

M-2A

Con 8/09. 34

GEOSTATISTICAL RESERVOIR MODELING

Clayton V. Deutsch

OXFORD
UNIVERSITY PRESS

2002

9/5/2002

Chapter 7

Object-Based Facies Modeling

Object-based facies models are visually attractive because the resulting facies realizations mimic idealized geometries interpreted in outcrops and modern analogues. Such models show facies belonging to clean geological shapes with realistic non-rectilinear continuity, which cannot be modeled with cell-based approaches. This chapter gives some details of modeling facies with object-based techniques.

Section 7.1 gives background on object-based techniques including an overview of some geological shapes that could be considered, the relevant data that object-based models must reproduce, and algorithms for object placement.

Section 7.2 presents the concepts of object-based modeling in the context of stochastic shales. Object-based modeling gained wide popularity in the context of modeling fluvial facies. Section 7.3 presents some details related to modeling abandoned sand-filled fluvial sand channels and related margin facies (levee and crevasse sands). Implementation details including the integration of well and seismic data are discussed. *

Section 7.4 presents considerations for other depositional systems. In particular, the application to deepwater depositional systems. Finally, Section 7.5 presents work flow diagrams for some operations related to object-based facies modeling.

7.1 Background

From a geological perspective, it is convenient to view reservoirs from a chrono-stratigraphic perspective. The sedimentary architecture is considered in light of a hierarchical classification scheme. The reservoir facies are divided into sequences, parasequences, bed sets, beds, and so on. For example, het-

heterogeneities in a fluvial setting are described by stratigraphic reservoir layers, channel complexes, channels, levees, and crevasses through additional smaller scale features. We consider modeling this genetic hierarchy of heterogeneities by successive coordinate transformations and geometric objects representing facies associations. Porosity and permeability models are then constructed at the appropriate scale using coordinate systems aligned with depositional continuity.

Significant effort has been focused on sedimentology and stratigraphy of reservoir systems. It is not possible to provide a reasonable overview of such studies here. Consider fluvial deposits as an example. The recent book by Miall [189] provides a well-illustrated description of fluvial sedimentary facies, basin analysis, and petroleum geology with more than 500 figures and 1000 references. The literature describing fluvial deposits is rich and varied. The history of quantitative computer models for fluvial systems is also extensive. Allen's early qualitative work in the 1960s and 1970s [7] led to quantitative computer simulations [8]. Leeder, at about the same time, was also building quantitative models [172]. Bridge published in this area with Leeder [30] and also published computer code [29] that was updated recently [179].

Although not specifically designed for fluvial facies, object-based models became popular in petroleum reservoir modeling in the mid-1980s due to the work of Haldorsen and others [117, 120, 236]. The importance of fluvial reservoirs in the Norwegian North Sea soon prompted the development of these object-based methods for fluvial facies [38, 45, 86, 113, 126, 203, 235]. The theory and implementation was refined over a number of years [101, 124, 134, 242, 250, 251, 253] with increasing practical application of these methods to Norwegian North Sea reservoirs [28, 252]. Such applications have set the standard for other oil producing regions of fluvial depositional setting. Other non-Norwegian oil companies also developed object-based modeling capability [6, 161].

* Object-based models are now created routinely in reservoir characterization. The three key issues to be addressed in setting up an object-based model are the (1) geological shapes and their parameter distributions, (2) algorithm for object placement modification, and (3) relevant data to constrain the resulting realizations.

We must exercise a little caution in this chapter. This is *not* a rigorous presentation of geological concepts for different depositional systems. A pragmatic approach is considered. Pragmatic from a geological, flow modeling, and geostatistical perspective. The idea of modeling reservoirs by genetic forward modeling is attractive, but not considered here. Object-based models are *pseudo-genetic* in the sense that erosional rules (younger rocks on top) and other geological principles are used as much as possible.

There is the inevitable question of semantics and word choice. I choose to use "object-based" modeling. Of course, these object-based models end up as a collection of cells, but they are not called cell-based. The reference to object-based is in the method of creation and not the format of the result. Some

prefer "Boolean" models, which has a connotation of greater statistical rigor. Others prefer "marked point process" connoting a statistical point process for object centroids and then object properties such as size, orientation, and facies type are attributed to the point process.

Geological Shapes

There is no inherent limitation to the shapes that can be modeled with object-based techniques. The shapes can be specified by equations, a raster template, or a combination of the two. The geological shapes can be modeled hierarchically, that is, one object shape can be used at large scale and then different shapes can be used for internal small-scale geological shapes. Some evident shapes:

- Abandoned sand-filled fluvial channels within a matrix of floodplain shales and fine-grained sediments. The sinuous channel shapes are modeled by a 1-D centerline and a variable cross section along the centerline. Levee and crevasse objects can be attached to the channels. Shale plugs, cemented concretions, shale clasts, and other non-net facies can be positioned within the channels. Clustering of the channels into channel complexes or belts can be handled by large-scale objects or as part of the object placement algorithm.
- Lower energy meandering fluvial systems can be modeled as sand lenses within a non-net background. We sometimes consider modeling the entire channel (as above) and then assigning sand and shale within the channel in some realistic manner.
- Other channelized depositional systems including deepwater and estuarine systems are often modeled by adapting fluvial channel modeling techniques to system-specific considerations such as channel size, width to thickness ratios, and internal heterogeneities. The transition from channels to other facies types can also be handled by some rule-based scheme, that is, a channel may evolve into a more lobe-like geometry at some distance into the modeling domain (as energy is lost and the sediments disperse).

Channel-type systems are discussed in detail in Section 7.2.

- Eolian dune shapes are obvious candidates for object-based modeling. Although the 3-D geometry of such object is not trivial to define analytically, we can make the necessary assumptions for practical modeling.
- Remnant shales may be modeled as disk or ellipsoid objects within a matrix of sand. This may be appropriate in high net-to-gross reservoirs. Although such shales may have a low proportion they will significantly affect the vertical permeability, hence, horizontal well production and coning.

In the case of a single secondary variable (Y_2), the simple cokriging estimator of $Y(u)$ is written:

$$Y_{COK}^s(u) = \sum_{\alpha_1=1}^{n_1} \lambda_{\alpha_1} Y(u_{\alpha_1}) + \sum_{\alpha_2=1}^{n_2} \lambda'_{\alpha_2} Y_2(u'_{\alpha_2}) \quad (5.15)$$

where the λ_{α_1} s are the weights applied to the n_1 Y samples and the λ'_{α_2} s are the weights applied to the n_2 Y_2 samples. Expression (5.15) is written as simple cokriging for standardized variables, that is, the means of Y and Y_2 are assumed known and zero.

Kriging requires a model for the Y covariance. Cokriging requires a joint model for the matrix of covariance functions including the Y covariance $C_Y(h)$, the Y_2 covariance $C_{Y_2}(h)$, the cross Y - Y_2 covariance $C_{Y Y_2}(h)$, and the cross Y_2 - Y covariance $C_{Y_2 Y}(h)$.

When K different variables are considered, the covariance matrix requires in all generality K^2 covariance functions. Such inference becomes extremely demanding in terms of data and the subsequent joint modeling is particularly tedious [110]. This is the main reason why cokriging has not been extensively used in practice. Algorithms such as collocated cokriging and Markov models (see hereafter) have been developed to shortcut the tedious inference and modeling process required by cokriging.

Other than tedious variogram or covariance inference, cokriging is the same as kriging. The cokriging equivalents of OK and UK exist where the mean values are implicitly estimated from the neighborhood data. The reader is referred to [34, 68, 110, 157, 195, 194, 259] for details.

Collocated Cokriging

A reduced form of cokriging consists of retaining only the collocated secondary data $y_2(u)$ or relocated data $y_2(u')$ to the nearest node u being estimated [68, 274]. This is not a problem if the distance $|u - u'|$ is small with respect to the volume of influence of $y_2(u)$. The cokriging estimator is written:

$$Y_{COK}^s(u) = \sum_{\alpha_1}^{n_1} \lambda_{\alpha_1} Y(u_{\alpha_1}) + \lambda' Y_2(u) \quad (5.16)$$

The corresponding cokriging system requires knowledge of the Y covariance $C_Y(h)$ and the Y - Y_2 cross covariance $C_{Y Y_2}(h)$. The latter can be approximated through the following model:

$$C_{Y Y_2}(h) = B \cdot C_2(h), \quad \forall h \quad (5.17)$$

where $B = \sqrt{C_Y(0)/C_{Y_2}(0)}$, $\rho_{Y Y_2}(0)$, $C_Y(0)$, $C_{Y_2}(0)$ are the variances of Y and Y_2 , and $\rho_{Y Y_2}(0)$ is the linear coefficient of correlation of collocated $y - y_2$ data.

* The approximation consisting of retaining only the collocated secondary datum does not affect the estimate (close-by secondary data are typically

very similar in values), but it may affect the resulting cokriging estimation variance: that variance is overestimated, sometimes significantly. In a kriging estimation context this is not a problem, because kriging variances are of little use. In a simulation context, see Chapter 8.2, where the kriging variance defines the spread of the conditional distribution from which simulated values are drawn, this may be a problem. The collocated cokriging variance should then be reduced by a factor (assumed constant $\forall \mathbf{u}$) to be determined by trial and error.

5.2 Sequential Gaussian Simulation

There are many algorithms that can be devised to create stochastic simulations: (1) matrix approaches (LU Decomposition), which are not extensively used because of size restrictions (an $N \times N$ matrix must be solved where N , the number of locations, could be in the millions for reservoir applications), (2) turning bands methods where the variable is simulated on 1-D lines and then combined into a 3-D model; not commonly used because of artifacts, (3) spectral methods using FFTs can be CPU fast, but honoring conditioning data requires an expensive kriging step, (4) fractals, which are not used extensively because of the restrictive assumption of self-similarity, and (5) moving average methods, which are infrequently used due to CPU requirements.

The common approach adopted in recent times for reservoir modeling applications is the sequential Gaussian simulation (SGS) approach [137]. This method is simple, flexible, and reasonably efficient. Let's review the theory underlying SGS. Recall the simple kriging estimator:

$$Y^*(\mathbf{u}) = \sum_{\beta=1}^n \lambda_{\beta} \cdot Y(\mathbf{u}_{\beta})$$

and the corresponding simple kriging system:

$$\sum_{\beta=1}^n \lambda_{\beta} C(\mathbf{u}_{\alpha} - \mathbf{u}_{\beta}) = C(\mathbf{u} - \mathbf{u}_{\alpha}), \quad \mathbf{u}_{\alpha} = 1, \dots, n$$

The covariance between the kriged estimate and one of the data values can be written as:

$$\begin{aligned} \text{Cov}\{Y^*(\mathbf{u}), Y(\mathbf{u}_{\alpha})\} &= E\{Y^*(\mathbf{u}), Y(\mathbf{u}_{\alpha})\} \\ &= E\left\{\left[\sum_{\beta=1}^n \lambda_{\beta} \cdot Y(\mathbf{u}_{\beta})\right] \cdot Y(\mathbf{u}_{\alpha})\right\} \\ &= \sum_{\beta=1}^n \lambda_{\beta} \cdot E\{Y(\mathbf{u}_{\beta}) \cdot Y(\mathbf{u}_{\alpha})\} \end{aligned}$$

The cov
kriging e
between
variance
selves is
estimate
ues in su
Altho
correct
function
constant

The sm
larly at
kriging is

This te
This mis
reproduc
An in
be added

The cov
used in t

Cov(Y

note that
of any da
E(R|u

fact that the conse-
 assigning the wrong
 facies as non-net

ability of the pre-
 assigned an 30%

ilities may be sum-

(10.5)

ty of the true facies.
 plete information and

how "right" the prob-
 t when a similar facies
 ally different facies are
 count for these conse-
 facies i is to facies j .
 ame facies, and $c(i, j)$
 channel sand and $j =$
 qualitatively consider-
 alitative assignment is
 account for the fuzzy

ilities. The local prob-
 ect the true fraction of
 der all locations where
 then 65% of those lo-
 about 0.65; we check
 actual fraction departs
 em. These results may

es prediction. Seismic
 uncertainty. Indicator
 lock cokriging (BC) of
 were considered. The
 o three methods (note
 an updating approach
 kriging approach is
 here. Figure 10.7 shows
 probabilities.

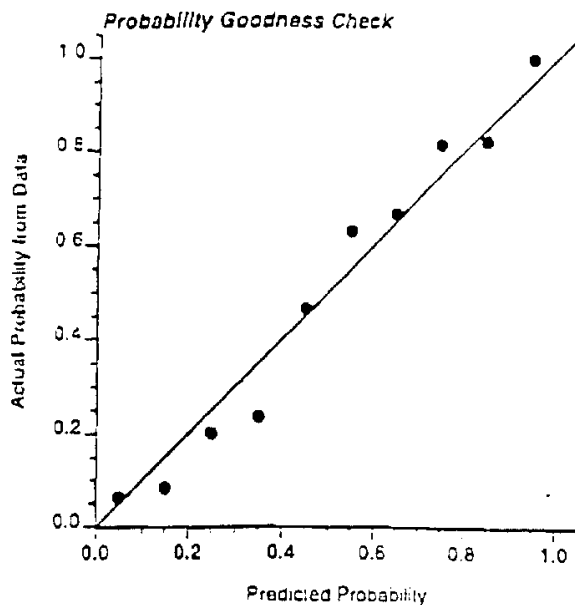


Figure 10.7: Example cross plot of actual probabilities versus the predicted probabilities. The closeness of the results to the 45 degree line attests to the goodness of the probabilities. These results are for block cokriging in all facies. We could also look at the results on a by-facies basis.

The main uses for the diagnostic tools presented here are: (1) detecting implementation errors, (2) quantifying uncertainty, (3) comparing different simulation algorithms (for example, Gaussian-based algorithms versus indicator-based algorithms versus simulated annealing based algorithms), and (4) fine-tuning the parameters of any particular probabilistic model (for example, the variogram model used). These tools provide basic checks, that is, necessary but not sufficient tests. They do not assess the multivariate properties of the simulation. Care is needed to ensure that features that impact the ultimate prediction and decision making, such as continuity of extreme values, are adequately represented in the model.

* 10.4 How Many Realizations?

An important contribution of geostatistical reservoir modeling is a framework for uncertainty assessment through Monte Carlo sampling from the space of uncertainty in reservoir responses. Monte Carlo sampling proceeds by (1) drawing L realizations from a probabilistic model, (2) processing the L realizations through some performance calculation, and (3) assembling a histogram of the L responses to represent a distribution of uncertainty in the output(s). Classical Monte Carlo simulation requires the L realizations to

be drawn randomly; therefore, they each go into the distribution of uncertainty with equal probability. The question of how many realizations must be addressed.

Significant CPU time is required to process realizations through a flow simulator, which is the most common performance calculator. Often, a small number of realizations are considered for this reason alone. The following discussion on the required number of realizations is for those cases that we can consider more realizations for a better assessment of uncertainty.

The required number of realizations depends on the (1) aspect of uncertainty or the "statistic" being quantified, and (2) the precision with which the uncertainty assessment is required. Few realizations may be required to assess an average statistic such as the average porosity. A large number of realizations are required, however, to assess the 1% and 99% percentiles of the oil in place distribution.

A quantitative basis must be established to determine the required number of realizations. The reporting procedures of environmental standards and political opinion polls will help us define the basis for the results of geostatistical analysis. Consider a poll to answer the question "Have recent price increases for energy caused any financial hardship for you or your household?" The answer would be reported as percentages of yes and no responses, for example, 56% yes and 44% no. Responsible polling agencies would add a caveat to convey the uncertainty in this result. They may report the number of respondents or, more likely, they would apply basic statistics to more completely communicate uncertainty in the response, for example:

The true percentage is within 3% of this reported result (56% yes)
18 times out of 20.

These two additional numbers are used to summarize uncertainty in the reported result. The left side of Figure 10.8 shows a schematic example of the uncertainty in the statistic of 56%. If more people were polled (analogous to more realizations being generated) the uncertainty would be less, see right side of Figure 10.8. This notion of "uncertainty in reported statistics" will be extended to quantify the uncertainty in our results for a fixed number of realizations.

The reported statistics from geostatistical analysis are particular quantiles of some response variable (for example, oil in place), say, the $F(p) = 0.1, 0.5,$ and 0.9 quantiles. As introduced above, two tolerance parameters are also required: (1) a tolerance for the quantiles, say $\Delta_F = \pm 0.01$, and (2) the minimum probability of being within probability Δ_F , say $t_F = 0.8$ or 80%. A typical case would be to require the P_{10}, P_{50} , and P_{90} of recovery factor within 2%, 50% of the time. A more stringent requirement would be the P_{10} and P_{90} of net present value within 0.1%, 95% of the time ($\Delta_F = 0.001$ and $t_F = 0.95$).

Thus, two parameters are used: (1) Δ_F , which is the deviation, in units of cumulative probability, from the reported quantile, and (2) t_F , which is

Figure
sponds
of this
reduced
availabl

the frac
In this
a certa
the pa
possibi
realiza

The
culated
tributi
numbe
indep
centra
toward
consider
that th

Th
the fir
sampl

where
the w
under

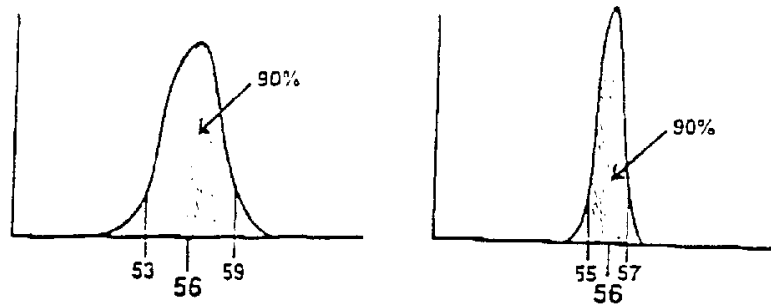


Figure 10.8: Uncertainty in a reported statistic of 56%. The left side figure corresponds to the example in the text where the true value is reported to be within 3% of this number 18 times out of 20 (90% of the time). The right side figure shows reduced uncertainty if a larger number of samples were taken (more realizations available) for uncertainty reporting.

the fraction of times the true value will fall within \pm of the reported statistic. In this context, we can derive the required number of realizations L to meet a certain specified standard, that is, a specified Δ_F and t_F . Alternatively, the parameters Δ_F and t_F can be established for a given L , that is, it is possible to calculate the uncertainty of our results due to a small number of realizations.

The number of realizations L to meet specified Δ_F and t_F can be calculated. The sampling distributions for the cdf values tend to normal distributions. This is not surprising since the cdf value is the sum of a large number of values (the indicator transform at the correct threshold) that are independent (the realizations are random) and identically distributed. The central limit theorem tells us that the sampling distribution in this case tends toward a normal distribution as the number increases. The number we are considering (L) is large for the central limit theorem; therefore, it is expected that the distribution will be normal. This has been verified numerically.

The mean and variance of the sampling distribution for a cdf value (say, the first quantile of interest $F = 0.1$) can be determined theoretically. The sampled cdf value F^* is the average of an indicator value:

$$F^* = \frac{1}{L} \sum_{i=1}^L i(u_i)$$

where the indicator function is 1 if random drawing u_i is less than or equal to the value F and 0 otherwise. The mean or expected value of F^* is the true underlying cdf value, that is, F . The variance of F^* is calculated as:

$$\sigma_{F^*}^2 = \frac{1}{L} \sigma_i^2 = \frac{1}{L} E\{I^2\} - E\{I\}^2 = \frac{1}{L} F - F^2 = \frac{F(1-F)}{L}$$

This analytical result for the uncertainty of a cdf/quantile has been verified by numerical results. The simplicity of these results makes it straightforward to calculate the t_F statistic for specified L , Δ_F , and F values.

$$\begin{aligned} t(L, \Delta_F, F) &= \text{Prob}\{F^* \geq (F - \Delta_F)|L\} - \text{Prob}\{F^* < (F + \Delta_F)|L\} \\ &= G\left(\frac{\Delta_F}{\sqrt{F(1-F)/L}}\right) - G\left(\frac{-\Delta_F}{\sqrt{F(1-F)/L}}\right) \\ &= 2 \cdot G\left(\frac{\Delta_F}{\sqrt{F(1-F)/L}}\right) - 1 \end{aligned} \quad (10.6)$$

There are numerous $G(y)$ functions available, for example, the gcum and ginv functions in GSLIB. This is an important result. The precision for a given number of realizations can be directly calculated. We can invert this relationship to get an equation that gives us the number of realizations L to achieve a certain precision specification:

$$L = \frac{F(1-F)}{\left(\frac{\Delta_F}{G^{-1}(\frac{t+1}{2})}\right)^2} \quad (10.7)$$

The number of realizations for a specified precision can be calculated directly.

Figure 10.9 shows curves of t_F versus the number of realizations L for $\Delta_F = 0.1, 0.2, 0.3, 0.4$, and 0.5 . These five curves were also calculated numerically and the results match. These curves relate to the 0.5 quantile. These relations tell us "how many realizations are needed for a specified precision" and "how good are probability values for a given number of realizations?"

Practical Results: How Many Realizations?

The analytical result presented above should be used if there is a requirement for quantitative measures of uncertainty. This may lead to an unrealistically large number of realizations. In general, a staged approach should be considered.

- Enumerate the geological scenarios to be considered and generate a single geostatistical realization of all variables for each scenario. Validate that each realization honors the input data within acceptable statistical fluctuation. Calculate the response variables (oil in place, flow performance, and so on) with each realization.
- Generate five realizations for all important scenarios. Important scenarios have a large probability of occurrence or response variables that are unusually low or high. The expert probabilities assigned to each scenario and the response variable from the first realization must be used to make this judgement. Calculate the response variables with these realizations.

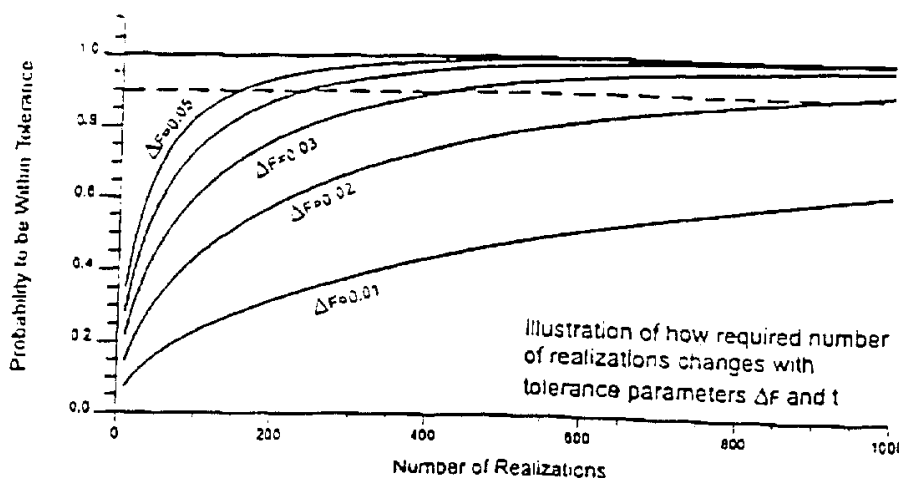


Figure 10.9: Illustration of how the t parameter is related to the number of realizations for $\Delta F = 0.1, 0.2, 0.3, 0.4,$ and 0.5 . The precision of the uncertainty assessment increases as more realizations are considered.

- Consider more realizations for those scenarios where the first five random realizations are quite different from each other. The response variables could be calculated on all of these additional realizations.

Calculating the response variables of interest can involve running a flow simulator at significant CPU and professional cost. The generation of geostatistical realizations is relatively quick. This leads to the idea of generating a large number of realizations, ranking them by some fast-to-calculate statistic, and then processing a limited number through the full flow simulator.

*10.5 Ranking Realizations

Each geostatistical realization reproduces the input data equally well and is equally probable given the scenario or random function model under consideration. Nevertheless, the realizations may be ranked according to some criteria or measure not used as data. For example, the volume of hydrocarbon-filled pore space connected the well locations can be used to rank a large number of geostatistical realizations.

There is no unique ranking index when there are multiple flow response variables and no ranking measure is perfect. Nevertheless, the value of ranking realizations will be a reduction in the number of realizations that must be processed to arrive at the same level of uncertainty.

A good ranking statistic correctly identifies low and high realizations. Before describing a number of ranking measures, consider some cases where ranking is problematic or unnecessary:

1. when each realization leads to nearly the same answer;
2. when the aspect of uncertainty being assessed is easy to calculate, for example, the uncertainty in oil-in-place may be simply assessed by calculating the pore volume of all realizations;
3. when there are many independent reservoir responses of interest, that is, it is impossible to conceive of a ranking index that would lead to a unique reliable ranking.

There are times, however, when significant professional or CPU time is required to evaluate each realization and the number of realizations considered must be limited. In these situations, it is worthwhile to consider ranking the realizations to limit the number of fine-scale simulations and yet to obtain an idea of uncertainty in the flow response.

Most simple ranking measures require each cell in all geostatistical realizations to be assigned a *net* indicator. This indicator is one if the cell is reservoir quality and zero otherwise. Reservoir quality is achieved in certain facies by porosity and permeability simultaneously above specified thresholds. The net-to-gross ratio for each realization can be easily calculated and used as a simple ranking measure. Another simple ranking measure is the total oil volume, that is, the product of the net-to-gross indicator, the oil saturation, and the pore volume. This accounts for porosity and saturation variations.

Flow performance depends on the connectivity of the reservoir-quality rock; there are many measures of connectivity that may be quickly calculated without running a full simulator [18, 65, 23]. A first measure of connectivity is available by determining the sets of *net* geological modeling cells that are connected in 3-D space. There are fast algorithms to scan through 3-D binary net/non-net realizations aggregating those reservoir-quality cells that are connected net cells. The result will be a 3-D specification of the number N_{geo} of geo-objects or connected objects each with an associated volume $V_{geo,j}$, $j = 1, \dots, N_{geo}$. Figure 10.10 shows the geo-objects for three SIS facies realizations.

The geo-objects of a set of realizations can be used in a number of ways to rank the realizations:

- The fraction of cells within the first, say, 5, geo-objects could be used as a ranking measure. A realization is more "connected" when this fraction is large.
- The connected hydrocarbon volume within some radius of the production wells could be considered if the well locations are known. This ranking measure accounts for local information about connectivity.
- The connected volume between "pairs" of injection and production well pairs could be used if the well pair locations are known.



Figure 10.10: Slices through three facies models. The black region on each realization is the largest geo-object. Smaller geo-objects are shaded in gray.

An alternative use of geo-objects is for the selection of well locations. An optimization scheme could be devised to select the well locations that maximizes the geo-object volume within some drainage radius.

Geo-objects measure static connectivity, that is, they do not account for tortuosity, permeability, attic oil, and the interaction between multiple producing well locations. There are some more sophisticated alternatives that measure dynamic connectivity and, yet, are still simpler than running the full flow simulator. There are random walk algorithms that measure "dynamic" continuity between injecting and producing locations. These methods often call for a solution to the pressure field (single phase flow) given assumed well rates. Particles are then tracked through the media and the distribution of "times" or "lengths" between injecting and producing wells provides a measure of connectivity that could be used for ranking.

There are other relatively simple and fast flow models including (1) tracer simulation, (2) simulation based on a network of 1-D streamlines, and (3) a water flood simulation in lieu of a more complex miscible or compositional-type simulation. The time of 5% watercut is likely a good measure for the breakthrough of other miscible components in a more complex process.

Another ranking approach is to use the correct physics or flow equations but with the geological models scaled to such a coarse resolution that the computer time is acceptable. The coarseness of the underlying grid will compromise the direct usefulness of the results. Nevertheless, the relative ranking of the results may be used to rank the underlying geological realizations.