realization does not reproduce the same correlation between the normal score values prior to cosimulation. This is illustrated in Figure 10.

The reproduction of the cross plot and the individual variable distributions clearly indicate that the application of the stepwise conditional transformation and corresponding back transform produces results comparable (or superior) to conventional cosimulation, without the added complexities associated with models of coregionalization.

Conclusion

The application of Gaussian simulation techniques require that model variables be multivariate Gaussian. The conventional method involves independently transforming each variable into normal score values, modeling the variograms and the cross variograms, and performing sequential cosimulation.
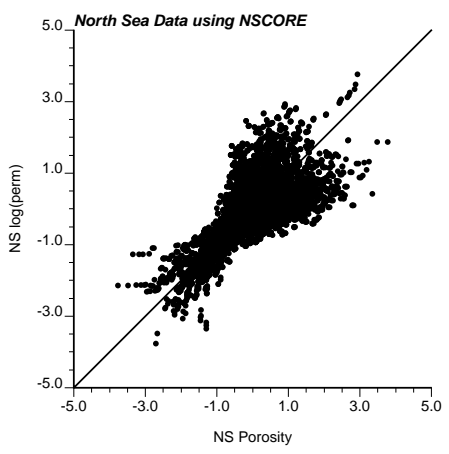
By employing the stepwise conditional transformation,via `stepcon`, prior to variography, cosimulation can be avoided. This tranformation removes the correlation between the model variables producing independent model variables. Each variable can be simulated independently, thereby simplifying the simulation process. The correlation between the variables is injected back during back transformation, `backstep`.
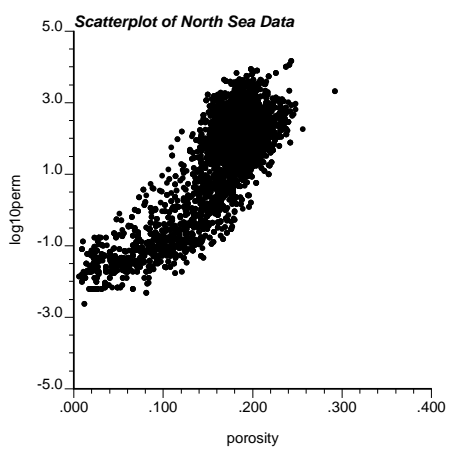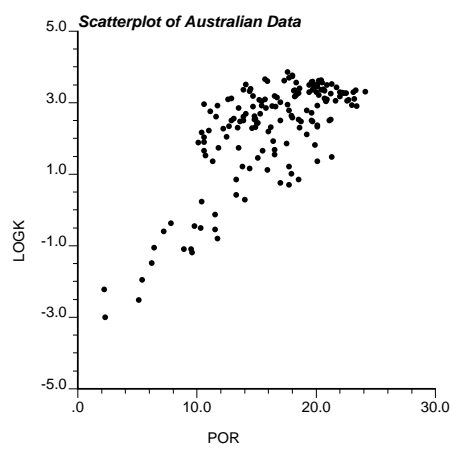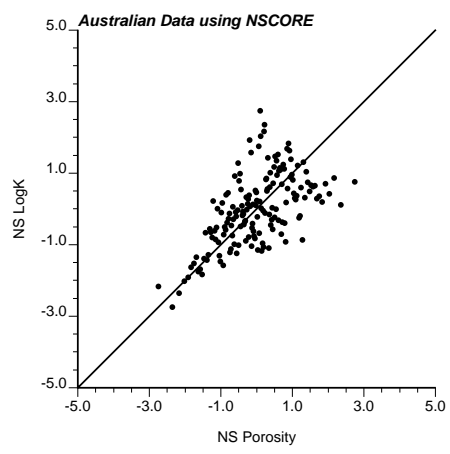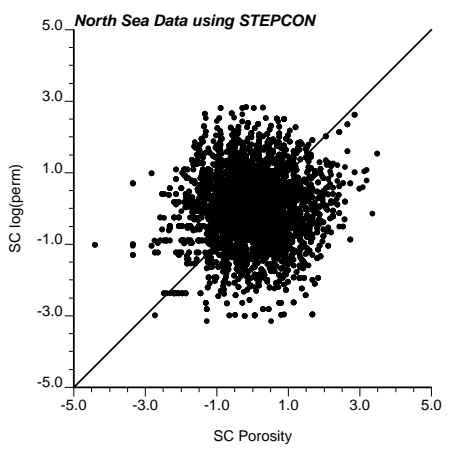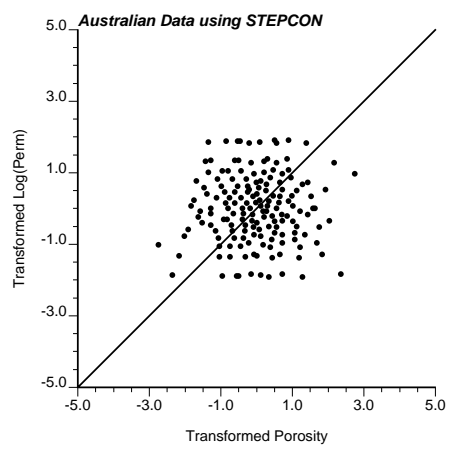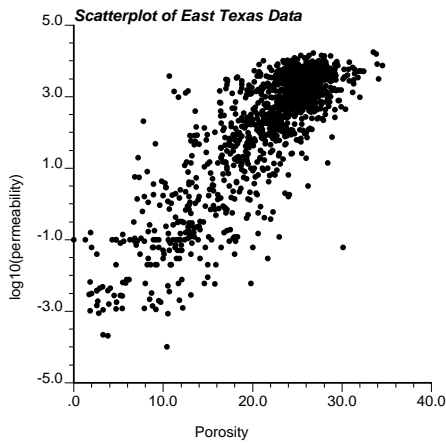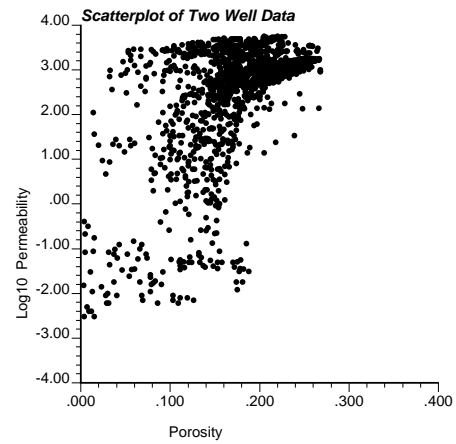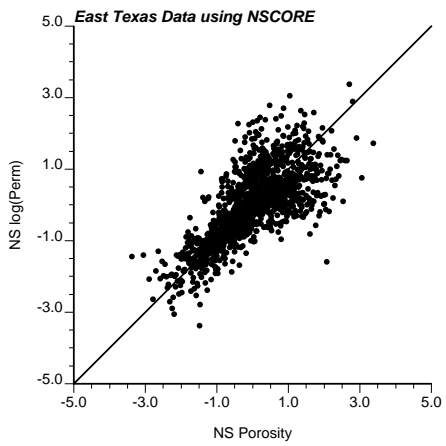
Figure 5: Comparative illustration of cross plots of the original data, normally transformed data and the stepwise conditionally transformed data for porosity and log(permeability) for North Sea data (left side) and an Australian data set (right side).
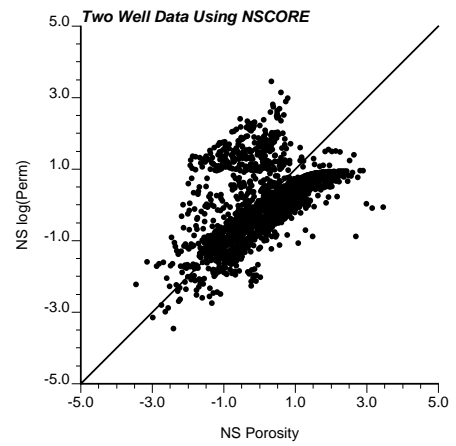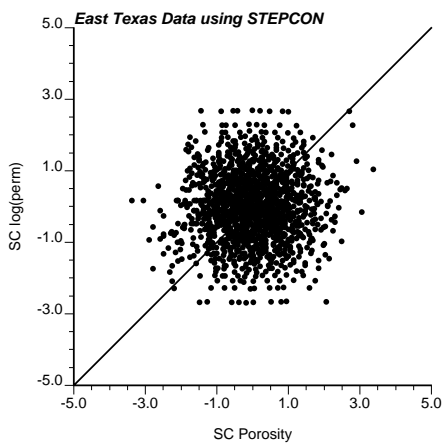
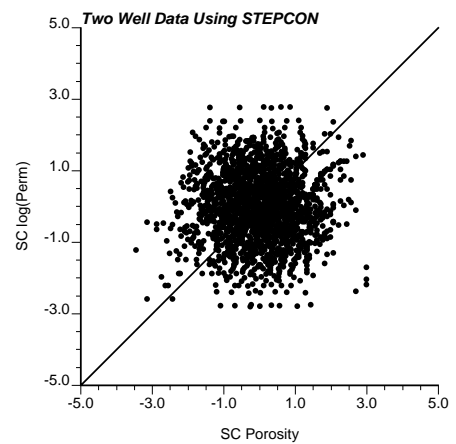Figure 6: Illustration of cross plots of the original data, normally transformed data and the stepwise conditionally transformed data for porosity and log(permeability) for East Texas core data (left side) and the Two Well data set (right side).

Cosimulated Porosity (NSCORE)

Cosimulated log(Permeability) (NSCORE)

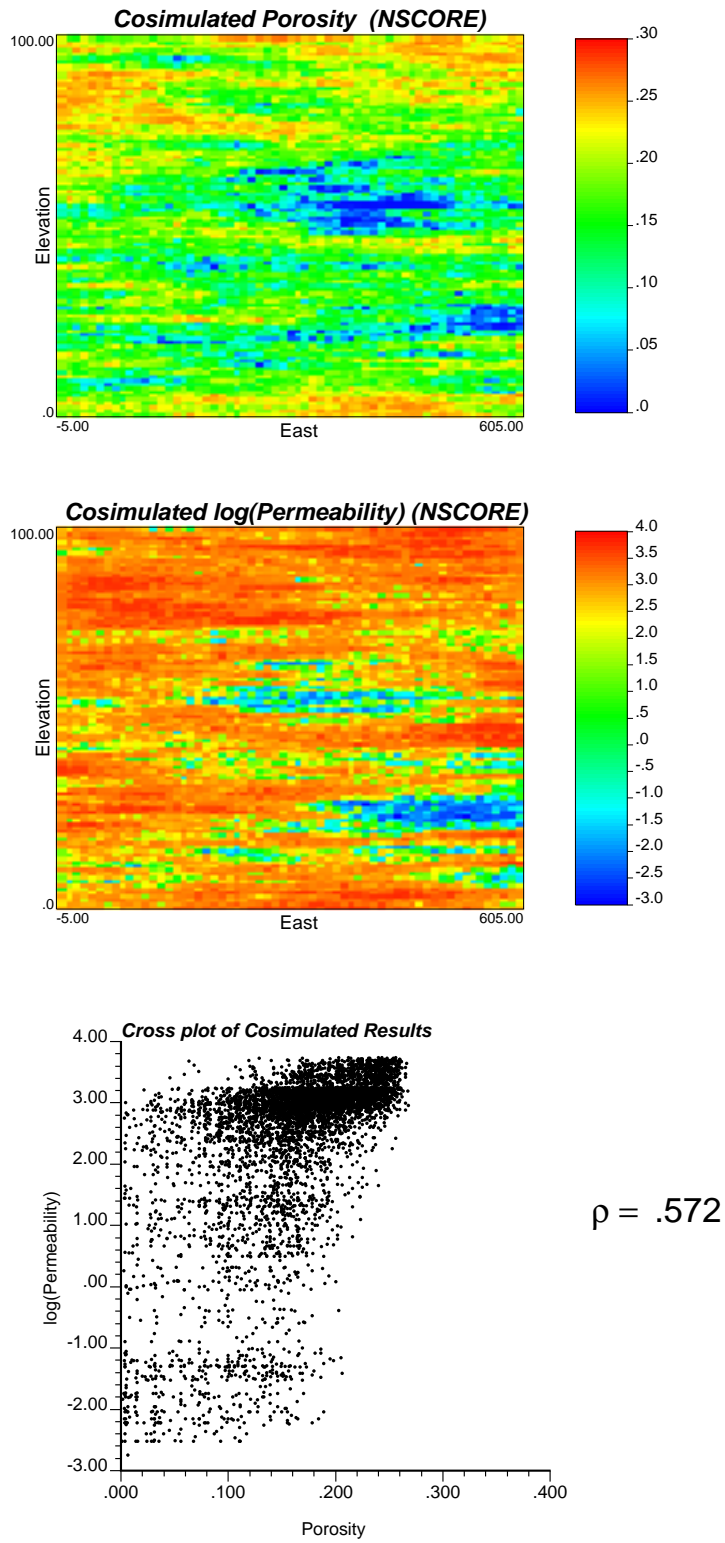**Cross plot of Cosimulated Results**

$\rho = .572$

Figure 7: One realization of cosimulation is shown for the porosity variable and the log(permeability) variable, after application of Normal Transformation. The bottom figure shows a cross plot the cosimulated results.
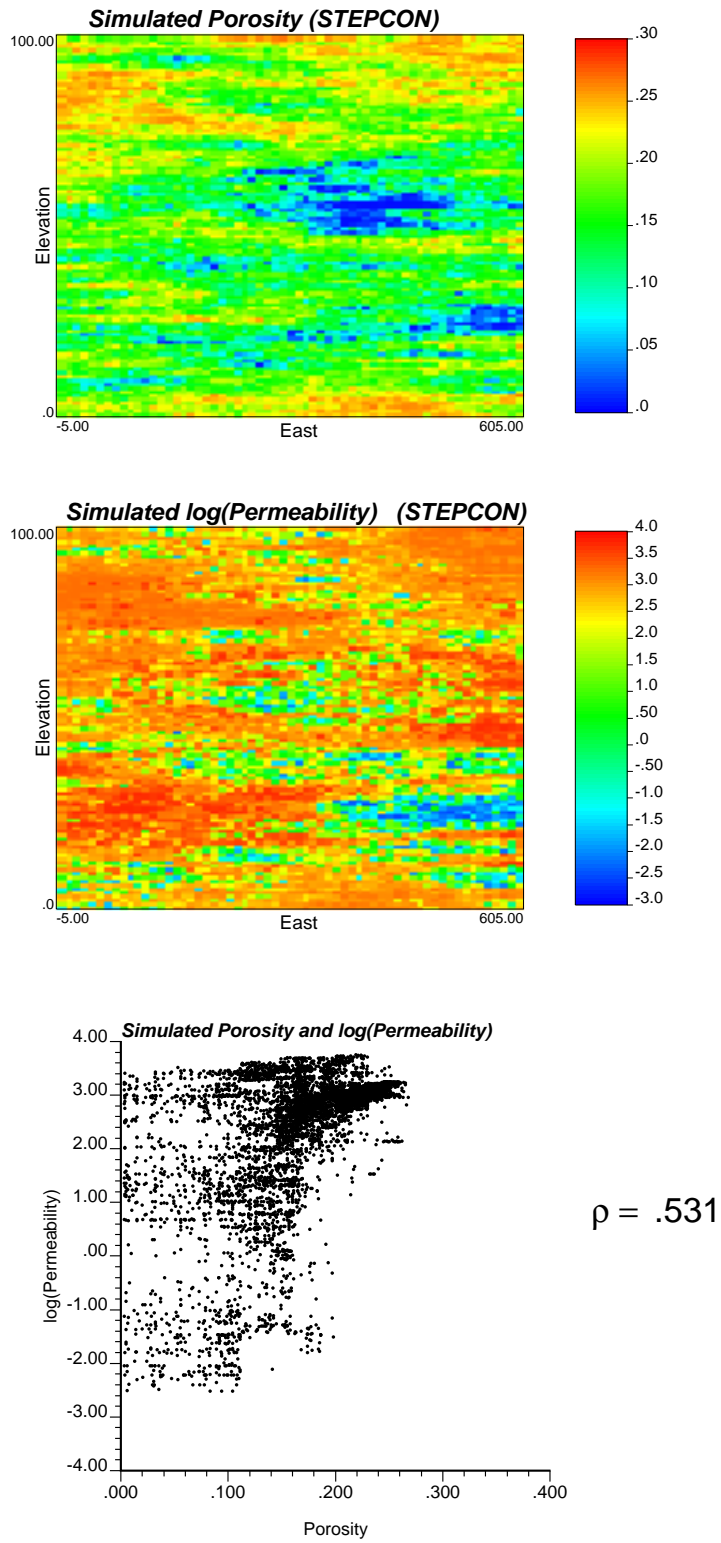
10

Figure 8: One realization of `sgsim` is shown for the porosity variable and the conditionally transformed log(permeability) variable, after the application of the Stepwise Conditional Transformation . The bottom figure shows a cross plot of the simulated results.
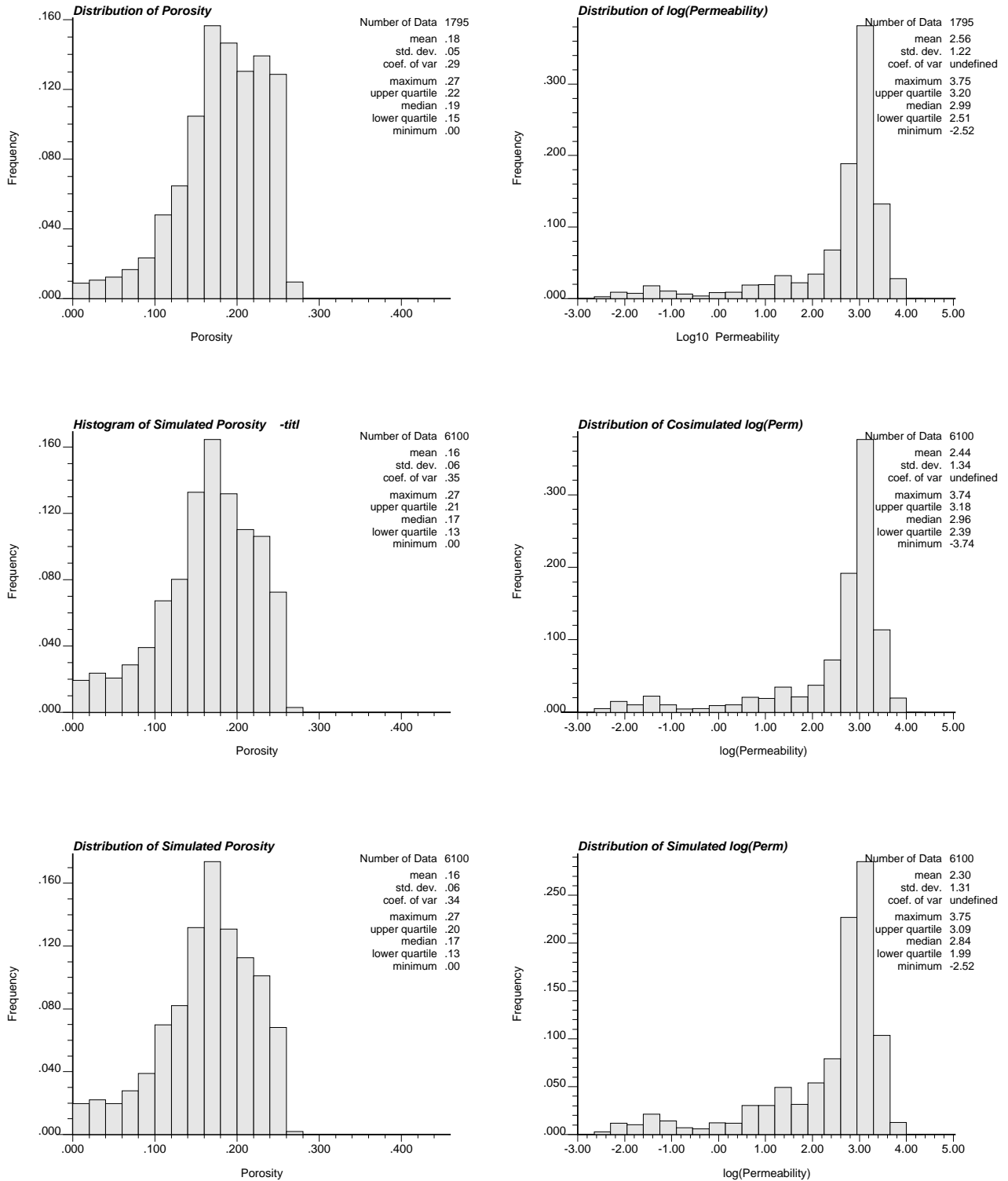
Figure 9: Comparison of Distribution of Model Variables. The top two histograms give the distribution of the original data; the corresponding distribution of porosity and log(permeability) for the cosimulated realization follows; and the last two histograms give the distribution of the model variables for the independently simulated realizations.
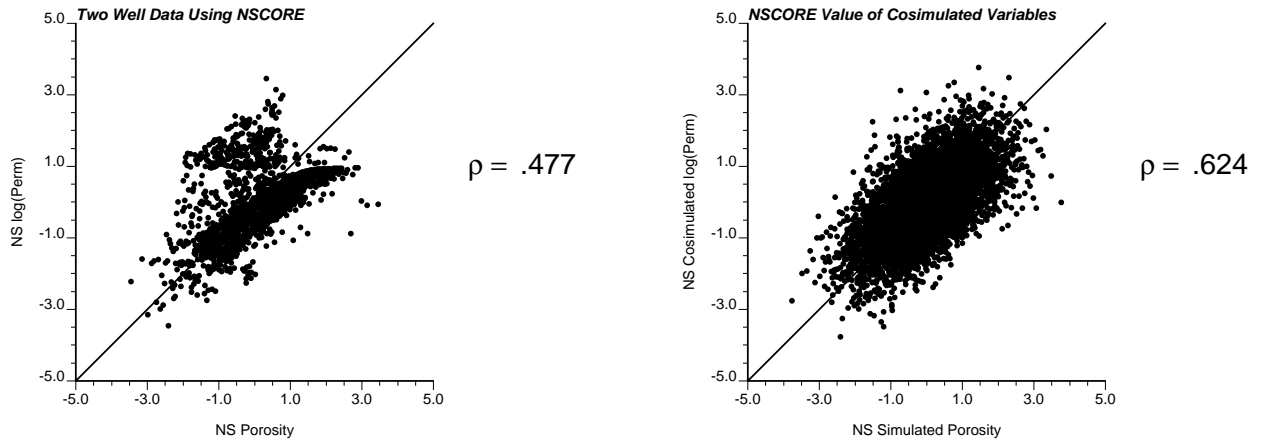
Figure 10: Comparison of cross plots generated by normally transforming the model variables and the normal score values of the cosimulated results.

References

[1] T. Anderson. *An Introduction to Multivariate Statistical Analysis.* John Wiley & Sons, New York, 1958.

[2] C. V. Deutsch and A. G. Journel. *GSLIB: Geostatistical Software Library and User's Guide.* Oxford University Press, New York, 1992.

[3] P. Goovaerts. *Geostatistics for Natural Resources Evaluation.* Oxford University Press, New York, 1997.

[4] E. H. Isaaks. *The Application of Monte Carlo Methods to the Analysis of Spatially Correlated Data.* PhD thesis, Stanford University, Stanford, CA, 1990.

[5] G. R. Luster. *Raw Materials for Portland Cement: Applications of Conditional Simulation of Coregionalization.* PhD thesis, Stanford University, Stanford, CA, 1985.

[6] M. Rosenblatt. Remarks on a multivariate transformation. *Annals of Mathematical Statistics*, 23(3):470–472, 1952.

Appendix

Example parameter files `stepcon.par` and `backstep.par` are shown in Figures 11 and 12. The parameters are self explanatory.

```
                    Parameters for STEPCON
                    *********************

START OF PARAMETERS:
../data/cluster.dat             -file with data
3                               -number of variables to transform
13 14 15                        -columns for variable transformation
-1.0e21  1.0e21                 -trimming limits
10                              -number of classes
stepcon.out                     -file for output
stepcon.trn                     -file for output transformation table
```

Figure 11: Parameters for `stepcon`.

```
                    Parameters for BACKSTEP
                    *********************

START OF PARAMETERS:
2                               -number of variables
file1.dat                       -data file number 1
file2.dat                       -data file number 2
-1.0e21  1.0e21                 -trimming limits
10                              -number of classes
stepcon.trn                     -file with transformation table
backstep.out                    -file for output
```

Figure 12: Parameters for `stepcon`.