Risk Analysis Speed-up with Surrogate Models
Tiago Amorim, SPE, Petrobras; and Denis J. Schiozer, SPE, UNICAMP

Abstract
Risk analysis is crucial in investment decisions. A more accurate risk analysis for a field development study can demand a large number of simulation runs, which can lead to high computational time. Some techniques have been developed to reduce the number of runs, such as experimental design with surface response methodology. One problem usually associated with this technique is the possibility of lower reliability associated with complex problems. Furthermore, they might not properly represent the problem when there are changes in the uncertain parameters, in this case a complete restart in the process may be necessary. An alternative is proposed here through the use of fast surrogate simulation models that generate results similar to the base model, even with changes in the reservoir attributes.

The surrogate simulation model has the same data as the base simulation model, but with a much coarser grid. The coarse grid parameters are adjusted automatically with a gradient-based optimization algorithm, minimizing the difference between the responses from the base and the surrogate models. Due to the large number of variables to adjust, several techniques were incorporated in the optimization algorithm: simultaneous perturbation stochastic approximation, response surface methodology and data partition. After this adjustment, a risk analysis can be conducted with the surrogate model.

Simulation models were constructed to test the results generated with the proposed surrogate model methodology. Sensitivity analysis for several factors has shown acceptable adherence of the coarse and base models. A risk analysis was conducted with both coarse and base models, the results generated with the coarse models were close to those with the base models. Overall, the time spent in adjusting the coarse model and generating responses for the risk analysis was smaller than directly using the base model in a risk analysis.

The main contribution of this work is to develop a methodology to construct fast surrogate models and to show that they can help to reduce the time needed to build a risk analysis, generating results that are similar to the full simulation model.

Introduction
Petroleum production is a risky business. Investment decisions must be made under uncertainty and a proper risk analysis of the development of a new petroleum field is essential. A large number of simulation runs must be made to generate a probabilistic estimate of production. Different techniques were developed to reduce the computational time: simplifications in the statistical analysis, for instance, can be applied through the derivative tree technique; reduction in the number of simulation runs can be applied through the use of experimental design with surface response methodology (proxy models).

The use of proxy models has increased recently but a possible problem related to this technique is the lack of flexibility and lower reliability associated with complex problems. It is not uncommon for the reservoir engineer to perform changes in the uncertain parameters and due to the lack of flexibility, the process may need to be restarted.

An alternative to proxies are the reduced-order models which try to describe the main effects of a high-order system with a low-order model. An application has been developed to replace the simulation model in an optimization process (Cardoso and Durlofsky, 2010). A major difficulty in generating reduced-order models is that it is not always possible do access the mathematical models used in commercial flow simulators (Heijin et al., 2003).

An option to try to overcome these difficulties is proposed here with the use of a surrogate model, instead of proxies generated by the response surface methodology, to replace the base model. The surrogate model proposed in this study is a simulation model that uses the same data as the base model, but has a much coarser grid. The grid properties and relative permeability curves are adjusted so that the surrogate model yields results similar to the base model. Due to the large number of variables to adjust, several different techniques were incorporated in the optimization process to reduce the number of
simulation runs needed. With a much faster simulation model, sensitivity and risk analysis studies can be performed with minimal computational effort.

**Fast Surrogate Model**

Proxy models, such as polynomial regression and multivariate kriging models, are widely used in risk analysis studies. These proxies are regression models that are fitted to the results of a series of experiments performed with the simulation model. Their effectiveness to emulate the simulator responses is closely related to the quality of the experimental points generated and the complexity of the problem (Kleijnen, 2005). A history matching or an optimization process may need more reliable responses from the proxy models. Due the inherent non-linearity of a simulation model, proxies generally do not generate good results for complex problems (Zubarev, 2009).

The work presented here investigates the use of a faster version of the simulation model as a proxy. This faster simulation model uses the same Black-oil simulator as the base model, but the grid is much coarser (Fig. 1). All the other information present in the base simulation model is copied: PVT data, well constraints, contacts etc. To avoid confusion with the classic proxy models, this faster simulation model will be named surrogate model in this report. Unlike the proxy models, the proposed surrogate model uses the same formulation as the base model.

The new grid is built starting with the well locations (Fig. 2). A Delaunay Triangulation (de Berg et al., 2008) is performed with the well locations, and additional points are inserted between the well locations and around the convex hull of this group of points. Finally, a new triangulation is performed with all the points created. Each point is associated to a group of cells from the base model, thus forming the cells of the new grid. And each edge between the points represents the connections between the cells. The new grid is inserted in the Black-oil simulator as a line of cells of varying geometry, connected to one another through non-neighbor connections, divided vertically into 3 layers. Due to this simplification the simulation suffers from large grid effects; therefore, not only the grid properties need to be adjusted, but also the relative permeability curves. The geometry of each new cell (top, bottom, porous volume) is adjusted in order to represent the group of cells from the fine grid associated to it. The variables of the surrogate model represent the main characteristics of flow in porous media phenomena: volumes in place, transmissibility between cells, well productivity indexes and relative flow. Details on the definition of the surrogate model variables will not be discussed in this paper, but can be found in the complete report of this study (Amorim, 2012).

The surrogate model responses are adjusted to the base model with a gradient-based optimization technique. The objective function of the optimization process is the difference in terms of production/injection data between the models: cumulated production/injection of oil/gas/water and bottom-hole pressures of each well. Due to its robustness and the fact that only the gradient of the objective function is necessary, the Broyden-Fletcher-Goldfarb-Shanno (BFGS) Quasi-Newton method was selected (Gao and Reynolds, 2004).

Although the grid has a small number of cells when compared to the base model, the number of variables in the optimization process is very large. In a model with eight wells, there are approximately 400 variables. Due to the large number of variables, several techniques were incorporated to the optimization algorithm to reduce the number of simulation runs needed.

The optimization process is comprised of several stages (Fig. 3). The objective function is successively split in several components as the number of variables increases. On the first stage all variables of the same kind are treated as one, e.g., all cells will have the same pore volume modifier. On the second stage the elements are divided into two groups, and all variables of the same kind in each group will have identical values. The division is repeated until there is only one well in each group. At every stage the objective function is split between each group of elements, and the gradients of each component of the objective function are calculated simultaneously. Since the objective function is comprised of the deviations of each well in the model, the gradient associated to any group of elements is the sum of the gradients of the wells that belong to this group (Ding, 2010).

Another time-consuming operation, in the optimization process, is the numerical calculation of the gradient. Two different approaches were used to reduce the number of simulations needed. When the number of variables to be adjusted is equal to or smaller than 15, a proxy model is built around the point of interest to calculate de gradient of the objective function (Barton, 2009). If more than 15 variables are adjusted, a simultaneous perturbation (Spall, 1998) is used to estimate the direction of the gradient.

**Applications**

To access the quality of the results generated with the proposed methodology, two risk analyses were made with each flow model. First, a risk analysis was created with the base model, using either Latin Hypercube or Experimental Design and Response Surface Method, whichever seemed more appropriated for the flow model. Then a surrogate model was adjusted to the base model and used in a risk analysis with the same uncertain variables as the first analysis. The resulting risk curves were compared to verify the misfits in the risk analysis performed with the surrogate model.

Three synthetic models were tested with the proposed methodology. All models are Black-oil simulations of a water injection application. Since the algorithm developed to generate the surrogate model has some limitations, the simulation models have some limiting characteristics: all wells are vertical, there is only one reservoir region with a single oil-water contact and the models do not have any faults.
The first simulation model, model #01, has 17 wells: 9 oil producers and 8 water injectors (Fig. 4). Both producers and injectors are constrained only by bottom-hole pressure limits. The simulation grid is 51x28x6, with 7482 active cells. Since both models are very fast – the base model runs in 14s, and the surrogate in approximately 1.2s – a risk analysis with 5 uncertain variables was conducted directly with each model using a Latin Hypercube design with 2000 simulation runs.

Model #02 has a larger simulation grid, 150x60x18, with 125118 active cells (Fig. 5). There are 9 oil producers and 7 water injectors. As in model #01, the wells only have BHP constraints. The base model runs in 275s, and the surrogate model in 1.9s. After adjustment of the surrogate model to the base simulation model, both were used to generate a risk analysis with experimental design and response surface methodology with the same 6 uncertain parameters.

The last model, model #03, has a larger number of wells: 59 oil producers and 17 water injectors (Fig. 6). The simulation grid is 90x162x22, with 136797 active cells. The wells have both rate and BHP constraints, as well as group rate constraints. The base model runs in 767s. Since the surrogate model grid size is closely related to the number of wells in the base model, it is not as fast the two other surrogate models, and runs in 14.5s. A risk analysis with proxy models generated by Experimental Design and Response Surface Method was used with each model.

Results
As discussed previously, the adjustment process of the surrogate model is performed in stages (Fig. 7). This approach, together with the simplifications introduced in the gradient calculation, enabled the process to be performed in adequate computer time, while generating acceptable results. All three surrogate models generated satisfactory results after the adjustment (Fig. 8, Fig. 9 and Fig. 10). In terms of cumulative oil production, the surrogate models predict values no more than 7% apart from the base models. While a 7% error might be unacceptable for an optimization process, it is good enough in a risk analysis. The range of results in a risk analysis is very wide, and random errors tend to be cancelled out. Still, systematic errors can have a meaningful impact on the results and must be properly addressed.

With model #01 it was possible to generate a risk analysis directly with the base model, using a Latin Hypercube design. The surrogate model estimated the percentile values 2 to 3% higher than the actual values generated with the base model (Fig. 11). Since the difference between the base and the adjusted surrogate model was 2%, this discrepancy between the percentiles is within the expected error. The simulation runs created for the risk analysis were compared (Fig. 12) and the results indicate that nearly all 2000 runs of the surrogate model deviated no more than 9% from the base model (Fig. 13).

For both models #02 and #03, comparison between the results from the base and surrogate models with their respective proxies show very good agreement (Fig. 14, Fig. 15, Fig. 16 and Fig. 17). The risk analysis generated with the surrogate models yielded results approximately 3 to 7% lower than the values generated with the base model (Fig. 18 and Fig. 19).

A quick inspection of the shape of the risk analysis curves shows the surrogate curve is almost parallel to the base curve (Fig. 11, Fig. 18 and Fig. 19), and this indicates the surrogate models have a sensitivity to the uncertain variables very close to that of the base models. Since the surrogate models used to construct the risk analysis did not generate the exact same results as the base models, deviations in the risk curves results are expected. One possible solution to improve the estimates is to correct the surrogate models responses in the risk analysis by the same deviation seen between the adjusted surrogate and base models. The percentile values generated with the surrogate models were divided by the ratio of cumulated oil production between the base and the adjusted surrogate models, e.g., the percentile values of the risk analysis for model #01 were divided by 1.02. The estimated risk analysis curves after this correction are improved for models #01 and #02 (Fig. 20 and Fig. 21). For model #03 this correction of the surrogate model risk analysis was not enough to create a good fit of the risk analysis generated with the base model (Fig. 22).

To access the quality of the estimates generated with the surrogate models, additional runs of the base model with the most influential parameters indicate the expected deviations in the risk analysis. A sensitivity analysis performed with the surrogate model for model #03 indicates porosity as one of the most influential parameters (Fig. 23). Comparison between the surrogate and base models after changes in the overall porosity (Fig. 24 and Fig. 25) corroborates the difference between the risk analysis curves (Fig. 19), i.e., bigger deviations were expected around the smaller values of accumulated oil production. These comparisons between the models were used to correct the surrogate model risk analysis. Sensibility to the water-oil contact and porosity were used to construct a correlation to correct responses from the surrogate model (Fig. 26). The results show a very good agreement between the risk analyses generated with the base and surrogate models (Fig. 27).

An analysis of the deviations between the base and surrogate models risk analyses shows why model #03 required a correlation to correct the values estimated with the surrogate model (Fig. 28). For models #01 and #02 the curves are almost vertical, and so a single correction value for all points is enough to correct the estimates. Model #03 needed a closer inspection of the deviations of the surrogate model, and the proposed correction was able to reduce the gap between the risk analysis curves (Fig. 29). Since these error curves will not be available in an application of this methodology, the procedure exemplified with model #03 to estimate the deviations of the surrogate model should always be applied.

In terms of computational effort, the studies conducted in this work show a great reduction in the total simulation time needed to compute a risk analysis with the surrogate models (Table 1). These gains are even larger if more than one risk analysis study is to be performed and for more complex models.
Table 1 – Time spent to generate each risk analysis.

<table>
<thead>
<tr>
<th>Model</th>
<th>Time to run [s]</th>
<th>Adjust.</th>
<th>Risk Analysis Type</th>
<th>Runs</th>
<th>Total time [min]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Base</td>
<td>Surr.</td>
<td>Var.</td>
<td>Type</td>
<td>Sur.**</td>
</tr>
</tbody>
</table>
| 1     | 14   | 3     | 78  | 5    | LHC  
|       |      |       |     |      | 2000 465 171 |
| 2     | 275  | 5     | 99  | 6    | ED+RSM  
|       |      |       |     |      | 45  206 102 |
| 3***  | 767  | 11    | 338 | 6    | ED+RSM  
|       |      |       |     |      | 45  576 397 |

* Risk analysis only.
** Adjustment and risk analysis.
*** Includes additional simulations to correct responses from surrogate model.

Conclusions
The results generated with surrogate models in risk analysis studies have shown good adherence to the results generated with the full simulation models, but with great reduction in the computational effort. Several sensitivity and risk analysis studies can be undertaken with the surrogate models prior to performing one final risk analysis with the full model, greatly reducing the time needed to perform all simulation runs.

The algorithm developed to construct the surrogate model can only handle a limited range of applications, but has performed well with the three tests carried out in this study. The quality of the responses generated with the surrogate models must be accessed to ensure reliable estimates.

The research has shown the surrogate model is a promising substitute to proxies, but additional studies must be undertaken. The next step is to further enhance the generation of the surrogate model: study the flow lines of the base model to better define the new grid geometry, include new features such as deviated wells and faults, and develop specialized objective functions to different classes of variables of the adjustment process.

Acknowledgments
The main author would like to thank Petrobras for allowing publication of the results presented here.

References
Fig. 1 – Comparison between fine and coarse grids.

Fig. 2 – Construction of the new simulation grid.

1 group 2 groups

4 groups 8 groups

Fig. 3 – Stages of the optimization process.

Fig. 4 – Permeability for model #01.

Fig. 5 – Porosity for model #02.

Fig. 6 – Initial oil saturation for model #03.

Fig. 7 – Adjustment process for model #01.
Fig. 8 – Comparison between the base and surrogate models after adjustment for model #01.

Fig. 9 – Comparison between the base and surrogate models after adjustment for model #02.

Fig. 10 – Comparison between the base and surrogate models after adjustment for model #03.

Fig. 11 – Risk analysis generated with the base and surrogate models for model #01.

Fig. 12 – Comparison between the base and surrogate models.

Fig. 13 – Histogram of the errors generated with the surrogate model.

Fig. 14 – Comparison between the base model and proxy constructed with it for model #02.

Fig. 15 – Comparison between the surrogate model and proxy constructed with it for model #02.
Fig. 16 – Comparison between the base model and proxy constructed with it for model #03.

Fig. 17 – Comparison between the surrogate model and proxy constructed with it for model #03.

Fig. 18 – Risk analysis generated with the base and surrogate models for model #02.

Fig. 19 – Risk analysis generated with the base and surrogate models for model #03.

Fig. 20 – Effect of proposed correction on the surrogate model for model #01.

Fig. 21 – Effect of proposed correction on the surrogate model for model #02.

Fig. 22 – Effect of proposed correction on the surrogate model for model #03.

Fig. 23 – Surrogate model sensitivity of the predicted oil production for model #03.
Fig. 24 – Comparison between the base and surrogate models with smaller porosity.

Fig. 25 – Comparison between the base and surrogate models with greater porosity.

Fig. 26 – Correlation used to correct responses from the surrogate model.

Fig. 27 – Effect of correlation on the surrogate model for model #03.

Fig. 28 – Error in risk analyses generated with the surrogate models, before correction.

Fig. 29 – Error in risk analyses generated with the surrogate models, after correction.