Use of Machine Learning in Petroleum Production Optimization under Geological Uncertainty

Obiajulu J. Isebor
Ognjen Grujic

December 14, 2012

1 Abstract

Geological uncertainty is of significant concern in petroleum reservoir modeling with the goal of maximizing oil production. Stochastic simulation allows generating multiple reservoir models that can be used to characterize this uncertainty. However, the large computation time needed for flow simulation (e.g., for use in production forecasting) impedes the evaluation of flow on all reservoir models. In addition, performing a formal optimization of the well controls to maximize say NPV leads to hundreds or thousands of function evaluations, each of which requires tens to hundreds of reservoir simulations depending on the number of reservoir models available.

In this work we apply machine learning techniques to provide computational savings on two fronts. We use kernel k-means clustering to select a small representative set of earth models that characterize the geological uncertainty so as to reduce the number of simulations for each optimization function evaluation, and use a kriging surrogate in the optimization to reduce the required number of function evaluations.

2 Introduction

The task of optimizing the production from an oil field is a tough one fraught with theoretical and computational challenges centered around the uncertainties that are an intrinsic part of the problem. Firstly, as the petroleum reservoir from which we plan to produce is below the surface of the earth, we only have approximate models of what we think the reservoir looks like. In attempts to capture this geological uncertainty in the reservoir description, tens to hundreds of probable earth models with different reservoir properties and property distributions are typically generated using stochastic simulation.

Generating these different realization models is not a goal in and of itself but the use of the models to understand the uncertainty in modeling some response function is what we are after. The response function is typically a measure of the performance of the field, e.g., cumulative oil production or net present value (NPV), that is of interest to decision makers in charge of managing the field's production. In recent years, techniques from the field of numerical optimization have been applied to optimize our measure of field performance so that we can plan to get the best out of our field moving forward in time, given the information we know about the field right now.

Our project involved the investigation of the use of machine learning methods to classify the tens to hundreds of earth models, and select a small representative subset, that still represents the uncertainty range, to use for forecasting or optimization. This is done in order to not run numerical reservoir simulations (which can be quite expen-

sive) on all the models but only a subset of them, while still spanning the range of uncertainty in the desired response (e.g., NPV).

In addition to earth model selection, we would also like to investigate the use of machine learning methods as proxies for the objective function evaluations during the optimization process. For example, having a supervised learning algorithm, like a kriging surrogate, trained to reproduce the response might provide significant savings in an optimization process that requires running hundreds or thousands of simulations by replacing a good number of these simulations (which could take minutes to hours to run each, depending on the size of the reservoir model) with approximations from the kriging surrogate (which runs in a fraction of a second).

3 Problem statement

The problem we aim to solve is the optimization of well controls in a petroleum field produced under waterflooding, where water is pumped through injection wells in order to maintain reservoir pressure and displace the resident oil toward production wells. Simulating the flow through the petroleum reservoir, in order to ascertain injection and production volumes for NPV calculations, is done with the use of a reservoir simulator (in this work, we use the Stanford-developed reservoir simulator called GPRS [4]).

The well controls (part of the inputs to the reservoir simulator) being optimized are the injection and production bottomhole pressures (BHPs) and the objective is to maximize undiscounted NPV from the petroleum field. Due to the uncertainty in the reservoir description, the optimization is performed in a robust manner where the objective function is actually the expected value of the projected production given a set of well controls, characterized as the average of NPV over several geological realizations. Formally stated, the optimization problem we aim to solve is given as

$$\max_{\mathbf{x}\in X} \; J(\mathbf{x}) = \frac{1}{N_r}\sum_{j=1}^{N_r} \mathrm{NPV}(\mathbf{x}, \mathbf{m}_j), \qquad (1)$$

where $\mathbf{x}$ represents the vector of well control variables (well BHPs), $X = \{\mathbf{x} \in \mathbb{R}^n;\ \mathbf{x}_l \leq \mathbf{x} \leq \mathbf{x}_u\}$ represents the box constraints for the control variables, N_r is the number of geological realizations, $\mathbf{m}_j$ represents the geological model parameters for realization j, and the undiscounted NPV from each model is given by

$$\mathrm{NPV}(\mathbf{x}, \mathbf{m}_j) = \underbrace{p_o Q_o(\mathbf{x}, \mathbf{m}_j)}_{\text{oil revenue}} - \underbrace{c_{wp} Q_{wp}(\mathbf{x}, \mathbf{m}_j)}_{\text{water production cost}} - \underbrace{c_{wi} Q_{wi}(\mathbf{x}, \mathbf{m}_j)}_{\text{water injection cost}}, \qquad (2)$$

where p_o, c_{wp} and c_{wi} are the price of oil and costs of produced and injected water per barrel, respectively. Q_o, Q_{wp} and Q_{wi} are the cumulative oil and water production and water injection in barrels, respectively. These are the outputs from the reservoir simulator that are required for the NPV computation. Figure 1 below presents the workflow for the average NPV calculation for running example used in the project. This example involves using 45 geological realizations (15 each produced from stochastic simulations assuming 3 different depositional environments). The field has eight wells, four production (red circles in figure) and four injection (blue circles in

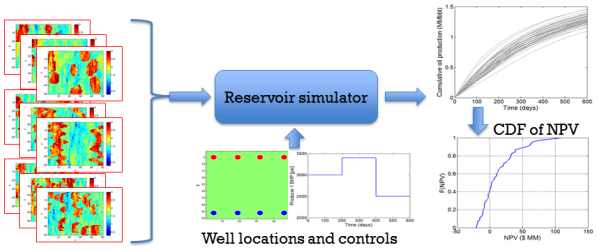


Figure 1: Illustration of computation of NPVs for full set of realizations.

figure) wells. With these well locations together with the initial specified well controls of constant injection and production, the 45 reservoir simulations are run, resulting in 45 injection and production profiles (only oil production profiles shown in Figure 1) from which we can calculate 45 NPV values, shown as an empirical cumulative density function (CDF) in Figure 1. From these 45 NPV values, we can calculate the average NPV, <NPV>, which is the objective function value for the initial set of controls.

4 Machine learning for computational savings

From equations (1) and (2) we see that each evaluation of the objective function in the optimization requires running N_r reservoir simulations (45 simulations in our example). In addition, the optimization process can require hundreds to thousands of function evaluations, depending on the complexity of the problem and the optimization algorithm used. It is no surprise that this optimization process considering the different geological realization can be quite expensive. As a result, we investigate using machine learning techniques to provide computational savings on two major fronts: reducing the number of realizations using clustering, while still effectively characterizing the geological uncertainty in the problem, in order to reduce the simulations needed for each objective function evaluation; and using an optimization approach that uses a kriging surrogate to reduce the number of function evaluations that require reservoir simulations.

4.1 Clustering and earth model selection

The idea of using k-means clustering for earth model selection to represent uncertainty in model responses using a subset of the original models was introduced by Scheidt and Caers in 2009 [6] (details of the approach can be found in this paper). In Figure 2 we present an illustration of the process for our running example.

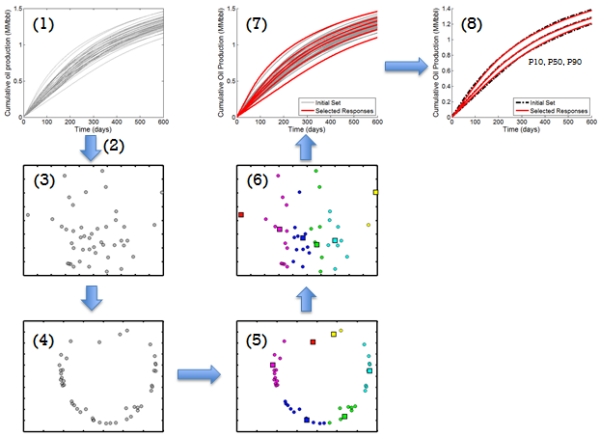


Figure 2: Earth model selection process.

The process illustrated in Figure 2 shows our clustering and model selection process and the different parts of the process are as follows

1. Given the initial well controls, we generate the model responses (cumulative oil production profiles) for all 45 reservoir models using the process illustrated in Figure 1.

2. Find pairwise distances between the the model responses.

3. Map with multi-dimensional scaling (MDS) into a 2D MDS space, as is done in [6].

4. Transform using a Gaussian kernel into a low-dimensional feature space where the points for each model are more separable.

5. Perform k-means clustering to cluster the data into six clusters (different colors in Figure 2) and identify the models that are the closest to the cluster centroids.

6. Transform back to the MDS space and identify the selected earth models.

7. Identify six selected model responses and show that they span the same uncertainty range as the initial 45 models.

8. Show that the selected models reproduce the ensemble statistics (P10, P50 and P90) as the initial set of models, again for the initial well controls.

4.2 Surrogate-based optimization

After performing the model selection explained in the preceding subsection, we now need to run the optimization in order to determine the optimal well controls that maximize <NPV>. We applied a surrogate-based search-poll optimization procedure introduced by Booker et al. [3], with a modified implementation from Abramson [1] and illustrated below in Figure 3.

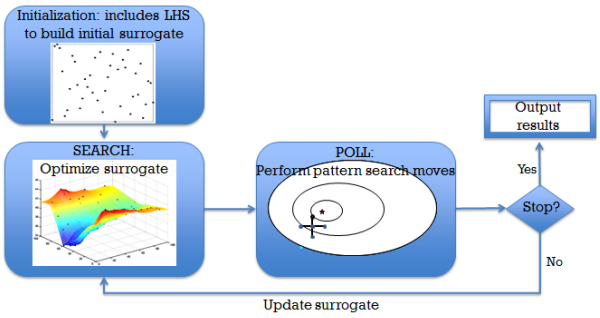
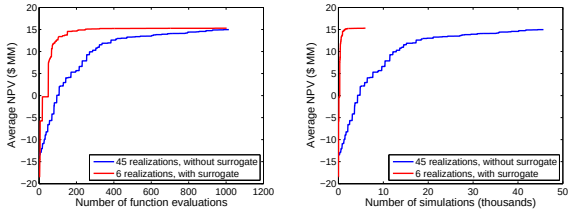


Figure 3: Surrogate-based optimization workflow.

The poll step involves using the generalized pattern search (GPS) algorithm of Audet and Dennis Jr. [2], which uses a stencil-based approach to identify points in optimization parameter space to evaluate the objective function. Before each poll step, there is a search step where a fast-running surrogate is optimized and the resulting well controls applied to the true objective function evaluation. We use a kriging surrogate (from the DACE kriging toolbox [5]), which is essentially a response surface approximating the true objective function surface. The initial surrogate is built from points from a Latin hypercube search (LHS) and updated, every time the polling step is unsuccessful, using the new points that were evaluated in the polling process.

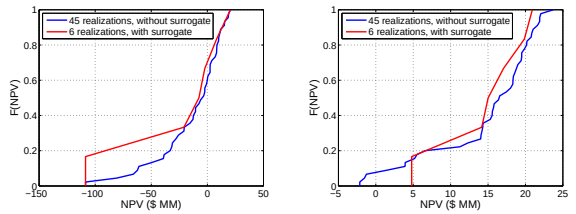
5 Results

We ran the optimization codes on our well control problem for several cases, with the original 45 models and with the selected 6 models, as well as with and without the use of the kriging surrogate in the optimization process. Our project results are summarized below in Figure 4. In Figure 4a we



(a) Improvement due to surrogate use

(b) Improvement due to model number reduction



(c) Initial NPV CDFs

(d) Final NPV CDFs

Figure 4: Performance result comparisons.

show the improvement in computational efficiency of optimization with and without surrogates. We see that with the use of the kriging surrogate, there is a rapid increase in <NPV> after only a few true function evaluations because we are doing a lot of the optimization on the surrogates and running the results on the true function evaluation. To get an idea of the timing, one surrogate function evaluation is a small fraction of a second, where each true function evaluation is about 30 seconds, for this simple case. The point of Figure 4a is that we get about the same optimal solution but with much less function evaluations.

In Figure 4b we show the improvement in computational efficiency due to reduc-

ing the number of reservoir models that characterize the uncertainty in the problem. For the case with N_r realizations, each function evaluation corresponds to N_r reservoir simulations and thus reducing the number of models needed to characterize uncertainty directly leads to a reduction in number of simulations needed for the optimization. Combine this with the improvement in computational efficiency from including surrogate use in the optimization and we get the comparison shown in Figure 4b.

Figures 4c and 4d present the initial and final CDFs respectively for the cases where we optimize <NPV> from all initial 45 models without using a surrogate and where we optimize <NPV> from only the six models with surrogate use in the optimization. The figures show that the six-model representation of uncertainty is just as good as the one with 45 models. The optimization is successful in improving the performance from the reservoir. Initially, <NPV> was negative meaning that the project is unprofitable under the scenario where we implement the initial well controls. After optimization, we improve <NPV> to about $15 million. Our optimized well controls are also more robust because the range of NPVs in the final CDF is $20 million, compared with about $130 million for the initial well controls.

6 Concluding remarks

In this work, we successfully implemented an approach for optimizing well controls in petroleum production under geological uncertainty. This process can be computationally expensive but we present machine learning approaches to improve the computational efficiency of the optimization approach.

We showed that MDS and kernel k-means clustering is suitable for selecting a representative subset of earth models. For our waterflooding well control optimization example, we showed that surrogate-based optimization, using a kriging surrogate, significantly reduces the optimization computational cost. Our results show that the optimized well control obtained with our surrogate-based approach run on the selected subset of reservoir models is comparable in terms of quality of solution with the full optimization with 45 models, while requiring significantly less computational expense.

References

[1] M. A. Abramson. *NOMADm version 4.6 User's Guide*. Department of Mathematics and Statistics, Air Force Institute of Technology, 2007.

[2] C. Audet and J. E. Dennis Jr. Analysis of generalized pattern searches. *SIAM Journal on Optimization*, 13(3):889–903, 2002.

[3] A. J. Booker, J. E. Dennis Jr., P. D. Frank, D. B. Serafini, V. Torczon, and M. W. Trosset. A rigorous framework for optimization of expensive functions by surrogates. *Structural Optimization*, 17:1–13, 1999.

[4] H. Cao. *Development of Techniques for General Purpose Simulators*. PhD thesis, Department of Petroleum Engineering, Stanford University, 2002.

[5] S. N. Lophaven, H. B. Nielsen, and J. Sondergaard. DACE: A Matlab kriging toolbox, version 2.0. Technical report, Technical University of Denmark, 2002.

[6] C. Scheidt and J. Caers. Uncertainty quantification in reservoir performance using distances and kernel methods: Application to a west africa deepwater turbidite reservoir. *SPE Journal*, 14(4):680–692, 2009.