Abstract

The goal of a reservoir study is to help to decide the future development of a field based on technical and economic criteria. To reach this goal, one would like to quantify the impact of uncertainty on production and economic forecasts to take the decision while considering the risk. Practically it would correspond to supply to the manager the uncertainty distribution (or P10, 50 and 90) of the production forecasts associated to each scenario.

The uncertainty on the production forecasts is linked to a specific scenario and to the knowledge of the reservoir. For a mature field two kinds of knowledge exist:

- Static parameters used to build the numerical model: geological concept, variograms, correlation lengths, permeability and porosity distributions, etc. The static parameters are associated with "a priori uncertainties" defined by their probability distributions.

- Dynamic data: measurements related to the dynamic behavior of the reservoir, such as measured pressure, oil/water/gas rates at the wells, 4d seismic, etc.

The bayesian formalism enables to reduce, in a statistical framework, the static parameter uncertainties by taking into account the dynamic data. These "a posteriori" distributions of the static parameters can then be used to compute probabilistic production forecasts for each possible scenario honoring static and dynamic knowledge of the reservoir. However, this formalism involves the determination of the likelihood function, which can lead to a prohibitive cost in terms of reservoir simulations.

To drastically decrease the amount of reservoir simulations needed to determine the "a posteriori" uncertainties we propose to approximate the likelihood function by a non-linear proxy model combining experimental design, universal kriging and dynamic training techniques. These "a posteriori" distributions can then be used into the classical experimental design approach to compute probabilistic production forecasts constrained by dynamic data.

The proposed methodology will be illustrated on a field case.

Introduction

The uncertainty associated to the production forecasts of a mature field can be obtained by considering the history matching uncertainty as stochastic [1], as follows:

- Building a reservoir simulation model, with given "a priori" uncertainties on the parameters Θ,

- Determining several history matched model, through the optimization of some parameters Θ in an history matching procedure [2],

- Deduction of the production forecasts uncertainties using the Joint Modeling method [3] with the different matched models.

---

European Conference on the Mathematics of Oil Recovery — Cannes, France, 30 August - 2 September 2004
However, due to the high number of simulation and computation time required, it is rather impossible to obtain a representative number of history matched models. Therefore, the uncertainty on the forecasts is not very reliable if it is obtained with a small sample of the possible solutions. Moreover, it is not possible to obtain the probability distribution on the uncertain parameters taking into account the production history ("a posteriori" distribution).

The Bayesian formalism solves this problem in a more rigorous way, enabling to obtain the posterior distribution of the parameters $\Theta$. The uncertainty on the production forecasts of the reservoir, taking into account the production history, can then be obtained by sampling from the posterior distribution of the parameters. Even if the Bayesian formalism is more rigorous it is very computer intensive in its initial formalism, because determining the posterior distribution of the parameters requires sampling the parameters uncertain domain and leads to a prohibitive number of simulations.

The proposed method drastically decreases the computation time required by the Bayesian formalism to quantify the production forecast uncertainties. The different steps are:

1) Approximate the "a posteriori" distribution of the parameters by a non linear proxy model, obtained using a limited number of simulations,

2) Approximate the production forecasts on which we want to quantify the uncertainties, by other proxy models,

3) Generate a random sampling of the parameters drawn from the "a posteriori" distribution approximated in (1) and then deduce the uncertainty on the forecasts using the proxy models determined in (2).

We are going to describe the proposed methodology and to show the results obtained on a two-parameter synthetic case based on the PUNQ dataset.

**Methodology**

1) **Approximation of the "a posteriori" distribution**

The "a posteriori" distribution of the parameters $p(\Theta|D)$ is the probability density function of the parameters taking into account the production history (or dynamic data $D$). This distribution can be obtained using the Baye's theorem:

$$p(\Theta|D) = c \cdot L(\Theta) \cdot p_0(\Theta)$$

Where $c$ is a normalization constant, $L(\Theta)$ the likelihood function and $p_0(\Theta)$ the "a priori" distribution of the parameters.

Therefore, to determine the "a posteriori" distribution we need to compute the likelihood function. Assuming that the difference between the simulation results and the measurements are gaussian distributed with zero mean and $C_D$ as the covariance matrix:

$$L(\Theta) = (2\pi)^{-m/2} \cdot (\det C_D)^{-1/2} \cdot \exp\{-OF\}$$

With the Objective Function

$$OF = \frac{1}{2} \left[ D - S(x_D, \Theta) \right]^T C_D^{-1} [D - S(x_D, \Theta)]$$

Where $m$ is the number of measured data and $S(x_D, \Theta)$ the simulation result of the model $S(\cdot, \Theta)$ at the point $x_D$ in space and time.

The proposed method will enable to obtain a proxy model of the "a posteriori" distribution through the approximation of the OF. The proxy model will be composed by a mean part obtained with a polynomial model and a residual part obtained by a kriging method [4], [5].

A detailed description of the method is shown on **Figure 1** and in the following.

1) **Initial experimental design:**

An initial step, based on an experimental design, allows to obtain a set of OF values for different parameter combinations. This experimental design is suited for the quadratic model chosen at step (2) to
determine the mean of the proxy model. To ensure that the proxy model will respect the positive nature of the OF a logarithmic transformation of the OF is performed.

2) Mean determination:
The mean of the proxy model is obtained using classical linear regression such as the least square method:

- Let $Y$ be the log(OF) vector associated to the $k$ simulations performed: $Y = [Y_1, \ldots, Y_k]$
- Let $X_c$ be the matrix of the known points, i.e. the parameter values associated to the simulations performed: $X_c = \begin{bmatrix} \Theta_{1,1}^{sim} & \Theta_{1,2}^{sim} & \cdots & \Theta_{1,n}^{sim} \\ \vdots & \vdots & \ddots & \vdots \\ \Theta_{k,1}^{sim} & \Theta_{k,2}^{sim} & \cdots & \Theta_{k,n}^{sim} \end{bmatrix}$ with $\Theta_{1,1}^{sim}, \ldots, \Theta_{n,n}^{sim}$ for $n$ parameters.

- Let $Z_c$ be the matrix associated to the shape of the mean and to the known points:

$$Z_c = \begin{bmatrix} 1 & \Theta_{1,1}^{sim} & \Theta_{1,2}^{sim} & \cdots & \Theta_{1,n}^{sim} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \Theta_{k,1}^{sim} & \Theta_{k,2}^{sim} & \cdots & \Theta_{k,n}^{sim} \end{bmatrix}$$

for an initial full experimental design.

The mean at the point $\Theta = (\Theta_1, \ldots, \Theta_n)$ is given by:

$$\text{mean}(\Theta) = Z \cdot \beta$$

with $Z = \begin{bmatrix} 1 & \Theta_1 & \Theta_2 & \cdots & \Theta_n \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \Theta_1 & \Theta_2 & \cdots & \Theta_n \end{bmatrix}$ and

$$\beta = (Z^t \cdot Z)^{-1} \cdot Z^t \cdot Y$$

3) Covariance assumption:

To be able to use kriging techniques we need to make some assumptions on the covariance of the results $Y$: the covariance is function of the distance between the associated points within the parameter space. We have used an exponential covariance type defined by:

$$C(\Theta_1, \Theta_2) = \sigma^2 \cdot \exp \left( -\frac{||\Theta_1 - \Theta_2||}{l} \right)$$

where $l$ is the correlation length.

The matrix $R$ is defined as: $\text{Var}(Y) = \sigma^2 \cdot R$

4) Kriging of the residuals:

The vector of the residuals at the known points $X_c$ is $\text{Res}_c = Y - Z_c \cdot \beta$. The residual estimation at the point $\Theta = (\Theta_1, \ldots, \Theta_n)$ is obtained using a kriging algorithm: residual*(\Theta) = $\lambda(\Theta)$: $\text{Res}_c$. The kriging coefficients are obtained through $\lambda(\Theta)^t = R^{-1} \cdot C(\Theta)$ where $C(\Theta) = [C_1 \cdots C_k]$ with $C_i = C(\Theta - \Theta_i, l)$. It is also possible to compute the kriging estimation variance.

We are then able to compute the proxy model as:

$$\text{proxy}(\Theta) = \text{mean}(\Theta) + \text{residual}^*(\Theta)$$

The prediction parameter $Q2$ is computed as follows:

$$Q2 = 1 - \frac{\sum_{i=1}^k \text{Diff}(\Theta_i^{sim})}{\sum_{i=1}^k [Y_i - \text{mean}(Y)]^2}$$

with $\text{Diff}(\Theta_i^{sim}) = [\text{proxyWithout}Y(\Theta_i^{sim}) - Y_i]^2$ where $\text{proxyWithout}Y(\Theta_i^{sim})$ corresponds to the value of the proxy model at the point $\Theta_i^{sim}$ determined without the known result $Y_i$.

5) Covariance optimization:

We have seen that the proxy model depends on the covariance assumption. In this step, the objective is to optimize the covariance correlation length to maximize the prediction criteria $Q2$. An iterative gradient method is used. Once the optimal value, $l_{\text{opt}}$, is determined we can get the optimal values of the kriging coefficients, $\lambda_{\text{opt}}$. 

---

European Conference on the Mathematics of Oil Recovery — Cannes, France, 30 August - 2 September 2004
6) **Addition of J new points:**
If necessary, several points can be added to improve the proxy model. They can be selected through different criteria such as: maximum variance of the proxy model, maximum or minimum of the proxy model, bad predictive zones, etc. A threshold is used to avoid the selection of a point too close of the known ones.

7) **Proxy model of the OF/likelihood/"a posteriori":**
When they are enough points: Q2 stable and close to 1, maximum number of simulations reached, we obtain the final proxy model of log(OF) and the approximate OF and likelihood are deduced: \( \text{proxyOF}(\Theta) = \exp(\text{proxy}(\Theta)) \), \( \text{proxyLkd} \approx \exp(- \text{proxyOF}(\Theta)) \). Then the proxy model of the "a posteriori" distribution is obtained using: \( \text{proxyPosterior}(\Theta) = \int_{\Omega} \frac{\text{proxyLkd}(\Theta) \times p(\Theta)}{\int_{\Omega} \text{proxyLkd}(\Theta) \times p(\Theta) d\Theta} \) where \( p(\Theta) \) is the "a priori" distribution and \( \Omega \) is the uncertain domain.

2- **Approximation of the production forecasts**

We are using experimental design technique and response surface modeling, such as described in [6], to approximate the production forecast responses of the model. Therefore, for each response, on which we want to quantify the uncertainty, we determine an accurate and predictive proxy model, function of \( \Theta \), to approximate the simulation model.

3- **Uncertainty quantification on the production forecasts**

A Monte-Carlo sampling of the parameters from the "a posteriori" distribution approximated in (1) is performed. Then, using the proxy models determined in (2) and the parameter samples the distribution of the production forecasts are deduced.

**Field case study**

1. **The PUNQ filed case**

The PUNQ test case is a 3D synthetic reservoir model derived from real field data. It was already used for comparative inversion studies in the European PUNQ project [7], and for validation of constrained modeling and optimization scheme development methods [8], [9], [10].

The top structure of the reservoir is presented in Figure 2. The reservoir is surrounded by an aquifer in the north and the west, and delimited by a fault in the south and the east. A small gas cap is initially present. The geological model is composed of five independent layers. The layers 1, 3, 4 and 5 are assumed to be of good quality, while the layer 2 is of poorer quality. The field has an estimated OOIP of 16 MMm3. Eight years of production history is available at 6 wells.

The objective of this study was to determine the production forecasts uncertainty using the proposed methodology, while determining the "a posteriori" distribution of uncertain parameters.

Two uncertain parameters have been determined and are described in **Table 1**.

<table>
<thead>
<tr>
<th>Short Name</th>
<th>Long Name</th>
<th>Min</th>
<th>Max</th>
<th>A priori pdf</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>MPH</td>
<td>Horizontal permeability multiplier for all layers</td>
<td>0.8</td>
<td>1.2</td>
<td>Uniform</td>
<td>1.09</td>
</tr>
<tr>
<td>SORW</td>
<td>Residual oil saturation after water swept</td>
<td>0.15</td>
<td>0.25</td>
<td>Uniform</td>
<td>0.18</td>
</tr>
</tbody>
</table>

Table 1: Range of the parameter uncertainties

The uncertainty domain is defined according to the Min and Max possible values of the parameters. The "a priori" distributions for each parameter are assumed to be uniform. The 8-years synthetic production history was simulated using the reference values and is shown on Figure 3 for the cumulative oil production.
The OF includes synthetic measurements of Gas Oil Ratio, Bottom Hole Flowing Pressure and Water Cut values for each well.

2. Results of the study

For validation purpose, a mapping of the OF was done on a 25x25 regular grid in the uncertainty domain and is shown on Figure 4 for log(OF). We can see that the optimum values of the two parameters are equal to their reference.

1) Approximation of the "a posteriori" distribution

The mean of the proxy model, that approximates log(OF), was chosen to be a full quadratic model:

\[ \text{mean}(\text{MPH}, \text{SORW}) = \beta_0 + \beta_1 \cdot \text{MPH} + \beta_2 \cdot \text{SORW} + \beta_3 \cdot \text{MPH}^2 + \beta_4 \cdot \text{SORW}^2 + \beta_5 \cdot \text{MPH} \cdot \text{SORW} \]

and was associated to an initial full "Central Composite - Face Centered" experimental design with 9 experiments.

At each step (6) of the iterative methodology 3 new points are simulated. One corresponds to the maximum of the proxy model estimation variance and two correspond to the minimum and maximum of the proxy model.

The initial proxy model of log(OF), obtained with 9 simulation results, is represented on Figure 5. It is of poor quality with a Q2 coefficient close to zero. Therefore, we need to add more points to improve it.

After 5 iterations (a total of 24 simulations) a log(OF) proxy is obtained, with a Q2 coefficient equal to 0.9. It is shown on Figure 7. When comparing Figure 4 and Figure 7, we can see that the final proxy model is a good approximation of log(OF), and therefore of the likelihood function Figure 6.

Since the "a priori" distribution laws for the uncertain parameters are uniforms, the "a posteriori" distribution of the parameters is directly deduced from the likelihood function.

2) Approximation of the production forecasts

Our objective is to quantify the uncertainty on the cumulate oil and water produced after 25 additional years of production (cumOil, cumWater). Therefore we want to approximate these simulator responses by 2 quadratic polynomial models with respect to MPH and SORW. The 9 simulations, obtained from the initial experimental design to approximate the likelihood function, are used again. The following, accurate and predictive, proxy models are obtained:

\[ \text{cumOil} = 4.984 + 0.07 \cdot \text{mph} - 0.224 \cdot \text{SORW} - 0.003 \cdot \text{mph} \cdot \text{SORW} - 0.01 \cdot \text{mph}^2 - 0.005 \cdot \text{SORW}^2 \]
\[ \text{cumWater} = 5.831 + 0.184 \cdot \text{mph} + 0.419 \cdot \text{SORW} + 0.008 \cdot \text{mph} \cdot \text{SORW} - 0.024 \cdot \text{mph}^2 + 0.008 \cdot \text{SORW}^2 \]

where mph and sorw are the normalized values of MPH and SORW between -1 and 1.

3) Uncertainty quantification on production forecasts

The parameter "a posteriori" probability density function is used to quantify the uncertainty on the production forecasts constrained to the production history. A total of 10000 parameters samples are obtained from the "a posteriori" pdf using the Monte-Carlo method. Then, using the proxy models of cumOil and cumWater with the 10000 sample values, the histograms and the percentiles (P10, P50 and P90) are computed, as shown on Figure 8 for cumOil.

To quantify the impact of production history on uncertainty quantification, we have computed the histograms and percentiles on cumOil and cumWater using the "a priori" distribution of the uncertain parameter, as shown on Figure 9 and on Table 2. We can see that the uncertainty on production forecasts was significantly reduced while taking into account the production data.
Table 2: Percentile values for cumOil and cumWater using "a posteriori" and "a priori" distributions.

<table>
<thead>
<tr>
<th></th>
<th>P10</th>
<th>P50</th>
<th>P90</th>
<th>Uncertainty</th>
</tr>
</thead>
<tbody>
<tr>
<td>CumOil in MMm³</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;A posteriori&quot; distribution</td>
<td>5.05</td>
<td>5.11</td>
<td>5.17</td>
<td>2.35%</td>
</tr>
<tr>
<td>&quot;A priori&quot; distribution</td>
<td>4.79</td>
<td>4.98</td>
<td>5.16</td>
<td>7.43%</td>
</tr>
<tr>
<td>CumWater in MMm³</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;A posteriori&quot; distribution</td>
<td>5.62</td>
<td>5.72</td>
<td>5.81</td>
<td>3.32%</td>
</tr>
<tr>
<td>&quot;A priori&quot; distribution</td>
<td>5.47</td>
<td>5.82</td>
<td>6.18</td>
<td>12.2%</td>
</tr>
</tbody>
</table>

**Conclusion**

The goal of a reservoir study is to help to decide the future developments of a field based on technical and economic criteria in considering the risk. Using the methodology proposed we are able to quantify, in a bayesian framework and with a limited number of simulations, the uncertainty on production and economic forecasts taking into account static and dynamic data.

**References**


Initial experimental design:
k points of \( \Theta \) \( \Rightarrow \) k values of OF.

Mean determination \( m(\Theta) \):
- Least square method,
- …

Universal kriging

Covariance assumption

Kriging of the residuals

Covariance optimization, while:
- Maximizing a predictive criteria (Q2),
- …

Convergence

No

Yes

Add points?

No

Yes

Addition of J new points:
- maximum variance of the proxy model,
- maximum/minimum of the proxy model,
- bad predictive zones,
- …

Proxy model of the OF/likelihood/"a posteriori"

Figure 1: Principle schema of the "a posteriori" distribution approximation

Figure 2: Top structure of the PUNQ field

Figure 3: production history
Figure 4: Mapping of log(OF)

Figure 5: Initial proxy model of log(OF) function with 9 known points

Figure 6: Final proxy model of the likelihood function with 24 known points

Figure 7: Final proxy model of log(OF) with 24 known points

Figure 8: Histogram and percentile for the cumOil response (in MMm³) using the "a posteriori" distribution on MPH and SORW

Figure 9: Histogram and percentile for the cumOil response (in MMm³) using the "a priori" distribution on MPH and SORW