

**BEYOND STEP TESTING AND PROCESS
REACTION CURVES: INTRODUCING
MEANINGFUL SYSTEM IDENTIFICATION
CONCEPTS IN THE UNDERGRADUATE
CHEMICAL ENGINEERING CURRICULUM**

D.E. Rivera¹ and M.E. Flores

*Department of Chemical, Bio and Materials Engineering,
and*

Control Systems Engineering Laboratory, Manufacturing Institute

Arizona State University, Tempe, Arizona 85287-6006

phone:(480)-965-9476, email:daniel.rivera@asu.edu

Abstract: This paper describes efforts at Arizona State University to introduce substantive topics in system identification to undergraduate chemical engineering students. Specifically, the paper focuses on how system identification issues relevant to industrial practice have been incorporated into a simulated gas-oil furnace experiment that is part of a senior-level process dynamics and control course (ChE 461).

Keywords: education, identification, chemical process control

1. INTRODUCTION

The lack of training in system identification concepts is a problem faced by nearly every chemical engineering program in the United States. While system identification is routinely used in many process industries (notably the petrochemical and refining industries) to obtain models for control design, the extent of identification training received by the great majority of chemical engineering graduates consists of the classical step-testing/process reaction curve technique made popular by (Ziegler and Nichols, 1942) and (Cohen and Coon, 1953). In this approach, model parameters are graphically determined from the process response to a step change in the actuator.

While the classical step test/process reaction curve method is useful when the process is subject to low variance, high frequency noise, it fails miserably under conditions commonly found in industrial practice, notably in the presence of integrating disturbances and significant drifts. It is under these circumstances that advanced identification approaches produce the most benefits. A rigorous treatment of system identification concepts that can address these circumstances is beyond the scope of the undergraduate curriculum; however, there are still many identification topics that are well within the grasp of the chemical engineering undergraduate. Some of the topics that we have been able to introduce in the curriculum at ASU include:

¹ to whom all correspondence should be addressed

- (1) the use of linear model structures more advanced than the simple first-order deadtime model (e.g., ARX, Output Error structures),
- (2) input signal design and execution under noisy conditions using “plant-friendly” signals such as Pseudo-Random Binary Sequences (PRBS),
- (3) meaningful validation of the predictive ability of models using crossvalidation and other statistically-based techniques,
- (4) an introduction to modern computer-aided tools for system identification, notably graphical tools for identification relying on the System Identification toolbox in Matlab.

The application of these topics is done with a simulated gas oil furnace experiment (Figure 1) that is studied in the laboratory component of ChE 461: **Introduction to Process Dynamics and Control**. The simulation is implemented in real-time on an industrial-scale plant information and control system, a Honeywell TDC3000/TotalPlant distributed control system (Figure 2).

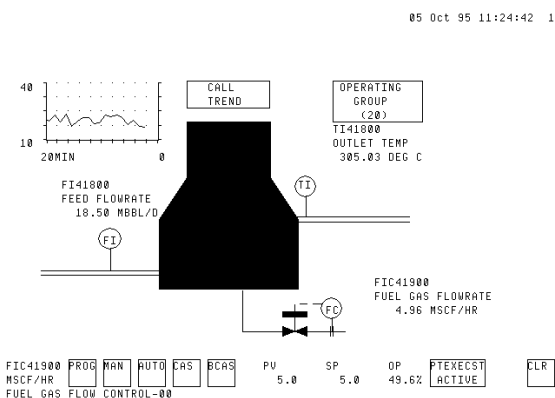


Fig. 1. Gasoil Furnace Schematic



Fig. 2. Universal Station nodes forming part of the Honeywell TDC3000 system at ASU's Control Systems Engineering Laboratory

2. A MOTIVATING PROBLEM

Teaching system identification concepts to chemical undergraduates is challenging because they lack the appropriate background. Many signal and systems fundamentals that are part of the basic educational training in other engineering disciplines (e.g., electrical engineering) are absent in the chemical engineering curriculum. Z -transforms, stationarity, autocorrelation, cross-correlation, power spectral density and cross-spectra are mostly foreign terms to the chemical engineering undergraduate. However, the typical ChE student possesses sufficient background in first-principles dynamic modeling, regression analysis, optimization, and statistics to understand important identification concepts, at least conceptually. In most situations, the only identification method

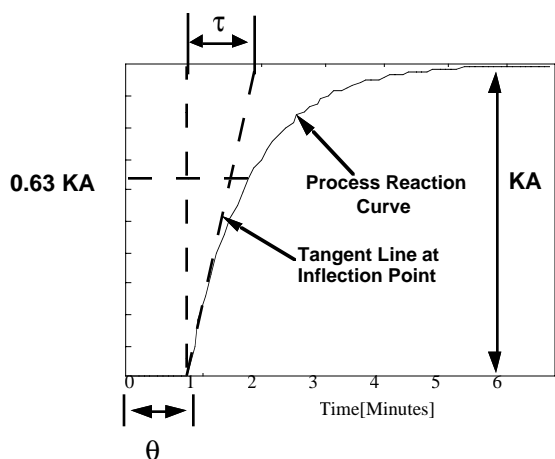


Fig. 3. Inflection point analysis on a process reaction curve from a step of magnitude A

that chemical engineering students are exposed to is inflection-point analysis on a process reaction curve (Figure 3) which consists of introducing a step input to the plant, and then using the output response (i.e., the process reaction curve) to graphically estimate the gain, delay, and time constant of a first-order with delay model:

$$\frac{y(s)}{u(s)} = p(s) = \frac{K e^{-\theta s}}{\tau s + 1} \quad (1)$$

This identification technique is very effective when the plant is under low-noise or disturbance-free conditions. However, in the presence of significant noise or drifting disturbances, graphical analysis may lead to misleading results.

A simple example can be used to demonstrate this point. The true model for this system consists of a first order with delay plant with parameters $K = 1$, $\tau = 10$ and $\theta = 5$ with sampling time $T = 1$. In an open-loop, uncontrolled setting, a potential drifting disturbance associated with this

plant can be characterized by an autoregressive, integrated model according to

$$\nu(t) = H(z)e(t)$$

$$H(z) = \frac{z^2}{(z - 0.91)(z - 1)}, \quad \sigma_e^2 = 0.0005 \quad (2)$$

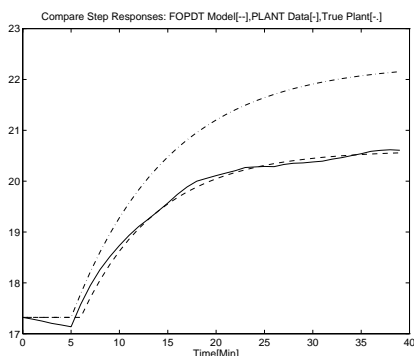


Fig. 4. Process reaction curve (solid) and comparison of the true plant step response (dashed-dotted) versus a graphical fit (dashed).

Imagine that a process reaction curve is generated for this system by applying a +5 magnitude step input to the manipulated variable. One possible graphical fit to the response is shown in Figure 4. The model parameters corresponding to this curvefit are $K_{est} = 0.651$, $\tau_{est} = 7.425$ and $\theta_{est} = 6$. Compared to the true plant, the estimated plant has errors between 20% and 35% in all its parameters, which in turn would require detuning of the subsequent control system. An erroneous model is obtained, even though it appears to be a “near-perfect” fit to the data.

The problem can be counteracted in two ways. One is to increase magnitude of the step input change such that the process output substantially exceeds the magnitude of the noise or disturbances. A step input which satisfies these conditions may be too large to be tolerated by operating personnel and therefore deemed “plant hostile.” A second alternative is to apply multiple steps, or preferably, to rely on identification signals that operate under low signal to noise ratios, such as pseudorandom binary signals (PRBS) (Godfrey, 1993). The use of a PRBS sequence enables “plant-friendly” operation, since the input signal can now be implemented over lower signal-to-noise ratios. However, the use of PRBS data calls for more sophisticated analysis techniques than graphical estimation. Detailed treatment of the multitude of techniques that can be used to solve this problem is clearly beyond the scope of an undergraduate course. Nonetheless, a structured approach can be pursued, which is discussed in the ensuing section.

3. METHODOLOGY

A summary of the identification procedure taught in ChE 461 is found in Figure 5. The methodology accepts a data from a wide number of inputs (either step, pulse, or PRBS). The appropriateness of a particular input signal choice used can be justified intuitively on the basis of the asymptotic variance expression:

$$\text{Model Parameter Variance} \sim \frac{n}{N} \frac{\lambda}{\mu} \quad (3)$$

where n is the number of model parameters, N is the number of data points (length of data set), λ is the variance of the disturbance signal ν , μ is the variance of the input signal u . μ/λ , the input signal-to-noise ratio, influences the variability in the estimated model. High noise conditions require either long data sets, high input magnitudes, or both; in the absence of noise, a simple step or pulse test will suffice.

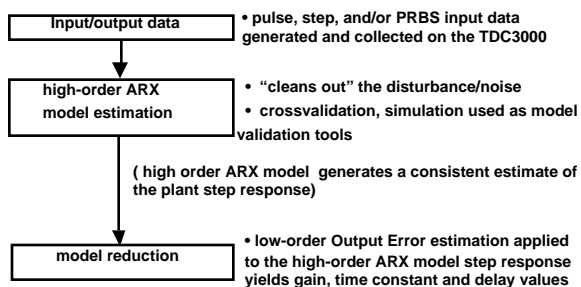


Fig. 5. Ident. procedure taught in ChE 461

To deal with noisy data measurements, a model structure is required which can “clean up” the noise in the data and lead to a model estimate with a higher signal-to-noise ratio. For this purpose, AutoRegressive with eXternal (ARX) models are used; these are conveniently represented by the difference equation

$$y(t) + a_1y(t-1) + \dots + a_{n_a}y(t-n_a) \quad (4)$$

$$= b_1u(t-nk) + \dots + b_{n_b}u(t-nk-n_b+1) + e(t)$$

where n_a is the order of the AR polynomial, n_b is the order of the eXternal input polynomial, and nk is the delay (expressed as an integer multiple of sampling time T).

Prediction-error estimation according to (4) constitutes a multidimensional linear regression problem, a technique the students are well acquainted with. *Consistent* (i.e., bias-free) parameter estimation is obtained when the following are true:

- (1) *The selected model structure is of “sufficiently high” order to captures the true plant dynamics.* This requires adequate specification of n_a , n_b , and nk in the ARX model structure.

- (2) *u* shows persistent excitation. The input signal *u* must exhibit variation over “enough” frequencies.

While “sufficiently high” and “enough” are formally defined in the identification literature, the undergraduate student can still grasp these general ideas and use this working knowledge in the laboratory.

The selection of the ARX model structure can be accomplished without substantial user intervention. Because ARX estimation consists of solving a linear regression problem, it can be applied exhaustively to a large number of model structures without demanding significant computational time. Plotting the loss function (the normalized sum of squared prediction errors) versus the number of model parameters over a crossvalidation data set (i.e., a data set different than the one used for estimation) can help pinpoint an “optimal” model structure with little need for iteration. If a crossvalidation data set is not available, a parsimonious model order can be determined via other measures, such as the Akaike Information Theoretic Criterion (AIC) or Rissanen’s minimum description length (MDL) principle (Ljung, 1999).

The final step in the methodology is to estimate parameters of a *continuous* first-order with delay model from the step response of the validated ARX model. This is accomplished by estimating an Output Error model with structure $[nb = 1 \quad nf = 1 \quad nk]$ and difference equation

$$w(t) = -f_1 w(t-1) + b_1 u(t-nk) \quad (5)$$

$$y(t) = w(t) + e(t) \quad (6)$$

which can be related a first-order with delay model per Equation 1 according to:

$$K = \frac{b_1}{(1-f_1)} \quad \tau = -\frac{T}{\log(f_1)} \quad \theta = nk * T$$

The values for *K*, τ and θ can now be used in model-based PID tuning rules that are applied in future experiments.

The combined methodology is carried out using a software package entitled *pIDfurn*. *pIDfurn* is a modified version of a commercial package entitled *pIDtune* developed by the authors. It is featured in the SYSID 2000 software demonstration sessions (Flores and Rivera, 2000).

4. GAS-OIL FURNACE EXPERIMENT

A principal objective of the laboratory portion of ChE 461 is to provide students with an appreciation for the diversity of tasks that must be carried out by practicing control engineers when building a process control system (Rivera *et*

al., 1996). A simulated gas/oil furnace (schematic shown in Figure 1) is used for this purpose. The goal of the control system is to maintain the outlet temperature of a gas oil stream at setpoint by manipulating fuel gas flow to the furnace. Disturbances enter the system via changes in the feed flowrate or as noise on the temperature; these can be specified as either random and deterministic in nature. The temperature dynamics are modeled by second-order with delay transfer functions:

$$y(s) = \frac{K_p e^{-\theta_p s}}{\tau_{1p}^2 s^2 + 2\tau_{1p}\zeta_p s + 1} u(s) + \frac{K_d e^{-\theta_d s}}{\tau_{1d}^2 s^2 + 2\tau_{1d}\zeta_d s + 1} d'(s) + \nu(s) \quad (7)$$

where *y* is the change in the outlet temperature, *u* is the change in the fuel gas flow, *d'* is the change in the feed flowrate, and ν represents an output disturbance.

The tasks carried out by students on this simulated system span three laboratory sessions; one session (Lab *D*) is devoted to identification, while two sessions (Labs *J1* and *J2*) are devoted to control design and implementation issues. The sequence of identification tasks carried out by students in Lab *D* is as follows:

- (1) Step testing under low noise conditions,
- (2) Step testing under drift conditions,
- (3) PRBS testing under drift conditions.

We examine each of these cases in the ensuing discussion.

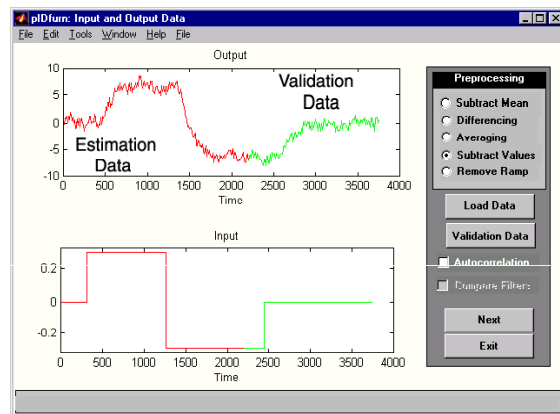


Fig. 6. Furnace “double pulse” data under low noise conditions.

4.1 Identification Analysis with Low Noise

The initial task of step testing is performed by introducing a small ($\pm 3.0\%$) change from the nominal value of 5.0 MSCF/hr in the fuel gas flowrate. Students are asked to visually ascertain if the test has been sufficiently informative; a

−6% change (from 5.3 MSCF/hr to 4.7 MSCF/hr) in the fuel gas flow is then introduced. A final +3% step change brings the process to its original operating condition. The result of this series of step tests is equivalent to a “double pulse” input, as seen in Figure 6.

The identification data is collected in the TDC3000 and transferred via floppy to a personal computer, where *pIDfurn* is used. The data is divided into estimation and validation data sets, and ARX estimation is performed for a range of values for n_a , n_b , and n_k and reported as seen in Figure 7. Each asterisk corresponds to a unique model structure. The student can graphically inspect the choices and select the structure leading to the lowest loss function value on the validation data set. Having obtained the optimal structure, the student can examine simulations of the best fit model on the estimation and validation data sets, and step responses. The ARX model step response is then fit via Output Error estimation to a first-order with deadtime model, as described previously. In this step, the only adjustable parameter is the value of the delay, which the student is asked to adjust until a best fit is obtained (Figure 8).

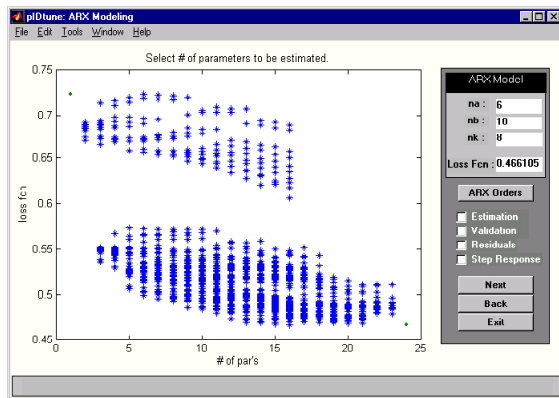


Fig. 7. ARX estimation results, low noise conditions

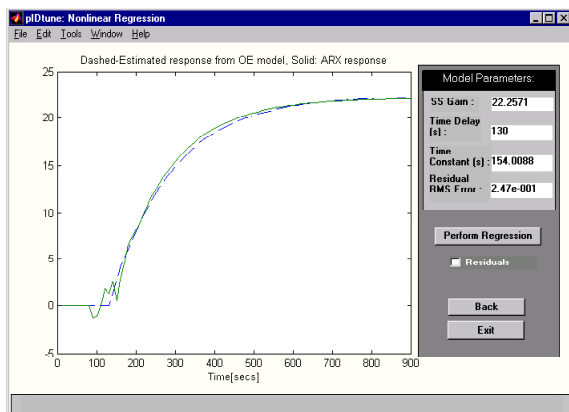


Fig. 8. ARX model step response (solid) and first-order deadtime OE estimate (dashed)

4.2 Identification Analysis with Drift using Double Pulse Input

Identification under low noise conditions is not challenging, but allows the student to gain some familiarity with the *pIDfurn* software and the general principles of ARX and OE estimation and validation. The problem becomes much more challenging when a drifting (integrated) disturbance $\nu(t)$ similar to Equation (2) is introduced. A double pulse data set under these circumstances is shown in Figure 9. Students rapidly experience a myriad of problems brought about by the drift in the data and are motivated to explore the option of PRBS testing under these conditions.

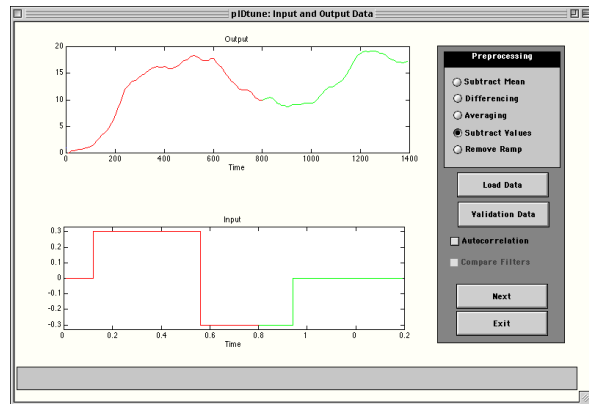


Fig. 9. “Double pulse” data under drift conditions.

4.3 Identification Analysis with Drift using PRBS Testing

Under drifting conditions, the variance analysis per Equation 3 indicates that either higher input magnitudes or a longer data set are needed; students are asked to pursue the latter option using a PRBS input because of its “plant-friendly” properties. The design variables for the PRBS input need to be specified, and these must come from *a priori* information obtained by the user. Following the analysis in (Gaikwad and Rivera, 1996), the guidelines for specifying T_{sw} (switching time) and n_r (number of shift registers) are used by the students:

$$T_{sw} \leq \frac{2.78\tau_{dom}^L}{\alpha_s} \quad N_s = 2^{n_r} - 1 \geq \frac{2\pi\beta_s\tau_{dom}^H}{T_{sw}}$$

n_r and N_s must be integer values, while T_{sw} must be an integer multiple of the sampling time T . τ_{dom}^H and τ_{dom}^L are high and low estimates of the dominant time constant, β_s is an integer factor representing the settling time of the process (e.g., for $T_{95\%}$, $\beta_s = 3$; for $T_{99\%}$, $\beta_s = 5$, etc.), while α_s is a factor representing the closed-loop speed of response, written as a multiple of the open-loop response time. Students are asked to use the

pIDfurn results from low noise testing as a basis for specifying this information.

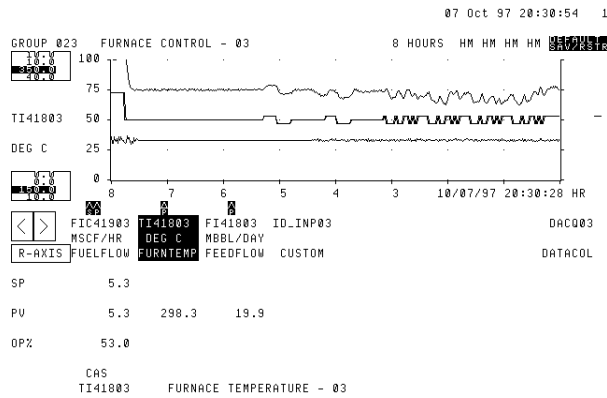


Fig. 10. 8 hr TDC3000 trend figure showing low noise step, high noise testing, and PRBS testing.

The PRBS sequence is allowed to run for enough cycles to cover an eight hour shift; a TDC3000 trend showing the sequence of low noise step, high noise step, and PRBS testing is shown in Figure 10. The PRBS data is imported into *pIDfurn*, as shown in Figure 11. Differencing is used to achieve stationarity in the data, and then equal number of PRBS cycles are used for estimation and validation (Figure 12). The sequence of ARX estimation, validation, and subsequent fit to an Output Error model follows as before; note the estimated ARX model compared against the OE model is shown in Figure 13. A excellent result (comparing favorably with the low noise modeling results) is obtained.

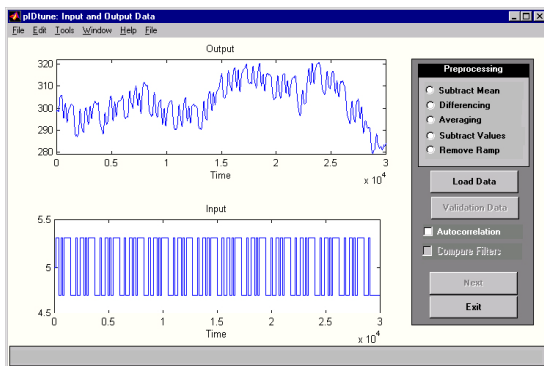


Fig. 11. PRBS data imported into *pIDfurn*

5. REFERENCES

Cohen, G.H. and G.A. Coon (1953). Theoretical considerations of retarded control. *Trans. ASME* **75**, 827.

Flores, M. E. and D.E. Rivera (2000). *pIDtune*: A graphical package for integrated system

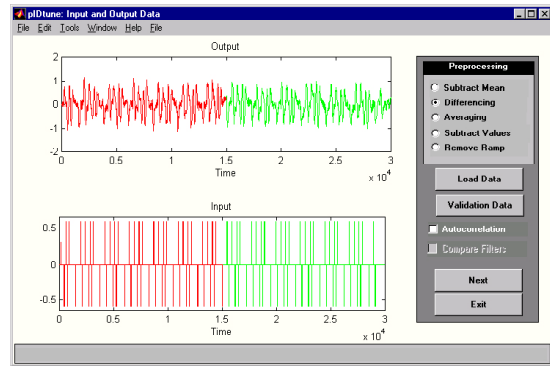


Fig. 12. PRBS data after differencing

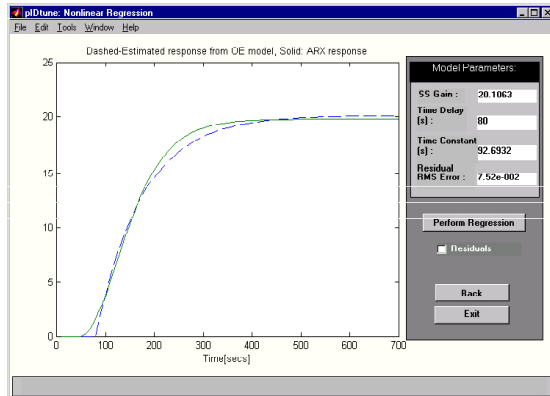


Fig. 13. ARX (solid) and OE (dashed) model step responses, PRBS data identification and PID controller design. In: *IFAC SYSID 2000*.

Gaikwad, S.V. and D.E. Rivera (1996). Control-relevant input signal design for multivariable system identification: application to high-purity distillation. *Proc. IFAC World Congress, San Francisco Vol. M*, 349–354.

Godfrey, K. (1993). *Perturbation signals for system identification*. Prentice-Hall. New York.

Ljung, L. (1999). *System Identification: Theory for the user, 2nd. Edition*. Prentice-Hall. New Jersey.

Rivera, D.E., K.S. Jun, V.E. Sater and M.K. Shetty (1996). Teaching process dynamics and control using an industrial-scale real-time computing environment. *Comp. Appl. in Eng. Education* **4**, 191.

Ziegler, J.G. and N. B. Nichols (1942). Optimum settings for automatic controllers. *Trans. ASME* **64**, 759.