# Abstract

Type wells are growing in acceptance in the industry as a means of forecasting production in low-permeability reservoirs. Although the conceptual understanding is becoming clearer, there are still many challenges when we try, as a means of forecasting production in low-permeability reservoirs, to apply type wells in practice, chief among them being how to build type wells that accurately reflect original well production data. At the same time, machine learning techniques have become more and more popular all over the world. There are also more and more researchers in petroleum engineering industry incorporating machine learning techniques into their own researches. In this thesis, we will mainly combine type wells and machine learning techniques together to implement the production profile prediction.

In this research, we constructed type wells based on well estimate ultimate recovery (EUR). The results of type wells were evaluated. For the sake of effective application of these type wells, we adopted 3 advanced machine learning techniques, neural networks (NNet), support vector machine (SVM) and random forest (RF) to classify type wells. 200 predicted well production data from Barnett Shale is used as input to the 3 algorithms. By using a single well production data with only short production history from the same geologic area, we can predict the EUR range of the well.

The results show that the 200 sorted EUR values predicted through the usage of ValNav follows a lognormal distribution as indicated in the log-probability paper. The P90, P50 and P10 type wells were picked up and the low P10/P90 value of 2.3 shows a low variance of EUR in this geologic area, which further implies that the type wells are representative.

The production data were processed before fed into the machine learning algorithms. The 4 fold cross-validation technique was also employed to reduce the generalization error of the trained classifiers. The details of these 3 algorithms were also introduced. NNet performed best with higher average test accuracy among the three machine learning algorithms employed.

Other properties of the three classifiers specifically for this project are also explored. In NNet, the number of hidden neurons at 163 is found to give the best performance. In SVM, the C value, the penalty parameter of the error term is found to affect the soft margin SVM performance. For different node splitting criterions in RF algorithms, the performance of adopting entropy is superior than using Gini Index for our problem.

The results in this project can be used to help oil & gas companies make financial decisions based on available production data in the same geologic area. Also, this project can also help provide research basis for researchers who are interested in this direction.

# Introduction

Unconventional resources have become of dominant in the oil and gas world in recent years. As a carry-over from the era of conventional resources, decline curve analysis (DCA), which is dominated by the Arps decline model (Arps 1945), is the main method used to predict production of these resources.

Unconventional resources, including shale gas, shale oil, and coalbed methane, are gaining increasing attention from many researchers in the petroleum industry. As defined by Holditch (Holditch 2003), an unconventional reservoir is one “that cannot be produced at economic flow rates or that does not produce economic volumes of oil and gas without assistance from mass stimulation treatments or special recovery processes and technologies”. This definition manifests the difficulties of unconventional resources extraction. Unlike extraction of conventional resources, the lengthy transient flow production period in unconventional resources limits the applicability of the Arps decline model, which was designed and validated in the conventional resources era.

Figure 1 below used viscosity and permeability to distinguish conventional and unconventional resources. Besides the characteristics of the unconventional resources themselves, unconventional reservoirs themselves are relatively more heterogeneous. There are more geologic uncertainties as well (e.g., different geologic layering of reservoirs).

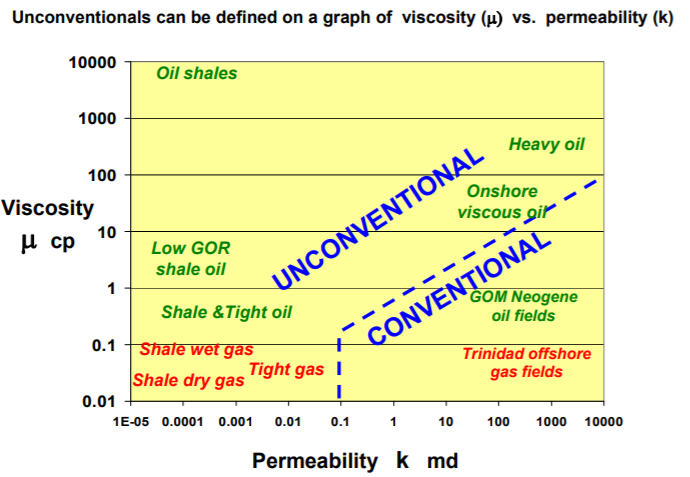


Figure 1 Conventional and Unconventional Resources (BP 2012)

These conditions have caused unexpected errors when applying conventional reservoir production forecasting methods to unconventional reservoirs. Based on Arps’ model (Arps 1945), some researchers have proposed alternative decline models to generate more accurate production data predictions. These modified models include Duong’s method, and stretched exponential and power law models, which can work well under certain constrained conditions for unconventional resources. In addition, the Arps model has been modified in an effort to provide more accurate forecasts. Unfortunately, all these approaches can result in ultimate error due to the many uncertainties in unconventional resources.

To tackle this issue, industry has begun to employ type wells, which should be distinguished from “type curves” in which a set of dimensionless flow rate versus dimensionless time curves are generated to predict future production.

## 2.1 Type Wells

The core idea of constructing type wells is to construct a well representing a set of wells that are being recognized as “analogous.” It aims to extract the inherent characteristics of multiple wells in a certain geologic area by producing one or a family of type wells which should be representative enough to represent all wells in an area of interest. The approach commonly employed in industry to construct type wells is to arithmetically average the production histories of a set of producing wells in the field. As pointed out by Freeborn (Freeborn et. al. 2012), there is a serious flaw in this approach. He maintains that “the type well calculation must include the SI well count as though those SI wells continue to produce at a rate of zero”. Figure 2 (Freeborn et. al. 2012), provides a graphical explanation of how to avoid this flaw.

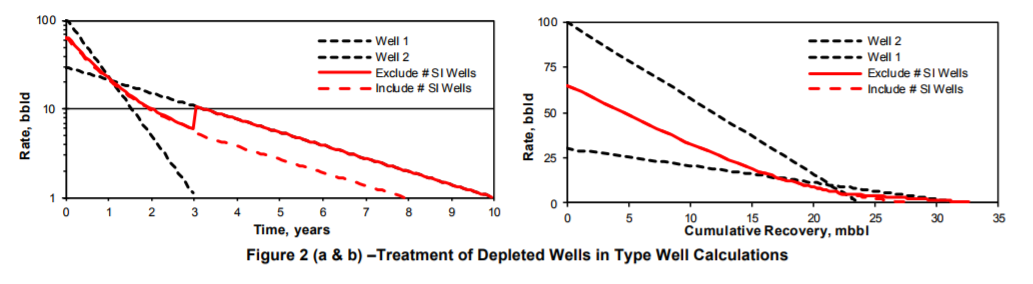


Figure 2 Treatment of Depleted Wells in Type Well Construction (Freeborn et. al. 2012)

In his paper, Freeborn also addressed other issues that may be encountered when constructing type wells. Those issues include the date of first production in which he proposes to refer to the shifted data as having been normalized, multi well testing and type well using history data. To test the theory, Freeborn (Freeborn et. al. 2012) gave 4 different field examples. Rastogi (Rastogi and Lee 2015) extended the methodology of Freeborn in constructing type wells, and they also expressed their considerations on the method of normalizing production data and removing outlier data. Rastogi suggested applying a type well production profile generated in one area to another area in the same resource with different geologic characteristics, to determine the universality of type well method.

In terms of how to construct type well accurately and efficiently, Freeborn (Freeborn and Russell 2015) used aggregation methods, which he proposed as a way to construct type wells that “accurately simulate the average rate-time production profiles expected from a drilling program of specific size and certainty.” They also improved the normalization method by building several scaling rules that scale a rate-time profile. Apart from that, Chaudhary (Chaudhary and Lee 2017) chose to generate type wells by cluster-weighted modeling, in which one type well is constructed as one “global weighted probabilistic superposition of local models” where local models consist of several wells combined using simple cubic spline techniques. As researchers are increasingly focusing on type well methods to evaluate the future production of unconventional plays, engineers have reached a consensus that the current challenge of type well construction is concentrated in how to accurately create type wells that are representative of wells in a certain geologic area.

## 2.2 Machine Learning Application in Petroleum Industry

In March of 2016, AlphaGo (Wikipedia 2017) from DeepMind won the Go championship over Lee Sedol. From then on, artificial intelligence has made unprecedented progress. The AI industry, particularly the machine learning technique, has attracted more and more people to this area. Today, progress in the AI industry and machine learning is based mainly on deep learning, which is largely supported by the convolutional neural network work inspired by Fukushima’s neocognition proposed in 1980. Yann LeCun then proposed convolutional neural networks. Convolutional neural networks are now widely accepted and popularly employed in image processing. Figure 3 and figure 4 graphically illustrate neocognition and convolutional neural networks, respectively.

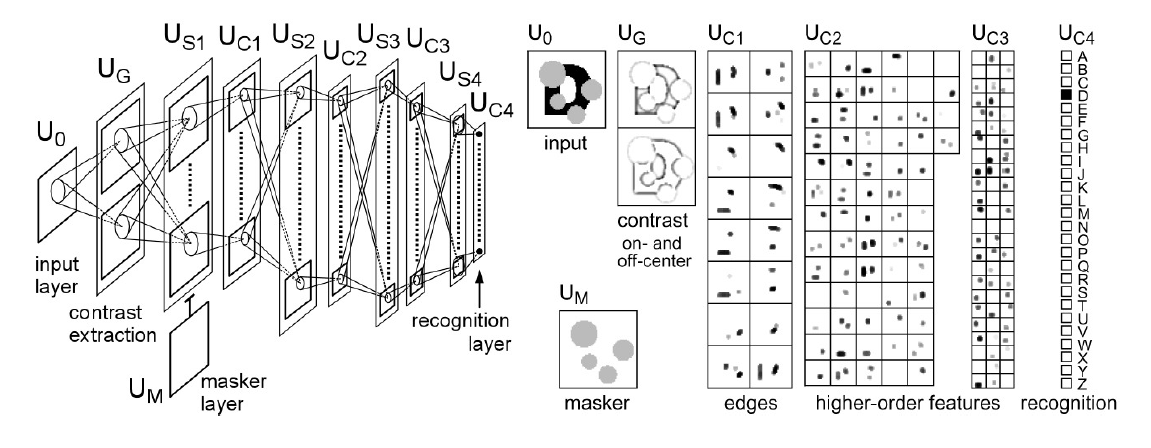


Figure 3 Fukushima's Neocognitron

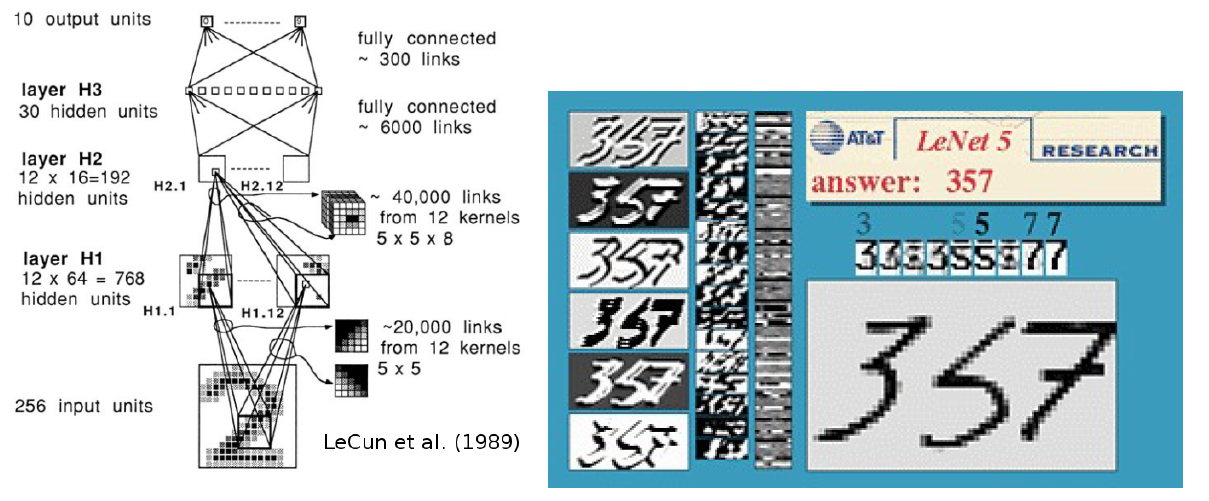


Figure 4 LeCun's Convolutional Neural Nets

Recent researchers have applied AI, including machine learning techniques, to the oil and gas industry. Ali (Ali 1994) reviewed the application of neural networks, a major machine learning technique, in the petroleum industry at that time. He pointed out that the capabilities of neural networks include “pattern recognition, classification of noisy data, nonlinear feature detection, market forecasting and process modeling,” which makes them well suited to solve problems in petroleum industry. In particular, we know that neural networks are quite useful to deal with nonlinear relationships between various variables.

Mohaghegh (Mohaghegh 2000) reviewed general applications of virtual artificial intelligence and artificial neural networks in the oil and gas industry. Ramgulam (Ramgulam et al. 2007) developed a trained artificial neural network to specifically address the history matching problem. Liu (Liu and Horne 2011) adopted data-mining techniques to analyze permanent downhole gauges. Liu (Liu and Horne 2011) exploited the least-mean-squares method and used the stochastic gradient descent method to train the parameters in a given polynomial equation to evaluate the relationship between pressure and flow rate. In his paper (Liu and Horne 2013b), Liu continued to approach the problems in interpreting pressure and flow rate data from permanent downhole equipment using data mining and machine learning methods, respectively. Tian (Tian and Horne 2015) extended Liu’s work further. He applied the “kernel ridge regression based machine learning” to further interpret data from multi-well tests. In addition, he completed flow rate reconstruction.

One of the main objectives of my research is to explore the advantages of machine learning in type well application. In particular, neural networks (NNet), support vector machine (SVM), and random forest (RF) are adopted to aid the application of type well in predicting well production profile.

Previous studies related to this area include the work of Subrahmanya (Subrahmanya et. al. 2014) who gave us a review of machine learning algorithms applied to production data pattern recognition. In this paper, he employed semi-supervised learning and active learning algorithm to analyze acquired data. More recently, Jia (Jia and Zhang 2016) used neural networks to forecast production from the Barnett Shale, and he achieved much more accurate predictions than that with conventional empirical models. Figure 5 illustrates the simple neural network structure that Jia employed (Jia and Zhang 2016).

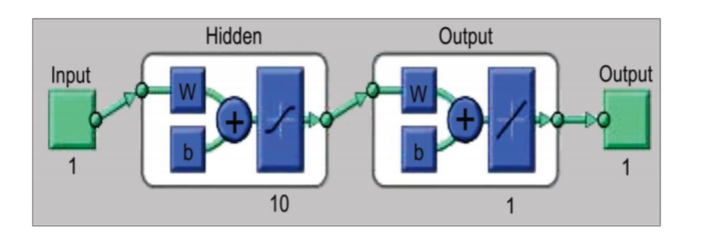


Figure 5 Simple Neural Network Structure (Xinli Jia et. al. 2016)

Aulia (Aulia et al. 2014) states that a subset of the bottom hole pressure (BHP) can be a contributing factor to the oil recovery factor in the field, so Aulia uses Latin Hypercube Monte Carlo (LHMC) and random forest (RF) to identify such subset. Aulia also points out that due to the fact that RF can identify the importance of RF’s independent variables based on a collection of uncertainty runs, LHMC and RF combined can be a global approach to sensitivity analysis. Hedge (Hedge et al. 2015) shows the possibilities of using statistical learning methods like trees, bagging, and random forest to predict rate of penetration during real-time operations. Hedges also anticipates that random forest and bagging techniques can be employed to determine the relative importance of input parameters, which will further provide sound information for drilling parameter predictions and optimizing the on-the-fly rig floor changes.

Zhao (Zhao et al. 2006) trained an -intensive support vector machine (SVM) to implement the regression of water saturation from seismic data by using a water saturation curve calculated from density and resistivity logs of a gas well at the Gulf of Mexico. This provides a way to estimate the water saturation away from wellbore, which will be applied to distinguish between commercial and low saturation gas. Zhao (Zhao et al. 2014) gave an example of using proximal SVM (PSVM) to classify lithofacies, specifically Zhao used the PSVM to differentiate limestone from shale in a Barnett Shale gas play. Zhao’s result was based on the two applications of PSVM one for waveform classification and the other for the classification of well data. The promising results in both seismic and well log data demonstrated the validity of PSVM classifier in binary classification. In Anifowose’s paper (Anifowose et al. 2012), artificial neural networks (ANN) and are both employed to predict porosity and permeability of oil and gas reservoirs with carbonate platforms. The results show that SVM performs better than ANN. In Anifowose’s implementation, six datasets were used through the stratified sampling rather than the conventional static method of data division. SVM algorithm was anticipated to assist petroleum exploration engineers to estimate various properties with better accuracy, leading to reduced exploration time and increased production.

## 2.3 Problems and Objectives

As unconventional resources have been playing a more and more important role in the international energy market, the effective and accurate methods to predict the well production in unconventional reservoirs have been in a great need. Type wells method, as a method used to generate a well or a family of wells. Those wells are representative of wells in the same geologic area. Finding these type wells is the primary problem we will resolve in this thesis. After we identify the type wells, the evaluation of the type wells also needs to be given.

Although type wells method is gaining attention from researchers in petroleum engineering field, as far as I know there are few reliable applications of type wells in forecasting well production currently. So, we also need to find the effective application of these type wells in this thesis, with which machine learning techniques get involved.

Machine learning techniques are powerful in solving regression and classification problems. For different specific problems, the performance of different machine learning algorithms varies. In this thesis, we need a model to deal with the high nonlinear relationship between variables. Thus, the performance of the machine learning algorithms’ involvement also needs to be evaluated.

To sum up, we mainly have the following 3 objectives to complete in this project:

1. Construct and evaluate type well production profiles

2. Apply type wells to forecast production of wells with only short production histories with the assistance of machine learning techniques

3. Evaluate the performance of different machine learning techniques in the application

## 2.4 Significance to Industry

This is the first trial in petroleum engineering field that using machine learning techniques and type wells method combined to forecast production profile. Machine learning techniques involved in this project aim to reliably classify the wells with only short production history, and type wells are responsible of indicating the estimate ultimate recovery (EUR) of wells with production history of limited length. The results of this project can be beneficial for oil & gas companies to make financial decisions based on the production of available wells.

## 2.5 Outline of thesis

In this thesis, after the state-of-art introduction, we first show the preprocess 200 wells’ production data from the same geologic area – Barnett Shale. The details of forecasting each well’s EUR and the results are given. We use the results to construct type wells (i.e. P10, P50 and P90 type wells). The type wells constructed are evaluated. Then we introduce the data processing techniques we adopted before we implement the machine learning algorithms.

3 machine learning algorithms, neural networks, support vector machine and random forest, are discussed. We implement the 3 algorithms to classify the wells with only short production history to one of 4 types. The Implementation details and algorithms’ performance are evaluated.

# References

Ali, J.K. 1994. Neural Networks: A New Tool for the Petroleum Industry? . Presented at the European Petroleum Computer Conference, Aberdeen, United Kingdom, 15—17 March. SPE-27561-MS. <http://dx.doi.org/10.2118/27561-MS>.

Anifowose, F., Ewenla, A., and Eludiora, S. 2012. Prediction of Oil and Gas Reservoir Properties Using Support Vector Machines. Presented at the International Petroleum Technology Conference, Bangkok, Thailand, 7—9 February. IPTC-14514. <http://dx.doi.org/10.2118/>IPTC-14514.

Arps, J.J. 1945. Analysis of Decline Curves. SPE Journal 160 (1). SPE-45228. <http://dx.doi.org/10.2118/45228>.

Aulia, A., Rahman, A., and Velasco, J.J.Q. 2014. Strategic Well Test Planning Using Random Forest. Presented at the SPE Intelligent Energy Conference and Exhibition, Utrecht, Netherlands, 1—3 April. SPE-167827. <http://dx.doi.org/10.2118/167827>.

BP 2012. AAPG , <http://www.searchanddiscovery.com/documents/2012/80217cander/ndx_cander>.

Chaudhary, N.L. and Lee, W.J. 2017. Constructing Enhanced Type Wells Using Cluster-Weighted Modelling. Presented at the SPE/AAPG/SEG Unconventional Resources Technology Conference, Austin, Texas, 24—26 July. SPE-URTEC-2697554-MS. <http://dx.doi.org/10.2118/URTEC-2697554-MS>.

Freeborn, R. and Russell, B. 2015. Creating More Representative Type Wells. Presented at the SPE/CSUR Unconventional Resources Conference, Calgary, Canada, 20—22 October. SPE-175967-MS. http://dx.doi.org/10.2118/175967-MS.

Freeborn, R., Russell, B., and Keinick, W. 2012. Creating Analogs, Fact and Fiction. Presented at the SPE hydrocarbon, Economics and Evaluation Symposium, Calgary, Canada, 24—25 September. SPE-162630. <http://dx.doi.org/10.2118/162630>.

Hedge, C., Wallace, S., and Gray, K. 2015. Using Trees, Bagging, and Random Forest to Predict Rate of Penetration During Drilling. Presented at the SPE Middle East Intelligent Oil&Gas Conference & Exhibition, Abu Dhabi, United Arab Emirates, 15—16 September. SPE-176792. <http://dx.doi.org/10.2118/176792>.

Holditch, S.A. 2003. The Increasing Role of Unconventional Reservoirs in the Future of the Oil and Gas Business. Journal of Petroleum Technology 55 (11): 34—79.

Jia, X. and Zhang, F. 2016. Applying Data-Driven Method to Production Decline Analysis and Forecasting. Presented at the SPE Annual Technical Conference and Exhibition, Dubai, United Arab Emirates, 26—28 September. SPE-181616-MS. http://dx.doi.org/10.2118/181616-MS.

Liu, Y. and Horne, R.N. 2011. Interpreting Pressure and Flow Rate Data from Permanent Downhole Gauges Using Data Mining Approaches. Presented at the SPE Annual Technical Conference and Exhibition, Denver, Colorado, 30 October—2 November. SPE-147298-MS. <http://dx.doi.org/10.2118/147298-MS>.

Liu, Y. and Horne, R.N. 2013. Interpreting Pressure and Flow Rate Data from Permanent Downhole Gauges with Convolution-Kernel-Based Data Mining. Presented at the SPE Annual Technical Conference and Exhibition, New Orleans, Louisiana, 30 September—2 October. SPE-166440-MS. <http://dx.doi.org/10.2118/166440-MS>.

Mohaghegh, S. 2000. Virtual-Intelligence Applications in Petroleum Engineering: Part 1—Artificial Neural Networks. Journal of Petroleum Technology 52 (9). SPE-58046-JPT. <http://dx.doi.org/10.2118/58046-JPT>.

Ramgulam, A., Ertekin, T., and Flemings, P.B. 2007. An Artificial Neural Network Utility for the Optimization of History Matching Process. Presented at the Latin American&Caribbean Petroleum Engineering Conference, Buenos Aires, Argentina, 15—18 April. SPE-107468-MS. <http://dx.doi.org/10.2118/107468-MS>.

Rastogi, A. and Lee, W.J. 2015. Methodology for Construction of Type Wells for Production Forecasting in Unconventional Reservoirs. Presented at the Unconventional Resources Technology Conference, San Antonio, Texas, 20—22 July. SPE-URTEC-2152273-MS. http://dx.doi.org/10.2118/URTEC-2152273-MS.

Subrahmanya, N., Xu, P., El-Bakry, A. et al. 2014. Advanced Machine Learning Methods for Production Data Pattern Recognition. Presented at the SPE Intelligent Energy Conference&Exhibition, Utrecht, Netherlands, 1—3 April. SPE-167839-MS. <http://dx.doi.org/10.2118/167839-MS>.

Tian, C. and Horne, R.N. 2015. Machine Learning Applied to Multiwell Test Analysis and Flow Rate Reconstruction. Presented at the SPE Annual Technical Conference and Exhibition, Houston, Texas, 28—30 September. SPE-175059-MS. <http://dx.doi.org/10.2118/175059-MS>.

Wikipedia 2017. <https://en.wikipedia.org/wiki/AlphaGo_versus_Lee_Sedol>.

Zhao, B., Zhou H., Li, X., and Han, D., 2006. Water Saturation Estimation Using Support Vector Machine. SEG Annual Meeting, New Orleans, Louisiana. SEG 2006-1693.

Zhao, T., Jayaram, V., Marfurt, K., and Zhou., H. 2014. Lithofacies classification in Barnett Shale Using Proximal Support Vector Machines, Danver, Colorado. SEG 2014-1491