Computers & Geosciences 96 (2016) 109–123

Contents lists available at ScienceDirect

Computers & Geosciences

journal homepage: www.elsevier.com/locate/cageo

Research paper

A general method to select representative models for decision making and optimization under uncertainty

Mehrdad G. Shirangi *, Louis J. Durlofsky

Department of Energy Resources Engineering, Stanford University, Stanford, CA 94305, United States

A R T I C L E I N F O

Article history:
Received 4 March 2016
Received in revised form
24 June 2016
Accepted 1 August 2016
Available online 3 August 2016

Keywords:
Reservoir simulation
Subsurface flow
Representative realizations
Robust optimization
Optimization under uncertainty
Production optimization
K-means clustering
K-medoids
Unsupervised learning
Feature selection
Well placement
Model selection

A B S T R A C T

The optimization of subsurface flow processes under geological uncertainty technically requires flow simulation to be performed over a large set of geological realizations for each function evaluation at every iteration of the optimizer. Because flow simulation over many permeability realizations (only permeability is considered to be uncertain in this study) may entail excessive computation, simulations are often performed for only a subset of 'representative' realizations. It is however challenging to identify a representative subset that provides flow statistics in close agreement with those from the full set, especially when the decision parameters (e.g., time-varying well pressures, well locations) are unknown a priori, as they are in optimization problems. In this work, we introduce a general framework, based on clustering, for selecting a representative subset of realizations for use in simulations involving 'new' sets of decision parameters. Prior to clustering, each realization is represented by a low-dimensional feature vector that contains a combination of permeability-based and flow-based quantities. Calculation of flow-based features requires the specification of a (base) flow problem and simulation over the full set of realizations. Permeability information is captured concisely through use of principal component analysis. By computing the difference between the flow response for the subset and the full set, we quantify the performance of various realization-selection methods. The impact of different weightings for flow and permeability information in the cluster-based selection procedure is assessed for a range of examples involving different types of decision parameters. These decision parameters are generated either randomly, in a manner that is consistent with the solutions proposed in global stochastic optimization procedures such as GA and PSO, or through perturbation around a base case, consistent with the solutions considered in pattern search optimization. We find that flow-based clustering is preferable for problems involving new well settings (e.g., time-varying well bottom-hole pressures) or small changes in well configuration, while both permeability-based and flow-based clustering provide similar results for (new) random multiwell configurations. We also investigate the use of efficient tracer-type simulations for obtaining flow-based features and demonstrate that this treatment performs nearly as well as full-physics simulations for the cases considered. The various procedures are applied to select realizations for use in production optimization under uncertainty, which greatly accelerates the optimization computations. Optimization performance is shown to be consistent with the realization-selection results for cases involving new decision parameters.

1. Introduction

In subsurface flow operations, decisions such as where to locate new wells and how to operate existing wells are best made by evaluating flow simulation results over an ensemble of realizations intended to capture the current state of geological knowledge. Evaluation of the consequences of different sets of decision parameters technically requires computing flow responses, for each case considered, over a large set of realizations. Because computational cost scales directly with the number of realizations employed, it is preferable to use as few realizations as possible. If too few realizations are considered, however, results may not represent the response from the full set, because geological uncertainty is not properly modeled. It is thus evident that, in order to achieve the optimal balance between cost and 'representivity,' the subset of geological realizations used for flow simulation must be selected carefully.

The issue of realization selection is particularly important in computational optimization under geological uncertainty. In commonly used derivative-free algorithms (discussed below), for

* Corresponding author.
E-mail address: mehr@stanford.edu (M.G. Shirangi).

http://dx.doi.org/10.1016/j.cageo.2016.08.002
0098-3004/

example, each iteration may involve, say, 100 function evaluations. However, in order to optimize expected reservoir performance over a set of N_R realizations (to account for geological uncertainty), a single function evaluation requires flow simulation to be performed over all of the realizations considered. If an optimization requires (say) 1000 iterations, this corresponds to $10^5 \times N_R$ flow simulations. If we take N_R to be 100, a total of 10^7 simulations will be required. However, if we can find n_r 'representative' realizations (with $n_r << N_R$) that can approximate the expected flow performance of the full set of N_R realizations, then we will achieve computational savings of a factor of N_R/n_r, which can be very substantial. Consistent with this, our intent here is to present a general framework that can be used to appropriately select a representative set of n_r realizations for use in optimization or decision making. Because the amount of computation required in optimization is so large, it is cost-effective to perform some number of flow simulations in determining the n_r representative realizations.

In this work, we introduce a new realization-selection method based on clustering techniques. The method is intended to provide a set of realizations that are most representative in terms of their flow solutions for new decision parameters such as well controls (e.g., time-varying well injection or production rates, or bottom-hole pressures) or well locations. By representative, we mean that flow results for the subset of realizations are in close agreement with those computed for the full set of realizations. Our procedure is quite general, and incorporates flow-based and permeability-based features (either separately or in combination) in the clustering. We define a low-dimensional flow-response vector that concisely characterizes simulation results and enables us to quantify the representivity of any subset of realizations relative to the full set. The most appropriate features (to be used in the clustering) will then be determined for several different problems involving new sets of well controls or well locations in oil reservoirs.

The problem of selecting a representative subset of realizations from a large set has been previously investigated. In the context of uncertainty assessment for future reservoir production, Scheidt and Caers (2009a) introduced a realization-selection method using kernel k-means clustering and streamline simulation. With this method, a few representative realizations are selected for flow simulation, with the goal that results for particular statistics characterizing future oil production are similar to those for the entire set. Scheidt and Caers (2009b) also proposed a distance kernel method to select a subset of reservoir models that provide an uncertainty range for a particular production response (such as cumulative oil production versus time) in agreement with that of the full set for a base operating scenario. Yeh et al. (2014) applied a similar approach using flow-based features from streamline simulation. Meira et al. (2015) and Rahim et al. (2015) introduced optimization-based methods for selecting a subset of realizations that are intended to be representative of the full set in terms of net present value (NPV) distribution and simulation results. These approaches were applied for a particular well configuration and set of well controls. Armstrong et al. (2013) presented a multistage programming with recourse procedure for selecting a representative subset of realizations in a mineral deposit problem.

Reservoir management often involves investigating the impact (on, e.g., oil recovery) of new decision parameters. This could include sensitivity analysis, where the effect of heuristic changes in decision parameters is investigated, or computational optimization, where decision parameters associated with well locations and/or controls are varied algorithmically to maximize an expected objective such as NPV. Within the context of well placement optimization, a variety of derivative-free approaches have been considered. These include global stochastic search methods such as particle swarm optimization (PSO) (Onwunalu and Durlofsky, 2010, 2011; Nwankwor et al., 2013; Humphries et al., 2014) and genetic algorithms (GAs) (Güyagüler et al., 2002; Yeten et al., 2003; Artus et al., 2006; Bouzarkouna et al., 2012). Local optimization methods, such as NEWUOA (Zhang et al., 2015) and pattern search techniques (Wilson and Durlofsky, 2013), as well as gradient-based methods (Zandvliet et al., 2008), have also been applied. For the optimization of (continuous) well settings, both gradient-based methods (Brouwer and Jansen, 2004; Sarma et al., 2006; Chen et al., 2012) and pattern search procedures (Echeverria Ciaurri et al., 2011) have been considered. A recently developed hybrid algorithm, which entails particle swarm optimization and mesh adaptive direct search (PSO–MADS), has been shown to outperform its component methods for combined well location and control problems (Isebor et al., 2014a,b). This indicates that both global random and local deterministic search components are beneficial for reservoir optimization. Thus, for optimization under uncertainty, representative realizations will need to be identified for both types of searches.

Robust optimization of subsurface operations, in which geological uncertainty is considered by optimizing over multiple realizations, has been addressed in a number of studies (e.g., Yeten et al., 2003; Bayer et al., 2008; van Essen et al., 2009; Tartakovsky, 2013; Isebor and Durlofsky, 2014). Because subsurface flow simulation is usually computationally expensive, a small number of realizations is typically used. Various strategies have been applied in this context to select a representative subset of realizations. For well control optimization, Shirangi and Mukerji (2012) selected representative realizations by applying k-medoids clustering using some flow-based features, while Yasari et al. (2013) selected realizations based on the ranking of NPVs obtained from an initial control strategy. For well placement optimization, Wang et al. (2012) applied k-means clustering, using a few static and simulation-based quantities. Torrado et al. (2015) applied a similar approach using only static features. Yang et al. (2011) selected realizations for the robust optimization of SAGD operations by ranking models in terms of NPV for a base well location and control strategy. Recently, Shirangi and Durlofsky (2015) introduced an 'optimization with sample validation' (OSV) procedure to determine the number of realizations to adequately represent the entire set in optimization problems. Representative realizations were selected from the NPVs computed for a base well configuration and control strategy. Under OSV, the NPVs are re-evaluated, and the number of realizations used in the optimization is increased, if a validation criterion is not satisfied. We note finally that there does not appear to have been extensive study of the impact of the realization-selection procedure on optimization results.

While several approaches for selecting small subsets of realizations have been suggested, it is important to recognize that the appropriate selection method may be different in different contexts. For example, a selected subset that is the most representative for a well control optimization problem (with fixed well locations) may not be the best choice for a well placement optimization problem, as these two problems are sensitive to different geological details. In this work, we attempt to address the realization-selection problem systematically. Toward this end, we first describe a procedure for quantitatively assessing different approaches using a flow-response variable. We then introduce a general clustering method for the selection of representative realizations. In this clustering, each realization is represented by a feature vector composed of a weighted combination of flow-based and geological quantities. Principal component analysis (PCA) is used to express the geology (permeability field) in terms of a small number of features, while flow-based features are obtained by solving one or more flow problems. The use of both full-physics