Mathematical Geology, Vol. 30, No. 5, 1998

Accounting for Estimation Optimality Criteria in Simulated Annealing[1]

P. Goovaerts[2]

This paper presents both estimation and simulation as optimization problems that differ in the optimization criteria, minimization of a local expected loss for estimation and reproduction of global statistics (semivariogram, histogram) for simulation. An intermediate approach is proposed whereby an initial random image is gradually modified using simulated annealing so as to better match both local and global constraints. The relative weights of the different constraints in the objective function allow the user to strike a balance between smoothness of the estimated map and reproduction of spatial variability by simulated maps. The procedure is illustrated using a synthetic dataset. The proposed approach is shown to enhance the influence of observations on neighboring simulated values, hence the final realizations appear to be "better conditioned" to the sample information. It also produces maps that are more accurate (smaller prediction error) than stochastic simulation ignoring local constraints, but not as accurate as E-type estimation. Flow simulation results show that accounting for local constraints yields, on average, smaller errors in production forecast than a smooth estimated map or a simulated map that reproduces only the histogram and semivariogram. The approach thus reduces the risk associated with the use of a single realization for forecasting and planning.

KEY WORDS: estimation, stochastic simulation, loss function, flow characteristics, mean absolute error.

INTRODUCTION

Until the late 1980s, a typical geostatistical study proceeded in three steps: exploratory data analysis, modeling of the spatial variability (semivariogram) and, last, prediction of attribute values at unsampled locations. Most users are now aware that least-squares interpolation algorithms such as kriging tend to smooth out local details of the spatial variation of the attribute, with small values typically overestimated and large values underestimated. The use of smooth interpolated maps is inappropriate for applications sensitive to presence of ex-

[1]Received 30 June 1997; accepted 29 October 1997.

[2]Department of Civil and Environmental Engineering, The University of Michigan, Ann Arbor, Michigan 48109-2125. e-mail: goovaert@engin.umich.edu

0882-8121/98/0500-0511$15.00/1

treme values and their patterns of continuity, such as the evaluation of recoverable resources in mining deposits (Journel and Alabert, 1990; Nowak, Srivastava, and Sinclair, 1993), the modeling of fluid flow in porous media (Schafmeister and De Marsily, 1993), or the delineation of contaminated areas (Desbarats, 1996; Goovaerts, 1997a). Because it allows the generation of maps (realizations) that reproduce the sample variability, stochastic simulation is increasingly preferred to estimation.

One may generate many realizations that all match reasonably the same statistics (histogram, semivariogram); this contrasts with the uniqueness of an estimated map for a given optimality criterion. The set of alternative realizations is particularly useful to assess the uncertainty about the spatial distribution of attribute values, and investigate the performance of different scenarios of, e.g., mine planning or pollution remediation. Nevertheless, to reduce the computational cost stochastic modeling studies often use a single realization as a basis for forecasting and planning (Srivastava, 1996), a potentially hazardous option as illustrated for categorical variables by Goovaerts (1996). Another option consists of correcting *a posteriori* the unique estimated map to make it more variable.

To remove the smoothing of kriging, Olea and Pawlowsky (1996) proposed a two-step approach they called "compensated kriging": the relation between smoothed kriging estimates and data values is first modeled using cross-validation and linear regression, the model is then inverted to correct the estimates at unsampled locations. If some corrected values lie outside the admissible interval, the parameters of the linear model are modified using a trial-and-error procedure. A case study showed that compensated kriging has properties intermediate between ordinary kriging and conditional simulation in terms of mean square error and reproduction of histogram and semivariogram. An alternative would consist of postprocessing the estimates using the rank-preserving algorithm developed by Journel and Xu (1994) to improve the reproduction of a target histogram, in this case the sample histogram. This variant of rank-transform allows the honoring of data values without significant modification of the spatial pattern of the original values.

A common shortcoming of these correction algorithms is the lack of control on the reproduction of spatial statistics such as the semivariogram and on the preservation of desirable properties of kriging such as local accuracy. In this paper, both estimation and simulation are formulated as an optimization problem: their common objective is to generate map(s) that match a set of constraints specified by the user. For estimation, the constraints are *local* in that they involve each grid node separately, e.g., the minimization of a local error variance. In contrast, the constraints implemented in simulation algorithms, such as reproduction of histogram or semivariogram, are *global* because they involve all grid nodes jointly. We propose to combine the different local and global constraints

into a single objective function which is then used to postprocess an initial random image or an estimated map using simulated annealing. The focus is not on the construction of an hybrid between estimated and simulated maps, but rather on the minimization of the risk associated with the use of a single realization through the incorporation of local estimation constraints in simulated annealing.

The proposed technique is illustrated using a synthetic 2-D dataset. Permeability maps are generated conditional to 60 well locations using either estimation or simulation algorithms with different weighting schemes for local vs. global constraints. Prediction performance is assessed by comparing the reference permeability values with the estimated or simulated ones. The set of maps are also submitted to a waterflood flow simulation, and the production forecasts (water cuts and recovered oil) are compared to the values provided by the reference dataset.

OPTIMUM STOCHASTIC ESTIMATION

Consider the problem of estimating the value of a continuous attribute z at an unsampled location $\mathbf{u}$ conditional to the dataset $\{z(\mathbf{u}_\alpha),\ \alpha = 1, \ldots, n\}$. Estimation can be formulated as an optimization process; the objective is to select, within the range of possible z-values at $\mathbf{u}$, a single value, denoted $z^*(\mathbf{u})$, that is "optimal" for some criterion (Journel, 1989, p. 27–28; Christakos, 1992, p. 341–343). A possible criterion is the minimization of the impact attached to the estimation error $e(\mathbf{u}) = z(\mathbf{u}) - z^*(\mathbf{u})$ that is likely to occur. Such impact can be expressed as a function $L(\cdot)$ of that error, e.g., $L(e(\mathbf{u})) = [e(\mathbf{u})]^2$. Given that particular loss function, the estimate $z^*_L(\mathbf{u})$ should be chosen as to minimize the resulting loss $L(z(\mathbf{u}) - z^*(\mathbf{u}))$.

Because the actual value $z(\mathbf{u})$ is unknown, the actual loss $L(z(\mathbf{u}) - z^*(\mathbf{u}))$ cannot be computed in practice. However, the uncertainty about $z(\mathbf{u})$ can be modeled by the conditional cumulative distribution function (ccdf) of the random variable Z at $\mathbf{u}$:

$$F(\mathbf{u}; z|(n)) = \text{Prob}\{Z(\mathbf{u}) \leq z|(n)\} \tag{1}$$

where the notation "$|(n)$" expresses conditioning to the local information, say, $n(\mathbf{n})$ neighboring data $z(\mathbf{u}_\alpha)$. Ccdf values can be determined using multi-Gaussian or indicator algorithms (Deutsch and Journel, 1998, p. 76; Goovaerts, 1994a). The idea is then to use this model of uncertainty to determine the expected loss:

$$\begin{aligned} \varphi_L(z^*(\mathbf{u})|(n)) &= E\{L(Z(\mathbf{u}) - z^*(\mathbf{u}))|(n)\} \\ &= \int_{-\infty}^{+\infty} L(z - z^*(\mathbf{u}))\, dF(\mathbf{u}; z|(n)) \end{aligned} \tag{2}$$

The "L-optimal" estimate for the loss function $L(\cdot)$ is the z-value that minimizes the expected loss (2).

Unlike interpolation algorithms such as simple or ordinary kriging, the determination of an optimal estimate here proceeds in two steps:

1. the uncertainty about the unknown value $z(\mathbf{u})$ is first modeled by the conditional cdf $F(\mathbf{u}; z|(n))$,
2. from that model, an estimate $z_L^*(\mathbf{u})$ is deduced according to a specific optimality criterion.

This dichotomy between assessment of uncertainty and estimation makes clear that there is no best estimate for all situations. For a given model of uncertainty, different estimates can be obtained depending on the loss function chosen (Journel, 1984; Srivastava, 1987; Goovaerts, 1997b, p. 340–346).

Examples of Loss Functions

In this section, one briefly reviews three types of loss functions that allow a straightforward (analytical) determination of the optimal estimate. In absence of such an analytical solution, the expected loss can be computed for a series of z-values, and the one yielding the smallest expected loss is retained as the optimal estimate.

A common approach consists of modeling the loss as a quadratic function of the estimation error:

$$L(e(\mathbf{u})) = [e(\mathbf{u})]^2 \tag{3}$$

The optimal estimate is shown to be the expected value of the ccdf at location $\mathbf{u}$, also called E-type estimate:

$$z_L^*(\mathbf{u}) = z_E^*(\mathbf{u}) = \int_{-\infty}^{+\infty} z \, dF(\mathbf{u}; z|(n)) \tag{4}$$

The corresponding expected loss is but the variance of the conditional cdf:

$$\sigma^2(\mathbf{u}) = \int_{-\infty}^{+\infty} [z - z_E^*(\mathbf{u})]^2 \, dF(\mathbf{u}; z|(n)) \tag{5}$$

Instead of a quadratic function of the estimation error, one might consider a linear function such as:

$$L(e(\mathbf{u})) = \begin{cases} \omega_2 \cdot e(\mathbf{u}) & \text{for } e(\mathbf{u}) \geq 0 \text{ (underestimation)} \\ \omega_1 \cdot |e(\mathbf{u})| & \text{for } e(\mathbf{u}) < 0 \text{ (overestimation)} \end{cases} \tag{6}$$

where the nonnegative parameters ω_1 and ω_2 are the relative impacts attached to overestimation and underestimation, respectively. The optimal estimate is

then shown to be the p-quantile of the ccdf (Journel, 1984):

$$z_L^*(\mathbf{u}) = q_P(\mathbf{u}) = F^{-1}(\mathbf{u}; p|(n)) \quad \text{with}$$

$$p = \frac{\omega_2}{\omega_1 + \omega_2} \in [0, 1] \tag{7}$$

If $\omega_1 = \omega_2$, the loss function (6) is symmetric, which means that underestimation and overestimation are penalized equally as in expression (3). The optimal estimate is then the median $q_{0.5}(\mathbf{u})$. If $\omega_1 \neq \omega_2$, both sign and magnitude of the estimation error are accounted for. For example, if $\omega_1 > \omega_2$, the impact of overestimation is larger than that of underestimation of the same magnitude. Thus, $p < 0.5$, and the optimal estimate is smaller than the median. The expected loss associated with the optimal estimate is:

$$\varphi_L(q_p(\mathbf{u})|(n)) = \int_{-\infty}^{+\infty} |z - q_p(\mathbf{u})| \, dF(\mathbf{u}; z|(n))$$

$$= z_E^*(\mathbf{u}) - m(q_p(\mathbf{u})) \tag{8}$$

where $m(q_p(\mathbf{u}))$ is the mean of the truncated distribution:

$$m(q_p(\mathbf{u})) = \frac{1}{F(\mathbf{u}; q_p(\mathbf{u})|(n))} \int_{-\infty}^{q_p(\mathbf{u})} z \, dF(\mathbf{u}; z|(n))$$

A third type of loss model is the indicator function:

$$L(e(\mathbf{u})) = \begin{cases} 0 & \text{if } e(\mathbf{u}) = 0 \\ 1 & \text{otherwise} \end{cases} \tag{9}$$

The optimal estimate is the most plausible outcome of the conditional distribution, that is the mode of the corresponding density function $f(\mathbf{u}; z|(\mathbf{n}))$. The expected loss is the probability of nonoccurrence of this estimate:

$$\varphi_L(z_L^*(\mathbf{u})|(n)) = 1 - dF(\mathbf{u}; z_L^*(\mathbf{u})|(n)) \tag{10}$$

Differences between estimates arise from both the ccdf model and the optimality criterion. For symmetric conditional distributions, the minimization of the second criterion (6) with $p = 0.5$ yields the same estimate as the first criterion (3); the mean and median are the same. If the distribution is Gaussian, the three types of loss functions yield the same estimate; the mean, median and mode of the ccdf are identical.

Map of Optimal Estimates

Estimation rarely concerns a single location $\mathbf{u}$. Most often, the estimation is performed at N grid nodes $\mathbf{u}_j'$ discretizing a study area $\mathcal{A}$. For a given loss

function $L(\cdot)$, the selection of an optimal estimate at $\mathbf{u}$ is independent of the estimation at other grid nodes. Thus, the unique solution to the optimization problem is the set of N "locally optimal" estimates $\{z_L^*(\mathbf{u}_j'), j = 1, \ldots, N\}$, denoted $\mathbb{S}_L^*$. The global expected loss associated with the "L-optimal" estimation grid is the sum of local expected losses:

$$\varphi_L(\mathbb{S}_L^*|(n)) = \sum_{j=1}^{N} \varphi_L(z_L^*(\mathbf{u}_j')|(n)) \tag{11}$$

For example, for the loss function (3), the global loss (11) is the sum of the variances of the N ccdfs $F(\mathbf{u}_j'; z|(n))$. The uniqueness of the "L-optimal" estimation grid entails that any other set of estimates, $\mathbb{S}^* = \{z^*(\mathbf{u}_j'), j = 1, \ldots, N\}$, with $\mathbb{S}^* \neq \mathbb{S}_L^*$, yields a larger global expected loss:

$$\varphi_L(\mathbb{S}^*|(n)) = \sum_{j=1}^{N} \varphi_L(z^*(\mathbf{u}_j')|(n)) > \varphi_L(\mathbb{S}_L^*|(n))$$

STOCHASTIC SIMULATION

Consider now the simulation of the continuous attribute z at N grid nodes $\mathbf{u}_j'$ conditional to the dataset $\{z(\mathbf{u}_\alpha), \alpha = 1, \ldots, n\}$. The creation of a stochastic image can also be formulated as an optimization problem; the objective is to generate a set of z-values $\{z^{(l)}(\mathbf{u}_j'), j = 1, \ldots, N\}$ which match approximately constraints such as reproduction of a target histogram or semivariogram model. For example, if the objective is to reproduce the semivariogram model $\gamma(\mathbf{h})$ over the first S lags, the set of simulated values should be such as to lower the objective function to a value close to zero:

$$O = \sum_{s=1}^{S} [\gamma(\mathbf{h}_s) - \hat{\gamma}(\mathbf{h}_s)]^2 \tag{12}$$

where $\gamma(\mathbf{h}_s)$ is the value of the target semivariogram model at lag $\mathbf{h}_s$, and $\hat{\gamma}(\mathbf{h}_s)$ is the corresponding experimental semivariogram value of the realization $\{z^{(l)}(\mathbf{u}_j'), j = 1, \ldots, N\}$.

Whereas the loss functions introduced for estimation involve each grid node $\mathbf{u}_j'$ separately, the objective function (12) involve many grid nodes simultaneously because the computation of the semivariogram values $\hat{\gamma}(\mathbf{h}_s)$ calls for many different pairs of z-values. Thus, an optimum cannot be reached if the simulated values are derived independently from one another. A second difference with estimation is that there are usually many solutions (i.e., sets of simulated values) to the optimization problem, which contrasts with the uniqueness of the "L-optimal" estimation grid. Consequently, the minimization of an objective function of type (12) is not as straightforward as the minimization of local loss functions, and requires iterative algorithms such as simulated annealing.

Simulated Annealing

Simulated annealing is a generic name for a family of optimization algorithms based on the principle of stochastic relaxation (Geman and Geman, 1984; Farmer, 1988). Once the objective function has been established, the optimization process amounts to systematically modifying an initial image or realization so as to decrease the value of that objective function, getting the realization acceptably close to the target statistics.

There are many possible implementations of simulated annealing, depending on the way the initial realization is generated and then perturbed, on the components that enter the objective function, and on the type of decision rule and convergence criterion that are adopted for the iterative algorithm (Deutsch and Cockerham, 1994). The following procedure is here used to generate realizations that reproduce both the histogram and semivariogram model:

1. Generate an initial realization $\{z^{(l)}_{(0)}(\mathbf{u}'_j), j = 1, \ldots, N\}$ by freezing data values at their locations and assigning to each unsampled grid node a z-value drawn at random from the target cdf $F(z)$. This approach is fast and yields a set of initial images that already honor the conditioning data and match the target histogram.
2. Compute the initial value of the objective function corresponding to that initial realization:

$$O(0) = \sum_{s=1}^{S} \frac{[\gamma(\mathbf{h}_s) - \hat{\gamma}_{(0)}(\mathbf{h}_s)]^2}{[\gamma(\mathbf{h}_s)]^2}$$

   where $\hat{\gamma}_{(0)}(\mathbf{h}_s)$ is the semivariogram value at lag $\mathbf{h}_s$ of the initial realization. The division by the square of the semivariogram model at each lag $\mathbf{h}_s$ gives more weight to reproduction of the semivariogram model near the origin.
3. Perturb the realization by swapping z-values at any two unsampled locations $\mathbf{u}'_j$ and $\mathbf{u}'_k$ chosen at random: $z^{(l)}_{(0)}(\mathbf{u}'_j)$ becomes $z^{(l)}_{(1)}(\mathbf{u}'_k)$ and *vice versa*. Assess the impact of the perturbation on the reproduction of target statistics by recomputing the objective function, $O_{\text{new}}(0)$, accounting for the modification of the initial image.
4. Accept all perturbations that diminish the objective function. Unfavorable perturbations are accepted according to a negative exponential probability distribution:

$$\text{Prob}\{\text{Accept } i\text{th perturbation}\} = \begin{cases} 1 & \text{if } O(i) \leq O(i-1) \\ \exp \dfrac{[O(i-1) - O(i)]}{t(i)} & \text{otherwise} \end{cases}$$

The idea is to start with an initially high temperature $t(0)$, which allows a large proportion of unfavorable perturbations to be accepted at the beginning of the simulation. As the simulation proceeds, the temperature is gradually lowered so as to limit discontinuous modification of the stochastic image. Two important issues are the timing and magnitude of the temperature reduction, which defines the *annealing schedule.* According to Deutsch and Cockerham's typology, a fast annealing schedule was used; that is, the initial temperature was set to 1 and lowered by a factor 20 (reduction factor = 0.05) whenever enough perturbations ($5 \times N$) have been accepted or too many ($50 \times N$) have been tried.

5. If the perturbation is accepted, update the initial realization into a new image $\{z^{(l)}_{(1)}(\mathbf{u}'_j), j = 1, \ldots, N\}$ with objective function value $O(1) = O_{\text{new}}(0)$.
6. Repeat steps 3 to 5 until either the target low value $O_{\min} = 0.001$ is reached or the maximum number of attempted perturbations at the same temperature has been reached three times.

Other realizations $\{z^{(l')}(\mathbf{u}'_j), j = 1, \ldots, N\}$, $l' \neq l$, are generated by repeating the entire process starting from different initial realizations. Typically, the number of nodes N is so large and the semivariogram is so little constraining that there exist many solutions to the optimization problem.

AN INTERMEDIATE APPROACH

As mentioned previously, for a given loss function $L(\cdot)$, estimation at the N grid nodes $\mathbf{u}'_j$ can be viewed as the selection of N values, $\mathcal{S}^* = \{z^*(\mathbf{u}'_j), j = 1, \ldots, N\}$, that yield the smallest global expected loss:

$$\varphi_L(\mathcal{S}^*|(n)) = \sum_{j=1}^{N} \varphi_L(z^*(\mathbf{u}'_j)|(n)) \tag{13}$$

For example, for the loss function (3), the optimal estimation grid is but the set of E-type estimates $\{z^*_E(\mathbf{u}'_j), j = 1, \ldots, N\}$ as defined in (4). The same solution can be reached by processing any initial random image[3] $\{z_{(0)}(\mathbf{u}'_j), j = 1, \ldots, N\}$ using simulated annealing with the following objective function to be lowered:

$$O(i) = \sum_{j=1}^{N} \varphi_L(z_{(i)}(\mathbf{u}'_j)|(n)) \tag{14}$$

[3]To simplify notation, the superscript (l), which refers to a particular realization, is omitted in this section.

where $z_{(i)}(\mathbf{u}'_j)$ is the z-value at node $\mathbf{u}'_j$ at the ith perturbation. The perturbation mechanism amounts to selecting randomly a single location $\mathbf{u}'_j$ and replacing the corresponding z-value $z_{(i-1)}(\mathbf{u}'_j)$ by a new value $z_{(i)}(\mathbf{u}'_j)$, say a value randomly drawn from the marginal cdf $F(z)$.

Consider the situation where the L-optimal estimate and the corresponding expected loss are known *a priori*, e.g., the mean and variance of the ccdf for the loss function (3). Thus, minimizing the objective function (14) amounts to lowering to zero the following objective function:

$$O(i) = \sum_{j=1}^{N} |\varphi_L(z_L^*(\mathbf{u}'_j)|(n)) - \varphi_L(z_{(i)}(\mathbf{u}'_j)|(n))| \tag{15}$$

One retrieves the usual formulation of an annealing-type objective function as the difference between target and actual statistics of the image.

Formulation of the estimation process as the minimization of an objective function of type (15) allows a straightforward incorporation of estimation optimality criteria in stochastic simulation. Consider the problem of generating a realization that reproduces a target cdf $F(z)$ and a semivariogram $\gamma(\mathbf{h})$ while keeping the desirable features of estimation such as the minimization of a local loss function. The proposed approach consists of processing an initial random image that already matches the target cdf using simulated annealing and a two-components objective function:

$$O(i) = \frac{\lambda_1}{O_1(0)} O_1(i) + \frac{\lambda_2}{O_2(0)} O_2(i) \tag{16}$$

where:

$$O_1(i) = \sum_{j=1}^{N} |\varphi_L(z_L^*(\mathbf{u}'_j)|(n)) - \varphi_L(z_{(i)}(\mathbf{u}'_j)|(n))|$$

$$O_2(i) = \sum_{s=1}^{S} \frac{[\gamma(\mathbf{h}_s) - \hat{\gamma}_{(i)}(\mathbf{h}_s)]^2}{[\gamma(\mathbf{h}_s)]^2}$$

$$\sum_{k=1}^{2} \lambda_k = 1, \qquad \lambda_k \geq 0$$

To prevent the component with the largest unit from dominating the objective function, each component O_k is standardized by its initial value $O_k(0)$. The relative importance of each component is controlled by the weights λ_k that sum to 1, which allows the user to strike a balance between a local criterion (minimization of a local expected loss) and a global criterion (reproduction of a target semivariogram).

Rescaling of E-Type Estimates

This new formulation allows an interesting interpretation of the rescaling of E-type estimates using the correction algorithm proposed by Journel and Xu (1994). Consider the problem of generating a realization that reproduces a target cdf $F(z)$ while minimizing the global expected loss for the quadratic loss function $[e(\mathbf{u})]^2$. This can be done by processing an initial random image that already matches the target cdf using simulated annealing and the objective function (16) with $\lambda_1 = 1$ and $\lambda_2 = 0$.

An alternative consists of starting with the set of E-type estimates $\mathcal{S}_L^*$ which minimizes the global expected loss and transforming (rescaling) these estimates so as to reproduce the target cdf while keeping the global expected loss as small as possible. A straightforward way to perform this *a posteriori* correction of optimal estimates is to apply the rank-preserving transform:

$$z_c^*(\mathbf{u}_j') = F^{-1}[F_L(z_L^*(\mathbf{u}_j'))] \qquad j = 1, \ldots, N \tag{17}$$

where $z_c^*(\mathbf{u}_j')$ is the value corrected from the original estimate $z_L^*(\mathbf{u}_j') = z_E^*(\mathbf{u}_j')$, and $F_L(\cdot)$ is the cdf of the N estimates. One can show that the set of corrected values $\mathcal{S}_c^*$ is the unique optimum for the joint constraints of histogram reproduction and minimization of global expected loss. Let $d(\mathbf{u}_j')$ be the difference between corrected and initial values at $\mathbf{u}_j'$, $z_c^*(\mathbf{u}_j') - z_E^*(\mathbf{u}_j')$. Accounting for definitions (2)–(5), the expected loss associated with the corrected value $z_c^*(\mathbf{u}_j')$ is:

$$\begin{aligned}
\varphi_L(z_c^*(\mathbf{u}_j')|(n)) &= \int_{-\infty}^{+\infty} L(z - z_c^*(\mathbf{u}_j'))\, dF(\mathbf{u}_j'; z|(n)) \\
&= \int_{-\infty}^{+\infty} [z - z_E^*(\mathbf{u}_j') - d(\mathbf{u}_j')]^2\, dF(\mathbf{u}_j'; z|(n)) \\
&= \sigma^2(\mathbf{u}_j') + [d(\mathbf{u}_j')]^2 - 2d(\mathbf{u}_j') \\
&\quad \cdot \int_{-\infty}^{+\infty} [z - z_E^*(\mathbf{u}_j')]\, dF(\mathbf{u}_j'; z|(n)) \\
&= \sigma^2(\mathbf{u}_j') + [d(\mathbf{u}_j')]^2
\end{aligned}$$

The global expected loss computed over the N nodes $\mathbf{u}_j'$ is thus:

$$\sum_{j=1}^{N} \varphi_L(z_c^*(\mathbf{u}_j')|(n)) = \sum_{j=1}^{N} \sigma^2(\mathbf{u}_j') + \sum_{j=1}^{N} [d(\mathbf{u}_j')]^2 \tag{18}$$

Minimizing the global loss (18) amounts to minimizing the sum of squares of differences between corrected and initial values, which is ensured by the rank-preserving transform (17).

CASE STUDY

Throughout this paper, the map shown at the top of Figure 1 (left graph) is considered as the reference exhaustive distribution of permeability in a 2-D section of a reservoir (Goovaerts, 1996). The reference dataset comprises 50 × 50 permeability values on a regular square grid. The corresponding histogram is positively skewed and indicates the presence of a few large values. Sixty locations were drawn at random and form the sample dataset available to reconstruct the reference image. Because of the sparsity of sampling, the sample mean and variance deviate from the reference statistics (Fig. 1 middle graph).

Generation of Permeability Maps

Fifty initial realizations of the spatial distribution of permeability values were generated by freezing data values at their locations and assigning to each unsampled grid node a permeability value drawn at random from the sample histogram. Each of these realizations was postprocessed using simulated annealing and the two-components objective function (16). The local expected loss in the first component O_1 was computed using the quadratic function (3) and ccdf models provided by ordinary indicator kriging; ccdf values were estimated for nine thresholds corresponding to the deciles of the sample histogram, and the resolution of the discrete ccdf was increased by performing a linear interpolation between tabulated bounds provided by the sample cdf (Deutsch and Journel, 1998, p. 134–138). Reproduction of semivariograms of highly skewed variables generally entails edge effects in the final realization, that is, extreme values tend to be pushed to the edges because they contribute only once to the semivariogram calculation (Deutsch and Cockerham, 1994). To avoid such artifacts, the semivariogram of logarithms shown at the bottom of Figure 1 was used as target in the second component. A geometric anisotropy model was fitted with a slightly smaller range in the E–W direction (anisotropy ratio = 0.77).

To investigate the relative influence of both components on the final realizations, five different sets of weights were considered. Figure 2 shows, for each weighting scheme, the first realization generated. E-type estimates were also computed from the ccdf models using the following approximation:

$$z_E^*(\mathbf{u}) \simeq \sum_{k=1}^{K+1} \bar{z}_k \cdot [F(\mathbf{u}; z_k|(n)) - F(\mathbf{u}; z_{k-1}|(n))] \tag{19}$$

where z_k, $k = 1, \ldots, K$, are K threshold values discretizing the range of variation of z-values. By convention, $F(\mathbf{u}; z_0|(n)) = 0$ and $F(\mathbf{u}; z_{K+1}|(n)) = 1$. Other thresholds z_k were identified to p-quantiles corresponding to regularly spaced ccdf increments, i.e., $z_k = F^{-1}(\mathbf{u}; k/[K+1]|(n))$. $\bar{z}_k$ is the mean of the

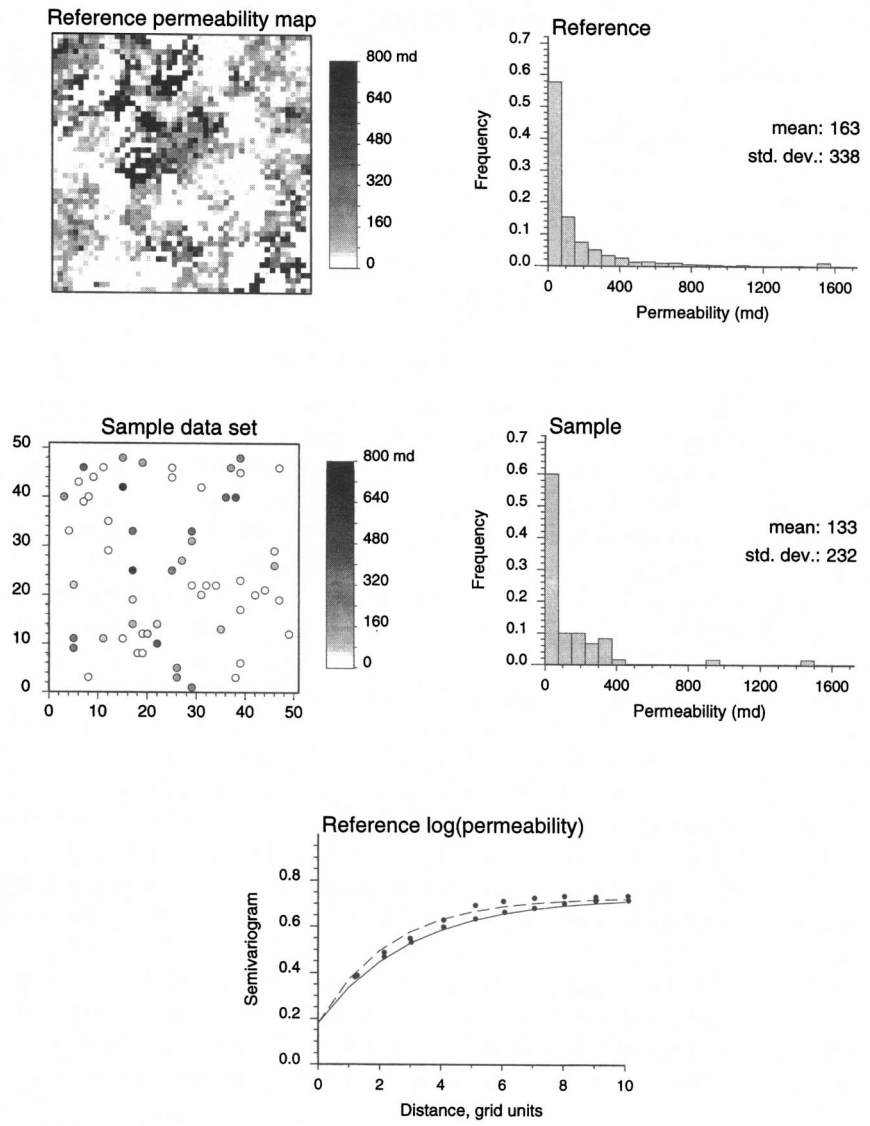

Figure 1. Reference permeability map with the corresponding histogram and semivariogram of logarithms computed in two directions: N–S (solid line) and E–W (dashed line). The information available consists of 60 randomly drawn values and the sample histogram.

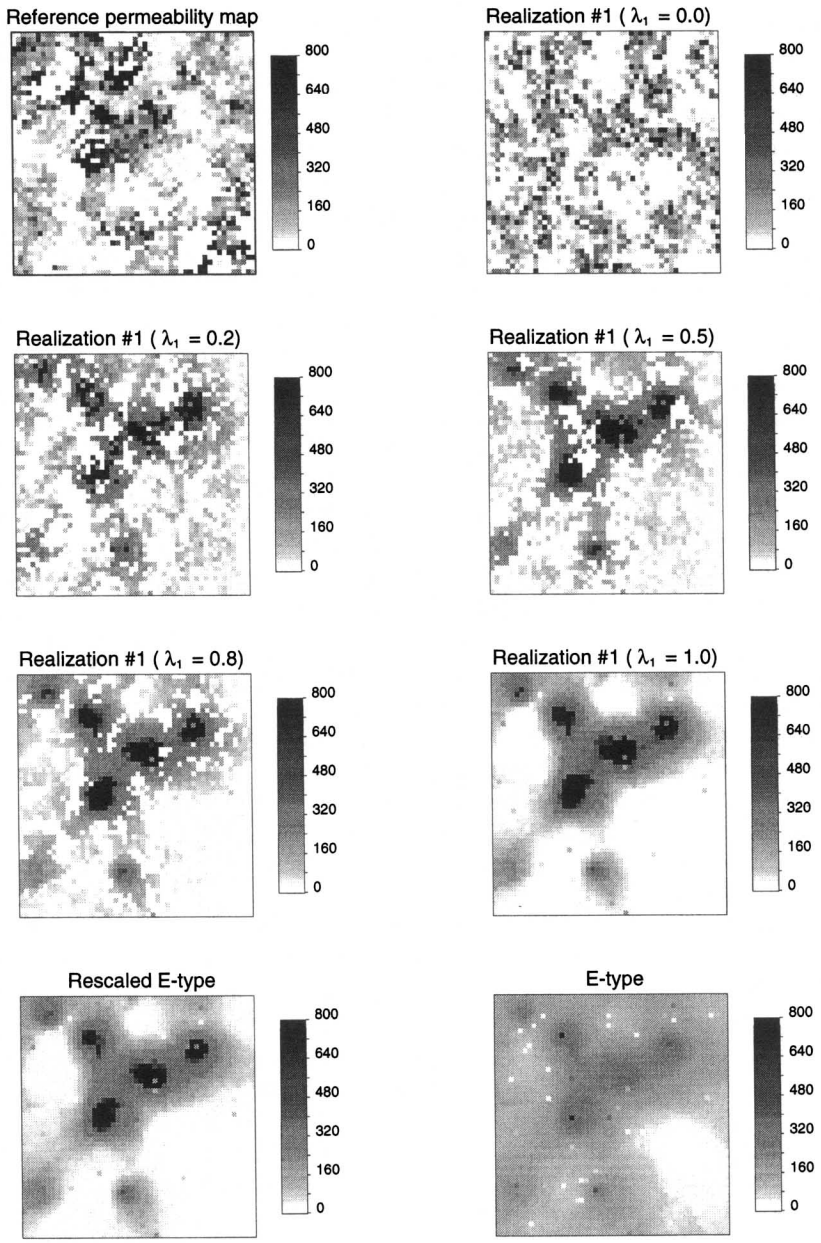

Figure 2. The reference permeability map, the first realization generated using an increasing weight λ_1 for the component that controls the minimization of the local expected loss, and the maps of rescaled and original E-type estimates.

class $(z_{k-1}, z_k]$ which depends on the intraclass interpolation model, e.g., for the linear model: $\bar{z}_k = (z_{k-1} + z_k)/2$. To correct for the smoothing effect, the E-type estimates were transformed (rescaled) using relation (17) which ensures reproduction of the sample cdf. Figure 2 (bottom graphs) shows the maps of E-type estimates after and before rescaling.

Impact of the Weighting Scheme

The semivariogram of logarithms was computed in the E–W and N–S directions for each of the map of simulated or estimated values. For each weighting scheme, the 50 semivariograms were averaged and displayed with the target model in Figure 3. When all the weight is given to the reproduction of semivariogram ($\lambda_1 = 0$), the spatial pattern of the reference image is not well reproduced by the first realization, in particular the clustering of high values is not apparent. A visual improvement is obtained by giving a small weight to the minimization of local expected loss ($\lambda_1 = 0.2$). In fact, the incorporation of local constraints based on ccdfs enhances the influence of each observation on the neighboring simulated values, hence the final realization appears to be "better conditioned" to the sample information. As the weight given to the first component increases, the realization becomes smoother while deviation from the semivariogram model increases. When all the weight is given to the minimization of local expected loss ($\lambda_1 = 1$), the realization is very close to the smooth map of rescaled E-type estimates (Fig. 2, left bottom graph), which has been shown to be the optimum for the joint constraints of histogram reproduction and minimization of local expected loss, recall previous section. As expected, the semivariogram model is not reproduced anymore. Figure 2 (right bottom graph) shows the map of E-type estimates, which is the optimum for the first component (i.e., minimum expected loss) but reproduces neither the histogram nor the semivariogram.

The set of 50 alternative realizations generated by simulated annealing provides a measure of uncertainty about the spatial distribution of permeability values. Local differences between realizations can be depicted by mapping some measure of the spread of the distribution of 50 simulated values at each grid node. Figure 4 shows, for each weighting scheme, the maps of the standard deviation of the distribution of the 50 simulated values at each location. White pixels correspond to single-valued distributions at the sixty data locations. Giving more weight to the minimization of expected loss reduces differences between realizations which become more similar to the unique optimum, that is the map of rescaled E-type estimates.

Prediction Performances

Table 1 (2nd column) gives the global expected loss obtained on average over 50 realizations for the different weighting schemes. As expected, the loss

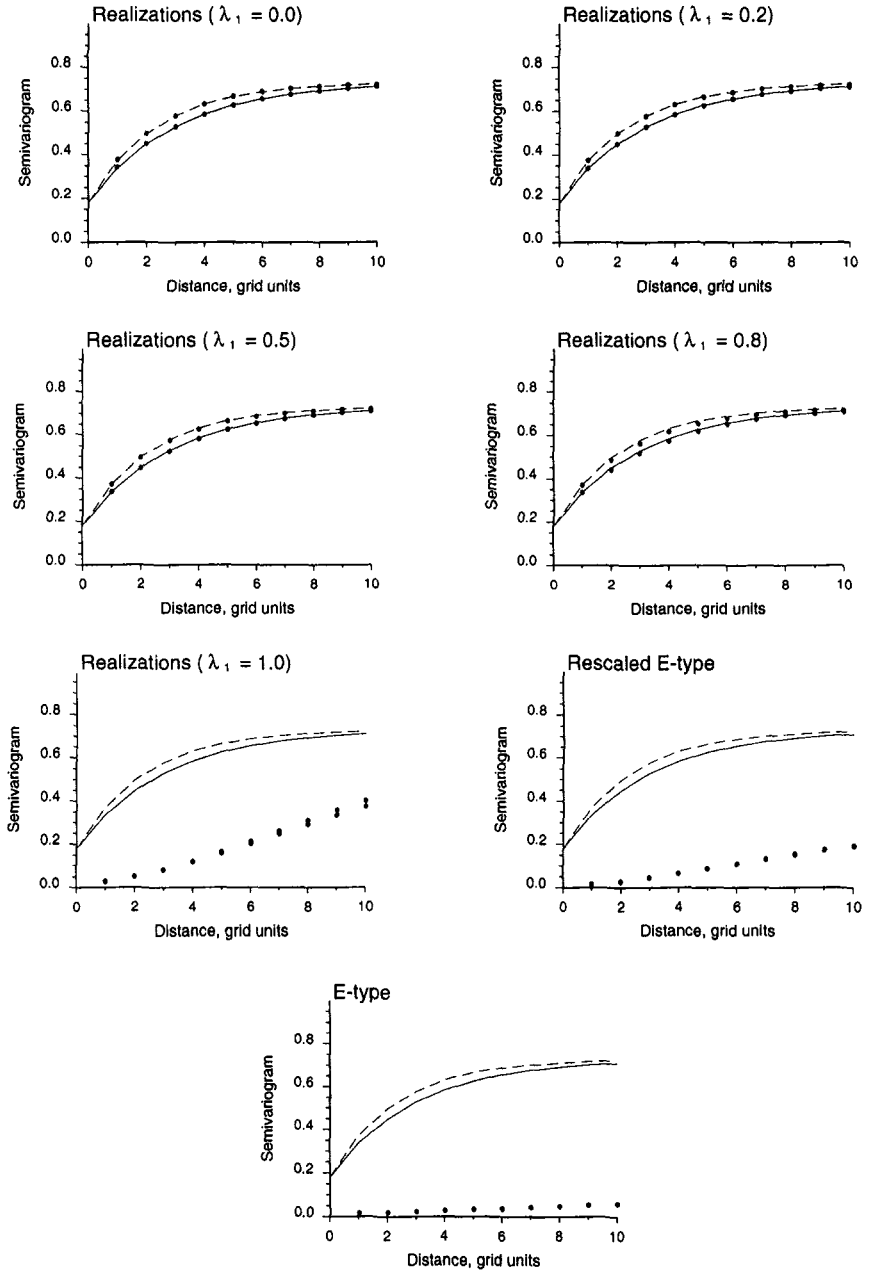

Figure 3. Reproduction of the anisotropic semivariogram model of logarithms obtained on average over 50 realizations generated using an increasing weight λ_1 for the local constraint, and for the maps of rescaled and original E-type estimates.

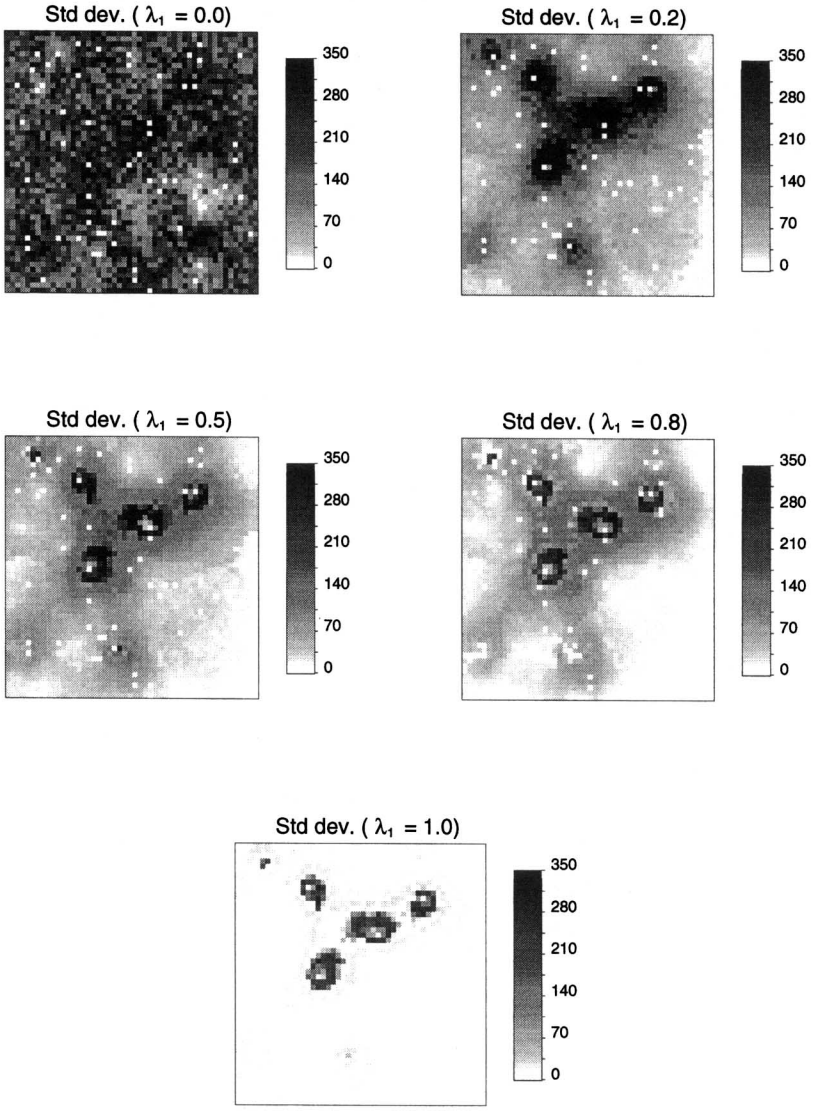

Figure 4. Maps of the standard deviations of the local distributions of permeability values generated using simulated annealing with an increasing weight λ_1 for the local constraint.

Table 1. Statistics Measuring the Prediction Performances Obtained on Average over 50 Realizations Generated Using an Increasing Weight λ_1 for the Local Constraint, and for the Rescaled and Original E-Type Estimates. For the Global Expected Loss and the Mean Square Error of Prediction, Results Are Expressed as Percentage of the Simulation Score When Local Constraints Are Ignored ($\lambda_1 = 0$)

			Aver. [Ref. values–predicted values]			
Algorithm	Loss (%)	MSE (%)	$\overline{K}^x_{\text{eff}}$ (md)	$\overline{K}^y_{\text{eff}}$ (md)	Water (year)	Oil (year)
Simulation						
$\lambda_1 = 0.0$	100	100	4.68	1.59	2.22	3.19
$\lambda_1 = 0.2$	87	88	2.26	2.22	1.38	3.01
$\lambda_1 = 0.5$	84	86	2.95	2.08	1.28	2.01
$\lambda_1 = 0.8$	83	85	4.81	2.18	0.78	0.83
$\lambda_1 = 1.0$	81	83	20.1	8.86	1.59	1.05
Rescaled E-type	74	80	36.1	28.6	1.49	1.40
E-type	41	69	75.7	73.9	1.65	2.67

decreases as the weight λ_1 given to the first component increases. When all the weight is given to the minimization of local expected loss ($\lambda_1 = 1$), the loss is 81%, which is larger than the score of the rescaled E-type estimates (74%). Consequently, if the objective is to minimize the local expected loss while reproducing a target histogram, better results are obtained by rescaling E-type estimates using a transform of type (17). This unique optimum cannot be reached by simulated annealing and the realizations appear to be trapped in suboptimal situations. In the absence of constraint of histogram reproduction, the optimum is the map of E-type estimates which yields a global loss of 41%.

Each generated map was compared with the reference permeability map, and the mean square error of prediction was computed as:

$$\text{MSE} = \frac{1}{N} \sum_{j=1}^{N} (z(\mathbf{u}'_j) - z^{(l)}(\mathbf{u}'_j))^2$$

Table 1 (3rd column) gives the error obtained on average over 50 realizations for the different weighting schemes. Prediction errors were also computed for the E-type estimates before and after rescaling. The smallest errors are obtained using an estimation approach (E-type) that seeks only the minimization of local expected loss. When the constraint of histogram reproduction is included, the prediction error increases. As for the global expected loss in the second column, better results are obtained by rescaling E-type estimates (MSE = 80%) instead of processing an initial random image using simulated annealing and a weight $\lambda_1 = 1$ (MSE = 83%). When a third constraint of semivariogram reproduction is included, the prediction error increases even more. In other words, reproduc-

tion of spatial variability as modeled by semivariogram is achieved at the expense of larger errors of prediction. Giving a small weight to the first component ($\lambda_1 = 0.2$), however, suffices to reduce significantly the mean square error while ensuring a fairly good reproduction of the semivariogram (Fig. 3, right top graph).

Flow Properties

The effective permeability of each map was computed in the E–W and N–S directions using the pressure solver *flowsim* (Deutsch and Journel, 1992). The reference values are $\overline{K}^x_{\text{eff}} = 38.6$ md and $\overline{K}^y_{\text{eff}} = 39.3$ md. The absolute differences between the reference and predicted values were computed in average over the 50 realizations (different weighting schemes), and for the maps of rescaled and original E-type estimates (see Table 1, 4th and 5th columns). Worst results are obtained for the smooth map of E-type estimates which severely overestimates the effective permeability. Rescaling of E-type estimates significantly reduces prediction error, but the best results are obtained when semivariogram reproduction is accounted for. Along the N–S direction, the minimization of the local expected loss ($\lambda_1 = 0.2$ or 0.5) improves the prediction performances over the "classical" simulation approach that ignores local constraints ($\lambda_1 = 0.0$). To check whether this improvement is stable over the 50 realizations, prediction errors were compared for each of the 50 pairs of realizations that were generated by postprocessing the same initial random images using simulated annealing and weights $\lambda_1 = 0.0$ or 0.5. Results are displayed on the scattergram at the top of Figure 5. Forty-three out of 50 black dots are below the 45° line, which means that accounting for local constraints yields smaller prediction errors in 86% of cases.

A waterflood simulator (Eclipse, 1991) was applied to the reference permeability map using the five spot injection/production pattern shown in Figure 6. The fractional flow of water and the proportion of oil recovered were computed for different time steps. Ten reference time values were retrieved for fractional flows of water of 5%, 15%, . . . , and 95%. Five reference time values were also retrieved for proportions of oil recovered of 10%, 20%, . . . , and 50%. The absolute differences between these reference values and predicted values were computed in average over the 50 realizations (different weighting schemes), and for the maps of rescaled and original E-type estimates (see Table 1, last two columns). Best results are obtained when both the reproduction of the semivariogram and the minimization of local expected loss are incorporated into the objective function. When one of these two components is ignored, prediction error increases. As for the N–S effective permeability, results are quite stable over all realizations. For example, scattergrams at the bottom of Figure 5 indicate that accounting for local constraints ($\lambda_1 = 0.8$) yields smaller prediction errors for 84% of the realizations.

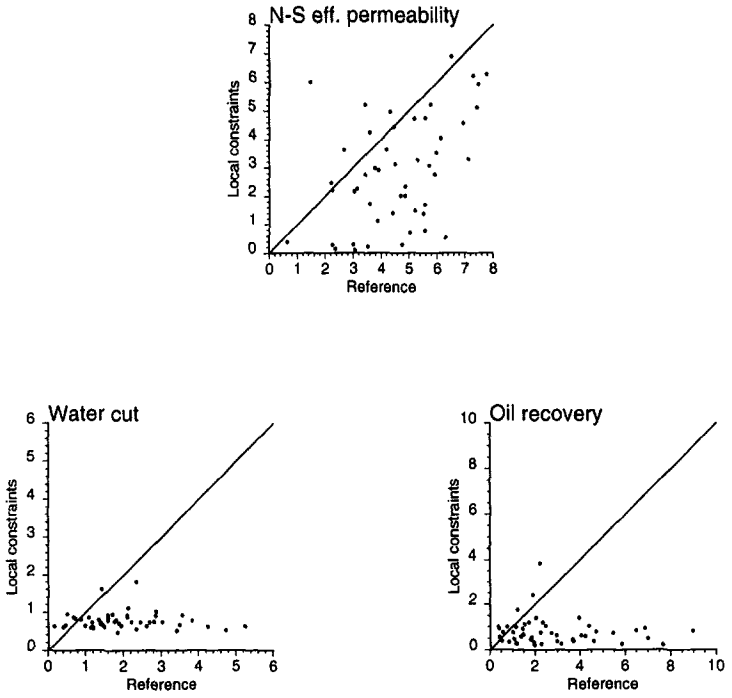

Figure 5. Scattergrams of prediction errors for realizations generated using the reference approach ($\lambda_l = 0.0$) vs. realizations that account for local constraints ($\lambda_l = 0.5$ or 0.8).

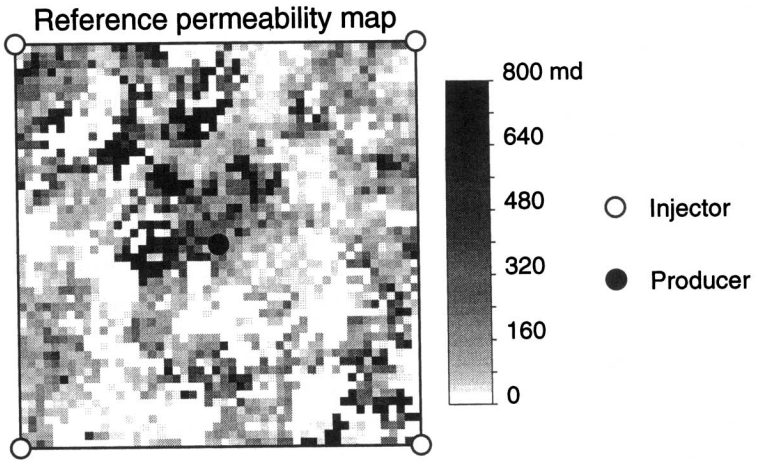

Figure 6. Locations of the producer (black dot) and the four injectors (open circles) on the reference permeability map.

Figures 7 and 8 show histograms of N–S effective permeabilities and three flow characteristics for 50 realizations generated using three different sets of weights. The black dot in the box plot below each histogram is the true value obtained from the reference image, and the three vertical lines are the 0.025 quantile, the median and the 0.975 quantile of the output distribution. A good simulation algorithm should generate an output distribution that is both accurate and precise. An output distribution is accurate if some fixed probability interval, e.g., the 95% probability interval, contains the true response. The precision of the output distribution is measured by its variance, the larger the spread of that distribution the larger the uncertainty of any prediction. Accounting for local constraints generally reduces the space of uncertainty, which agrees with previous results obtained for categorical variables (Goovaerts, 1994b) and with the variance maps of Figure 4. Except for effective permeability, the reduction is drastic when a large weight $\lambda_1 = 0.8$ is given to local constraints. It is worth noting that most of the distributions still contain the true value: increase in precision is not achieved at the expense of accuracy.

CONCLUSIONS

Both estimation and simulation approaches can be formulated as the selection of a set of attribute values that are optimal for specific criteria. Estimation amounts at minimizing *local* criteria such as a conditional estimation variance, whereas stochastic simulation aims at reproducing *global* statistics such as histogram or semivariogram. The search for a map that would have the local accuracy of kriging and the nonsmoothing effect of simulation amounts to finding a balance between local and global constraints which are usually conflicting. Such a compromise can be achieved using simulated annealing and a weighted combination of components that measure deviations from local or global features of interest.

Two major issues of optimization processes are the number of solutions that exist and the way these solutions can be reached. Local constraints imposed for estimation involve each grid node separately and so, for a given optimality criterion, the solution is unique and readily accessible, e.g., the map of E-type or p-quantile estimates. The number of solutions, hence the space of uncertainty, increases when several grid nodes are involved by the components in the objective function. For example, there are many different realizations that match the same target histogram or semivariogram. One must be aware that simulated annealing provides only approximate solutions to the optimization problem in that the objective function is rarely lowered to zero. A good illustration is the map of rescaled E-type estimates which has been shown to be the optimum for the joint constraints of histogram reproduction and minimization of local expected loss. This unique solution is not reached using simulated annealing, and

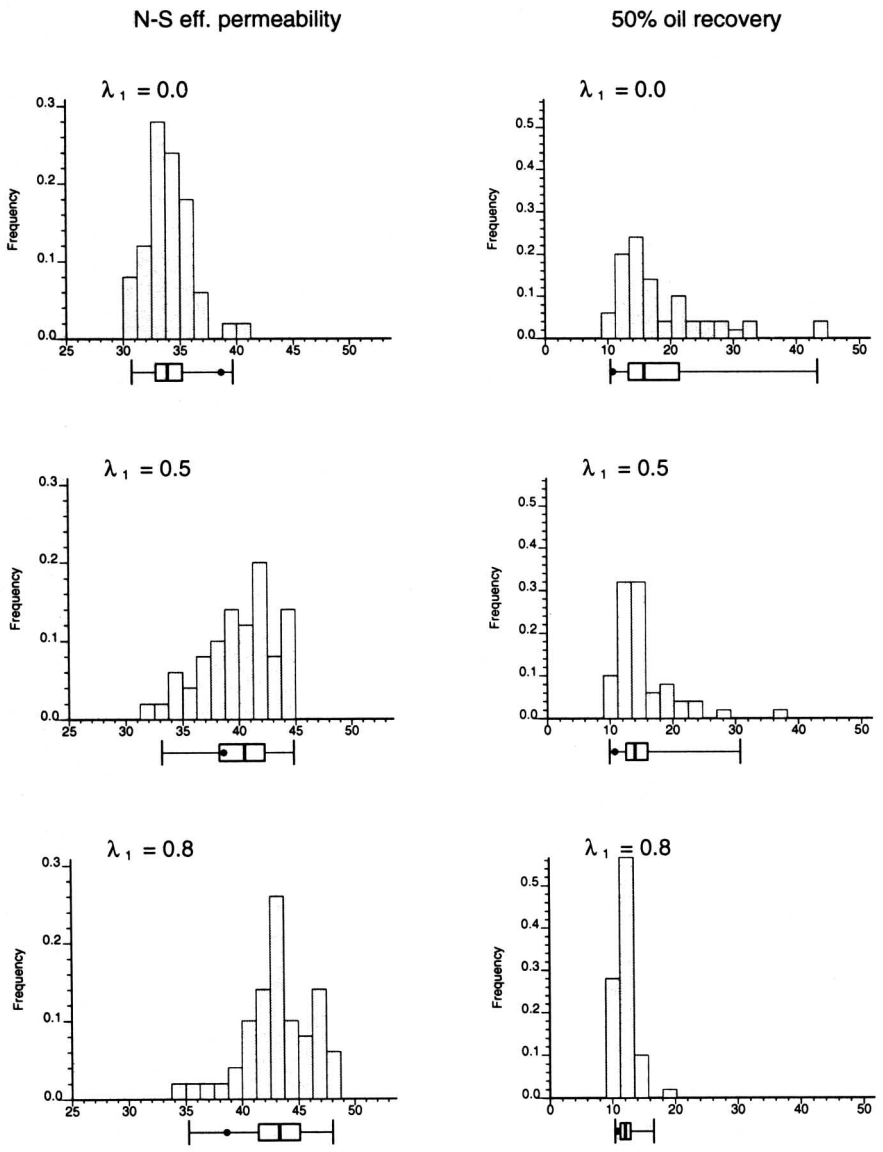

Figure 7. The distributions of N–S effective permeability and time to recover 50% of the oil derived from 50 realizations generated using the reference approach ($\lambda_1 = 0.0$) and accounting for local constraints ($\lambda_1 = 0.5$ or 0.8). The black dot in the box plot below each histogram is the true value obtained from the reference image, the three vertical lines are the 0.025 quantile, the median and the 0.975 quantile of the output distribution.

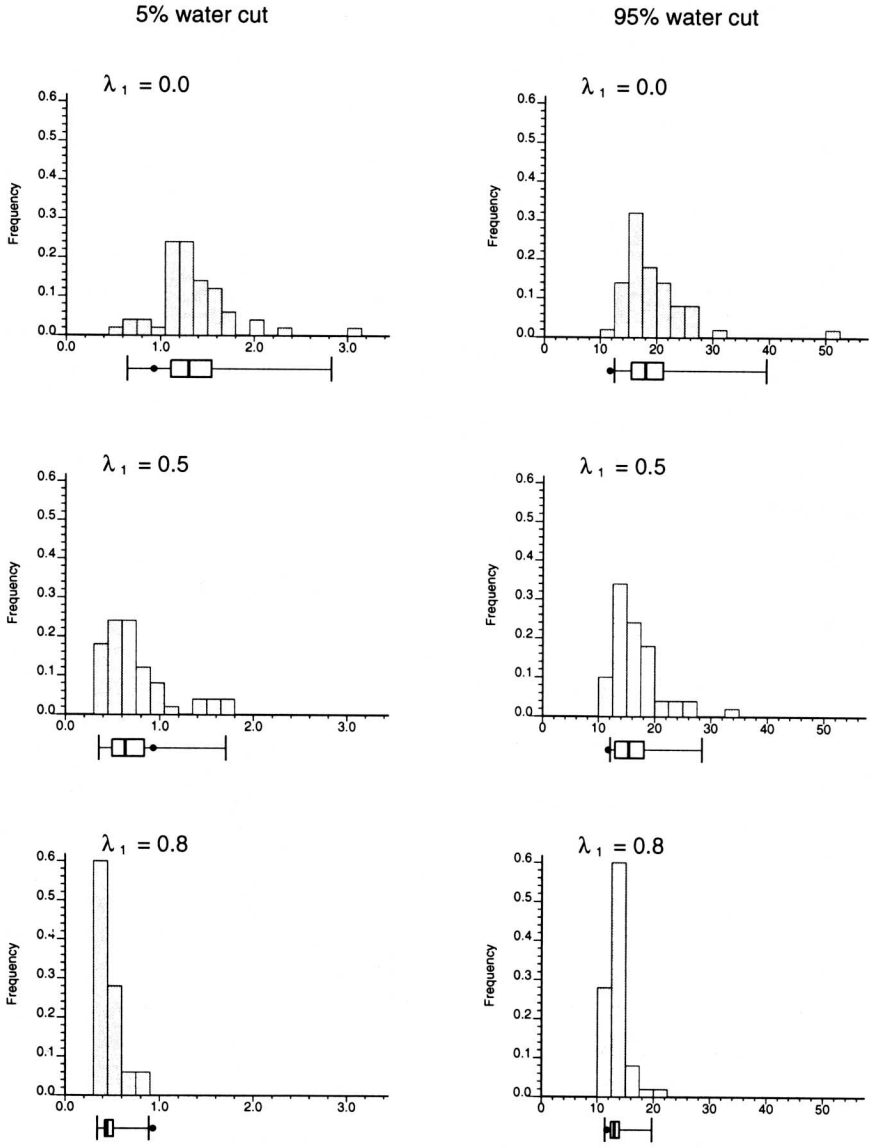

Figure 8. The distributions of the time to reach 5% or 95% water cut derived from 50 realizations generated using the reference approach ($\lambda_1 = 0.0$) and accounting for local constraints ($\lambda_1 = 0.5$ or 0.8). The black dot in the box plot below each histogram is the true value obtained from the reference image, the three vertical lines are the 0.025 quantile, the median and the 0.975 quantile of the output distribution.

fluctuations between realizations, as depicted by the variance map at the bottom of Figure 4, leads to an overestimation of the size of the space of uncertainty. Future studies should investigate the impact of factors such as the annealing schedule, the perturbation mechanism or the initial image (random image, estimated map, realization generated using other simulation algorithms) on the lowering of the objective function.

The case study shows that the incorporation of local criteria in stochastic simulation allows one to keep some desirable features of estimation such as smaller prediction errors while reproducing the pattern of spatial variability (semivariogram). The intermediate approach also yields, on average, smaller errors in production forecasts than algorithm that accounts only for semivariogram and histogram reproduction. One should avoid giving too much weight to local constraints because it can induce a dramatic reduction in the space of uncertainty and so generate a precise but possibly inaccurate distribution of outcomes. Note that if reservoir performance is to be predicted from a single realization, one is better off sampling a distribution of outcomes that is precise but slightly inaccurate than sampling a distribution that contains the true value but is very wide, which increases the risk of drawing a single realization far different from the reality. An alternative to reducing the spread of flow responses consists of integrating more information, such as soft data, in the determination of conditional cdfs, which can be easily done using indicator-based algorithms.

In the future, the approach should be extended to other types of loss function such as the asymmetric function (6) or the indicator function (9). In the latter case, the minimization of a local expected loss amounts to maximizing the local probability of occurrence of attribute values as determined from ccdf models, which is analogous to the procedure developed by Goovaerts (1994b) for categorical variables.

ACKNOWLEDGMENTS

This work was done while the author was with the Unité Biométrie, Université Catholique de Louvain, Belgium. The author thanks the National Fund for Scientific Research (Belgium) for its financial support.

REFERENCES

Christakos, G., 1992, Random field models in earth sciences: Academic Press, New York, 474 p.

Desbarats, A. J., 1996, Modeling spatial variability using geostatistical simulation, *in* Srivastava, R. M., Rouhani, S., Cromer, M. V., and Johnson, A. I., eds, Geostatistics for environmental and geotechnical applications: American Society for Testing and Materials, Philadelphia, p. 32-48.

Deutsch, C. V., and Cockerham, P., 1994, Practical considerations in the application of simulated annealing to stochastic simulation: Math. Geology, v. 26, no. 1, p. 67-82.

Deutsch, C. V., and Journel, A. G., 1992, Annealing techniques applied to the integration of geological and engineering data: Stanford Center for Reservoir Forecasting, Stanford University, Unpublished annual report No. 5.

Deutsch, C. V., and Journel, A. G., 1998, GSLIB: Geostatistical software library and user's guide, 2nd Ed.: Oxford University Press, New York, 369 p.

ECLIPSE 100 Reference Manual, 1991, Intera ECL Petroleum Technologies, Highlands Farm, Greys Road, Henley-on-Thames, Oxfordshire, England.

Farmer, C., 1988, The generation of stochastic fields of reservoir parameters with specified geostatistical distributions, *in* Edwards, S., and King, P., eds., Mathematics in oil production: Clarendon Press, Oxford, p. 235–252.

Geman, S., and Geman, D., 1984, Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images: IEEE Trans. Pattern Anal. Machine Intell. PAMI, v. 6, no. 6, p. 721–741.

Goovaerts, P., 1994a, Comparative performance of indicator algorithms for modeling conditional probability distribution functions: Math. Geology, v. 26, no. 3, p. 389–411.

Goovaerts, P., 1994b, Prediction and stochastic modelling of facies types using classification algorithms and simulated annealing: Stanford Center for Reservoir Forecasting, Stanford University, Unpublished annual report No. 7.

Goovaerts, P., 1996, Stochastic simulation of categorical variables using a classification algorithm and simulated annealing: Math. Geology, v. 28, no. 7, p. 909–921.

Goovaerts, P., 1997a, Kriging vs. stochastic simulation for risk analysis in soil contamination, *in* Soares, A., Gómez-Hernández, J. and Froidevaux, R., eds., geoENV I-Geostatistics for environmental applications: Kluwer Academic Publ., Dordrecht, p. 247–258.

Goovaerts, P., 1997b, Geostatistics for natural resources evaluation: Oxford University Press, New York, 483 p.

Journel, A. G., 1984, Mad and conditional quantile estimators, *in* Verly, G., David, M., Journel, A. G., and Maréchal, A., eds., Geostatistics for natural resources characterization: Reidel, Dordrecht, p. 261–270.

Journel, A. G., 1989, Fundamentals of geostatistics in five lessons, vol. 8, Short course in geology: American Geophysical Union, Washington, D.C.

Journel, A. G., and Alabert, F., 1990, New method for reservoir mapping: Jour. of Petroleum Technology, p. 212–218.

Journel, A. G., and Xu, W., 1994, Posterior identification of histograms conditional to local data: Math. Geology,. v. 26, no. 3, p. 323–359.

Nowak, M., Srivastava, R., and Sinclair, A., 1993, Conditional simulation: A mine planning tool for a small gold deposit, *in* Soares, A., ed., Geostatistics Tróia '92, quantitative geology and geostatistics: Kluwer Academic Publ., Dordrecht, p. 977–987.

Olea, R. A., and Pawlowsky, V., 1996, Compensating for estimation smoothing in kriging: Math. Geology, v. 28, no. 4, p. 407–417.

Schafmeister, M. T., and de Marsily, G., 1993, Regionalization of some hydrogeological processes and parameters by means of geostatistical methods—current status and requirements, *in* Dimitrakopoulos, R., ed., Geostatistics for the next century: Kluwer Academic Publ., Dordrecht, p. 383–392.

Srivastava, R. M., 1987, Minimum variance or maximum profitability?: Canadian Industrial Mining Bull., v. 80, no. 901, p. 63–68.

Srivastava, M. R., 1996, An overview of stochastic spatial simulation, *in* Mowrer, H. T., Czaplewski, R. L., and Hamre, R. H., eds., Spatial accuracy assessment in natural resources and environmental sciences: Second international symposium: General Technical Report RM-GTR-277, U.S. Department of Agriculture, Forest Service, Fort Collins, p. 13–22.