Abstract

In petroleum engineering, simulation models are used in the reservoir performance prediction and in the decision making process. These models are complex systems, typically characterized by a vast number of input parameters. Usually the physical state of the reservoir is highly uncertain, and thus the appropriate parameters of the input choices. The uncertainty analysis often proceeds by first calibrating the simulator against observed production history and then using the calibrated model to forecast future well production. Most models go through a series of iterations before being judged to give an adequate representation of the physical system. This can be a difficult task since the input space to be searched may be high dimensional, the collection of outputs to be matched may be very large, and each single evaluation may take a long time. As the uncertainty analysis is complex and time consuming; in this paper, a stochastic representation of the computer model was constructed, called an emulator, to quantify the reduction in the parameter input space due to production data over different production periods. The emulator methodology used represents a powerful and general tool in the analysis of complex physical models such as reservoir simulators. Such emulation techniques have been successfully applied across a large number of scientific disciplines. The emulator methodology was applied to evaluate the production data capacity to identify uncertain reservoir physical features over the production period for a synthetic reservoir simulation model. The synthetic model was built to represent a region of an injector and related producers. In the case studied; thousands of realizations were required to identify certain physical reservoir features. This justifies the use of emulation and shows the importance of this technique for the identification of regions of feasible input parameters. Moreover, the impact on the input space reduction due to different production periods was determined. The emulator methodology used assists in carrying out tasks that require computationally expensive objective function evaluation, such as identifying regions of feasible input parameters; making predictions for future behavior of the physical system and investigating the reservoir behavior.
Introduction

Reservoir simulators are important and widely-used in reservoir management; this tool is used in the reservoir performance prediction and in the decision making process. These simulators are computer implementations of high-dimensional mathematical models for reservoirs, where the model inputs are physical parameters and the outputs are observable characteristics such as well pressure measurements, fluid production and so forth. The uncertainties are always present in the reservoir characterization process; thus the input parameters are usually uncertain so is the simulator output.

The procedure to calibrate the reservoir simulation model is called history matching. Based on observed data, the set of possible input choices for the reservoir model is identified. Two different procedures can be used to perform the history match process: the deterministic and the probabilistic approach.

The deterministic approach involves running the initial simulation model with different input values to obtain one simulation model between many probable matches to the field data. According to Elrafie et al. (2009), this conventional procedure does not handle the uncertainty of all model variables and the possibility to identify and carry forward a set of multiple history match model scenarios to predictive forecasting.

In a probabilistic approach, in which several reservoir model scenarios are considered, the uncertainty analysis procedure is used in the process. Identifying the input parameters for which the simulation outputs match the observed data, can be a difficult task since the input space to be searched may be high dimensional, the collection of outputs to be matched may be very large, and each single evaluation may take a long time.

To deal with the large number of iterations and high computational resources commonly encountered in the probabilistic approach, proxy models are used. Zubarev (2009) define proxy models as a mathematically defined function that replicates the simulation model output for selected input parameters. Several papers show the use of different proxy-modeling algorithms in the history matching process (Cullick, 2006; Junker et al., 2006 and Slotte and Smorgrav, 2008).

As the history match process and uncertainty reduction quantification is complex and time consuming; this work shows the workflow used to quantify the reduction in the parameter input space due to production data over different production periods. This workflow comprises the construction of a proxy model called an emulator. This technique was applied to a synthetic reservoir simulation model, built to represent a region of an injector and related producers.

The emulator methodology used represents a powerful and general tool in the analysis of complex physical models such as reservoir simulators. Such emulation techniques have been successfully applied in reservoir-simulation problems, as seen in Cumming & Goldstein (2009), and references therein.

Objective

Describe a workflow used to evaluate the production data capacity to identify uncertain reservoir physical features over the production period using the emulation technique. Moreover, show the application for a synthetic reservoir simulation model built to represent a region of an injector and related producers.

Proposed Methodology

This topic presents the workflow used to construct the emulator. It is important to highlight that this study used a synthetic reservoir model. There is no historical data available in the process, thus the production data considered as historical data derived from a hypothetical reality selected from possible scenarios. These scenarios were obtained through an uncertainty analysis performed on the initial reservoir simulation model.

The workflow was designed to quantify the simulation reservoir model uncertainty reduction due to production data. The objective was to identify the inputs of a reservoir simulation model, within a possible input parameter space, whose outputs match to the hypothetical historical production data. The workflow used is shown in Figure 1. Each stage is described as follows.

Input and Output Parameter Definition

In reservoir simulation, uncertain inputs are physical parameters determined through an uncertainty analysis performed on the base model. The outputs of the model, for a given choice of inputs, are observable characteristics such as well-bottom-hole pressure, water rate at production wells and water saturation maps. The input variable selection depends on the underlying
problem and knowledge of the engineer.

The physical state of the reservoir uncertainty varies due to the amount of information available and production period. As in this study, the analysis is being performed in the field development phase; the uncertainty of the appropriate choices of the input parameters for the reservoir model is high.

**Input Data Set Sampling**
The input data set sampling is an important stage in creating an adequate emulator. Different sampling methods exist and have been applied in reservoir simulations. The Latin Hypercube Design (LHD) is an efficient design and was selected as a sampling method to this work. Scenarios were generated based on the input parameter space and sampled using the LHD. The selected scenarios were simulated using commercial simulation software to obtain the production outputs. The sampled input parameters and resulting simulation outputs were used to construct the emulator.

**Emulator Estimation**
The emulator is an approximation of the existing numerical reservoir model. It should be able to replicate the response of the simulation model. The general structure to develop the emulator was based on Cumming & Goldstein (2009) and Vernon et al. (2013) and is as follows.

The simulator is represented by a vector function, taking inputs $x$ which represent the vector of reservoir input parameters, and return the output parameter $f(x)$. The output parameter $f(x)$ intends to represent the real physical system output $y$. The field observed outputs is represented by $z$, as field observation is susceptible to measurement errors, the difference between $z$ and $y$ is represented by the relation in $z=y+e$

$$ z = y + e $$ Equation 1

where $e$ is the vector of random observational errors, taken to be independent of $y$. If $f(x)$ was a perfect representation of the system, then an input vector $x^*$ only would be accepted as representing the system if $f(x^*) = y$. In practice, however, the simulation reservoir model $f$ simplifies the physics and approximates the solution of the resulting equations. Therefore, the structural discrepancy is represented by $y=f(x^*)+e$

$$ y = f(x^*) + e $$ Equation 2

where $e$ is the random structural discrepancy vector and is independent of $f(x^*)$. Combining $z=y+e$ and $y=f(x^*)+e$ the input parameters $x^*$ is acceptable if it is probabilistically consistent with the relation shown in $z=f(x^*)+e+e$

$$ z = f(x^*) + e + e $$ Equation 3

The objective is to identify all choices of $x^*$ which would give acceptable fits to available production data or identify a wide range of elements $x^*$ belonging to the input parameter space $X(z)$. If the input parameter space was low dimensional, and the function was very fast to evaluate, then it would be possible to estimate $X(z)$ by evaluating the function in the entire space and identify the collection of all $x^*$ choices consistent with $z=f(x^*)+e+e$. However, for a reservoir simulation model it is infeasible to evaluate the simulator at enough choices to search the input space exhaustively. Therefore, a representation of the output uncertainty at each input choice must be constructed. This representation is termed an emulator.

The emulator both suggests an approximation to the function and also contains an assessment of the likely magnitude of the $f_i$ is presented in $f_i|x = f_i y g_{ij}(x) + u_i(x)$

$$ f_i(x) = \sum_j \beta_{ij} g_{ij}(x) + u_i(x) $$ Equation 4

where $x$ are the input variables, $i$ is the output being emulated, $j$ is

- Linearity: expected value of $u(x)$ must be equal zero, $E(u)=0$;
- Homoscedasticity and Independence of Errors: $\text{Var}(u) = \sigma^2$ and $\text{Cov}(u_i,u_j) = 0$. 


Emulator Diagnostics

Emulator diagnostics is the process of assessing an emulator’s prediction accuracy and quality. The response values predicted by the emulator must comprise the results of the full numerical simulation for the input dataset. Moreover, two measures were evaluated. The first is the squared multiple correlation ($R^2$); according to Rice (1995) this coefficient is used as a crude measure of the strength of a relationship and the second measure is the standard error ($\sigma$) which offers a first handle on how well the fitted equation fits the sample data. These measures are defined in $R^2 = 1 - \frac{RSS}{RYY}$ and $\sigma = \sqrt{\frac{RSS}{n-p}}$

where $RSS$ is the residual sum of squares obtained by calculating the square difference between the fitted and observed value; $RYY$ is the total sum of squares obtained by calculating the square difference between the fitted and mean observed value; $n$ is the number of data points and $p$ is the number of parameters to be estimated (Weisberg, 2005).

Rice (1995) comments that, it is necessary to evaluate the residuals to assess the quality of the fit. Plots of the residuals versus the fitted values were used to find failures of assumptions. Ideally the residual should show no relation to the $x$ values, and the plot should look like a horizontal blur.

Implausibility Analysis

The implausibility analysis is performed to obtain the input parameters whose outputs match the hypothetical historical data. The hypothetical historical data is derived from a hypothetical reality selected from all possible scenarios generated in the uncertainty analysis; moreover these inputs are obtained to improve the emulator reliability and to evaluate the uncertainty reduction at the end of the process.

The range of input parameters that are member of $X(z)$ is determined through the implausibility value calculation ($I$). For each set of input parameters an emulator output $\hat{f}(x)$ is obtained; with this data the implausibility value is determined using $I^2(x) = \frac{(z_i - E[\hat{f}(x)])^2}{Var[\hat{f}(x)] + Var[z_i] + Var[e_i]}$

where $z_i$ is the hypothetical historical output value, $E[\hat{f}(x)]$ is the emulator output expected value and $Var[\hat{f}(x)], Var[z_i]$ and $Var[e_i]$ are the variances of the emulator output value, structural discrepancy ($e$) and observational errors ($e$) respectively.

Large values of $I(x)$ suggest that it is implausible that $x \in X(z)$. As for each vector of inputs $x$ there are many implausibility values, one for each output, the implausibilities are then combined. The implausibility value $I(x)$ for a vector of inputs $x$ is considered as being the maximum value among all $I_i(x)$ obtained.

The input parameters $x$ that satisfy $x \in X(z)$ are called non-implausible parameters, since in the next iteration the same input may be found to be no longer plausible. If the emulator is not accurate enough or $X(z)$ does not enable a better understanding of the reservoir’s physical features, more simulation runs are designed within the ‘non-implausible’ regions in the input space and the emulation analysis is repeated iteratively; each iteration is called a Wave (Cumming & Goldstein, 2009; Vernon et al., 2013)

The maximum acceptable implausibility value cutoff determines whether an input parameter vector ($x$) is viewed as non-implausible or not. This value can be defined based on various considerations as discussed in Vernon et al. (2013), but often the cutoff used is equal to the critical value of some appropriate distribution, for example the standard normal distribution.

Non-Implausible Inputs Evaluation

The ‘non-implausible’ input parameters obtained at the end of the process represent the input parameters of the reservoir
simulation model, whose outputs match to the hypothetical historical production data. These parameters are evaluated to identify how much production data improved the reservoir model understanding.

While carrying out these analyses considering different production periods, it is possible to evaluate the impact of the production period over the reservoir uncertainty reduction.

![Figure 1 - Process to perform uncertainty reduction quantification.](image-url)

**Results**

This item shows the application of the workflow described in the previous topic to a synthetic reservoir model built to represent a region of an injector and related producers. The uncertainty reduction was quantified considering two different production periods: the first at an early stage of production (1000 days) and the second at an intermediate stage of production (3500 days).

**Base Model**

The reservoir simulation model designed in the field development phase is called base model. In this study the base model consists of a five-spot configuration and is structurally represented by a horizontal top at -1000 m, discretized with a 45 x 45 x 1 grid in the x, y and z directions, respectively, with a dimension of 40 m in the three directions, totaling 2025 blocks. A light oil and Black Oil fluid model was used and presents a constant permeability equal to 500 mD and a constant porosity equal to 20%. This model takes approximately 10 seconds to be simulated. The base model were built by Risso (2007) and modified by Machado (2010). The permeability map is presented in Figure 2:
Input and Output Definition
The reservoir model uncertain input parameters that make up the vector x and that parameterize the reservoir geology containing a channel, are shown in Figure 3; a description and the ranges of these inputs are shown in Table 1.

Table 1 – Input parameters and associated ranges.

<table>
<thead>
<tr>
<th>Input Parameter</th>
<th>Description</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>(x_c)</td>
<td>channel Cartesian x center value</td>
<td>grid cell 5</td>
<td>grid cell 41</td>
</tr>
<tr>
<td>(y_c)</td>
<td>channel Cartesian y center value</td>
<td>grid cell 5</td>
<td>grid cell 41</td>
</tr>
<tr>
<td>(\theta)</td>
<td>channel angle</td>
<td>0</td>
<td>(\pi)</td>
</tr>
<tr>
<td>(w_c)</td>
<td>channel width</td>
<td>(2\sqrt{2})</td>
<td>(5\sqrt{2})</td>
</tr>
<tr>
<td>(L_c)</td>
<td>channel length</td>
<td>0</td>
<td>26 grid cells</td>
</tr>
<tr>
<td>(k_c)</td>
<td>channel permeability</td>
<td>1000mD</td>
<td>3000mD</td>
</tr>
<tr>
<td>(k)</td>
<td>reservoir permeability</td>
<td>200mD</td>
<td>600mD</td>
</tr>
</tbody>
</table>

Seventeen production output parameters were selected to evaluate the impact of production data acquisition in the reservoir model uncertainty reduction. The definition of each is as follows:

- \(f_1(x)\) to \(f_4(x)\): production well 01 to 04 bottom-hole pressure (BHP);
- \(f_5(x)\): injector well bottom-hole pressure (BHP);
- \(f_6(x)\) to \(f_9(x)\): production well 01 to 04 water rate;
- \(f_{10}(x)\) to \(f_{13}(x)\): production well 01 to 04 time to breakthrough (BT).

Input Data Set Sampling
The selection of the first input data sampling was obtained through the Latin Hypercube sampling method. Two hundred vectors of inputs $x$, from the initial input space, was sampled generating 200 scenarios. The probability distribution of all uncertainties was considered uniform. Figure 4 shows 2 dimensional projections of the locations of the 200 input vectors used.

The generated scenarios were simulated to obtain the production outputs $f(x)$. The sampled input parameters and resulted simulation outputs were used to estimate the emulator in the first iteration (Wave 1). The initial input space is reduced at the end of the Wave 1 analysis due to the imposition of the implausibility cutoff; the new input space then consists of the non-implausible input parameters; those whose outputs may match the hypothetical historical data.

In order to improve the emulators’ quality and reduce even more the input space, a new data sample is obtained using the LHD from the non-implausible input space derived from the Wave 1 analysis. The Wave 2 analysis consists of estimating new emulators using this new Wave 2 data sample. The quantity of iterations (Waves) depends on the emulator quality needed, reduction in the input parameter space, computational and time resources.

**Figure 4 – First set of input data values, generated from a LHD of size 200.**

**Emulator Estimation and Diagnostic**

To estimate the emulator some assumption were adopted:

- As the study is performed with hypothetical historical data, no measurement errors are considered. Therefore, in this study the observational errors $e$ is equal zero;
- There is no structural discrepancy between the simulator and real physical system output, thus, $\varepsilon$ is equal to zero.

Three interactions (Wave 1, 2 and 3) were needed to obtain the non-implausible input space $X(z)$ for each of the production period analyzed. For the period of production equal 1000 days, the output $f_{10}(x)$ to $f_{13}(x)$ were not used, since there is no water breakthrough at the production wells up to this period. Moreover, in both cases for Wave 1 analysis only BHP outputs were emulated; the water rate linear models obtained were not judged to be accurate enough based on their diagnostics.

**Implausibility Analysis**

To determine the ‘non-implausible’ inputs a hypothetical reality was selected from the initial input space. The hypothetical reality used has a high permeability channel; its position and permeability values are shown in Figure 5 and its input values are presented in Table 2.
Table 2- Hypothetical reality input parameters values.

<table>
<thead>
<tr>
<th>Input Parameter</th>
<th>Description</th>
<th>Hypothetical Reality Value</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_c$</td>
<td>channel Cartesian x center value</td>
<td>19.6</td>
<td>grid cell</td>
</tr>
<tr>
<td>$y_c$</td>
<td>channel Cartesian y center value</td>
<td>36.4</td>
<td>grid cell</td>
</tr>
<tr>
<td>$\theta$</td>
<td>channel angle</td>
<td>2.47</td>
<td>rad</td>
</tr>
<tr>
<td>$w_c$</td>
<td>channel width</td>
<td>17.7</td>
<td>grid diagonal</td>
</tr>
<tr>
<td>$L_c$</td>
<td>channel length</td>
<td>5.4</td>
<td>grid</td>
</tr>
<tr>
<td>$k_c$</td>
<td>channel permeability</td>
<td>2000.5</td>
<td>mD</td>
</tr>
<tr>
<td>$k$</td>
<td>reservoir permeability</td>
<td>274.7</td>
<td>mD</td>
</tr>
</tbody>
</table>

Each vector from the input parameter space is evaluated to determine if the output parameter obtained using the emulator may match the hypothetical reality output. This evaluation is performed by analyzing the implausibility value obtained. In the case studied the maximum implausibility value cutoff was chose to be equal to the 99% critical value of the corresponding standard normal distribution, and hence set to 2.576.

A vector from the initial input parameter space is considered non-implausible if the implausibility value is less than the cutoff using the emulators obtained in Wave 1, 2 and 3 analyses. Table 3 shows the reduction in the parameter space as the implausibility analysis is performed over the three waves. The results obtained show the importance of using an emulator in the uncertainty quantification reduction. The volume of input parameter space considered non-implausible was found to be a small proportion of the original input space: a reduction of $10^{-6}$ in the 3500 days case.

Table 3- Number of input parameters considered non-implausible.

<table>
<thead>
<tr>
<th>Analysis Phase</th>
<th>Period of Production Evaluated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1000 days</td>
</tr>
<tr>
<td>Initial Input Space</td>
<td>1,000,000</td>
</tr>
<tr>
<td>Wave 1</td>
<td>11,948</td>
</tr>
<tr>
<td>Wave 2</td>
<td>617</td>
</tr>
<tr>
<td>Wave 3</td>
<td>3</td>
</tr>
</tbody>
</table>

Non-Implausible Inputs Evaluation

To obtain a significant number of non-implausible input parameters an initial space of $8e+08$ vectors were used. The input space considered non-implausible after all Waves analyses for the production periods equal to 1000 and 3500 days is shown in Figure 6 and Figure 7, respectively.

Each square shows the relation between the corresponding variables; the colors are related to how close the emulator output obtained using a certain input parameter value match the hypothetical output value. Red and pink colors represent values closer and not so close, respectively, to the hypothetical output value.
At an early stage of production, Figure 6, it was possible to identify with high accuracy the field permeability. The range of x and y channel position was narrowed, but channel length values equal to zero was possible to obtain. Zero values to channel length indicate no channel exists. It was not possible to obtain significant information about the channel permeability, angle and width at this stage.

For an intermediate production period, Figure 7, a significant uncertainty reduction is obtained using production data. In addition to the field permeability, x and y channel position are close to the hypothetical reality values. It was possible to better understand the channel length, however no significant information was obtained about the channel permeability and width, perhaps suggesting a limit to the amount of information that can be obtained from this production data (or that more waves could be required).

The uncertainty reduction can also be seen in Figure 8, Figure 9 and Figure 10. The cumulative oil production, bottom-hole pressure and water rate for the production wells are presented for the initial input data set (red lines), scenarios obtained after the uncertainty reduction at 1000 (green lines) and 3500 (cyan lines) days, with the reality model shown as a single dark blue line.

There is a significant uncertainty reduction using production data up to 1000 days for most of the wells, however for production well 03 there are still high uncertainty. The uncertainty at production well 03 was reduced using production data up to 3500 days. Note the strong agreement for several outputs between the cyan lines and the dark blue line of the reality model, implying that we have found many locations in input space that are consistent with the observed data. The agreement at late times also implies that we could make accurate predictions of the future behaviour of the reality model based solely on the data at 1000 and 3500 days.

Figure 6 - Non-implausible inputs for period of production equal 1000 days.
Figure 7 - Non-implausible inputs for period of production equal 3500 days.

Figure 8 – Uncertainty reduction: cumulative oil.
Figure 9 - Uncertainty reduction: bottom-hole pressure.
Conclusions

A workflow to determine the input parameters whose output values match to historical data using emulation techniques was presented. The workflow was successfully applied to a five-spot synthetic case that was built to represent a region of an injector and related producers. The uncertainty reduction of a reservoir model due to new information acquisition for different production periods was quantified. The field production data used was obtained by considering a hypothetical reality among all possible scenarios, since the analysis was performed at the development stage and used a synthetic model. Two periods of production were evaluated: at an early production stage (1000 days) and at an intermediate production stage (3500 days).

The results obtained showed the importance of using emulators in the uncertainty reduction quantification and history matching process. The number of input parameters considered non-implausible was a small set of the initial input space. At an early stage it was possible to reduce the uncertainty by identifying the hypothetical real field permeability and identifying possible values for channel positioning. However, other important physical features were not identified, such as the channel permeability and width. At an intermediate stage, the uncertainty reduction was higher. However, still some important physical features that impact on production prediction, such as channel permeability and width were not identified; therefore, further steps of this research will test the application of the emulation technique with seismic 4D data to reduce uncertainty.

Nomenclature

LDH  Latin Hypercube Design
x    input vector
f(x)  output vector
y    real physical outputs vector
z    field observed outputs vector
e    random observational errors vector
ε    random structural discrepancy vector
X(z)  input parameter space
\( \hat{f}(x) \)  emulated output
i    output emulated
j    number of functions
β    scalar
g    deterministic function
u    local variation
E    expected value
Var  variance
Cov  covariance
σ    standard deviation
R²   squared multiple correlation
RSS  residual sum of squares
RYY  total sum of squares
n    number of data points
p number of parameters to be estimated
I implausibility value
x, y, z Cartesian directions
x_c channel Cartesian x center value
y_c channel Cartesian y center value
θ channel angle
w_c channel width
L_c channel length
k_c channel permeability
k reservoir permeability

References


