A NEW SIGNAL DESIGN TOOL FOR PROCESS MODEL IDENTIFICATION

V. C. Machado, J. O. Trierweiler, and A. R. Secchi

Group of Integration, Modeling, Simulation, Control and Optimization of Processes (GIMSCOP)
Department of Chemical Engineering, Federal - University of Rio Grande do Sul (UFRGS)
Rua Duque de Caxias, 1303 apt. 603 CEP: 90010-283 - Porto Alegre - RS - BRAZIL,
Fax: +55 51 3316 3277, Phone: +55 51 3316 4072
EMAIL: {vinicius, jorge, arge}@enq.ufrgs.br

Abstract: A new low-cost systematic procedure to obtain high-quality data for model identification is presented. Two major steps compose this methodology: first, a simplified process model is built using a PIMS (Plant Information and Management System) database, which contains limited information about the process behavior. Second, the signal amplitudes and its directions are calculated, as well as the total time of the experiment. The signal is designed to do not remove the system from its normal operating point, reducing the identification costs. The new procedure was applied to an industrial distillation unit.

Keywords: Signals, Dynamic Models, Identification, Distillation Columns, Linear Control Systems

1. INTRODUTION

Model identification is the key part of modern advanced process control systems, since most process controllers are based on a model.

Generally, a model can be obtained through rigorous modeling based on the first principles or by empiric modeling and parameter estimation using plant data. In most cases, it is very difficult to create an accurate first-principles model, due to the complexity of the process and the lack of information. On the other side, the empirical modeling, although easier to obtain than the rigorous modeling, demands an intensive identification test, where all manipulated variables must be excited to produce some effects in the controlled variables. Several well-known model structures, like FIR (Finite Impulse Response), *ARX*, *OE* (Output Error), *BJ* (Box-Jenkins), (Ljung, 1999), *SUBID* (Subspace Identification) (Overschee and De Moor, 1996), etc., may be used to fit the plant data. The problem is that, normally, the identification tests remove the plant from its normal operating point and spend much time and money to be accomplished. It is important that the signal of the manipulated variables during the test be strong enough to excite the process output, producing a good signal-to-noise relation (*SNR*), given by:

$$SNR = \frac{\phi_{yy}}{\phi_{uu}} \qquad (1)$$

where ϕ_{yy} is the signal spectra of the output variable and ϕ_{uu} is the signal spectra of the input variable. High SNR guarantees a good quality of the identified model, but may drive the process away from the operating point. In this way, it is necessary to have a tool to produce good data for model identification at low cost and time.

In the literature, there is a variety of signals to dynamic model identification (Cutler, 1988; Luyben

1992). The most used signal in identification tests is the step signal, which is characterized by an abrupt level change in the value of the system inputs and is easy to implement. However, the step signal removes the plant from its normal operating point, and is time consuming, because the plant needs to achieve a new steady state after the input variation. The random binary signals and *Pseudo Random Binary Signals* (*PRBS*) (Tulleken, 1990) have the advantage to present zero mean during the identification test, maintaining the industrial unit in the operating point. However, they are also time consuming tests and are not suitable for multivariable systems as explained below. Still there are ramp signals and wave signals to identification test, but these signals are not so used in the practice.

For multivariable systems, usually, the step test is preferred in process industries, where each manipulated variable is changed individually. Clearly, this procedure is time consuming and the multivariable characteristic is not perfectly captured. Besides, the signal amplitude of each input variable is heuristically determined. The multivariable PRBS uses a lot of signal directions to excite the system. This fact is not good, because during the test a lot of data are repeated, bringing undesired correlation among all variables and the parameter estimation becomes a difficult task. These characteristics were the motivation to develop a new tool to signal design for model identification.

In the next section the proposed methodology is presented. In section 3, dynamic models of a distillation unit are identified with data generated from the proposed signal design methodology.

2. SIGNAL DESIGN METHODOLOGY

Unfortunately, most available identification tests do not use the historical plant data. Nowadays, with advent of *PIMSs* (Plant Information and Management Systems), a plenty of plant data can be stored and easily retrieved. Normally, these data are noisy and do not contain good dynamic information about the process. However, it is still possible to obtain some information about the process from these data.

In order to take into account the historical plant data in the signal design, two procedures are proposed in this work: characterization of the input and output noises and correlation analysis between inputs and outputs. The correlation analysis is used to select good plant data intervals to identify a simplified process model.

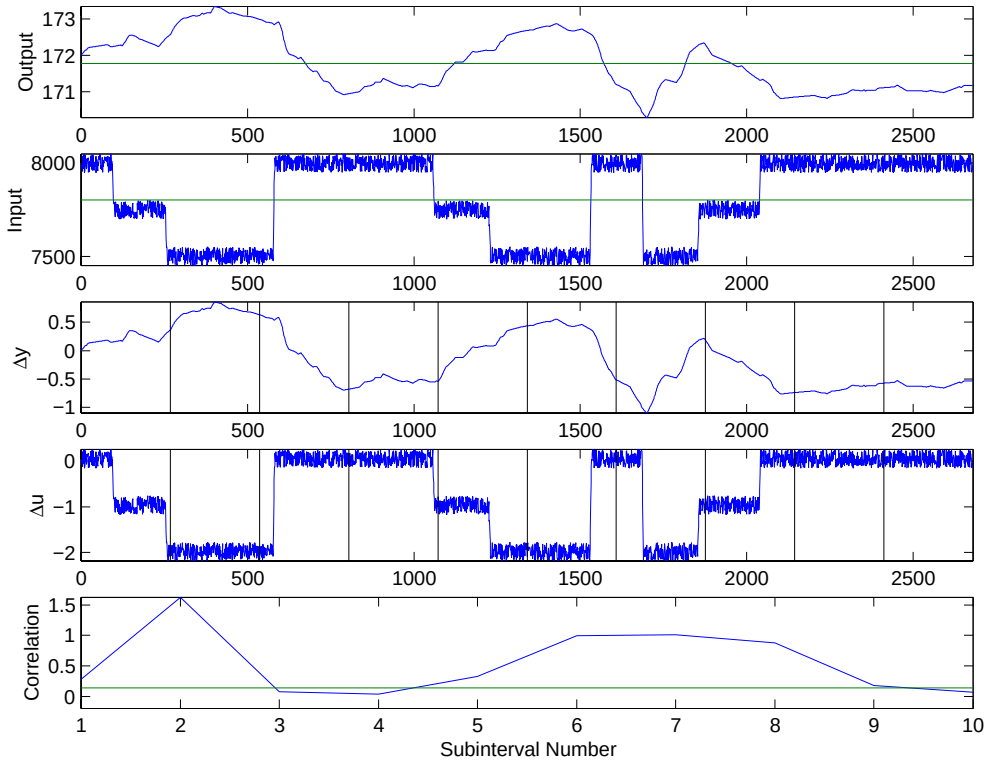

Fig. 1: Correlation analysis to select the best plant data intervals to model identification. The two first graphics show the vector *y(t)* and *u(t)* and their mean values, respectively. The third and fourth graphics show the vectors *Δu(t)* and *Δy(t)*, and each subinterval in which the correlation function was calculated. the last graphic shows the maximal correlation function values for each data subinterval and its threshold value.

2.1 Simplified process model identification

The first step of the multivariable signal design methodology is the search for plant data intervals to identify a simplified process model. It is based on a sweeping procedure to observe how inputs and outputs are correlated by calculating the correlation function, given by:

$$Corr(k) = \frac{1}{N}\sum_{t=1}^{N} \Delta u(t)\Delta y(t+k) \qquad (2)$$

where k means the output shifting (in sampling intervals), t means the actual sampling interval and t+1 means de next sampling interval. N is the total number of points and the variables Δu and Δy are the two normalized signals being analyzed. In case of industrial plant data, Δu (related to the inputs) and Δy (related to the outputs) are calculated by :

$$\Delta y(t) = \frac{y(t) - \mu_y}{\max\left(\left|y(t) - \mu_y\right|\right)} \qquad (3)$$

$$\Delta u(t) = \frac{u(t) - \mu_u}{\max\left(\left|u(t) - \mu_u\right|\right)} \qquad (4)$$

where μ_y and μ_u are the output and input mean values, respectively, *y(t)* and *u(t)* are the original output and input vectors of plant data, respectively.

To illustrate how this correlation analysis works to select the best plant data intervals, Figure 1 presents real plant data and the result of the calculated correlation function.

It is important to note that the correlation function is calculated for all time intervals, and the best intervals are selected to compose the data set to identify a simplified process model. To decide what intervals will compose the identification data set, a threshold value, L, should be defined. If the maximal correlation value is greater than L, then the subinterval is selected. This value is related to the noise level, and it was defined as a value proportional to the geometric mean between the standard deviations of the input and output noise, given by:

$$L = 2.6\sqrt{\sigma_u \cdot \sigma_y} \qquad (5)$$

where σ_u and σ_y are the standard deviation of the input and output noise, respectively. These values are obtained from the less correlated subinterval as described below.

Not only plant data time intervals need to be found, but also the input and output noises amplitudes. It is necessary to keep in mind that the input variation effects should be great than the noise amplitude to identify a good quality plant model. Although the plant data, in special the output variables, are corrupted by stochastic variations, it is convenient to affirm that the difference between the output signal (y) and its mean (μ_y), named as *ydiff* (*ydiff* = $y - \mu_y$), follows the normal distribution, which is represented by:

$$p(ydiff \mid \mu_y, \sigma_y) = \frac{1}{\sigma_y\sqrt{2\pi}} e^{\left(\frac{-(ydiff-\mu_y)^2}{2\cdot\sigma_y^2}\right)} \qquad (6)$$

where p is the probability density of the data (*ydiff*) with deviation σ_y and mean μ_y. If the data follows a normal distribution, then 95% of the plant data must lie between –2.6 σ_y and +2.6 σ_y. Considering this confidence interval, then the noise amplitude is 2.6 σ_y. Therefore, comparing the distribution of the plant output data with the normal distribution, it is possible to determine the noise level. Similar to the correlation analysis, the plant data is divided in time intervals, as shown in Figure 2, and a comparison between data distribution and the normal distribution is made in each interval. Figure 3 shows the comparison for the second time interval of Figure 2. The same comparison is done for all time intervals, and the interval which the data distribution is closer to the normal distribution is selected to evaluate the noise level. The metric used to make this selection is the absolute integral of the difference between the normal distribution and the data distribution. Observing the data of Figure 2, by visual inspection it is clear that the second time interval better represents the output noise level, which was the same result obtained by applying the proposed procedure, leading to an output noise level of 13.83 units

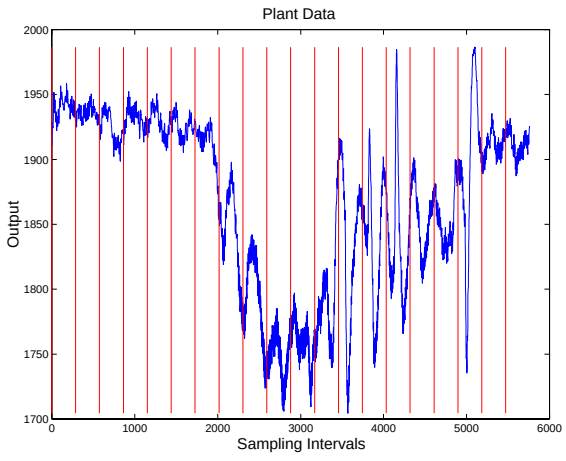

Fig. 2: Plant data and subintervals.

This procedure is repeated to the others input and output variables. As already was commented, the noise level is one of the most important information to design good signal to models identification. With

the selected data, a simplified model can be identified by any identification algorithm, such as *ARX*, *OE*, *BJ*, **s**ubspace **i**dentification, etc. At this stage it is only necessary a *First Order with Dead Time* (*FODT*) model to design a good input signals, having a reasonable information about the process time constant and process gain. In the next section it is shown how to determine the manipulated variables profile to be applied during the identification test.

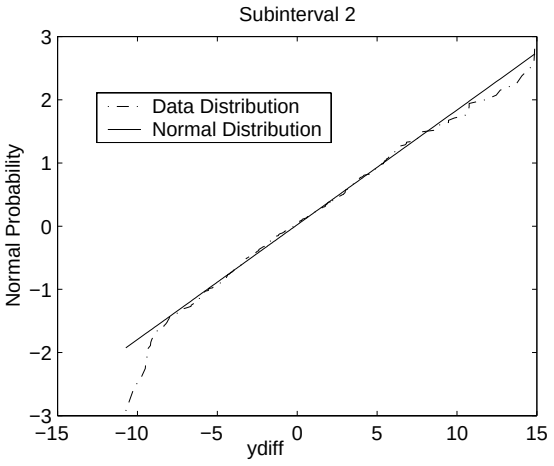

Fig. 3: Comparison between the normal distribution and the output data distribution, for the second time interval showed in Figure 2.

2.2 Input signal design

An important information in signal design is how much time is necessary to keep the system excited in order to capture the whole system response. In the proposed methodology, this information is obtained from the simplified model identified in the first step. The time constant of each channel of the transfer matrix contributes to determine the system time constant. To calculate a representative system time constant, it was chosen a geometric mean (τ_{GEO}) among all time constants present in the simplified model, calculated by:

$$\tau_{GEO} = \sqrt[N]{\prod_{i=1}^{N} \tau_i} \tag{7}$$

where N is the total number of the nonzero system time constants, and τ_i is the i^{th} system time constant. The geometric mean gives a tradeoff between large and small time constants. A simple arithmetic mean tends to be closer to the high values of time constants. Using the representative time constant of the system, τ_{GEO}, it is defined three relevant frequencies: $\omega_1 = 0.5\,\tau_{GEO}^{-1}$, $\omega_2 = \tau_{GEO}^{-1}$, and $\omega_3 = 1.5\,\tau_{GEO}^{-1}$, to determine how much time each manipulated variable will be kept in a certain level during the identification test. These frequencies were chosen in order to design signals with high probability to capture the true system dynamics. Then, the time that each signal should be kept in a determined level is:

$$time\_of\_disturbance = \frac{\pi}{\omega_i} \tag{8}$$

For each frequency, it is necessary to determine the most important directions of the input signals. It can be done by evaluating the frequency response, $G(j\omega_i)$, of the simplified model G (which was determined in the first step of the methodology). However, as it is very common to observe a great difference among the system variables magnitudes, it is important to consider the scaling of $G(j\omega_i)$, given by:

$$G_S(j\omega) = L_S . G(j\omega) . R_S \tag{9}$$

where L_S and R_S are the input and output scaling matrices, respectively. All necessary directions to design the best signals for model identification can be obtained by applying the Singular Value Decomposition (*SVD*) to $G_S(j\omega_i)$ yielding:

$$G_S(j\omega) = W \sum V^T \tag{10}$$

where W and V^T are the system output and input direction matrices, respectively, and Σ is diagonal matrix of the system singular values. Each column of the V^T is an input direction vector, named as U_d. The amplitudes, Δu, of the designed input signals are given by:

$$\Delta u = E_F . R_S . U_d \tag{11}$$

where E_F is an expansion factor to guarantee that all outputs will be excited, and is given by:

$$E_F = \max\left(\frac{L_S^{-1} \cdot \Delta y_{noise}}{\left| G_s(j\omega_i) \cdot U_d \right|} \right) \tag{12}$$

where Δy_{noise} is the noise amplitude (determined in the first step of the methodology). It should be clear that the above procedure has to be applied for each designed frequency ω_i. In the next section, the results of applying the proposed signals design methodology to model identification of a distillation unit.

3. INDUSTRIAL APPLICATION

The proposed methodology was applied to an industrial naphtha distillation unit. A brief

description of the process: the crude petroleum feed is heated up in a heat exchanger battery, the salt is removed, and the oil goes to the first separation unit, a flash drum, to obtain a mixture of naphtha, diesel, and solvents. The mixture goes to the second and main separation unit, a distillation column, which produces heavy naphtha, diesel, and small amount of gasoline. The control objectives in this unit are to keep the high quality of diesel, observed through temperatures and volumetric flows, and to reduce the atmospheric residue, which is the fraction of raw materials that did not transform in commercial products. This residual is produced in the distillation column. To attain these control objectives, a MPC (model predictive controller) calculates the setpoints of some important variables in the system, such as the petroleum feed, the flash temperature, and the top temperature of the distillation column. The studied distillation system has 8 inputs and 14 outputs. To make a good signal design for a block diagonal controller, it is necessary to group some related input variables. In this case, three blocks were created: block 1 (input 1, petroleum feed), block 2 (inputs 2 and 6, flash drum variables) and block 3 (inputs 3,4,5,7, and 8, related to the distillation column).

The signal design to identify a dynamic model to the distillation unit starts with a generation of a simplified model, based on historic plant data. Ten historic plant data sets were analyzed, and a correlation analysis was carried out according to the methodology described in Section 2.1. The noise amplitudes were also determined, whose results are presented in Table 1.

Table 1. Noise level of the system outputs.

Variable	Noise Level	Variable	Noise Level
1	0.075	8	0.299
2	6.688	9	0.378
3	8.151	10	0.130
4	0.148	11	0.248
5	0.408	12	0.289
6	3.056	13	12.100
7	3.011	14	10.101

The designed and applied signals can be seen in Figure 4.

With the identification test data set, a dynamic model, using the discrete output error algorithm of IDENT Toolbox (Ljung, 1989), was identified. The estimated model parameters were then converted to the continuos time domain. Figure 5 shows the comparison between the identified model and the identification test data set for two output variables for each signal block. Because the system has many output variables, only two variables were chosen to show the results.

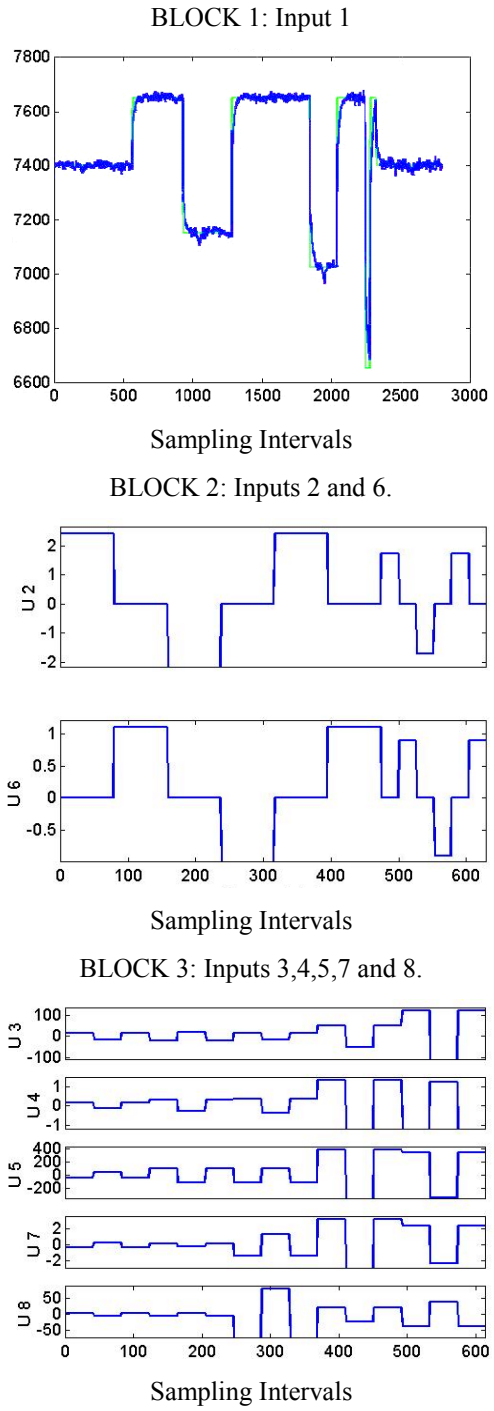

Fig. 4: Designed signals to model identification of the distillation unit.

It can be observed in Figure 5 that the obtained results were satisfactory. The identified model represents well the system behavior. It is important to mention that the process did not leave its normal operating point during the test, fact that can be verified by the plant data average and standard

deviation, shown in Table 2.

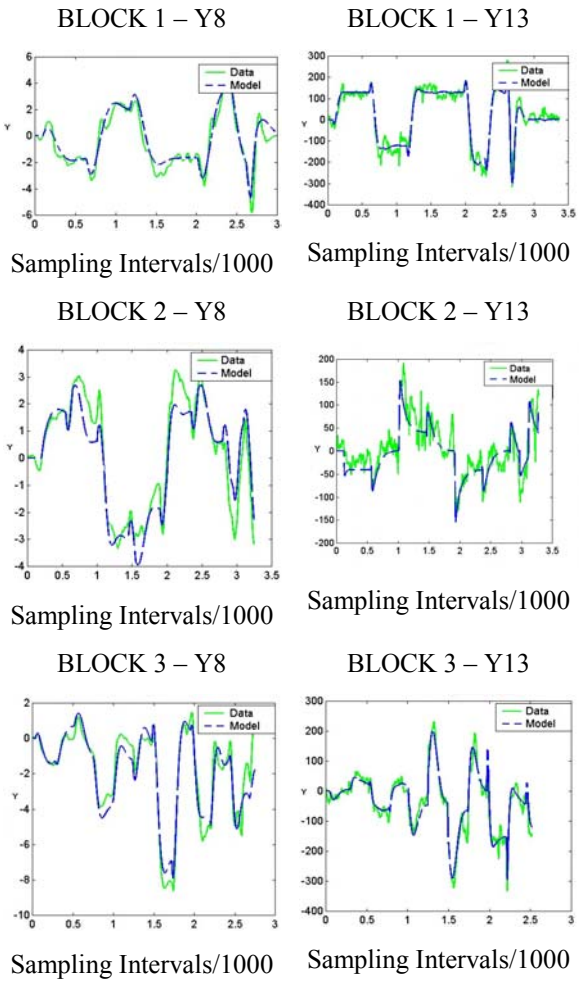

Fig. 5: Comparison between identified model and plant data for some output variables.

CONCLUSIONS

A new methodology to design signals for dynamic model identification was developed and applied to an industrial unit, generating good quality data without removes the plant from its normal operating point, at a low cost. The implemented methodology does a consistent and rational analysis of data sets contained in PIMS databases, to extract some important system information to guide a correct design of the identification signals. These signals are oriented to excite the system not only in the high gain direction but also in the low gain, to better determine the multivariable system gain.

ACKNOWLEDGMENT

The authors thank CAPES/FINEP and PETROBRAS for the financial support. In special, we thank RECAP (Capuava Refinary – São Paulo, Brazil) and its staff, by giving us the opportunity to test the proposed signal identification methodology there.

Table 2. Mean and standard deviation values of the plant data during the identification test.

Y	Mean			Deviation		
	Block 1	Block 2	Block 3	Block 1	Block 2	Block 3
1	144.07	142.64	143.80	0.55	3.01	0.51
2	758.86	753.91	783.37	17.17	62.95	12.32
3	1212.43	1252.62	1201.30	16.33	83.57	32.76
4	33.43	34.63	35.33	0.13	3.52	0.10
5	115.15	111.74	116.06	2.26	2.70	3.50
6	173.07	146.16	157.12	15.87	35.36	37.83
7	1874.53	2127.83	2227.72	106.47	118.07	92.07
8	46.35	42.53	41.86	2.06	2.85	2.39
9	249.02	235.88	227.05	4.33	3.21	4.19
10	54.03	53.03	56.48	1.92	3.91	2.56
11	-9.73	-10.05	-9.63	0.83	1.50	1.5
12	347.69	350.07	350.41	0.95	2.41	1.94
13	1839.43	1924.13	2035.30	130.03	72.47	85.50
14	3423.85	3441.15	3585.73	153.91	40.84	42.37

REFERENCES

Cutler, C.R.; Hawkins, R. B. Application of a large predictive multivariable controller to a hydrocracher second stage reactor, Proceedings of American Control Conference, 1988, 284-291.

Ljung, L. MATLAB®/IDENT-System Identification Toolbox (1989), The MathWorks.

Ljung, L. System Identification: *Theory for the User*, Prentice Hall, New Jersey. (1999)

Luyben, W. L., *Practical Distillation Control*, Van Nostrand Reinhold, New York, 1992.

Overschee, P. V., Bart D. M. *Subspace Identification for Linear Systems*. Kluwer Academic Publishers, Netherlands, 1996.

Tulleken, H. J. A. F., Generalized binary noise test-signal concept for improved identification-experiment design, Automatica, 26(1),1990, 37-49.