What in the Reservoir is Geostatistics Good For?

Dr. Clayton V. Deutsch

Centre for Computational Geostatistics (CCG)
School of Mining and Petroleum Engineering
University of Alberta

Abstract

This paper was written as an invited paper for the Journal of Canadian Petroleum Technology (JCPT). New developments are not presented – it is a type of overview paper.

Geostatistics provokes strong reactions. There are champions who believe the application of geostatistics adds value in almost any reservoir modeling situation. There are skeptics who do not think that a geostatistical model will have a meaningful impact on reservoir management decisions. The majority of engineers and geoscientists, however, are seeing an increasing use of geostatistics and are not sure when geostatistics should be used and how the results affect reservoir decisions.

There are three specific cases where geostatistics can provide valuable support for decision making: (1) calculating maps of uncertainty over large areas to support resource calculations and well placement, (2) reconciling well and seismic data into high resolution reservoir models, and (3) constructing representative models of heterogeneity to provide input to flow simulation and support reservoir forecasting. These three cases are developed without excessive theoretical detail. Realistic examples are presented without getting lost in the details of a particular reservoir. Limitations and pitfalls are discussed.

Framework of Geostatistics

Geostatistics refers to the theory of regionalized variables and the related techniques that are used to predict variables such as rock properties at unsampled locations. Matheron formalized this theory in the early 1960's (Matheron, 1971). Geostatistics was not developed as a theory in search of practical problems. On the contrary, development was driven by engineers and geologists faced with real problems. They were searching for a consistent set of numerical tools that would help them address real problems such as ore reserve estimation, reservoir performance forecasting, and environmental site characterization. Reasons for seeking such comprehensive technology included (1) an increasing number of data to deal with, (2) a greater diversity of available data at different scales and levels of precision, (3) a need to address problems with consistent and reproducible methods, (4) a belief that improved numerical models should be possible by exploiting computational and mathematical developments in related scientific disciplines, and (5) a belief that more responsible decisions would be made with improved numerical models. These reasons explain the continued expansion of the theory and practice of geostatistics. Problems in mining, such as unbiased estimation of recoverable reserves, initially drove the development of geostatistics. Problems in petroleum, such as realistic heterogeneity models for unbiased flow predictions, were dominant from the mid 1980s through the late 1990s. More recently, the problems of realistic geologic modeling and reliable uncertainty quantification are driving development.

The main focus of geostatistics is constructing high-resolution 3-D models of categorical variables such as facies and continuous variables such as porosity and permeability. It is necessary to have *hard* truth measurements at some volumetric scale. All other data types including geophysical data are considered *soft* data and must be calibrated to the hard data. It is neither possible nor optimal to construct models at the resolution of the hard data. Models are generated at some intermediate geological modeling scale, and then scaled to an even coarser resolution for resource calculation or flow simulation. A common goal of geostatistics is the creation of detailed numerical 3-D geologic models that simultaneously account for a wide range of relevant data of varying degrees of resolution, quality, and certainty. Much of geostatistics relates to data calibration and reconciling data types at different scales.

Conventional mapping algorithms were devised to create smooth maps to reveal large-scale geologic trends; they are low pass filters that remove high frequency property variations. The goal of such conventional mapping algorithms is *not* to reveal the variability. For many practical problems, however, this variability has a large affect on the predicted response. Geostatistical simulation techniques, conversely, are devised with the goal to reveal the full variability that could be encountered, that is, create maps or realizations that are neither unique nor smooth. Although the small-scale variability of these realizations may mask large-scale trends, geostatistical simulation is more appropriate for most engineering applications.

At any instance in geological time, there is a single true distribution of facies and reservoir properties. This true distribution is the result of a complex succession of geological processes. Although some of these processes may be understood quite well, we do not completely understand all of the processes and their interaction. Further, we could never have access to the boundary conditions in sufficient detail to model the unique true distribution of properties. We can only hope to create numerical models that mimic physically significant features. Uncertainty exists because of our lack of exhaustive knowledge and the variability. Geostatistical techniques allow alternative realizations to be generated, which are often considered jointly as a model of uncertainty.

Geostatistical modeling requires spatial statistical control on the nature of the variability. Often, however, there are often insufficient data to provide reliable statistics. For this reason, data from reservoir analogues are used to help infer spatial statistics that are impossible to calculate from the available data. There are general features of each geological setting that can be transported to other reservoirs of similar geological setting.

The conceptual framework of geostatistics starts from an admission that the true variability of reservoir properties is important, but will never be accessible. Numerical tools are used to create numerical models that mimic the patterns of variability that we believe exist in the reservoir. These models are rarely based on depositional or diagenetic principles, but they are useful to appreciate variability and the consequent uncertainty.

Prerequisites

A sequential or hierarchical approach is often followed for geostatistical reservoir modeling. The areal geometry and large scale stratigraphic layering are defined first, perhaps deterministically. The facies are modeled within each reservoir layer. Continuous variables are modeled within the chosen facies types. Repeating the entire process creates multiple equally probable realizations.

Reservoir rock types or facies are almost always modeled when high resolution 3-D models are constructed. Typically, two to six grouped facies types are specified and facies models are

constructed one grid cell at a time using statistical control (cell-based modeling) or by collections of cells that represent depositional or diagenetic units (object-based modeling). Figure 1 illustrates these two modeling approaches. Continuous reservoir properties, generically denoted Z, are assigned within reservoir layers/zones that are deemed reasonably homogeneous.

The uncertainty about an unsampled value z is modeled through the probability distribution of a random variable (RV) Z. The probability distribution of Z after data conditioning is usually location-dependent; hence the notation $Z(\mathbf{u})$, with $\mathbf{u}$ representing the location. Inference of any statistic requires repetitive sampling. For example, repetitive sampling of the variable $z(\mathbf{u})$ is needed to evaluate the probability function: $F(\mathbf{u};z) = \text{Prob}\{Z(\mathbf{u}) \leq z\}$ from experimental proportions; however, at most one sample is available at any single location $\mathbf{u}$. Therefore, the paradigm underlying geostatistics is to trade the unavailable replication at location $\mathbf{u}$ for replication over other locations within the same stratigraphic layer or facies. This trade of replication corresponds to a decision of stationarity. The decision to pool data for statistical calculations is an indispensable feature of geostatistics.

Essential Tools of Geostatistics

The tools mentioned in this section are described more fully in a number of books and technical papers. My purpose is to briefly introduce the different tools before mentioning them in the applications.

Wells are not drilled to be statistically representative of the reservoir; they are often intended as locations for production. Even in exploration, there is a desire to delineate interesting (high quality) areas of the reservoir. It is critical to establish a representative distribution for each variable being modeling. This includes facies proportions and the histograms of porosity and permeability within each facies type. *Declustering* techniques weight the data such that wells drilled close together are given less weight. Wells drilled farther apart are given more weight. Declustering is suitable when there are sufficient data to sample areas of high and low quality. Sometimes there are too few wells. There may be areas of relatively poor reservoir quality that have not been drilled. *Debiasing* techniques are used to establish representative distributions based on a secondary variable such as seismic or a geologic trend. The results of declustering and debiasing include representative facies proportions and representative histograms of each continuous variable under consideration. A large-scale trend model may have been built for debiasing – this trend model will also come into subsequent geostatistical calculations.

A goal of geostatistics is to construct local conditional distributions of uncertainty. The word *conditional* is used to emphasize that the uncertainty is conditioned to local well data and other relevant soft data coming from seismic or other sources. These local distributions of uncertainty can be used to appreciate variability and for decision making. The standard approach to estimate conditional probabilities is Bayes Law, which has been used for more than 200 years. Conditional probabilities require inference of the multivariate probability of the conditioning data and the multivariate probability of the conditioning data and the event being predicted.

The multivariate probabilities required for continuous variables are difficult to infer from data. A multivariate Gaussian model is systematically adopted. The continuous variable is transformed to a Gaussian distribution, and then all multivariate distributions are assumed to be Gaussian. We would wish for some alternative probabilistic models to choose from; however, the multivariate Gaussian probability distribution is remarkably tractable and used almost exclusively. Figure 2 illustrates transformation of a continuous variable from an arbitrary distribution to a Gaussian

distribution. The distributions are shown as cumulative distributions. In Gaussian units (the right side), all distributions are Gaussian in shape. The uncertainty in original units must be established by back transformation.

The conditioning to different data events requires a measure of proximity or closeness between (1) the data events and what we are trying to predict and (2) between the data events. An important feature of the multivariate Gaussian probability distribution is that closeness is always measured by the covariance. If possible, covariances are calculated directly from the available hard and soft data. Often, we do not have enough data and the covariances must be inferred from a statistical model. For a variety of historical reasons, spatial covariances are rarely calculated directly. A measure of variability called the variogram is calculated and fit, then covariances are inferred from the fitted variogram model.

Bayes Law is used to calculate conditional probability distributions in presence of multiple hard and soft data. The results of Bayes Law are obtained *exactly* with simple kriging in a Gaussian framework. Simple cokriging is used when multiple hard and soft data are used. Kriging is often thought of as a geostatistical procedure for estimating the value of variable at an unsampled location. This is true, but a more insightful interpretation of kriging is that it solves Bayes Law for the parameters of conditional distributions of uncertainty. Conditional distributions of uncertainty in Gaussian units are back transformed to original units (see the bottom of Figure 2). In many cases, such distributions of uncertainty are used for decision making. Simulation must be used to assess uncertainty in more than one location at a time.

Simulated realizations are required for two reasons. Firstly, they provide numerical models of heterogeneity for process evaluation, for example, calculating recovery factors. Secondly, they permit input uncertainty to be transferred to output uncertainty, for example, calculating uncertainty in in-place hydrocarbon. There are a number of implementations that generate multiple realizations. Sequential methods such as sequential Gaussian simulation are popular. This paper cannot replace textbooks on the subject of geostatistical modeling, but a small example will be shown to illustrate these concepts.

A Small Example

Consider a square grid of 101 – 16m grid cells that cover just over one regular Section of land. Let's directly model porosity. The global representative distribution will be taken as lognormal with a mean $m=0.15$ and a standard deviation $\sigma=0.075$. The global representative distribution would be obtained by declustering and/or debiasing using the available well and seismic data. Consider an average data of 0.15 in the northwest corner of the area and a high data of 0.25 in the southeast corner of the area.

Uncertainty is characterized in Gaussian units. The transformation to a standard Gaussian distribution is defined analytically in this case:

$$y = \frac{\log(z) - \alpha}{\beta} \quad \text{where } \beta = \sqrt{\log\left(1 + \frac{\sigma^2}{m^2}\right)} \text{ and } \alpha = \log(m) - \beta^2/2$$

In our case $\alpha=-2.01$ and $\beta=0.472$. The back transform is also defined analytically: $z=exp(y\beta+\alpha)$. The porosity data values of 0.15 and 0.25 are transformed to 0.236 and 1.317, respectively.

A fitted variogram model of the Gaussian transformed values is required. This would be obtained from the available data and analogue information. The variogram will be taken as an exponential function with an effective range of 2000m: $\gamma(h)=1-exp(3h/2000)$. In fact, $\gamma(h)$ is the semivariogram or one half of the variogram. Under a decision of stationarity, the covariance function is simply $C(h)=1-\gamma(h)=exp(3h/2000)$.

Local conditional distributions are defined everywhere by a local conditional mean and variance that are computed by simple kriging. Figure 3 shows these results. The locations of the wells are evident on the conditional variance map – the conditional variance is zero at the two well locations. These results are in Gaussian units. We back transform these conditional distributions to original units by back transforming a large number of quantiles, say 200. Figure 4 shows maps of the conditional mean, conditional variance, P_{90} low value and P_{10} high value in original units. Note how the conditional variance in original units is higher in the south and east because the mean is higher; the conditional variance in original units depends on the data as well as the data configuration.

Multiple realizations of porosity are generated by Gaussian simulation. Five realizations are shown on the left of Figure 5. The two well data are reproduced by all realizations. The original oil in place (OOIP) was calculated on each realization assuming a thickness of 10m and a residual water saturation of 0.2. The distribution of OOIP is shown at the right of Figure 5. These realizations allow us to visualize heterogeneity as well as assess uncertainty. The realizations could be ranked by their OOIP and select realizations (say the ones with the P_{90}, P_{50} and P_{10} OOIP) could be input to flow simulation.

This little example shows a hint of what geostatistics is aimed at. In practice, we must consider multiple layers, multiple facies, multiple data types, and multiple variables such as saturation and permeability. The following examples are somewhat larger in scope and illustrate work flows that have proven useful.

Case One: Mapping Uncertainty

We are often interested in mapping over a large area with sparse well control, a number of seismic variables, and some conceptual geologic maps. One goal is to predict the best values of reservoir variables such as storativity (ϕh) and transmissibility (kh) and risk qualified values such as the P_{90} and P_{10} estimates. Geostatistics does not help much in a purely exploration setting; good geologic sense and past experience are required. A requirement of geostatistics is enough data to calculate correlations and perform statistical analysis; in the context of this example, a minimum of 5-10 wells are required.

The example presented here is fashioned after a number of real examples. Consider the 117 wells shown on Figure 6. The area is about 16 km by 19km. The top reservoir structure, the base reservoir structure and the thickness (which is directly derived from the top and base structure) will be considered as independent secondary variables. There are four variables we must map with uncertainty: gross pay thickness in the reservoir thickness, net pay thickness, net porosity, and water saturation in the net reservoir. The relationships between these variables are shown on Figure 7. The cross plots show one point per well. The cross plots are essentially unreadable; however, we look at them for strange values that do not belong, non-linear trends relationships and other features. These cross plots do not reveal anything unusual. The variables were transformed to Gaussian units and the matrix of correlation coefficients was calculated; see the

lower left table on Figure 7. Mapping of the four variables of interest should respect these correlations.

Mapping also requires measures of spatial correlation. The isotropic variograms for the Gaussian transform of the four variables are shown on Figure 8. The lack of short scale information is typical. Experience is required to infer $\gamma(\mathbf{h})$ for small $\mathbf{h}$. The intercept near $\mathbf{h}=0$ is low for most reservoir variables. As before, the required covariance values $C(\mathbf{h})$ are calculated as $1-\gamma(\mathbf{h})$.

There are a variety of geostatistical techniques to simultaneously account for multiple variables. A practical and robust Bayesian Updating technique will be illustrated here. The method is theoretically equivalent to the common collocated cokriging implementation (see Xu et al, 1992) with a Bayesian interpretation (see Doyen et al, 1996). As with virtually all geostatistical techniques, the data variables are transformed to be Gaussian, all calculations are performed in Gaussian units, and the results are back transformed. There are three steps to calculate the local uncertainty in a primary variable ($p=1,...,N_P$) at each location $\mathbf{u}$ in the area A:

1. Calculate the uncertainty in each primary variable at each location based on the available secondary data ($N_S(\mathbf{u})$ available at location $\mathbf{u}$):

$$\left.\begin{aligned} \overline{y}_{L,p}(\mathbf{u}) &= \sum_{i=1}^{N_S(\mathbf{u})} \nu_{i,p}(\mathbf{u}) \bullet y_{S,i}(\mathbf{u}) \\ \sigma^2_{L,p}(\mathbf{u}) &= 1 - \sum_{i=1}^{N_S(\mathbf{u})} \nu_{i,p}(\mathbf{u}) \bullet \rho_{S_i,p} \\ \sum_{j=1}^{N_S(\mathbf{u})} \nu_{i,p}(\mathbf{u}) \bullet \rho_{S_i S_j} &= \rho_{S_i,p},\ i = 1,...,N_S(\mathbf{u}) \end{aligned}\right\} p = 1,....,N_P, \forall\, \mathbf{u} \in A \qquad (1)$$

   The mean of the distribution of uncertainty is a linear combination of the available secondary data. The variance depends on how well correlated the secondary data are to the primary variable we are estimating. The weights used to calculate the mean and variance are established with the well known normal equations. We could interpret these distributions of uncertainty as likelihood distributions (denoted with an L).

2. Calculate the uncertainty in each primary variable based on the available surrounding well data ($N_P(\mathbf{u})$ available near location $\mathbf{u}$):

$$\left.\begin{aligned} \overline{y}_{P,p}(\mathbf{u}) &= \sum_{i=1}^{N_p(\mathbf{u})} \lambda_{i,p}(\mathbf{u}) \bullet y_P(\mathbf{u}_i) \\ \sigma^2_{P,p}(\mathbf{u}) &= 1 - \sum_{i=1}^{N_p(\mathbf{u})} \lambda_{i,p}(\mathbf{u}) \bullet \rho_{p,\mathbf{u}_i,\mathbf{u}} \\ \sum_{j=1}^{N_p(\mathbf{u})} \lambda_{i,p}(\mathbf{u}) \bullet \rho_{p,\mathbf{u}_i,\mathbf{u}_j} &= \rho_{p,\mathbf{u}_i,\mathbf{u}},\ i = 1,...,N_p(\mathbf{u}) \end{aligned}\right\} p = 1,....,N_P, \forall\, \mathbf{u} \in A \qquad (3)$$

   There may be some confusion between "P" for prior distribution and the use of "p" as an index for a primary variable. As with the likelihoods, the mean is a linear combination of the data and the variance depends on the spatial correlation. The weights are established by simple kriging. We could interpret these distributions of uncertainty as prior distributions (denoted with a P).

3. The likelihood and prior distributions must be merged to provide updated distributions of uncertainty that account for all secondary data and the nearby primary data. Bayes Law in a multivariate Gaussian setting leads to some very neat results. The equations to merge the likelihood and prior distributions of uncertainty is one neat result:

$$\left.\begin{aligned} \overline{y}_{U,p}(\mathbf{u}) &= \frac{\overline{y}_{L,p}(\mathbf{u})\bullet\sigma^2_{P,p}(\mathbf{u}) + \overline{y}_{P,p}(\mathbf{u})\bullet\sigma^2_{L,p}(\mathbf{u})}{\left(1-\sigma^2_{L,p}(\mathbf{u})\right)\left(\sigma^2_{P,p}(\mathbf{u})-1\right)+1} \\ \sigma^2_{U,p}(\mathbf{u}) &= \frac{\sigma^2_{L,p}(\mathbf{u})\bullet\sigma^2_{P,p}(\mathbf{u})}{\left(1-\sigma^2_{L,p}(\mathbf{u})\right)\left(\sigma^2_{P,p}(\mathbf{u})-1\right)+1} \end{aligned}\right\} p = 1,....,N_P, \forall\, \mathbf{u} \in A \qquad (5)$$

The updated distributions of uncertainty (denoted with a *U*) describe the local uncertainty at each location of interest.

These results are shown on Figure 9. Only the mean values are shown. The likelihood means are shown in the center, the prior means are shown to the left and the updated means are shown on the right. There are four rows: one for each variable under consideration. These results are like those on Figure 3 – they represent uncertainty in the Gaussian units. The uncertainty in original units is retrieved by back transformation. The updated maps largely reflect the prior maps; however, the influence of the secondary data can have a significant local influence.

For illustration purposes, the distributions of uncertainty were back transformed to original units; Figure 10 shows maps of the P_{90} low values and the P_{10} high values for each variable. The P_{90} low value map is used to identify areas that are surely high – where the P_{90} low value is high, then the variable is surely high. The P_{10} high value map is used to identify areas that are surely low – where the P_{10} high value is low, then the variable is surely low. P50 maps are not significantly different from maps of local mean values. These maps can be used for planning well locations.

Geostatistical simulation techniques could be applied to assess joint uncertainty. Simulation must be conducted jointly in space (as illustrated on Figure 5) and jointly between multiple variables. There are a number of cosimulation techniques for this purpose. Assessing uncertainty over a large area with multiple variables is an important goal in petroleum geostatistics. Constructing detailed 3-D models that reproduce all of the available data is another important goal.

Case Two: Data Integration

A reservoir model is required for resource assessment and to help with devising a reservoir management strategy. There is almost always seismic data and a limited number of well data. These data must be integrated under a sound conceptual geologic model. A small example leading to optimization-based well placement is developed.

Figure 11 shows a map of a seismic attribute and two wells: one well is in a higher quality area and the other is in a poorer quality area. There are a number of algorithms to simulate the variability of facies and petrophysical properties. Illustrations of facies modeling techniques are deferred to the next example. The porosity and permeability for this example are simulated using sequential Gaussian simulation (SGS). The SGS algorithm is as an industry standard algorithm that assumes a multivariate Gaussian distribution between all variables (seismic, porosity, and permeability) and all locations, after each variable is transformed to have a Gaussian histogram. A unique feature of the multivariate Gaussian distribution is the sole requirement for covariances or variogram values between all variables and locations. The variogram for the seismic attribute

(recall Figure 11) is not a problem since it is available (virtually) everywhere. The variograms of porosity and permeability are more of a problem.

The vertical variogram of porosity is well established by just two well data; however, it is not possible to calculate a horizontal variogram from the well data. The horizontal variogram is calculated from the seismic data. The normalized variograms are shown in Figure 12. There is a strong assumption that the seismic attribute is providing a good measure of continuity for porosity. This seems reasonable in this case given the close correspondence between the well average porosity and the collocated seismic value. Analogue data would be used if this was deemed unacceptable (see Chapter 4 of *Geostatistical Reservoir Modeling*). Details of permeability modeling will not be shown here because of space constraints. A conventional SGS cosimulation approach was used.

An additional parameter is required to use the seismic – the correlation between the seismic attribute (at a relatively coarse vertical resolution) and the porosity at a small scale. A correlation coefficient of 0.5 was chosen for the modeling. This led to a correlation of 0.8 between the vertically averaged porosity and the seismic attribute, which is considered reasonable.

A total of 50 realizations were created in flattened stratigraphic coordinates. Figure 13 shows a horizontal slice and two vertical slices through the first realization. The spatial variability away from the well locations looks natural relative to the well locations. A variety of uncertainty maps could be created. The realizations could be ranked by increasing oil in place and selected quantiles, say the P_{10}, P_{50} and P_{90} used in flow simulation. These applications are common practice in petroleum geostatistics. A slightly different application will be demonstrated here.

The placement and timing of production and injection wells is a significant decision in reservoir development planning. Well placement is difficult with a single deterministic model of reservoir structure and petrophysical properties. The duration of plateau production must be maximized, water handling must be minimized, recovery should be maximized and key economic indicators should also be optimized. These response variables depend in complex nonlinear ways on the placement/timing of the wells, their operating conditions and the subsurface reservoir description. Many of the interactions are resolved by a combination of sound engineering judgment and flow simulation. The problems associated with well placement become more complex in presence of multiple geostatistical realizations. The number of flow simulation runs becomes intractable and it becomes impossible to visualize all possibilities. There is a need for numerical measures to assist in the optimization of well locations to minimize risk and maximize reservoir performance.

The conventional approach is to consider a few *reasonable* well plans and perform some limited perturbations/optimization. This resulting solution is likely to be close to the optimal because experienced reservoir engineers can quickly rule out many bad well configurations. Nevertheless, there is still room for improving the well plan by considering an optimization scheme. The consequences of a minor improvement in the well plan are significant; an incremental improvement in recovery translates to a large monetary value. It is impossible to run flow simulation for many well configurations and geological models. Thus, the following idea is proposed:

1. Run some flow simulations with a number of different well configurations and different geological models. The optimization will be better with more flow simulations; 20 different models would provide a starting point. Summary flow response variables, such as discounted value of the production, are calculated from each flow simulation.

2. Propose static reservoir quality measures that capture the local *goodness* of the reservoir, for example, connected pore volume discounted by distance and permeability:

$$Q_S = \sum_{iw=1}^{nw} \sum_{j=1}^{n_{iw}} V_j \cdot \phi_j \cdot \left(1 - S_{w,j}\right) \cdot \left(\frac{1}{d_{j,iw}}\right)^{w_d} \cdot k_{j,iw}{}^{w_k} \tag{6}$$

   Where Q_S is the static reservoir quality for a particular well plan and geological model, nw is the number of producing wells, V_j is the volume of cell j that is close to well iw, ϕ_j is the porosity of cell j, $S_{w,j}$ is the water saturation of cell j, $d_{j,iw}$ is the distance of cell j to well iw, and $k_{j,iw}$ is the harmonic average of the permeability from cell j to well iw in the shortest path. In this formalism, we only need to calibrate ω_d and ω_k.

3. Calculate the static measures on the models used for flow simulation in step one. Calibrate the static reservoir measures to the flow response variables using classical multivariate statistical tools. This amounts to determine the ω_d and ω_k parameters that lead to the best correlation between the static reservoir quality and the flow performance. The result is a static reservoir proxy for flow simulation that can be optimized very quickly.

4. Optimize a set of well locations that maximize the calibrated measure of static reservoir goodness over all realizations accounting for their probability. This optimization can be repeated considering changes to the number of wells, initial configuration, and weighting of different factors.

5. Validate the results by performing flow simulation on the optimal configuration and reasonable alternatives proposed by the reservoir engineer *before* (and after) the optimization.

As with all numerical short cuts, there are limitations including: (1) it is difficult to capture the effect of timing and the incremental knowledge gained during the drilling program, (2) the physics of flow are not accounted for directly in the optimization, (3) the location of injectors and the pore volume replaced by water injection are not accounted for, and (4) the specifics of well completion are not accounted for. Notwithstanding the long list of limitations, there is significant value in the optimization. Attention is focused on what makes a well plan good and important features of the reservoir are better understood. Application to a number of reservoir studies indicates 1-5% improvements in hydrocarbon recovery.

Returning to the example, a number of flow simulations were performed with different cases. Three different well patterns were chosen: square, five-spot and random. Three realizations were chosen: P_{10}, P_{50} and P_{90}. Three different numbers of vertical wells were chosen: 10, 25 and 35. These decisions were made arbitrarily for the purpose of this example. In practice, the choices are made by the reservoir engineers with reservoir-specific considerations. The field oil production rates were discounted by 10% per year and a cumulative discounted oil production was calculated. A single number, summarizing the flow simulation results, characterizes each realization. 27 runs were used for calibration (based on three geostatistical realizations). The ω_d and ω_k parameters were optimized to be approximately 1.0 and 0.7, respectively. Figure 14 shows the results. The calibrated static measure very closely predicts the actual flow response. The results are robust in the sense that minor changes in the calibration parameters do not make significantly change this excellent correlation.

Optimization was then undertaken with 10, 20, 30 and 40 wells. The results are shown on Figure 15. The underlying color scale map is one of hydrocarbon thickness – red is high and blue is low.

This measure is closely related to the static quality defined above. An initial choice of well locations is iteratively perturbed to maximize the goodness of the well plan. The objective function is to maximize the goodness of the well plan, in terms of the static quality measure, in expected value over all realizations. Each realization is considered equally probable; however, a subjective weight could be assigned to favor specific realizations. The optimized well locations can be tweaked to account for complex considerations not easily coded in an optimization algorithm. Flow simulation with the optimized locations is required to verify that optimizing the static quality measure also improves the actual production.

The optimized well locations may not perform significantly better than those chosen by a thoughtful engineer; however, the advantages are the ability to optimize over a set of alternative geostatistical realizations with a repeatable quantifiable measure of goodness. There will always be multiple realizations when a development plan is being established – the uncertainty will diminish as additional wells are drilled, but it will not go away. Optimizing over multiple realizations and achieving well locations that are robust over the entire space of uncertainty is desirable. There may be a need to customize the optimization to adapt to reservoir-specific factors such as depletion/displacement mechanism.

This example highlights two points: (1) constructing reservoir models that integrate multiple types of data, and (2) using models of uncertainty for decision making.

Assessing uncertainty over large areas and constructing detailed 3-D models that reproduce all of the available data are important goals in petroleum geostatistics. Yet another goal is the assessment of the impact of heterogeneity on flow performance. Running flow simulation on realistic high resolution geostatistical models is a worthy goal.

Case Three: Models of Heterogeneity

A reservoir model is required input to flow simulation for forecasting and recovery calculations. Although flow simulation is performed in a minority of reservoirs, virtually all major developments rely heavily on flow simulation for planning and forecasting. There is almost always seismic data, a limited number of well data and some flow testing. Geostatistics does not work well if there are a large number of wells with extensive production history; the available techniques do not reliably account for extensive historical production data in the geostatistical reservoir models. There is a critical time, however, after expensive exploration and before significant production when a good geostatistical reservoir model combined with flow simulation can answer important questions and facilitate decision making.

The large scale structure of a reservoir including volumetrics and compartmentalization has the greatest affect on reservoir performance. The connectivity of high permeability pathways and low permeability barriers, however, can have a large affect on dynamic performance. Figure 16 shows some flow simulation results for production from a horizontal well. Oil production is shown at the top and the water cut is shown below. The flow behavior of a simplistic layercake model and an interpolated (kriged) model are not even in the range of uncertainty from the 20 geostatistical simulations. The affect of heterogeneity on flow predictions is variable.

There are times when heterogeneity has a dominant affect on flow performance and there are times when the heterogeneity averages out and the model behaves in a *heterogeneously homogeneous* manner. Running the flow simulator with a number of candidate heterogeneous models will establish the importance for a particular reservoir. There are a number of geostatistical tools aimed at creating heterogeneous models. Most geostatistical practitioners

agree that the major heterogeneity is captured in facies models. Facies are modeled as categorical variables by a variety of techniques based on cell, object, process-mimicking and multiple point statistics. Continuous properties like porosity and permeability are assigned once facies realizations have been constructed.

Figure 1 shows a schematic cell-based and object-based realization. Cell-based techniques are based on statistical controls such as indicator variograms. A sequential simulation approach is commonly followed where the cells are assigned a facies in a sequential approach considering the well data, seismic data, and previously assigned cells. Cell-based methods capture the coarse features, but they do not always appear realistic. Object-based techniques proceed by filling a model with facies objects. The shape, size, orientation and relationships between objects are chosen to appear realistic and match the available data.

It is becoming increasingly common to model facies with numerical techniques that have been adapted to mimic complex features of the depositional and diagenetic processes. These techniques are often referred to as process-mimicking techniques. They are hybrids of object based modeling, cell based modeling and depositional processes. Figure 17 illustrates two such models. Both models represent fluvial processes; however, the bottom illustration represents facies in a point bar setting with significant lateral accretion (Pyrcz, 2004).

As mentioned, a limitation of cell-based techniques is that they do not appear geologically realistic. A limitation of object-based and process mimicking techniques is a poor reproduction of all available data. A novel workflow has been proposed to overcome these two limitations. Firstly, object-based or process mimicking techniques are used to generate unconditional realizations, that is, realizations that reflect the spatial statistics, but not the well and seismic data. These unconditional realizations are called training images. Then, multiple point statistics (MPS) are extracted from these training images. Finally, sequential or iterative algorithms are used to generate realizations that reproduce the MPS and the well and seismic data. This is an active area of research.

There are two additional subjects that should be mentioned: scale-up and ranking. Geostatistical models should not be generated directly at the scale required for flow simulation. Virtually all geostatistical simulation algorithms assign facies and properties at the scale of the hard data, that is, the scale of the well data. These models must be scaled-up to more reasonably affect the less variable properties of flow simulation grid blocks. Scale up techniques are well established and outside the scope of this paper; those techniques work well if the geostatistical model has sufficient resolution, say, more than three simulated grid points per flow simulation grid block in each coordinate direction. The second additional subject to be discussed is ranking.

Geostatistics makes it relatively easy to generate multiple realizations. Most of the effort goes into setting up a reasonable workflow and establishing the required parameters. It is not possible, however, to process a large number of realizations through flow simulation. The computer requirements are prohibitive. It is common to rank the realizations from low to high based on a measure of static quality. The static quality could be the same as that used for well optimization (Equation 6). More complex ranking measures could be used; however, the calibration shown on Figure 14 is hard to improve upon. It is common to take the 10^{th}, 50^{th} and 90^{th} percentile realizations and process them through flow simulation.

A cohesive example has not been presented in this section. Glimpses of different techniques have been given. Heterogeneity is modeled by a set of realizations generated by a variety of statistical techniques. The realizations are often ranked to limit the number of realizations that have to be

used in flow simulation. Geostatistical techniques generate models at the small scale of the well data; they are scaled up to more accurately reflect the heterogeneity at the flow simulation scale. Assessing the affect of heterogeneity on recovery and flow performance is an important application of geostatistics.

When to Avoid Geostatistics

There are some inescapable realities of reservoir modeling: (1) data cleaning, formatting and checking take much longer than anyone believes possible, (2) bad data lead to a bad model regardless of the techniques used, (3) all numerical models depend on an appreciation of the features being modeled – understanding the geology is critical, and (4) reasonable models are only obtained if they are built for a particular purpose – there is no universal model or modeling approach. There are some additional sources for concern with geostatistical techniques.

There must be ground truth hard data for model construction. An essential feature of geostatistics is the calibration of extensive soft data with hard data, which results in an assessment of uncertainty at unsampled locations. Geostatistics should not be used when there are no hard data measurements of the variables being modeled.

Another essential feature of geostatistics is the exploitation of multivariate and spatial structure in measurements. At times, there is too little data or too widely spaced data to observe any structure. Perhaps this is the time when geostatistics is most needed to provide a quantitative measure of uncertainty; however, the results will be entirely model driven. Geostatistics should be avoided when there is too little data to make a meaningful decision of statistical populations and probabilistic predictions.

Geostatistical tools are inherently statistical – no physics or process information is embedded in the prediction. It is a mistake to apply geostatistics directly to variables such as pressure, flow rate or reservoir production particularly if the wells interact with each other. Geostatistics should be used to construct the input static models. Flow simulation, or some suitable process model, should be used for the dynamic predictions.

Geostatistics should not be considered when there is too little time and expertise to effectively apply the techniques and validate the resulting models. Geostatistical techniques are time consuming and finicky to apply. Slick demos by software vendors have not helped; they misrepresent the amount of time it really takes to construct a verified and useful set of geostatistical realizations. It takes time to establish a reservoir-specific workflow, choose reasonable modeling parameters, undertake reasonable sensitivity studies, verify the results, and apply the models to the problem at hand.

Conclusions

Geostatistical reservoir models are useful to transfer uncertainty in geological parameters through process evaluation to output uncertainty. Basic tools of the Monte Carlo simulation paradigm are adapted to spatially correlated variables. They are often applied hierarchically in an attempt to capture geologic structures and reproduce all of the available data.

What in the reservoir is geostatistics good for? Three examples were given. Firstly, the prediction of uncertainty in reservoir properties at unsampled locations using a large number of soft geophysical, geological and engineering data at many locations combined with a limited number of hard data at a few well locations. Secondly, the integration of seismic and well data

into plausible geological scenarios for resource assessment and well placement. Thirdly, the construction of realistic heterogeneity models for recovery predictions. The examples were fabricated to mimic the features of real reservoirs. The data and programs for all synthetic examples are available from the author.

References

- Deutsch, C.V. and Journel, A.G., 1998: GSLIB - Geostatistical software library and users guide. Oxford University press, 2nd Edition.
- Deutsch, C.V. 2002, *Geostatistical Reservoir Modeling*. Oxford University Press, New York.
- Doyen, P. M., L. D. den Boer and W. R. Pillet, 1996, Seismic Porosity Mapping in the Ekofisk Field Using a New Form of Collocated Cokriging, Society of Petroleum Engineers Paper Number 36498.
- Goovaerts, P. 1997, *Geostatistics for Natural Resources Evaluation*. Oxford University Press, New York.
- Matheron, G. 1971, *The theory of regionalized variables and its applications*. Les cahiers du CMM. Fasc. No. 5, Ed. Ecole Nationale Superieure des Mines de Paris, Paris.
- Pyrcz, M.J. 2004, *Integration of Geologic Information into Geostatistical Models*, Ph.D. Thesis, University of Alberta, Edmonton.
- Xu, W., T. T. Tran, R. M. Srivastava and A. G. Journel, 1992, Integrating Seismic Data in Reservoir Modeling: The Collocated Cokriging Alternative, Society of Petroleum Engineers Paper Number 24742.

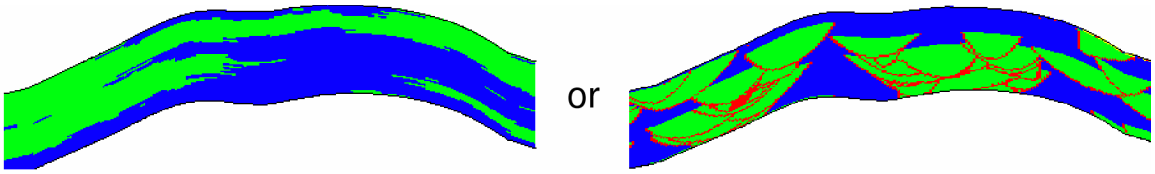

Figure 1: A schematic cross section through a cell based facies model at the left and an object-based model at the right.

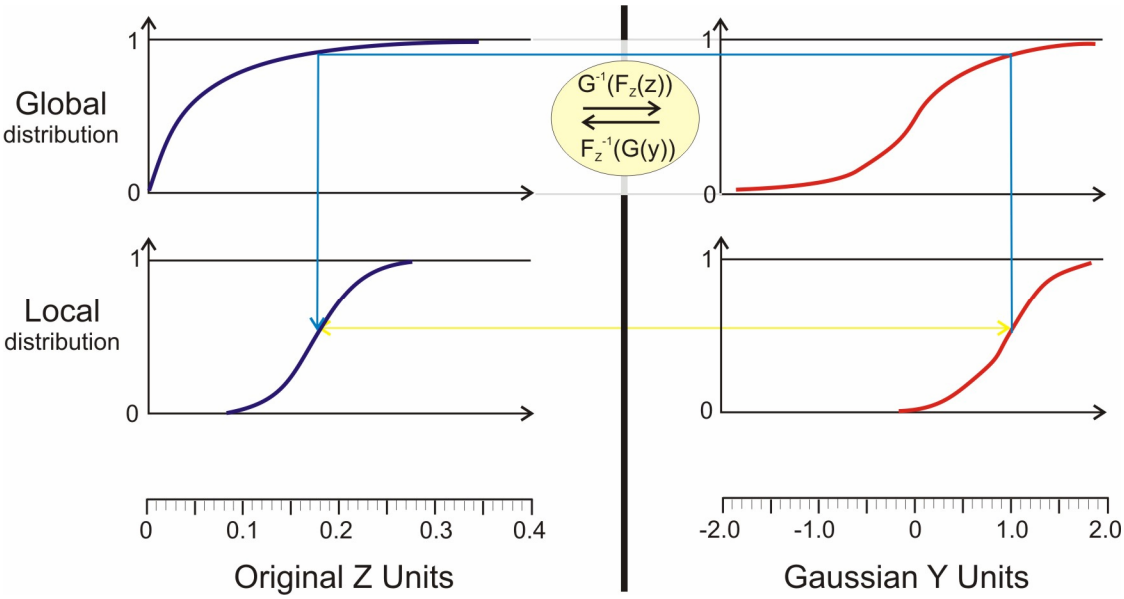

Figure 2: Schematic illustration of normal score transform. The original Z- data are on the left and the Gaussian Y-values are on the right. The top figures are the global CDFs and the bottom figures represent local CDFs. Quantiles are transformed using the global distribution (the three part blue line).

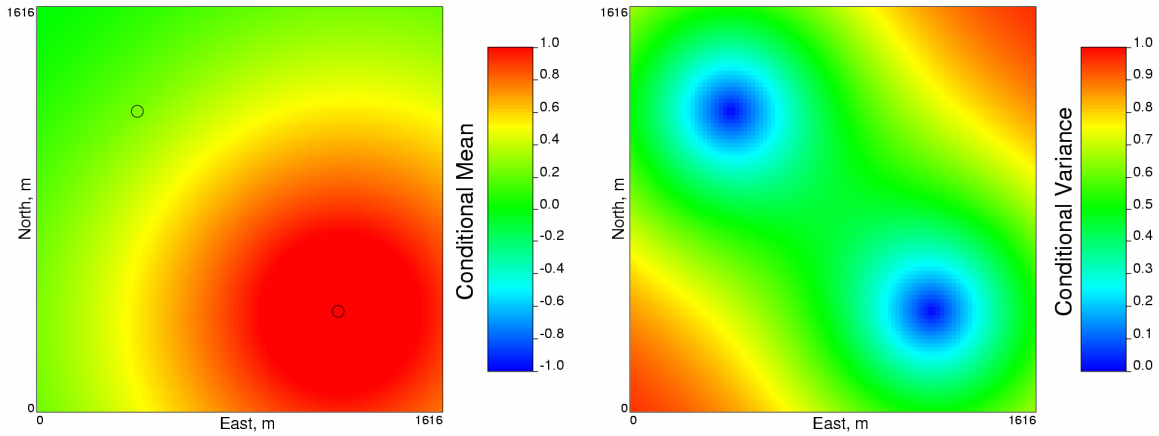

Figure 3: Map of the conditional mean (left side) and conditional variance (right side) for the Small Example.

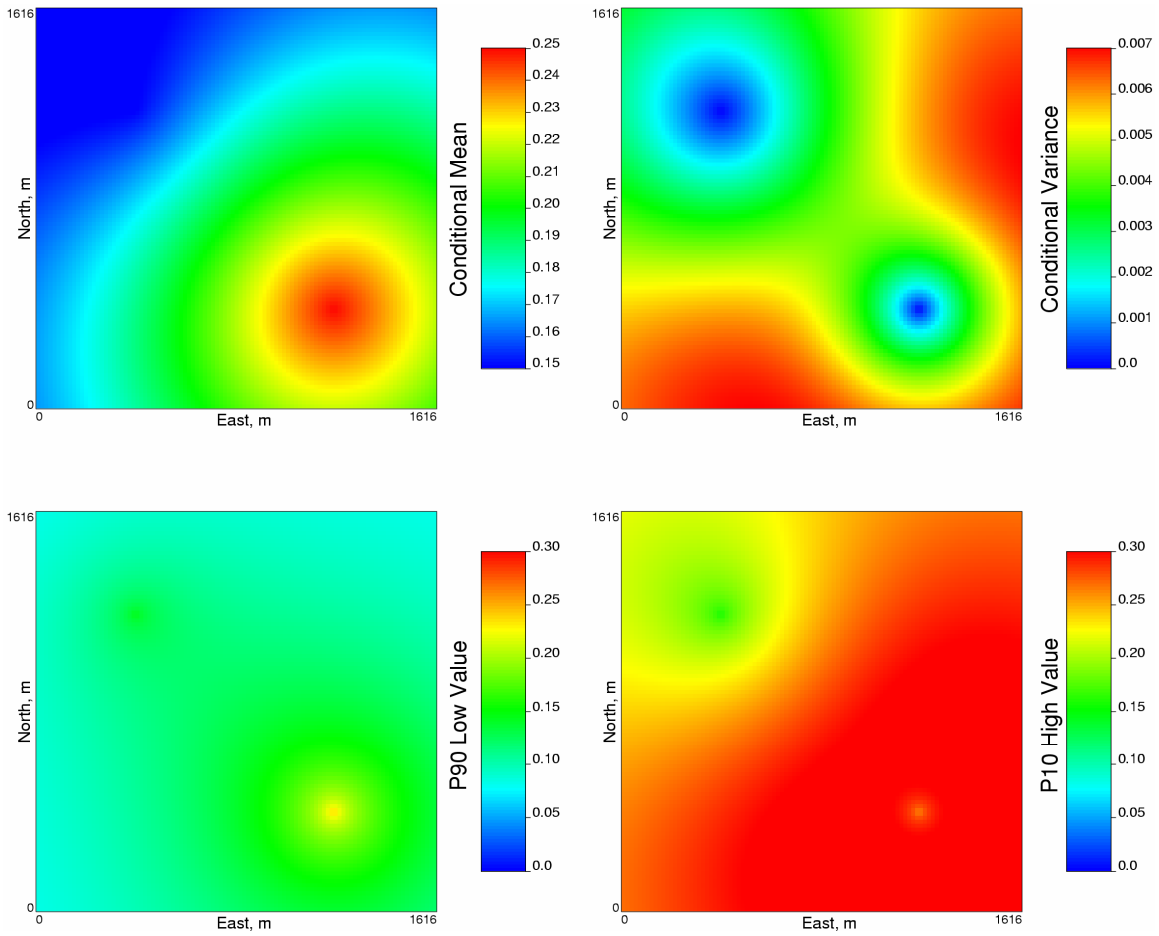

Figure 4: Map of the conditional mean (upper left) and conditional variance (upper right) in original units. Maps of the P90 low value and P10 high values are shown in the lower left and right.

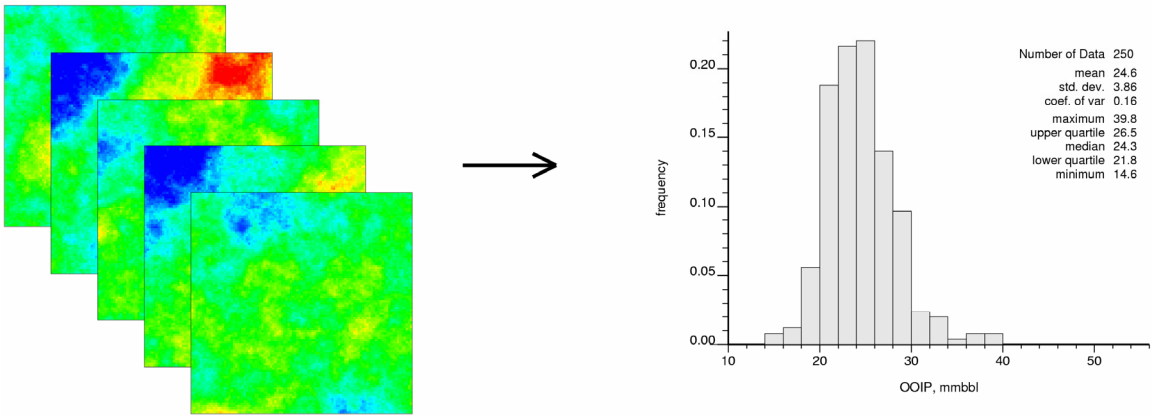

Figure 5: Multiple realizations (5 out of 250) are illustrated on the left and a histogram of the OOIP for the 250 realizations is shown to the right.

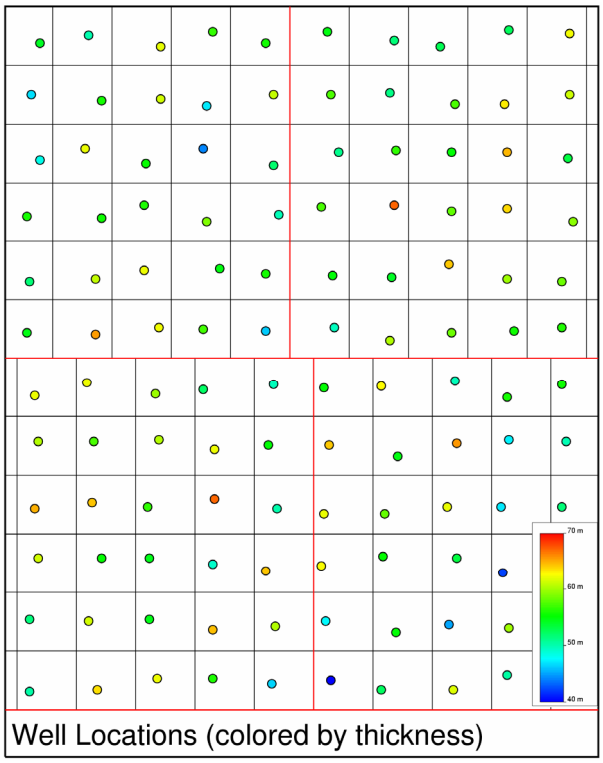

Figure 6: Location map of 117 wells for Case Study One. The wells locations are colored by gross reservoir thickness in meters.

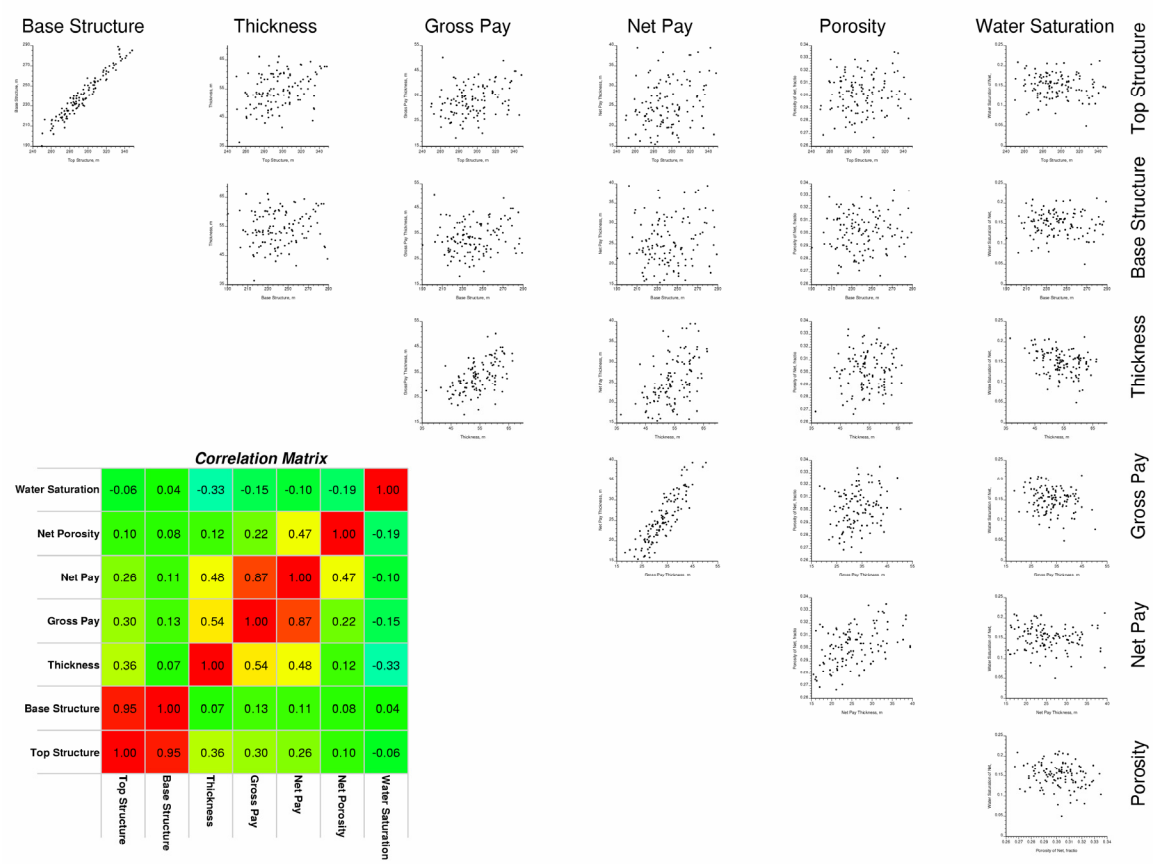

Figure 7: Bivariate summary of multiple variables. The cross plots are shown in the upper right and the matrix of correlation coefficients is shown in the lower left.

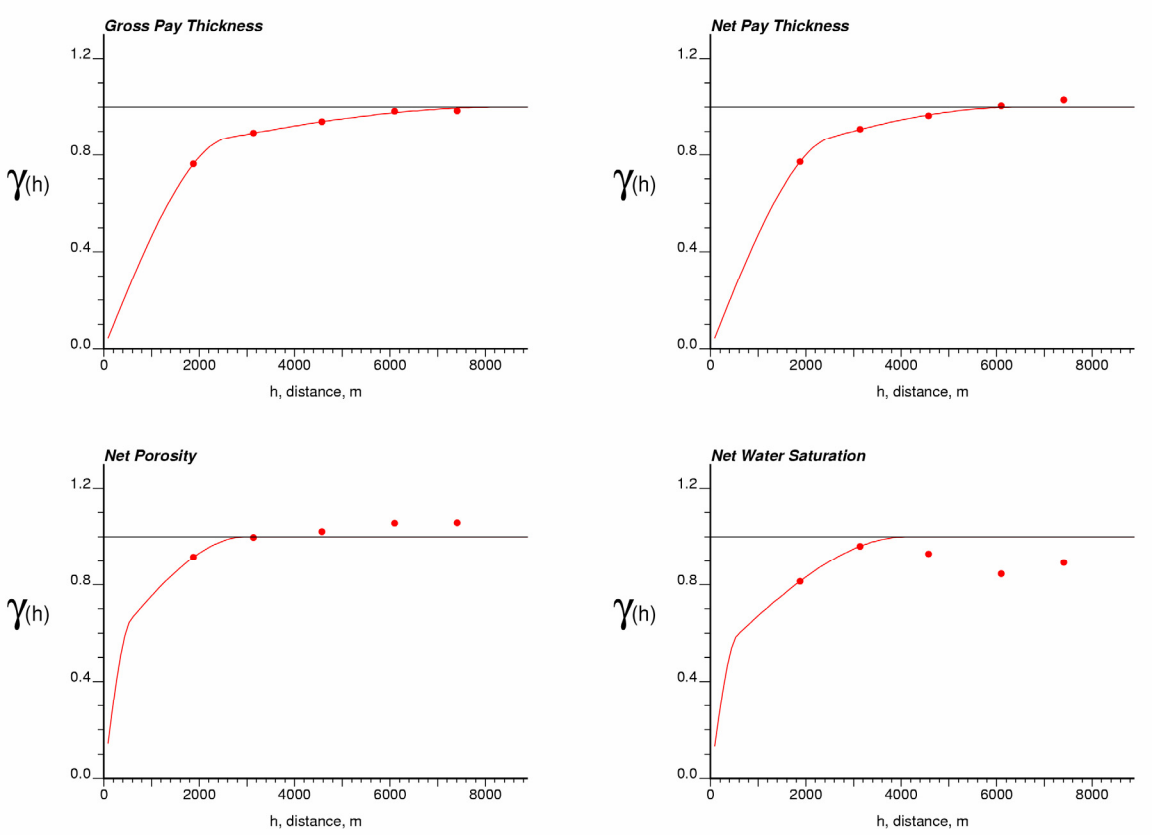

Figure 8: Variograms of the four variables under consideration. The dots are the calculated points and the solid line is the fitted curve. These variograms are isotropic.

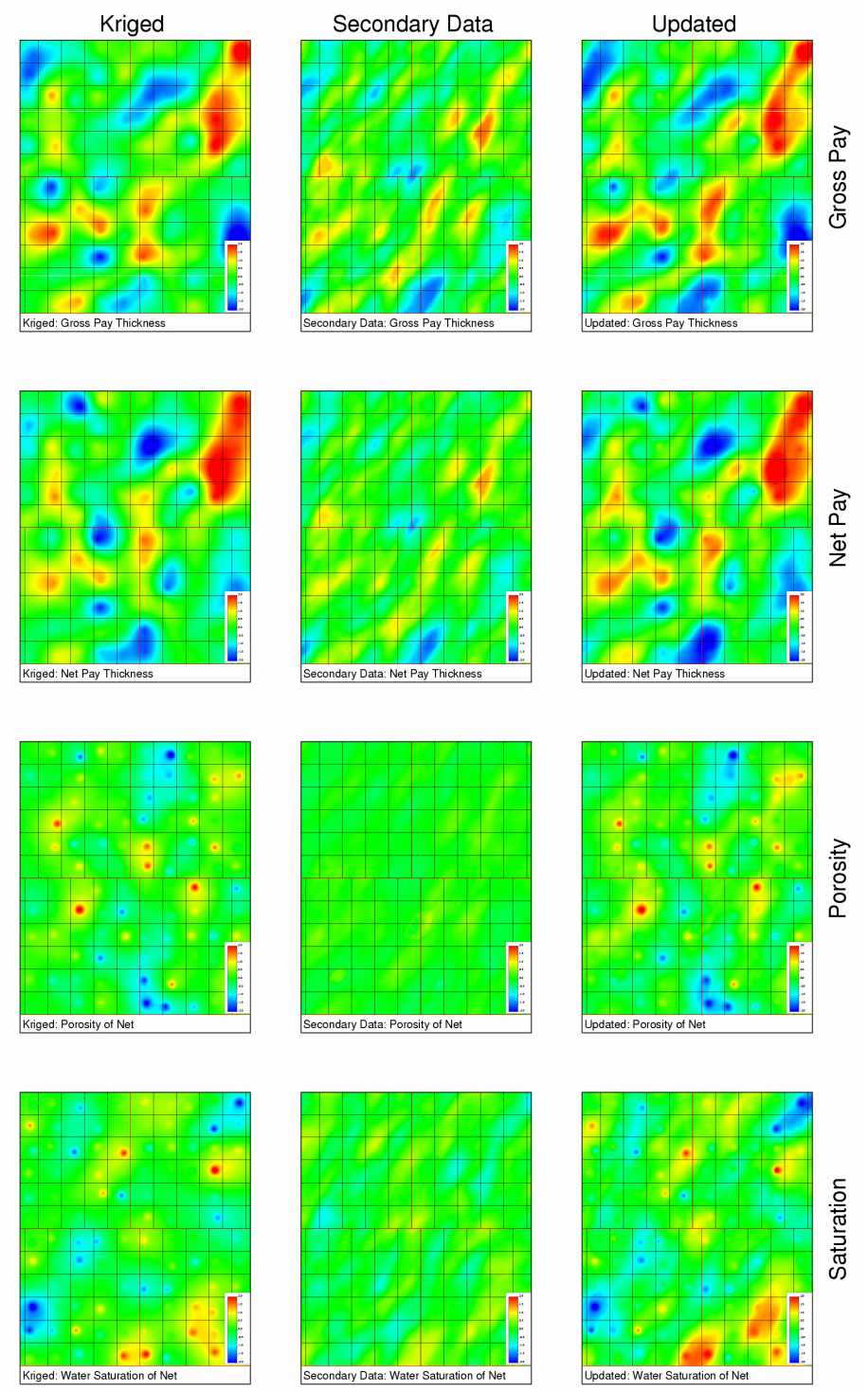

Figure 9: Each row shows maps of a variable being predicted. The map on the left is from well data alone, the center map is from secondary structural information, and the map on the right is updated map accounting for well and secondary data. The maps are in Gaussian units.

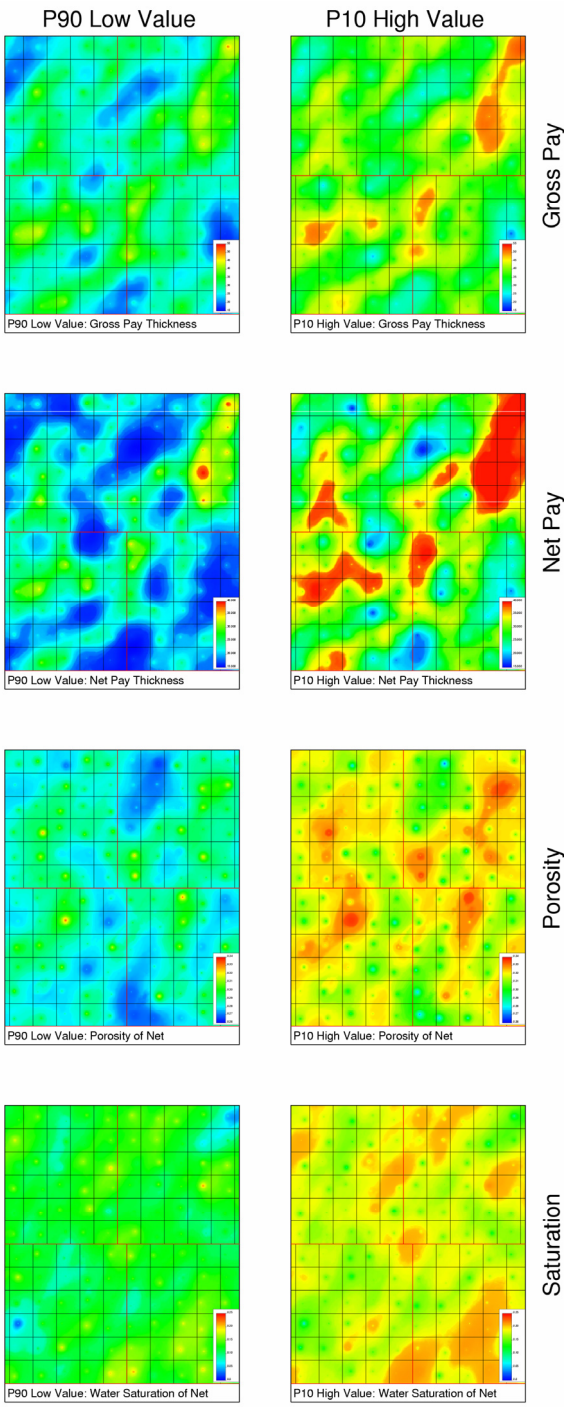

Figure 10: P90 low and P10 high values for each variable being predicted. The maps are in the units of the original data variables.

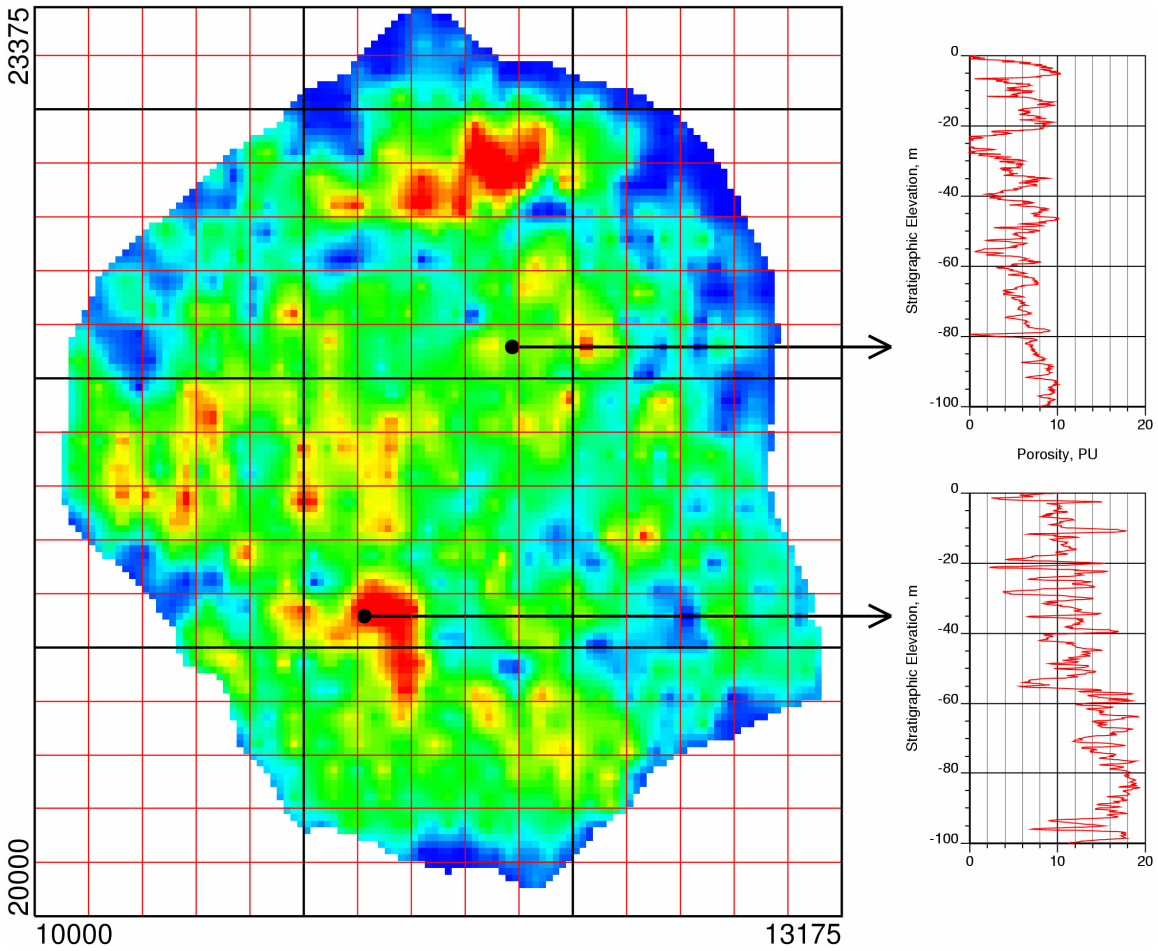

Figure 11: Seismic data attribute and two wells for Case Study Two. The bright colors on the seismic map are high porosity and the dark colors are lower.

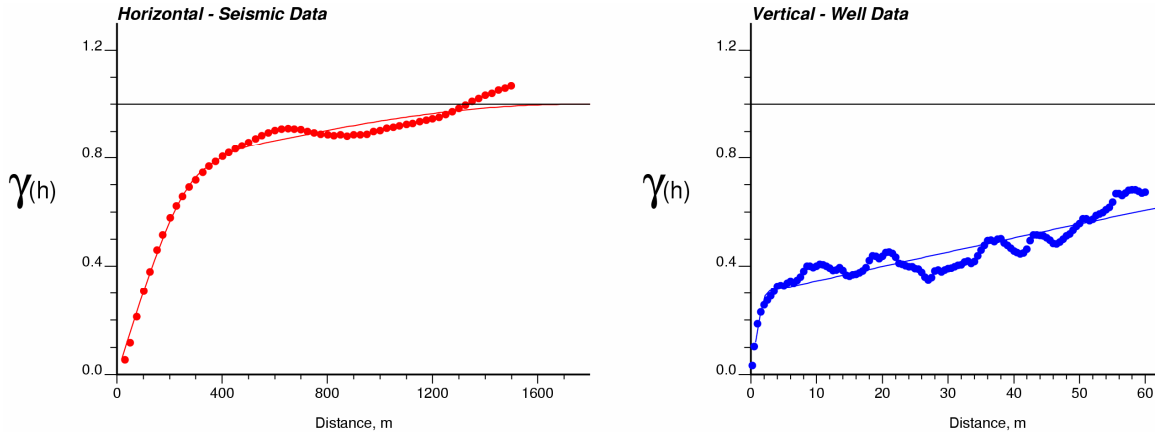

Figure 12: Horizontal and vertical variograms from the seismic and well data, respectively. The fitted variogram model toboth simultaneously.

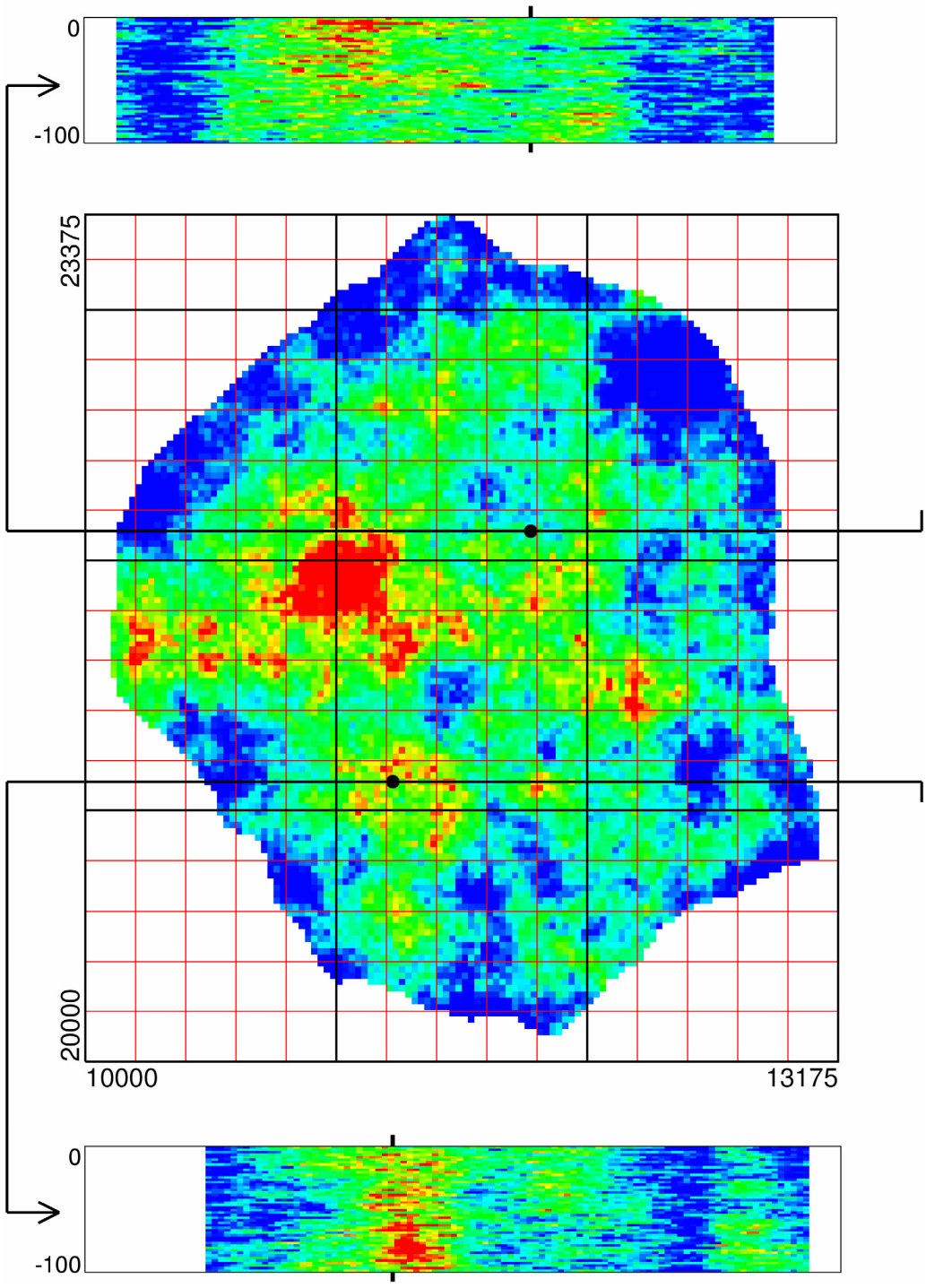

Figure 13: An areal slice and two vertical slices through an SGS realization of porosity. The two cross sections are through the well locations

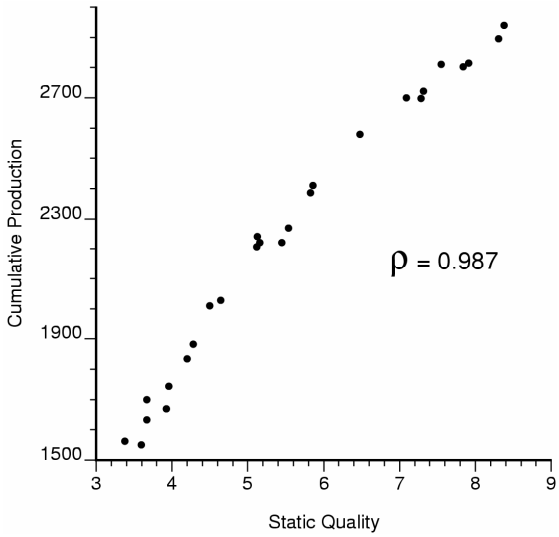

Figure 14: Calibration between calibrated/optimized static reservoir quality and discounted cumulative oil production.

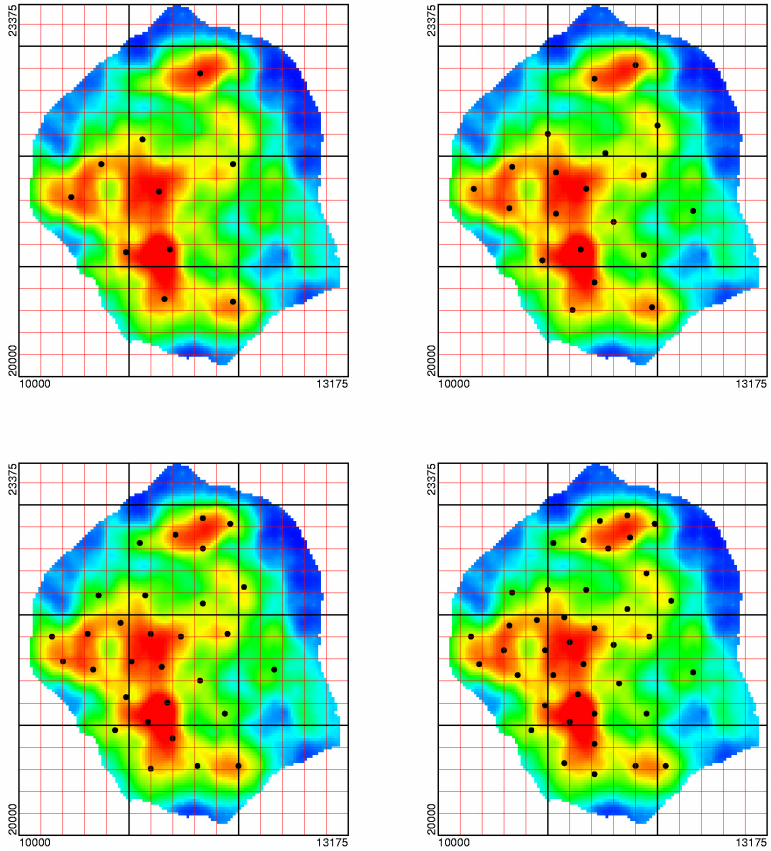

Figure 15: Optimized well locations for 10, 20, 30 and 40 wells. The underlying color scale map is the expected oil column thickness (average over all 50 realizations).

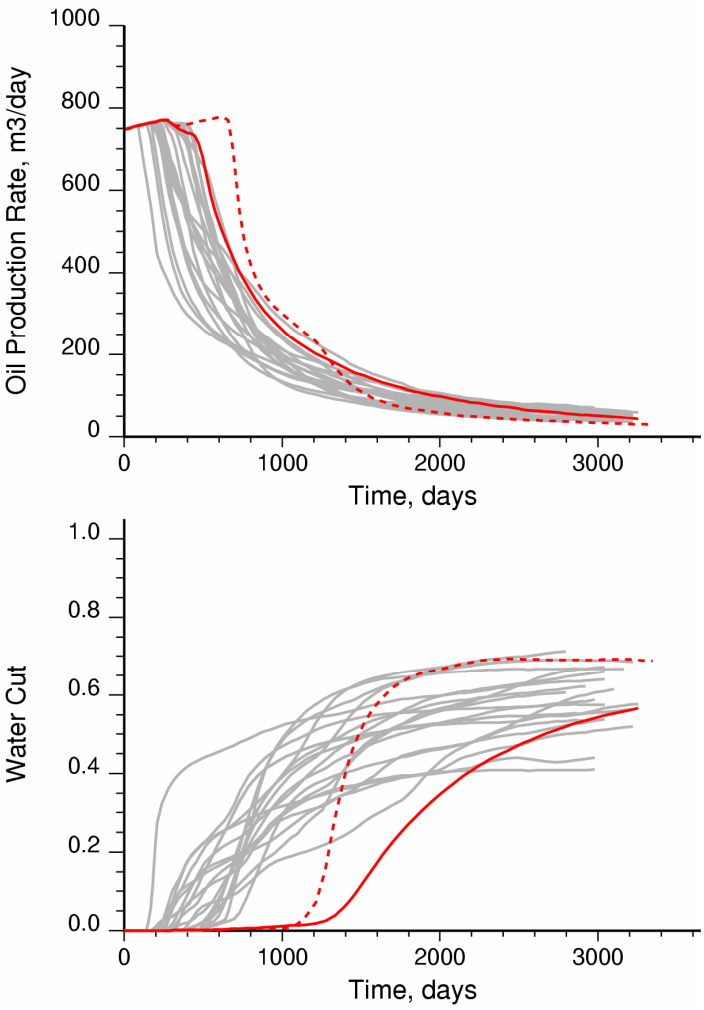

Figure 16: The solid red lines are the flow simulation results from a simplistic layercake model. The dashed lines are from an interpolated (kriged) model. The gray lines show the results from 20 geostatistical simulations.

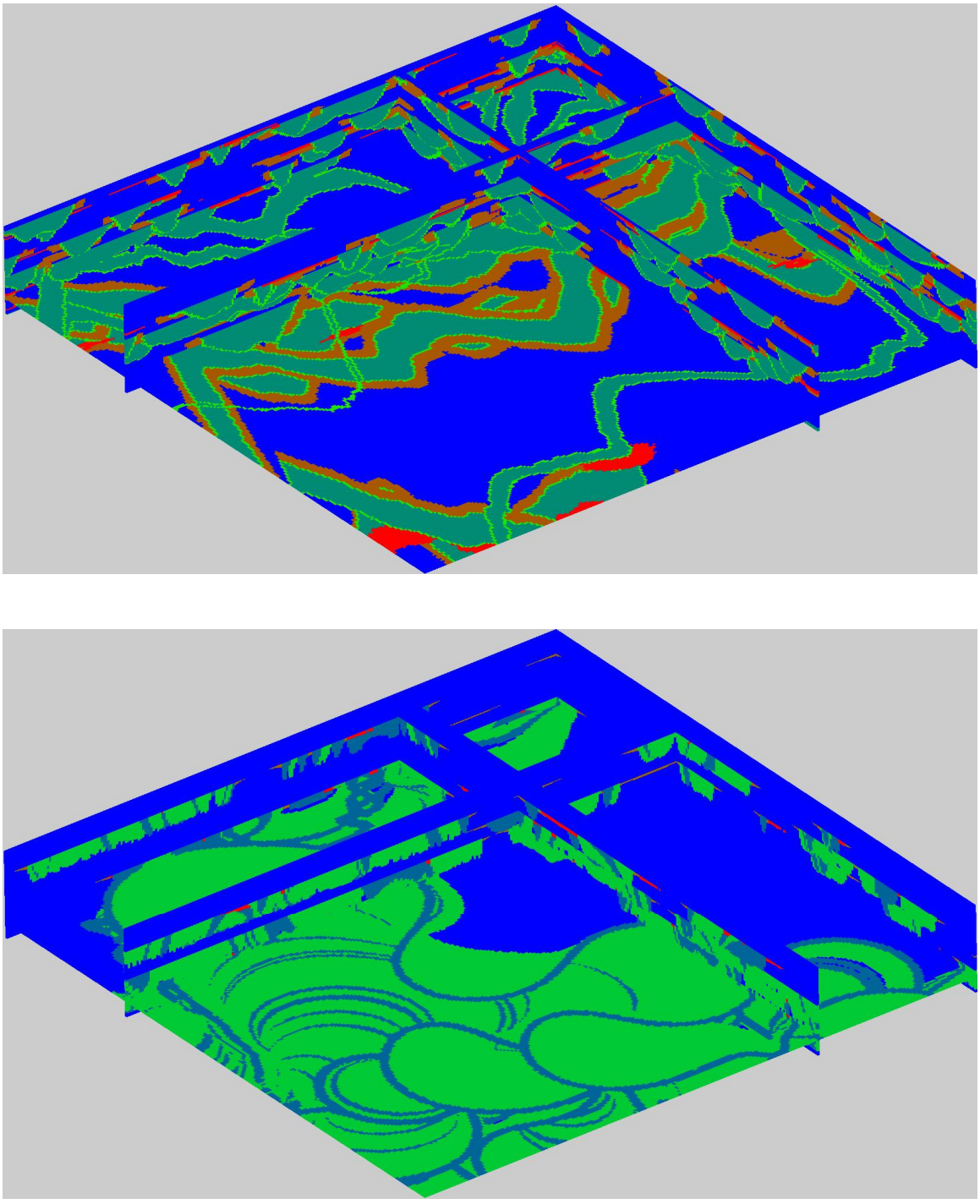

Figure 17: Two cutaway block diagrams of fluvial facies models constructed with process-mimicking techniques.