SPE 63069

Multivariate Stepwise Transformation for Stochastic Reservoir Modeling

O. Leuangthong[1], C. V. Deutsch[1], A. Haas[2], and A. Shtuka[3]

[1]University of Alberta, Edmonton, Alberta, Canada T6G 2G7
[2]Elf Aquitaine, CSTJF, Avenue Larribau, 64018 Pau Cedex, France
[3]National School of Geology (ENSG) Nancy, France

Copyright 2000, Society of Petroleum Engineers Inc.

This paper was prepared for presentation at the 2000 SPE Annual Technical Conference and Exhibition held in Dallas, Texas, 1–4 October 2000.

This paper was selected for presentation by an SPE Program Committee following review of information contained in an abstract submitted by the author(s). Contents of the paper, as presented, have not been reviewed by the Society of Petroleum Engineers and are subject to correction by the author(s). The material, as presented, does not necessarily reflect any position of the Society of Petroleum Engineers, its officers, or members. Papers presented at SPE meetings are subject to publication review by Editorial Committees of the Society of Petroleum Engineers. Electronic reproduction, distribution, or storage of any part of this paper for commercial purposes without the written consent of the Society of Petroleum Engineers is prohibited. Permission to reproduce in print is restricted to an abstract of not more than 300 words; illustrations may not be copied. The abstract must contain conspicuous acknowledgment of where and by whom the paper was presented. Write Librarian, SPE, P.O. Box 833836, Richardson, TX 75083-3836, U.S.A., fax 01-972-952-9435.

Abstract

Reservoir characterization requires simultaneous modeling of multiple correlated variables including seismic attributes, porosity, permeability, and water saturation. The relationship between these variables is often non-linear with complex variability patterns. Classical normal scores transformation and least squares regression fails to capture these realistic features. A multivariate stepwise transformation procedure is proposed whereby the original data variables are transformed to Gaussian variables with multivariate Gaussian properties. This procedure leads to much improved reservoir models with conventional geostatistical procedures. These models, in turn, lead to improved reservoir production forecasting and development planning.

This normal or Gaussian transformation is essential prior to geostatistical modeling with Gaussian techniques. The procedure proceeds with a stepwise transformation using conditional probability distributions. Once all variables have been transformed the collocated Gaussian variables are strictly multivariate Gaussian. Conventional transformation procedures only ensure that the univariate distributions are Gaussian; cross plots (bivariate) and higher order distributions are not Gaussian.

This transformation approach is important and relevant to modern reservoir characterization due to the widespread use of Gaussian techniques. Aside from object-based facies modeling, more than 90% of geostatistical applications use Gaussian techniques.

Introduction

Petroleum reservoir characterization increasingly uses geostatistical tools. Large-scale lithofacies modeling is typically followed by smaller scale modeling of more heterogeneous features, such as porosity and permeability. Different geostatistical 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[1,2,3].

The use of Gaussian techniques, such as sequential Gaussian simulation, depends on the mathematical characteristics of a Gaussian variable[4,5,6]. 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 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.

The stepwise conditional tranformation has three important advantages that make it a practical technique that may catch on in practice. First, it leads to transformed variables that are exactly multivariate Gaussian and not just univariate Gaussian; thus, the back transformed results of Gaussian simulation better mimic geological features. Second, the stepwise transformed variables have no linear or non-linear correlation, which greatly simplifies simulation. The relationship between the variables is captured in the transformation and back-transformation. Third, it is very simple to apply and does not introduce any artifacts in resulting models.

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, $i=1…N$ must be transformed to Gaussian variables Y_i, $i=1…N$. The merit of any transformation technique is dependent on how well the transformed variables, the Y_i's, follow the multivariate Gaussian distribution of the following orders.

1. *First order*: univariate case where each Y_i, $i=1…N$ should have a Gaussian density function with zero mean and unit variance and characteristic bell shape, see **Fig. 1**. The probability density must follow the Gaussian distribution:

$$f(y) = \frac{1}{\sqrt{2\pi}} e^{\frac{-y^2}{2}} \qquad (1)$$

2. *Second order*: bivariate case 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 bottom of **Fig. 1**. The bivariate probability must follow the Gaussian distribution:

$$f(y_1, y_2) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}\left(y_1{}^2 - 2\rho y_1 y_2 + y_2{}^2\right)\right]$$

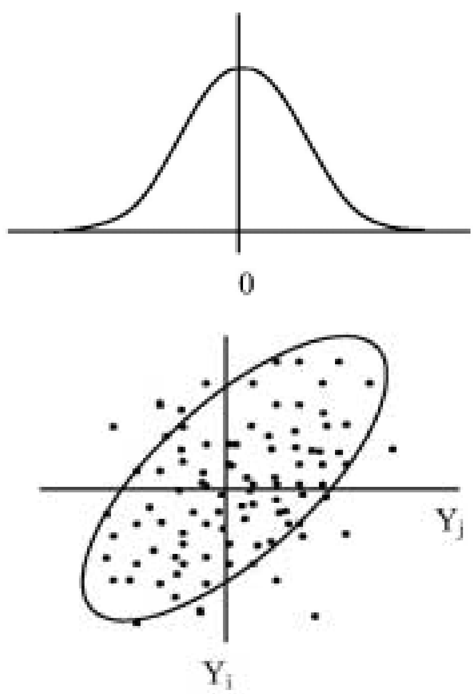

Fig. 1. Schemetic illustration of a univariate and bivariate Gaussian distributions.

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

With no loss of generality, we will consider first and second order Gaussianity. The technique we propose, however, works for any number of variables 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 vector **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 Gaussian and independent.

Normal Score Transform: A Recall

A standard normal distribution has a mean of zero and a unit variance, see **Eq. 1** above. 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)[9,10] are described below and illustrated in **Fig 2**.

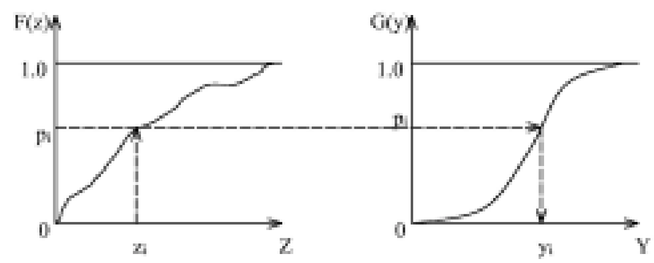

Fig. 2. Illustration of normal score transform. The cumulative probability p_i of the original data z_i is determined, then the matching normal value y_i is calculated.

1. The original data are ranked in ascending order.
2. The sample cumulative distribution function of the original data variable, Z, is calculated.
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:

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

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

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 are used for Gaussian simulation. The results of simulation must then be back transformed. **Fig. 3** gives a sketch of the procedure: the original data variables are transformed to normal distributions (starting from the right), the variables are cosimulated to honor the relationship between the variables, then, the simulated values and transformed data are back transformed to the correct units.

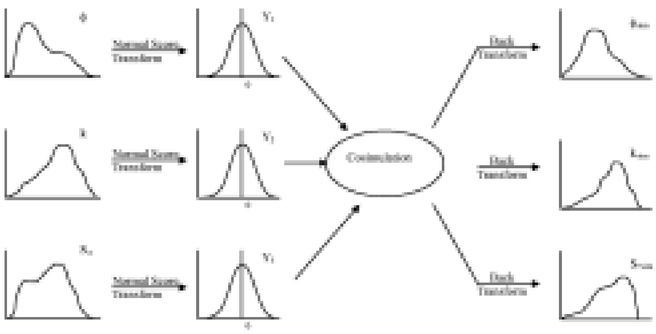

Fig. 3. Procedure for conventional Gaussian simulation for reservoir modeling: the original data are normal score transformed to normal variables, these variables are cosimulated accounting for correlation, and then back transformed to the correct units.

Stepwise Conditional Transformation:

Rosenblatt[7,8] first introduced this technique in 1952. It bears resemblance to the normal score transformation technique. In the univariate case, the stepwise-conditional technique is identical to the normal score transform.

In the bivariate case, the normal transformation of the second variable is made conditional to the probability class of the first variable. Correspondingly, for *k*-variate problems, the k^{th} variable is conditionally transformed based on the *k-1* first variables. For three variables:

$$\begin{aligned} Y_1 &= \Pr(Z_1 \le z_1) \\ Y_2 &= \Pr(Z_2 \le z_2 | Z_1 = z_1) \\ Y_3 &= \Pr(Z_3 \le z_3 | Z_2 = z_2, Z_1 = z_1) \end{aligned} \quad (2)$$

Fig. 4 shows the steps to accomplish this conditional transformation. Once the data are binned based on their conditional probabilities, each group of data is normal score transformed. Each variable follows a univariate normal distribution. The bivariate distribution between all pairs of variables follows a bivariate normal distribution with zero correlation. This is a consequence of each conditional distribution being normal with a mean of zero and a variance of 1, that is, each conditional distribution is a standard normal distibution. All multivariate distributions are independent standard Gaussian distributions.

The independence of the transformed variables has two very important consequences. First, there is no need for cokriging or cosimulation. All variables *have* no correlation and *should* be simulated independently. Second, the back-transformed variables have the correct behavior, that is, any non-linear correlation or heteroscedastic (unequal variance of the conditional distributions in real coordinates) features are reproduced in the back-transformed result.

Simulation is then performed on the transformed Y_i's and the final back transformation introduces the right correlation. The steps for porosity-permeability simulation:

1. Transform porosity by normal score transformation.
2. Transform permeability according to the conditional distributions of permeability given porosity.
3. Simulate the porosity normal score values.
4. Independently simulate the permeability stepwise transformed data.
5. Back transform the simulated "permeability" values using the correct conditional distributions.

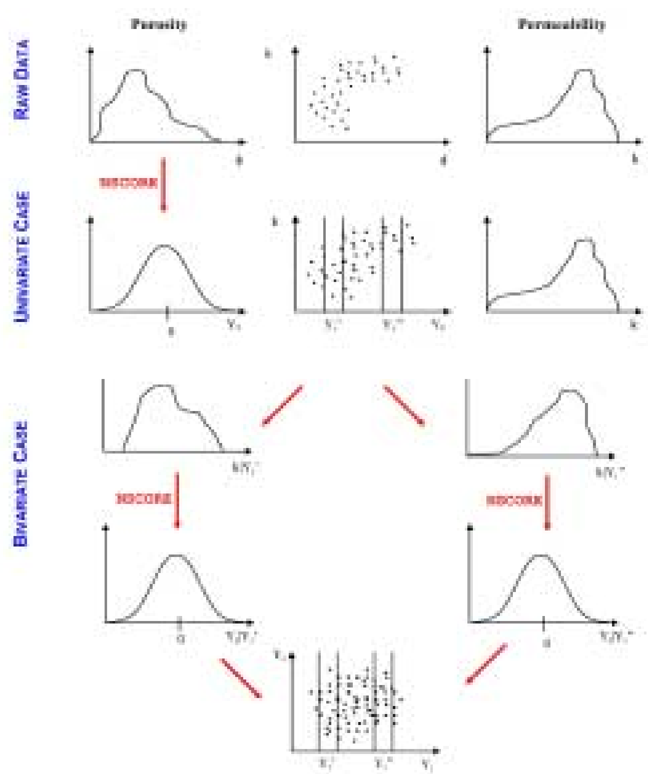

Fig. 4. Illustration of stepwise conditional transformation for two-variables (porosity and permeability). The top row of the Figure shows the porosity histogram, porosity-permeability cross plot, and permeability histogram. The second row from the top labeled Univariate Case shows the normal score transform of porosity; the permeability is left unchanged, but the central cross plot changes because the porosity has been transformed to normal space. The third and fourth rows from the top labeled Bivariate Case shows how the permeability values are transformed according to conditional distributions given the porosity values. The left side shows the transformation of permeability values belonging to a "low" class of porosity; the right side shows transformation of permeability values belonging to a "high" class of porsosity.

6. Back transform the simulated porosity values.
7. Check the results to ensure reproduction of the histograms, variograms, and cross plot relations.

As mentioned above, the multivariate spatial relationship of the original model variables is not transformed. That is, there is no modification to bivariate spatial distributions $Y(\mathbf{u})$ and $Y(\mathbf{u}+\mathbf{h})$ or trivariate spatial 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 since each class of Y_2 data is independently transformed to a normal distribution removing any correlation 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 Transfomation

There are few limitions of this method; however, there are three considerations to be mentioned.

The transformed 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 reverse order using the correct conditional transform.

In presence of many variables (N>3) 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 $10^N – 20^N$ data where N is the number of variables. As we develop later, a smoothing or modeling of the conditional distributions may be adequate.

The data are partitioned into classes. There must be sufficient data to determine reliable classes; otherwise, artifacts could result. We suggest smoothing the multivariate relations to provide adequate data. This may be particularly important to use this technique with seismic data where few calibration data points are available. Log and core data typically provide sufficient data.

Application

It is straightforward to develop programs to perform the stepwise conditional transformation and the back transformation. Data for a porosity / permeability modeling example are shown in **Fig. 5**. These are real data from a deltaic depositional setting.

The conventional approach to geostatistical modeling would be to normal score transform both variables, determine the variogram of each, and to perform some form of cosimulation (typically collocated cosimulation). The porosity – permeability cosimulation would require that the cross plot relation between the normal score of porosity and the normal score of permeability be bivariate Gaussian. We see from **Fig. 6** that this is not the case. The correlation coefficient is 0.59.

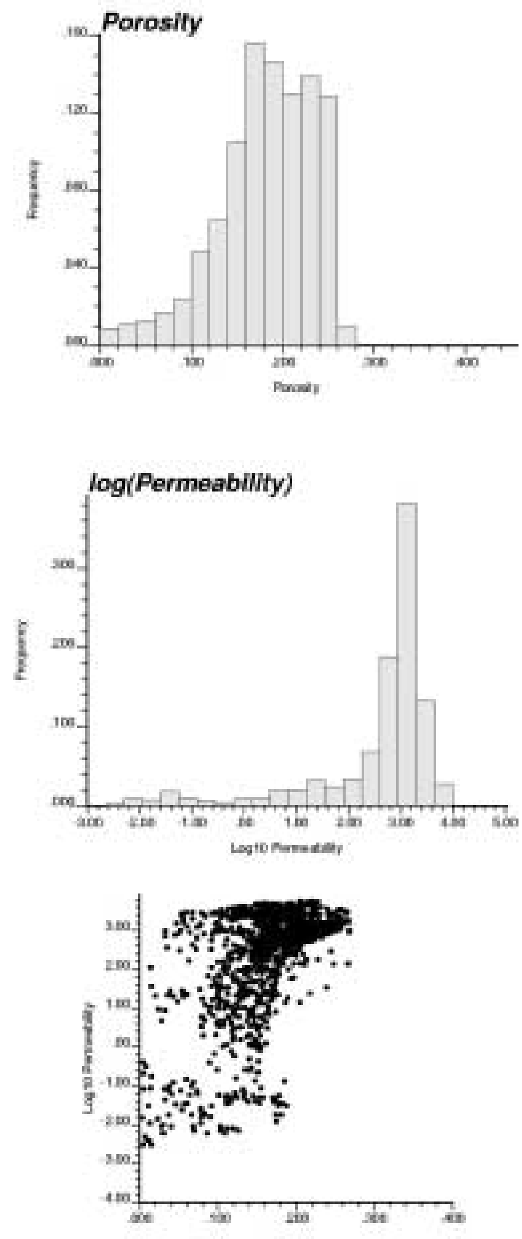

Fig. 5. Porosity / permeability data: top – porosity histogram, middle – permeability histogram, and bottom – porosity – permeability cross plot.

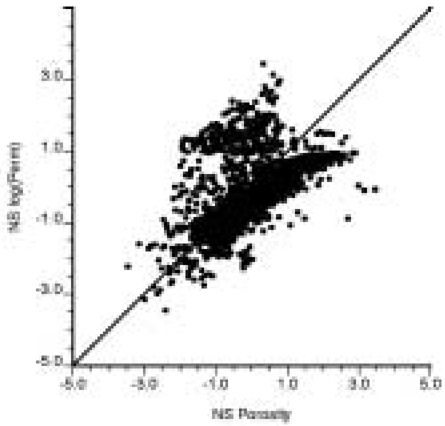

Fig. 6. Cross plot of normal score transforms of porosity and permeability. Note that these points do not follow the elliptical probability contours of the bivariate Gaussian distribution.

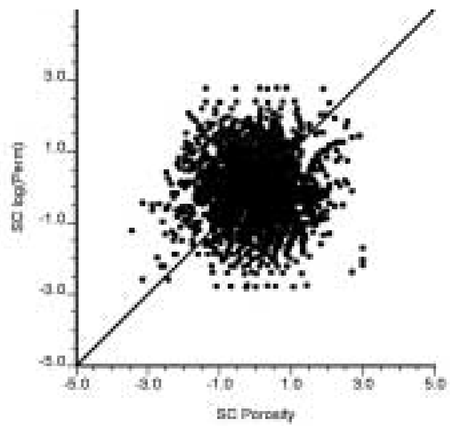

Fig. 7. Cross plot of stepwise conditional transforms of porosity and permeability. Note that these points follow a bivariate Gaussian distribution with zero correlation. There are some minor visual artifacts due to the class structure used in the transformation.

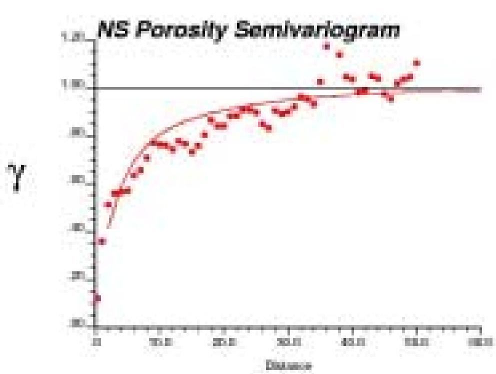

Fig. 8. Variogram of the normal score transform of porosity. This variogram is used for the simulation of porosity in both the conventional and the stepwise approach.

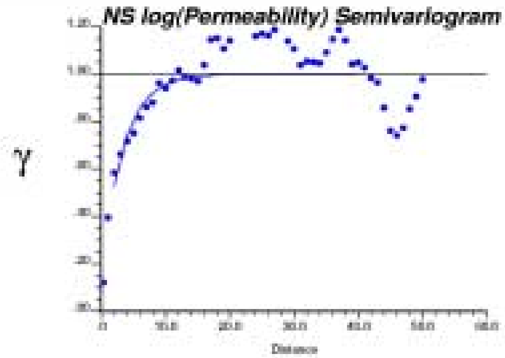

Fig. 9. Variogram of the normal score transform of permeability. This variogram is only used for the simulation of permeability in the conventional approach.

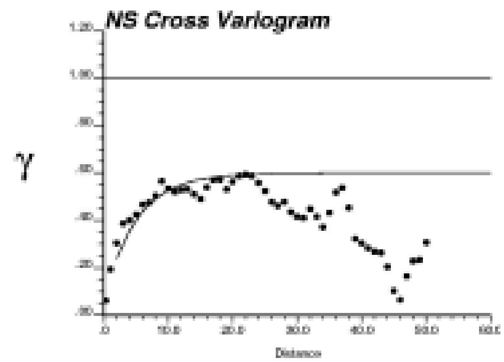

Fig. 10. Cross variogram between the normal score transform of porosity and the normal score transform of permeability. This variogram would only used for a full cosimulation approach. In practice the collocated cosimulation approach is used and this variogram is never explicitly used in reservoir modeling.

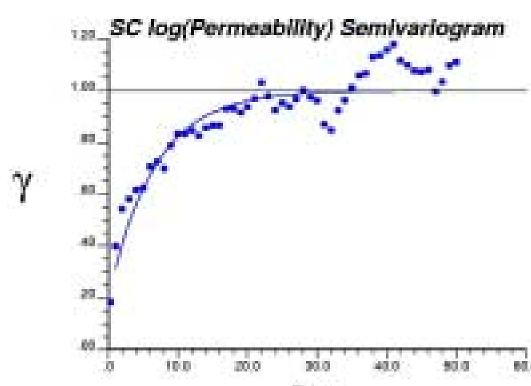

Fig. 11. Variogram of the stepwise conditional transform of permeability: used for the independent simulation of the transform of permeability.

The stepwise conditional transforms of porosity and permeability are plotted on **Fig. 7**: note the circular probability contours of a bivariate Gaussian distribution with zero correlation.

Fig 8 shows the vertical normal scores variogram of porosity that is used for both the conventional and stepwise conditional approaches to simulation. There are only two wells available in this data set; therefore, there is no horizontal variogram and the horizontal to vertical anisotropy ratio is taken from analogue information.

The Gaussian approach to permeability modeling would also require the variogram of the normal score transform of permeability, see **Fig. 9**. The cross variogram between porosity and permeability (see **Fig. 10**) is only needed for a full cokriging approach, which is not used very much due to heavier modeling and computer requirements. The collocated cosimulation approach would only require the sill of the cross variogram or the correlation coefficient of 0.59.

The variogram of the stepwise conditional transform of permeability is shown on **Fig. 11**. As we have mentioned before, there is no need to cosimulate this permeability variable since the correlation is captured in the transformation.

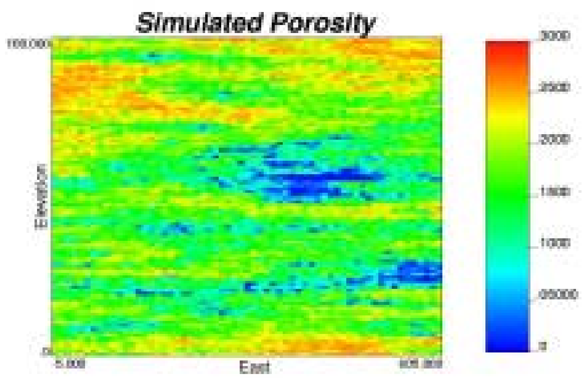

Fig. 12. Color scale image of simulated porosity. This is a vertical cross section. The two vertical wells are at either end of the cross section.

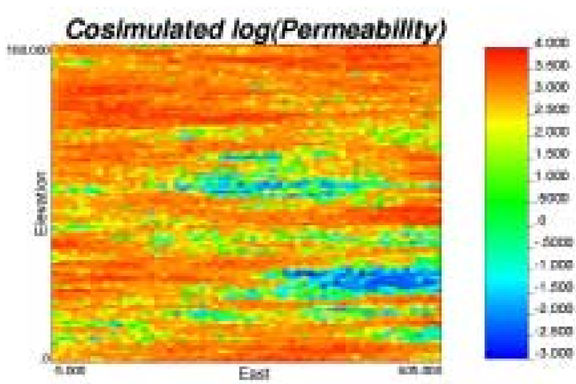

Fig. 13. Color scale image (in log scale) of simulated permeability using collocated cosimulation with correlation coefficient of 0.59. Note the correlation with porosity on Fig. 12.

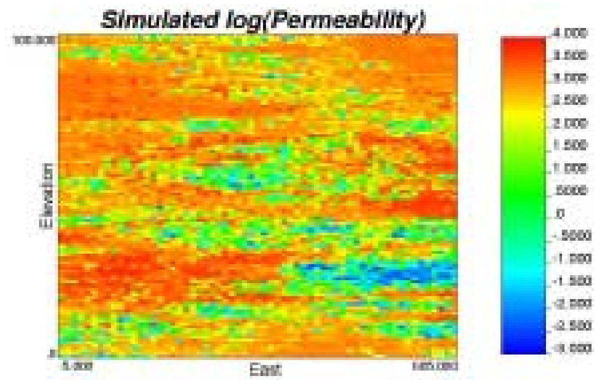

Fig. 14. Color scale image (in log scale) of simulated permeability from stepwise approach, that is, independent simulation of the stepwise conditional transform and then back transform according to correct conditional distributions.

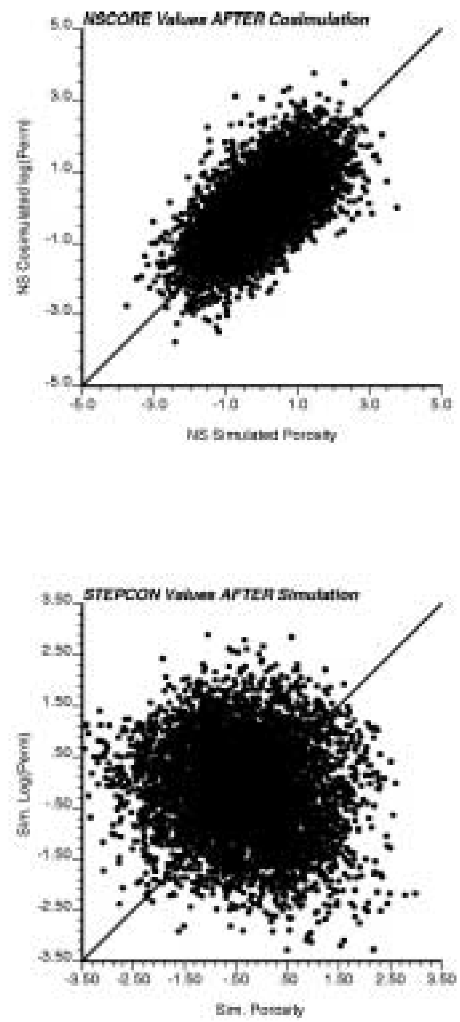

Fig. 15. The "normal-space" cross plots between porosity and permeability. In the top case (conventional) the relation is bivariate Gaussian with the correct correlation coefficient, but does not have the same features as Fig. 6. The lower cross plot shows a bivariate Gaussian relation with no correlation, which is required by the stepwise conditional approach.

Fig. 12 shows a simulated porosity model. No prior facies model was used in this case; however, often porosity and permeability model is modeled on a by-facies basis. Note that the porosity model is identical for both the conventional and stepwise approaches.

Fig. 13 shows a permeability model constructed by collocated cosimulation and the correlation coefficient of 0.59. There are no evident problems with this model although there can be difficulty in reproducing the correct histogram of permeability if the first variable has a variogram model with greater spatial continuity.

The permeability model constructed by after stepwise transformation, independent simulation, and back transformation is shown on **Fig. 14**. This model has no evident problems; it has the right histogram, variogram, and bivariate correlation to porosity.

The bivariate relation resulting from the stepwise conditional transformation is correct provided enough classes have been used in the transformation. The conventional approach often introduces an artifact due to the implicit assumption that the normal score transformed variables are multivariate Gaussian. **Fig. 15** shows the "normal-space" cross plots between porosity and permeability. In the conventional case the resulting bivariate Gaussian relationship has the correct correlation coefficient, but does not have the same features as the data distribution shown on **Fig. 6**. The lower cross plot shows a bivariate Gaussian relation with no correlation, which is required by the stepwise conditional approach, see **Fig. 7**. The cross plot looks larger because there are more points.

The original cross plot is better represented by the stepwise approach. Visualization of the many points on the cross plot of simulated values is awkward because the plot is almost entirely covered by points.

Implementation Details

There are a number of implementation details that must be addressed. They center around the number of classes to use for the conditional distributions and dealing with sparse data.

All data within the same class are treated the same; therefore, no correlation between variables can be captured if only one class is used. Considering the data used above, a cross plot of the correlation coefficient versus the number of classes is shown on **Fig. 16**. 20 classes provides excellent convergence of the correlation coefficient to that of the data.

Of course, there must be sufficient data within each class to provide a reliable conditional distribution. Too few data lead to unnecessary random variations in the transformation. **Fig. 17** illustrates one problem; two data pairs with nearly the same values could appear quite different in "transformed space" because they fall in different classes that are poorly defined by too few data.

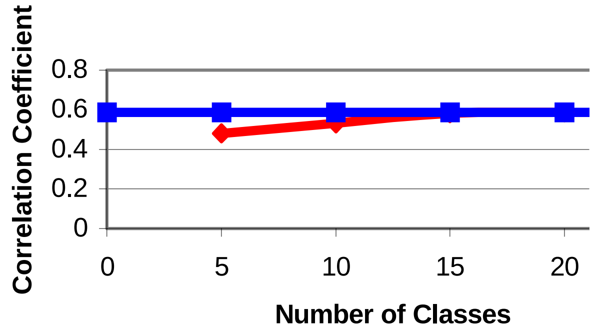

Fig. 16. Convergence of the correlation coefficient with increasing number of classes in the stepwise conditional transformation procedure.

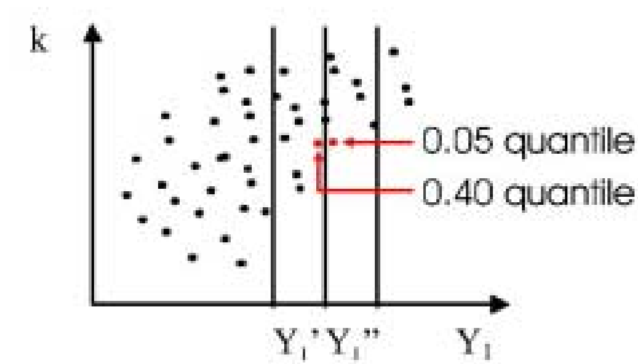

Fig. 17. Schematic illustration showing how two data points of essentially the same values can lead to different transformed values if the number of classes and data are too few.

Stable conditional distributions are defined with 20 to 100 data. This implies that 400 to 2000 data are required in the bivariate case. Although we may not have this many data, we could consider a procedure to *smooth*, *fill in*, or *model* the multivariate distributions. Fig. 18 shows an example where there are only 12 hard data, but 1000 additional data have been added to make fill in the relationship. There are a number of techniques to accomplish this smoothing. The simplest is adding stochastic data points (see **Fig. 18** below) although kernel smoothing[11] and simulated annealing[10] algorithms could be employed.

Adding stochastic data points proceeds in three steps: (1) fit the conditional mean of the second variable given the first, (2) fit the conditional variance and assume a distribution shape, and (3) draw with Monte Carlo simulation many points from the conditional distributions. Of course, the fitted distributions could be used directly.

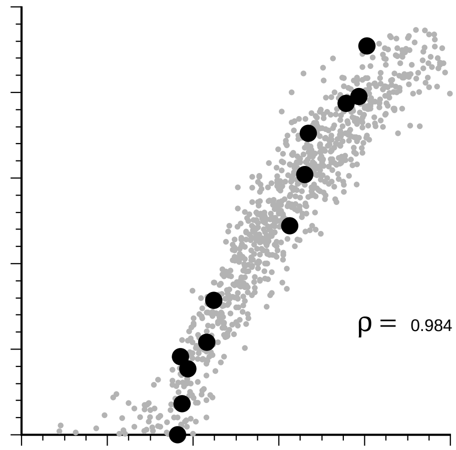

Fig. 18. Twelve hard data (large black dots) and modeled data to permit reliable inference of conditional distributions.

Similarities to P-Field Technique

There are similarities between this stepwise conditional transformation technique and the p-field[3] or cloud transform technique. In particular, the back-transformation according to conditional distributions; however, the procedure is quite different in practice. P-field separates data conditioning from simulation. The stepwise approach requires conditional simulation using the transformed values. This avoids the local minima and local maxima artifacts that result from p-field simulation.

Summary and Future Work

We have presented a technique that will transform multivariate data (e.g., porosity, permeability, water saturation) to transformed variables that are Gaussian *and* independent. This removes the need for cokiriging or cosimulation and greatly simplified reservoir modeling in presence of multiple variables.

Implementation of this technique is straightforward. In presence of sparse data (less than 500 pairs) is is necessary to "fill-in" the cross plot and multivariate relations with synthetic data or a statistical model. Reliable techniques exist for this.

Nomenclature

Y = Gaussian random function
Z = original data variable, e.g., porosity
$\mathbf{u}$ = location in space
$\mathbf{h}$ = separation lag distance in space

References

1. Xu, W., T. T. Tran, R. M. Srivastava and A. G. Journel, Integrating Seismic Data in Reservoir Modeling: the Collocated Cokriging Alternative, 1992 SPE Annual Technical Conference and Exhibition, Washington, DC, SPE Paper Number 24742, pp 833-842
2. Doyen, P. M., L. D. den Boer, and W. R. Pillet, Seismic Porosity Mapping in the Ekofisk Field Using a New Form of Collocated Cokriging, 1996 SPE Annual Technical Conference and Exhibition, Denver CO, SPE Paper Number 36498, pp 21-30
3. Srivastava, R. M., Reservoir Characterization with Probability Field Simulation, 1992 SPE Annual Technical Conference and Exhibition, Washington, DC, SPE Paper Number 24753, pp. 927-938
4. Anderson, T. *An Introduction to Multivariate Statistical Analysis*, John Wiley & Sons, 1958
5. Verly, G., Sequential Gaussian Cosimulation: A Simulation Method Integrating Several Types of Information, Geostatistics Troia '92, vol 1., ed. A. Soares, Kluwer Academic Publishers, 1993, pp. 543-554
6. Isaaks, E. H., *The Application of Monte Carlo Methods to the Analysis of Spatially Correlated Data*, Ph.D. Thesis, Stanford University, 1990
7. Rosenblatt, M., 1952, Remarks on a Multivariate Transformation, Annals of Mathematical Statistics, vol 23, no 3, pp 470-472.
8. Luster, G. R., *Raw Materials for Portland Cement: Applications of Conditional Simulation of Coregionalization*, Ph.D. Thesis, Stanford University, 1985
9. Goovaerts, P., Geostatistics for Natural Resources Evaluation, Oxford University Press, 1997
10. Deutsch, C. and Journel, A.G.: *GSLIB: Geostatistical Software Library and User's Guide*, 2nd edition, Oxford University Press, 1998.
11. Scott, D. W., Multivariate Density Estimation: Theory, Practice, and Visualization, John Wiley & Sons, 1992