

# Artificial Neural Network Model to Predict Production Rate of Electrical Submersible Pump Wells

Ahmed Sabaa<sup>1\*</sup>, Mahmoud Abu El Ela<sup>2</sup>, Ahmed H. El-Banbi<sup>3</sup>, and Mohamed H. M. Sayyounh<sup>2</sup>

<sup>1</sup>Agiba Petroleum Company

<sup>2</sup>Cairo University

<sup>3</sup>The American University in Cairo

## Summary

Production data are essential for designing and operating electrical submersible pump (ESP) systems. This study aims to develop artificial neural network (ANN) models to predict flow rates of ESP artificially lifted wells. The ANN models were developed using 31,652 data points randomly split into 80% (25,744 data points) for training and 20% (5,625 data points) for testing. Each data set included measurements for wellhead parameters, fluid properties, ESP downhole sensor measurements, and variable speed drive (VSD) sensors parameters. The models consisted of four separate neural networks to predict oil, water, gas, and liquid flow rates. Sensitivity analyses were performed to determine the optimum number of input parameters that can be used in the model. The best performance was achieved with ANN models of 16 input parameters that are readily available in ESP wells.

The results of the best ANN configuration indicate that the mean absolute percent error (MAPE) between the predicted flow rates and the actual measurements for the testing data points of the oil, water, gas, and liquid networks is 3.7, 5.2, 6.4, and 4.1%, respectively. In addition, the correlation coefficients ( $R^2$ ) are 0.991, 0.992, 0.983, and 0.979 for the estimated oil, water, gas, and liquid flow rates for the testing data points, respectively. The performance of the ANN models was compared against performance of published physics-based models and the results were comparable. Unlike the physics-based methods, the ANN models have the advantage that they do not require periodic calibration. The ANN models were used to predict the flow rate curves of an oilwell in the Western Desert of Egypt. The results were compared to the actual separator test data. It was clear that the model results matched the actual test data.

The ANN model is useful for predicting individual well production rates within wide variety of pumping conditions and completion configurations. This should allow for continuous monitoring, optimization, and performance analysis of ESP wells as well as quicker response to operational issues. In comparison to traditional separators and multiphase flowmeters (MPFMs), the use of the developed ANN models is simple, quick, and inexpensive.

## Introduction

ESP systems provide efficient artificial lift to many wells. They have improved significantly since they were invented in the 1910s. One of the key monitoring techniques for ESP-lifted wells is the determination of surface flow rate generated by the pumps utilizing two-phase separators or MPFMs. The infrequent measurement of flow rates may prevent faster reactions to pumps' operational problems, especially in mature fields with a large number of wells. Before installing a pump, flow rate data are required for proper design and equipment selection. Following pump installation, flow rate is required to validate the design and ascertain whether any formation damage had an impact on the design productivity index. Without equipment removal, this formation damage is probably beyond repair, so it is necessary to change the pump's operating parameters as well as the wellhead parameters to keep the pump running within its operating range. Among the main failure reasons of ESPs is the missing or wrong flow rate information during the artificial lift design stage. The cost impact of ESP failure is significant as is the cost of flow rate measurements using MPFMs or test separators (Alhanati et al. 2001).

Real-time flow rate estimation is now achievable thanks to the advancements in communication, well technology, and field equipment (Gupta et al. 2016). This can lead to establishing approaches for continuous estimation of multiphase well rates utilizing models and indirect field sensor readings, rather than employing multiphase orifices, venturis, or different commercial three-phase meters. Pump operations, reservoir management, production allocation, and cost optimization all benefit from real-time flow rate predictions. For ESP-lifted wells, there are two ways to estimate flow rate: physics-based models and data-driven models.

In the physics-based models, the multiphase flow rates are estimated by simplified hydrodynamical models (approaches) using wellhead and ESP sensors data as input parameters. Olsen et al. (2012) presented an algorithm that uses ESP real-time monitoring data to estimate the liquid production flow rate of a heavy oil field. In another study, Krylov et al. (2019) developed an algorithm that can estimate the missing values of flow rates using power consumption and hydraulic calculations data of gas-liquid mixture flow rates in the tubing. Data validation and reconciliation algorithm is another approach in the physics-based methods. The data validation and reconciliation process embedded in a commercial software package describes the process of adjusting the virtual flowmeter model parameters such that the virtual flowmeter model output matches the measured field data (Couput et al. 2008; Haouche et al. 2012).

Machine learning techniques have been utilized in various aspects of the upstream oil business in recent years. Data-driven modeling has been closely tied with machine learning, and the two terminologies have been used interchangeably to describe the same concept (Temizel et al. 2017). The data-driven models depend on the available data in the field by performing statistical analysis on the data and deriving relations between input features and quantities of interest.

\*Corresponding author; email: ahmed.sabaa@agiba.com

Copyright © 2022 Society of Petroleum Engineers

Original SPE manuscript received for review 15 June 2022. Revised manuscript received for review 10 August 2022. Paper (SPE 212284) peer approved 18 August 2022.

Supplementary materials are available in support of this paper and have been published online under Supplementary Data at <https://doi.org/10.2118/212284-PA>. SPE is not responsible for the content or functionality of supplementary materials supplied by the authors.

There is limited work for data-driven flow rate estimation of ESP-lifted wells. Al-Jasmi et al. (2013) used various tools based on neural networks to model the behavior of ESP and gas lift wells and to predict the flow rates. Binder et al. (2015) implemented the moving horizon estimation technique to estimate the flow rate for one ESP well. The model estimates also the productivity index and the viscosity. Moreover, Hallo et al. (2017) developed a mathematical model to estimate the oil flow rate of one ESP well using genetic programming. Ricardo et al. (2018) developed another model to estimate the gas flow rate only in a two-phase (liquid-gas) mixture being pumped through an ESP using the support vector machine learning algorithm. The learning machine's hidden parameters are determined with a genetic algorithm.

In general, the bulk of the virtual flowmeter systems for ESP wells are commercial in nature. As a result, vendors do not provide details of the models used in their software packages (Bikmukhametov and Jäschke 2020).

### Commonly Used Flow Rate Prediction Methods

Caicedo and Montoya (2012) developed a model to estimate the ESP flow rate using the ESP sensor readings, basic fluid properties data, and the head equation. In this method, the knowledge of the ESP model and the total number of pump stages is mandatory. The calibration technique for the ESP head performance curve requires correcting the entire pump performance curve with a single test point to force the new curve to pass through the calibration point. As a result, this method cannot be automated and thus cannot be applied to real-time data.

Camilleri introduced an analytical method for obtaining real-time liquid flow rate. The main principle in this method considers that the power absorbed by the pump is equal to that generated by the motor. This method has been implemented in several field applications. However, the model needs to be recalibrated periodically using actual production test data. In addition, it does not predict the gas flow rate (Camilleri and Zhou 2011; Camilleri et al. 2015).

Denney et al. (2013) described the performance of a commercial software in field conditions using ESP data. This software is used to estimate the flowrates of ESP systems by applying the neural network approach. However, there is limited information about the model development and the choice of the input features.

The objective of this work is to develop a neural network model that can predict multiphase flow rates of ESP wells using readily available field data. The model will cover a wide variety of pumping conditions and wells configurations and will not require periodic calibration.

### Methodology

The procedure used to achieve the objective of this work is illustrated in Fig. 1. The methodology includes the following steps:

- **Data Collection:** ESP sensor real-time data and flow rate historical data were collected and combined from several oil fields in the Western Desert of Egypt with more than 6 years of production. The collected data were used to build and validate the ANN model and compare its results with other existing physics-based methods.
- **Data Preprocessing:** The data were checked and cleaned. In addition, the missing data for some data sets were completed using statistical techniques with field knowledge.
- **Model Structure and Configurations:** The model was designed and developed with several configurations. MATLAB® software was used in the ANN model development.
- **Model Testing and Validation:** The ANN models were tested and validated. The results were then compared with other existing physics-based models.

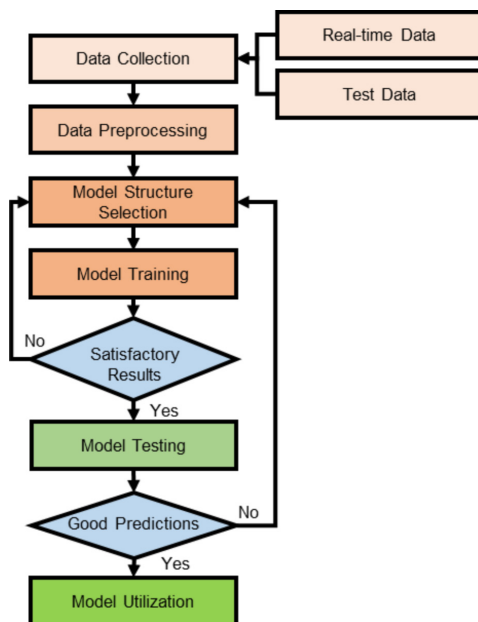


Fig. 1—Flow chart of the proposed approach for flow rate measurement.

### Model Development

**Data Collection.** Recent ESPs are often equipped with gauges and sensors for continuous health monitoring and analysis. This research began with collection of 31,970 data points of actual field measurements. The data were collected from 79 wells in the Western Desert

of Egypt covering production from 12 different reservoirs. It should be highlighted that the data sets were gathered from a variety of sources, each with its own set of formats and updating frequencies. The data sets included several parameters as shown in **Tables 1 and 2**. The real-time data of approximately 133 ESP pump installations (some wells have several ESP installations due to pump failures or well repairs) were combined with corresponding configuration and historical data [well test data, pressure/volume/temperature (PVT) data, and well completion data].

Wellhead Parameters	VSD Sensors	ESP Downhole Sensors	Fluid Properties	Water-Cut Percentage
Wellhead pressure	Drive frequency	Intake pressure	API gravity	Water cut
Wellhead temperature	Motor current	Intake temperature	Gas gravity	
Separator pressure	VSD current	Motor temperature	Oil viscosity	
	Output voltage	Discharge pressure	Gas/oil ratio	

Table 1—Features classification of the input parameters.

Element	Unit	Min	Max	Average	Median	Standard Deviation
Wellhead pressure	psig	60	3,200	740	509	538.1
Wellhead temperature	°F	85	249	198	210	33.8
Separator pressure	psig	50	1,520	308	210	290.8
Oil gravity	API	10	54	39	39	4.6
Water cut	%	0	100	35	28	31
Gas gravity	–	0.65	1.20	0.94	0.95	0.08
Gas/oil ratio	scf/STB	0	34,069	332	200	721.3
Oil viscosity	Cp	0.18	2.19	0.97	0.79	0.65
Intake pressure	psi	65	4,392	1,691	1,630	871.4
Intake temperature	°F	218	275	245	247	11.0
Motor temperature	°F	226	353	289	294	20.3
Drive frequency	Hz	32	63	49	50	4.5
Motor current	Amp	19	102	55	60	16.3
Discharge pressure	psi	600	7,087	3,952	4,007	998.6
VSD current	Amp	35	820	476	535	161.6
Output voltage	Volt	264	418	363	373	35.3
Records of data per well	–	11	8,496	395	123	999
Oil rate, $Q_o$	BOPD	0	6,697	2,003	1,656	1,572.2
Water rate, $Q_w$	BWPD	0	6,478	859	552	972.4
Gas rate, $Q_g$	MMscf/D	0	2.2	0.4	0.3	0.4
Liquid rate, $Q_{liq}$	BPD	0	6,819	2,862	2,755	1,474.9

Table 2—Statistical analysis for the collected database.

Creating each record for training and testing requires combining data from well tests and ESP sensors. Production test data are at frequency of 30-minute intervals (i.e., not real time) and are sporadic according to test separators availability. This means 31,970 well test data points were collected; then, the corresponding ESP sensors and PVT data were added for the periods covered with well tests. The time span of all data was more than 6 years.

**Data Preprocessing.** The collected data were checked and analyzed to remove erroneous values (e.g., water cut > 100%, zero values for pump intake and pump discharge pressures, and zero values for pump intake temperature and motor temperature) and frozen sensor readings. The missing data were then handled by either deleting the entire row of data or by filling the empty cells using statistical practices and field knowledge. Less than 2% of the data were discarded during the preprocessing phase with no impact on the data used for modeling.

The next step to process the database was to remove the outliers and either replace the eliminated data (e.g., replacing the eliminated data with previous or nearest nonoutlier element) or removing the entire record. The preprocessed data (31,652 data points) were then divided into two sets: The first set (80% of the full database) was used to train the model, and the second set (20% of the database) was used to test the model.

**Feature Selection.** Feature selection is a technique to reduce the computational complexity of the machine learning model and to avoid overfitting problems. This is accomplished by reducing the dimensionality of the data and creating a model with only a subset of measured features (predictor variables). Even if all features are relevant and contain information about the response variable, using too many features can degrade prediction performance.

The filter-type feature selection algorithms in MATLAB measure feature importance based on the characteristics of the features, such as feature variance and feature relevance to the response. Important features are determined as part of the data preprocessing step, and the model is then trained using the selected features. As a result, the selection of filter-type features is unrelated to the training algorithm.

The Regressional ReliefF algorithm was used in this work to rank features with K-nearest neighbors. This algorithm is suitable for estimating feature importance in distance-based supervised models that predict the response using pairwise distances between observations. Predictor ranks and weights usually depend on number of nearest neighbors, specified as a positive integer scalar ( $k$ ). Regressional ReliefF algorithm penalizes the predictors that give different values to neighbors with the same response values and rewards predictors that give different values to neighbors with different response values. Regressional ReliefF uses intermediate weights to compute the final predictor weights. A useful technique can be to assign weights to contributions of neighbor, so that the nearer neighbors contribute more to the average than the more distant ones. A common weighting scheme, for example, is to assign a weight of  $1/d$  to each neighbor, where  $d$  is the distance between them.

The features were reduced from 25 to 16 that represent the most important predictors for oil, water, gas, and liquid rates. These input features were then classified into five main groups as shown in **Table 1**: wellhead parameters, VSD monitoring, main ESP downhole sensors, fluid properties or PVT, and finally water cut. Well geometry and completion features were neglected. This selection is also in line with the parameters used in the physics-based methods.

The statistical analysis of the selected features is shown in **Table 2**. It is obvious that a broad range of reservoir characteristics are covered, and the flow rates are in a representative range for wells artificially lifted with ESPs. The wide range of parameters will lead to the creation of a generalized model.

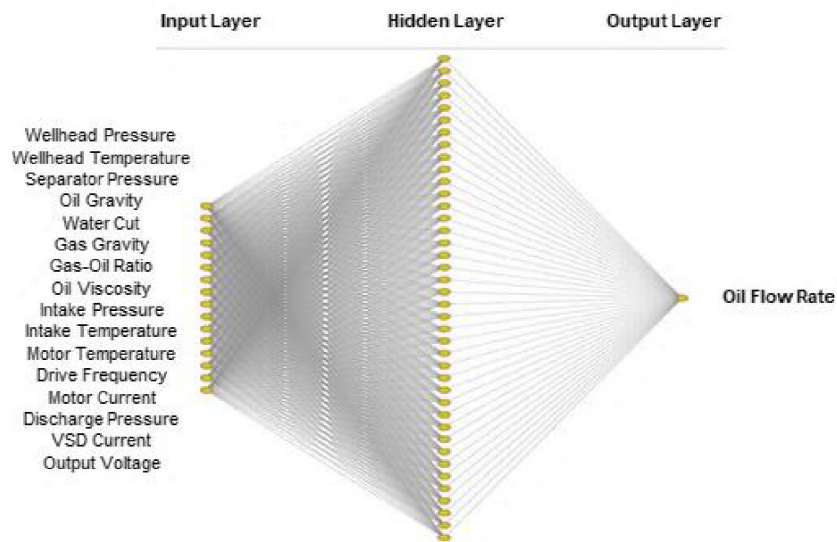
## Model Structure

The ANN is a machine learning technique that mimics the human biological neural network behavior. It is widely used in modeling complex relationships between inputs and outputs in nonlinear systems. The network consists of units called neurons which are arranged in layers and are interconnected to each other through weights. Learning or adaptation in the network occurs when the weights are adjusted to make the network produce correct outputs. The sum of the weights times the inputs is called the linear combination of the inputs.

The neuron takes the linear combination and puts it through a so-called activation or transfer function. The transfer functions calculate a layer's output from its net input. In this work, a combination of sigmoid, linear, and rectified linear unit activation functions was tried and the one yielding the smallest error was finally chosen.

Attempts to create a single ANN model with three outputs (oil, water, and gas flow rates) and with four outputs (considering also liquid flow rate) showed high accuracy for liquid, oil, and water flow rates estimation but low accuracy for predicting gas flow rate. To improve the accuracy, a new model with four separate networks was developed. Each network predicts only one output (oil flow rate, water flow rate, gas flow rate, or liquid flow rate).

Each of the developed ANN networks consists of three layers: input layer, hidden layer, and output layer with log-sigmoid in the hidden layer and hyperbolic tangent sigmoid transfer function in output neurons, respectively. Schematic representation for one of the developed ANNs is shown in **Fig. 2**.



**Fig. 2—Three-layer neural network architecture.**

The input layer represents the input parameters of the problem. Five configurations for the input layer were used to solve the problem. Each configuration depends on specific input parameters as shown in **Table 3**. The strategy was to test the effect of changing the input features on the model performance as follows:

- Configuration 1: It includes 16 input parameters (including wellhead parameters, fluid properties, ESP downhole sensors main readings, VSD sensors features, and finally water cut).
- Configuration 2: It includes 12 input parameters [all above features except the fluid properties (PVT) group]. This configuration was proposed to be used when PVT properties are believed to be erroneous.
- Configuration 3: It includes 11 input parameters (including wellhead parameters, ESP downhole sensors main readings, and VSD sensors features). This configuration was proposed to test the possibility of having a fully automated data-driven model using ESP and wellhead real-time data only.

Configuration	Input Parameter															
	Wellhead Pressure	Wellhead Temperature	Separator Pressure	Drive Frequency	Motor Current	VSD Current	Output Voltage	Intake Pressure	Intake Temperature	Motor Temperature	Discharge Pressure	API Gravity	Gas Gravity	Oil Viscosity	Gas/Oil Ratio	Water Cut
Configuration 1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Configuration 2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓					✓
Configuration 3	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓					
Configuration 4	✓	✓	✓	✓	✓	✓	✓					✓	✓	✓	✓	✓
Configuration 5	✓	✓	✓	✓	✓	✓	✓									

Table 3—Input parameters and features of each ANN configuration.

- d. Configuration 4: It includes 12 input parameters (all data excluding the ESP downhole sensors main readings). This configuration was proposed to develop a model that can be used in cases of downhole sensors failure, which can be a common problem in some ESP wells.
- e. Configuration 5: It includes seven input parameters (considering wellhead parameters and VSD sensors main readings only).

The hidden layer is the one where the computational process occurs. The number of neurons in the hidden layer is problem-dependent (Abdalla et al. 2020). An iteration process was used to identify the ideal number of neurons in the hidden layer of each network until the optimum number of neurons was identified as shown in Fig. 3. Finally, the output layer consists of one neuron which represents one response variable for each network and configuration (oil, water, gas, or liquid flow rate).

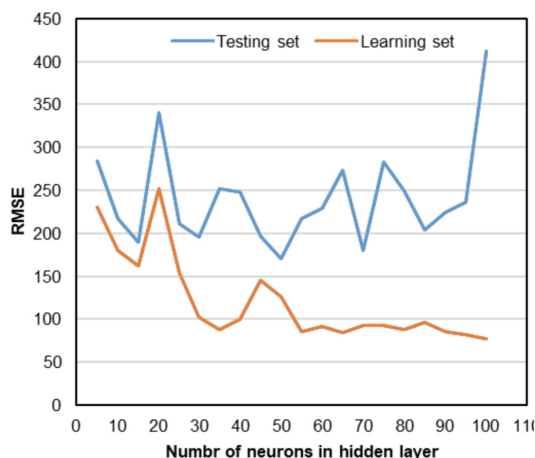


Fig. 3—ANN performance with different numbers of neurons in the hidden layer.

The ANN models were developed to predict oil, water, gas, and liquid flow rates based on the feed-forward approach in which information or signals would propagate only in one direction from input to output. The model was trained by applying a backpropagation procedure using the Levenberg-Marquardt algorithm. To guarantee the optimum performance of the network and avoid overfitting, each time a new configuration was evaluated, the neural network was carefully trained and optimally adjusted in terms of hidden layer dimensions.

In this study, we used the following performance indicators to assess the quality and the accuracy of the developed model: root mean squared error (RMSE), mean absolute error (MAE), and MAPE (Kolassa and Schütz 2007).

### Model Results

It should be highlighted that the preprocessed data were divided into two sets: The first set (80% of the database – 25,744 records) was used to train the model, and the second set (20% of the database – 5,625 records) was used to test the model. Table 4 shows the final optimized features and the results of the developed ANN models. The predicted phase flow rate, the ANN hyperparameters, and the outcomes of the training and testing data sets are demonstrated for each configuration.

The ANN model with 16 input neurons (Configuration 1) achieved the best results. The MAPE between the predicted flow rates and the actual measurements for the training data points of the oil, water, gas, and liquid networks is 3.4, 3.8, 3.6, and 3.7, respectively. However, the MAPE between the predicted flow rates and the actual measurements for the testing data points of the oil, water, gas, and liquid networks is 3.7, 5.2, 6.4, and 4.1%, respectively. Fig. 4 is the graphical presentation of the predicted oil, water, gas, and liquid flow rates using the ANN model against the actual measurements for the testing data points. The deviation of the predicted values from the measured flow rates is shown with a 45° line. It is clear that the ANN model predictions are closer to the 45° straight line and equally distributed around it.

In Configuration 2, the fluid properties group was eliminated from the input layer. This reduces the number of inputs to 12 input parameters. The ANN model of this configuration achieved MAPE between the predicted flow rates and the actual measurements for the testing data points of the oil, water, gas, and liquid networks of about 3.6, 7.1, 12.2, and 4.1%, respectively. It should be highlighted that the gas flow rate prediction of this configuration was the most affected as the MAPE is almost doubled compared to the results of Configuration

Configuration	Phase	Inputs	Number of Hidden Layers	Neurons in Hidden Layer	Training				Testing			
					RMSE	$R^2$	MAE	MAPE	RMSE	$R^2$	MAE	MAPE
1	Oil	16	1	50	126	0.993	64	3.4%	171	0.991	93	3.7%
	Water	16	1	70	64	0.995	33	3.8%	108	0.992	47	5.2%
	Gas	16	1	20	0.03	0.995	0.016	3.6%	0.04	0.983	0.019	6.4%
	Liquid	16	1	30	182	0.984	101	3.7%	230	0.979	142	4.1%
2	Oil	12	1	46	94	0.996	57	3.0%	171	0.991	90	3.6%
	Water	12	1	48	109	0.986	50	5.8%	138	0.986	63	7.1%
	Gas	12	1	50	0.05	0.985	0.030	6.7%	0.10	0.891	0.038	12.2%
	Liquid	12	1	48	141	0.990	87	3.2%	306	0.963	141	4.1%
3	Oil	11	1	50	171	0.987	102	5.4%	303	0.972	161	6.3%
	Water	11	1	60	131	0.980	74	8.7%	512	0.812	153	17.2%
	Gas	11	1	55	0.05	0.982	0.034	7.5%	0.12	0.835	0.046	14.9%
	Liquid	11	1	44	197	0.981	107	3.9%	260	0.973	148	4.3%
4	Oil	12	1	44	104	0.995	65	3.5%	292	0.974	115	4.6%
	Water	12	1	48	91	0.990	41	4.9%	240	0.959	66	7.4%
	Gas	12	1	40	0.03	0.994	0.018	4.0%	0.06	0.964	0.024	7.6%
	Liquid	12	1	40	198	0.981	122	4.4%	257	0.974	164	4.8%
5	Oil	7	1	48	226	0.977	150	8.0%	282	0.975	191	7.5%
	Water	7	1	60	182	0.961	108	12.7%	262	0.951	148	16.5%
	Gas	7	1	56	0.06	0.975	0.041	9.3%	0.12	0.823	0.052	16.8%
	Liquid	7	1	49	229	0.974	126	4.6%	287	0.967	166	4.9%

Table 4—Results of the developed ANN models.

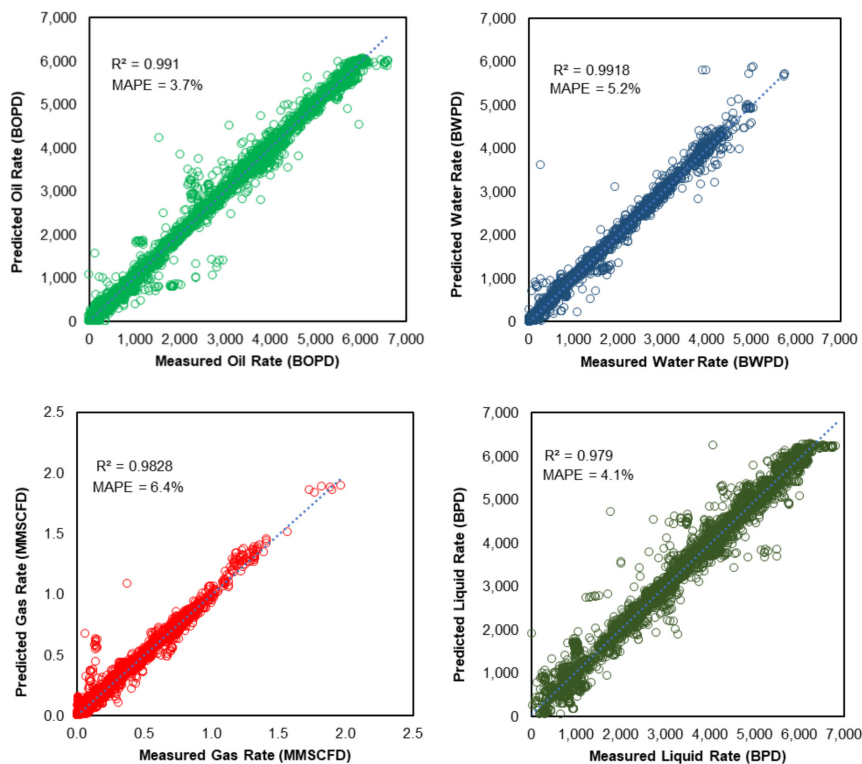


Fig. 4—Crossplot of Configuration 1 flow rate prediction for the testing data points.

1 as demonstrated in Table 4. One of the challenges during collecting the data for this study was filling out the missing values of the gas/oil ratio, because the majority of ESP-lifted wells are in low gas/oil-ratio wells. In such conditions, measuring the gas flow rates with two-phase separators is either hard or has little value.

In addition to eliminating PVT features, removing the water cut percent in the input layer of Configuration 3 increased the oil, water, and gas flow rate prediction errors (MAPE) in the testing data points to 6.3, 17.2, and 14.9 (almost double the results of Configuration 1) as presented in **Table 4**. The MAPE of liquid flow rate prediction remained as low as 4.3%. Therefore, a fully automated ANN model that uses ESP and wellhead real-time data only and capable of predicting liquid flow rates is actually possible.

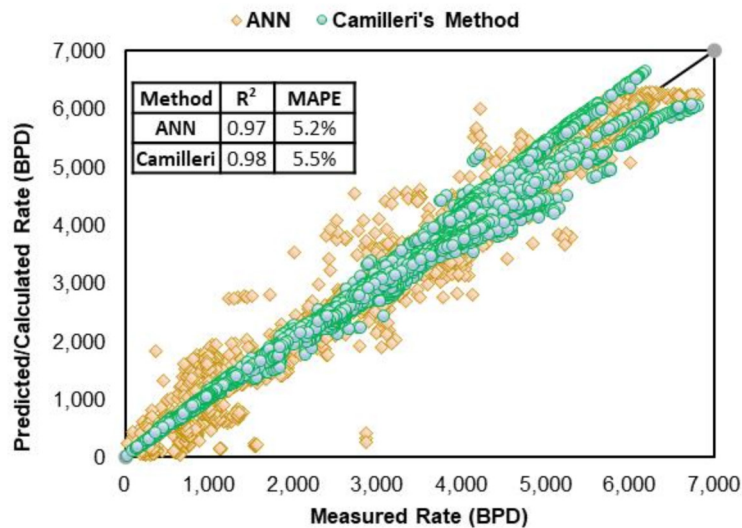
It is common to lose the real-time data of the ESP downhole sensors due to a short in one phase of the electrical cable. Therefore, Configuration 4 was proposed. In this configuration, the oil and liquid flow rate predictions are achieved with reasonable accuracy as presented in **Table 4**. The MAPE between the predicted oil, water, gas, and liquid flow rates and the actual measurements for the testing data points are 4.6, 7.4, 7.6, and 4.8%, respectively.

Configuration 5 depends on the wellhead parameters and the VSD features only (i.e., seven inputs). In this configuration, the model can still predict ESP flow rates with MAPE of 5% for liquid, 7.5% for oil, and around 16.5% for both water and gas flow rates in the testing data points. The results of Configuration 5 are presented in **Table 4**.

## Model Validation

The testing data sets were used to validate the developed ANN models against published physics-based methods (namely, Camilleri and Caicedo methods). In general, the physics-based methods can work well for ideal cases, where the operational point is located on the catalog's performance curve. However, measurement errors, pump wear, fluid properties approximations, gas or viscosity effects, and reservoir properties changes over time move the operational point off the catalog curve. Accordingly, continuous and consistent well model correction (calibration) and update, principally with the test separator results, should be made to match the model with fields measurements (Petukov et al. 2011).

Most of the testing data points (5,521 data points) were utilized to run Camilleri's model. **Fig. 5** displays comparison between the results of Configuration 1 ANN model for liquid flow rate prediction and Camilleri's method. The MAPE between the measured and calculated liquid flow rate for the ANN model and Camilleri's approach is 5.2 and 5.5%, respectively.



**Fig. 5—Comparison between the results of the ANN model (Configuration 1) and Camilleri's method (5,521 data points).**

Caicedo's equations rely largely on the availability of several input parameters such as water cut, API gravity, and the free gas rate into the pump. Owing to the complex process of model calibration in this method, it was not possible to use the entire testing data set. Consequently, only 130 data points from the testing data set had enough data to run Caicedo's process. The results indicate that the MAPE between the measured and calculated liquid flow rates for the ANN model (Configuration 1), Camilleri's approach, and Caicedo method for these 130 data points are 7.5, 5.5, and 16%, respectively. The results of the comparison are demonstrated in **Table 5** and **Fig. 6**. The reason behind the high MAPE of Caicedo's method is believed to be the high sensitivity of this method to water-cut changes and to the presence of free gas. These results confirm the strength of the developed ANN model to predict the production flow rates of ESP wells using readily available field data.

Method	5,521 Data Points				130 Data Points			
	RMSE	R <sup>2</sup>	MAE	MAPE	RMSE	R <sup>2</sup>	MAE	MAPE
ANN	269	0.97	165	5.2%	319	0.945	200	7.5%
Camilleri	240	0.98	176	5.5%	228	0.972	148	5.5%
Caicedo			N/A		609	0.8	419	16.0%

N/A: not applicable

**Table 5—Comparison between the results of the developed ANN model (Configuration 1) and physics-based models.**

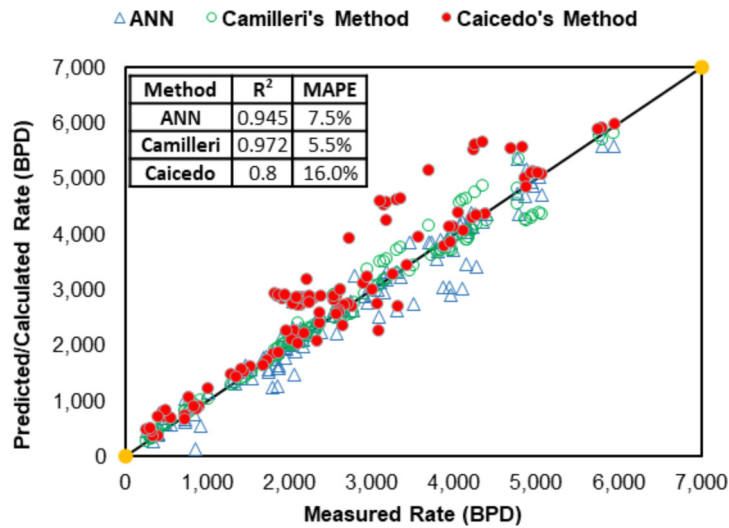


Fig. 6—Comparison between the results of the ANN model (Configuration 1), Camilleri, and Caicedo's methods (130 data points).

### Field Application

Well-X is an onshore oil well located in the Western Desert of Egypt. The well was drilled vertically to a depth of 16,500 ft. The original reservoir pressure was 6,200 psi, while the reservoir temperature was 274°F. The well was completed with an ESP pump to produce 1,000 STB/D. The ESP was set at 9,900 ft due to limitations of electrical cable length and downhole temperature. The well was subject to several ESP failures due to solids production. The problem was handled by the installation of a downhole sand filter system.

It should be highlighted that real-time data from ESP downhole sensors of Well-X was captured over 1-minute intervals. The model was used to work through 160 K points (~100 days) of data to predict the flow rate at 1-minute intervals. The range of the data for each parameter is presented in Table 6. Constant water-cut value (50%) was considered during the evaluation period which was extended for 100 days. The well produces light oil with an average oil gravity of 37.6 °API and oil viscosity of 2.2 cp. The gas/oil ratio is equal to 25 scf/STB and the flowline pressure is about 800 psi, which is high due to network restrictions. Consequently, the quantity of produced gas was negligible.

Parameter	Unit	Min	Max	Mean	Remarks
Wellhead pressure	psig	997	1,002	1,000	Real time
Wellhead temperature	°F	111	111	111	Real time
Separator pressure	psig	827	832	830	Real time
Oil gravity	API	37.6	37.6	37.6	Nonreal-time
Water cut	%	50	50	50	Real time
Gas gravity	AIR = 1	0.85	0.85	0.85	Nonreal-time
Gas/oil ratio	scf/STB	25	25	25	Nonreal-time
Oil viscosity	cp	2.187	2.187	2.187	Nonreal-time
Intake pressure	psi	434	1,422	1,022	Real time
Intake temperature	°F	242	4,824	246	Real time
Motor temperature	°F	246	4,822	288	Real time
Drive frequency	Hz	38	43	43	Real time
Average amps	A	41	58	47	Real time
Discharge pressure	psi	1,018	5,380	4,788	Real time
VSD current	A	452	641	528	Real time
Output voltage	V	255	292	287	Real time

Table 6—Field case Well-X input parameters.

The developed ANN model (Configuration 1) was run using the data of Well-X to predict the low rate curves. The predicted liquid, oil, and water rates were compared to actual separator test data. The results of the comparisons are demonstrated in Fig. 7. It is clear that the model results match the actual test data. The results of the model simulate the actual fluctuations in production rates (e.g., the liquid rate fluctuations between 800 and 1,100 BFPD).

The results of the field application indicate that the model has good potential for predicting individual well production rates. This should allow for continuous monitoring, optimization, and performance analysis.



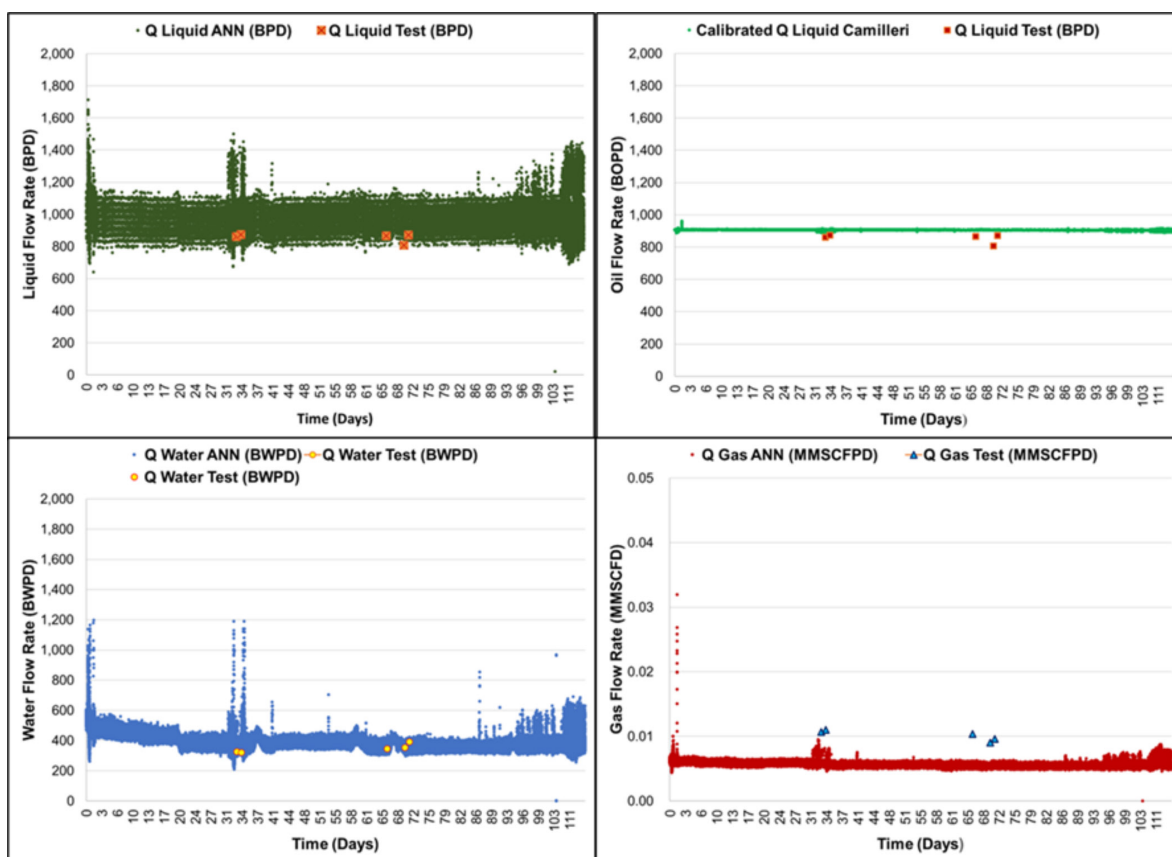


Fig. 7—Comparison between the results of the ANN model (Configuration 1) and the actual test data of Well-X.

## Conclusions

- ANN models were developed to predict the flow rates of ESP-lifted wells using actual field data from the Western Desert of Egypt. Each model consists of four neural networks with one output neuron each to predict oil, water, gas, and liquid flow rates separately. Feature engineering and sensitivity analysis helped to determine the optimum inputs.
- Five different configurations are proposed according to the number of input parameters in each configuration. The best performance is achieved with the configuration of 16 input parameters (Configuration 1) that include wellhead parameters, fluid properties, ESP downhole sensors readings, and VSD sensors parameters. The results of this configuration indicate that the MAPE between the predicted flow rates and the actual measurements for the testing data points (5,521 data points) of the oil, water, gas, and liquid networks is about 3.7, 5.2, 6.4, and 4.1%; respectively. In addition, the correlation coefficients ( $R^2$ ) are 0.991, 0.992, 0.983, and 0.979 for the estimated oil, water, gas, and liquid flow rates using the testing data points; respectively.
- The performance of the developed ANN model was compared against the performance of physics-based models. The ANN model almost matched the performance of Camilleri's method with MAPE of 5.2 and 5.5%, respectively. It was a challenging task to apply Caicedo's approach to real-time data. The ANN model has a major advantage as it does not require periodic calibration. The MAPE between the measured and calculated liquid flow rates for the ANN model, Camilleri approach, and Caicedo method for 130 data points are 7.5, 5.5%, and 16%; respectively.
- The developed ANN model was used to predict the flow rate curves of Well-X (located in the Western Desert of Egypt). The results of the model matched the actual separator test data.
- The model has good potential for predicting individual well production rates under a wide variety of pumping conditions and completion configurations. This should allow for continuous monitoring, optimization, and performance analysis of ESP wells as well as quick response to operational issues.

## Acknowledgments

The authors thank the management team of Agiba Petroleum Co. for permission to publish this article. Special thanks to the Petroleum Engineering team for their support during the data gathering stage.

## References

- Abdalla, R., El Ela, M. A., and El-Banbi, A. 2020. Identification of Downhole Conditions in Sucker Rod Pumped Wells Using Deep Neural Networks and Genetic Algorithms (Includes Associated Discussion). *SPE Prod & Oper* 35 (2): 435–447. SPE-200494-PA. <https://doi.org/10.2118/200494-PA>.
- Alhanati, F. J. S., Solanki, S. C., and Zahacy, T. A. 2001. ESP Failures: Can We Talk the Same Language? Paper presented at the SPE Gulf Coast Section Electric Submersible Pump Workshop, Houston, Texas, USA, 25–27 April. SPE-148333-MS. <https://doi.org/10.2118/148333-MS>.
- Al-Jasmi, A., Goel, H. K., Nasr, H. et al. 2013. A Surveillance “Smart Flow” for Intelligent Digital Production Operations. Paper presented at the SPE Digital Energy Conference, The Woodlands, Texas, USA, 5–7 March. SPE-163697-MS. <https://doi.org/10.2118/163697-MS>.

