Using NARX model with wavelet network to inferring the polished rod position

Emanuel Cajueiro, Ricardo Kalid, and Leizer Schnitman

***Abstract*—** Although several studies have been conducted on the sucker-rod pumping system, even today, the acquisition of the polished rod position is carried out by using position transducers. In this paper, we present experimental results showing that the dynamic position of the polished rod can be inferred from the torque current of a three-phase induction motor (which is given by sensorless vector AC-drive) using nonlinear autoregressive model with exogenous input (NARX) with wavelet network (wavenet). The results obtained in the validation stage show that, on the basis of the experimental set used in this work, the best estimated model is suitable to represent the dynamic behavior of the polished-rod of the pumping unit.

***Keywords*—** NARX model, polished rod position, sucker-rod pumping, wavelet network, wavelet frames.

I. Introduction

SUCKER-ROD pumping is the most widely-used artificial lift method for onshore oil wells all over the world [1]. This system transforms a rotary motion provided by prime mover in a reciprocating motion that is transferred to the downhole pump by a set of rods, in the form of a vertical displacement, which provides mechanical energy to the fluid by allowing its elevation to the surface [2]. The dynamic position of the polished rod is an aid in carrying out a diagnosis of onshore wells. These diagnoses are generally conducted through an interpretation of the pump dynamometer card, which is a plot of the calculated loads at various positions of pump stroke, and represents the load the pump applies to the bottom of the rod string [3]. It is usually calculated from the surface dynamometer card, which is the plot of the measured or predicted rod loads at the various positions throughout a complete stroke [2].

Since the 1960s a number of studies have been undertaken to build models that are computer-aided and able to reproduce the dynamic behavior of sucker-rod pumping. In [3] the rod string is represented by a linear model, by means of the one-dimensional damped wave equation. The boundary conditions are given for the polished rod and the pump. Thus, with the aid of partial differential equations these systems were solved using a finite differences method. On the basis of a concept of [3] more efficient models were devised that take account of the dynamics of the fluid and rod [4], [5], [6]. Another useful technique was outlined in [7] with the aim of analyzing the dynamic behavior of the sucker rod string. However, due to the highly nonlinear nature of the relationship between the oil well and sucker-rod pump, a considerable effort is required to apply these studies for prediction or on-field simulation in a real system , and as a result,, even today, the acquisition of the polished rod position is carried out by using transducers.

According to [8] and [9], it is possible to model a dynamic system directly from measured input and output data , by employing a structured model that does not use any priori information about the behavior of the system. This type of structure is known as a black-box model. According to [10], there are two main reasons for obtaining a model from a dynamic system using the input-output data in a black box approach: "1) Even the best theoretical model built from equations of motion is still only an approximation of reality; 2) Due to situations where the theoretical model is more complicated, or the physics of the process is poorly understood, the only reliable information is experimental data".

It is well known that artificial neural networks (ANN) constitute a powerful tool in nonparametric estimates due to their ability to be used as an arbitrary function approximation mechanism that 'learns' from observed data [11], [12], [13]. In recent years, a special class of artificial neural networks, known as the wavelet networks (also called wavenet), have aroused considerable interest [14], [15], [16]. This new class of ANN, was first studied by [17] and improved by [18], as an alternative to the classical feedforward neural network (FNN), for approximating arbitrary nonlinear functions. Basically, the wavenet is a wavelet decomposition implemented in a one-hidden-layer neural network, where the wavelets are activation functions of hidden neurons [18]. One of the main advantages of wavelet network compared with ANN is that whereas the wavelet theory provides useful guidelines for the construction and initialization of the networks, ANN use random data [18], [19].

Basically, wavelet networks can be built by adopting two different approaches, which are as follows: a multiresolution approach [20] and the single-scaling radial wavelet frames [21]. In the first approach, the wavelets used in the expansion form an orthogonal basis. In the second, the expansion is built by using nonorthogonal wavelets. In this paper, we consider the wavelet network based on the single-scaling radial wavelet

Manuscript received September 23, 2011.

E. B. A. C thanks the CAPES for financial support.

E. B. A. C is with the Postgraduate Program in Industrial Engineering, Federal University of Bahia, Salvador, Brazil (emanuelbenicio@ufba.br).

R. A. K, is with the Postgraduate Program in Industrial Engineering, Federal University of Bahia, Salvador, Brazil (kalid@ufba.br).

L. S. is with the Postgraduate Program in Industrial Engineering, Federal University of Bahia, Salvador, Brazil (leizer@ufba.br).

frames. This choice simplifies the use of wavelet networks in high dimensional systems (see [21], Section III A, for more details).

In this paper, we use the wavelet network as a nonlinearity estimator of a nonlinear autoregressive model with exogenous input (NARX) [22] to infer the position of the polished rod from torque current of the three-phase AC induction motors. The main advantage of using torque current is that this information is given by sensorless vector AC-drive, which is used in most onshore wells for starting and speed control of AC induction motors.

II. NARX Model

We set out from the linear autoregressive model with exogenous input (ARX) [23], which is given by the following equation:

$$y(k) + a_1 y(k\text{-}1) + \cdots + a_{na} y(k\text{-}na) = b_1 u(k\text{-}nk) + \cdots + b_{nb}\, u(k\text{-}nk\text{-}nb+1) \quad (1)$$

where, y and u are, respectively, the output and input data, k is the discrete time instant; n_a is the number of autoregressors, n_b is the number of exogenous regressors, and n_k are the delay from input to output in terms of the number of samples. Isolating $y(k)$ in (1), it is possible determine the next output, through a weighted sum of its previous observation, called regressors

$$\hat{y}(k) = -a_1 y(k\text{-}1) - \cdots - a_{na} y(k\text{-}na) + b_1 u(k\text{-}nk) + \cdots + b_{n_b} u(k\text{-}nk\text{-}nb+1) \quad (2)$$

We can then rewrite the equation (2) using the compact notation

$$\hat{y}(k) = x^T(k)\theta \quad (3)$$

where, $x^T(k)$ is the vector of regressors given in (4), and θ is the vector coefficients given in (5)

$$x(k) = [y(k\text{-}1), \ldots, y(k\text{-}na), u(k\text{-}nk), \ldots, u(k\text{-}nk\text{-}nb+1)]^T \quad (4)$$

$$\theta = \left[-a_1, \ldots, -a_{na}, b_1, \ldots, b_{nb}\right]^T \quad (5)$$

Equation (4) represents the memory of the model, called vector of regressors.

Based on the linear ARX model, the nonlinear ARX (NARX) model use nonlinear mapping F between the input and output data, rather than a weighted sum that represents linear mapping, such as ARX.

$$\hat{y}(k) = F(x(k)) \quad (6)$$

where, $x(k)$ is the vector of regressors given in (4), and F is a nonlinear function. In this paper, F is a wavelet network, which is composed of radial wavelet frames [24].

In order to capture the linearity in the sampled data set more effectively, it is possible to rewrite (6) as follows,

$$\hat{y}(k) = F(x(k)) + x(k)\theta^T + l \quad (7)$$

where, θ represents the linear coefficients, and l is the output offset. In other words, the NARX structure considered in this work is a series-parallel combination, as shown Fig. 1.

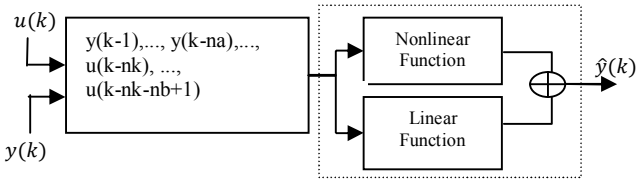


Fig. 1 NARX model.

The capacity of linear function approximation, block $(x(k)\theta^T + l)$, is discussed in full detail in [25].

III. Wavelet

In this section, we briefly outline some of the basic concepts of wavelet, which will allow an understanding of the wavelet network in Section IV.

A. Wavelet and Radial Wavelet

Wavelets ("little waves") are functions located both in time and frequency. These functions allows decompose a finite energy signal into different frequency components by superposition of functions obtained after scaling and translating an initial function over test function ψ, known as Mother Wavelet [26].

Wavelet Analysis is different from other techniques in that it analyses frequency components with a resolution that matches their scale. This attractive theory has been successfully employed in many applications, including artificial neural network, communication, numerical analysis (see, for example, [21], [27] and [28]), among others.

For a function to be classified as a wavelet, it must meet the following requirements:

1) The average value of the wavelet in the time domain should be zero

$$\int_{-\infty}^{+\infty} \psi(t)dt = 0 \quad (8)$$

2) The function must have finite energy

$$\int_{-\infty}^{+\infty} |\psi(t)|^2 dt = 1 \quad (9)$$

3) The function must be admissible as a wavelet

$$C_\psi = 2\pi \int_{-\infty}^{+\infty} \frac{\left|\hat{\psi}(\omega)\right|^2}{|\omega|} d\omega \quad (10)$$

where $\hat{\psi}$ is the Fourier transform of the wavelet function ψ, and $\psi \in L^2(\mathbb{R})$. It is worth noting that the inverse wavelet transform only exists for $0 < C < \infty$. This means that the analyzed signal can be reconstructed without loss of information [26].

To obtain a wavelet in $L^2(\mathbb{R}^d)$, as shown in [24], the

symmetrical radial wavelet (11) will be considered

$$\{\hat{\psi}(\omega) = \hat{\phi}(\|\omega\|) : \omega \in \mathbb{R}^d\} \quad (11)$$

where $\hat{\psi}(\omega)$ is the Fourier transform of $\psi(x)$, with $x \in \mathbb{R}^d$, and $\|x\| = \sqrt{x^T x}$ is the Euclidian norm, and it is necessary for the radial wavelet function (11) to satisfy the admissibility condition

$$C_\psi = (2\pi)^d \int_0^\infty \frac{\left|\hat{\psi}(h\omega)\right|^2}{h} dh < \infty \quad (12)$$

The reader should refer to [24] and [26] for a detailed account of these requirements.

B. The Continuous Wavelet Transform

Any finite energy signal $f \in L^2(\mathbb{R}^d)$ can be decomposed and represented by wavelets functions ψ that satisfy the condition (12)

$$(T^{wav} f(x))(a,b) = w(a,b) = \int_{\mathbb{R}^d} f(x) \frac{1}{|\sqrt{a^d}|} \psi\left(\frac{x-b}{a}\right) dx \quad (13)$$

and reconstructed by its inverse

$$f(x) = \frac{1}{C_\psi} \int_0^\infty \int_{\mathbb{R}^d} \frac{1}{a^{d+1}} \frac{1}{|\sqrt{a^d}|} w(a,b) \psi\left(\frac{x-b}{a}\right) dbda \quad (14)$$

where, (13) is the *continuous wavelet transform* (CWT) and (14) is the *inverse wavelet continuous transform* (IWT); $a \in \mathbb{R}_+$ and $b \in \mathbb{R}^d$ are respectively, *dilation* and *translation* parameters. The term given in (15) that was taken from CWT in (13), is called mother wavelet function.

$$\psi(x) = \frac{1}{|\sqrt{a^d}|} \psi\left(\frac{x-b}{a}\right) \quad (15)$$

For more details about continuous wavelet transform, see [26].

C. Frame Condition

For an IWT to be implemented in a digital machine, it must be discretized. When the IWT is discretized, some conditions are required to ensure that this discrete version of the reconstruction of f can actually hold. One of the solutions that meet these requirements is called Frames which was originally developed by [29] and discussed in [26]. It is used to analyze the stability, integrity and redundancy of a discrete representation of a signal using wavelets. A frame is a set of "redundant bases" $\{\Psi(a,b);\ a > 1, b \epsilon \mathbb{Z}\}$ that satisfies the *frame condition*

$$A\|f\|^2 \le \sum_{m,n} \left|\langle f, \psi_{m,n}\rangle\right|^2 \le B\|f\|^2 \quad (16)$$

and thus allows a signal $f \in L^2(\mathbb{R}^d)$ to be described through a finite sum of nonorthogonal terms. Where $\|\cdot\|$ denotes the norm of the function, $\|f\|^2$ is the energy of the function f; $A > 0$ and $B < \infty$ denote, respectively, the lower and upper frame bounds; $\psi_{m,n}$ (with $m \in \mathbb{Z}$ and $n \in \mathbb{Z}^d$) is a radial wavelet frame function that satisfies the condition (5).

The wavelet frame theorem was first developed by [26] to ensure the construction of wavelets in $L^2(\mathbb{R})$ space. The extension to the $L^2(\mathbb{R}^d)$ space was successfully developed in [24] by adopting two approaches: the single-scaling and the multiscaling wavelet frames. In the case of the single-scaling, a single dilation parameter is used in all the dimensions of a wavelet, whereas in the multiscaling, an independent dilation parameter is used in each dimension. Thus the former is structurally simpler than the latter. We only consider single-scaling wavelet frames here.

D. Discrete Wavelet family and Inverse Discrete Wavelet Transform

Once the condition (16) is satisfied, to obtain the discrete version of (15), its dilation and translation parameters need to be replaced by sampled values. In this case, a is replaced by a_0^m and b is replaced by $nb_0a_0^m$, which are, respectively, the dilation and translation step

$$\{\psi_{m,n}(x) = a_0^{-dm/2} \psi(a_0^{-m}x - nb_0) : m \in \mathbb{Z}, n \in \mathbb{Z}^d\ \} \quad (17)$$

where, $a_0 > 1$ and $b_0 > 0$ are two scalar constants. Then, the inverse wavelet transform (14) can be discretized into

$$f(x) = \sum_{m,n} w_{m,n}\, \psi_{m,n}(x) \quad (18)$$

where, $w_{m,n}$ is the wavelet coefficient; $\psi_{m,n}$ satisfies the condition (12) and constitutes a radial wavelet frame of $L^2(\mathbb{R}^d)$.

IV. Wavelet Network and Its Structure

If the parameters of the inverse continuous wavelet transform (IWT) in (14) (w, a and b) are discretized according to the sample data (resulting in (17)), the adaptive discretization of the inverse wavelet transform given in (18), can be called a wavelet network. As mentioned earlier, this structure can be regarded as a neural network, whose wavelets are the activation function of hidden neurons.

In practical applications, it is unnecessary and impossible to construct a wavelet network using infinite radial wavelet frames, and hence, the decomposition (18) must be truncated into a finite set. If the truncation uses a regular wavelet lattice $\{a_0^{-m}x, nb_0\}$, obviously, we can expect that the truncation does not take into account the sparseness of the training data; as a result, many wavelets whose support does not contain any data point, may remain in the truncated frame [21]. Reference [21] seeks to overcome this problem by using an algorithm able to avoid "empty" wavelets which can improve the quality of the truncated wavelet frame for wavelet network construction.

In the following section, we summarize the basic procedures for building and training the wavelet network:

Considering a truncated family of (17)

$$\{\psi(a_0^{-m}x - nb_0) : m \in S_a, n \in S_t(n)\ \} \qquad (19)$$

where, $S_t \subset z^d$ is a finite set of n, and $S_a \subset \mathbb{Z}$ corresponding to a finite set of m to desired resolution level estimation, are both a finite subset of $\mathbb{Z}$.

1 For each sample data $ze = \{(u(1), y(1)), \dots, (u(k), y(k))\}$ of the finite set of dimension k, determine the wavelet in (19) whose supports contain the sample point, and build the set of wavelets candidates to be used on the wavelet network (library *W*).

2 Considering the wavelets present in *W* as regressors. To remove the redundant wavelets in *W* for the estimation *f*, use the heuristic algorithms given in [21], Section III, to select the best set of wavelets belonging to the library *W*.

3 Assuming that the most suitable set of wavelets $\Psi = [\psi_1, \dots, \psi_L]$ employed to represent the system output $Y = (y_1, \dots, y_L)^T$ is known. Then, basically, the problem consists of how to use the last square solution in $Y = \Psi\omega$ to determine the weight (ω) of the wavelet network.

A number of considerations are necessary to ensure that the above steps are effected:

- we consider the wavelet function as the second derivative of a Gaussian function (the "Mexican Hat")

$$\psi(x) = \left(\dim(x) - ||x||^2\right) e^{-\frac{||x||^2}{2}} \qquad (20)$$

Where, x is the vector of regressors (given in (4)) dilated, translated and well conditioned (applying principal component analysis techniques - PCA; see [22] for more details), and $||x||^2 = x^T x$. In this case, the Mexican Hat was chosen because it offers good time-frequency localization [26];

- we take $\{a_0^{-m}x, nb_0\}$ in the form of a dyadic grid ,i.e., a_0 = 2 and $b_0 = 1$;
- The choice of the number of wavelets functions used in the structure wavenet of this paper was performed by using the generalized cross-validation algorithm (GCV): $GCV(S) = \frac{1}{N}\sum_{k=1}^{n}(\hat{f}(u_k) - y_k)^2 + 2\,S\,\frac{1}{N}\sigma_e^2$. This algorithm selects the best functions that serve to minimize the estimation error. Where, GCV is minimized with respect to S. $\hat{f}$ is the approximation given by wavenet, *S* is the number of wavelets functions of the network, N is the sample length of the input vector u_k. y_k are input and output sampled data belonging to the estimation data set, and σ_e^2 is the noise variance estimated from u_k and y_k.

It should be remembered that in this paper the wavelet network is used to estimate the regression model given in (6). Thus, on the basis of the previous theory about the wavelet network construction, and in the NARX model (7), we can rewrite expansion (18) as follows:

$$f(x) = \sum_{k=1}^{ns} a_{sk}\varphi(D_{sk}(x - r)Q - t_{sk}) + \sum_{k=1}^{nw} \omega_{wk}\,\psi(D_{wk}(x - r)Q - t_{wk}) + x(k)P\theta^T + l \qquad (21)$$

where, the *scaling* function (22) is used in wavelet network to improve the regularity of the estimator. Scaling functions and wavelet functions are both used at different dilation scales, in a structure similar to the wavelet packet decomposition [21], [27], [28].

$$\varphi(x) = e^{\frac{-||x||^2}{2}} \qquad (22)$$

and $\psi(x)$ was defined in (20), Where r is the mean value of the regressor vector computed (1 x m), P is a matrix of the linear subspace (m x p), Q is a matrix of nonlinear subspace (m x q) obtained by principal component analysis applied to the regressors in x. θ is the vector of linear coefficients (p x 1), a_{sk} is the vector of coefficients associated with the scalons (s x 1), ω_{wk} is the vector of coefficients associated with the wavelons (nw x 1); D_{sn} is the dilatation matrix associated with scalons (s x 1), D_{wn} is the dilatation matrix associated with wavelons (nw × 1); t_{sn} is the translation matrix associated with scalons (ns x q), t_{wn} is the translation matrix associated with wavelons (nw x q), l is output offset (scalar), ns is the number of scalons and nw the number of wavelons.

The wavelet network architecture used in this work is shown in Fig. 2.

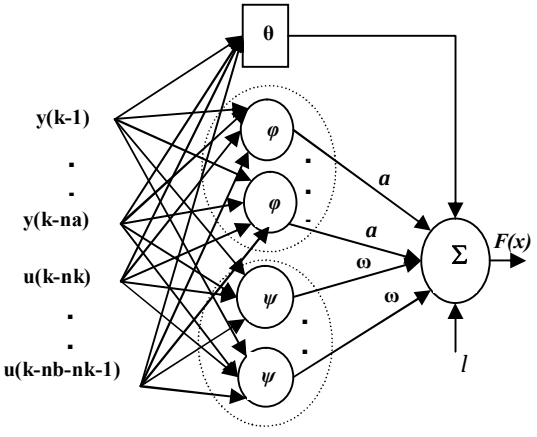


Fig. 2 Structure of Wavelet Network based on NARX model.

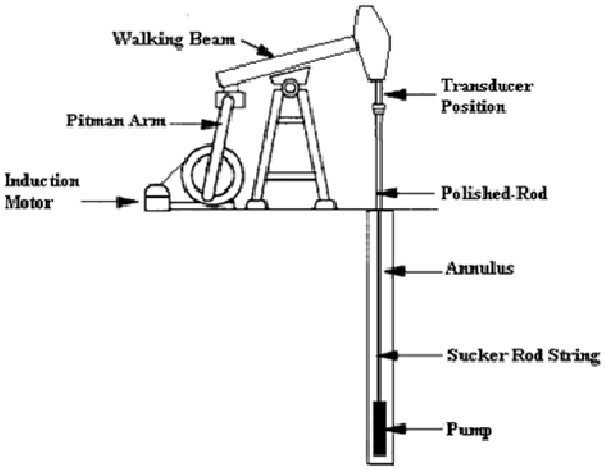
It should be remembered that the wavenet structure used in this work represents the nonlinear function of the NARX model. Therefore, for the nonlinear and linear part of the NARX model and wavenet, respectively, it can be verified that: $F(.)$ (of the NARX model) = $\sum_{k=1}^{ns} a_{sk}\varphi(.) +$

$\sum_{k=1}^{nw} \omega_{wk}\, \psi(.)$ (of the wavenet); and $x(k)\theta^T + l$ (of the NARX model) = $x(k)P\theta^T + l$ (of the wavenet).

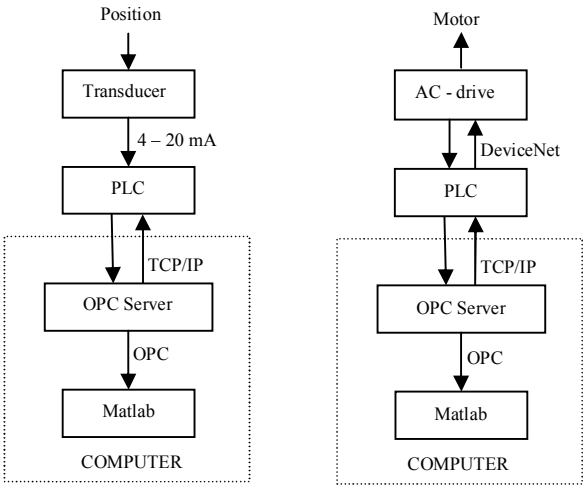
V. Case Study

A. Oil Well Pump Unit

The experiment set-up involves a pumping unit, an oil well that has been developed for tests (as shown in Fig. 3 (a)) and a data acquisition system to sample the position and torque current of the motor (generated by AC-drive), as shown in Fig. 3 (b) and 3 (c), respectively.



(a)



(b) (c)

Fig. 3 Oil well and pumping unit (a); data acquisition system of position (b) and data acquisition system of torque current (b).

A fluid with low viscosity, such as light oil, was used in the well. A three-phase asynchronous induction motor with 5 HP, 6 poles, and speed nameplate of 1160 rpm controlled by an AC-driver, was used as the primary power for the pump unit, and a string potentiometer as position transducer. The experiment was conducted under controlled conditions, and simulated actual operating conditions. In this case, the well starts with the annulus within the high level range. It then changes to normal operating conditions, arriving at the condition in which the level in the annular is insufficient to complete the task of filling the piston chamber during the upward cycle. The frequency of the pumping unit was also randomly varied in order to excite the dynamics of the system and simulate normal operating conditions.

.

B. Choice of Input Signal

The torque current of the motor (torque-producing current) was chosen as an input variable in the system identification because it can be acquired directly from the read parameter of the sensorless vectorial AC-drives, and also because it is strongly correlated to the change of position in the sucker rod.

It should be remembered that the input current motor can be decomposed into two components (see Fig. 4): the flux-producing current (IM) and the torque-producing current (IR), as shown in (23),

$$I_s = \sqrt{I_M^2 + I_R^2} \tag{23}$$

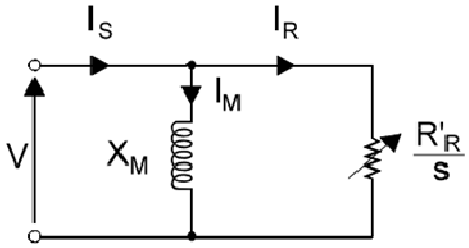


Fig. 4 Simplified equivalent circuit of an AC induction motor. Where V is the stator suply voltage, X_M is the magnetizing inductance and the relation R'R/S is the variable rotor resistence.

where, I_s is the stator current [30]. The behavior of induction motors in a pumping unit motor is dynamic and cyclic. Throughout the stroke, torque and speed are continually changing as the motor reacts to the dynamic load. The torque produced in the motor is proportional to the cross product of I_M and I_R [30]. Increasing the load on the motor shaft also increases the strength of the magnetic field in the stator and rotor. In an AC-motor, the torque-producing current *(also called, torque current)* increases in proportion to the increase in the load torque. Naturally these changes can be captured through I_R.

The interested reader is referred to [30] for further details about sensorless vector AC-drives.

C. Data Collection

The torque current of the motor (generated by Ac-drive) and the position of the polished rod, input (u) and output (y) data respectively, were sampled with a period of 0.05 seconds and a total of 31000 data were collected. The data set was split into two subsets, to create two independent data sets. The first range ($ze^N = \{u(1), y(1), \ldots, u(17000), y(17000)\}$) was used for model estimation, and the second range ($zv^N = \{u(17001), y(17001), \ldots, u(31000), y(31000)\}$) was used

for model validation. Fig. 5 shows a window with 3000 samples of filtered input-output data.

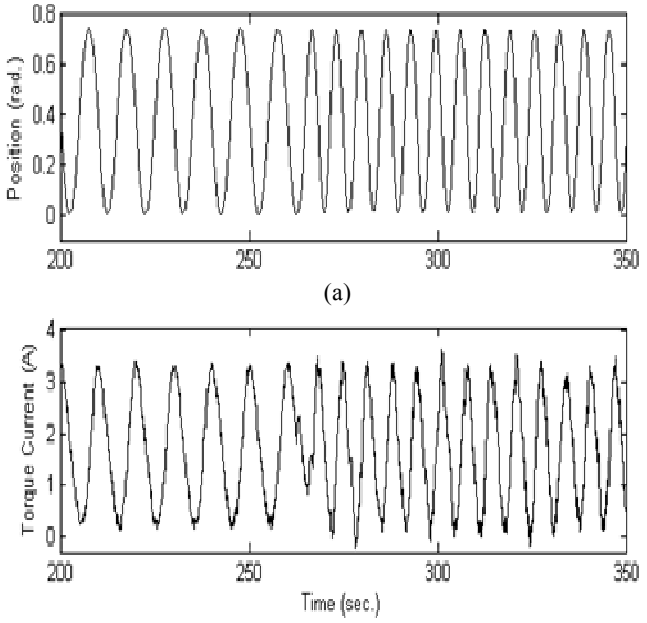


Fig. 5 Polished rod position (a), and torque current of the motor (b).

D. Model Order selection

According to [22], [23], the choice of a model orders is a combination of prior knowledge of the system and trial and error. A high dynamic order of the model can result in high variance error and increases its complexity, whereas underestimating the orders will result in a biased model [23].

In this paper, several orders were suggested and evaluated through Akaike's final prediction error (FPE). The representation of FPE is defined as,

$$FPE = \left[\frac{N+n_\theta}{N-n_\theta}\right] V_n(\theta|Z^N) \qquad (24)$$

where, $V_n\ (\theta|ze^N) = \det\left(\frac{1}{N}\sum_{t=1}^{N}\varepsilon(t,\theta_n)\left(\varepsilon(t,\theta_N)\right)^T{}_a\right)$ is the loss function, ze^N is the estimation data, N is the length of the input or output data. $\varepsilon(t,\theta_N) = y_1(t) - \hat{y}_1(t|\theta_N)$ is the prediction error parameterized by θ_N, between measurement output y(t) and predicted output $\hat{y}(t|\theta)$, and n_θ is the number of estimated parameters.

E. Evaluating the Estimated Model

It is possible evaluate how much (in percentage terms) each estimated model is able to infer the polished rod position from the torque current of the motor. For this purpose, we use the following equation:

$$fit = \left(1 - \frac{\sum_{k=1}^{N}|y_k - ys_k|}{\sum_{k=1}^{N}|y_k - \bar{y}|}\right) \times 100 \qquad (25)$$

where, y_k (y, ..., y_N) is the measurement output, ys_k (ys_1, ..., ys_N) is the model output, and $\bar{y}$ is the mean of vector y_k. For (25), the value of 100% corresponds to a perfect fit.

Remembering that for an appropriate evaluation, the torque current of the motor (u(k)), present in the validation data set (zv), was used as input variable into each estimated model.

VI. Results and Discussion

Table 1 summarizes the main results obtained after the estimation of the models. In our experiments, the model M2 with a number of 43 wavelets function selected by GCV, for construction of the wavelet network, was indicated by FPE as being the most appropriate means of representing the dynamics of the system. Thus, after simulating model output $\hat{y}(k)$, applying the torque current of the motor $u(k)$ that is present in validation data set *zv,* into equation (26)

$$\hat{y}(k) = \hat{f}(\hat{y}(k-1), \hat{y}(k-2), u(k-1), u(k-2), u(k-3)) \qquad (26)$$

and making a comparison with the measured position present in *zv*, a fit of 90.18% was obtained. However, after an exhaustive search to find a more appropriate model than the M2 model, the results obtained show that the model M4 in which only the regressors *y(t-1),y(t-2),y(t-3),u(t-1)* and *u(t-2)* are input from the nonlinear part of the NARX model (in other words, the regressor u(k-3) are input only of $x(k)\theta^T$), afforded a better fit to the experimental data than M2 (see Fig. 3).

VII. Conclusion

This paper has outlined an estimate of a pumping unit system. The validity of the model was obtained by comparing the simulated output with the measured output. The results obtained showed that the nonlinear ARX model based on wavenet with 34 wavelets outperformed all the other models that were estimated, and was sufficient to infer the dynamic position of the polished rod from torque current of three-phase induction motor, which is given by the sensorless vector AC-drive. This shows that it is possible to replace the position sensor used in the sucker-rod pumping with a nonlinear black-box model, which can be programmed in a microprocessor system.

Table I Summary of estimates for NARX models

NARX Model	Standard Regressors [na nb nk]	Nonlinear Regressors	Number of wavelets	FPE	Fit (%)
M1	[1 2 1]	y(k-1), u(k-1), u(k-2)	110	$3.58\ 10^{-6}$	64.47
M2	[2 3 1]	y(k-1), y(k-2), u(k-1), u(k-2),u(k-3)	43	$5.67\ 10^{-7}$	90.18
M3	[3 3 1]	y(k-1),y(t-2),y(k-3),u(k-1),u(k-2),u(k-3)	40	$5.79\ 10^{-7}$	85.95
M4	[3 3 1]	y(k-1),y(k-2),y(k-3),u(k-1),u(k-2)	34	$5.69\ 10^{-7}$	92.48

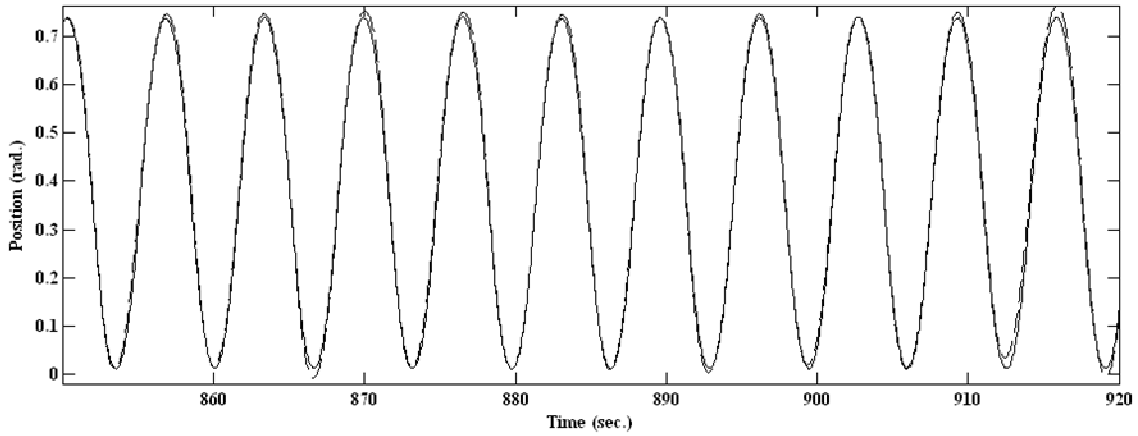


Fig. 6 A comparison between measured output *y(k)* (solid line) and the simulated output $\hat{y}(k)$ (dashed line) with the NARX model M4. A window of 70 seconds.

ACKNOWLEDGMENT

The authors would like to express their thanks to the technical team at the Artificial Lift Techniques Laboratory, Federal University of Bahia for their assistance in conducting the experiments.

REFERENCES

[1] M. Cook and M. Graham, *Hydrocarbon exploration and production*, Developments in Petroleum Science, No. 46, Elsevier, 2003.
[2] H.B. Bradley, *Petroleum Engineering Handbook*, Third Printing Society of Petroleum Engineers, U.S.A., 1992.
[3] S.G. Gibbs, "Prediction of the behavior of a sucker-rod pumping system," J. Pet, Eng., 1963, pp.769-778.
[4] D.R. Doty and Z Schmidt, "An improved model for sucker-rod pumping," Soc. Pet. Eng., 1983, pp 34-41.
[5] S. D. L. Lekia and R. D. Evans, "Coupled Rod and Fluid Dynamic Model for Predicting the Behavior of Sucker - Rod Pumping System," Society of Petroleum Engineers - 21664, 1991, 297-312.
[6] S. Miska, A. Sharaki, and J.M. Rajtar, "A simple model for computer-aided optimization and design of sucker-rod pumping systems," Journal of Petroleum Science and Engineering, 17, 1997, 303-312.
[7] M.H. Hojjati, S.A. Lukasiewicz, "Modelling of sucker rod strings," Journal of Canadian Petroleum Technology 44 (12), 2005, 55–58.
[8] L. Ljung, "Black-box models from input-output measurement," IEEE Instrumentation and Measurement, Technology Conference, Budapest, Hungary, May 21-23, 2001.
[9] T. Söderström and P. Stoica, *System Identification*, Prentice-Hall Int., London, 1989.
[10] M. F. RAHMAT, Y. K. HOE, S. USMAN and N. A. WAHAB, "Modeling of PT326 Hot Air Blower Trainer Kit Using PRBS Signal and Cross-Correlation Technique," Jurnal Teknologi, jun., 42(D), 2005, 9–22.
[11] S. Haykin, Neural Networks: *A Comprehensive Foundation*, Pearson Education, 2nd Edition, 1998.
[12] C. Alippi and V. Piuri, "Experimental neural networks for prediction and identification," IEEE Transactions on Instrumentation and Measurement," Vol. 45, No. 2, April, 1996.
[13] A. ZAK, "Neural Model of Underwater Vehicle Dynamics," International Journal of Mathematics and Computers in Simulation, Volume 1, 2007.
[14] S. Pittner, S. V. Kamarthi, and Q. G. Gao, "Wavelet networks for sensor signal classification in flank wear assessment," Journal of Intelligent Manufacturing, vol. 9, 1998, pp. 315–322.
[15] W. Zhong Li, W. Liao, and P. Han, "Application of Wavelet Network for Detection and Localization of Power Quality Disturbances," WSEAS Transactions on Power Systems, Volume 1, 2006.
[16] A.A. Safavi, J.A. Romagnoli, "Application of Wave-nets to Modelling and Optimisation of a Multidimensional Chemical Process," IEEE Xplore Digital Library, Australia, 2006.
[17] Y. C. Pati and P. S. Krishnaprasad, "Discrete Affine Wavelet Transforms for Analysis and Synthesis of Feed-forward Neural Networks," in R.P. Lippmann, J.E. Moody, D.S. Touretzky (Eds.), Advances in Neural Information Processing Systems III, 1991, pp. 743-749.
[18] Q. Zhang, and A. Benveniste, "Wavelet networks", IEEE Trans. Neural Networks, NN-3(6),1992, 889-898.
[19] S. Postalcıoğlu, K. Erkan, and E. D. Bolat, "Comparison of Wavenet and Neuralnet for System Modeling," Proceedings of 9th International Conference, KES, Australia, Part II, 2005, pp. 100-107.
[20] A.A. Safavi and J.A. Romagnoli, "Application of Wavelet-based Neural Networks to the Modelling and Optimisation of an Experimental

Distillation," Engng Applic. Artif. lntell, Vol. 10, No. 3, 1997,pp. 301-313.
[21] Q. Zhang, "Using wavelet network in nonparametric estimation," IEEE Transactions on Neural Network, Vol. 8, No. 2, March, 1997.
[22] O. Nelles, Nonlinear System Identification: From Classical Approaches to Neural Network and Fuzzy Models, Springer, 2001.
[23] L. Ljung, System Identification - Theory for the User, 2nd edn., Prentice-Hall, Englewood Cliffs, N.J., 1998.
[24] T. Kugarajah and Q. Zhang, "Multidimensional wavelet frames," IEEE Trans. Neural Networks, vol. 6, pp. 1552–1556, Nov. 1995.
[25] E. D. Sontag, Feedforward Nets for Interpolation and Classification, J.Comp.Syst.Sc, 1991.
[26] I. Daubechies, Ten Lecture on Wavelets, Society for Industrial and Applied Mathematics, Pennsylvania, 1992.
[27] S. Siripanadorn, W. Hattagam, and N. Teaumroong, "Anomaly Detection in Wireless Sensor Networks using Self-Organizing Map and Wavelets," International Journal of Communications, Volume 4, pp. 74-83, 2010.
[28] A. Cohen, Numerical Analysis of Wavelet Methods, Elsevier, 2003.
[29] J. Duffin and A. C. Schaeffer, "A class of nonharmonic Fourier series," Amer. Math. Soc., vol. 72, pp. 341–336, 1952.
[30] M. Barnes, Practical Variable Speed Drives and Power Electronics, Elsevier, 1st edn, 2003.