Introduction

This paper addresses the characterization of the global net-to-gross NTG (sand to non-sand ratio) of a reservoir, and the assessment of the corresponding uncertainty. The available data are a few wells, possibly biased towards the high pay zones, and a cube of seismic impedance.

In traditional geostatistics, the probability distribution for the global NTG is obtained by generating a set of $L$ simulated fields $G_i(s_k), i = 1, \ldots, L$ which honor a particular geological scenario $s_k$. Such a scenario combines the observed quantitative data $d_0$ and, traditionally, a variogram model inferred from these data. Most often, the simulated fields $G_i(s_k)$ are constrained by prior global distributions for the simulated values (interpolated and smoothed sample histogram, Gaussian distribution, Dirichlet distribution), which amounts to freeze somewhat the target value whose uncertainty is to be evaluated. Each of the simulated fields $G_i(s_k), i = 1, \ldots, L$ then yields a global NTG value $a_i^{(k)} = \varphi(G_i(s_k))$ computed from some algorithm or transfer function $\varphi(\bullet)$. This whole process can be repeated for $K$ different geological scenarios, e.g., having different variogram parameters. The distribution of the resulting NTG values $a_i^{(k)}, i = 1, \ldots, L; k = 1, \ldots, K$ is then used for making decisions.

Freezing the target statistics to some initial guess amounts to fix target values to global parameters whose uncertainty is precisely to be assessed. The resulting distribution is then but a measure of how imperfectly the simulation algorithm reproduces the target. Such approach is questionable, especially when the quantitative information used for estimating this target is sparse.

In this work, we propose an uncertainty workflow based on a Bayesian inversion, which does not start by freezing the global target. The main sources of uncertainty for early exploration data are described in Section 1. The uncertainty model for a given geological scenario is presented in Section 2. The incorporation of alternative geological scenarios into the uncertainty model is then discussed in Section 3.

1. Parameters of the uncertainty model

In any uncertainty modeling task, the prior determination and ranking of the main sources of uncertainty is possibly as significant as the implementation model itself.

In this proposal, the uncertainty model is built on a classical Bayes inversion approach, considering three jointly related sources of uncertainty and corresponding random variables:

A workflow to assess uncertainty in early hydrocarbon reservoir development

A. G. JOURNEL and G. CAUMON

Petroleum Engineering Dept, Stanford University, 367 Panama St, Stanford, CA 94305-2220, USA

To be presented at the EAGE Conference Workshop W6, 7th June 2004, Paris
the geological scenario $S$, the global NTG value $A^*$ estimated from the data available and, last, the true value $A$ itself.

**Notations**

Random variables (RV’s) modeling the uncertainty about related deterministic variables are denoted with capital letters, e.g., $A$ for the global NTG value. The $L$ realizations of a random variable are denoted by a set of indexed small letters, e.g. $a_l$, $l = 1, \ldots L$ for the random variable $A$. Vector random variables and the associated vector realizations are denoted in bold, e.g., $\mathbf{D}$ and $\mathbf{d}$ for the data, see hereafter.

1.1 Geological scenarios

In early exploration, the main uncertainty is related to the geological settings of the domain under study. We propose to represent this unknown by a set of geological scenarios $s_k$, $k = 1, \ldots K$, considered as the possible outcomes of a discrete random variable $S$. Each scenario $S = s_k$ is defined by:

- the vector of actually observed quantitative data $\mathbf{d}_0$;
- a particular geological interpretation of these data.

This interpretation can result in a 2-point variogram model or a prior conceptual image of multiple-point spatial patterns depicted by a training image $TI_k$. The uncertainty model presented hereafter is not specific to any particular representation of the spatial variability. We do, however, propose to use training images, since these can carry geological interpretation more comprehensively than variograms.

In what follows, we define the following equivalence of notations:

$$s_k \equiv \{\mathbf{d}_0, TI_k\}$$

The inference of a scenario from sparse exploration data is an expert, necessarily subjective process which has a significant influence over the uncertainty model. The consequent training images $TI_k$, $k = 1, \ldots K$ could be obtained by runs of process-based forward simulations (Bowman and Vail, 1999; Cross and Lessenger, 1999), or more simply by digitizing and interpolating sketches suggested by sedimentologists.

Each geological scenario $s_k$ is then numerically represented by a set of $L$ simulated fields $G_l(s_k)$, $l = 1, \ldots, L$. These fields need not be conditioned strictly to any prior estimate $a$ of the global value $A$, but their selection is done from the observed data $\mathbf{d}_0$. This last point is essential to our workflow: a scenario $s_k$ includes both the training image $TI_k$ and the data $\mathbf{d}_0$.

1.2 Estimation algorithm

The observed data set $\mathbf{d}_0$ is used to produce an estimate $A^* = a^*$ of the global value $A$. The estimation method can be represented by a function $\varphi(\mathbf{d})$ mapping any particular set of data $\mathbf{d}$ into a “best” estimate $A^* = a^*$. For NTG, this algorithm could for instance:

- average the observed sand / non-sand values at well locations;
- estimate the sand / non-sand indicator values on the domain, using some indicator co-kriging of well data and seismic impedance$^1$ (Goovaerts, 1997, p. 297);
- perform Bayesian inversion of impedance data based on well calibration (Fournier and Derain, 1995; Bitanov, 2003).

$^1$In the case of co-kriging, the transfer function $\varphi$ between the quantitative data $\mathbf{d}$ and the estimate $a^*$ accounts for variograms and cross-variograms which are specific to each scenario $s_k$. 

Note that the estimate \( A^* = a^* = \varphi(d) \) appears as a summary of the data set \( d \). One way to randomize the estimate \( a^* \) is to randomize the data \( d \) themselves. Therefore, we will use the probabilistic notation \( D \) for the data event, with \( D = d_0 \) denoting the actual quantitative data values observed. In our case, these data comprise both seismic impedances and well observations. \( D = d \) denotes an alternative data event, when some aspect of the data is changed, e.g., the well locations hence the well data values are changed while freezing the data size and drilling strategy. Randomizing the data \( d \) could have consisted also in considering alternative seismic processing and/or different combinations of seismic attributes. Randomizing the estimate \( a^* \) into a random variable \( A^* \) could call also for randomizing the estimation algorithm \( \varphi(*) \). In this paper, only the locations of the wells are varied, the estimation algorithm, the seismic data and the well drilling strategy are frozen. This particular decision does not affect the generality of the workflow proposed.

1.3 Unknown global value

Most error assessment algorithms, including bootstrap resampling (Efron, 1979), consider only a randomization \( A^* \) of the estimate \( a^* \). Instead, we suggest that the unknown global value \( A \) should be randomized as well, conditionally to each geological scenario \( S = s_k, k = 1, \ldots, K \).

Selecting a prior probability \( P(A = a) \) for \( A \) taking any value \( a \in [0,1] \) is difficult, since it should almost never be a uniform distribution. The idea of the proposed conditional randomization is to update such necessarily arbitrary prior probability by each proposed geological scenario \( s_k \).

2. Distribution of the true value \( A \) given the observed data \( D = d_0 \)

It would be illusory to look for some prior unconditional “objective” probability \( P(A = a) \). In practice, the determination of a distribution of the true value \( A \) would depend on the observed data \( D \) summarized by the “best” estimate \( A^* = \varphi(D) \), and on the geological scenario \( S \).

In early exploration, the only reliable source of information is the actually observed data \( d_0 \). Therefore, we propose to condition the distribution of the true value \( A \) to this information only:

\[
P(A = a \mid A^* = \varphi(d_0)) = \sum_{k=1}^{K} P(A = a \mid S = s_k, A^* = \varphi(d_0)) \cdot P(S = s_k)
\]

(1)

The main hurdle in equation (1) is the definition of a prior probability \( P(S = s_k) \) for each proposed geological scenario, probably the most consequential decision for the final result. Ideally, this scenario probability should be chosen independently of the observed quantitative data \( d_0 \). A convenient solution consists of taking a uniform distribution:

\[
P(S = s_k) = 1 / K, \quad \forall k = 1, \ldots, K
\]

One might consider decreasing the weight of very unlikely scenarios. This “unlikeliness” should not be determined solely from the quantitative data \( d_0 \), but also from some additional (expert) qualitative information, see Appendix B for further discussion.

The novelty of our proposal is to displace the somewhat arbitrary decision about a prior probability for \( A \) into that for the geological scenario \( S \). Geological expertise could give \( P(S = s_k) \), it generally cannot give \( P(A = a) \).
Each of the $K$ conditional probabilities under the right hand side summation of expression (1) need now be estimated, see next expression (2).

### 3. Distribution of the true value $A$ given the data $d_0$ and a geological scenario $S = s_k$

We propose to compute the probability for the global property $A$ to take any value $a$ given a specific geological scenario $s_k = \{d_0, TI_k\}$ using the Bayesian inversion:

$$P(A = a \mid S = s_k, A^* = \varphi(d_0)) = \frac{P(A^* = \varphi(d_0) \mid S = s_k, A = a) \cdot P(A = a \mid S = s_k)}{P(A^* = \varphi(d_0) \mid S = s_k)}$$

(2)

The evaluation of the various right-hand terms appearing in formula (2) is now described, see also Figure 1.

#### 3.1 Prior probability $P(A = a \mid S = s_k)$

This prior probability should be assessed from the available data $d_0$ and their interpretation under a given geological scenario $S = s_k = \{d_0, TI_k\}$, see 2nd row of Figure 1.

For any prior geological scenario $S = s_k$, one can simulate, conditional to the observed data values $d_0$ at their actual locations, a series of $L$ fields and observe the resulting simulated distribution $P(A = a \mid S = s_k)$ of the global property $A$. These simulations should not start by freezing the expected value of $A$ to $a^* = \varphi(d_0)$ or to any other estimated value.

If reproduction of multiple point spatial statistics is called for, then a simulation algorithm such as per the program snesim (Strebelle, 2002) could be used. If the training images are reduced to variogram models, traditional algorithms such as sequential Gaussian or indicator simulation would suffice (Goovaerts, 1997).

**Remarks**

- Let $A$ relate to a global NTG ratio. The NTG ratios of the $L$ simulated realizations will all be somewhat different from the NTG of the training image $TI_k$, because of conditioning to the observed data $d_0$ and fluctuations from realization to realization. It is the distribution of the simulated NTG ratios which is taken as prior probability $P(A = a \mid S = s_k)$ in the Bayesian inversion (2).

- The resulting prior probability distribution $P(A = a \mid S = s_k)$ depends upon the simulation algorithm chosen and its input parameters. Such a subjective algorithmic decision is inherent to any uncertainty model. Our proposal accepts any simulation algorithm as long as it produces realizations honoring the data $d_0$ and the prior conceptualization of spatial patterns $TI_k$ produced by expert geological interpretation of these data (see Appendix A for further discussion).

- However, leaving to an algorithm the determination of the prior probability may be questionable; one may instead leave the burden of an explicit definition of $P(A = a \mid S = s_k)$ to the sedimentologists.
Initial data $D = d_0$

$K$ different training images

Scenario $S = s_k = \{ d_0, TI_k \}$

$L$ simulations

$N$ resamplings

Figure 1: Flow chart of the simulations needed to evaluate Equation (2).

3.2 Likelihood probability $P(A^* = \varphi(d) \mid S = s_k, A = a)$

The $L$ simulated fields as obtained from Section 3.1 can be classified into $M \ll L$ classes of global simulated property value $a_1, a_2, \ldots, a_M$. For any class $a_m, m = 1, \ldots, M$, one or more representative simulated fields is retained to resample the data set $D$, see last row of Figure 1. Resampling the data can for instance be achieved by moving the data configuration over the field (spatial bootstrap, Journel, 1993); any other relevant resampling procedures can be considered.

Applying the estimation method $\varphi$ repetitively on the various resampled data $D = d$ gives the likelihood probability used in expression (2):

$$P(A^* = \varphi(d) \mid A = a_m, S = s_k)$$

3.3 Estimator distribution $P(A^* = \varphi(d) \mid S = s_k)$

The denominator in expression (2) can be calculated by summing up the previously obtained prior probabilities

$$P(A \in a_m \mid S = s_k)$$

and likelihood probabilities.
\[ P(A^* = \varphi(d) \mid S = s_k, A \in a_m) \]

which gives:

\[
P(A^* = \varphi(d_0) \mid S = s_k, ) = \sum_{m=1}^{M} P(A^* = \varphi(d_0) \mid S = s_k, A \in a_m) \cdot P(A \in a_m \mid S = s_k) \tag{3}\]

3.4 The final result

Expression (2) gives the uncertainty model for any given scenario \(S = s_k\). These results are then combined into the final uncertainty model expression (1).

4. Preliminary summary

As opposed to bootstrap resampling (Efron, 1979) or the Dirichlet formalism (Haas and Formery, 2002), the proposed model of uncertainty (1) accounts for spatial correlation and spatial patterns between data, as delivered by training images (Journel, 1993; Bitanov, 2003).

The novelty of the proposal (2) is the Bayesian updating of the prior probability \(P(A = a \mid S = s_k)\) by the specific estimation process \(A^* = \varphi(d_0)\) retained. Note that the estimation algorithm \(\varphi(\bullet)\) itself could also be a significant source of uncertainty. Therefore, the uncertainty model could be made more complete by randomizing \(\varphi(\bullet)\). Also uncertainty about the data values \(d_0\) themselves could be accounted for by randomizing the seismic data acquisition and processing.

The critical prior probability \(P(A = a)\), of traditional approaches, too often casually assumed uniform, is replaced by the probability \(P(A = a \mid S = s_k)\) obtained from the fluctuations of the \(L\) simulated global values \(a_1, a_2, \ldots, a_L\), these simulations being constrained by a specific geological scenario \(s_k \equiv \{d_0, TI_k\}\). The arbitrariness of the prior \(P(A = a)\) is replaced by that of the prior \(P(S = s_k)\). We believe the latter to be easier to get from geological expertise.

References


**A. Accounting for algorithm-specific variations**

The input to our uncertainty model can be split into the observed quantitative data values $d_0$, and the qualitative interpretation of these data resulting in a spatial variability model represented by a set of alternative training images $TI_i$, $i = 1, ..., K$.

The training image brings a quantitative yet visually explicit representation of the qualitative geological interpretation. In the case of NTG, each specific training image $TI_i$ carries a global NTG. Any stochastic simulation can be constrained to honor a global target value. Removing or relaxing this constraint may conveniently provide us with the prior distribution $P(A = a | S = s_i)$ used in equation (2).

This latter distribution is clearly algorithm-specific. The spread of the distribution comes from the inability of the simulation algorithm to reproduce exactly the NTG and spatial patterns of the training image. Changing the size of the data conditioning window or the size of the training image are likely to result in significant variations of the prior distribution $P(A = a | S = s_i)$. This is an inevitable drawback associated with the challenge of evaluating a distribution for an entity $a$ (the true global NTG) which is essentially unique and unknown (Journel and Huijbregts, 1978, p. 410).

One might be interested in separating the algorithm-specific variations from the data-specific variations. For example, run unconstrained simulations, to get the distribution $P(A = a | TI = TI_i)$, and compare these results with $P(A = a | S = s_i)$ obtained by constraining the simulations to the data $d_0$.

**B. Assessing the probability distribution of geological scenarios**

Determining the prior probability of a given geological scenario $S = s_i$ is a very difficult task, yet extremely consequential for the final uncertainty model. Consider the case where a geological scenario is simulated by a process-based forward modeling algorithm. One might want to define simple probability laws for the input parameters of that algorithm (subsidence rate, distance to sediment source, duration of deposit, etc.), and come up with a good alternative to the bland unrealistic uniform distribution suggested for expression (1).