26th International Congress of Mechanical Engineering
COBEM 2021
ABCM Associação Brasileira de Engenharia e Ciências Mecânicas
November 22-26, 2021, Virtual Congress, Brazil

COBEM2021-1618
WATER CUT ESTIMATION IN ELECTRICAL SUBMERSIBLE PUMPS USING ARTIFICIAL NEURAL NETWORKS

Matheus Paris Orsi
Alberto Luiz Serpa
Jorge Luiz Biazussi
University of Campinas, Campinas, Brazil
Rua Mendeleyev, 200 - Cidade Universitária, Campinas-SP, Brasil, CEP 13083-860
matheusorsi@gmail.com
alserpa@unicamp.br
biazussi@unicamp.br

Abstract. *An artificial lift is a method used to obtain a higher oil flow rate from the well, through some scheme that reduces the pressure at the bottomhole. Electrical submersible pumping is a common method in petroleum industry. The main component of this method is the electrical submersible pump (ESP), that can operate with complex flows involving mixtures of oil, water and gas. The presence of water in oil fields is a problem because it favors the formation of emulsions, which are the mixture of oil and water. Emulsions can be found in the form of oil-in-water and water-in-oil emulsions, depending on which phase is the continuous one and which is the dispersed one. Water-in-oil emulsions increase considerably the viscosity of the mixture and affect the pump's efficiency, diminishing its pumping capacity. The increase or decrease of the water fraction in the process may cause the phenomenon called catastrophic phase inversion (CPI), in which the dispersed phase becomes the continuous one and rapidly alters the physical properties of the flow, causing operational instability throughout the production system. In order to identify and predict this important phenomenon in complex multiphase flows, the usage of advanced identification tools, based on experimental data, has been used in recent years. In this work, artificial neural networks are used to estimate the water fraction in a flow that runs through an ESP. For that, data like inlet and outlet pressures, temperature, vibration and the correspondent water cut values, among others, were collected from an ESP operating with water and oil. Single-phase and two-phase tests were performed with the purpose of collecting data with different water cut values, ranging from 0% (single-phase oil) to 100% (two-phase water and oil). From the laboratory experiments, it was possible to build a data-driven computational tool capable of estimating the water fraction that runs through the pump, based on an optimized artificial neural network structure, which achieved an R-score of 0.9987.*

Keywords: *machine learning, electrical submersible pump, liquid-liquid flow, artificial neural networks, system identification.*

1. INTRODUCTION

In the petroleum industry, ensuring unconventional oil reserve production is a challenging task. The fluids are highly viscous and heavy, which brings high losses on flowing pressures along the production lines. When drilled, some wells (called flowing wells) contain enough inner pressure to raise the fluids to the surface, however, it decreases with time, to the point where the production can be interrupted. In most cases, the initial pressure of the well is already lower than the necessary to raise the fluids, or the flow is low to the point of not being economically viable. In these situations, to ensure that the fluids may reach the surface, artificial lift methods are applied. Among them, the electrical submersible pumping is widely used by oil companies because of its high productivity potential and versatility: electrical submersible pumps (ESP) are suitable for producing high to extremely high liquid volumes (reaching 30,000 barrels per day) and can be used in many applications, with a highlight to offshore operations because of the equipment's low space requirements (Takacs (2018)).

The most important component of the system, focus of this work, is the electrical submersible pump (ESP), which provides the fluid the energy needed so that the elevation to the surface may occur. It transforms kinetic energy from the pump's rotation into pressure energy. As it usually occurs in the industry, the equipment pumps an emulsion, which is a multiphase mixture of two immiscible fluids, water and oil. This mixture can be classified into a dispersion of droplets of oil in water or a dispersion of droplets of water into oil, depending on which phase is the continuous one. The estimation of the water fraction in the mixture is important to determine the flow's properties like effective viscosity, production flow rate and pressure drops. The formation of emulsions represents a problem because they increase the mixture viscosity and decreases the pump's performance considerably by lowering its lifting capacity and energy efficiency (Bulgarelli (2018)). The presence of water in oil fields also represents an environmental hazard, since the liquid produced is plenty of chemicals and must be treated before getting discarded (Gomes (2014)).

Predicting the flow characteristics that goes through the production lines is a challenging task since the offshore oil's exploration occurs in increasingly deeper waters, which increases the complexity of developing instrumentation that can withstand the pressure and temperature levels of this challenging environment. The costs of sending probes to perform the field's recognition or for maintenance purposes are extremely high. Thus, the creation of remote tools that enable the fluid's characterization becomes essential. Widely used nowadays, machine-learning algorithms are a great alternative to solve this problem (Shirley *et al.* (2012), Dasari *et al.* (2013), Dang *et al.* (2019)). They are data-driven algorithms that, with enough previous data, are able to predict values of variables of interest, e.g. the water fraction in the flow. The bigger the database is, the higher is the probability of the algorithm to achieve a good performance. Many researches about these methods are being made by the scientific community, though the identification using machine-learning techniques in the oil industry has not yet been too explored.

This work presents applications of machine-learning known as neural networks. They can recognize patterns in rough data, enabling its classification and grouping. Neural networks are versatile and are vastly applicable, from predicting house prices to providing medical diagnostics. Here, it is explored the functionality of identification, when the user wishes to estimate values in a data input of two or more variables. The algorithm is used to identify values of water fraction in a database produced experimentally through an ESP operating with a two-phase flow of oil and water.

2. METHODOLOGY

2.1 Experimental Facility

The test rig located at LABPETRO, from the "Center for Petroleum Studies" (CEPETRO - UNICAMP), is composed of an 8-stage ESP manufactured by Baker-Hughes, model P100L, booster pumps that provide water and oil flows, flow meters, pressure gauges, temperature sensors, water cut meter, among others. Figure 1 shows a complete scheme of the test rig.

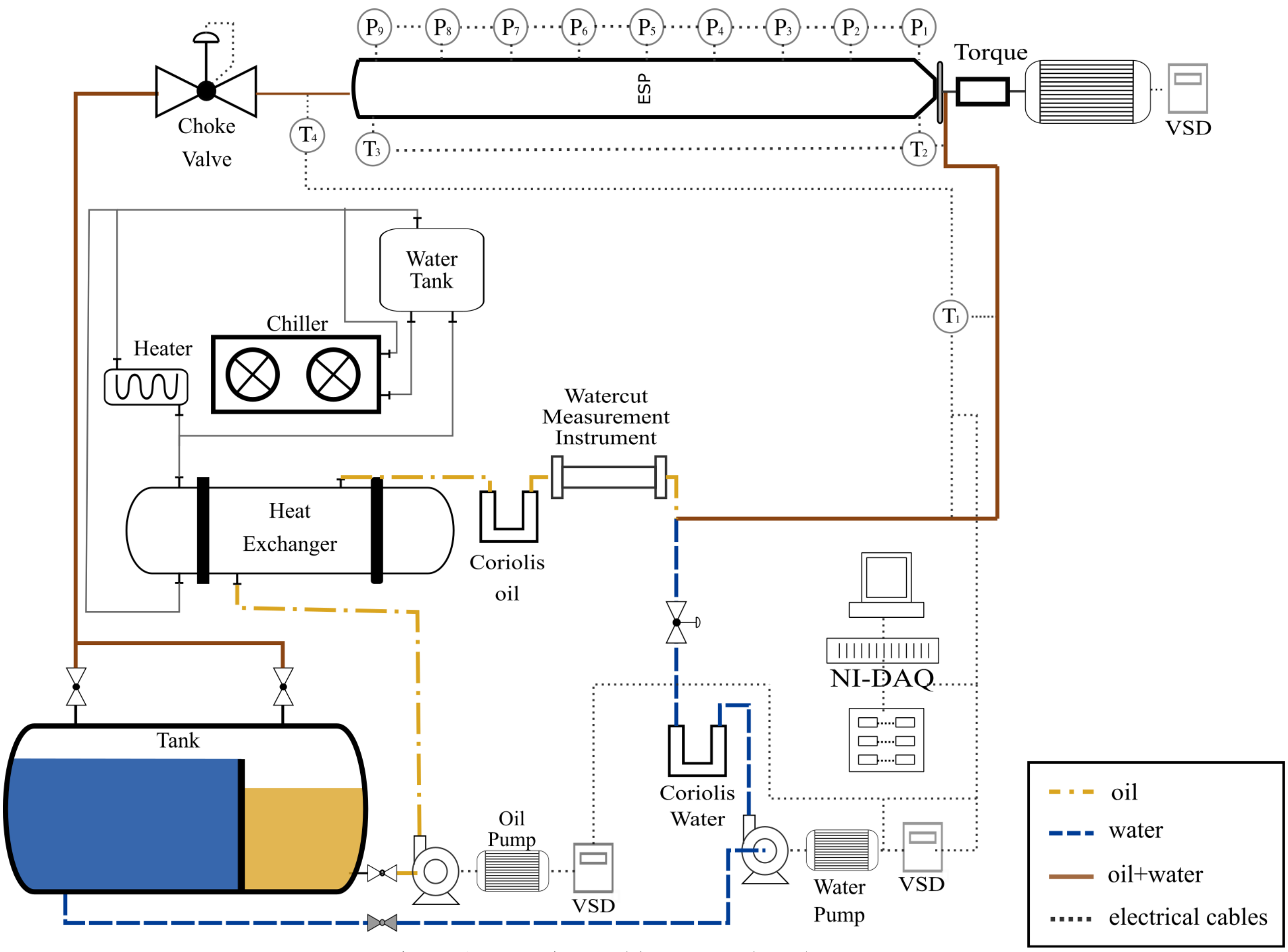

Figure 1: Experimental layout used on the tests.

The test lines begin at the separation tank. It is divided in two chambers, where the larger one corresponds to the outlet of the emulsion pipeline after the fluids go through the ESP. It is responsible for holding the fluids, performing a gravitational separation of the oil. When it separates from the water, it overcomes the barrier inside the tank and falls over the smaller chamber, where there is only oil with a small proportion of water mixed into it. Oil is inserted into the pipelines by a two-screw booster pump while water is pumped by a centrifugal pump. The fluids are mixed at a "T" intersection prior to the ESP inlet.

Single-phase and two-phase tests were performed on the pump. The first ones were designed to find the best efficiency point (BEP) for each of the rotations and temperatures analyzed. The temperature was varied so as to vary the oil's viscosity. Since water does not present a significant viscosity change with the temperature, for the single-phase tests with water only the ESP rotation was varied. Table 1 disposes the single-phase tests performed on the ESP.

Table 1: Single-phase tests planning.

Fluid	Rotation [RPM]	Temperature [°C]	Viscosity [cP]
Oil	1800	20	352
	2400	25	244
	3000	30	175
	3500	45	77
Water	1800	25	1
	2400		
	3000		
	3500		

After identifying the best efficiency points for each rotation and viscosity, the two-phase tests were designed. In this case, both rotation and viscosity were varied, but the total flow rate was fixed at points near the BEP, with a variation of $\pm 20\%$. The tests began with a single-phase flow of oil and water was introduced while the oil flow was reduced, keeping the total flow rate constant, until the flow was composed by water alone. Table 2 contains the tests performed with the mixture of oil and water. This procedure was important to analyze the ESP behavior with the full range of the variable of interest of this project, from nearly 0% to 100% of water cut. Also, the phase inversion phenomenon and its consequences to the system were observed.

Table 2: Two-phase tests design.

Viscosity [cP]	ESP Rotation [RPM]	Efficiency
352	1800	80% BEP
244	2400	100% BEP
175	3000	
77	3500	120% BEP

The water cut, which is the target variable to be determined in this work, was calculated by Equation 1:

$$WC[\%] = \frac{Q_{water}}{Q_{water} + Q_{oil}} \times 100 \qquad (1)$$

where Q_{water} is the water's flow rate (m^3/h), Q_{oil} is the oil's flow rate (m^3/h), and WC is the water cut (%).

2.2 Artificial Neural Network (ANN) Algorithm

For the identification of the water fraction at each experimental point, an ANN algorithm was built using *Tensorflow* (Abadi *et al.* (2015)) and *Keras* (Chollet *et al.* (2015)), which are open source libraries available for *Python*. These are considered to be the most used tools for machine-learning applications and are widely studied by the scientific community nowadays. The neural network is composed of dense layers (i. e., all of one layer's neurons are connected to all of the next layer's neurons) with backpropagation, which speeds up the learning process. Figure 2 shows the example of the dense neural network's architecture used in this work. In the next section, we explain how this architecture was found. Each node represents a neuron that transmits its information to the next layer's neurons, according to specific weights, represented by the arrows, and activation functions. The input layer receives the input data, the hidden layers are known as a “black box” and the output layer contains the output values predicted by the network (in our case, the water cut). A neural network tries to minimize the error between the real and predicted values on the output of the algorithm by updating the weights at each cycle of training (or *epoch*).

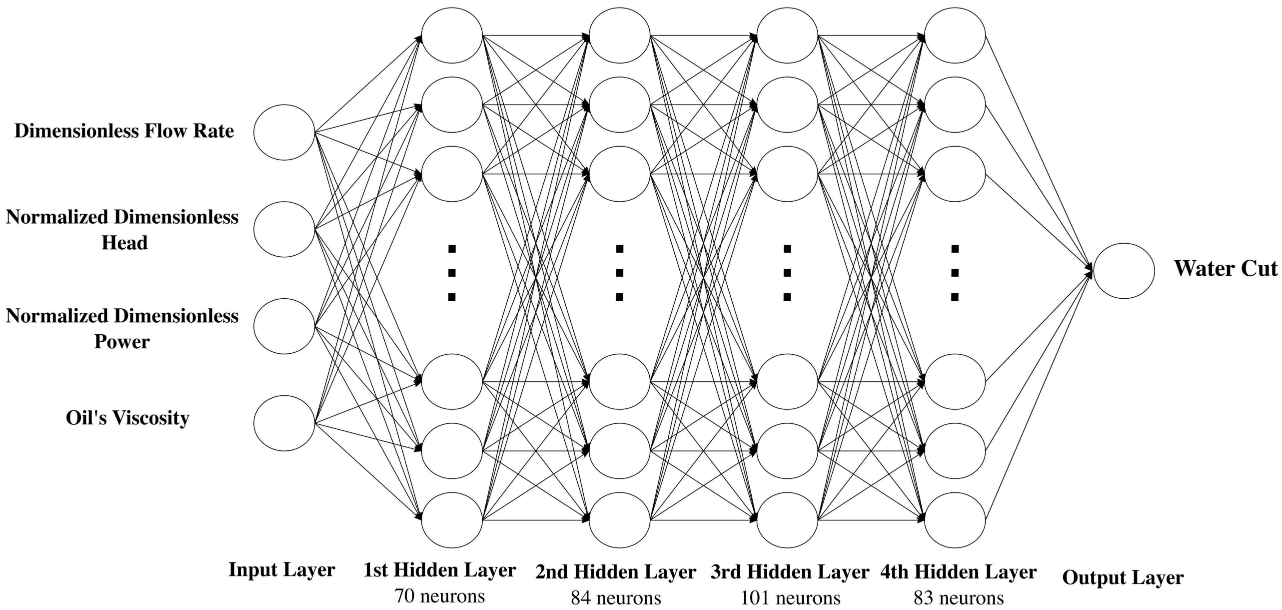

Figure 2: Example of a neural network architecture.

The downside of using ANNs is that there is not a rule that defines the best architecture for a model. The process of choosing the architecture is made through trial and error efforts or by applying an optimization of hyperparameters algorithm, such as the number of hidden layers, number of neurons per layer and the number of epochs of training.

With the conclusion of experimental tests, some of the variables collected from the ESP on two-phase tests were used to train a neural network's algorithm to estimate the water fraction that went through the pump. The variables selected were the oil's viscosity and dimensionless numbers, because they reduce the amount of variables needed to describe the problem and can be used to transfer the model results to data collected in oil fields. In total, 16 curves with 718 samples were collected on the two-phase tests.

The dimensionless numbers used to train the algorithm are described next:

- Dimensionless Total Flow Rate:

$$\Phi = \frac{q}{\omega D^3} \tag{2}$$

where q is the total flow rate (m^3/s), ω is the pump's rotation (rad/s) and D is the pump's diameter (m). For this pump, $D = 0.108m$.

- Normalized Dimensionless Head:

$$\Psi_{norm} = \frac{\Psi_{oil}}{\Psi_{water}} \tag{3}$$

where Ψ_{water} is the dimensionless head with water as the working fluid, calculated by Equation 4 (Biazussi (2014)),

$$\Psi_{water} = \frac{1}{4} - k_4 + (-k_1 - Xk_2 + 2k_4k_5)\Phi + \left[-\left(\frac{X}{\Phi}\right)^n k_3 - k_4k_5^2 - k_6\right]\Phi^2 \tag{4}$$

The coefficients k_1 to k_6 and n are geometric parameters of the pump. They were obtained by Bulgarelli *et al.* (2021) and are presented in Table 3. X is the reverse rotational Reynolds, given by Equation 5,

$$Re_w = \frac{\rho\omega D^2}{\mu} = \frac{1}{X} \tag{5}$$

where μ is the fluid's viscosity ($Pa \cdot s$) and ρ its specific mass (kg/m^3).

Table 3: Pump's geometric coefficients (Bulgarelli *et al.* (2021)).

Parameter	**Value**
k_1	1.50
k_2	6024.98
k_3	264.96
k_4	0.14
k_5	7.25
k_6	28.22
n	0.45

Ψ_{oil} is the fluid's dimensionless head, given by Equation 6,

$$\Psi_{oil} = \frac{\Delta P}{\rho_{oil}\omega^2 D^5} \tag{6}$$

where ΔP is the pressure differential between the pump's inlet and outlet (Pa).

- Normalized Dimensionless Power:

$$\Pi_{norm} = \frac{\Pi_{oil}}{\Pi_{water}} \tag{7}$$

where Π_{water} is the dimensionless power with water as the working fluid, calculated by Equation 8 (Biazussi (2014)). The constants were fitted with data obtained by Bulgarelli (2018) and are presented in Table 4.

$$\Pi_{water} = \Psi_{oil} z_0 + z_1 + \Phi_{oil} z_3 + \Phi_{oil}^2 z_5 + \Phi_{oil}^3 z_7 \tag{8}$$

Table 4: Coefficients of the dimensionless power curve.

Parameter	**Value**
z_0	0.11392
z_1	0.000754
z_3	0.100872
z_5	-0.87167
z_7	2.82834

Π_{oil} is the dimensionless ESP power with the mixture of oil and water, given by Equation 9, where BHP is the ESP shaft power.

$$\Pi_{oil} = \frac{BHP}{\rho_{oil}\omega^3 D^5} \tag{9}$$

Table 5 summarizes the inputs selected for analysis. The water cut was used to compare the real values and the ones predicted by the algorithm.

Table 5: Variables selected for training the ANN algorithm.

Inputs	**Units**	**Output**
Oil's Viscosity (μ_{oil})	[cP]	Water cut [%]
Normalized Dimensionless Power (Π_{norm})	[-]	
Dimensionless Total Flow Rate (Φ)	[-]	
Normalized Dimensionless Head (Ψ_{norm})	[-]	

3. RESULTS AND DISCUSSION

Before applying the data collected to the algorithm, they had to be scaled. This step is known as feature scaling and it is necessary because some variables vary in a range which is much larger than the others. For example, in our data set there are variations of the ESP rotation in a range from 1800RPM to 3500RPM, while the water cut varies from 0% to 100%. If the data is not scaled, the algorithm does not work efficiently because it considers that variables that present larger variation influences the phenomenon in a stronger way, which is not always correct. With that, the data must be transformed into a more homogeneous set, with means and standard deviations that truly represent the influence of each input to the outcome. The feature scaling technique used in this work was the standardization, given by Equation 10. The standardization transforms the data into a distribution centered around zero with a unitary standard deviation,

$$X_{new} = \frac{X - \mu}{\sigma} \tag{10}$$

where X_{new} is the value of the variable after scaling, X is the variable before scaling, μ is the mean of the feature data and σ its standard deviation.

We divided the full data into training and test data sets, selected randomly. The training dataset contained 80% of the shuffled data, with 574 samples, while the test dataset contained the remaining 20% of the data with 144 samples.

With the standardized and divided data, it was possible to perform tests to determine the best ANN architecture for the purpose of this work. This process was made by applying an optimization of hyperparameters algorithm using the *Optuna* package, also available for *Python* (Akiba *et al.* (2019)). This package allows the user to determine which hyperparameters are to be optimized and the number of trials that the algorithm should perform in order to achieve the best value for a given metrics. In this work, we use the R-squared, given by Equation 11,

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \tag{11}$$

where N is the number of samples in the data, y_i is the individual predicted value, $\hat{y}_i$ is the individual actual value and $\bar{y}$ is the mean of the data. The R^2 metric represents the accuracy of a model while fitting the data into a line or space. Values close to $R^2 = 1$ represent a fine model, while values close to $R^2 = 0$ are only achieved by poor models.

We programmed the optimization algorithm to select the best architecture within the range of 4 to 6 hidden layers and 64 to 128 neurons per layer, using the Adam optimizer's default learning rate and ReLU neurons. The function of the optimizer is to perform the network's training. The rectified linear unit activation function or ReLU for short is a piece-wise linear function that will output the input directly if it is positive, otherwise, it will output zero, according to Equation 12.

$$f(x) = max(0, x) \tag{12}$$

After nearly 2000 trials performed by the *Optuna* algorithm, the result was an architecture with 4 hidden layers with different numbers of neurons per layer. The optimized architecture is displayed on Table 6.

Table 6: Optimized architecture used to produce the results.

Hidden layer	**Neurons per layer**	**Learning rate**
1	70	
2	84	0.001
3	101	
4	83	

Using the optimized architecture, we tested the algorithm for 2000 epochs. The objective is to reduce the error after each epoch with the update on the weights. The resulting model created by the network was used to predict the water cut on all of the samples acquired from the laboratory to check if the model is accurate. Figure 3 displays the results after the training. It consists on the crossing between real and predicted values. The green line on the figure represents the ideal values with 0% error, while the yellow and red lines represent error of 5% and 10% respectively.

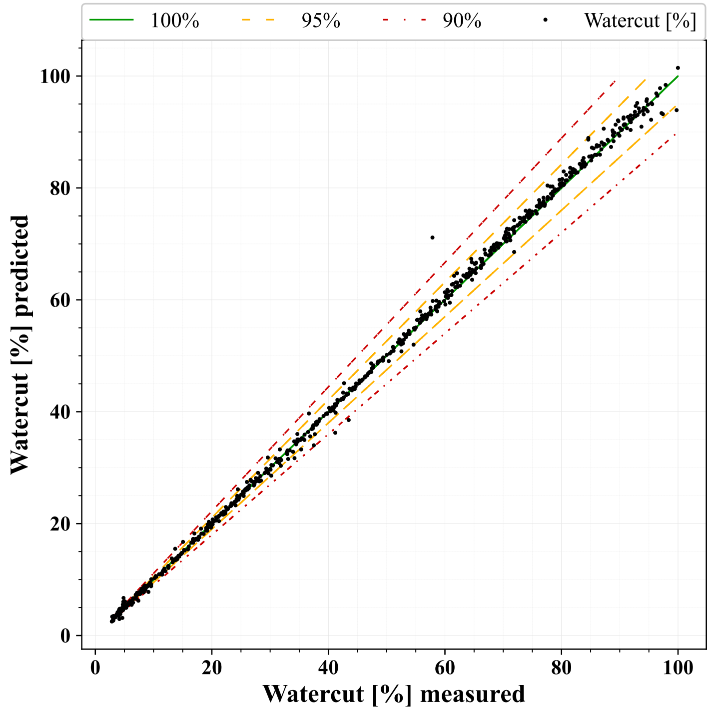

Figure 3: Results obtained after 2000 epochs of training.

For a better visualization of the results, we have generated Figures 4 and 5. It contains two curves obtained by the algorithm, where the first one shows the best one and the second displays the worst curve where the algorithm failed to predict some values after the inversion point. On the figures, the black line represents the curve of differential pressure per stage accordingly to the water cut obtained from the laboratory and the red dots corresponds to the ones produced by the network.

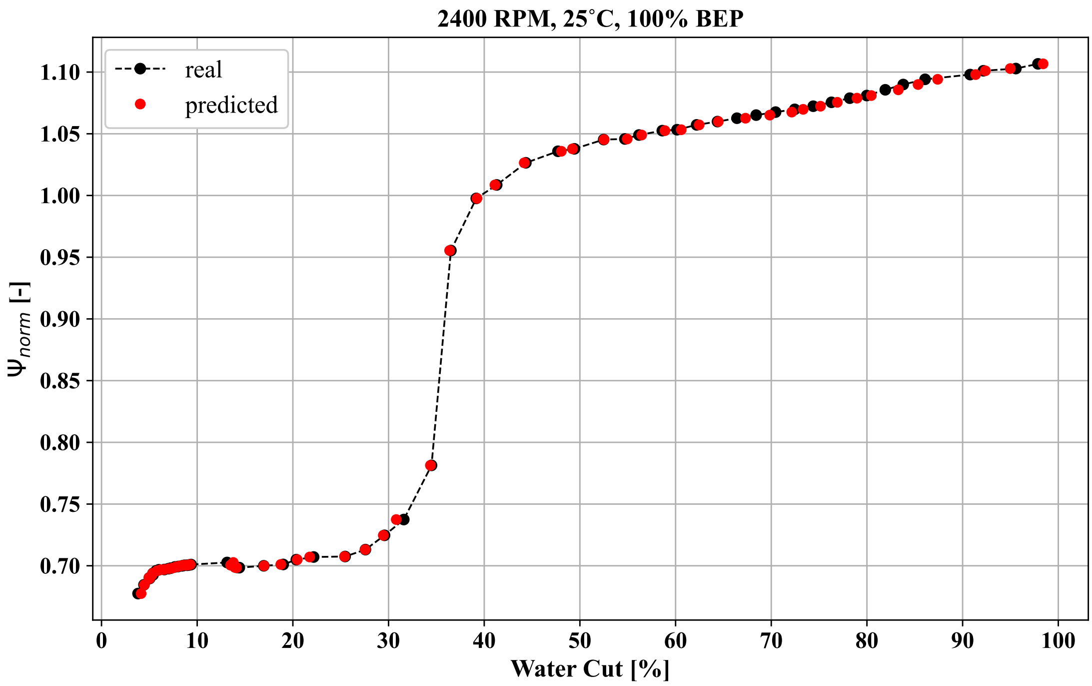

Figure 4: Best curve obtained by the algorithm.

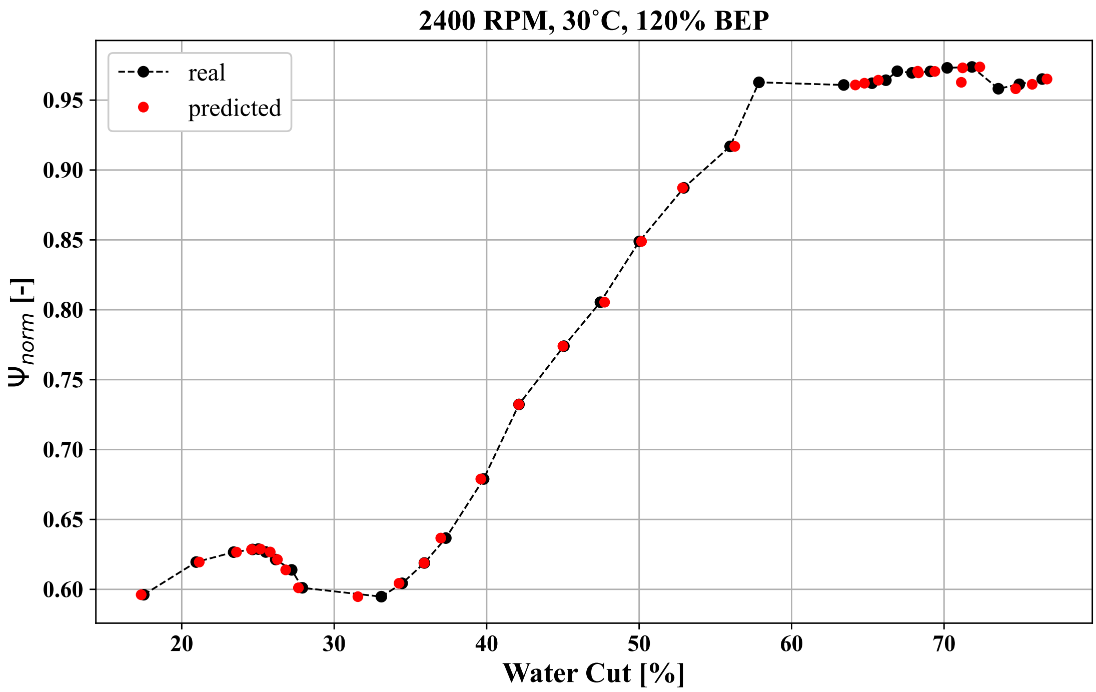

Figure 5: Curve with a few wrong predictions.

The algorithm produced an R-score of $R^2 = 0.9987$, indicating that the model works fine and achieves high levels of accuracy. It can seen that most of the water cut values predicted by the algorithm are very close to the ones obtained on laboratory. However, some point were predicted wrongly by the algorithm. This problem would probably be fixed with a larger dataset.

4. CONCLUSIONS

In this work, a neural network's algorithm was designed to estimate the water fraction that runs through an electrical submersible pump. Curves of the phase inversion phenomenon predicted by the algorithm were very similar to the ones obtained on laboratory. Some points were predicted wrongly by the network, but with errors considered to be low.

The architecture's optimization process resulted in a neural network configuration that would hardly be achieved through trial and error. This was an essential step to the model creation, since an automated search for the best architecture improved greatly the algorithm's efficiency.

5. ACKNOWLEDGEMENTS

We gratefully acknowledge the support of EPIC – Energy Production Innovation Center, hosted by the University of Campinas (UNICAMP) and sponsored by Equinor Brazil and FAPESP – São Paulo Research Foundation (2017/15736-3 and 2019/08446-4). We acknowledge the support of ANP (Brazil's National Oil, Natural Gas and Biofuels Agency) through the R&D levy regulation. Acknowledgements are extended to the Center for Petroleum Studies (CEPETRO), School of Mechanical Engineering (FEM), and ALFA Research Group.

We also acknowledge Mauricio Barrios Castellanos, Felipe de Castro Teixeira Carvalho and Natache do Socorro Dias Arrifano Sassim for their support to this work.

6. REFERENCES

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y. and Zheng, X., 2015. "TensorFlow: Large-scale machine learning on heterogeneous systems". URL `https://www.tensorflow.org/`. Software available from tensorflow.org.

Akiba, T., Sano, S., Yanase, T., Ohta, T. and Koyama, M., 2019. "Optuna: A next-generation hyperparameter optimization framework". In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*. pp. 2623–2631.

Biazussi, J.L., 2014. *Modelo de Deslizamento para Escoamento Gás-Liquido em Bomba Centrífuga Submersa Operando com Líquido de Baixa Viscosidade*. Ph.D. thesis, University of Campinas - UNICAMP.

Bulgarelli, N.A.V., 2018. *Experimental Study of Electrical Submersible Pump (ESP) Operating with Water/Oil Emulsion*. Master's thesis, University of Campinas - UNICAMP.

Bulgarelli, N.A.V., Biazussi, J.L., Verde, W.M., Perles, C.E., de Castro, M.S. and Bannwart, A.C., 2021. "Experimental investigation on the performance of electrical submersible pump (esp) operating with unstable water/oil emulsions". *Journal of Petroleum Science and Engineering*, Vol. 197, p. 107900.

Chollet, F. *et al.*, 2015. "Keras". `https://keras.io`.

Dang, W., Gao, Z., Hou, L., Lv, D., Qiu, S. and Chen, G., 2019. "A novel deep learning framework for industrial multiphase flow characterization". *IEEE Transactions on Industrial Informatics*, Vol. 15, No. 11, pp. 5954–5962.

Dasari, A., Desamala, A.B., Dasmahapatra, A.K. and Mandal, T.K., 2013. "Experimental studies and probabilistic neural network prediction on flow pattern of viscous oil–water flow through a circular horizontal pipe". *Industrial & Engineering Chemistry Research*, Vol. 52, No. 23, pp. 7975–7985.

Gomes, A.P., 2014. *Gestão Ambiental da Água Produzida na Indústria de Petróleo: Melhores Práticas e Experiências Internacionais*. Master's thesis, Federal University of Rio de Janeiro.

Shirley, R., Chakrabarti, D.P. and Das, G., 2012. "Artificial neural networks in liquid-liquid two-phase flow". *Chemical Engineering Communications*, Vol. 199, No. 12, pp. 1520–1542.

Takacs, G., 2018. *Electrical submersible pumps manual: design, operations, and maintenance*. Gulf professional publishing, 2nd edition.

7. RESPONSIBILITY NOTICE

The authors are solely responsible for the printed material included in this paper.