IDENTIFICATION OF POLINOMIAL NARMAX MODELS FOR AN OIL WELL OPERATING BY CONTINUOUS GAS-LIFT

Pagano, D. J. * **Dallagnol Filho, V.** * **and Plucenio, A.** *

* *Departamento de Automação e Sistemas, Universidade Federal de Santa Catarina, 88040-900 Florianópolis-SC, Brazil*
e-mail: {valdemar, daniel, plucenio}@das.ufsc.br

Abstract: Two nonlinear models (polynomial NARMAX) are identified for a simulated oil well operating by continuous gas-lift. The chosen input/output pair (injected gas mass flow rate/pressure drop in the production tubing) used in the identification can be applied in a control strategy decoupling injection from production choke control. The model derived with data obtained by exciting the plant around three different operating points compares well with another using a more aggressive excitation.

Keywords: Nonlinear Identification, NARMAX models, oil well production, continuous gas-lift

1. INTRODUCTION

In order to control a well operating by continuous gas-lift, a mathematical model of the well is usually needed. However, physical modelling of the input and output relations is complex, encompassing partial differential equations, which are hard to manipulate. An alternative is to use *identification* techniques, which try to find mathematical relations between the input and the output series of a system, without prior knowledge of its internal behavior.

The ultimate goal is to control the wellhead flow-rate. In an effort to avoid using expensive multiphase flowmeters, this is obtained indirectly by controlling other variables like the pressure in front of the perforations. The idea is to control the pressure in the wellhead and the pressure drop in the production tubing in such a way as to have a desired pressure in front of the perforated zone. The control of the pressure in the wellhead is done with a local controller and is part of the setup used in the identification.

The system under analysis has a clearly nonlinear behavior, making any linear model valid only inside a narrow operating region. The specific type of nonlinear model chosen is the polynomial NARMAX (*Nonlinear AutoRegressive Moving Average model with eXogenous inputs*). An arsenal of simple and robust algorithms is available to estimate the parameters of this kind of models.

This paper is organized as follows: first of all, the polynomial NARMAX model is presented; then the system under analysis is described. Following the identification procedure is described and finally conclusions are drawn.

2. NARMAX MODELS

A NARMAX model is represented like follows (Leontaritis and Billings, 1985):

$$y(k) = F[y(k-1), \ldots, y(k-n_y), u(k-1), \ldots, u(k-n_u), \nu(k), \nu(k-1), \ldots, \nu(k-n_\nu)], \quad (1)$$

where F is a nonlinear function, $u(k)$ is the input signal, $y(k)$ is the output signal, $\nu(k)$ is the noise in the system, n_y, n_u and n_ν are the largest delays in y, u e ν, respectively. However the determination of the function F is a hard task.

A polynomial NARMAX model is an expansion of the function F in a polynomial function with degree of nonlinearity ℓ. It is considered that the system does not have pure time delay and that none of the parameters to be estimated depends on $\nu(k)$. The polynomial approximation with degree of nonlinearity ℓ is given by (Chen and Billings, 1989):

$$y(k) = \theta_0 + \sum_{i_1=1}^{n} \theta_{i_1} x_{i_1}(k) + \sum_{i_1=1}^{n} \sum_{i_2=i_1}^{n} \theta_{i_1 i_2} x_{i_1}(k) \cdot x_{i_2}(k) + \ldots + \sum_{i_1=1}^{n} \cdots \sum_{i_l=i_{l-1}}^{n} \theta_{i_1 \ldots i_l} x_{i_1}(k) \ldots x_{i_l}(k) + \nu(k) \quad (2)$$

where:

$$x_1 = y(k-1) \qquad x_{n_y+1} = u(k-1)$$
$$\vdots \qquad\qquad \vdots$$
$$x_{n_y} = y(k-n_y) \qquad x_n = u(k-n_u) \quad (3)$$

being $n = n_y + n_u$ and θ constant parameters.

The use of a polynomial NARMAX representation may be justified by the following reasons: it is a global representation, allowing the global dynamics of the system to be represented, and not only the dynamics around a certain equilibrium point; it is easy to quantify the complexity of the model, based on the degree of non-linearity, number of terms and maximum delay used; it may deal with moderated levels of noise; analytical information about the model is easy to acquire; it is possible to have NARMAX models with a good fit to the data, as long there are not abrupt variations in the signals (Leontaritis and Billings, 1985); simple and robust algorithms may be used to estimated the parameters (since the model is linear in the parameters).

3. SYSTEM DESCRIPTION

The continuous gas-lift works by reducing the gravity term of the production tubing pressure drop. This is accomplished by injecting gas inside the production tubing through a gas-lift valve. Gas, being much lighter than the liquid in the production tubing, moves up, gasifying the flowing fluid, reducing its average density and, consequently, the pressure in front of the perforated zone.

In most wells, several gas-lift valves are distributed along the production tubing in such a way as to permit gas to enter progressively from top to bottom valve when injecting gas in the annular tubing-casing. The deepest valve is the only one which remains in operation while the other valves are only used for the start-up of the well. This work proposes a different set-up in an effort to avoid the utilization of mechanical gas-lift valves. In this approach an orifice valve is installed downhole, substituting the classical gas-lift valves and

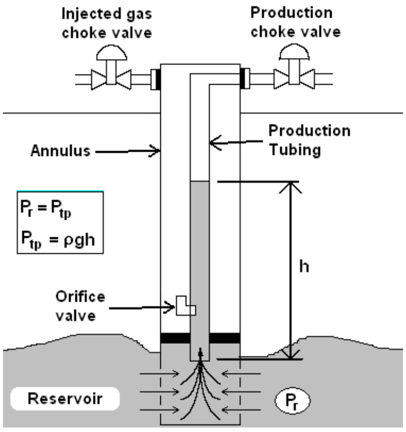

Fig. 1. Oil well

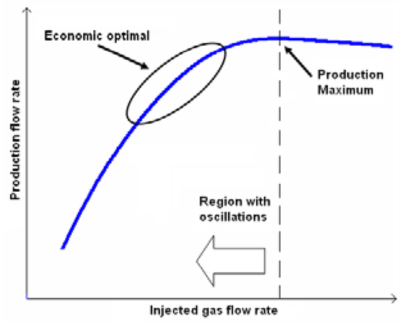

Fig. 2. Production flow x Injected gas flow curve, with the area of largest economic interest signaled

the control is done in the surface acting on the gas-lift and production chokes. Figure 1 shows the main components of the gas-lift oil well set-up considered in this work. The start-up procedure for this set-up is not studied but it could possibly be done with a high pressure compressor.

There is an optimal operating region for the well, economically speaking, shown in Figure 2, which is related to the fluid fraction flow-rates produced by the well, its current market prices, and the costs of gas-compression and so on. This region, however, has the inconvenient of presenting oscillations when the system operates in open-loop, reducing the well productivity and affects the oil, water, gas separation efficiency.

Several works have appeared in the literature (Eikrem *et al.*, 2004), proposing different strategies to stabilize the oscillations in wells operating via gas-lift using similar set-up acting in the production choke.

In (Plucenio, 2002) a control strategy is proposed using the mass flow-rate measured on the surface, and acting in the gas injection mass flow-rate. Linear ARX models are identified in three different points of operation in order to develop a robust control.

In this paper, the well is treated as a SISO system, with the mass flow rate of injected gas (Q_i) as the input and the pressure in the production tubing (P_{tp}) as the output (see Fig. (3)). The input of the system

Q_i is actually the setpoint of a controller actuating in the injection valve opening. This controller has the standard PI structure, with $K_p = 5 \times 10^{-3}$ and $K_i = 0.1$. The pressure in the production tubing may be decomposed as $P_{tp} = P_{wf} - P_h$, where P_{wf} is the pressure measured in the bottom of the well and P_h is the pressure measured in the head of the well. The main advantage of considering P_{tp} as the output of the system is that P_{tp} is relatively isolated of disturbances in the pressure on the boundaries of the system (in the separator). P_{wf} and P_h will react similarly to these disturbances and compensate for these disturbances when P_{tp} is calculated. The pressure in the head of

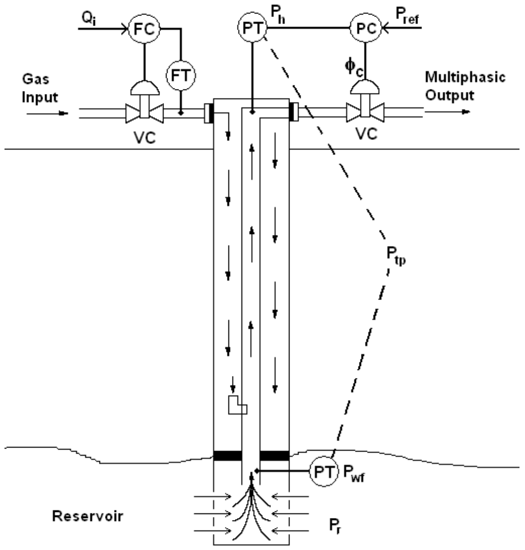

Fig. 3. Measuring and actuation points in the oil well

the well is also controlled by a local controller that, by acting on the opening of the production choke, guarantees that P_h remains constant (which is desirable). The setpoint for the P_h controller is 2.24 MPa, being the structure a standard PI with $K_p = -1 \times 10^{-5}$ and $K_i = 0.01$.

This definition of input and output variables have the advantage of allowing easy implementation, since instrumentation for measuring the pressure in the head and bottom of the well is common in modern wells (Veneruso *et al.*, 2000). Besides that, measuring pressure is trivial, on the contrary to the instrumentation needed for measuring the flow rate of a multiphasic fluid, which is very expensive.

The system possesses an obvious nonlinear behavior, which can be observed in Figure 4, showing the output corresponding to the application of a sequence of steps in the input of the system. It may be observed that not only the transitory response changes depending of the region of operation, but also the steady state response, and the signal of the static gain, which changes from negative to positive when the injected gas flow rate increases beyond a certain point. The desired operating region lies in a region with negative gain.

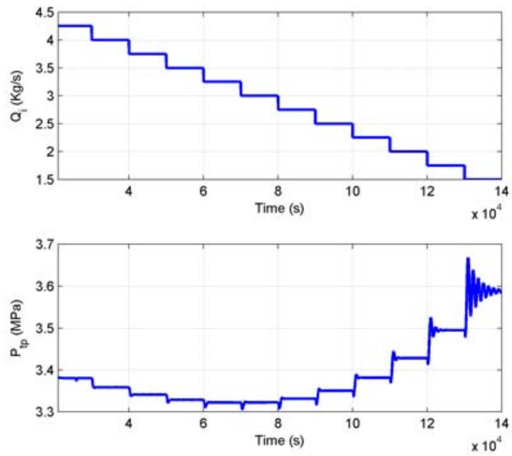

Fig. 4. Top: sequence of steps applied in the input of the system (Q_i). Bottom: Corresponding output (P_{tp}) showing the nonlinearity of the system

Besides the nonlinear characteristic, the system possesses a non-minimum phase response (see Fig.(4)), which makes harder the synthesis of a controller. [1]

It must be noted that the model identified to quantify the relation $P_{tp} \times Q_i$ is influenced by the choice of the parameters of the local controllers (for gas injection and for the pressure in the head of the well). Any change in the structure or the parameters of this controllers demand a new identification of the system.

The data used for identification was generated with the software OLGA® 2000, by Scandpower Co., version 4.10.1. The system used in the simulator is a modification of a model supplied by Scandpower, representing a real well operating in deep waters in the Mexican Gulf. The well has the following characteristics:

- Reservoir static pressure = 33.094 MPa
- Reservoir temperature = 82.2^oC
- Reservoir productivity index = 2×10^{-6} kg/s/Pa
- Pressure in the separator = 2.585 MPa
- Temperature in the separator = 26.7^oC
- Gas pressure at compressor output= 9.652 MPa
- Gas temperature at compressor output = 20^oC

4. NONLINEAR IDENTIFICATION

First of all, the original model in (2) was changed, including in the candidate terms those containing $u(k)$. The presence of a term containing $u(k)$ indicates that there may exist a direct transfer of information from the input to the output of the system, in other words, a part of the dynamic may be fast enough to reflect

[1] The term "non-minimum phase" is generally used in the context of linear systems meaning the presence of a zero outside the unit circle (in the discrete case). This notion was extrapolated here, where the term "non-minimum phase response" was used to state that the response of the system presents a behavior similar to the one that could be found in a linear system with non-minimum phase

“immediately” in the output. The model is therefore given by:

$$\begin{aligned} y(k) = \theta_0 &+ \sum_{i_1=1}^{n} \theta_{i_1} x_{i_1}(k) \\ &+ \sum_{i_1=1}^{n} \sum_{i_2=i_1}^{n} \theta_{i_1 i_2} x_{i_1}(k) \cdot x_{i_2}(k) + \ldots \\ &+ \sum_{i_1=1}^{n} \cdots \sum_{i_l=i_{l-1}}^{n} \theta_{i_1 \ldots i_l} x_{i_1}(k) \ldots x_{i_l}(k) + e(k) \end{aligned} \tag{4}$$

where:

$$\begin{aligned} x_1 &= y(k-1) & x_{n_y+1} &= u(k) \\ x_2 &= y(k-2) & x_{n_y+2} &= u(k-1) \\ &\vdots & &\vdots \\ x_{n_y} &= y(k-n_y) & x_n &= u(k-n_u) \end{aligned} \tag{5}$$

being $n = n_y + n_u + 1$, n_y the maximum delay in y and n_u the maximum delay in u.

As input signals for the system, two strategies where used: the first one used an “aggressive” signal, with more abrupt variations, which tries to excite a large range of frequencies and reach different operating regions of the system. The second signal is more “well behaved”, using small variations around three operating points, reducing the risk of damage to the plant.

4.1 *Aggressive signal*

The “aggressive” signal was obtained by keeping the input signal constant at $Q_i = 2.15$ kg/s, until the initialization transitory of the system was over. After it, it was added to the constant signal a random signal with zero mean and unitary variance, being each step kept for 200 seconds. The use of this random signal tries to excite a broad range of frequencies. Before adding the random signal to the constant, it is multiplied by a crescent value, such that the system starts operating around the operating point and move away from it as time passes. Figure 5 shows the input signal applied and Figure 6 shows the corresponding output. The test duration was 15000 seconds, with a sampling rate of 40 seconds.

Another signal with the same characteristics but with another realization of random numbers was used as input of the system to produce data to validate the identified models. The desired model has a degree of nonlinearity $\ell = 2$, $n_y = n_u = 5$, resulting in 78 candidate terms. Besides these terms, 10 linear noise moving average terms were added to avoid biasing of the estimates.

Among the candidate terms, there are 6 term clusters ($\Omega_0, \Omega_y, \Omega_{y^2}, \Omega_u, \Omega_{u^2}$ and Ω_{yu}). The term cluster Σ_{y^2} was eliminated from the candidate terms set, because

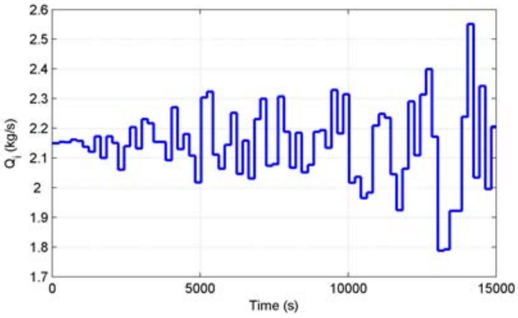

Fig. 5. Aggressive input signal of the system used to estimate the parameters of the nonlinear model

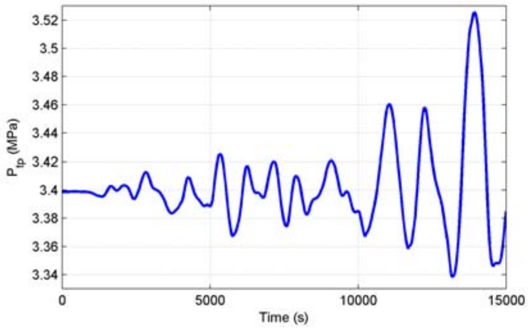

Fig. 6. Output signal resulting from the aggressive input

the desired polynomial NARMAX model should have only one fixed point. The location of the fixed points of the model (which has degree $\ell = 2$) is the solution of the equation:

$$(\Sigma_{y^2})\bar{y}^2 + (\Sigma_y + \Sigma_{yu}\bar{u} - 1)\bar{y} + (\Sigma_0 + \Sigma_u\bar{u} + \Sigma_{u^2}\bar{u}^2) = 0. \tag{6}$$

Therefore, by eliminating the term cluster Ω_{y^2}, there will exist only one fixed point located in:

$$\bar{y} = -\frac{(\Sigma_0 + \Sigma_u\bar{u} + \Sigma_{u^2}\bar{u}^2)}{(\Sigma_y + \Sigma_{yu}\bar{u} - 1)} \tag{7}$$

The Error Reduction Ratio (ERR) criterium (Chen and Billings, 1989) was used to sort the sequence that the terms should be included in the model, but the actual number of terms in the final model was determined by using the Akaike Information Criterium (AIC) (Akaike, 1974). The parameters of the estimated model were then checked for statistical significance, by comparison with the standard deviation of the estimate. A 99% level of significance was used, meaning that each parameter should satisfy $-3\sigma_i \leq \hat{\theta}_i \leq 3\sigma_i$, where σ_i is the standard deviation of the estimate of the parameter i and $\hat{\theta}_i$ is the estimate of the parameter i. An iterative process was then performed, were the terms that were not significant (but were still included in the model by the ERR criterium) were excluded from the set of candidate terms and a new model was identified and checked for significance.

The final model identified has 6 deterministic terms, shown in table 1, plus 10 linear moving average terms

Table 1. NARMAX model terms (aggressive input) ordered by the ERR value.

Order	Term	$\hat{\theta}_i$	σ
1	$y(k-1)$	$+2.01356$	$+2.23668\times10^{-2}$
2	$y(k-2)$	-1.01002	$+2.24378\times10^{-2}$
3	$u(k-4)y(k-2)$	-3.49843×10^{-2}	$+2.93181\times10^{-3}$
4	$u(k-4)y(k-5)$	$+2.37093\times10^{-2}$	$+3.52717\times10^{-3}$
5	$u(k-4)$	$+3.09134\times10^{-2}$	$+2.84186\times10^{-3}$
6	$u^2(k)$	$+8.50101\times10^{-4}$	$+7.68678\times10^{-5}$

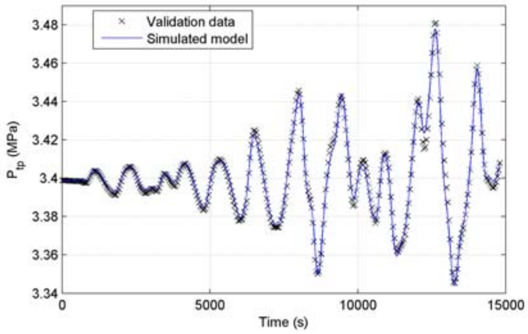

Fig. 7. Free simulation of the NARMAX model identified with the data from an aggressive input, compared with validation data

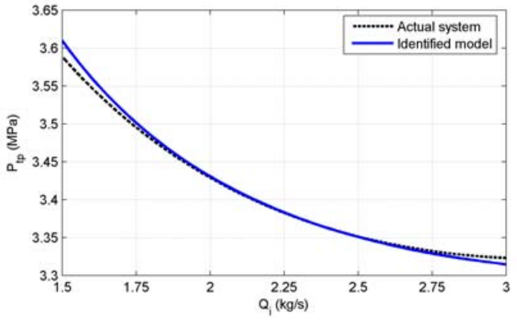

Fig. 8. Fixed points of the NARMAX model identified for the aggressive input

of the noise signal, which had the sole purpose of avoiding biasing of the estimates, being discarded afterwards. To quantify the quality of a model, it was used the fit index defined by (Ljung, 2004):

$$\text{fit} = 100 * \left(1 - \frac{\|\hat{\boldsymbol{y}} - \boldsymbol{y}\|}{\|\boldsymbol{y} - \text{mean}(\boldsymbol{y})\|}\right), \qquad (8)$$

where $\hat{\boldsymbol{y}}$ is the vector with the output of the model and $\boldsymbol{y}$ is the vector with the real output of the system. The equation (8) compares the quality of prediction of a model with the mean of the data as a trivial predictor.

By using the validation data to evaluate this model, the output of the model had a fit $= 87.78\%$, as seen in Figure 7. The steady-state characteristic of the model, compared to the steady-state characteristic of the system may be seen in Figure 8. It may be seen that the model represents well the system under analysis in the defined operating region (from $Q_i = 1.5$ kg/s to $Q_i = 3$ kg/s).

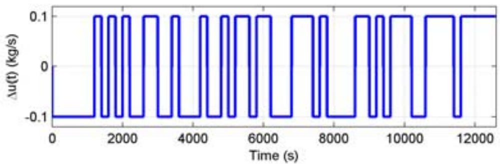

Fig. 9. PRBS Signal $\Delta u(t)$ applied in the system

4.2 *Well behaved signal*

The model identified in the previous section was able to reproduce adequately the dynamics of the system under analysis. However, the input used to generate the identification data is too "aggressive", presenting big changes and may be risky to use in the real plant. In order to avoid this risks, a new NARMAX model was identified, using data acquired from the use of a "well-behaved" input (with smaller changes in the signal).

The system was carefully brought to three operating points and when in steady-state, a PRBS signal (*Pseudo Random Binary Signal*) was applied in the input. Figure 9 depicts the PRBS signal used ($\Delta u(t)$). The actual signal applied in the input is $u(t) = \Delta u(t) + u_0$, where u_0 is the operating point. The three chosen operating points where $u_0 = 1.5$, $u_0 = 2.15$ and $u_0 = 2.8$ kg/s.

Obviously the data used for estimating the parameters of the model is not ideal in a theoretical point of view, because the input is restricted to three small operating regions, not passing through all the desired operating region (from 1.5 kg/s to 3 kg/s). However, the use of this data set has two advantages over the "aggressive" signal used in the previous section:

- it is less risky to the plant, for having less abrupt variations;
- production can still be carried on during the execution of the tests, because there is only a slight disturbance over the steady-state inputs.

The candidate models searched have the same characteristics of the ones searched in the previous section, being the degree of nonlinearity $\ell = 2$, $n_y = n_u = 5$, and 10 linear moving average terms used to avoid biasing of the estimates. From the set of candidate terms the term cluster Ω_{y^2} was eliminated too.

After repeating an iterative procedure which includes: sorting the remaining candidate terms with the use of the ERR criterium, defining the number of terms to be included in the final model with the Akaike Information Criterium, verifying the statistical significance of the estimated parameters and validating statistically the model by residual analysis, the model shown in table 2 was found. The model has a fit $= 91.2\%$ to the validation data, which is an excellent performance. Figure 11 shows the steady-state characteristic of the model identified compared with the actual steady-state

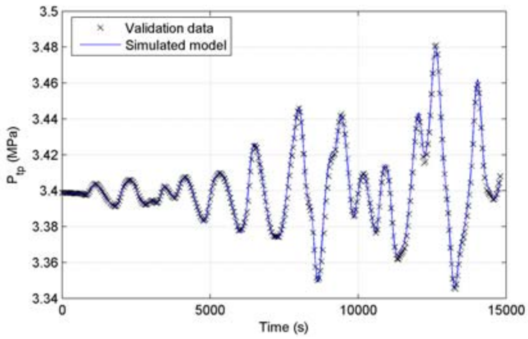

Fig. 10. Free simulation of the NARMAX model identified with the data from an well-behaved input, compared with validation data

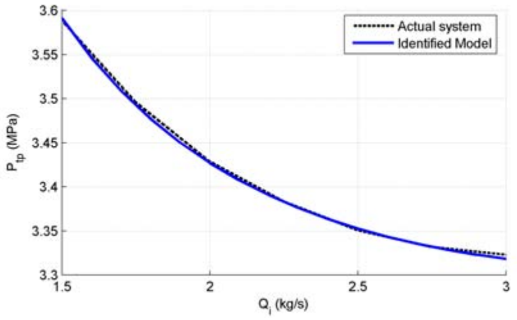

Fig. 11. Fixed points of the NARMAX model identified for the well-behaved input

characteristic of the system. It is seen in the figure that the model is a good representation of the oil well.

Table 2. NARMAX model Terms (well-behaved input) ordered by the ERR value.

Order	Term	$\hat{\theta}_i$	σ
1	$y(k-1)$	$+2.84619$	$+3.94354\times10^{-2}$
2	$y(k-2)$	-2.23221	$+5.70014\times10^{-2}$
3	$y(k-4)$	$+3.88183\times10^{-1}$	$+1.83896\times10^{-2}$
4	$u(k-2)y(k-1)$	-7.26036×10^{-3}	$+5.33763\times10^{-4}$
5	$u(k)$	$+5.29649\times10^{-3}$	$+1.78812\times10^{-4}$
6	$u(k-5)$	$+2.78799\times10^{-3}$	$+6.50890\times10^{-4}$
7	$u(k-3)y(k-1)$	-2.67362×10^{-1}	$+2.23946\times10^{-2}$
8	$u(k-2)$	$+1.25873\times10^{-2}$	$+2.18152\times10^{-3}$
9	$u(k-5)u(k-3)$	-1.22951×10^{-3}	$+2.48621\times10^{-4}$
10	$u(k-3)y(k-2)$	$+3.64800\times10^{-1}$	$+3.17758\times10^{-2}$
11	$u(k-3)y(k-4)$	-9.75874×10^{-2}	$+9.86630\times10^{-3}$
12	$u(k-2)u(k-2)$	$+1.73758\times10^{-3}$	$+2.33335\times10^{-4}$

5. CONCLUSIONS

In this paper, two models of an oil well operating by continuous gas-lift were identified, relating the pressure in the production tubing (output) with the mass flow rate of injected gas (input). The presented strategy has the advantage of allowing an easy implementation on existing oil wells, where the needed instrumentation is widely available (Veneruso *et al.*, 2000).

The two polynomial NARMAX models identified showed to represent adequately the system, which would be impossible to do with linear models. The absence of a stronger nonlinearity, in the considered range of gas-lift injection flow rate, made it possible to use a well behaved input signal, which is not ideal in a nonlinear identification viewpoint, but is preferred for presenting less risk to the plant during the test procedure.

The model identified with the well-behaved signal showed better performance when near the boundaries of the operating region, because two of the three operating points chosen to apply the PRBS signal are at the boundaries. The aggressive signal, in the other hand, concentrates the input in the middle of the operating region and so the model identified with has a slightly worse performance near the boundaries of the operating region.

As a next step in research, the models identified will be used to design a controller to the simulated plant in the OLGA simulator, as a previous step to the implementation of this control strategy in a real oil well.

6. ACKNOWLEDGMENTS

V. A. Dallagnol Filho was funded by CNPq. D. J. Pagano and A. Plucenio were funded by Agencia Nacional do Petroleo (ANP), Brazil, under project aciPG-PRH No 34 ANP/MCT. The authors would also like to acknowledge Scandpower for providing an academical OLGA software license.

7. REFERENCES

Akaike, Hirotugu (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19**(6), 716–723.

Chen, S. and S. A. Billings (1989). Representation of non-linear systems: the narmax model. *International Journal of Control* **49**(3), 1013–1032.

Eikrem, G. O., B. Foss and L. Imsland (2004). Stabilization of gas lifted wells based on state estimation. *IFAC ADCHEM2004, Hong Kong*.

Leontaritis, I. J. and S. A. Billings (1985). Input-output parametric models for non-linear systems. *International Journal of Control* **41**(2), 303–344.

Ljung, Lennart (2004). *System Identification Toolbox For Use With Matlab 6.0*. The MathWorks, Inc.. Natick, MA.

Plucenio, A. (2002). Stabilization and optimization of an oil well network operating with continuous gas-lift. *SPE ATCE, San Antonio, Texas*.

Veneruso, A. F., S. Hiron, R. Bhavsar and L. Bernard (2000). Realiability qualification testing for permanently installed wellbore equipment. *SPE 62955, ATCE, Dallas, Texas*.