Conditioning a surface-based model to wells and thickness data.

Antoine Bertoncello, Jef Caers, Tao Sun, Hongmei Li and Gregoire Mariethoz

Abstract

Surface-based models imitate a sequence of depositional events in time. By considering sedimentation processes, these algorithms produce highly realistic subsurface structures from a variety of environments. However, since depositional events are forward-modeled, they cannot be directly conditioned to data. Instead, conditioning requires inferring the right combinations of parameters that match data. Solving such inverse problem is time-consuming and expensive because those models are complex and highly parameterized. In this context, we propose an optimization methodology that enables to fit data efficiently. The key idea behind this method is to decrease the dimensionality of the problem. Three complementary approaches are developed for this purpose. The first approach aims at identifying the leading uncertainty. The second one is a sensitivity analysis on the input parameters. The third one is a re-formulation of the full inverse problem. This methodology is general in the sense that it can be used with any kind of data and environment of deposition. In our work, the efficiency of the methodology is illustrated by using a forward model reproducing lobe structures and a set of borehole data and thickness information extracted from a realistic turbidite reservoir.

1. Introduction

Turbidites are crucial hydrocarbon reservoirs and building their 3D numerical representation is critical for reservoir development (Weimer & al., 2000). Currently, however, no single technique in reservoir characterization is able to satisfactorily reproduce the complex geological structures of those depositional systems (Michael et al., 2009). Traditional geostatistical methods are convenient for integration of well and seismic data but they are not able to produce realistically the complex interactions between the various architectural elements observed in such depositional systems. Physics-based models (also termed process-based models) allow simulating highly realistic reservoir architectures but their use is limited due to the computational burden imposed by the technique (Miller, et al., 2008). To overcome those two issues, a new family of stochastic models, termed surface-based models or event-based models, has recently been developed (Pyrcz, et al., 2004; Pyrcz, & Strebelle, 2006; Miller, et al., 2008; Biver et al. 2008; Zhang et al., 2009; Michael et al., 2009). These models do not fully reproduce the sedimentological processes but mimic them with considerable realism, producing key geological features at a fraction of the process-based models CPU cost.
However, at present, conditioning surface-based methods to reservoir data remains an issue. Because of the complexity of these models, they are difficult to directly constrain to data, as many geostatistical methods do by construction. It is indeed not possible to generate a conditional model by deduction or derivation of the observations. To overcome this problem, different approaches have been developed. Pyrcz et al. (2005) propose a method based on a local adjustment of the lobe surface to fit the neighboring well data. If the surface geometry contradicts data outside a tolerance, the geometry is rejected. The main limitation of such approach is the computational time associated with the rejection method. The approach based on direct interpolation between logs used by Zhang (2009) necessitates a large amount of wells, which is unrealistic in deep offshore development. Michael et al. (2010) propose a method for conditioning using well correlation. Each of the depositional units needs to be identified in well data and ordered according to the different depositional periods. However, interpreting wells is time consuming and subject to uncertainty. In addition, the interpretation correlation needs to be consistent with the forward stratigraphic model.

A possible solution for conditioning is to use an iterative approach, where the output model is compared with the data and the input parameters updated in accordance to the mismatch. The solution of such inverse problems has been extensively studied in 2D stratigraphy modeling (Charvin, et al., 2009) but the developed methods are not tailored to fit efficiently reservoir-type data, and therefore cannot be easily applied in a reservoir modeling workflow. Achieving an efficient match requires the development of a robust optimization scheme.

The contribution of this paper is to show that conditioning through optimization is possible. Although our approach is general, we will focus our work on modeling lobe architecture of turbidites. The first section describes surface-based models and the challenges of conditioning them to data. The second section discuss the implementation of an optimization scheme. A real example aims at demonstrating the efficiency of the method.
Fig. 1. Surface-based modeling workflow. First, the forward model computes over the full domain a 2D thickness map of the lobe. Then, this thickness map is used to generate the 3D structure of the lobe. All the generated lobes are sequentially stacked to produce the final 3D earth model.

2. Methodology

2.1. Surface-based model concepts

In this part, we present the surface-based model used in the paper. The algorithm, based on Michael’s work (Michael et al., 2010), mimics the deposition of lobes. For a given topography, the model starts by generating a new lobe with lobe length, width and height drawn for inputs distributions (1). Then, its location is selected according to the deposition rules defined in the model (2). The new lobe is stacked on top of the current topography. The lobe’s top surface is merged with this topography. This new surface becomes then the current topography (3). The process is repeated until the reservoir is filled.
2.1.1. Lobe generation

In a surface-based approach, geobodies such as lobes are simulated by using a predefined geometry template. The actual size of the lobes is drawn from probability distributions. Based on this shape, a 2D thickness property is defined such that the center of the lobe reaches a maximum thickness, creating a 3D structure. This idealized thickness map is stochastically perturbed by adding a Gaussian correlated noise generated with a Sequential Gaussian Simulation, (SGS, Goovaerts, 1997). The noise models the small scale variability of the lobe’s structure (Fig. 2).

![Initial lobe geometry + Gaussian noise to perturb lobe thickness = Final lobe](image)

**Fig. 2.** The initial thickness of a lobe is perturbed by adding a Gaussian noise. It allows reproducing small scale variability of the lobe structure.

The location of a lobe is randomly drawn from a probability map (Fig. 3). Two sets of rules are used to compute the map. Both of them aim at reproducing the tendency of geobodies to fill topographic lows. The first approach computes probabilities based on topographic elevation and distance between lobes. Those rules are obtained from an existing process-based model representing a typical lobe system (Michael, et al., 2010). The second approach is only based on topographic features. Both approaches produce a different probability map. Those two maps are then combined in a single one, using the tau model (Journel, 2002). The variation between 0 and 1 of the tau value emphasizes more or less one of the probabilities, hence one of the depositional rules. With the first approach, the distance between a new lobe and the previously simulated one is partly controlled by predefined statistics. As a consequence, both of the lobes tend to be more closely together. With the second approach, the distance between the two lobes is only function of the topography, leading to a higher degree of freedom in the lobe placement.
Fig. 3. Each of the two deposition models produces a different probability map of lobe locations. Using the Tau model, it is possible to combine those two probability maps into a single one (Journal 2002). The Tau value controls the relative importance of each model in the final map. It influences therefore the stacking patterns of the lobes. (CDF=cumulative distribution function)

2.1.2. Lobe stacking and erosion

The generated geometry represents the thickness of the lobe. Each generated lobe needs to be stacked on top of the depositional surface to generate the spatial architecture of the lobe. At this stage, another process that needs consideration is the erosion created by the deposition of the geobody (Fig. 4). Erosion may modify indeed the thickness of the underlying geobodies and create some connected flow paths by eroding flow barriers that separate lobe structures. In our model, the intensity of erosion is assessed by the curvature and gradient of the deposition surface. Locations with high gradient magnitude will be eroded more those with low gradient, since the flow energy is assumed to be lower in region with low gradient. The curvature profile is used in such a way that positive curvature indicates less erosional power than negative curvature, point at which the flow of sediments collides into surfaces barriers.
Fig. 4. The erosion process associated with a lobe’s sedimentation is computed in four steps. The first one consists on delimiting the region where the geobody is being deposited. Then, based on the geological model, the intensity of erosion is computed. An erosion surface is then created by modifying the geometry of the underlying geobodies. Lastly, the new lobe is deposited on top of this erosion surface. In general, the erosion process is more pronounced where the topography has a positive curvature, a high slope and a high elevation.

2.1.3. Input parameters

In order to perform a single forward simulation, three sets of parameters are required (Table 1). The first set controls the shape of the lobes. The second describes the geological processes. The final set models the initial condition of the system. All those parameters are uncertain and represented in the form of probability distributions.
Table 1. Summary of the rules and parameters needed to run a simulation (see also Michael et al., 2009). The parameters are highly uncertain because difficult to infer. They are represented in the form of probability distributions. To generate a realization, the parameter values are drawn from these probability distributions. (CDF=cumulative density function)

<table>
<thead>
<tr>
<th>Parameters and rules</th>
<th>Representation</th>
<th>Uncertainty</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Lobe geometry</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-Template geometry</td>
<td>Deterministic equation</td>
<td>Uncertain because inferred from analog models and/or data. The lobe geometry is modified by an erosion process that is difficult to quantify.</td>
</tr>
<tr>
<td>-Width</td>
<td>CDF</td>
<td></td>
</tr>
<tr>
<td>-Length</td>
<td>CDF</td>
<td></td>
</tr>
<tr>
<td>-Thickness</td>
<td>CDF</td>
<td></td>
</tr>
<tr>
<td>-Variance/Covariance of the perturbation</td>
<td>CDF</td>
<td></td>
</tr>
<tr>
<td><strong>Geological process</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-Progradation intensity</td>
<td>CDF</td>
<td>Abstraction of complex physical processes into simplified rules leads to an additional error or uncertainty. The statistics are inferred from process-based models analysis.</td>
</tr>
<tr>
<td>-Migration intensity</td>
<td>CDF</td>
<td></td>
</tr>
<tr>
<td>- Influence of topography on deposition</td>
<td>Deterministic set of rules</td>
<td></td>
</tr>
<tr>
<td>-Tau value</td>
<td>CDF</td>
<td></td>
</tr>
<tr>
<td>-Frequency of intermediate shale</td>
<td>Deterministic set of rules</td>
<td></td>
</tr>
<tr>
<td>-Erosion intensity</td>
<td>Deterministic set of rules</td>
<td></td>
</tr>
<tr>
<td>-Number of lobes</td>
<td>CDF</td>
<td></td>
</tr>
<tr>
<td><strong>Initial condition</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-Coordinates X Y of the sediments source</td>
<td>CDF</td>
<td>Uncertain because this location is not recorded in any geological structures, there can be multiple sources including those outside the domain of study</td>
</tr>
</tbody>
</table>

2.2. Optimization framework for constraining to data

Surface-based models contain two stochastic components: parameter uncertainty and spatial uncertainty (Fig. 5). The first one is due to the random selection of the input parameters values from probability distributions (size of the lobes, source locations …). Some of those parameters (the ones controlling the model of deposition) are then used to compute a spatial distribution of lobe locations. The exact location of the lobes is drawn from this probably map. This stochastic process models the spatial uncertainty. This is the uncertainty on a lobe location given a specific set of input
parameters. In inverse modeling, one attempts to create 3D models that match the data by varying both input parameters and spatial location of the geological structures.

**Fig. 5.** Surface-based models present two stochastic processes. The first one is due to the input parameters uncertainty. To run a simulation, input values are indeed randomly drawn from probability distributions. Some of those parameters are then used to define a probability map of lobe locations. This is the spatial uncertainty: a set of input parameters only narrows the possible locations of a lobe over the domain and the exact location is randomly selected from this probability map. Both of those stochastic processes control the output variability.

Solving such inverse problem by means of optimization is not straightforward. Since the deposition of each geobody requires defining a new set of values (depending on the model, it could be the geobody location, orientation, size...), the total number of parameters to optimize becomes rapidly prohibitive. For a simple model containing ten lobes, each lobe being specifically controlled by three parameters, the optimization problem involves thirty parameters if one optimizes for lobe definition. This number doesn’t yet account for the parameters controlling the geological process and the parameters setting the initial condition of the system.

Decreasing the number of parameters is therefore vital to develop an efficient optimization scheme. Three complementary approaches are investigated to address this challenge: an evaluation of the impact of spatial uncertainty, a sensitivity analysis of the input parameters and a reformulation of the optimization problem.

**2.2.1. Weighting parameter uncertainty versus spatial uncertainty**

Determining the parameters impacting the data-fit is essential because, once identified; one can focus on optimizing them and ignore the less impacting ones. Such model simplification makes the computational burden of optimization more supportable.
The first step is then to study the sensitivity of the model to families of parameters. In our surface-based model, two families of parameters exist. The first family contains the model input parameters: lobe size, source location... The second family contains the spatial parameters, which are the lobes locations. For both families, the parameter values are uncertain: hence, one refers to parameter uncertainty and spatial uncertainty. The first step of the sensitivity analysis is to determine which family of parameters impacts the most the data-fit. If the input parameter uncertainty is dominant, the optimization of the lobe locations is unnecessary. Contrarily, a dominant spatial uncertainty allows focusing the optimization only on the lobe locations. To check what the dominant uncertainty is, two separate sets of runs are performed; each consists of generating a series of stochastic models.

- Set 1: Various input parameters are randomly drawn but the random seed for selecting the lobes locations is kept constant. As a consequence, only parameters uncertainty is considered in the simulations.

- Set 2: The same input parameters are used for all the simulations but the random seed to select lobe locations is changed. As a consequence, only spatial uncertainty is considered.

For each set of models, a variance in the data-mismatch is computed. A high variance means that the corresponding uncertainty strongly impacts the model variability in term of data fitting. Based on the two values, it is possible to evaluate the relative importance of each uncertainty.

### 2.2.2. Sensitivity analysis of the input parameters

Performing a sensitivity analysis of the input parameters can help determining which ones are most impacting the data fit; hence one can focus on those few parameters to solve the conditioning problem. One main issue complicates however the task. Even if the spatial uncertainty is not leading, it makes the output results non-unique: one set of inputs can give different results if the lobes locations are changed. It adds variability to the output (data-fit) that is not linked to the input parameters uncertainty. It makes therefore interpreting what input parameters are consequential to fitting the data more difficult. A commonly used method for sensitivity analysis is experimental design combined with response surface analysis (Gosh and Rao, 1996). In order to be efficient, such methods requires models that vary relatively smoothly with regard to the input parameters, which is not the case of surface-based models. Instead, a distance-based model selection and sensitivity analysis is applied. This method has been successfully applied by Scheidt and Caers (2009) to study the sensitivity of a reservoir flow behavior in term of channel characteristics. The key idea behind this method is to define a dissimilarity distance between simulated models. This distance indicates how similar two models are in terms of the response of interest (in our case the data-fit). Based on these dissimilarities, the models are mapped in a low dimensional space, where clustering is applied. Such mapping enables extracting the underlying structure and relations between models.
Each cluster consists therefore of models with similar characteristics. For each cluster, one can define the subset of parameters values used to generate the models. By studying how these subsets compare to each other, it is possible to identify the most influencing parameters.

### 2.2.3. Optimization scheme

Once the sensitive parameters have been identified, the following step consists of optimizing them. To perform this step efficiently, three issues have to be addressed. First, the model response is non-smooth to input variations. Secondly, the Gaussian noise perturbing the lobes thicknesses must respond smoothly to input parameters changes. Lastly, the number of parameters to optimize can be significant, carrying out a prohibitive computational cost. Each problem is addressed in the following parts.

#### Choice of the optimization algorithm

Since the forward model does not respond continuously to small variation of the input parameters, the resulting evaluation function is non smooth. A derivative free optimization algorithm is therefore required. In this work, the Nelder-Mead approach (polytope method) is chosen because of its simplicity and its robustness for problems with discontinuities (Nelder and Mead, 1965).

#### Gaussian noise generation

Generating a gaussian noise requires defining a variance, a covariance and a random seed controlling the stochastic components (random number generation) of the algorithm. To develop an efficient optimization approach, it is important to be able to modify the noise gradually and smoothly. For the same seed, a small perturbation in the variance and covariance leads to a small perturbation in the resulting Gaussian noise. However, for given variance and co-variance, a method is required to gradually change the Gaussian noise. To achieve this, we use the gradual deformation method (Hu, 2000). The basic idea behind the gradual deformation is to modify the stochastic simulation with a continuous parameter. A small variation of this parameter induces a small, gradual deformation of the noise structure, while preserving the spatial variability expressed by the given variance and covariance functions (Fig. 6). This approach is well suited for optimization because it will induce smooth variations in the objective function at least in terms of perturbing the Gaussian noise.
Fig. 6. Traditionally, a Gaussian noise is generated by defining a variance, covariance and seed. This seed controls the stochastic component of the algorithm. In term of optimization, generating a noise using this approach is not convenient because a small perturbation of the seed changes completely the noise structure, inducing discontinuities in the objective function. Generating a Gaussian noise using the gradual deformation method allows avoiding this issue (Hu, 2000). Thanks to this method, the stochastic process is controlled by a continuous parameter and not a random seed: a small change in this parameter generates a slight deformation of the noise, causing therefore a smooth variation in the objective function in terms of perturbing the Gaussian noise.

Reformulation of the inverse problem

After the sensitivity analysis, the number of parameters to optimize may still remain significant. In some other cases, the sensitivity analysis may not identify any leading parameters, with no simplification of the problem possible. In such cases, simultaneously optimizing all the parameters may be extremely costly. An alternative approach is to solve sequentially the problem. The idea is to divide the optimization problem into smaller ones that are easier to solve. The general scheme is presented in the following workflow (Fig. 7). Some specific details on how this workflow is tailored to our model are also provided.
**Initialization step:** As a prerequisite, two sets of parameters need to be differentiated. The first set includes general parameters describing the initial condition of the system and the geological process as well as the overall length, width and thickness of the lobes. These parameters are randomly drawn once during a forward simulation. The second set includes the parameters specific to an individual lobe, such as coordinates and the Gaussian noise. The lobe specific parameters are drawn each time a new lobe is being deposited. The first set of parameters is initialized by random drawing from the input probability distributions. For the second set, we associate a random seed with each lobe. This seed allows drawing lobe locations and geometrical features during the forward model.

**Step 0:** In this step, parameter set 1 is optimized using the polytope method. The seed associated with the lobes remains the same. Depending on the sensitivity analysis results, the inverse problem can even be limited to the most sensitive parameters. Since only few parameters are required to be optimized, the problem is relatively faster to solve. Once optimized, these parameters values are kept constant in the following optimization steps.

**Step 1 (lobe 1):** The location and shape of the first lobe are optimized. This requires optimizing eight parameters: two parameters controlling the lobe location, three for the lobe overall shape (width, length and thickness), and three for the Gaussian noise. The seeds specified in step 0 are used to generate the remaining lobes. As a consequence, the algorithm evaluates and optimizes the influence of the first lobe on the sequence of deposition. In general, perturbing the first lobe of the sequence of deposition induces possibly large changes in the model. Indeed, the perturbation of the bottom lobes propagates sequentially when more lobes are added. At the end of this step, the first lobe of the deposition sequence is frozen. Its geometry is added to the initial paleotopography. As a consequence, the following step of optimization simulates only N-1 lobes.

**Step i (lobe i):** The location and geometry of lobe i is optimized. Since the first i-1 lobes are frozen, a forward model requires simulating the deposition sequence from lobe i to the last lobe, which means $N_{\text{lobes}} - i + 1$ lobes to generate.

This approach allows simplifying the parameter optimization problem by dividing it into similar steps. For each step, the optimization problem is faster to solve because only one subset of parameters is being perturbed. This approach, termed “greedy” reduces the computational time taken by the forward simulations. The cost of one forward model with $N_{\text{lobes}}$ is $O(N_{\text{lobes}})$. The lobes are sequentially deposited and the complexity of the problem increases linearly with the number of lobes. The computational time of a forward simulation is therefore $\alpha \times N_{\text{lobes}}$ with $\alpha$ the time required to simulate a single lobe. This means also that, for a traditional, the traditional time for iterating is $N_{\text{iterations}} \times \alpha \times N_{\text{lobes}}$.

In our approach, the problem is divided in $N_{\text{steps}} = N_{\text{lobes}} + 1$ (one step for the general parameters and one for each lobe). Step 0 (optimization of the general parameters)
and step 1 (optimization of lobe 1) requires simulating the full sequence of deposition. The computational time of a forward simulation is $\alpha \cdot N_{\text{lobes}}$. For the second step, the first lobe has been merged with the topography. Since the forward model starts the sequence of deposition from the second lobe, the computational time becomes $\alpha \cdot (N_{\text{lobes}} - 1)$. Similarly, the computational time of a forward simulation at the step $i$ is $\alpha \cdot (N_{\text{lobes}} - i + 1)$.

\[
T = \frac{N_{\text{lobes}}}{N_{\text{lobes}} + 1} \alpha \cdot (1 + \frac{N_{\text{lobes}} + 1}{2})
\]

It means that if the simulation of a lobe takes 1 second, the model runs 10 lobes and the optimization stops after 1000 iterations, then the duration of the optimization scheme is 5400 seconds or 5.4 seconds per forward model on average. A traditional optimization would take for the same number of iterations 10000 seconds, in other words 10 seconds per forward model. With no improvement in the forward-model implementation and no increase in CPU-speed, our approach allows performing almost twice as much forward simulations as a typical optimization approach does for the exact same computational time.
Fig. 7. Optimization workflow. The main inverse problem is solved sequentially. General parameters are first optimized, then, the optimization focuses on the geometry and location of the first lobe. Each step requires optimizing a relatively small number of parameters.
3. Case study

3.1. Presentation of the data set

The workflow is applied to a real data set originating from a Pleistocene turbidite system in the East Breaks mini-basin, western Gulf of Mexico (Miller et al., 2008). This mini-basin is 8 km wide by 16 km long and is filled with lobe deposits (Fig. 8). The data is composed of 3 well logs and 2 surfaces (the bottom and top surfaces of the mini-basin). The log data provides information about thicknesses of individual lobes at the wells locations. Because the mini-basin is very shallow, we assume that the morphology of the surfaces does not need any restoration. As a consequence, the bottom surface is used directly as the paleo-seafloor on which the simulated lobes are deposited. The top surface is then used to compute the thickness map of the sediments. The log data and thickness map are consistent with each other in the sense that the sum of all the lobes thicknesses recorded in a well matches the corresponding thickness defined in the map. In a traditional reservoir study, surfaces, also called horizons, are derived from seismic survey, and individual lobes thicknesses are deduced from well logging measurements. The data used in the case-study are not raw seismic data but have been processed and interpreted; they are therefore subject to uncertainty. The data uncertainty plays a critical role in problems where risks in exploration, development or production of reservoirs have to be assessed (Thore et al., 2002). In general, such uncertainty is handled by working on different alternative data sets. In other words, considering data uncertainty in our particular optimization scheme would require generating top surface models for each of the alternative data sets. In our study, we assume we are dealing with one such top surface, or one alternative thickness data set.

The case study is twofold. First, the workflow is applied to generate a surface-based model that fit the East Breaks wells and thickness data. Next, the computational performance of the method is evaluated. The main challenge brought by East Breaks is caused by the different types of data: matching the thickness map means reproducing the overall shape of the sediments package. Matching log data requires reproducing a precise internal layering at the well locations.
3.2. Specification of the input parameters

The parameters required to run a forward model are uncertain and represented with probability distributions. In our case, the wells and thickness map don’t provide enough information to infer those probabilities. The distribution of the source location, the progradation and migration intensity are borrowed from the information given by Miller et al (2008).

Since no data is available about the lobes geometry in Miller’s paper, the distributions of lobes sizes are based on the work of Seller et al. (2008) who presents an extensive study of similar deep-water fan lobes. Note that in our model, the thickness variable is used to build the initial geometry of a lobe, before any erosion occurs. The lobes observed by Seller et al. have already been eroded; hence the observed thickness does not equal the initial one. To circumvent this problem, the thickness distribution entered in our model is tuned such that the generated simulations display lobe thicknesses after erosion similar to the ones observed by Seller et al. (2008).

The parameters generating the Gaussian noises should be chosen such that the perturbed lobes geometries are always geologically consistent. Defining a criterion that evaluates automatically the geological consistency of a lobe is however extremely difficult. Instead, a visual approach is used. Various generated lobes are perturbed with different noise, each with different variance and covariance structure. Based on the overall shape of the lobe (if it compares visually well with a typical lobe shape), the geometry is accepted or rejected. The rules are defined by studying which parameters preserve the geometry and which parameters don’t. As a result, we estimate that the correlation length should be between 1000m and 7000m for all lobes. The variance is more critical since it defines the intensity of the perturbation

Fig. 8. a) bottom surface. b) top surface. c) sediment thickness. Sedimentation occurs mainly around the sediments source, which is assumed to be unique and located inside the red window.
and has a larger impact on the lobe overall geometry, hence consistency. The maximum variance is defined such that the created perturbation cannot exceed (in thickness units) one third of the maximum lobe thickness.

Running the above defined forward model also requires defining the number of lobes present in the system. This number is important because it works as a stopping criterion for the model: the forward simulation ends when the specified number of deposited lobes is reached. In our case study, the stopping criterion is different. Indeed, the thickness map gives information on the volume of deposited sediments, not the number of lobes. A forward simulation stops when the volume of deposited lobes matches the reference volume. Hence, defining the number of lobes is not necessary.

Because of the lack of data describing the East Breaks system, the input parameters are highly uncertain and their associated probability distributions present a large range of variability. This results in a slower convergence towards matching the well and thickness data since the search space is very large.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lobe Geometry</td>
<td></td>
</tr>
<tr>
<td><strong>Width</strong></td>
<td>Uniform [1000m 3000m]</td>
</tr>
<tr>
<td><strong>Length</strong></td>
<td>Uniform [2000m 7000m]</td>
</tr>
<tr>
<td><strong>Thickness</strong></td>
<td>Uniform [5m 35m]</td>
</tr>
<tr>
<td><strong>Variance of the noise</strong></td>
<td>Uniform [0 $x$] $x$ depending on the maximum lobe thickness</td>
</tr>
<tr>
<td><strong>Covariance of the noise</strong></td>
<td>Uniform [1000m 7000m]</td>
</tr>
<tr>
<td>Geological process</td>
<td></td>
</tr>
<tr>
<td><strong>Progradation intensity</strong></td>
<td>Uniform [500m 1000m]</td>
</tr>
<tr>
<td><strong>Migration intensity</strong></td>
<td>Uniform [200m 700m]</td>
</tr>
<tr>
<td><strong>Tau value</strong></td>
<td>Uniform [0 1]</td>
</tr>
<tr>
<td>Initial condition</td>
<td></td>
</tr>
<tr>
<td><strong>Coordinates X Y of the sediments source</strong></td>
<td>Unique and uniformly distributed in a window of 3kms by 3kms situated at the top of the domain (fig. 8)</td>
</tr>
</tbody>
</table>

Table 2 Probability distributions associated with the input parameters.

3.3. Optimization

3.3.1. Definition of the objective function

An objective function determines how close a model is in reaching the objective of the optimization. In the case study, a matched model should fit at the same time the thickness map and the log data. This corresponds to two sub-objectives. Similarly,
matching the log data requires matching the lobes thickness recorded in each of the three logs.

\[ E_{\text{wells}} = \frac{1}{N_{\text{lobe}} \times N_{\text{data}}} \sum_{i=1}^{N_{\text{lobe}} \times N_{\text{data}}} ( \text{thick}_i \text{ data}_i \text{ lobe}_i - \text{thick}_i \text{ sim}_i \text{ lobe}_i )^2 \]

The result is a mean square error of the mismatch in term of lobe thickness. The errors associated with each well are then averaged to obtain a combined wells error \( E_{\text{wells}} \).

\[ E_{\text{wells}} = \frac{1}{3} \sum_{i=1}^{3} (E_{\text{wells}}) \]

The mismatch between the two thickness maps, \( E_{\text{seismic}} \), is computed by computing the mean square error of both surfaces discretized on a 2D grid. The final objective function is then defined as:

\[ E_{\text{total}} = \frac{E_{\text{wells}}}{E_{\text{wells}} \text{ initial}} + \frac{E_{\text{seismic}}}{E_{\text{seismic}} \text{ initial}} \]

\( E_{\text{wells}} \text{ initial} \) and \( E_{\text{seismic}} \text{ initial} \) are the mismatch values of the initial guess.

### 3.3.2. Weighting input parameter uncertainty versus spatial uncertainty

Two sets of model realizations (200 realizations each) are generated, one with constant parameter input and varying random seed for lobe location selection, and one with varying parameter input and constant random seed. As a criterion for deciding the number of realizations to generate, we use the histogram of the misfit with the data calculated for each set of models. After 200 realizations, the two histograms remain stable when new observations are added in them. We assume then that a set of 200 models is relevant to compare both uncertainties. The two series of runs shows that the parameter uncertainty is the leading uncertainty because the corresponding histograms of misfit have the largest variance (Fig. 9). This makes sense in the East-Breaks case: the lack of available prior geological knowledge to characterize the system results in a large uncertainty in input parameters (see Table 2). This high uncertainty strongly impacts the model output variability in term of data fitting.
Fig. 9. 1) Histogram representing the mismatch with data when the parameters uncertainty only is considered. The variance is 3.2. 2) Histogram representing the mismatch with data when the spatial uncertainty only is considered. The variance in mismatch is 0.83

3.3.3. Sensitivity analysis

The input parameters considered in the study are the ones presented in Table 2. The same ranges of variability are used. Our sensitivity analysis is based on 300 realizations. The linear interactions between two parameters are considered. We assume that the other interactions are an order of magnitude less important.

Fig. 11 shows importance of each parameter for the data-fit through a Pareto plot. The red line corresponds to the significance level, meaning that the parameter values that cross the line are statically influential to the model response.

The results show that the Tau parameter, the variance and covariance of the noise are the most influencing parameters in term of data fitting (Fig. 10). The tau parameter is important because it modifies the probability maps used to draw lobe locations. It impacts therefore directly the lobe stacking patterns. Two reasons explain the importance of the Gaussian noise. First, by perturbing the geometry of the lobes, the internal layering of the model changes. The second reason is related to the shape of the deposition surfaces. The placement of the lobes is mainly controlled by the
topography. The added noise modifies this topography; hence influences the placement of the lobes (Fig. 11).

Fig. 10. Pareto plot showing the importance of each parameter on the data-fit.
Fig. 11. The lobe $n$ is perturbed with two different noises. The first noise creates a low topography on the right of the lobe. The following lobe is logically filling it (left pictures). The opposite happened in the other case (right pictures). This example shows the importance of the noise in the placement of the lobes.

3.3.4. Presentation of the optimized model.

**Definition of the number of iterations**

The Tau value and the Gaussian noises added on top of the lobes are optimized by using the sequential (step-by-step) optimization. To validate the efficiency of this approach, one wants to make sure that the computational time carried out by the optimization process is not too large. It is even more critical in our case-study since the analysis of the model spatial uncertainty and the parameters sensitivity has already required considerable CPU time. We observed in the runs performed in the sensitivity analysis that the number of lobes in the models varies from 10 to 15. To control the total CPU-time, each optimization sequence of the conditioning approach is stopped after 50 iterations. The associated computational time for the full optimization scheme (around 1.5h) is assumed to be reasonable enough for the method to be integrated in a real modeling workflow. However, better matches would be obtained with longer optimization runs.

**Initial guess**

The workflow starts first by the simulation of an initial guess. The initial parameter values are randomly picked in the prior distributions. Logically, the initial model
doesn’t match the data set: the sediment package is thinner and covers a larger area of
the domain (Fig. 12 and Fig. 14) and the lobes thicknesses are not reproduced at the
well location.

Optimized model

The optimized model is composed of 12 lobes. It displays the main features present in
the East Breaks data (Fig. 12). Most of the deposited lobes are gathered around the
proximal part of the basin (near the source) and no deposition occurs in the distal part.
The high and low sedimentation areas are therefore accurately reproduced. The two
similar thickness maps also confirm the quality of the match. From the error map, we
can see that most of the mismatch is located near the sediment source area. At this
location, the environment of deposition is, in general, high energy and the associated
erosion process intense. A source of residual misfit might be the model of erosion,
which is probably not accurate enough.

The simulated lobes compare well with the ones recorded in the log data (Fig. 13).
However, the surface-based model produces slightly less variability in terms of lobe
thicknesses. This is probably due to the use of statistics that reproduce poorly the
actual thickness variability. Another reason can be the erosion model, which may be
too simplistic. A high variability in the recorded thickness can indeed mean high
variability in erosion intensities.

Fig. 14 displays the internal layering of the geological model at different stages of the
optimization process. The initial guess presents a poor match with the data. At the end
of step 5 (250 iterations), the obtained model starts to fit the log data and the
thickness map (represented by the location of top surface) with increased accuracy.
The fit with the logs data is especially good at the bottom part of the reservoir.
Indeed, at the end of step 5, the first four lobes of the deposition sequence have
already been perturbed and combined with the topography. Three of them are
recorded in the cross-section (the three bottom ones). The final model reproduces the
thickness of the reservoir observed at the cross-section location and fit accurately the
log data. Fig. 15 exhibits three matching models generated from different initial
guesses. Those models present a large variability in term of internal layering. It
means that the optimization algorithm converged towards different local minima.
Fig. 12 Result of the optimization method. The larger errors in term of misfit with data are located near the sources, where the erosion is very pronounced.
Fig. 13: Well log data recorded in the optimized model and data.
Fig. 14. Cross-section of the model at different steps of the optimization process. Each lobe is represented by a different colored zone.
Fig. 15. Three matching models generated from different initial guesses.

3.3.5. Computational performance of the workflow

The results of the case study outlined above show the possibility of fitting surface-based models to real data. The accuracy of the match is defined by the quality of the model itself. However, the efficiency in matching data is controlled by the workflow itself. This feature is very important because it determines the applicability of such methods.

Problem dimensionality

In our model, the general geometry of the lobes, the geological process and the initial condition of the system are defined by seven parameters. Each lobe is, in addition, controlled by eight specific parameters: two for the lobe coordinates in the domain (location), three to generate the Gaussian noise and three for the lobe thickness, width and length. Since twelve lobes are present in the model, fitting the East Breaks data by solving the full inverse problem necessitates the perturbation of 103 parameters (7+8*12). The parameter uncertainty has been shown to be dominant. The optimization of the lobes locations is not necessary. As a consequence, the number of parameters to optimize is reduced to 79 (7+6*12). The sensitivity analysis shows that the Tau parameter and the parameters generating the lobe noises are the most influential ones. Retaining only those influential parameters results in a set of 37 parameters (1+3*12). By using the developed sequential optimization, the problem is then divided into 13 different steps: an initial one to optimize the Tau value and one
for each lobe’s noise optimization. As a consequence, by combining an uncertainty-
sensitivity analysis and a reformulation of the conditioning problem, the initial 103
dimensional problem is reduced to one 1D optimization and twelve 3D optimizations.

Benchmarks
Several runs are generated using our step-by-step optimization approach. Those runs
start from different initial guesses. Since the total number of iterations cannot be
controlled, the number of function evaluation per step is set to 50. As a consequence,
The total number of iterations varies from 600 iterations (for 11 lobes) to 800
iterations (for 15 lobes)
To evaluate the efficiency of the method, we need first to verify that the generated
data-fit is one of the best fit the forward model can produce. It requires therefore
evaluating what is a “good fit”. To address this issue, our approach is compared with
a random sampling approach. The 103 parameters are randomized and 10000
realizations are generated. The minimal mismatch is set as a reference. This value
gives a reasonable indication of what a "good fit" can be. All the 103 parameters (and
not only the 37 most influencing) are randomized to ensure that the full variability of
the model is captured.
We also need to evaluate the speed of convergence of the developed method. The
comparison is done with a traditional approach where the 37 parameters are
optimized all at once. The Nelder-Mead method is chosen as the optimization
algorithm. In both cases, the total number of iterations is set also 700.

Results
Our conditioning workflow achieves a better match than a random sampling of 10000
iterations (Fig. 16). It also outperforms the traditional optimization approach in term
of speed of convergence and quality of fit. The traditional approach tends indeed to
be overwhelmed by the amount of parameters (37 at once).
The mismatch generated with our method dramatically decreases during the first 100
iterations. Those 100 iterations correspond to the two first steps of the optimization
scheme. During those two steps, the depositional model (Tau parameter) and the
geometry of the first lobe are perturbed, which induces large model changes. Finding
a local optimum is then easier.
The efficiency of the method is even more evident when its evaluation is based on
computational time and not number of iterations. Indeed, the simulation time of a
single realization decreases during the step-by-step optimization: less and less lobes
have to be simulated since the lobes of the previous iterations are combined with the
topography. This improvement is illustrated in fig 17. 650 iterations (12 lobes) of the
traditional method take 2.7 hours (2 hrs 42 mn). The computational time of a forward
simulation is 14.9 seconds for 12 lobes. With our conditioning workflow, 650
iterations take 1.38 hours (1 hrs 23mn). The average computational time of a forward
simulation is 7.6 seconds in average for 12 lobes. The improvement of 48% is
consistent with the theoretical one (2.2.3). The efficiency of the method can also be
assessed by comparing the computational time necessary to reach the 0.5 mismatch
threshold (considered as a ‘good fit’). With the brute force optimization, only one run reaches the threshold and in 1.25 hours. With our method, all the optimization runs reach the threshold, in 0.4 to 0.6 hours (100 to 200 iterations).

\[ \text{Fig. 16 Performance results of the different optimization approaches. On the left, the evaluation is based on the number of iterations. On the right, the evaluation is based on the computational time (hours).} \]

**Discussion**

While the approach is general in fashion and working well in the presented example, it is important to state some of the limitations. First, the variability between the generated matching models has not been evaluated. In our case, this variability is supposed to represent the remaining uncertainty on the reservoir geometry given the available data and our model. This uncertainty is important since most of the decisions in reservoir exploration and management are based on it. In such case, one wants to make sure that the full variability of the model is captured so that all the high pay zones are considered (in drilling for example) or all unsuited events hedged (in reservoir production). However, our approach perturbs a small subset of parameters and solves the inverse problem sequentially. As a consequence, the full variability of the model may not be entirely captured in the generated realizations. Future research should therefore focus on integrating such optimization method in an uncertainty assessment workflow.

Secondly, it seems that the proposed method would work better for fields with lower amount of data and with a high degree of uncertainty in the geological model (early
development). Indeed, the method has the ability to modify drastically the geometry of the model, changing the locations of the lobes and the thickness of the reservoir for example. In addition, matching hundreds of wells is probably difficult since the model may not able to reproduce the high complexity present in the data.

Thirdly, fully assessing the applicability of the workflow calls for new case studies with different environments of deposition (fluvial, deltaic, carbonate) and data (facies, production, well test).

The sequential optimization approach developed in this work can be useful in inverse basin modeling. Some of the input parameters in basin modeling, such as sediments supply or sea level are indeed defined as time series. Therefore they are not represented by one parameter but by several ones, the number depending on the time-step used to discretize the geological time. Optimizing and perturbing the entire curve at once is extremely costly, especially since those models are complex, highly non-linear and slow to run. By using the step-by-step approach, the curve would be divided in small parts. The optimization would then start by evaluating and optimizing the influence of the oldest part of the curve on the full sequence of deposition. Once the optimization is performed, the corresponding simulated structures would be integrated in the topography and the optimization would focuses on the next part of the curve, and the sequence would end when the totality of the curve has been optimized.

**Conclusion**

We presented a workflow that allows generating surface-based models that fit wells and seismic data. Solving such optimization problems is time-consuming and expensive. The key idea behind the method is to decrease the dimensionality of the problem. To this end, three complementary approaches are developed. The first one identifies the leading uncertainty. The second one is a sensitivity analysis on the input parameters. The third one is a reformulation of the optimization problem. The problem is solved sequentially and not all at once. This reformulation, by decreasing the number of lobes to simulate after each optimization sequence, reduces also the cost of running the model. The workflow is applied successfully to a turbidite data set composed of a thickness map and three well logs. The surface-based model used in the study reproduces the lobes structures. The workflow is general in the sense that it can be applied to any kind data and environments of deposition.
References


