SPE-198756-MS

Dynamic Production Forecasting using Artificial Neural Networks customized to historical well Key Flow Indicators

David N. Nnamdi, Niger Delta Petroleum Resources Ltd., Victor O. Adelaja, Geotrex Systems Ltd.

This paper was prepared for presentation at the Nigeria Annual International Conference and Exhibition held in Lagos, Nigeria, 5–7 August 2019.

Abstract

The existing decline curve analysis (DCA) equations, some with valid theoretical justifications, cannot directly react to changes in operating conditions. Thus, they all assume constant operating conditions over the flowing life of a well. This however is an obvious oversimplification.

This paper begins by briefly reviewing Gilbert's equation for flowrate prediction and then the C-curve and Logistic growth model DCA theories. The above review serves to identify well key flow indicators (KFI) and performance drivers. Subsequently, a forecasting approach which involves building artificial neural network (ANN) frameworks and training them on well KFI data is presented.

Using trained ANNs, production forecasts were generated for three oil wells in the Niger-Delta producing from separate reservoirs under different flow regimes. The results were compared to forecasts from traditional DCA methods and material balance simulation, as well as with future production from the wells themselves. The results indicated that trained ANNs are capable of generating better performance curves than traditional DCA, with forecasts tying closely with results of material balance simulation and measured future well production rates. The ability of trained ANNs to evaluate the effect of changes in operating conditions (i.e. FTHP, GOR and water-cut) on production profiles and reserves drainable by wells, allows for scenario forecasting which is invaluable in field development planning. This is illustrated with field cases.

This paper also presents a novel approach to evaluating the optimal hyperparameter configuration (i.e. the number of layers, neuron count per layer, dropout, batch size and the learning rate) required to minimize the loss function whilst training an ANN on any given dataset. This should prove invaluable to engineers and geoscientists integrating deep learning into sub-surface analyses.

Introduction

Further development planning of brown fields requires engineers to forecast production profiles and drainable reserves for flowing wells and generate decline profiles for proposed wells. In such scenarios, decline curve analysis has proven to be a very useful tool as it is less complex than building material balance or numerical simulation models and may give reasonable estimates of production. The wide acceptability of Arps decline equations proposed in 1945 thus stems from its simplicity and eventual theoretical justification by Fetkovich and Cox in separate works some 35 years later.

Arps (1945) described the exponential decline equation as valid for slightly compressible fluids produced from a bounded reservoir at constant bottom-hole flowing pressure, BHFP. His proposed equation made use of two inputs; production rate and time (or cumulative production), through which a decline factor was estimated and used to forecast future production. In cases where the above conditions are invalid, production rate is proportional to the reservoir pressure and the resulting decline is non-exponential. The proposed hyperbolic equation designed to accommodate this decline type still made use of rate and time to evaluate the associated b-exponent that in effect, skews the decline curve. In forecasting with Arps decline equations, the main assumptions are that the sum of energies helping to produce the oil or gas changes in a uniform manner and that the wells are operated at constant conditions (Stephen and Bobby 2008).

While the first assumption may be valid for depletion-type reservoirs, operating conditions during the life of a well are bound to change—a well may be choked back to satisfy handling constraints or to curtail water influx, both of which offsets the decline profile of the well. In fact, the onset of substantial water production or increasing gas saturations can alter the decline pattern of an oil well.

It follows then that forecasting production must be done holistically considering the effects of changing operating conditions and fluid ratios on the well's primary fluid production. Rate and time (or cumulative production) alone simply cannot suffice.

In 1954, Gilbert proposed his famous correlation relating gross liquid production at well head to choke size, gas-liquid ratio (GLR) and tubing head pressure, (THP). The correlation was said to be valid for critical flow (Gilbert 1954) and subsequently, several other authors modified his correlation to fit observed trends in different producing regions each modifying the constants (Baxendell 1958, Achong 1961). In trying to isolate and predict oil rate rather than liquid rate, Ghareeb modified the equation to replace the GLR with producing gas-oil ratio (GOR) and Water-cut (Ghareeb 2007). Most of these Gilbert-type correlations have been shown (in author's published results) to be useful in evaluating flowrate when good estimates of the dependent variables are available.

Furthermore, in the development of the C-curve decline equation, authors argued that an exponential decline mirrored the exponential decay of radioactive elements and is a function of the remaining reserves (Peter and Marie 2005). This ideology is supported in the Logistic Growth Model proposed by Clark et. Al. in 2011, where an estimate of the ultimate recovery is set as the carrying capacity and production progressively declines as NP approaches UR, until production becomes zero.

Well Key Flow Indicators (KFI)

Combined evaluation of Gilbert's correlation, C-Curve, and Logistic Growth model hypotheses suggests that well surface production rate is a function of the following parameters:

- Bean size
- FTHP
- GOR
- Water-cut, and
- Remaining UR

The above have been termed the well Key Flow Indicators, **KFI**.

It is important to note that time on its own does not affect production rate. It is instead a metric used to quantify the rate of change of the above-mentioned properties. It was therefore not considered part of a well's KFI.

Artificial Neural Networks

Artificial Neural Networks (ANNs) are computational systems whose design was largely inspired by the structure of and connections between neurons observed in the brains of biological organisms. Although the first computational models for neural networks were created in 1943 by Warren McCulloch and Walter Pitts, the familiar multi-layered variant of neural networks were not designed until 1965 by Ivakhnenko and Lapa. These designs however did not come into popular use until the creation of the backpropagation algorithm by Paul Werbos (in 1975), and a general increase in available computing power thanks to the proliferation of graphics processing units (GPUs) and distributed computing. Both of these advances made training ANNs much quicker and thus more feasible.

Although ANNs particularly excel at image and visual recognition tasks, they have also been applied to a wide range of classification and regression problems, often with excellent results. Figure 1 below shows the current structure of artificial neural networks in use today.

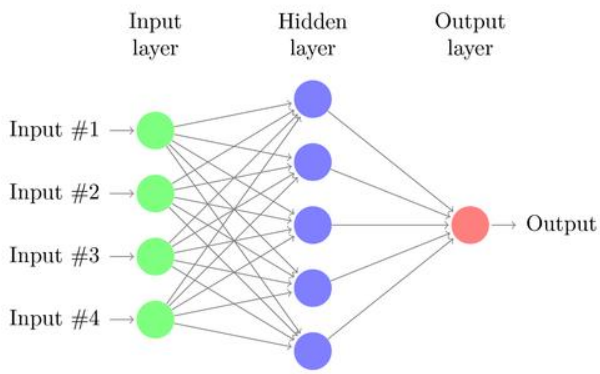

Figure 1: Typical structure of a neural network (source: texample.net).

ANNs are comprised of a series of layers – each containing a user-defined number of neurons. Each layer is connected to the succeeding layer by individual connections between every neuron in both layers (as shown the image above). Each individual connection is defined by the linear function:

$$y = w \cdot x + b$$

where:

> x is the input into the current neuron (output from the previous neuron)
>
> w is the current weight of the connection between both neurons
>
> b is the bias applied to the connection between both neurons, and
>
> y is the output from the current neuron (input into the next neuron)

Depending on the problem the ANN is intended to solve, a non-linearity may be introduced into this linear connection by using an activation function such as the rectified linear unit (relu), sigmoid or hyperbolic tan (tanh) function. The chosen activation function is typically applied to all the neurons in a given layer and changes the above equation to the form:

$$y = f(w \cdot x + b)$$

Where:

> f is the chosen activation function (to introduce a non-linearity)

This cycle of layers and subsequent connections begins at the input layer and continues throughout the network until the output layer is reached.

From the description above, the weights and biases assigned to the connections between neurons in an ANN can be seen to be the single most important factor in the accuracy of its predictions on any given input dataset. Since every ANN initially has randomly generated weight and bias values assigned to each connection, ANNs must first be trained using data from the problem domain before they can be used in predictive analysis.

Several aspects of the training process are governed by hyperparameters and the value of these hyperparameters has an enormous impact on the level of success of the ANN training process. Some of these hyperparameters include the learning rate, the number of neurons per layer, the overall number of layers in the ANN, the value of dropout used (if any – to prevent overfitting), the batch size (i.e. the number of input data samples that are passed through the neural network before backpropagation), and the number of training epochs (i.e. the number of times to pass all the input data through the network before training is considered complete). Choosing ideal values for the hyperparameters is one of the biggest challenges in the use of ANNs for predictive analysis (second only to data gathering and cleanup).

In order to obtain the optimal hyperparameter combination (or neural network structure, **NNS**) whilst training an ANN on any dataset, an algorithm was designed to take as input: ranges for certain ANN hyper-parameters and perform training simulations within the provided ranges, whilst evaluating and optimizing the NNS based on improvements in the validation loss at the end of each simulation. The development of this algorithm is outlined in Appendix A.

Methods and Results

Field Case Application

For this study, ANNs are applied to forecast oil production from three wells producing from separate reservoirs in the Niger Delta. Analysis carried out on the third well is shown in Appendix B. The flow regimes of the first two wells are described below.

1. Well-1 drains a saturated reservoir producing under boundary dominated or pseudo-steady state flow.

2. Well-2 drains an Under-saturated reservoir that is still experiencing steady state flow (or a constant pressure boundary).

Case 1 (PSS Flow)

Well-1

Well-history

Daily production data for Well-1 spanning 4 years was used in this study. During the early stages of production, the overlaying reservoir gas cap was blown-down giving rise to high GOR and FTHP (illustrated in Figure 2 below). The well KFI is compiled and the Remaining UR (RUR) generated by subtracting the cumulative daily oil production at each time step from a pre-estimated UR (from volumetric calculations).

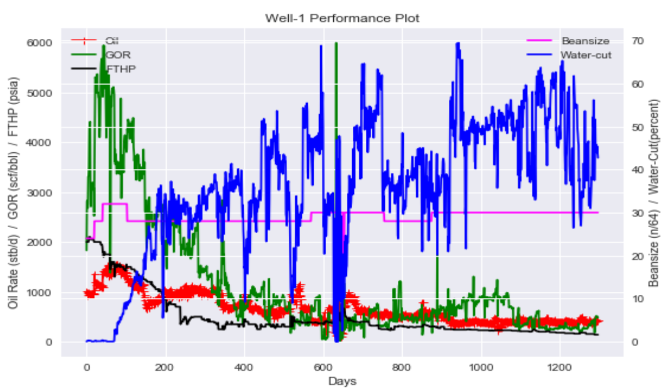

Figure 2: Well-1 Production performance plot showing well KFI and their changes overtime.

Data Analysis

Due to several factors which may include abrupt changes in operating conditions, human error or faulty production allocation algorithms, spurious data is sometimes entered into well information databases. High frequency data such as daily production figures are particularly prone to containing this kind of erroneous data which is detrimental to the learning process of an ANN. In trying to match this erroneous data, the ANN will learn incorrect weights and biases which will reduce prediction accuracy.

In order to address this issue, the daily production data was passed through an outlier detection algorithm (Appendix C) to identify and remove spurious data from the original input data and generate the final dataset passed to the ANN for training. The outlier detection algorithm used for the rest of this study centered on the Inter-quartile range (IQR). The IQR is able to easily identify and remove outliers because by definition, it is unaffected by extreme values in data sets. The original unfiltered dataset can be seen in Figure 3 below and the results of the application of the IQR outlier algorithm are illustrated in Figures 4 and 5.

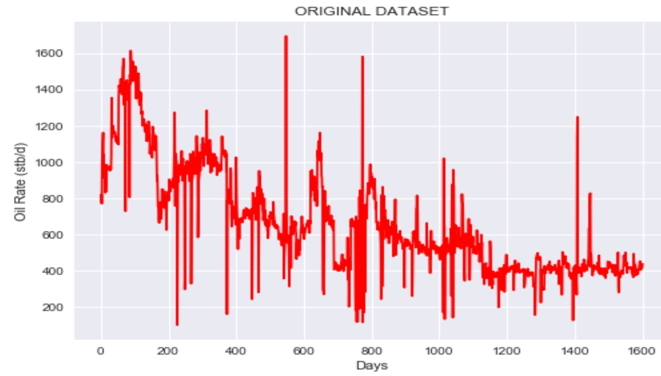

Figure 3: Well-1 Original dataset with spurious data points before application of the IQR outlier detection algorithm.

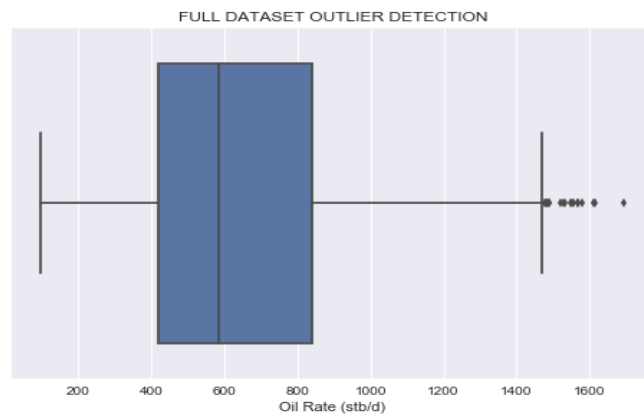

Figure 4: Box plot showing IQR Outlier detection on full dataset.

```
No. of dropped rows:  326
```

Figure 5: Well-1 Filtered dataset after application of IQR outlier detection algorithm, 326 rows of spurious data dropped.

KFI correlation panel

A correlation panel relating the entire well KFI was generated with the aim of visually identifying the key cause of decline in the well oil rate. This is given in Figure 6.

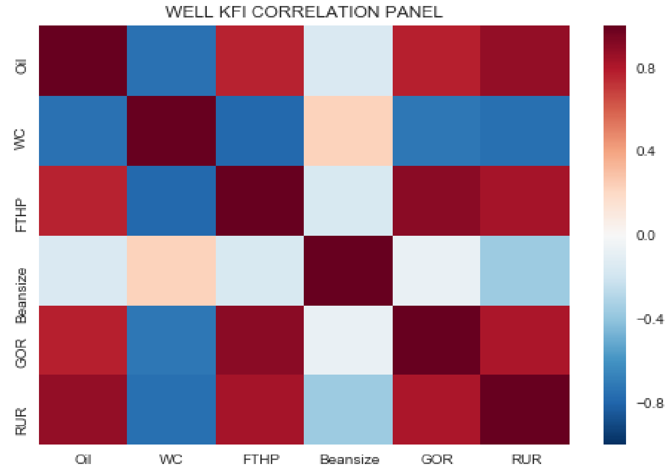

Figure 6: Well-1 KFI correlation panel with a side scale (Deep blue indicates a strong inverse correlation while dark red indicates a highly positive correlation)

In analyzing a KFI correlation panel, there are three major correlations to look out for:

a. **Oil rate correlation with water-cut**—a strong negative correlation indicates that water breakthrough is a key reason for oil rate decline.
b. **Bean size correlation with FTHP**
c. **Oil rate correlation with bean size**—if there is a positive correlation here and (b) above also has a strong negative correlation, it may indicate that the well's production has been constrained.

Based on analysis of Figure 6, the primary reason for the decline in well-1 oil rate has been attributed to an influx of water as is evident in the performance plot (Figure 2).

ANN Training

Well-1 KFI was passed into the NNS optimization algorithm for training and optimization. Based on the results of optimization, the following hyper-parameters were selected for training of the ANN for well-1:

Table 1: Well-1 ANN hyper parameter selection	
Parameter	**Well-1**
Split: Train/Test	85/15
Learning Rate	0.001
Number of Layers	2
Neurons per layer	64
Dropout (all layers)	0.5
Optimizer	Adam
Loss function	MSE
Batch size	300
EUR	2.85

The ANN was trained with 85 percent of the filtered dataset and the match to historical data is shown in Figure 7 below:

Figure 7: ANN oil rate prediction vs. actual historical production data.

ANN Forecast Comparison with Arps DCA

The trained ANN was used to predict oil flowrate on test data which comprised of the remaining 15 percent of the filtered dataset. Using the oil rate and time(days) from the training data, Arps exponential and hyperbolic decline curves were generated and used to predict oil flowrate for test data period. Comparison plots of the forecasts are given in Figures 8 and 9:

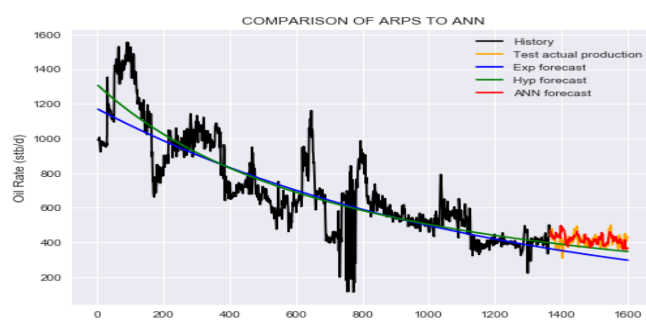

Figure 8: Comparison plot showing historical training data and test data predictions.

Figure 9: Zoom in on test data forecast by trained ANN vs. Arps DCA equations.

Analysis of Figure 9 reveals the superiority of the trained ANN to the traditional Arps decline curves. It is able to account for undulating changes in well KFI (i.e. FTHP, GOR, WC) before computing the corresponding oil rate. This superiority is also illustrated in the statistical analysis table below:

Table 2: Well-1 Statistical analysis of errors on forecast

Forecast Type	WELL-1	
	Error %	MAPE %
Arps exponential	-18.85	19.54
Arps hyperbolic	-11.29	10.16
Trained ANN	1.17	6.4

One key reason for forecasting production is to evaluate reserves drainable by producing wells. A significant under-estimation (or in some cases over-estimation) of the reserves can be made if other key factors (such as the water-cut profile) that account for oil production are not considered in forecasts. This is illustrated in Figure 10 below:

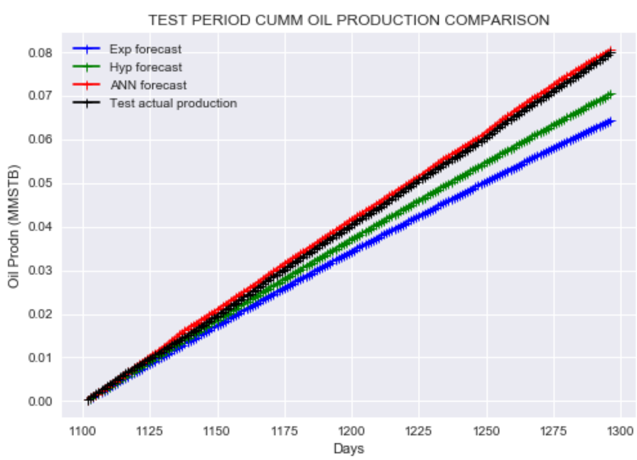

Figure 10: Corresponding cumulative oil production predicted within the test data period by trained ANN and Arps DCA equations.

Also, consider a field development plan in which drilling schedules and stream dates of new wells are controlled by demand/facility constraints and adequate forecast are the basis of this plan. If well flowrate forecasts are generated using Arps decline curves alone, an engineer may significantly under-estimate the gross field production resulting in new wells being drilled and streamed earlier than required thus creating facilities handling issues.

ANN Scenario Forecasting

Uncertainties may exist in predicting certain well and reservoir properties (such as THP and fluid ratios), as such one forecast profile may not be sufficient for field development and planning purposes. A trained ANN allows for scenario forecasting, by which the engineer can define well KFI changes overtime and then generate the corresponding oil rate profile.

In this study, sensitivities were carried out on the FTHP, Water cut and GOR forecast profiles and the following plots show the ability of a trained ANN to adequately adapt to account for these variations. The resulting oil rate profiles are smooth curves because smooth KFI forecast profiles were used as input to the trained ANN. In Figures 11-13 below, the lines with markers represent the oil rate profile while those without markers represent the KFI being varied:

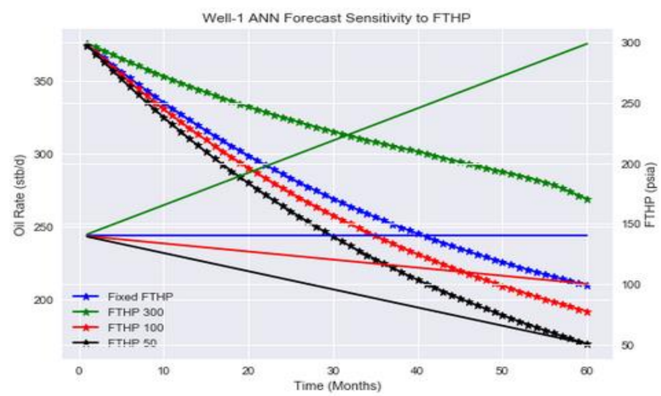

Figure 11: ANN scenario forecast with various FTHP profiles.

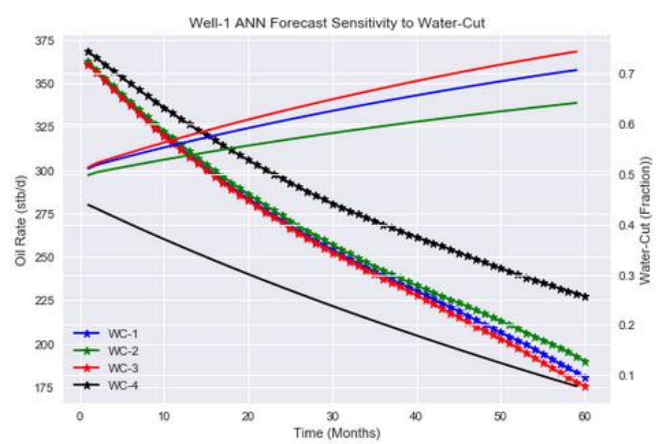

Figure 12: ANN scenario forecast with various Water-cut profiles.

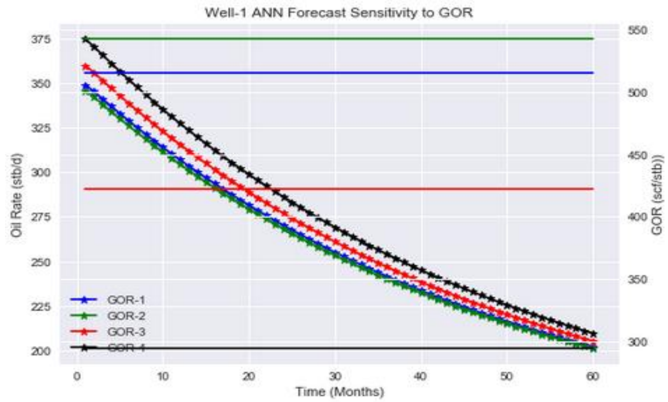

Figure 13: ANN scenario forecast at different fixed GOR.

As is illustrated in Figures 11-13, the trained ANN exhibits robustness in evaluating various scenarios of well decline and thus allows the engineer some flexibility in predicting well performance.

Case 2 (Steady State Flow)

Well-2

Well-history

Well-2 is a deviated well completed in an under-saturated reservoir with strong aquifer support and has produced between ~2500-3000bbls/d with minimal decline. In the flow history of well-2 illustrated in Figure 14 below, the water-cut has been < 0.5% (excluding the sharp spike assumed to be a measurement error).

Figure 14: Well-2 Production performance plot showing KFI

Data Analysis

Daily production data, spanning over a year was used in this study. The data was passed through the IQR outlier detection algorithm and the filtered data shown in Figure 15 was used for training the ANN. The KFI correlation panel for well-2 is shown in Figure 16.

```
No. of dropped rows:  137
```

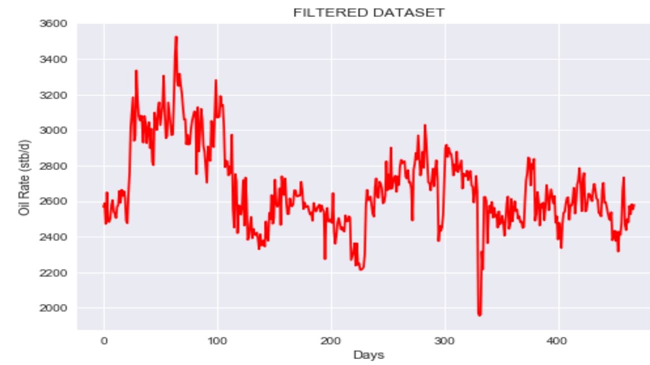

Figure 15: Well-2 Filtered dataset after application of IQR outlier detection algorithm, 137 rows of spurious data dropped.

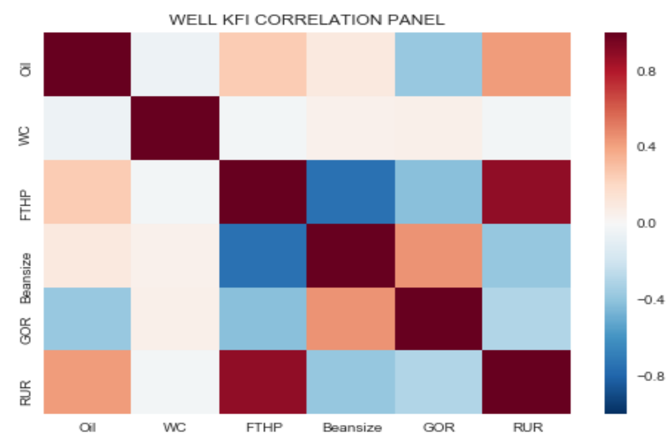

Figure 16: Well-2 KFI correlation panel.

Analysis of Well-2 KFI correlation panel reveals the following:

- Negligible correlation between oil rate and water-cut
- Slight positive correlation between the FTHP and oil-rate
- Slight positive correlation between the bean size and oil rate

The implications of the above are briefly discussed. Well-2 shows no clear observable reason for well decline apart from a slight correlation between FTHP and oil rate. This indicates that prior to water break-through in the well perforations, any decline in well production performance will be due to decline in the FTHP. However, this decline will be very gradual as the reservoir is still experiencing steady state flow and the reservoir pressure is maintained by a strong aquifer. Well-2 reservoir pressure history is shown in Figure 17 below.

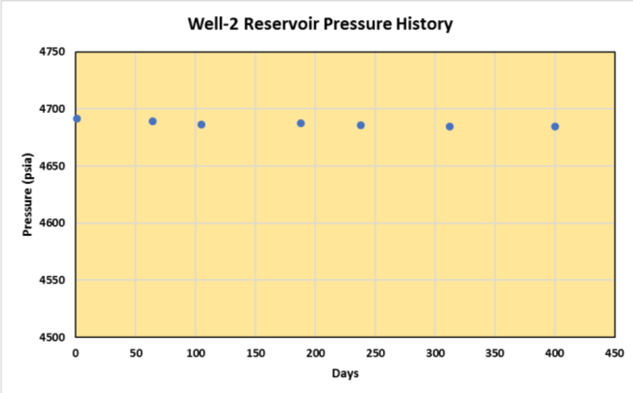

Figure 17: Well-2 reservoir pressure history

ANN Training

The NNS optimizer evaluated the following metrics as the best for well-2 training:

Table 3: Well-2 ANN hyper-parameter selection	
Parameter	**Well-2**
Split: Train/Test	85/15
Learning Rate	0.001
Number of Layers	2
Neurons per layer	64
Dropout (all layers)	N/A
Optimizer	Adam
Loss function	MSE
Batch size	100
EUR	7.03

Because the training data is small (i.e. <1000 KFI points), the model selected a smaller batch size for training, and the dropout was zero as all the neurons were able to learn (adjust weights and biases) adequately.

ANN Scenario Forecasting

In well-2, because the water-cut shows a near zero correlation with the oil rate, one would expect that the weights and biases attached to the term in the trained ANN will make its significance negligible. This is however an erroneous assumption. A forecast with various water-cut profiles is simulated and shown in Figure 18 to illustrate:

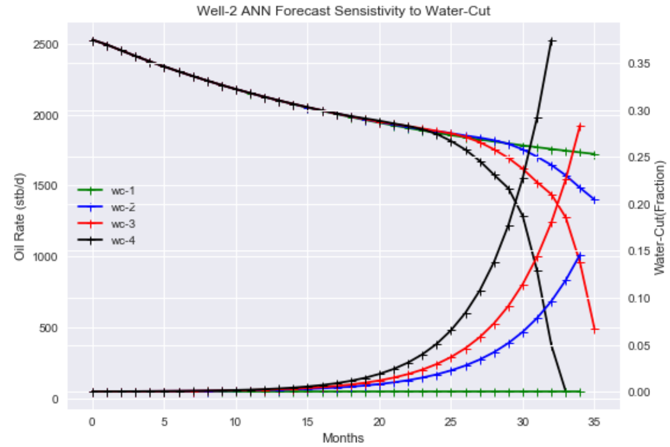

Figure 18: Well-2 ANN scenario forecast with various Water-cut profiles

From Figure 18, the ANN is able to simulate decline in oil rate production during water breakthrough periods despite historical data having no such pattern. This decline trend predicted is similar to that of a forecast generated for well-2 using a material balance simulator that captures all the flow physics that may be observed in reality.

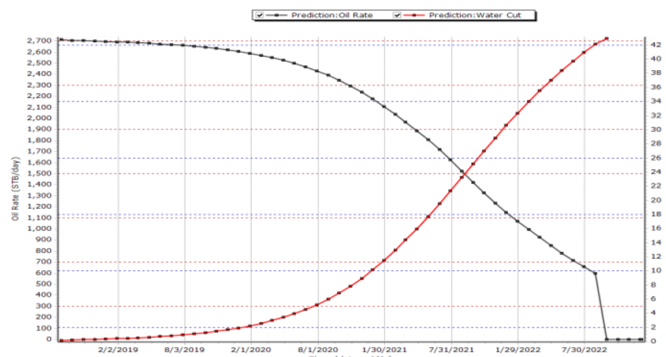

Figure 19: Well-2 production forecast simulated in PETEX MBAL

At this point, it is clear that even with limited data, the ANN generates a more realistic forecast of production when compared to Arps decline curves. This is simply because Arps curves cannot react to changes in operating conditions. Rather, the hyperbolic and exponential decline curves extrapolate production over a longer period and thus, (in this case) widely over-estimates the UR drainable by well. This can be very detrimental in field development and fiscal planning of the operating company.

Discussion and Recommendations

Neural networks are designed in such a way that they are able to learn from historical data and decipher patterns that result in one or multiple outcomes. This learning process gives rise to a major limitation when forecasting—pre-trained ANNs may perform poorly when tested on KFI profiles with patterns that differ widely from those in the training dataset. In sum, the forecast quality is only as good as your input dataset.

To adequately train an ANN and improve forecast accuracy, the following are recommended:

- Ensure well KFI is actual field measured data
- Training dataset should contain a minimum of 1000 KFI data points (even though <1000 KFI points were used in well-2, ANNs have been proven to work better when larger training datasets are used)
- The dataset should show an overall decline trend in well flowrate otherwise predictions may become unrealistic
- Always filter the dataset to remove outliers as spurious data points are detrimental to learning process. This can be done using multiple passes of the IQR algorithm until an acceptable data table is generated
- In Test/Train splitting, ensure the training data captures the most recent well decline trend as this significantly improves prediction accuracy

General recommendations on using an NNS optimizer:

- Multiple optimization runs using smaller ranges of hyperparameters tend to work better than single optimization using wide ranges
- Less neurons per layer (<128) tend to generate better results (depending on the size of your training dataset)
- Unless KFI data is greater than 5000 points, one or two ANN layers function best
- Smaller dropouts generally improve model performance (zero dropout may be best in some cases)

Conclusions

This paper introduces a newer method of forecasting production in flowing wells using trained ANNs. The method requires less data and is less rigorous than building a reservoir model but is more flexible than forecasting with traditional Arps decline curves. The ability to generate forecasts for different scenarios of changing well and reservoir properties makes it a very useful tool for field development planning.

Further work involves grouping KFI data from wells on decline in the same field and using them to train ANN models. Thus, production profiles may be generated for proposed wells for further field development.

Finally, in order to promote one key element of Data Science; "*Reproducibility*", the code used in this study has been made available on GitHub via the following link: https://github.com/davvic/ann-forecasting-well-kfi

Acknowledgements

The authors would like to thank Niger Delta Petroleum Resources Limited for their support and permission to use the well data presented in this paper. We are also incredibly grateful to Chief Edirin J. Abamwa for the inspiration to explore this research area.

Nomenclature

ANN – Artificial Neural Network

DCA – Decline Curve Analysis

FTHP – Flowing Tubing Head Pressure

KFI – Well Key Flow Indicators

RUR – Remaining Ultimate Recovery; this is the ultimate recovery less the total production from the well

WC – Water-cut

References

Achong, I. 1961. Revised Bean Performance Formula for Lake Maracaibo Wells. Internal Report, Shell Oil Co. Houston, TX.

Arps, J.J. 1945. Analysis of Decline Curves. Trans., AIME 160:228-247.

Baxendell, P.B. 1958. Producing Wells on Casing Flow—An Analysis of Flowing Pressure Gradients. Transactions of the American Institute of Mining, Metallurgical and Petroleum Engineers, Vol. 213: 202-205.

Clark, A.J. and Lake, L.W. and Patzek, T.W. 2011. Production Forecasting with Logistic Growth

Models. Paper SPE 144790 presented at the SPE Annual Technical Conference and Exhibition, Denver, 30 October-2 November.
Cox, D.O. 1978. Reservoir Limit Testing Using Production Data. The Log Analyst (March-April).
Fetkovich, M.J. 1980. Decline Curve Analysis Using Type Curves. JPT 32 (6): 1065-1077. SPE-4629-PA. DOI: 10.2118/4629-PA.
Ghareeb, M. and Shedid, A. 2007. A New Correlation for Calculating Wellhead Production Considering Influence of Temperature, GOR, and Water-Cut for Artificially Lifted Wells. Paper 11011 presented at the International Petroleum Technology Conference.
Gilbert, W.E. 1954. Flowing and Gas-Lift Well Performance. Dril. And. Prod. Prac. API 143: 127-154.
Neilson, M.A. 2015. Neural Networks and Deep Learning. Determination Press.
Peter, R.P.C. and Marie, N.L. 2005. C-curve and Independent Oil-Rate and Oil-Cut Decline. Paper SPE 96092 presented at Offshore Europe, Aberdeen, 6-9 September.
Stephen, W.P. and Bobby, D.P. 2008. Analysis of Production Decline Curves. Society of Petroleum Engineers, Richardson, TX.

Appendix A—ANN Hyper-parameter Optimizer

The following lines describe the algorithm developed for optimizing the neural network hyper-parameters and evaluating the best ANN configuration to use for predictive analysis on the input dataset.

1. The algorithm is provided with input data for training/validation, the index at which to split the data into training and validation datasets, and ranges for each hyperparameter. The algorithm supports varying the following: layer count, neurons per layer, learning rate, batch size and dropout.
2. The ranges provided are used to generate hyperparameter combinations in the following way:
    a. The first value in the layer count range is constantly incremented by one until the second value is reached. All values obtained (inclusive) are added to a list.
    b. The first value in the neuron count per layer range is constantly multiplied by 2 until the second value is reached. All values obtained (inclusive) are added to a list.
    c. The first value in the learning rate range is constantly multiplied by 10 until the second value is reached. All values obtained (inclusive) are added to a list.
    d. The various batch sizes to be tested are provided as a list.
    e. The dropout values to be tested are provided as a list.
    f. All lists obtained from this process are combined (mathematically speaking) to give a set containing all hyperparameter combinations to be tested.
3. The input data is scaled and split (using the provided index) into train and validation datasets.
4. For each hyperparameter combination, an ANN is generated and trained. The algorithm uses Early Stopping to determine the number of training epochs required for each ANN. Early stopping works by calculating the validation loss after each epoch and stopping training if the validation loss is no longer decreasing (after waiting a preset number of epochs). The best performing weights and biases (by validation loss) for each hyperparameter combination is saved to a file on the disk and the algorithm moves on to the next combination.
5. When the ANN corresponding to each hyperparameter combination has been trained and the best performing weights and biases stored, the algorithm compares all the validation losses to find the model with the lowest overall validation loss and returns this as the best performing model for the provided training data.

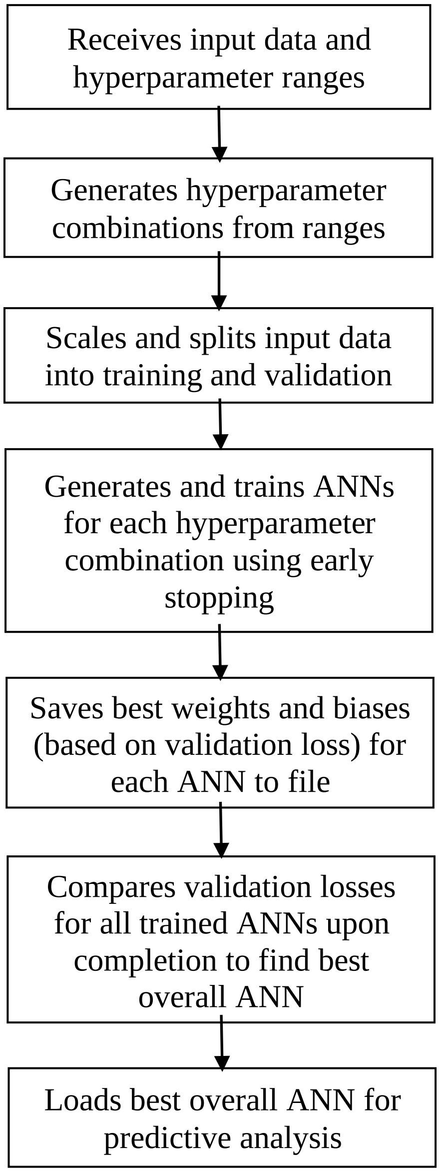

Figure A—1: Flow-chart showing ANN optimizer process

Appendix B—Further Field Case Application

Forecasting with ANN was applied to a third well producing from an initially under-saturated reservoir that was near saturation. The reservoir is bounded by faults and supported by a weak aquifer and thus with production, the reservoir pressure depletes below the saturation pressure and free gas is produced alongside the oil in the later stages of well life. The well goes into decline with significant water breakthrough.

A trained ANN was used to forecast production on test data and the resulting plots showing comparison with Arps decline equations are shown below:

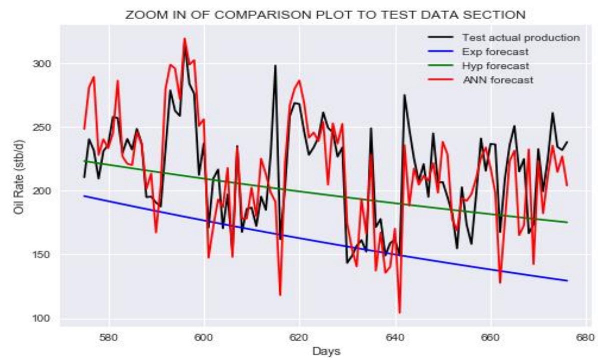

Figure B—1: Zoom in on test data forecast by trained ANN vs. Arps DCA equations.

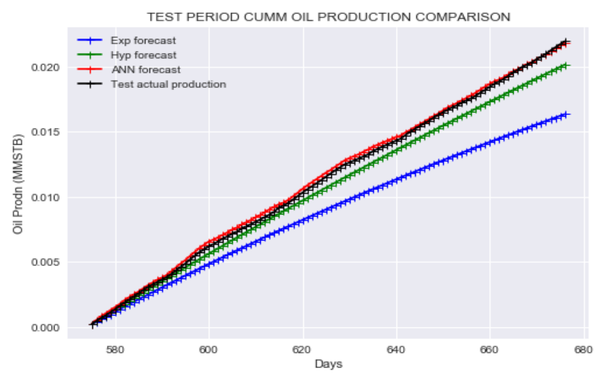

Figure B—2: Corresponding cumulative oil production predicted within the test data period by trained ANN and Arps DCA equations.

Appendix C—IQR Outlier Detection

The description below outlines the algorithm used in removing outliers from well KFI dataset.

Split dataset into groups; each group consisting of approximately 30 days of production data. For each group:

1. Take Oil flowrate as your primary target data and find the 25th and 75th percentile (i.e. Q1 and Q3).
2. Compute the IQR = Q3-Q1
3. Calculate an upper bound = Q3+1.5*IQR
4. Calculate a lower bound = Q1-1.3*IQR
5. Remove all data points greater than the upper bound or less than the lower bound.

Repeat step 2-5 using FTHP, GOR and Water-cut as your primary target data.