

# Optimizing Well Placements for Reservoirs undergoing Water Flooding through an A.I Based Proxy Reservoir Simulator

Sikandar Batra<sup>1</sup>, Alejandro Rodríguez-Martínez<sup>2</sup>, Ignacio E. Grossmann<sup>1</sup>

<sup>1</sup>Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh, PA, USA

<sup>2</sup>Integrated Asset Modelling, Total S.A., Pau, France

\*Corresponding author: Ignacio E. Grossmann; [grossmann@cmu.edu](mailto:grossmann@cmu.edu)

## Abstract

The absolute and relative positions of injector and producer wells, during a waterflooding operation, can have a profound impact on the oil production rate of the reservoir, the recovery factor and the feasible economic life of the project. The conventional approach of finding optimal well positions is highly laborious, computationally expensive, unreliable and requires experienced specialists. In this paper, we delineate multiple novel artificial intelligence based proxy reservoir simulators (PRS), which address these issues. PRS combine first engineering principles related to decline curve analysis and efficient machine learning algorithms to predict production profiles of potential wells placed across a discretized reservoir model. The simulators were applied and tested upon a conventional water drive sandstone twin dome structure reservoir model. The reservoir model was undergoing a water flooding operation through a singular fixed position injector well. The developed simulators were trained upon production profiles that represent less than 10% of potential positions available for producer wells. Subsequently, the PRS were able to predict the production profiles of all potential positions within the time order of milliseconds. Furthermore, the A.I based simulators were found to have an average predictive accuracy of 95%. The robust performance and inexpensive overhead of the PRS enables them to be effective tools in facilitating agile and accurate decision-making for well placement.

**Keywords:** Machine Learning; Optimization; Reservoir Modelling; Well Placement

## 1 Introduction

The volatility of oil prices and stiff competition from supplementary renewables have encouraged Oil and Gas companies to develop modelling tools that aid the increase of field productivity, minimize operational and capital cost, mitigate operational uncertainty, and automate traditional analytical processes. Traditional producer and injector well placement workflows in reservoirs undergoing secondary recovery operations are rarely versatile, and require continuous expert human intervention. This article in subsequent sections presents the approach and fundamental theory associated with two distinct A.I based PRS, which efficiently and with minimal supervision, determine the optimal position of producer wells for conventional reservoir models with a single fixed position water injector well. Moreover, this paper demonstrates the efficacy of both proxy simulators by exhibiting/evaluating the results generated from a case study that applies the PRS on a conventional water drive twin-dome sandstone reservoir model.

The major difference between both distinct proxy simulators is the way they optimize positions for multiple producer wells. The proxy simulator, named ‘Iterative Approach,’ determines optimal producer well locations in a sequential manner, whereas the simulator named ‘Pair Well Approach’ optimizes multiple producer well locations in a simultaneous manner. In both cases the simulators are designed to be highly interpretable, and make decisions based upon fundamental reservoir science principles, associated with decline curve analysis.

## 2 Literature Review

Optimal well placements are highly dependent on geological parameters, field development schedules, and subsurface fluid attributes. Such placement problems are inherently complex, non-linear and multimodal (Wang et al., 2016). To solve such intricate problems, the preferred method is to couple a reservoir numerical simulator with an optimization algorithm (Yeten et al, 2002). The reservoir numerical simulator in the ‘preferred’ method calculates objective functions, such as cumulative oil production or oil recovery, by finite differences (Sepehrnoori and Yu, 2014; Singh and Srinivasan, 2013, 2014). In this way, derivative free optimization algorithms, through methods

like simulated annealing, are used to drive optimization procedures (well placements and targets; drilling schedules etc.). To accelerate the optimization process, three distinct approaches are usually applied. The approaches include the development of streamline-based simulations, proxy models and reduced-order models (Afshari et al, 2011; Caers, 2003; Cruz et al, 2004; Ding et al, 2014; Jansen and Durlofsky, 2016; Gildin et al, 2006).

A streamline-based simulator models reservoirs by implicit saturation and explicit pressure formulations. Furthermore, streamline-based simulators explicitly calculate oil saturation through multiple one-dimensional streamlines (Afshari et al, 2011). Therefore, they require less computing time than conventional numerical simulators, hence accelerating the computation of the objective function. Proxy models are also frequently used to speed-up the computation of the objective function. The framework of a proxy model usually contains artificial neural networks, statistical methods, and Kriging algorithms (Knudsen and Foss, 2015; Onwunalu et al, 2008; AlQahtani et al, 2012). Proxy models expedite computation times by reducing the number of simulations needed to be run on the numerical reservoir simulator. Reduced order models speed up the computation times by reducing the set of reservoir variables through the implementation of the proper orthogonal decomposition based reduction and trajectory piecewise linearization procedures (He and Durlofsky, 2015; Jansen and Durlofsky, 2016; Gildin et al, 2006).

### 3 Approach & Motivation to Develop Proxy Reservoir Simulators (PRS)

Two machine learning based PRS (Iterative and Pair Well Approach) and a conventional discretized reservoir model are presented in this paper. The framework of both proxy simulators is similar. The PRS first extracts production data from a discretized reservoir model that was developed using the Computer Modelling Group (CMG) Black Oil Modelling Software Suite (CMG, 2019). The discretized reservoir model has active producer and injector wells. For every simulation run on the discretized reservoir model, the position of the single producer well (Iterative Approach) or pair producer wells (Pair Well Approach) changes, whereas the position of the injector well remains constant. In order to collect training data for the ML based PRS, we run multiple simulations on the discretized reservoir model.

The proxy simulators use the discretized reservoir model's production data and hyperbolic decline curve correlations to characterize well production profiles. The decline curve correlations are defined by three key parameters: Initial Production Rate ( $Q_i$ ), Decline Rate ( $b$ ), Production Rate at Time of Abandonment ( $Q_{ab}$ ). The PRS are, respectively, trained to predict the three defining parameters of the well production profiles for every possible producer single/pair well placement on the discretized reservoir model. Subsequent to the predictions, the proxy simulators calculate the expected ultimate recovery (EUR-approximation of the quantity of oil potentially recoverable from a well) of every potential single/pair producer well position. The proxy simulators choose optimal well placements on the basis of EUR calculations. For a more detailed/visual overview of the 'approaches', see Figures 13.1 and 13.2.

The dual approach presented in this section addresses the need to compare the Iterative Approach, in which wells are added one at a time in a sequential manner, with the simultaneous optimization of wells placement method (Pair Well Approach). It is obvious that the optimal development plan will always be closer to the output of a simultaneous well placement optimization, rather than results yielded from sequential optimization. This could disqualify the iterative approach as intrinsically suboptimal. Nonetheless, these methods are not in competition, since they correspond to two different development stages.

The Pair Well Approach is more relevant for green field planning in which facilities and piping are to be designed for a plateau optimization. In such case, many wells must be placed at once for financial project evaluations. Also, in such an early stage, the drilling campaigns may allow for parallel drilling operations.

The sequential optimization - placing the wells one by one - adapts better to mature fields in which production is optimized to maintain plateau or extend field life. In mature fields it is unusual to have large drilling campaigns with many wells being drilled at same time. The prospect drilling/development operations are more likely to consist of detailed/comprehensive well placement studies and field results analysis. Therefore the drilling of subsequent new wells becomes an arduous and long process. The advantage of the sequential approach over the simultaneous method in the mature field case is further reinforced by the customary management's dilemma: "Shall we drill a first well at the best possible spot, to then consider where an eventual second well should go, or shall we sacrifice on the positioning of the first well, so that a higher combined production may be obtained once an eventual second well is drilled?". Anyone familiar with the oil and gas industry and its rhythms knows that senior management is unlikely to approve the budget for a second well without even knowing the results of the first one.

Moreover, the use of decline curve analysis (DCA) to model wells response could be seen as a severe weakness when implementing the iterative well placement. Specifically, the problem could be that any interaction with other producers would invalidate prior DCA data, generated with individual well production. Nonetheless, the nature of production optimization tends to place wells away from each other's drainage regions, effectively solving the issue. Indeed, once wells drainage areas start to show significant overlap, it may be suspected that too many wells are being used, or that they are not placed in the right configuration. Also, if a pseudo steady state was reached in the simulation, the DCA of non-interfering wells would remain valid. In order to estimate the effect of the above considerations, a dual approach has been taken: single well placement for the implementation of a sequential algorithm, and a Pair well placement PRS. This allows for comparison of the Iterative and the Simultaneous placement strategies, within computationally affordable costs, while providing a tool that could do the job of simultaneous placement for venturous managers who may yield a sizable operational and drilling budget.

## 4 Problem Statement

Subsequent sections outline a solution to a specific case study. In this case study the two PRS are being applied on a conventional consolidated sandstone twin dome reservoir model undergoing water flooding (we developed the reservoir model by using conventional black oil modelling tools). The objective of the A.I based proxy simulators is to determine the optimal placement of single/pair producer wells perforated exclusively in layer 1 of the reservoir model. The optimal placement is based on the EUR forecast. Only one injector well is responsible for the water flooding operation, and the injector well location is assumed to be fixed over all cases. The position of the injector well is off the axis of the structure, equidistant from the peaks of both twin domes, and the well is exclusively perforated in layers 2 and 3 of the reservoir model. Furthermore, for single producer wells the cut off is set at 55% WCT, and for pair producer wells the **field** cut off is set at 65% WCT.

## 5 Discretized Reservoir Model

The two proxy reservoir simulators were applied to a twin dome consolidated sand stone conventional water drive reservoir model that was developed using conventional black oil reservoir modelling tools. The contour map of the reservoir was modelled with multiple libraries found on python (NumPy, SciPy, etc.). The reservoir model consists of 5 layers (depth range: 9802ft-10398ft), and the water oil contact was set at 10300ft. The size of the grid cells discretizing the reservoir model is 300 ft. by 300 ft. The thickness of the grid cells significantly varies across the reservoir model, with an average value of 25.4 ft. The resolution of the grid is currently limited in order to restrict the size of the optimization problem. The pressure gradient of the reservoir fluid is 0.465 psi/ft. The temperature gradient of the reservoir model is 1.2°F/100ft. The porosity and permeability values of grid cells in the same layer are constant. The porosity value range across the 5 layers is 0.13-0.15. The horizontal permeability range across the 5 layers is 6.27-17.09, with a permeability anisotropy of 10. The horizontal permeability values are derived from the 'Timur Empirical Equations' (Timur A.,1968) and the corresponding porosity values. The compressibility of the rock was modelled using the Newman empirical correlation (Newman G.H., 1973), and an empirical correlation characterizing the relationship between pressure and porosity.

The bubble point pressure ( $P_b$ ) across the reservoir is kept constant at 2000 psi.  $P_b$  is significantly lower than the average reservoir pressure to ensure the reservoir remains undersaturated, and the water injector is primarily responsible for increased oil recovery, and not subsurface pressure maintenance. The PVT model was developed using a multitude of correlations including the Vasquez and Beggs correlation (Vazquez M.E., 1976); Al-Shammasi correlation (Al-Shammasi A.A., 2001). The dynamic rock properties were modelled on the basis of Brook-Corey correlations (Brooks R.H. *et al.*, 1964). The well trajectory data files were developed on Python. All wells simulated in the reservoir model are placed in cell centers, are vertical, and the perforation range of the production wells is limited to the thickness of the top reservoir grid cell. The producer wells have a constant draw down pressure of 300 psi, and the injector well has a constant injection over pressure of 350 psi. Only water is being injected in the reservoir. The density of the water is 8.33 ppg. Figures 5.1 (2D Contour Map) and 4.2 (3D map) are comprehensive geological maps of the reservoir model. It should be noted that the fault showcased in Figure 5.2 has 100% transmissibility. The fault has a throw of 6 ft in the south east direction. We intend to alter and investigate the transmissibility parameter in future works, therefore in regards to this paper, the fault merely provides geometric heterogeneity.

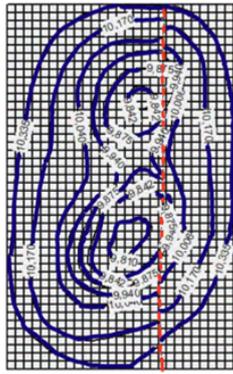


Figure 5.1- 2D Contour Map

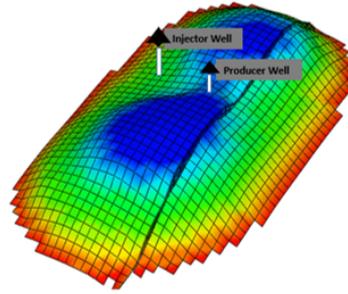


Figure 5.2- 3D Map

## 6 Data Sampling

### 6.1 Iterative Approach

We selected a sample data size of 108 pseudo wells placements. The size of the data set was determined through empirical means. The experimental process strived to minimize computational overhead, while giving a comprehensive spatial representation of the reservoir model. Only one injector well and one producer well are active per run. We developed a sampling algorithm that positions a higher concentration of producer wells placements within the shallowest contour that includes both twin domes. The concentration of wells follows a spatial skewed normal distribution, with just under 65% of pseudo wells positioned within the 'shallowest contour'. This sampling algorithm enabled the data resolution to be highest in the most likely location of the optimal well placement, which in turn enhanced our ability to develop an accurate, robust and stable predictive model.

For all 108 pseudo wells we determined the time step at which the wells exceed their water cut limit (55%). The shortest time step was specified as the abandonment time step for all pseudo wells. Hence, the characterization of the production decline curve for all psuedo wells was limited to the declared time of abandonment. This process bought a common production time period for all wells. Through Pandas (Pandas, 2021), a Python library, we automated the filtering out of grid cells that had a grid thickness value lower than 10 ft for  $Q_i$ ,  $b$  and  $Q_{ab}$  predictive algorithms. All the filtered grid cells are found on the fringes of the reservoir. Filtered grid cells displayed anomalous characteristics and a very low EUR; therefore examination is unnecessary. We initiated the filtration process to reduce variance within the training data set and in turn minimize the propensity of ML algorithms to over fit and loose versatility. Figure 6.1 displays the locations of the 108-pseudo wells and the location of the injector well.

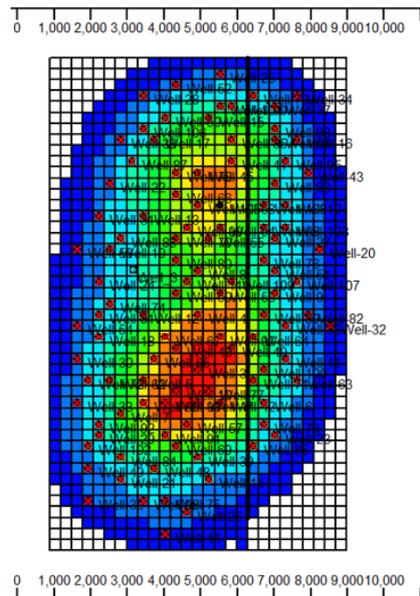


Figure 6.1- Contour Map showing 108 pseudo well positions that were used to develop iterative approach training/testing data set

## 6.2 Pair Well Approach

The sample data size is 110 pairs of pseudo wells. Only one injector well and two production wells are active per run. The same sampling algorithm as the Iterative Approach was used to position a higher concentration of pseudo production wells within the shallowest contour that includes both twin domes. Furthermore, the pair of wells chosen within the tightest contour is randomized, in order to give a fair representation of the spatial area. The production data of the pair wells were aggregated and the combined production decline curves were then analyzed. The approach to standardize the combined decline curves, and the removal of anomalous grid cells was similar to the 'Iterative Approach'. However it should be noted that the field water cut limit for pair wells approach was 65% as simultaneous pair well oil production yielded higher field production rates, which in turn allowed the case study's operator to incur higher water disposal costs.

## 7 Decline Curve Analysis

### 7.1 Iterative Approach

**Table 7.1- Summary of DCA equations used for Iterative Approach**

Iterative Approach Equations		
Equation No.	Type of Equation	Equation
1	Arps Production (Q) Decline Curve Equation as a function of time (days)	$Q(t) = Q_i(1 + bd_it)^{-1/b}$
2	Definition of $d_i$ in terms of $Q_{ab}$ and $t_{ab}$	$d_i = \frac{-\left(\frac{Q_{ab}}{Q_i}\right)^{-b} \left(\left(\frac{Q_{ab}}{Q_i}\right)^b - 1\right)}{b * t_{ab}}$
3	Time of production abandonment for all Pseudo Wells	$t_{ab} = 13 \text{ Years}$
4	Value Constraints for b	$0 < b < 1$
5	Expected Ultimate Recover (EUR) for single Pseudo well	$EUR = \frac{Q_i^b}{d_i(1-b)} (Q_i^{(1-b)} - Q_{ab}^{(1-b)})$

The 'Arps decline curve analysis' (Arps J.J., 1945) method was used to characterize the production decline curves for all pseudo wells. The Arps method was chosen, as it is highly versatile, computationally inexpensive and robust for conventional reservoirs. Table 7.1 provides a summary of all the equations used by the Iterative approach to characterize the decline curve, and to calculate the EUR of every pseudo well. Equation 1 and 2 in Table 7.1 stipulates that the 3 defining parameters for Arps curves are  $Q_i$ ,  $Q_{ab}$  and b. Equation 3 of Table 7.1 asserts that the time of production abandonment for all pseudo wells is 13 years when the iterative approach is applied to the case study outlined in section 4 (process of determining time of abandonment delineated in section 6.1). To implement the Arps production decline curve analysis method on every pseudo well, we used the well's production data and an algorithm named Curve-fit (Scipy V.1.5.4, 2020). Curve-fit used the least square and trust region method to determine the following Arps decline curve coefficients  $Q_i$ , b,  $Q_{ab}$ .

Figure 7.2 is a visual representation of the decline curve analysis process for a single pseudo well. For this particular well, the limit water content of 55% occurs at 13 years. Therefore, the production curve is truncated and characterized until the time of abandonment through curve-fit (Figure 7.2, top right). It should further be noted that the original production profile graph (Figure 7.2, top left) displays an increase in oil production rate post 13 years. The increase in production is caused by a surge in average pressure within layer 1 of the reservoir model. The pressure increase is due to the steady infiltration of injected water, from layers 2 and 3, into layer 1.

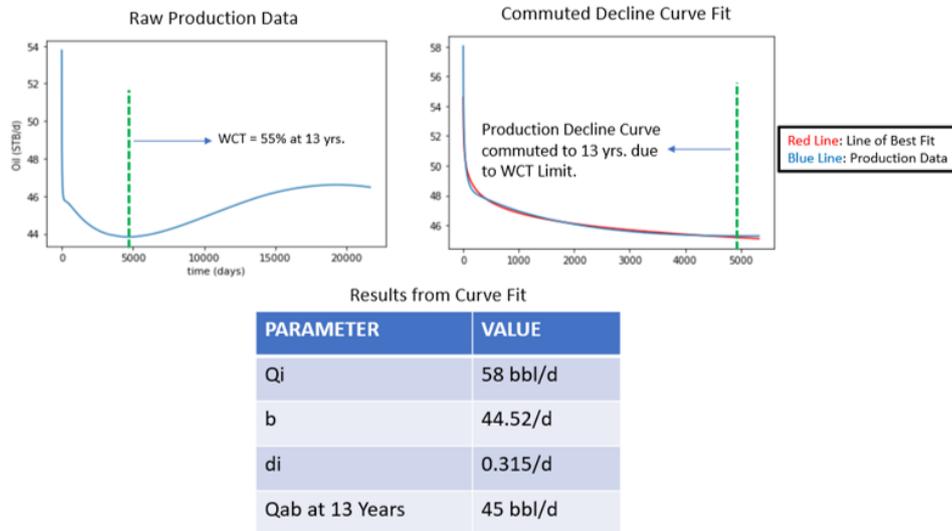


Figure 7.2- Step by Step Process of Decline Curve Analysis

## 7.2 Pair Well Approach

Table 7.3- Summary of DCA equations used for Pair Well Approach

Pair Well Approach Equations		
Equation No.	Type of Equation	Equation
1	Combined Production (Q) rate of both Producer wells as function of time (days)	$Q_{combined}(t) = Q_{p1}(t) + Q_{p2}(t)$
2	Combined Initial Production rate of both Producer wells at time = 0 days	$Q_{ci} = Q_{p1}(t_0) + Q_{p2}(t_0)$
3	Combined Production rate of both Producer wells at time of abandonment	$Q_{cab} = Q_{p1}(t_{ab}) + Q_{p2}(t_{ab})$
4	Combined Arps Decline Curve Equation	$Q_{combined}(t) = Q_{ci}(1 + bd_i t)^{-1/b}$
5	Definition of $d_i$ in terms of $Q_{cab}$ and $t_{ab}$	$d_i = \frac{-\left(\frac{Q_{cab}}{Q_{ci}}\right)^{-b} \left(\left(\frac{Q_{cab}}{Q_{ci}}\right)^b - 1\right)}{b * t_{ab}}$
6	Time of production abandonment for all Pseudo Wells	$t_{ab} = 22 \text{ Years}$
7	Value Constraints for b	$0 < b < 1$
8	EUR for pair Pseudo well	$EUR = \frac{Q_{ci}^b}{d_i(1-b)} (Q_{ci}^{(1-b)} - Q_{cab}^{(1-b)})$

Much like the iterative approach, the Pair Well approach also used the ‘Arps decline curve analysis’ method to characterize the combined production profiles of pair wells. Table 7.3 is a summary of all the equations used by the Pair-Well approach to characterize the combined decline curves and calculate the EUR of every pseudo pair well. Equation 1 to Equation 3 of Table 7.3 showcase that that production data of sub-component wells ( $p1$  &  $p2$ ) in a pair well set were aggregated to develop a combined production profile. This combined production profile was then analyzed through Equation 4 of Table 7.3. Like the iterative approach, the 3 critical defining parameters ( $Q_{ci}$ ,  $Q_{cab}$ ,  $b$ ) of the ‘Combined Arps Decline Curve Equation’ were determined using the Curve-fit algorithm and combined production data. However, unlike the iterative approach, the time of production abandonment was deemed to be 22 years (process of determining time of production abandonment delineated in section 6.1 and 6.2) when the pair well approach was applied to the case study outlined in section 4.

## 8 Machine Learning Based Predictive Models

Section 8 outlines the process we underwent to develop multiple A.I based models that can predict critical defining parameters of production profiles (Arps Decline Curve) for every possible well position on the discretized reservoir model. All the A.I models presented in the subsequent subsections are using non-linear regression based

supervised machine learning algorithms. Therefore, all the A.I models were trained on labeled data sets. The hyperparameters for all the A.I models were optimized through exhaustive grid search algorithms and the efficacy of different regression ML algorithms were compared using statistical significance tests and  $R^2$  based K-fold cross validation analysis. The feature engineering process was based on domain expertise and correlation coefficient analysis

## 8.1 Summary of ML Model Accuracy (Iterative approach)

**Table 8.1- Summary of ML Model Accuracy: Iterative Approach**

Type of ML Model (Outputs)	Algorithm Used	Accuracy of Model ( $R^2$ )- Cross Validation	Percentage Error of Test data (Tr:Te=80:20)
Qi	Random Forrest Regression	0.99 +/- 0.01	1.2%
Qab (at 13 Years)	Random Forrest Regression	0.99 +/- 0.01	2.3%
b	Decision Tree Regression	0.97 +/- 0.03	6.4%

Table 8.1 describes the average accuracy (over 5 cross validation folds) and percentage error with respect to test data for all ML models developed for the iterative approach. The results are very promising as they show all models to be highly stable, generalizable and accurate. We tried to develop the models using the SVM regression algorithm (Scikit-Learn 0.23.2, 2020), but the resultant models were not accurate. It should be noted that any expected ultimate recovery values derived from the models are only valid for the first 13 years of the time period as the production profiles used to train the ML algorithms were truncated in accordance to declared abandonment time step.

## 8.2 Summary of ML Model Accuracy (Pair Well approach)

**Table 8.2- Summary of ML Model Accuracy: Pair Well Approach**

Type of ML Model (Outputs)	Algorithm Used	Accuracy of Model ( $R^2$ )- Cross Validation	Percentage Error of Test data (Tr:Te=80:20)
Qi	SVM Regression	0.99 +/- 0.01	1.8%
Qab (at 22 Years)	Random Forrest Regression	0.96 +/- 0.06	0.8%
b	SVM Regression	0.93 +/- 0.09	6.1%

Table 8.2 describes the average accuracy (over 5 cross validation folds) and percentage error with respect to test data for all ML models developed for the pair well approach. It further stipulates that the abandonment time step for the pair well approach and the efficacy of the corresponding ML models is limited to 22 years. The significant difference in abandonment time steps for both pair well and iterative approaches is due to the disparity in their water cut limits.

## 8.3 Steps taken to develop the ML workflow for both simulators

### 1. Feature (Inputs to Model) Engineering:

- Developed heat maps that showcase correlation strength between input parameters and output parameters. Also plotted interactive visual aides (graphs) to visualize trends.
- Identified anomalous data through hierarchical clustering algorithm.
- Used domain expertise to identify features: Initial Pressures, Grid Cell Thickness, Depth Difference between perforated grid cells of Injector and Producer wells, 2D distance difference between Injector and Producer wells, etc.

### 2. Pre-processed Data Sets and tuned ML Algorithms:

- Standard Scalar; Hyper Parameter Optimization

3. Used different ML Regression Algorithms

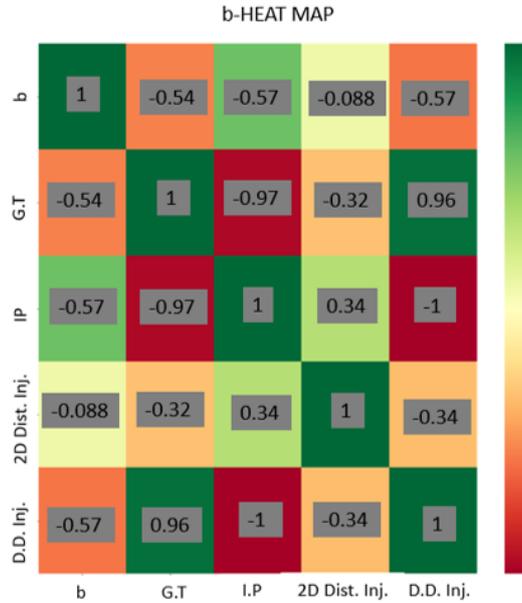
- Random Forrest; Decision Tree; Support Vector Machine (regression)

4. Evaluated and Compared the accuracy of different ML models

- Mean Square Error; Mean Absolute Error; Percentage Error Difference (with respect to testing data)
- Statistical Significance Testing (used to compare performance between two ML regression algorithms)

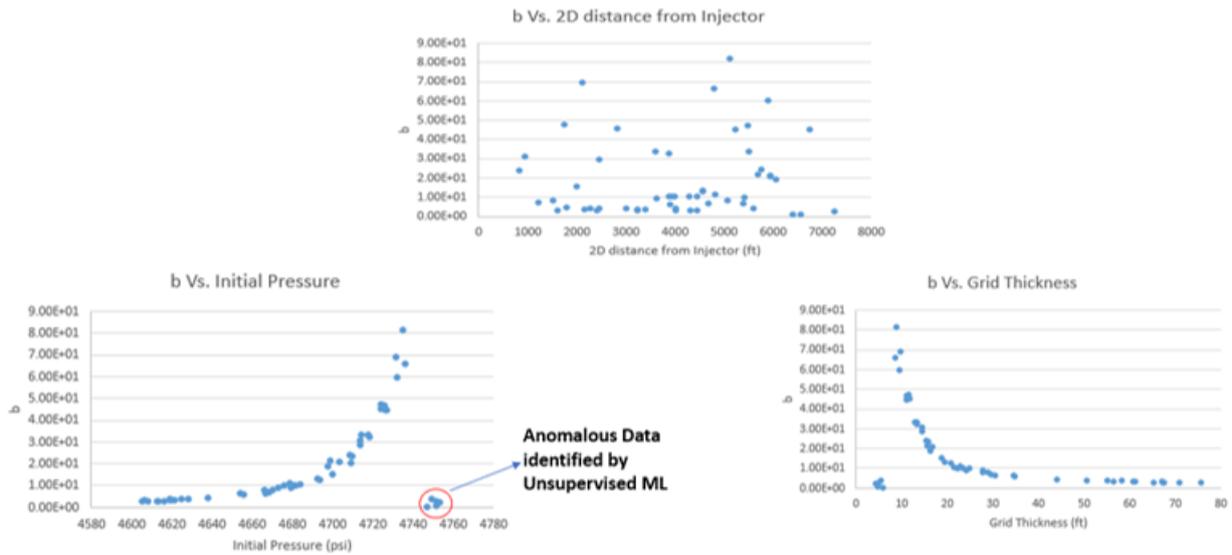
### 8.4 Feature Engineering for Iterative Approach

A Heat map, much like Figure 8.4, is a matrix which provides the correlation strength between two features of a specific ML model. The darker the color, the stronger the correlation. It should further be noted that the number housed in the matrix cells is a scalar that measures the strength of a linear correlation between two features. All numbers lie between -1 and 1. The closer to 0, the weaker the strength of the correlation.



**Figure 8.4- Heat Map showcasing correlation strength between 'b' and input features**

From various heat maps it was inferred that dependant variables  $Q_i$ ,  $Q_{ab}$  and  $b$  are all strongly correlated with the following input features: Initial Pressure, Grid Cell Thickness and Depth Difference between perforated grid cells of Injector and Producer wells. Therefore, these input features are apt for the subsequent development of the machine learning models.

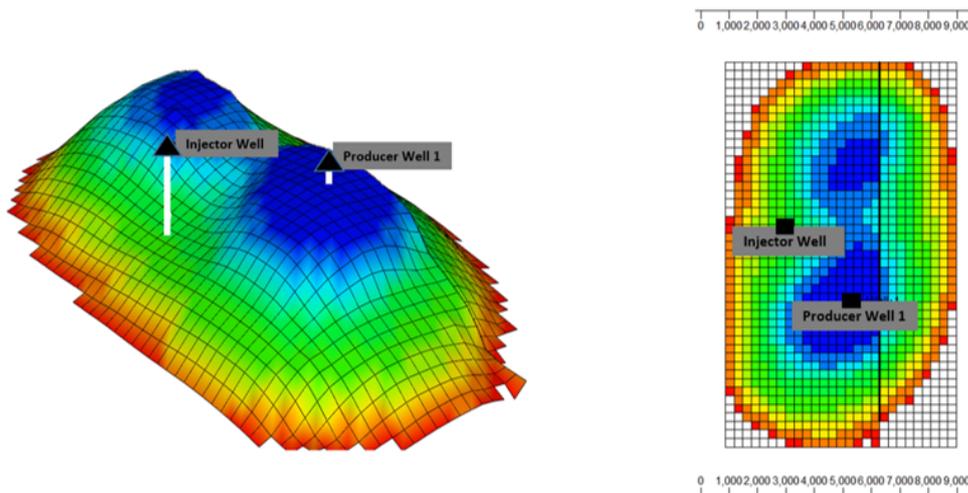


**Figure 8.5 is a sample set of 3 interactive visual graphs that plot  $b$  against Initial Pressure, Injector/Producer Depth Difference and Grid Cell Thickness. These graphs help identify anomalous data and trends.**

From the graphs in Figure 8.5 it can be understood that the relationships between  $b$  and the input features are nonlinear in nature.  $b$  has an inverse relationship with Grid Cell Thickness and Depth difference from Injector. However,  $b$  has a direct relationship with initial pressure. The red circle on the graph highlight anomalous data points. All the anomalous data points correspond to pseudo wells that are located on the fringe of the reservoir model, and have a lower than 10ft grid cell thickness. It should be noted that the anomalous grid cells/pseudo wells showed insignificant production potential, and were therefore disregarded when the training the A.I based PRS. Hence any predictions/forecasts made by the PRS for grid cells that have thickness lower than 10ft are inaccurate but also inconsequential towards the optimal solution.

## 9 Determining Optimal Well Location (Iterative Approach)

The ML models predict the  $Q_i$ ,  $Q_{ab}$  and  $b$  values for every active grid cell in the discretized reservoir model. Utilizing Equation 2 and 5 from Table 7.1 we calculated the  $d_i$  and EUR values for every active grid cell in the reservoir model. Subsequently, the calculated EUR data set was ranked in descending order. The highest EUR and its corresponding grid cell was found to be the optimal well location for the producer well. Figure 9.1 displays the optimal placement for the producer well on the reservoir model map. Well 1 is the producer well.



**Figure 9.1- 2D and 3D grid maps of discretized reservoir model showcasing optimal producer well placement outputted by Iterative Approach**

The optimal well location seems to be sensible and intuitively placed at the top of the largest structure. For further evaluation of the iterative approach, unsupervised machine learning clustering models were applied. The

clustering models classified production zones based on average EUR values in Figure 10.1. As expected, the highest production zones were found to be situated on top of structures.

## 10 Characterizing different reservoir production zones in the reservoir using Unsupervised ML

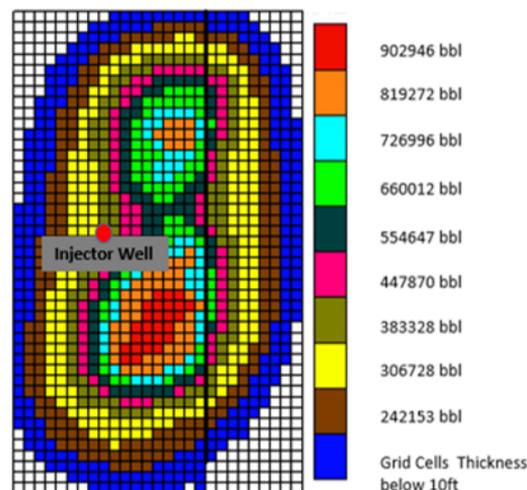
The following subsections utilize clustering unsupervised machine learning algorithms to spatially demarcate/classify production regions of the discretized reservoir model. The clustering algorithm is trained upon the ‘iterative approach’ EUR forecast data set. However, it should be noted that EUR forecast data set was unlabeled. Hyperparameters such as number of clusters were determined through domain expertise and the elbow correlation. The training time order of the clustering algorithms was in the time order of milliseconds. The K-means clustering algorithm was chosen by us over hierarchical clustering algorithms as we wanted the clustering models to scale up for large data sets accurately and efficiently.

### 10.1 Approach

We started off by using the data set that includes the calculated EUR values for each active grid cell in the reservoir model. We then clustered the EUR data set using a K-means cluster algorithm (Na, Xumin et al.). Each cluster developed by the unsupervised ML algorithm demarcates different production zones on the reservoir contour map as shown in Figure 10.1. The label of each cluster is the average EUR for each production zone. Optimal number of clusters is determined by domain knowledge (such as EUR is strongly correlated with grid cell thickness), and using the Elbow method (Marutho et al.). The Elbow method correlates percentage of variance explained (ratio of between-group variance to within-group variance) with the number of clusters. The plot of the Elbow correlation helps pinpointing a Goldilocks zone that minimizes number of clusters but maximizes percentage of explained variance. Once the results of the clustering ML algorithm were obtained, the production zones and corresponding labels were qualitatively evaluated for feasibility.

### 10.2 Results

Figure 10.2 illustrates the different production zones on the reservoir model map. Every color is a different production region. The labels (average EUR forecast) of the colors are displayed on the legend.



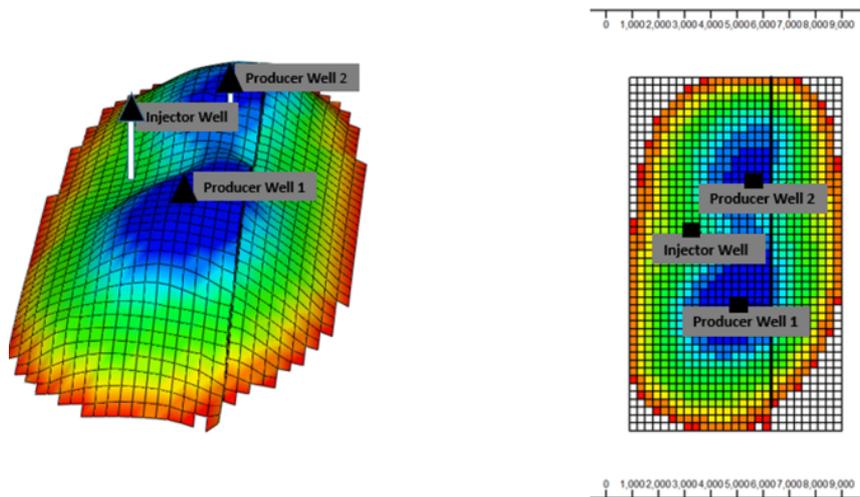
**Figure 10.2- Demarcating Production zones on reservoir grid map based on average EUR through Unsupervised ML**

As expected, Figure 10.2 shows that the highest production regions are on top of the domes. The highest production region (red color) coincides with the optimal producer well placement found by the Iterative proxy model. As one goes further away from the dome, the productivity of the regions decreases. The highest production region (red color) has an average EUR of 902,946 bbl over 13 years (assuming only one well is producing from the reservoir).

## 11 Determining Optimal Well Location (Pair Well Approach)

The process to determine the optimal pair well locations is similar to the one described for the iterative approach. One, major difference is that the cumulative EUR is not calculated for all possible pair well combinations, but rather it is only calculated for pair well combinations where both wells are situated in high production zones (classified by the K-means clustering algorithm developed in the previous section). Additional differences include, the size of the training/testing dataset of the ‘pair well approach’, being 110 pair wells instead of 108 single producer wells. While the ML based models for the pair well approach seem to display relatively high levels of average  $r^2$  accuracy (refer section 8.2) over 5 cross validation K-folds of their training/testing data, it should be noted that the pair well approach ML models were significantly less effective when tested upon pair well combinations in the bottom 4 production zones (in terms of average EUR) demarcated in Figure 10.2. The lack of accuracy of the pair well approach ML models in low production zones is due to its relatively small and spatially underrepresented training data sample. Hence, implying that without a significantly larger training dataset, which is sourced from expensive computational simulations, the pair well approach is much less versatile and accurate than the iterative approach for heterogeneous reservoir environments.

Figure 11.1 displays the optimal placements for pair producer wells on the reservoir model map. The optimal well placements were determined on the basis of which investigated pair well combination outputted the highest combined forecast EUR. The optimal well locations seem to be sensible and intuitively correct.



**Figure 11.1- 2D and 3D grid maps of discretized reservoir model showcasing optimal pair producer well placements outputted by Pair Well Approach**

## 12 Iterative vs. Pair Well Approach Results: Two Producer Well System

In this section we display the optimal well placements provided by the Iterative approach for a two producer and one injector well system. That is, after optimally placing the first producer well, iterations take place for the exact process, while keeping the producer well in the simulation that provides the production data for decline curve fitting used to place the second optimal well. This iterative process is clearly sub-optimal in comparison to simultaneous pair well placement. But we believe for mature or lean field development scenarios where no parallel drilling occurs, the iterative approach is a more accurate and efficient process than the pair well approach. The assumptions, constraints and reservoir environment of this specific application were consistent with the case study outlined in section 4. From Figure 12.1, it is evident that the optimal producer well placements provided by the Iterative approach differ slightly from the placements outputted by the Pair Well Approach (Figure 11.1). Moreover, as expected, it is evident in Table 12.2 that the combined EUR of the Pair Well Approach placements (assuming abandonment time step of 13 years) is 2.4 % or 41,113 bbl higher than the combined EUR of the Iterative Approach placements. Hence, implying that the Pair Well Approach is, from a mathematical perspective, more effective when trying to optimally position two producer wells simultaneously. That being said, over a 13 year production time period, the difference between both approaches falls well below the uncertainties and numerical errors of our limited case study.

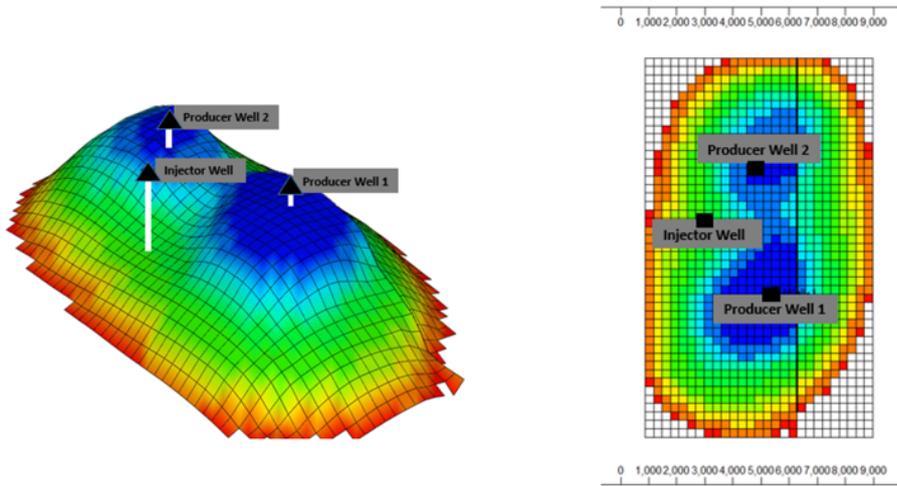


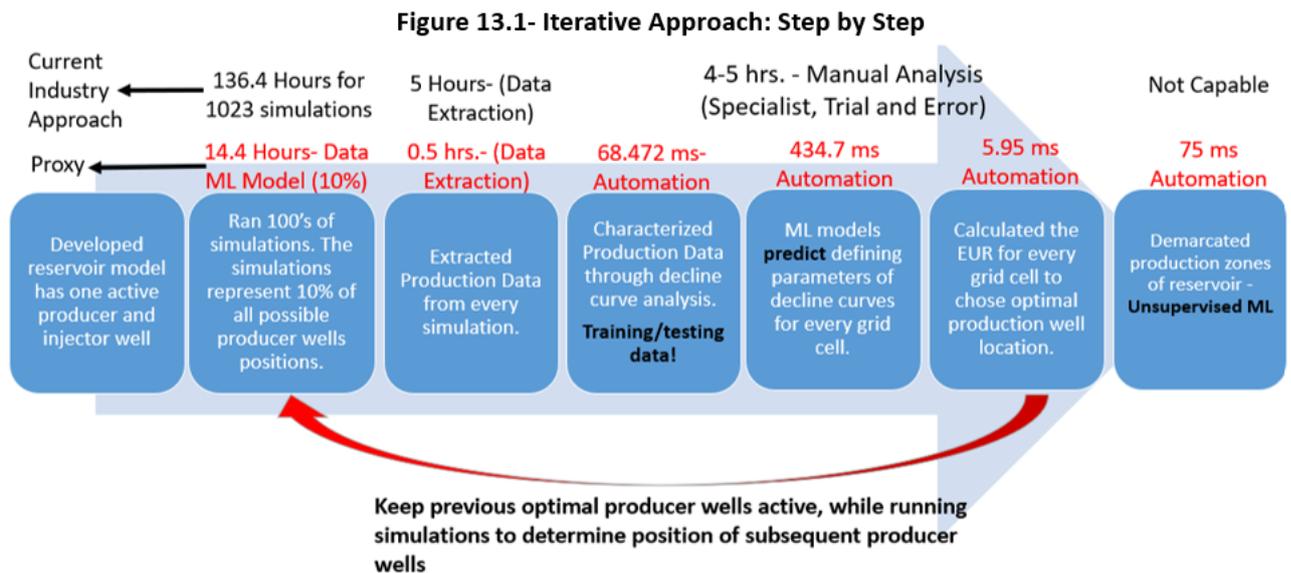
Figure 12.1- 2D and 3D grid maps of discretized reservoir model showcasing optimal pair producer well placements outputted by Iterative Approach

Table 12.2- Comparing results of both approaches for two producer well system

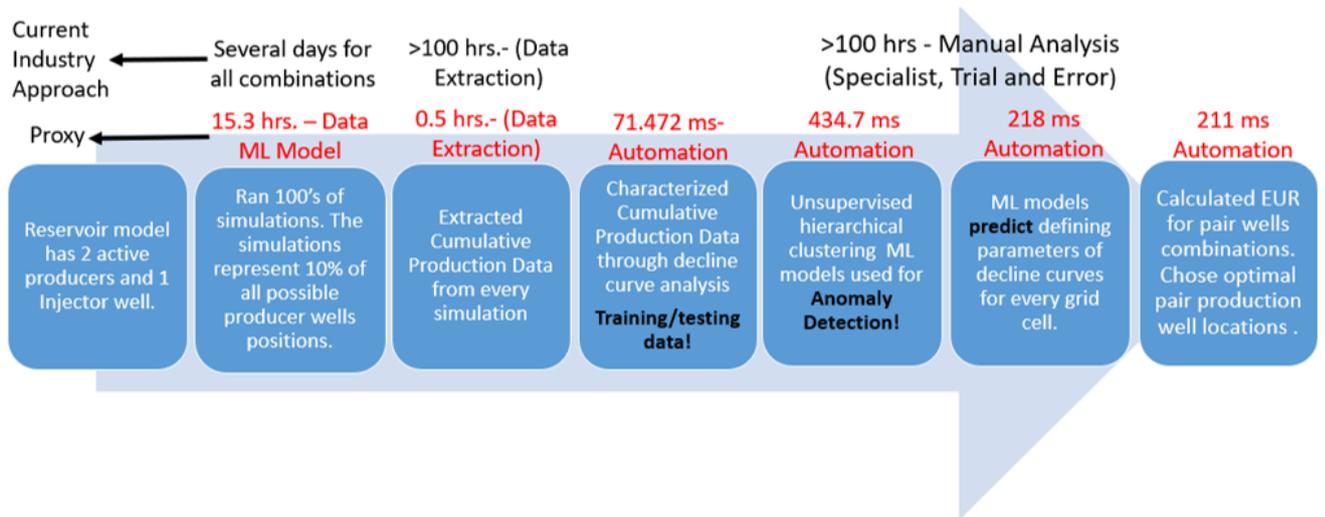
	Combined EUR for both Wells	Computational Time For Two Wells
Iterative Approach EUR- 13 Yrs.	1,726,122 bbl	29.8 hrs.
Pair Well Approach EUR- 13 Yrs.	1,767,235 bbl	15.8 hrs.

### 13 Proxy Reservoir Simulators Computational Cost

Figures 13.1 (Iterative Approach) and 13.2 (Pair Well Approach) display a detailed outline of both reservoir proxy simulator pipelines. Furthermore, the timestamps displayed in each figure compare the computation time of the proxy simulators against estimates for current conventional industry optimization approaches when applied to the problem outlined in section 4. The scale of said problem is small as the proxy reservoir simulators (PRS) search area is only 1150 active grid cells. However it should be noted that the run times for both proxy simulators were in the order of milliseconds and the models were implemented on a standard specification laptop. Moreover, the PRS requires minimal human intervention in comparison to conventional approaches.



**Figure 13.2- Pair Well Approach: Step by Step**



## 14 Novelty of Integrated Proxy Simulators

The PRS do not require any time series based input data to predict optimal well placement. Therefore, the proxy simulators are computationally inexpensive (in terms of data storage) relative to metaheuristic models. The PRS are not subject to the uncertainties and assumptions of analytical/fundamental flow transport models equations as the simulators are based on production data and decline curve analysis. Therefore, we hypothesize the codified process to train the proxy simulators is applicable to a diverse range of reservoirs. Furthermore, the PRS illustrated high levels of stability and accuracy when tested on our case study reservoir model, thus providing reliable proof of concept.

## 15 Conclusions

The proposed reservoir proxy simulators have proven to be computationally inexpensive and consistently accurate, in a limited scope, when optimizing producer well locations. It should further be noted that due to the autonomous nature of the machine learning and decline curve analytics workflow, both PRS require minimal supervision. Identifying the limitations of the proposed PRS in terms of versatility and scalability, still requires further investigation. In the near future we look forward to applying the proposed PRS to industry scale reservoir models that display heterogeneous geological features and properties. Moreover, we aim to compare the accuracy and efficacy of both PRS models when optimizing multiple producer well placements. We have also identified various promising potential extensions to the current workflow of the project:

- Develop a composite ML and GAMS based prescriptive model to schedule installations of artificial lift systems
- Develop an ML based classification model that determines optimal completion strategies
- Forecasting water production for drilled wells and related disposal costs
- Develop a prescriptive model for surface infrastructure design

## 16 References

1. Afshari, Saied, Babak Aminshahidy, and Mahmoud Reza Pishvaie. "Application of an Improved Harmony Search Algorithm in Well Placement Optimization Using Streamline Simulation." *Journal of Petroleum Science and Engineering* 78, no. 3-4 (2011): 664-78. doi:10.1016/j.petrol.2011.08.009
2. Al-Shammasi, A.a. "A Review of Bubblepoint Pressure and Oil Formation Volume Factor Correlations." *SPE Reservoir Evaluation & Engineering* 4, no. 02 (2001): 146-60. doi:10.2118/71302-pa.
3. Alqahtani, Ghazi, Ravi Vadapalli, Shameem Siddiqui, and Srimoyee Bhattacharya. "Well Optimization Strategies in Conventional Reservoirs." Paper Presented at the SPE Saudi Arabia Section Technical Symposium and Exhibition, Al-Khobar, Saudi Arabia, April 2012. doi:10.2118/160861-ms.
4. Arps, J.j. "Analysis of Decline Curves." *Transactions of the AIME* 160, no. 01 (1945): 228-47. doi:10.2118/945228-g.
5. Brooks, Royal Harvard, and A. T. Corey. *Hydraulic Properties of Porous Media*. Fort Collins: Colorado State University, 1964
6. Caers, Jef. "Efficient Gradual Deformation Using a Streamline-based Proxy Method." *Journal of Petroleum Science and Engineering* 39, no. 1-2 (2004): 57-83. doi:10.1016/s0920-4105(03)00040-8.
7. Chen, Hongwei, Qihong Feng, Xianmin Zhang, Sen Wang, Wensheng Zhou, and Yanhong Geng. "Well Placement Optimization Using an Analytical Formula-based Objective Function and Cat Swarm Optimization Algorithm." *Journal of Petroleum Science and Engineering* 157 (2017): 1067-083. doi:10.1016/j.petrol.2017.08.024.
8. Cruz, Paulo S. Da, Roland N. Horne, and Clayton V. Deutsch. "The Quality Map: A Tool for Reservoir Uncertainty Quantification and Decision Making." *SPE Reservoir Evaluation & Engineering* 7, no. 01 (2004): 6-14. doi:10.2118/87642-pa.
9. "Denney, Dennis. "Multivariate Optimization of Field-Development Scheduling and Well-Placement Design." *Journal of Petroleum Technology* 50, no. 12 (1998): 83-85. doi:10.2118/1298-0083-jpt.
10. Ding, Shuaiwei, Hanqiao Jiang, Junjian Li, and Guoping Tang. "Optimization of Well Placement by Combination of a Modified Particle Swarm Optimization Algorithm and Quality Map Method." *Computational Geosciences* 18, no. 5 (2014): 747-62. doi:10.1007/s10596-014-9422-2. 10.1115/1.4032547. Accessed 8 Sept. 2020.
11. "Formation Compressibility." *Formation Compressibility - an Overview* — ScienceDirect Topics. Accessed June 20, 2021. <https://www.sciencedirect.com/topics/engineering/formation-compressibility>
12. Gildin, Eduardo, Hector Klie, Adolfo Antonio Rodriguez, Mary Fanett Wheeler, and Robert H. Bishop. "Development of Low-Order Controllers for High-Order Reservoir Models and Smart Wells." Paper Presented at the SPE Annual Technical Conference and Exhibition, San Antonio, Texas, USA, September 2006. doi:10.2118/102214-ms.
13. He, J., and L. J. Durlofsky. "Constraint Reduction Procedures for Reduced-order Subsurface Flow Models Based on POD-TPWL." *International Journal for Numerical Methods in Engineering* 103, no. 1 (2015): 1-30. doi:10.1002/nme.4874.
14. "Home." Computer Modelling Group Ltd. Accessed June 20, 2021. <http://www.cmgl.ca/>.
15. Jansen, Jan Dirk, and Louis J. Durlofsky. "Use of Reduced-order Models in Well Control Optimization." *Optimization and Engineering* 18, no. 1 (2016): 105-32. doi:10.1007/s11081-016-9313-6
16. Jia, Hu, and Lihui Deng. "Water Flooding Flowing Area Identification for Oil Reservoirs Based on the Method of Streamline Clustering Artificial Intelligence." *Petroleum Exploration and Development* 45, no. 2 (2018): 328-35. doi:10.1016/s1876-3804(18)30036-3.
17. Knudsen, Brage Rugstad, and Bjarne Foss. "Designing Shale-well Proxy Models for Field Development and Production Optimization Problems." *Journal of Natural Gas Science and Engineering* 27 (2015): 504-14. doi:10.1016/j.jngse.2015.08.005.
18. Marutho, Dhendra, Sunarna Hendra Handaka, Ekaprana Wijaya, and Muljono. "The Determination of Cluster Number at K-Mean Using Elbow Method and Purity Evaluation on Headline News." 2018 International Seminar on Application for Technology of Information and Communication, 2018. doi:10.1109/isemantic.2018.8549751.
19. Na, Shi, Liu Xumin, and Guan Yong. "Research on K-means Clustering Algorithm: An Improved K-means Clustering Algorithm." 2010 Third International Symposium on Intelligent Information Technology and Security Informatics, 2010. doi:10.1109/iitsi.2010.74.
20. Newman, G.h. "Pore-Volume Compressibility of Consolidated, Friable, and Unconsolidated Reservoir Rocks Under Hydrostatic Loading." *Journal of Petroleum Technology* 25, no. 02 (1973): 129-34. doi:10.2118/3835-pa.

21. Nwankwor, E., A. K. Nagar, and D. C. Reid. "Hybrid Differential Evolution and Particle Swarm Optimization for Optimal Well Placement." *Computational Geosciences* 17, no. 2 (2012): 249-68. doi:10.1007/s10596-012-9328-9.
22. Onwunalu, Jerome Emeka, Michael Lev Litvak, Louis J. Durlofsky, and Khalid Aziz. "Application of Statistical Proxies to Speed Up Field Development Optimization Procedures." Paper Presented at the Abu Dhabi International Petroleum Exhibition and Conference, Abu Dhabi, UAE, November 2008. doi:10.2118/117323-ms.
23. "Pandas." Pandas. Accessed June 20, 2021. <https://pandas.pydata.org/>.
24. PetroWiki. "Relative Permeability Models." PetroWiki. January 19, 2016. Accessed June 20, 2021. <https://petrowiki.spe.org/Relativepermeabilitymodels>.
25. PetroWiki. "Calculating PVT Properties." PetroWiki. June 20, 2020. Accessed June 20, 2021. <https://petrowiki.spe.org/CalculatingPVTproperties>.
26. "Scipy.optimize.curvefit¶." Scipy.optimize.curvefit - SciPy V1.6.3 Reference Guide. Accessed June 20, 2021. <https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.curvefit.html>.
27. Siavashi, Majid, Mohammad Rasoul Tehrani, and Ali Nakhaee. "Efficient Particle Swarm Optimization of Well Placement to Enhance Oil Recovery Using a Novel Streamline-Based Objective Function." *Journal of Energy Resources Technology* 138, no. 5 (2016). doi:10.1115/1.4032547.
28. Siavashi, Majid, and Mohsen Yazdani. "A Comparative Study of Genetic and Particle Swarm Optimization Algorithms and Their Hybrid Method in Water Flooding Optimization." *Journal of Energy Resources Technology* 140, no. 10 (2018). doi:10.1115/1.4040059.
29. Singh, Harpreet, and Sanjay Srinivasan. "Uncertainty Analysis by Model Selection Technique and Its Application in Economic Valuation of a Large Field." Paper Presented at the North Africa Technical Conference and Exhibition, Cairo, Egypt, April 2013. doi:10.2118/164623-ms.
30. Singh, Harpreet, and Sanjay Srinivasan. "Scale up of Reactive Processes in Heterogeneous Media - Numerical Experiments and Semi-Analytical Modeling." Paper Presented at the SPE Improved Oil Recovery Symposium, Tulsa, Oklahoma, USA, April 2014. doi:10.2118/169133-ms.
31. "Sklearn.svm.SVR¶." Scikit. Accessed June 20, 2021. <https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVR.html>.
32. Timur, A. "An Investigation Of Permeability, Porosity, Residual Water Saturation Relationships For Sandstone Reservoirs." *The Log Analyst* 9 (1968).
33. Vasquez, M. "Correlations for Fluid Physical Property Prediction." MS Thesis, University of Tulsa, Tulsa, Oklahoma, 1976. doi:10.2118/6719-pa.
34. Wang, Xiang, Ronald D. Haynes, and Qihong Feng. "A Multilevel Coordinate Search Algorithm for Well Placement, Control and Joint Optimization." *Computers & Chemical Engineering* 95 (2016): 75-96. doi:10.1016/j.compchemeng.2016.09.006
35. Xu, Suxin, Fanhua Zeng, Xuejun Chang, and Hong Liu. "A Systematic Integrated Approach for Waterflooding Optimization." *Journal of Petroleum Science and Engineering* 112 (2013): 129-38. doi:10.1016/j.petrol.2013.10.019.
36. Yeten, Burak, Louis J. Durlofsky, and Khalid Aziz. "Optimization of Nonconventional Well Type, Location and Trajectory." *All Days*, 2002. doi:10.2118/77565-ms.
37. Yu, Wei, and Kamy Sepehrnoori. "Optimization of Well Spacing for Bakken Tight Oil Reservoirs." *Proceedings of the 2nd Unconventional Resources Technology Conference*, 2014. doi:10.15530/urtec-2014-1922108.