Mathematical Geology, Vol. 32, No. 1, 2000

Integrating Large-Scale Soft Data by Simulated Annealing and Probability Constraints[1]

C. V. Deutsch[2] and X. H. Wen[3]

Interpretation of geophysical data or other indirect measurements provides large-scale soft *secondary data for modeling* hard *primary data variables. Calibration allows such soft data to be expressed as prior probability distributions of nonlinear block averages of the primary variable; poorer quality soft data leads to prior distributions with large variance, better quality soft data leads to prior distributions with low variance. Another important feature of most soft data is that the quality is spatially variable; soft data may be very good in some areas while poorer in other areas. The main aim of this paper is to propose a new method of integrating such soft data, which is large-scale and has locally variable precision. The technique of simulated annealing is used to construct stochastic realizations that reflect the uncertainty in the soft data. This is done by constraining the cumulative probability values of the block average values to follow a specified distribution. These probability values are determined by the local soft prior distribution and a nonlinear average of the small-scale simulated values within the block, which are all known. For each realization to accurately capture the information contained in the soft data distributions, we show that the probability values should be uniformly distributed between 0 and 1. An objective function is then proposed for a simulated annealing based approach to enforce this uniform probability constraint. The theoretical justification of this approach is discussed, implementation details are considered, and an example is presented.*

KEY WORDS: geostatistical simulation, stochastic modeling, reservoir characterization.

INTRODUCTION

There are often few *hard* primary z data available to build a complete three-dimensional model of the z variable. In many cases, the limited hard data are supplemented by *soft* secondary data that are related to the z variable of interest. In a petroleum reservoir characterization context, seismic measurements, remote sensing, inversion of historical production data, or geological interpretation is used for soft data. We present our new method in the context of petroleum reservoir

[1]Received 24 August 1998; accepted 10 February 1999.
[2]Department of Civil & Environmental Engineering, 204 C Civil/Electrical Engineering Building, University of Alberta, Edmonton, Alberta, Canada, T6G 2G7. e-mail: cdeutsch@civil.ualberta.ca
[3]Chevron Petroleum Technology Company, P.O. Box 446, La Habra, California USA, 90633-0446. e-mail: xwen@chevron.com

modeling; however, it is widely applicable to any soft data integration problem encountered in earth science modeling.

Soft data are often at a large scale relative to the scale of the z variable being modeled. Another feature of many soft data types is that the precision or "hardness" of the soft data is locally variable. These two features make it difficult to apply conventional geostatistical algorithms to soft data integration.

Although the focus here is on integrating large-scale soft data the resulting realizations of the small scale z values must simultaneously honor additional data such as following: (1) local hard z data; (2) the histogram of small scale z data; (3) measures of z spatial variability, e.g., z-variogram and, possibly, indicator variograms for special continuity of extreme values; and (4) soft data from other sources.

The common situation of large-scale soft data is illustrated in Figure 1. Note two important features of soft data: (1) the volume of measurement is significantly larger than either the scale of the hard data or the modeling scale; and (2) the precision of the soft data derived distributions is better in certain areas, e.g., near the center of the area of interest where the calibration is more robust or where the interpretation of the raw secondary data, such as historical production data, is more reliable. These prior distributions are obtained by seismic-data calibration (Deutsch, Srinivasan, and Mo, 1996), inversion of historical production data (Oliver, 1994; Wen, Deutsch, and Cullick, 1997), or by geological interpretation.

Conventional geostatistical approaches have difficulty in handling this type of data (Dubrule, 1989; Journel and Huijbregts, 1978; Goovaerts, 1997). The kriging equations in Gaussian-based simulation methods could be written to include soft data of large volumetric support. The resulting cokriging equations, however, would call for the difficult inference of many statistical parameters: (1) the *small*-scale hard–soft cross variogram for each soft datum, (2) the cross variograms between all soft data, (3) a positive definite model of coregionalization that would fit all of these variogram models, and (4) linear averaging (perhaps after some ω-power transform).

The inference problem would be even more daunting for indicator-based methods. A common approach in indicator geostatistics is to transform the soft data to prior distributions of the hard primary variable of interest. Then, some model, such as the Markov–Bayes model (Zhu, 1991), is adopted to update these distributions to posterior distributions for geostatistical simulation. The large volumetric scale of the soft data considered here makes this approach intractable; the prior distributions are of nonlinear averages of the primary variable and *not* on the primary variable itself.

Another significant problem with kriging-based approaches is that there is no convenient way to handle the fact that the soft data have locally variable precision. One could imagine grouping soft data according to some measure of "goodness" and considering each grouping as a different type of soft data.

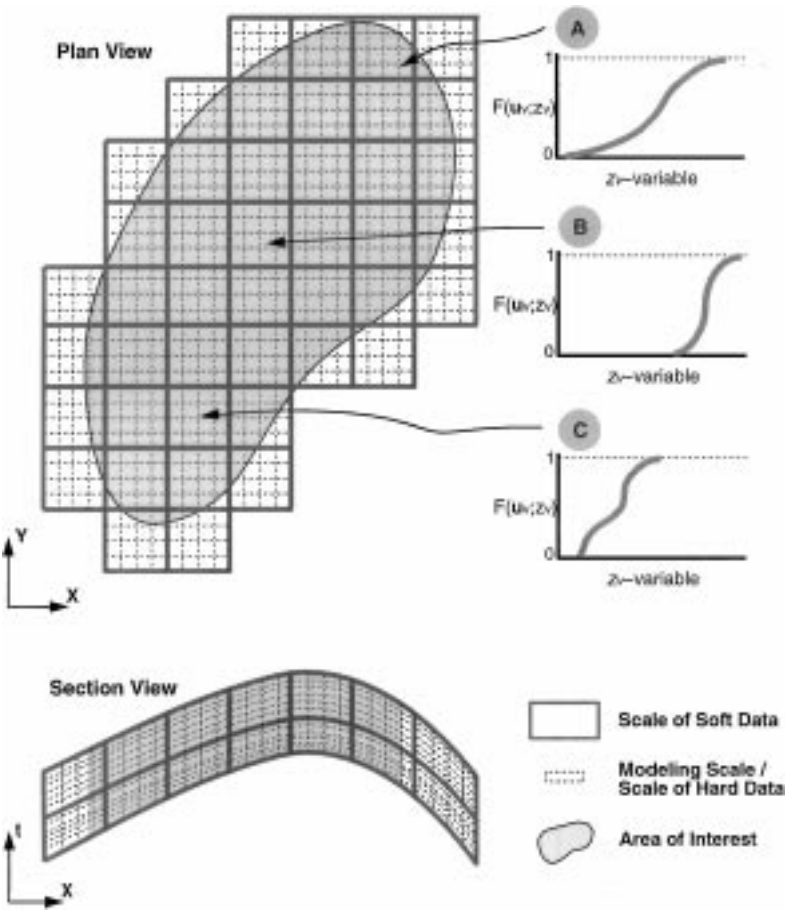

Figure 1. A schematic illustration of a reservoir showing the scale of soft data derived from seismic or production data and the scale of hard data and modeling scale. The large-scale soft data may be represented as prior distributions of z_v averages, $F(\mathbf{u}_v; z_v)$. At certain locations (A) the soft data does not provide much information and $F(\mathbf{u}_v; z_v)$ has a large variance. At other locations (B and C), the soft data provide information on whether z_v is high (B) or low (C).

This would make the inference problem for kriging-based approaches even more difficult.

An iterative *simulated-annealing* approach will be considered here to accomplish the integration of these diverse data types. The difficulty with annealing-based methods is often the delicate adjustment of many tuning parameters. Considering soft data by the methodology presented below speeds convergence and mitigates the difficulty of setting tuning parameters such as the annealing schedule.

The soft data values are reformatted as prior distributions of block z values. In some cases, such as large-scale inversion of historical production data, the soft data are already in the correct format. In presence of soft seismic data, the reformatting requires (1) knowledge of the averaging process, e.g., arithmetic, geometric, harmonic, . . . , and (2) a calibration cross plot between z-average values and collocated soft data. Knowledge of the averaging process comes from an external understanding of the physics. The calibration cross plot comes from locations where both block z values and the soft data variable are known, for example, at well data locations. The distribution of block z values is then inferred for each soft data value.

We simulate the z values at a small scale. It is possible, however, to calculate large-scale z block values at any block location in the iterative simulation process. These large-scale block values are calculated at the locations of a soft data derived prior probability distribution. The simulated values are considered as quantiles on the corresponding soft data distributions. Then, these z quantiles are transformed to cumulative probability (CDF) values, which are unit free and account for differences in the local distributions. The "differences" in the local distributions are due to the fact that the soft data are better in some areas and less informative in other areas.

It will be shown that these probability values must be uniformly distributed between 0 and 1 to reflect the "information" in the soft data. This notion was introduced in the context of cross-validating stochastic simulation algorithms (see Deutsch, 1996).

The approach presented in this paper is based on the *definition* of a probability distribution and quantiles of probability distributions. As such, the notation may be unusual to many geoscientists; it will be necessary to go between figures and the notation.

The methodology and theoretical framework will be presented first. Then, implementation details using an objective function in simulated annealing will be discussed. It is interesting to note that the theoretical background has no direct link to simulated annealing; other numerical algorithms could be used. Finally, an example is presented with ideas for future development.

METHODOLOGY

Consider the distribution over a field A of an attribute $z(\mathbf{u}), \mathbf{u} \in A$, where z could represent a categorical indicator variable or a continuous variable such as a mineral concentration or other petrophysical property. The location vector for the z-variable at the small 3D modeling scale is denoted $\mathbf{u} = (x, y, t)$, where t is the vertical "stratigraphic" coordinate. The goal here is to construct alternative, equally probable, high-resolution 3D models of the spatial distribution of $z(\mathbf{u})$; each realization is denoted with the superscript l: $\{z^{(l)}(\mathbf{u}), \mathbf{u} \in A\}, l = 1, \ldots, L$. Soft

secondary data available at a coarse scale (denoted v), provides prior probabilistic information on volumetric averages of z:

$$F(\mathbf{u}_v; z_v), \qquad \mathbf{u}_v \in A \tag{1}$$

The location $\mathbf{u}$ and hard data value z are subscripted by v to denote the scale of the location and the value being considered; $\mathbf{u}_v$ is a coarse grid block and z_v is an average of z over that block. These distributions are nonstationary—that is, dependent on the soft data around $\mathbf{u}$:

$$F(\mathbf{u}_v; z_v) = \text{Prob}\{Z_v(\mathbf{u}) \leq z \mid \text{soft data around } \mathbf{u}\}$$

where $z_v(\mathbf{u})$ could be represented as the ω-power (block) average of z values within v centered at $\mathbf{u}$:

$$z_v(\mathbf{u}_v) = \left[\int_{\mathbf{u}' \in v(\mathbf{u})} z^{\omega}(\mathbf{u}')\, d\mathbf{u}' \right]^{\frac{1}{\omega}} \tag{2}$$

where $\omega = 1$ for a categorical indicator variable and most continuous variables and $\omega \in [-1, 1]$, accounting for nonlinear averaging in the case of permeability. The specific value of ω must be established through a calibration exercise. Such a calibration exercise would consist of (1) generating fine z scale realizations of the variable, (2) scaling the fine-scale realization to a coarse v-scale realization, and (3) calculate the ω value that closely reproduces the v-scale values.

The $\omega = 1$ average for a categorical indicator variable would represent the fraction of that category at location $\mathbf{u}_v$. There would, of course, be one such average for each category. The procedure will be developed below for a continuous variable. An identical implementation would be used for a binary system (with only two categories); however, implementation for more categories would require increasing the dimensionality to consider multiple indicators, that is, multiple averages.

The soft data distributions are established through calibration with the hard data. This is a different type of calibration than the ω calibration. Here, the calibration exercise is that performed by the bicalib program (and related discussion) in the GSLIB book, that is, the soft data are directly compared to the hard data at the hard data locations. In the simplest case, the conditional distributions may be taken from the cross plot between the soft and hard data. In the presence of too few calibration data, the cross plot may need to be supplemented by small-scale "physics-based" simulations or expert interpretation (Alabert, 1989, or Deutsch, 1992).

The problem is to construct 3D realizations, $\{z^{(l)}(\mathbf{u}), \mathbf{u} \in A\}, l = 1, \ldots, L$, of z at the scale of the hard data (the modeling scale illustrated in Fig. 1) that honor the given prior distributions of $z_v(\mathbf{u}_v)$. For a given realization $\hat{l}$ at location $\mathbf{u}'_v$, there

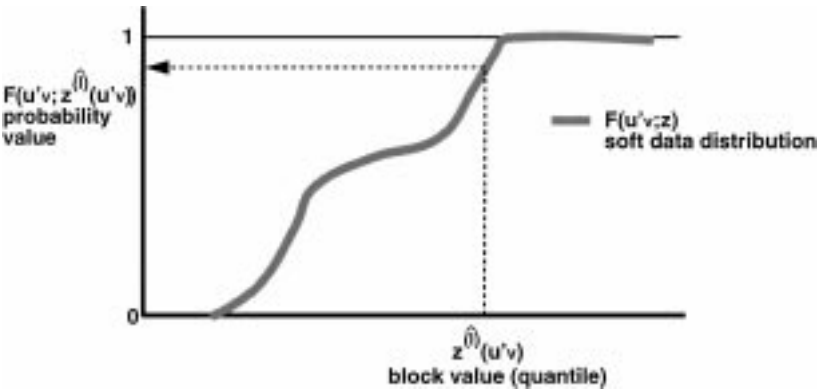

Figure 2. A schematic illustration showing how the probability value $F(\mathbf{u}'_v; z_v^{(\hat{l})}(\mathbf{u}'_v))$ is taken directly from the soft data prior distribution given the local hard data average $z_v^{(\hat{l})}(\mathbf{u}'_v)$.

is only one block average—$z_v^{(\hat{l})}(\mathbf{u}'_v)$—which corresponds to a particular probability value on the prior distribution, that is, $F(\mathbf{u}'_v; z_v^{(\hat{l})}(\mathbf{u}'_v))$. If $z_v^{(\hat{l})}(\mathbf{u}'_v)$ is less than the minimum value of the prior distribution, then $F(\mathbf{u}'_v; z_v^{(\hat{l})}(\mathbf{u}'_v)) = 0$; when $z_v^{(\hat{l})}(\mathbf{u}'_v)$ is below the median, then $F(\mathbf{u}'_v; z_v^{(\hat{l})}(\mathbf{u}'_v)) < 0.5$; when $z_v^{(\hat{l})}(\mathbf{u}'_v)$ is above the median, then $F(\mathbf{u}'_v; z_v^{(\hat{l})}(\mathbf{u}'_v)) > 0.5$; and when $z_v^{(\hat{l})}(\mathbf{u}'_v)$ is greater than the maximum value of the prior distribution, $F(\mathbf{u}'_v; z_v^{(\hat{l})}(\mathbf{u}'_v)) = 1$. The probability $F(\mathbf{u}'_v; z_v^{(\hat{l})}(\mathbf{u}'_v))$ is taken directly from the soft data derived prior distribution (see calibration discussed in previous paragraph) given the local hard data average $z_v^{(\hat{l})}(\mathbf{u}'_v)$ (Fig. 2).

If no other information is available, the distribution of average values (z_v), over many realizations, for each coarse grid block $\mathbf{u}_v$, should match the distribution provided by the prior soft information. That is, for a large number of realizations (L):

$$F_L(\mathbf{u}_v; z_v) = F(\mathbf{u}_v; z_v), \qquad \forall \mathbf{u}_v \in A, \forall z_v \in [z_{v,\min}, z_{v,\max}] \tag{3}$$

where

$$F_L(\mathbf{u}_v; z_v) = \sum_{l=1}^{L} i^{(l)}(\mathbf{u}_v; z_v); \qquad i^{(l)}(\mathbf{u}_v; z_v) = \begin{cases} 1, & \text{if } z_v^{(l)}(\mathbf{u}_v) \le z_v \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

Integrating additional local hard and soft data would amount to "update" the prior soft distributions; therefore, in practice, the distributions $F_L(\mathbf{u}_v; z_v)$ will not equal the soft data prior distributions $F(\mathbf{u}_v; z_v)$. An implicit assumption in the following development is that this updating is not significant, i.e., the distribution $F_L(\mathbf{u}_v; z_v)$ should be close to $F(\mathbf{u}_v; z_v)$ for most locations. This is a reasonable assumption in

the common case of sparse hard data and when the soft data have been calibrated to the hard data.

If we were to generate all "L" realizations at one time, it would be possible to compare the simulated distributions $F_L(\mathbf{u}_v; z(\mathbf{u}_v))$ to the prior distributions $F(\mathbf{u}_v; z_v)$.

By construction, the probability values $F_L(\mathbf{u}_v; z_v^{(l)}(\mathbf{u}_v))$ are uniformly distributed between 0 and 1. This is due to the definition of the cumulative distribution and quantile functions (see figures for graphical illustration). Thus, if we interpret the simulated values $z_v^{(l)}(\mathbf{u}_v)$ as quantiles of the soft data distributions $F(\mathbf{u}_v; z_v)$ (see Fig. 2 for an illustration of the meaning), the associated probability values

$$\mu^{(l)}(\mathbf{u}_v) = F\left(\mathbf{u}_v; z_v^{(l)}(\mathbf{u}_v)\right), \qquad l = 1, \ldots, L, \text{ fixed } \mathbf{u}_v \tag{5}$$

must also be uniformly distributed between 0 and 1 to meet the constraint expressed in Eq. (3). The distribution of these probability values, $F'(\mathbf{u}_v; \mu)$ is only defined when the quantile values for all realizations $z_v^{(l)}(\mathbf{u}_v), l = 1, \ldots, L$ are available; however, for practical reasons, we would like to generate only one realization at a time.

The central idea underlying our proposal is to trade the unavailable replication over all L realizations for replication available over the set of locations $\mathbf{u}_v$ in the area of interest A. That is, we consider the distribution defined by the probability values over all locations:

$$\mu^{(l)}(\mathbf{u}_v) = F\left(\mathbf{u}_v; z_v^{(l)}(\mathbf{u}_v)\right), \qquad \mathbf{u}_v \in A, \text{ fixed } l \tag{6}$$

Choosing the distribution over all locations instead of the distribution over all realizations of the random function is analogous to the common decision of *stationarity*. In statistical inference, *stationarity* amounts to replace the unavailable replication at a particular location $\mathbf{u}$ for replication elsewhere in the area of interest (Deutsch and Journel, 1992; Isaaks and Srivastava, 1989). The appropriateness of this decision can be checked *a posteriori* (as we do later in this paper); however, it cannot be proven. In all rigor we would like to impose constraint (5) and not (6).

Figure 3 presents this concept in graphical form. The first three soft data cumulative distributions $F(\mathbf{u}_v; z_v)$ are illustrated at the top of Figure 3; in practice, there are many such distributions—one for each coarse grid block location. These distributions are fixed, that is, they do not change during the simulation of z or for different realizations. The simulated values [labeled as $z_v(\mathbf{u}_v)$, $z_v(\mathbf{u}'_v)$, $z_v(\mathbf{u}''_v)$ on the figure] depend on the simulation method and realization. Each of these quantile values lead to different cumulative probability values (0.68, 0.11, and 0.93 in the figure). The histogram at the bottom of the figure represents the fraction (of the n block values) that fall into each class of probability. The soft data are honored, in this case, when 1/5 of the values falls into each of the 5 classes.

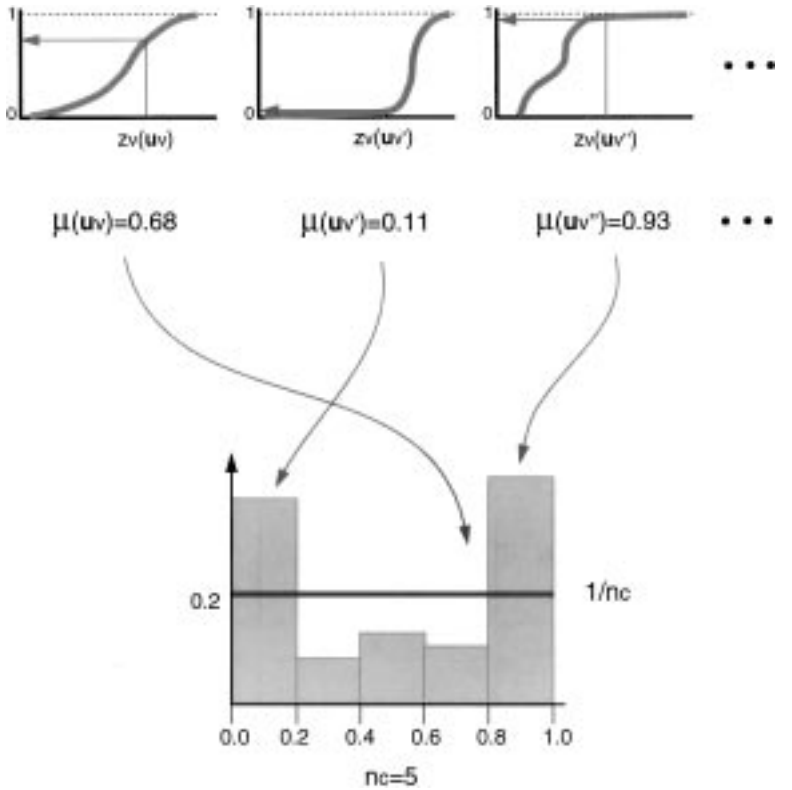

Figure 3. A schematic illustration of the idea of using probability constraints.

We now present an objective function that imposes constrain (6) and study its effectiveness at integrating large-scale soft data. Our examples will show that the soft data is honored accounting for both its scale and precision.

SIMULATED ANNEALING FOR GEOSTATISTICAL MODELING

Integration of large-scale soft data will be enforced through an objective function in a simulated-annealing based approach (Deutsch, 1992; Farmer, 1992; Hird, 1993; Ouenes and others, 1994; Perez, 1992). Conventional constraints such as the histogram, variogram, and indicator variograms will be included (see code in second edition of GSLIB; Deutsch and Journel, 1997). The multiple constraints will be weighted such that all reach low final values for each realization. Note, however, that no weighting permits low objective function values in presence of

conflicting objective function components. Conflicting data must be resolved prior to simulated annealing.

The basic elements of the simulated-annealing based approach are as follows:

1. Create an initial realization $z(\mathbf{u}_j), \mathbf{u}_j \in A, j = 1, \ldots, N$, where N is the number of grid cell locations in the fine-scale model (the scale of the z model being constructed).
2. Calculate a quantitative objective function that measures the mismatch between the desired features and those of the realization.
3. Perturb the model by changing the property value at some cell $j' \in [1, \ldots, N]$.
4. Calculate a new objective function by updating the current one.
5. Apply the simulated-annealing decision rule to decide whether or not to accept the perturbation (this decision rule requires a temperature parameter, which is reduced periodically to ensure ultimate convergence of the optimization problem; see Deutsch, 1992).
6. Return to step 3 until the objective function has been lowered to a point where the desired data have been satisfactorily honored.

In the context of integrating large-scale soft data, we need to write an equation for the objective function and show how it can be quickly updated after a local perturbation. For clarity, we will drop the (l) superscript in the following; we create one realization and then start all over again with a new random initial realization and sequence of random numbers.

Consider n soft data at locations $\mathbf{u}_{v_i}, i = 1, \ldots, n$. Given the initial realization, the hard variable $[z(\mathbf{u}_j), \mathbf{u}_j \in A, j = 1, \ldots, N]$ can be averaged to the scale of the soft data, perhaps nonlinearly, with (2) to yield $z_v(\mathbf{u}_{v_i}), i = 1, \ldots, n$. The probability value at each coarse grid block location $\mu(\mathbf{u}_{v_i}) = F(\mathbf{u}_{v_i}; z_v(\mathbf{u}_{v_i}))$ can be calculated from the known $F(\mathbf{u}_{v_i})$ distributions.

Closer Look at Soft Data Distributions

The n soft data distributions $F(\mathbf{u}_{v_i}; z_v), i = 1, \ldots, n$ are known through calibration. Thus, we can calculate the mean of each distribution:

$$\bar{z}_v(\mathbf{u}_v) = \int_{-\infty}^{+\infty} z_v \, dF(\mathbf{u}_v; z_v) \tag{7}$$

where the subscript $i = 1, \ldots, n$ is implicit to this equation.

These conditional means must be considered in our formulation of an objective function. For example, areas with large conditional means will have small starting cumularive probability values (recall that the initial realization of z's is

assigned randomly). Conversely, when the conditional mean is low, the cumulative probability values are likely to be high. See Figure 1 for an illustration. Although fully expected, these biased starting values could lead to a subtle bias in the final realizations. In particular, imposing a uniform distribution of probability values over all locations (6) will not necessarily ensure a uniform distribution over many realizations [Eq. (5)] in "low" and "high" regions. High z-valued areas could be underestimated and low z-valued areas could be overestimated.

The conditional mean values are known and constant for all locations (independent of the realization). Therefore, to avoid this bias we could simply classify all locations by their conditional mean values, for example, we could define 10 classes based on the 9 deciles of the conditional mean distribution. Then, the objective function to ensure uniformity can be written for each class. In particular, the distribution of conditional mean values $F_{\overline{z}}(z)$ could be established by taking the average [Eq. (7) above] of the soft data distributions at the soft data locations $\mathbf{u}_{v_i}, i = 1, \ldots, n$. An indicator variable would then be defined that classifies each location into 1 of n_m equal probability classes:

$$i_m(m; \mu_v) = \begin{cases} 1, & \text{if } \frac{m-1}{n_m} \leq F(\overline{z}(\mathbf{u}_v)) < \frac{m}{n_m} \\ 0, & \text{otherwise} \end{cases} \qquad m = 1, \ldots, n_m \tag{8}$$

Uniformity of the probability values will be enforced within each of the n_m classes; there are about n/n_m data in each class. Once again, the reason to partition the domain into these classes is to avoid systematic under or overestimation.

Objective Function

As described in Eq. (6), the goal is for the distribution of $\mu(\mathbf{u}_{v_i}), i = 1, \ldots, n$ values to be uniformly distributed. To avoid bias, our goal is for the n_m distributions of $\mu(\mathbf{u}_{v_{j,m}}, j = 1, \ldots, n/n_m; m = 1, \ldots, n_m)$ to be uniform. To lighten the notational burden, a global objective function for $\mu(\mathbf{u}_{v_i}), i = 1, \ldots, n$ will be written. Then, the procedure can be applied n_m times.

To enforce the distribution of $\mu(\mathbf{u}_{v_i}), i = 1, \ldots, n$ values to be uniformly distributed, we define $n_c + 1$ thresholds $\mu_c, c = 0, \ldots, n_c$ to equally divide the range of μ (0 to 1):

$$\mu_0 = 0.0; \qquad \mu_c = \frac{c}{n_c}, \qquad c = 1, \ldots, n_c \tag{9}$$

An indicator variable is defined for each class:

$$i_\mu(c; \mu) = \begin{cases} 1, & \text{if } \mu_{c-1} \leq \mu < \mu_c \\ 0, & \text{otherwise} \end{cases} \qquad c = 1, \ldots, n_c \tag{10}$$

The proportion in each class is then

$$p_c = \frac{\sum_{k=1}^{n} i_u\big(c; \mu(\mathbf{u}_{v_k})\big)}{n}, \qquad c = 1, \ldots, n_c \tag{11}$$

where n is the number of coarse block locations. The closeness to a uniform distribution is measured by

$$O_\mu = \sum_{c=1}^{n_c} \left[p_c - \frac{1}{n_c} \right]^2 \tag{12}$$

For practical application of simulated-annealing based algorithms, we must be able to quickly update the objective function after a local change. For a candidate change of $z(\mathbf{u}_{j'})$ to $z^{\text{new}}(\mathbf{u}_{j'})$, the z average in the coarse v block containing location $\mathbf{u}_{j'}$ is changed, i.e., $z_v(\mathbf{u}_{v_{i'}})$ changes to $z_v^{\text{new}}(\mathbf{u}_{v_{i'}})$ by

$$z_v^{\text{new}}\big(\mathbf{u}_{v_{i'}}\big) = \left[z_v\big(\mathbf{u}_{v_{i'}}\big)^\omega + \frac{1}{n_v}\big(z^{\text{new}}(\mathbf{u}_{j'})^\omega - z(\mathbf{u}_{j'})^\omega\big) \right]^{\frac{1}{\omega}} \tag{13}$$

where n_v is the number of fine scale cells within a coarse block v. The current class c' of $z_v(\mathbf{u}_{v_{i'}})$ and the class c'' of the candidate value $z_v^{\text{new}}(\mathbf{u}_{v_{i'}})$ are readily determined from the original probability value and the one considering the change, i.e., $F(\mathbf{u}_{v_{i'}}; z_v(\mathbf{u}_{v_{i'}}))$ and $F(\mathbf{u}_{v_{i'}}; z_v^{\text{new}}(\mathbf{u}_{v_{i'}}))$ using (10). The class probabilities (11) are changed as follows:

$$p_{c'}^{\text{new}} = p_{c'} - \frac{1}{n}$$

and

$$p_{c''}^{\text{new}} = p_{c''} + \frac{1}{n}$$

the updated objective function is

$$O_\mu^{\text{new}} = O_\mu - \left[p_{c'} - \frac{1}{n_c} \right]^2 + \left[p_{c'}^{\text{new}} - \frac{1}{n_c} \right]^2 - \left[p_{c''} - \frac{1}{n_c} \right]^2 + \left[p_{c''}^{\text{new}} - \frac{1}{n_c} \right]^2$$

This local updating is necessary for fast implementation of the proposed approach.

The sasim simulated annealing program in GSLIB (Deutsch and Journel, 1992) can be easily modified to include the objective function described in Eq. (12)

with the local updating rule. Recall that the objective function will be applied in n_m classes.

As a preliminary example, Figure 4 shows reproduction of the global distribution of uniform probability values (left) and the cross plot of the probability values vs the conditional mean. Note the desired lack of correlation between the probability values and the conditional mean. This example is a 100 by 100 realization.

In the example developed below we show that this objective function leads to a uniform distribution over many realizations (5). We have have checked for relationships with other characteristics, such as the local variance, but have observed none.

EXAMPLE

In this section we demonstrate the application of the proposed approach to integrate large-scale soft data.

The first step is to display the large-scale soft data that we are considering. A complete, but paper-consuming, approach, is to show a series of probability or quantile maps. A less complete, but more consise, approach is to show the mean and variance of each soft data-derived local distribution, i.e.,

$$E\{F(\mathbf{u}_v; z_v)\} \qquad \text{and} \qquad \text{Var}\{F(\mathbf{u}_v; z_v)\}$$

where $E\{\cdot\}$ and $\text{Var}\{\cdot\}$ are the expected value and variance operators.

Synthetic data are considered here. The data are from a study related to the integration or historical production data in geostatistical reservoir modeling. Figure 5 shows the expected value, $(E\{F(\mathbf{u}_v; z_v)\})$, and standard deviation, $(\sqrt{\text{Var}\{F(\mathbf{u}_v; z_v)\}})$, of the soft data distributions considered here. As commonly encountered with many data, the distributions in the central part of the area of interest have less variance. Note the continuity in the 20° direction, the two bands of high permeability, and the region of reduced permeability in the lower right corner. The variance map must be consulted to judge the reliability of the mean map. These coarse-scale distributions will be used, together with an assumed known histogram and variogram of the fine-scale z variable, for stochastic simulation. Locations (A) and (B) are highlighted for more detailed checking; the histograms of z_v at (A) and (B) are given at the bottom of Figure 5. (A) is in a region with relatively large uncertainty and (B) is in a location with low uncertainty.

Figure 6 shows two coarse-scale z realizations. These realizations were constructed with a revised sasim simulated annealing program to honor (1) the input target coarse-scale histogram, (2) the input target coarse-scale variogram, and (3) closeness of the probability values $\mu^{(l)}(\mathbf{u}_v) = F(\mathbf{u}_v; z_v^{(l)}(\mathbf{u}_v))$ to a uniform distribution. Each simulation took less than 30 CPU sec on an SGI Indigo 2. Each component in the three-part objective function was lowered to less than 0.01,

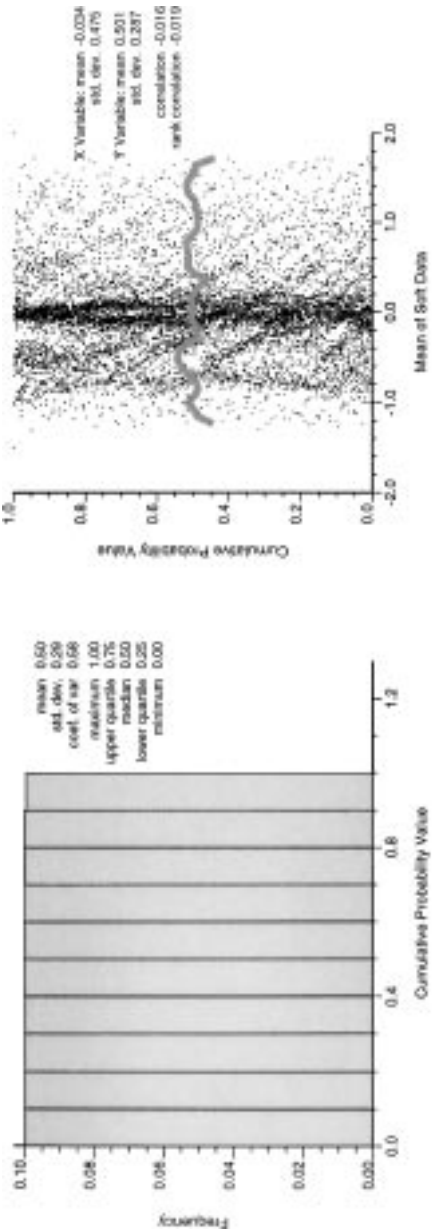

Figure 4. Simulated annealing reproduces the constraint of uniformly distributed probability values (left) for all classes of conditional mean (shown on right). The thick gray line is a moving window average of the probability values.

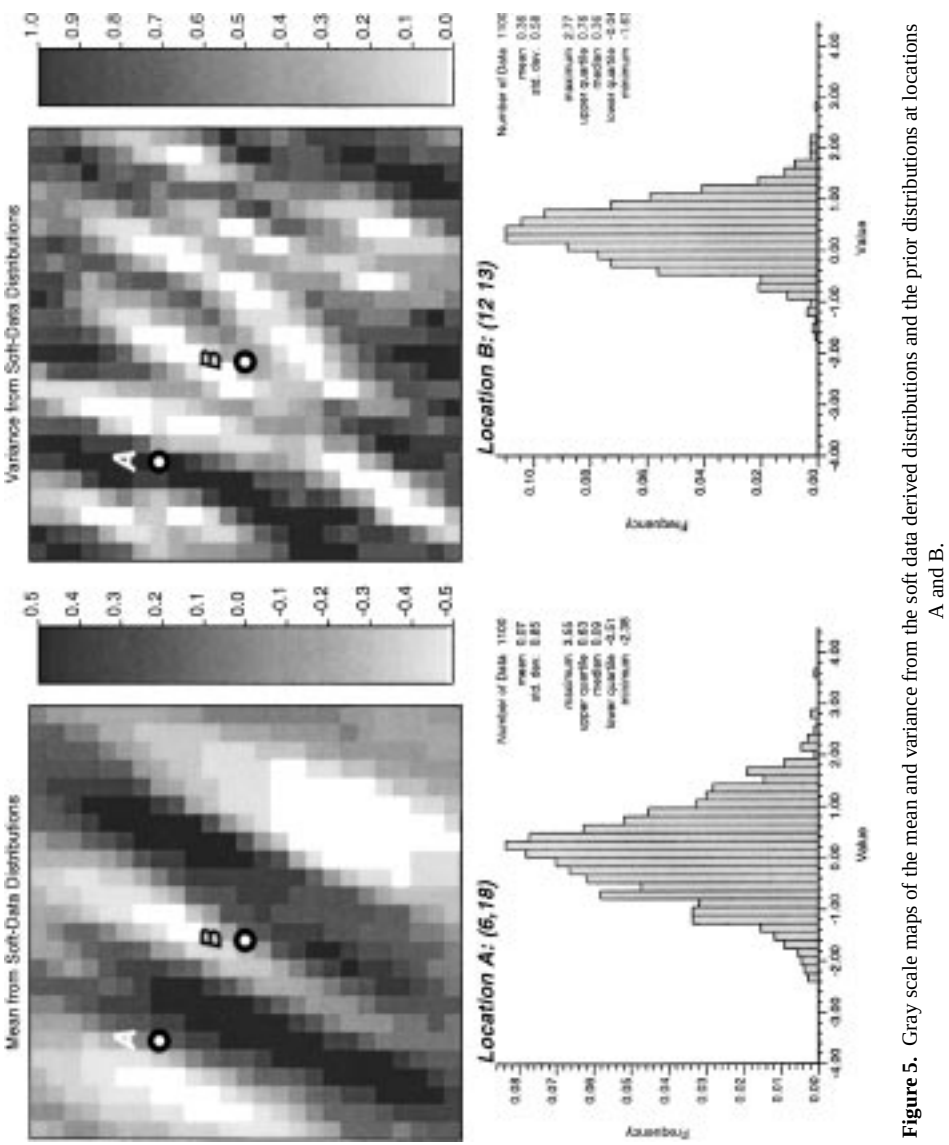

Figure 5. Gray scale maps of the mean and variance from the soft data derived distributions and the prior distributions at locations A and B.

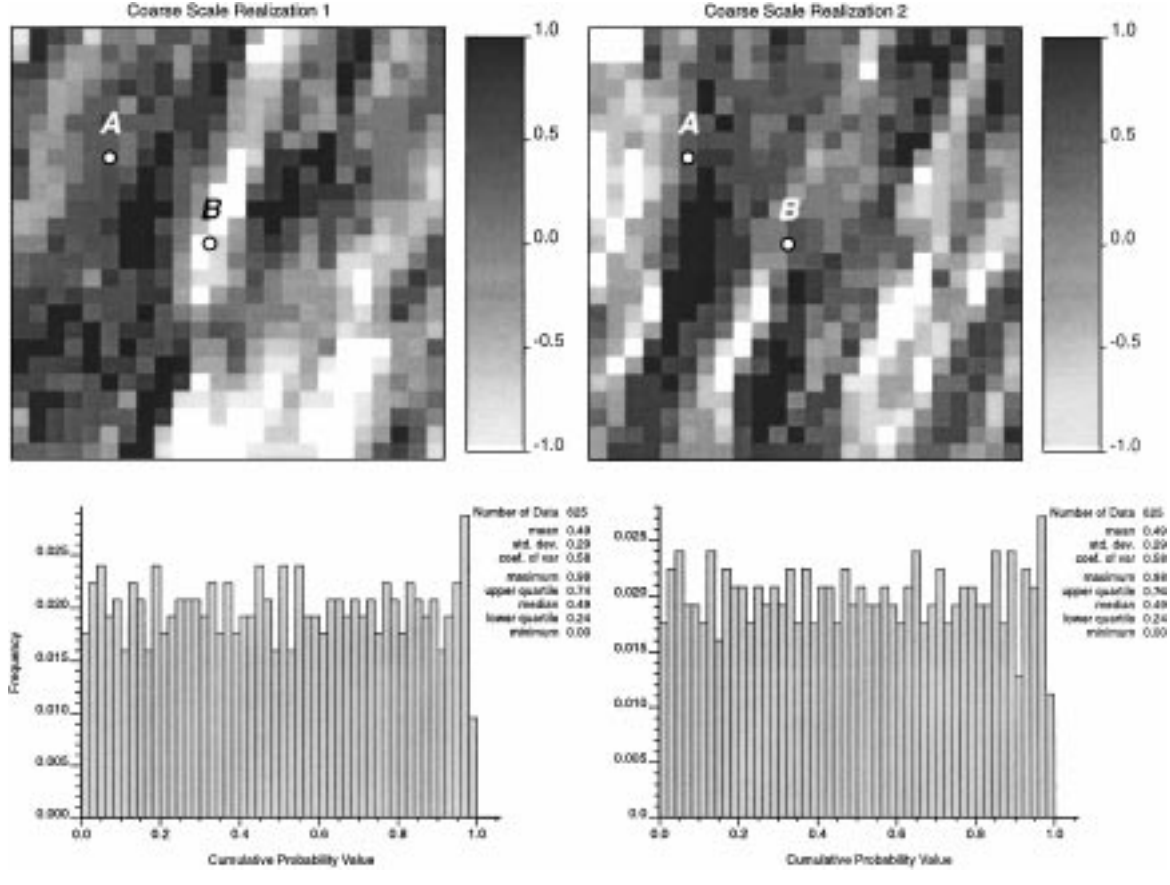

Figure 6. Two realizations at the same scale as the soft data shown in Figure 5. The histograms of the cumulative probability values, $F(\mathbf{u}_v; z_v^{(l)}(\mathbf{u}_v))$, are close to uniform between the bound values of 0 and 1.

where the objective function was initialized to 1 in the beginning. Histograms of $\mu^{(l)}(\mathbf{u}_v)$ corresponding to the two realizations are shown below each realization on Figure 6. The two high-permeability bands and the reduced permeability region have been reproduced within their specified reliability (refer back to the variance map on Fig. 5).

As described above, a key decision in this approach is to consider the distribution of $\mu^{(l)}(\mathbf{u}_v)$ over all locations of one realization instead of over all realizations. To check this assumption, 100 realizations were generated and the histograms at two locations (A) and (B) on Figure 5 were checked. Figure 7 shows the histograms of $\mu^{(l)}(\mathbf{u}_v)$ over the 100 realizations; note that the distributions are approximately uniform between the bounds of 0 and 1, as expected.

The most important novel aspect of the proposed approach is that the z variable can be simulated at a fine scale while imposing large-scale soft data constraints. Figure 8 shows two fine scale realizations that honor the coarse-scale soft data shown in Figure 5. A geometric average was considered to calculate the $z_v^{(l)}(\mathbf{u}_v)$ values for every coarse grid block; there are $4 \cdot 4 = 16$ fine-scale grid cells for each coarse-grid block. The small-scale z-variable target histogram and variogram in the three part objective function have greater variance than the coarse-scale input, because there is less volume averaging.

DISCUSSION

Conventional Gaussian and indicator-based geostatistical algorithms for data integration do not simultaneously account for the coarse-scale and variable precision of soft data. The simulated-annealing based objective function presented here is designed to account for such constraints. The objective function constrains the probability values $\mu^{(l)}(\mathbf{u}_v) = F(\mathbf{u}_v; z_v^{(l)}(\mathbf{u}_v))$ to follow a uniform distribution between 0 and 1. The block averages from realization l, $z_v^{(l)}(\mathbf{u}_v)$, must fall within the distribution predicted by the soft data $F(\mathbf{u}_v; z)$; sometimes on the low side $F(\mathbf{u}_v; z) < 0.5$ and sometimes on the high side $F(\mathbf{u}_v; z) > 0.5$, but, in expected value, uniformly between 0 and 1. This approach directly constrains the realization to the "information" contained in the soft data.

We expect the input soft distributions to come from remote sensing, seismic, or dynamic flow data. These data sources provide coarse-scale constraints with locally variable precision. The proposed method is no panacea; calibration of these soft data derived distributions and the inference of required small-scale statistics (histogram and variogram) remain important outstanding problems. When there are sufficient hard data (more than about 100), the calibration is reliable. In the presence of very few hard data it is essential to have a solid physical understanding of the soft data and how it relates to the variable under consideration. Of course, this is a problem encountered by all geostatistical methods.

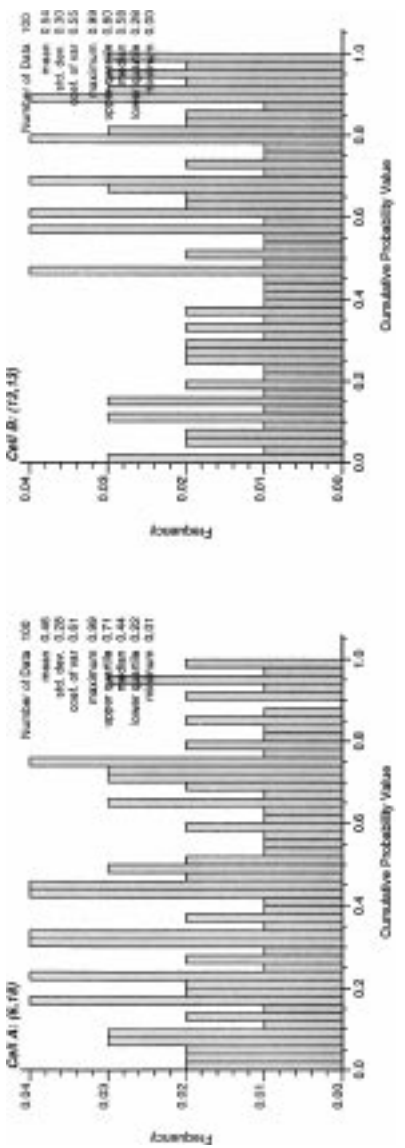

Figure 7. Histograms of the cumulative probability values, over 100 realizations, at locations A and B for the example shown in Figures 5 and 6.

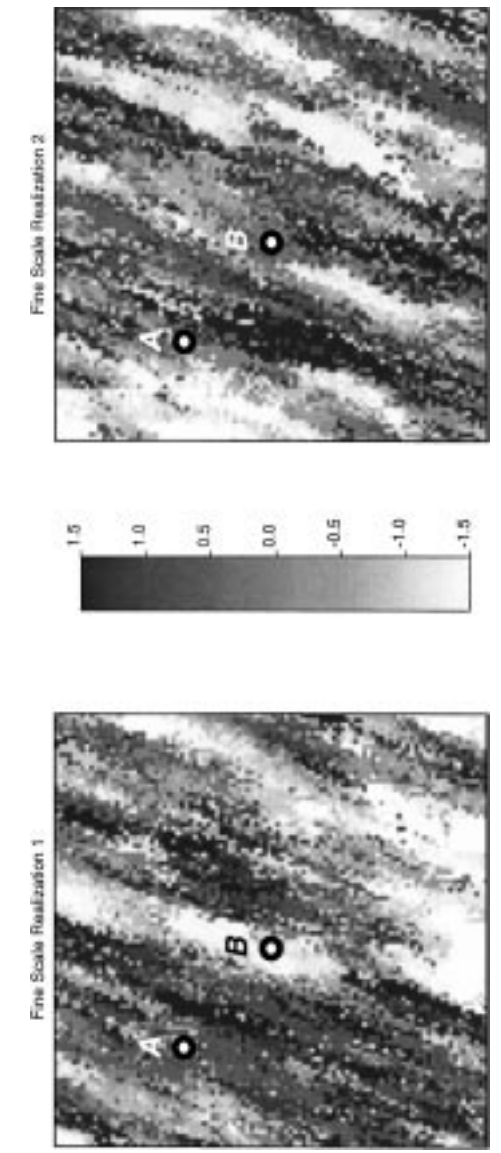

Figure 8. Two fine-scale realizations constructed with the large-scale probability constraints.

Implementation of the method is made practical because of the locally updatable objective function. In practice, there is no addictional CPU penalty to pay to for adding the soft data constraint to an annealing-based simulation program.

The methodology also permits nonlinear averaging. An ω-power averaging procedure was illustrated; however, any averaging law using, for example, a wavelet or kernel function derived from the physics could be considered.

REFERENCES

Alabert, F. G., 1989, Constraining description of randomly heterogeneous reservoirs to pressure test data: a Monte Carlo study, *in* 1989 SPE Annual Technical Conference and Exhibition Formation Evaluation and Reservoir Geology: San Antonio, TX, October 1996. Society of Petroleum Engineers.

Deutsch, C. V., 1992, Annealing techniques applied to reservoir modeling and the integration of geological and engineering (Well Test) Data, unpubl. doctoral dissertation, Stanford University, Stanford, CA, 306 p.

Deutsch, C. V., and Journel, A. G., 1997, GSLIB: Geostatistical software library and user's guide (2nd ed.): Oxford University Press, New York, 369 p.

Deutsch, C. V., Srinivasan, S., and Mo, Y., 1996, Geostatistical reservoir modeling accounting for the scale and precision of seismic data, *in* 1996 SPE Annual Technical Conference and Exhibition Formation Evaluation and Reservoir Geology: Denver, CO, October 1996. Society of Petroleum Engineers, SPE Paper Number 36497, p. 9–19.

Dubrule, O., 1989, Geostatistical reservoir characterization constrained by 3d seismic data: Petroleum Geoscience, v. 4, p. 121–128.

Farmer, C. L., 1992, Numerical rocks, *in* King, P. R., ed., The Mathematical generation of reservoir geology: Clarendon Press, Oxford, 1992 (Proceedings of a conference held at Robinson College, Cambridge, 1989).

Goovaerts, P., 1997, Geostatistics for natural resources evaluation: Oxford University Press, New York, 450 p.

Hird, K. B., 1993, A conditional simulation method for reservoir description using geological and well performance constraints: unpubl. doctoral dissertation, University of Tulsa, Tulsa, Oklahoma.

Isaaks, E. H., and Srivastava, R. M., 1989, An introduction to applied geostatistics: Oxford University Press, New York, 561 p.

Journel, A. G., and Huijbregts, C. J., 1978, Mining geostatistics: Academic Press, New York, 600 p.

Oliver, D. S., 1994, Incorporation of transient pressure data into reservoir characterization: In Situ, v. 18, no. 3, p. 243–275.

Ouenes, A., Bhagavan, S., Bunge, P. H., and Travis, B. J., 1994, Application of simulated annealing and other global optimization methods to reservoir description: Myths and realities, *in* SPE 69th Annual Conference and Exhibition, New Orleans, LA, Washington, DC, September 1994, Society of Petroleum Engineers, SPE paper number 28415, p. 547–561.

Perez, G., 1992, Stochastic conditional simulation for description of reservoir properties: unpubl. doctoral dissertation, University of Tulsa, Tulsa, OK.

Wen, X. H., Deutsch, C. V., and Cullick, A. S., 1997, High resolution reservoir models integrating multiple-well production data, *in* 1997 SPE Annual Technical Conference and Exhibition Formation Evaluation and Reservoir Geology, San Antonio, TX, October 1997. Society of Petroleum Engineers. SPE Paper Number 38728, p. 9–19.

Zhu, H., 1991, Modeling mixture of spatial distributions with integration of soft data: unpubl. doctoral dissertation, Stanford University, Stanford, CA.