Journal of Petroleum Science and Engineering 163 (2018) 463–475

Contents lists available at ScienceDirect

Journal of Petroleum Science and Engineering

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

# Fast evaluation of well placements in heterogeneous reservoir models using machine learning

Azor Nwachukwu [a], Hoonyoung Jeong [b,*], Michael Pyrcz [a], Larry W. Lake [a]

[a] University of Texas at Austin, United States
[b] Bureau of Economic Geology, United States

ARTICLE INFO

*Keywords:*
Surrogate reservoir modeling
Machine learning
Well placements

ABSTRACT

Surrogate models, or proxies, provide computationally inexpensive alternatives for approximating reservoir responses. Proxy models are routinely developed to generate spatially-varying output such as field pressures and saturations, or well responses such as production rates and bottom-hole pressures. In this study, a machine learning approach is adopted to predict reservoir responses based on injector well locations. The proxy developed in this work is trained to reproduce reservoir-wide objective functions, i.e., total profit, cumulative oil/gas produced, or net $CO_2$ stored.

Because of the geological complexity of most reservoirs, slight adjustments in injector well locations could yield dramatic changes in the objective function responses. Hence, most proxies do not include well locations as inputs in their formulation. This complex relationship between well locations and reservoir-wide responses makes non-parametric, machine learning-based methods an attractive option. We introduce a machine learning approach in which the primary predictors are physical well locations, and the primary response is a defined objective function such as NPV. The complexity of the response surface with respect to well locations necessitates that we augment the predictor variables with well-to-well pairwise connectivities, injector block permeabilities and porosities, and initial injector block saturations. Introducing well-to-well connectivities yields significant improvements in prediction accuracy. Connectivities are represented by 'diffusive times of flight' of the pressure front, which is computed using the Fast Marching Method.

A handful of training observations are obtained from numerical reservoir simulations. The Extreme Gradient Boosting method is then used to build an intelligent model for making predictions given any set of observations. The proposed approach is demonstrated using five synthetic case studies: i) a homogeneous reservoir waterflood, ii) a channelized reservoir waterflood, iii) a 20-model ensemble waterflood, iv) a $CO_2$ flood in a heterogeneous reservoir, v) a $CO_2$ flood in a heterogeneous reservoir with complex topography. Results show a significant correlation between proxy predictions and reservoir simulation results.

## 1. Introduction

Accurate prediction of reservoir behavior in response to changes in control parameters is an important aspect of reservoir management and optimization. Full-physics numerical simulations currently provide the most accurate approach for estimating reservoir response and performance. Computing resources to run such simulators are however significant, limiting their applicability for problems requiring many forward evaluations. Physics-based proxy models have historically been developed to address this issue. Sayarpour et al. (2007) applied capacitance-resistance models for a quick evaluation of waterflood performance during optimization. The fast marching method has been proposed by Sharifi et al. (2014), Zhang et al. (2014), Leem et al. (2015), Xie et al. (2015), among others, as a method to compute the shape of pressure front propagation in heterogeneous reservoirs. Jeong and Srinivasan (2016) used scaled connectivity analysis for predicting $CO_2$ plume migration. Nwachukwu et al. (2017) coupled a particle tracking proxy with a finite element solver to mimic geomechanical effects of $CO_2$ injection.

Recent advances in artificial intelligence and statistical learning methods have been accompanied by an increase in the development of non-physics based proxies. These approaches use purely data-driven tools to find complex patterns between control parameters and reservoir response. The non-physics based proxies serve as a 'black-box'

* Corresponding author.
*E-mail address:* nwchiaz1@umbc.edu (H. Jeong).

https://doi.org/10.1016/j.petrol.2018.01.019
Received 20 September 2017; Received in revised form 9 December 2017; Accepted 10 January 2018
Available online 17 January 2018
0920-4105/

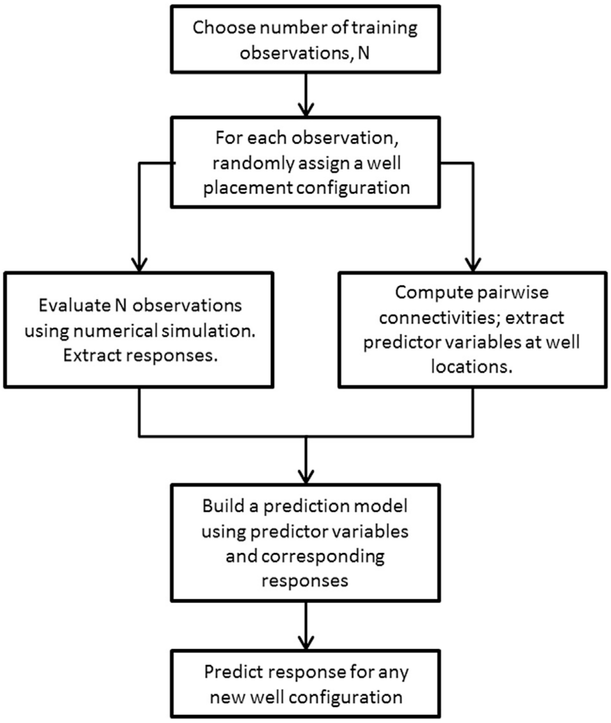


**Fig. 1.** Workflow of the proposed algorithm.

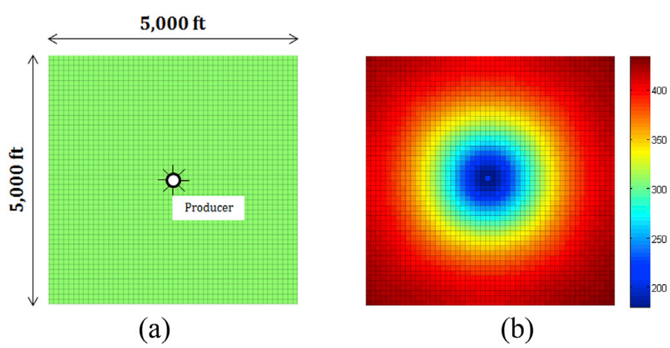


**Fig. 2.** (a) Homogeneous model used for validation, (b) true profit response (in MM$) as a function of injector location.

predictor that operates on input variables such as well rates, locations, and pressures, to produce response variables such as NPV. Sampaio et al. (2009) applied feed-forward artificial neural networks to reproduce the response surface of a heterogeneous reservoir model during history matching. Memon et al. (2014) used radial basis neural networks to predict bottom-hole pressures in an under-saturated reservoir. Amini et al. (2012) built a back propagating neural network to generate $CO_2$ and pressure distributions with high accuracy in a few seconds.

The proxy developed in this study uses Extreme Gradient Boosting (XGBoost), an algorithm based on decision trees, introduced by Chen and Guestrin (2016). XGBoost is a variation of the Gradient Boosting Method (GBM) (Friedman, 2001), designed to be more computationally efficient and flexible. Zhang and Haghani (2015) used GBM to analyze and model time travel along two freeway sections in Maryland. Moisen et al. (2006) applied the gradient boosting method to predict tree species presence and basal area in the mountain ranges of Utah. To the best of our knowledge, this study presents the first application of GBM in surrogate reservoir modeling. The training dataset used to build the proxy is obtained by running sample numerical simulations with different input sets. The goal is to be able to train a model with relatively few simulations such that the response surface can be accurately recreated.

### 1.1. Well placement evaluation

Determining locations for optimal well placement is a habitual issue in reservoir engineering projects. Güyagüler (2002) used the non-gradient-based genetic algorithm (GA) to optimize placement of injection wells in the Pompano field in the Gulf of Mexico. Forouzanfar and Reynolds (2014) proposed a gradient-based scheme to jointly optimize the number of wells, well locations and well controls. In the aforementioned studies, numerical simulators were used to evaluate the impact of well placement on the response variables. Parallel computing resources were employed to speed up the simulation process, however, the large size (number of cells) of the field models still posed significant challenges in terms of computational time.

Because of geological complexities, the problem of training a predictive model to generate response values using spatial well locations as input parameters is not straightforward. Surrogate models are typically developed with the restriction that well locations are predefined and unchanging. There haven't been many studies in which well locations are included as parameters in proxy development. A study by Hassani et al. (2011) presents a promising attempt to address this challenge. They introduce a data-driven approach in which spatial parameters of a horizontal well are used to predict its cumulative oil production. The study was however based on a single homogeneous model and did not present an application to diverse geologic scenarios.

In this study, we enhance the approach presented in Hassani et al. (2011) to account for heterogeneity, and propose a novel idea for improving prediction accuracy. To resolve the challenge posed by geologic complexity, we propose the idea of augmenting physical well locations with well-to-well connectivities as the primary predictor variables. Since connectivities carry implicit information about the reservoir geology, these measures correlate much better with the response surface than spatial locations. In addition to improving prediction accuracy, this also facilitates extrapolation between different geologic models, i.e., one predictive model can be used to accurately generate the response for different geologic models in an ensemble. A demonstration of this is shown in a later section.

### 1.2. Connectivity measures

Adequate characterization of connectivity (high-conductivity flow paths and low-conductivity flow barriers) is crucial when exploring potential locations for well placement. Knudby and Carrera (2004) suggest two requirements for any valid measure of connectivity: i) it should be quantifiable and measurable, ii) it should contain information on the characteristics of flow and transport in the medium.

A measure of connectivity presented by Renard and de Marsily (1997) is the exponent for power averaging of permeabilities. Effective conductivity of a medium, $K_{eff}$ can be obtained by power averaging of the point values of permeability $K$ by

$$K_{eff} = \left[\frac{1}{V}\int_V K(x)^c dV\right]^{\frac{1}{c}} \tag{1}$$

where V is the reservoir volume and $x$ is the location in space. If $K_{eff}$ is known, for example from a pumping test (Meier et al., 1998), then the exponent $c$ can be interpreted as a measure of field connectivity (Knudby and Carrera, 2004). The tractability of this measure makes it an attractive choice, however, it is solely reliant on permeability and neglects other crucial parameters like porosity, fluid compressibility etc. Moreover, it only provides a singular reservoir connectivity value for a complicated connectivity network, whereas pairwise point-to-point connectivities are needed for this study.

Capacitance-Resistance Modeling (CRM) has been implemented by Sayarpour et al. (2007), Nguyen et al. (2011), Cao et al. (2014), and others, as an approach to quantify interwell connectivity in waterfloods and $CO_2$ floods. This method is based on the continuity equation and