Identification for Control Purpose by Relay Techniques: Achievable Performance versus Complexity

G. Marchetti, C. Scali (*)

Dipartimento di Ingegneria Chimica, Università di Pisa - Via Diotisalvi, 56126, Pisa - Italy.

Abstract

In the paper two relay techniques are compared in terms of ease of application, duration of experimental tests and achievable performance. The first one is the Two Channel Relay (Friman and Waller, 1997) and makes use of two relays in parallel (one augmented by an integrator). The second one is the ATV+ technique (recently introduced by Scali et al. (1999), as a modification of the ATV technique by Li et al., 1991) and introduces an additional delay in the process. The two techniques are briefly recalled, putting into evidence their main features. Different indexes are introduced for a quantitative evaluation of the duration of tests and of achievable performance. From the comparison of results for a large number of sample processes, it can be concluded that the TCR is very effective, as it requires shorter times for the experimental tests and it allows to achieve reasonable performance for many cases. On the other side, the ATV+ requires about twice longer experimental times, but, being coupled with a an appropriate design of the controllers, allows to achieve better performance for all the cases with PI/PID controllers and further improvement with model based controllers.

Keywords: Relay feedback, Identification, Autotuning, PID control.

1.Introduction

Relay techniques are very appealing to perform identification of chemical processes for control purposes; being fast and easy to use, they can be frequently repeated in order to perform an autotuning of PI/PID controllers. Therefore for all the situations where there is lack of knowledge about a chemical process to build a mathematical model, or this is available but too complicated or non-linear, the procedure of controller design and tuning can be greatly simplified through an experimental identification of the process.

As illustrated by Åström and Hägglund (1984), bringing the system to stable oscillations under relay feedback gives the essential information to build a low order model and to perform the design of the controller. The main drawback of this technique is that when the knowledge is limited to the critical point, only the Ziegler-Nichols (1943) tuning rules can be adopted, but these settings do not guarantee stability even for simple classes of processes. Refinements of the tuning rules have been proposed by several authors but they do not provide a general solution to the problem.

Some more information about the process other than the position of the critical point should be obtained and, to achieve this goal, the relay test can be modified in several ways. Among them: the Two Channel Relay (Friman and Waller, 1997) and the ATV+ technique (Scali et al. 1999), briefly recalled and then compared in terms of ease of application, duration of tests and achievable performance.

2.Identification with the Two Channel Relay

From a simple relay test we obtain the knowledge of a point located on the negative real axis of the Nyquist plane. With a PI controller it is therefore impossible to specify a phase margin, because the

(*) e-mail: scali@ing.unipi.it

identified point is shifted towards the second quadrant as a result of the integral action. This obstacle can be removed with the identification of a point located in the third quadrant, at a certain angular distance from the negative real axis. To identify this point several modified relay identification techniques exist: we will concentrate on the identification via Two Channel Relay (Friman and Waller, 1997). The Two Channel Relay (TCR) is a device composed by two relays connected in parallel. Before one of the two relays an integrator is placed (as shown in figure 1-a), so that the whole structure is characterized by a negative phase angle.

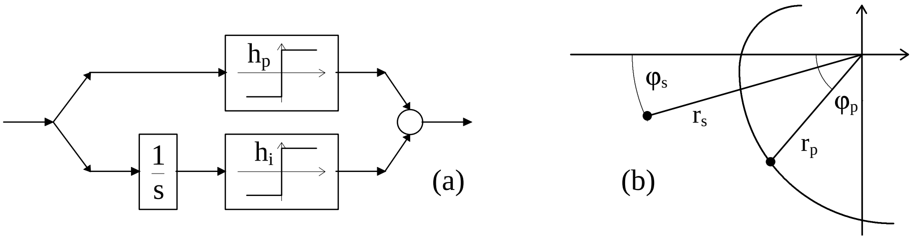

Figure 1: Structure of the Two Channel Relay (a) and position of the identified point (b)

The TCR is characterized by a phase which varies between zero and $-\pi/2$ depending on the amplitudes of the two relays. When the system *relay + process* reaches limit-cycle conditions the overall phase angle is $-\pi$ radians, and this means that the point representing the process has a phase between $-\pi/2$ and $-\pi$ (so it lies in the third quadrant). The tuning of the controller is accomplished simply by imposing that the identified point (whose coordinates are r_p and φ_p, see figure 1-b) is moved to a position with a specified gain and phase margin (new coordinates r_s and φ_s).

Friman and Waller (1997) suggest default values of amplitudes and angles for completely unknown processes, but they also propose a refinement of the technique when the high frequency behaviour of the process (strong or weak attenuation) is known. Processes with a weak high-frequency attenuation are conventionally referred to as “group 1 processes”, whereas group 2 contains the processes characterized by a strong attenuation. The recommended settings are shown in table 1.

	φ_p	φ_s	r_s
Default	30	15	½
Group 1	45	0	1/3
Group 2	0	60	1/3

Table 1: Summary of parameters for the Two Channel Relay identification

3.ATV$^+$ identification

The main limitation of the relay identification is that it only provides the knowledge about the critical point of the process: this piece of information is unfortunately not enough for a good tuning of the controller. The knowledge obtained from the relay test could be used to build a parametric model, following the ATV technique (Li et al., 1991). The original ATV technique cannot however be applied to completely unknown processes, because it requires the a priori knowledge of the process dead time. If this is not available the new ATV$^+$ identification technique must be adopted. This will be briefly described in the follow, pointing out differences and similarities with the original method.

The first step of the procedure is a standard relay test, which provides the critical parameters of the process. The identification of other points of the Nyquist curve is made possible by introducing a delay element between the relay and the process, as shown in figure 2.

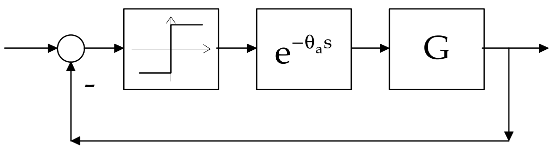

Figure 2: Setup of relay with additional delay (ATV).

When the system reaches stable oscillations the overall phase shift is again 180°, but only a part of this is attributable to the process, as the additional delay is also characterized by a phase lag. This last contribution is known and so the position of the identified point can be easily calculated:

$$G(i\omega'_u) = \frac{1}{k'_u} \cdot \exp(-i\pi + \theta_a \cdot i\omega'_u) = -\frac{1}{k'_u} \cdot \exp(\theta_a \cdot i\omega'_u)$$

In the previous expression k'_u and ω'_u represent the critical parameters obtained from the execution of the modified test. For the identification of additional points the values of θ_a are chosen so that they correspond to phase lags of the delay element of ≈45° and ≈75° respectively. Once the coordinates of the identified points are known, the type of model must be chosen, among the six candidates listed below:

$$G_1(s) = \frac{b_0 \cdot e^{-\theta s}}{a_1 s + 1} \qquad G_2(s) = \frac{b_0 \cdot e^{-\theta s}}{a_2 s^2 + a_1 s + 1} \qquad G_3(s) = \frac{b_0 \cdot e^{-\theta s}}{a_3 s^3 + a_2 s^2 + a_1 s + 1}$$

$$G_4(s) = \frac{(b_1 s + b_0) \cdot e^{-\theta s}}{a_1 s + 1} \qquad G_5(s) = \frac{(b_1 s + b_0) \cdot e^{-\theta s}}{a_2 s^2 + a_1 s + 1} \qquad G_6(s) = \frac{(b_1 s + b_0) \cdot e^{-\theta s}}{a_3 s^3 + a_2 s^2 + a_1 s + 1}$$

The models are characterized by a maximum of six parameters (b_1, b_0, a_3, a_2, a_1 and θ), because it has been assumed that in the identification three relay tests are executed and, as it will now be shown, each relay test supplies two equations. The method is explained for model 1, as the extension to the higher order models is obvious. Knowing the position of the identified point an equation can be written as:

$$G(i\omega) = \frac{b_0 \left(\cos(\omega_j \theta) - i \cdot \sin(\omega_j \theta)\right)}{i \cdot a_1 \omega_j + 1} = X_j + i \cdot Y_j$$

where the suffix j is 1,2 or 3 depending on the test considered. By separating the real and imaginary parts the previous gives two equations:

$$a_1 \omega_j Y_j + b_0 \cos(\omega_j \theta) = X_j \qquad\qquad -a_1 \omega_j X_j - b_0 \sin(\omega_j \theta) = Y_j$$

Repeating the procedure for the three points we obtain a system of six equations in the two unknowns a_1 and b_0. If the dead time θ is known (this is the case of the ATV identification) the system is linear, so it can be written in matrix form and solved in a least-squares sense through a pseudo-inversion. When the delay is not known the system must be solved for several trial values of θ. The trial values are chosen between zero and a maximum possible value which is calculated from the frequencies measured in the relay tests. As the process under examination is assumed to be open loop stable, the values of θ for which the resulting model is unstable are discarded, and the remaining ones are compared on the basis of the value of the residual square error evaluated after the solution of the system. The value of θ is chosen for which the residual error has its minimum, and this calculation is repeated for all the candidate models. It is finally possible to choose the model which best suits the experimental data by comparing again the value of the residual error. In a previous work (Scali, Marchetti and Semino, 1998), it has been shown that the new method can be considered equivalent to the previous one, concerning the accuracy of the identified models and the control performance obtainable with the described tuning technique. Therefore only the ATV^+ method, which allows to deal with completely unknown processes, will be considered in the sequel.

The controller can be tuned with a technique which is a generalisation of that proposed in Scali and Semino, 1998. The tuning procedure provides tuning parameters for conventional (PI or PID controller). The procedure is written for the most general case of model G_6, the steps involved are the following:

- Build the inverse-based IMC controller for step inputs:

$$q(s) = \frac{a_3 s^3 + a_2 s^2 + a_1 s + 1}{(\lambda s + 1)^2 \left(|b_1| s + b_0\right)}$$

- Calculate the structure of the equivalent feedback controller;

$$C(s) = \frac{q(s)}{1 - q(s)G(s)} = \frac{a_3 s^3 + a_2 s^2 + a_1 s + 1}{(\lambda s + 1)^2 \left(|b_1| s + b_0 \right) - e^{-\theta s} \cdot \left(b_1 s + b_0 \right)}$$

- Eliminate the exponential term due to the process delay by using a first order Padè approximation:

$$C'(s) = \frac{\left(a_3 s^3 + a_2 s^2 + a_1 s + 1 \right)(1 + \theta s / 2)}{(\lambda s + 1)^2 \left(|b_1| s + b_0 \right)(1 + \theta s / 2) - (1 - \theta s / 2) \cdot \left(b_1 s + b_0 \right)}$$

- Reduce the structure to that of a PID controller by neglecting terms of order higher than s^2 in the numerator, than s in the denominator:

$$C_{PID}(s) = \frac{\left(a_2 + a_1 \theta / 2 \right) s^2 + \left(a_1 + \theta / 2 \right) s + 1}{\left(2\lambda b_0 + 2 b_1 + \theta b_0 \right) s}$$

- The controller parameters have thus the following expression:

$$K_c = \frac{\left(a_1 + \theta / 2 \right)}{\left(2\lambda b_0 + 2 b_1 + \theta b_0 \right)} \qquad \tau_I = a_1 + \frac{\theta}{2} \qquad \tau_D = \frac{\left(a_2 + a_1 \theta / 2 \right)}{\left(a_1 + \theta / 2 \right)}$$

with $a_2 = 0$ for first order processes;

- The integral and derivative time are readily obtained from the previous expressions. The controller gain is not determined by choosing an appropriate value for the filter constant λ, but by imposing that the closed loop resonance peak has a value of 1.26 (2 dB):

$$\max_{\omega} \left| \frac{G(i\omega)C(i\omega)}{1 + G(i\omega)C(i\omega)} \right| = 2 \text{ dB}$$

It must be noted that for the same process both the PI and the PID controllers can be designed. The controller gains will be different, because the structure of $C(i\omega)$ that appears in the sensitivity function is also different.

4. Comparison between the proposed techniques

The proposed identification methods have been applied to a group of 45 sample processes, described by transfer functions up to six parameters, which include third order dynamics, overdamped and underdamped, with dead time and inverse response elements. The comparison will be carried out by analyzing the ease of identification, the duration of tests, and the achievable performance.

4.1. Ease of identification

For each of the 45 processes three different identifications have been performed:

1. ATV^+ identification, by means of a relay with additional delay ;
2. TCR identification, assuming that the processes were completely unknown and therefore adopting the “default” settings for the relay and the controller ;
3. TCR identification, using the appropriate (group 1-2) settings for the process under examination. Following the guidelines and the examples given in (Friman and Waller, 1997) the high-frequency attenuation has been evaluated examining the following index:

$$AI = \frac{K_p}{\left| G\left(i\omega_u \right) \right|} = K_p \cdot K_u$$

   where K_u and ω_u are the critical gain and frequency and K_p is the process steady-state gain. A process is considered to have a strong high-frequency attenuation (so it belongs to group 2) when it is characterized by a value of AI greater than 10. The value of AI was calculated for each of the processes, in order to provide the additional knowledge concerning the high-frequency behaviour: the identification was then carried out with the specific settings (group 1 or group 2) for the process under examination. This case will be later referred to as TCR2 identification.

The first difference between the techniques which emerges is that the ATV^+ method can be successfully applied to all the 45 processes examined, whereas the identification via TCR fails in some cases (namely 5 for TCR and 6 for TCR2), all dominated by large dead times or by the numerator dynamics. In these situations the oscillations induced by the relay do not become stable even after a very long time. An example of this behaviour is reported in figure 3.

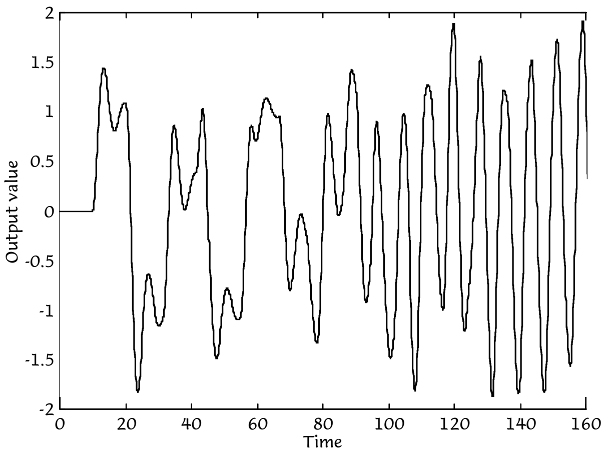

Figure 3: Example of a case in which the identification test does not lead to stable oscillations

4.2.Duration of tests

The attention has been focused on the time needed for the running of the relay tests. Every relay test was stopped only when the system had produced two complete stable cycles, to make sure that the transient was over and that the oscillations were fully developed. The first aspect considered is the influence of the transient on the total running time of the experiment. An index was defined to quantify this influence:

$$n_c = \frac{\text{total duration of test}}{\text{period of stable oscillation}}$$

Since each test contains at least the 2 last stable cycles, the amount by which n_c exceeds 2 can be seen as a measure of the length of the transient. The average values of n_c for all the processes are shown in table 2.

ATV^+			TCR	TCR2	
1st test	2nd test	3rd test		group 1	group 2
3.26	2.44	2.30	4.62	3.74	3.43

Table 2: Values of the n_c index for ATV^+ and TCR identification

For the ATV^+ method the index is calculated for each of the three relay tests. It can be seen that the transitory phase is considerable for the first test, but very short for the successive ones. This behaviour can be understood considering that the system is already oscillating when the additional delay is inserted or modified, and the system requires only a short time to adapt to the new conditions. On the contrary, at the beginning of the first test the system must bring up oscillations from a steady state condition.

The value of the index n_c for the TCR identification is considerably higher than those measured for the ATV^+ technique. This is probably due to the presence of the integrator, because the integral of the error usually needs a longer time to develop the stable oscillations. For the TCR2 case it must be

pointed out that only 5 out of 45 processes turned out to belong to group 2, and were therefore identified with the standard relay test (see table 1). The remaining 40 processes have been identified according to group 1 settings, that is at an angular distance of 45° to the negative real axis. In this case the height of the integral relay is greater than in the default case (identified point lies only 30° away from real axis), and it is therefore quite surprising to find out that the value of n_c is lower.
Figure 4 shows the output variations during the relay tests for one sample process. The first plot refers to the ATV^+ identification (the small circles indicate the time when the additional delay is switched), and the second one to an identification test via TCR. From figure 4 it can be seen that for ATV^+ the first cycle can be considered stable after a time of about 60, the second after about 130, and the third one after 220; for TCR the cycle is fully developed after about 140. Therefore the ATV^+ tests are characterized by shorter transients.

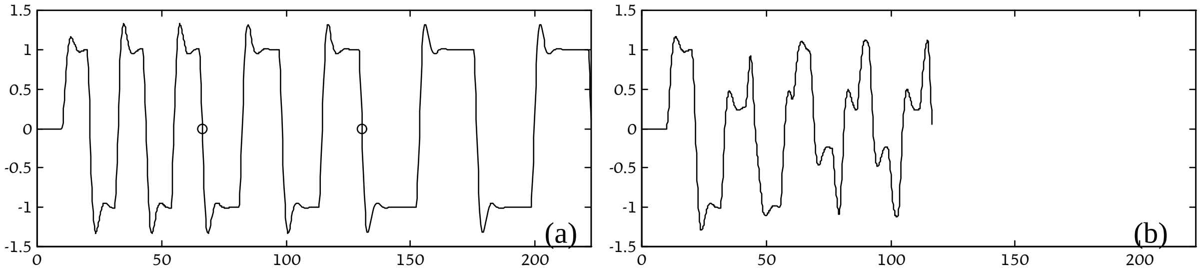

Figure 4: Example simulations of relay tests for ATV^+ (a) and TCR (b) identification

To examine the total duration of the identification procedure the index n_t has been defined as follows:

$$n_t = \frac{\text{total duration of identification}}{\text{duration of standard relay test}}$$

The value that appears in the denominator can be seen, for each process, as the shortest time needed for a relay identification and gives an idea of how more time-consuming are the improved techniques. The average values of n_t are shown in table 3.

ATV⁺	**TCR**	**TCR2**	
all tests		group 1	group 2
3.72	1.87	1.77	1.00

Table 3: Total duration of identification procedures

The ATV^+ requires clearly a longer time, but it must be remembered that three relay tests are accomplished, whereas for the TCR only one test is executed. For the TCR2-group 1 it can be seen again that the duration is less than for the default case. Obviously, for group 2 processes the duration is exactly the same as the normal Åström-Hägglund relay test.

4.3.Achievable performance

To evaluate the performance of the regulators designed after the identification the DISE index has been defined as follows:

$$\mathrm{DISE} = 100 \cdot \frac{\int_0^\infty e_{con}^2 \, dt - \int_0^\infty e_{ref}^2 \, dt}{\int_0^\infty e_{ref}^2 \, dt}$$

This index expresses the percentage differences in ISE (Integral of Square Error), and can be used more readily than the simple ISE value to evaluate performance differences between different controllers. In particular, negative values of the index mean that the specified controller performs better than the reference one, positive values mean that the reference controller works better. The comparison has been carried out between 4 type of controllers:

1. the PI controller based on the model obtained via ATV^+ identification;
2. the PID controller based on the model obtained via ATV^+ identification;
3. the PI controller designed as a result of the default TCR identification;
4. the PI/PID controller designed as a result of the group-specific TCR identification;

The ATV^+/PI controller has been chosen as the reference controller and the differences in performance in the other three cases have been expressed in terms of the DISE index. The performance of the controller has been studied for the two cases of set-point variation and rejection of a disturbance entering the system upstream of he process. The average values of DISE are reported in table 4.

	Set-point variation	**Disturbance rejection**
ATV^+/PID	-9.10	-23.71
TCR	37.21	23.00
TCR2	36.37	26.83

Table 4: Average values of the DISE index

It must be noted that the ATV^+/PI does not appear in the previous table because, being the reference controller, it would have all values of DISE equal to zero. The ATV^+ controllers appear to be superior to those obtained by means of the TCR identification, as it can be seen by the high values that appear in the last two rows of the table. This result is not surprising, if we consider that the latter are based only on the knowledge of one point of the Nyquist curve, whereas the former exploits the knowledge of a parametric model. Also to note is that, there is not great difference between the results obtained in the TCR cases with the default identification and the group-specific identification. From table 4 it can also be seen that the ATV^+/PID controller is characterized by negative values of the DISE index, and this means that it has an average performance which is better than that of the respective PI. It must however be pointed out that there are some cases in which the PID has a slower or very oscillating response, and even a case in which the resulting controller produces an unstable behaviour. However, this is a particularly “difficult” process, characterized by a dead time which is 10 times the dominant time constant and a damping factor of 0.25.

Figure 5 shows the responses of four processes to a set-point variation with the different controllers considered above: the first row of plots refers to the ATV^+ controllers (solid: PI, dashed: PID), the second row to the TCR controllers (solid: default PI, dashed: group specific PI/PID). Figure 6 shows the corresponding cases for rejection of a disturbance. By examining the plots on the first row it can be seen that the PID controller leads to a slight improvement of the achievable performance, especially in the case of disturbance suppression. From the second row of plots one can see that

Making a comparison between the plots on different rows, it can be seen that the ATV^+ controllers are generally characterized by a faster response, and reach asymptotic value in a shorter time; in some cases the response can be more oscillating in the beginning.

Further advantages for the ATV^+ procedure can be pointed out in the case of adopting a model based design for the controller, which allow to exploit the larger knowledge about the process obtained by the three points identification. This will be the object of next research work.

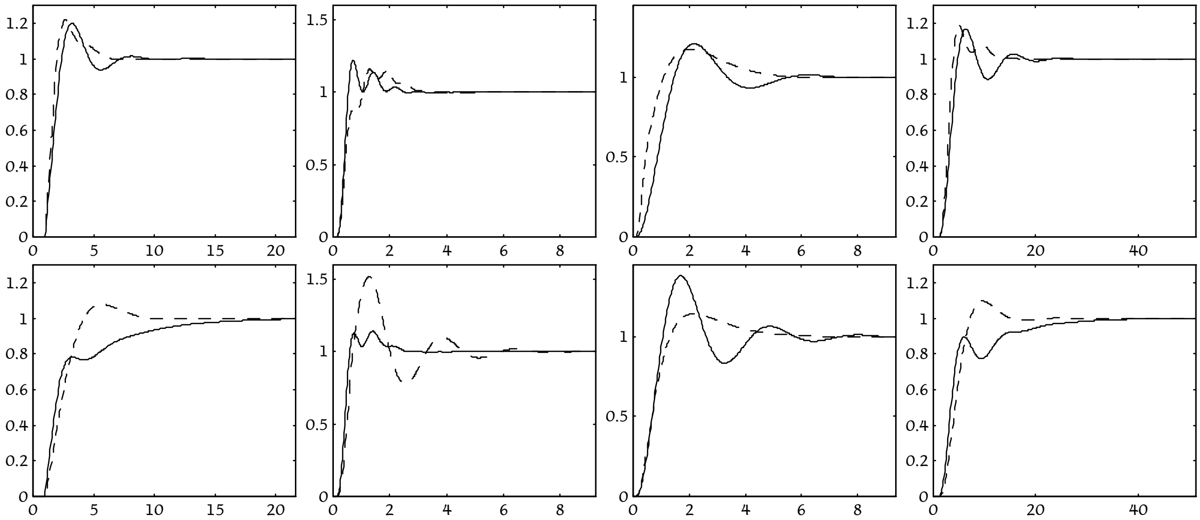

Figure 5: Simulation examples on four processes (set-point variation)

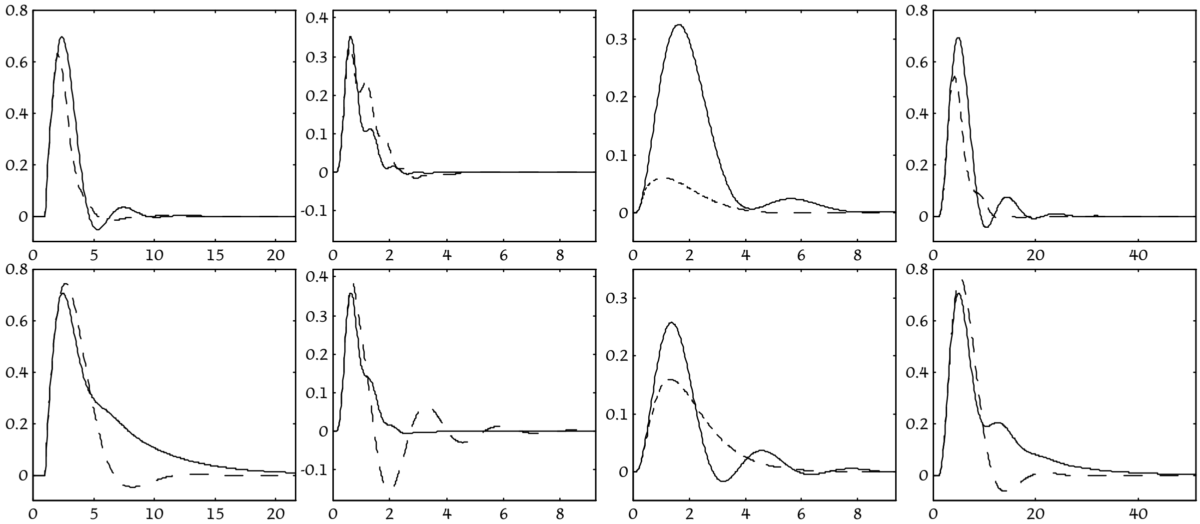

Figure 6: Simulation examples on four processes (disturbance rejection)

A design rule can finally be defined, which allows to choose between PI and PID, once the model has been identified. It has been found that a relatively good classification of the processes can be made according to the value of the Delay Dominance Index (DDI) defined as follows:

$$\mathrm{DDI} = \frac{\theta \cdot \omega_u}{\pi}$$

The value of DDI shows how much of the phase lag at the crossover frequency derives from the process dead time, and thus varies in a range from 0 to 1. Figure 7 is obtained by plotting the average values of the DISE index (between set-point variation and disturbance suppression) for the 45 examined processes versus the values of DDI. It can be seen quite clearly that, except for very few cases, the performance of the PID controller steadily improves as the value of DDI is reduced. This suggests that for processes characterized by high values of DDI (delay dominant) a PI regulator should be preferred, because no great improvement is obtained by the adoption of a PID structure. Conversely, for processes with low DDI (lag-dominant) the PID controller is the best choice, because the performance obtainable with the PI can be largely improved, with reduction of ISE up to 50% or more.

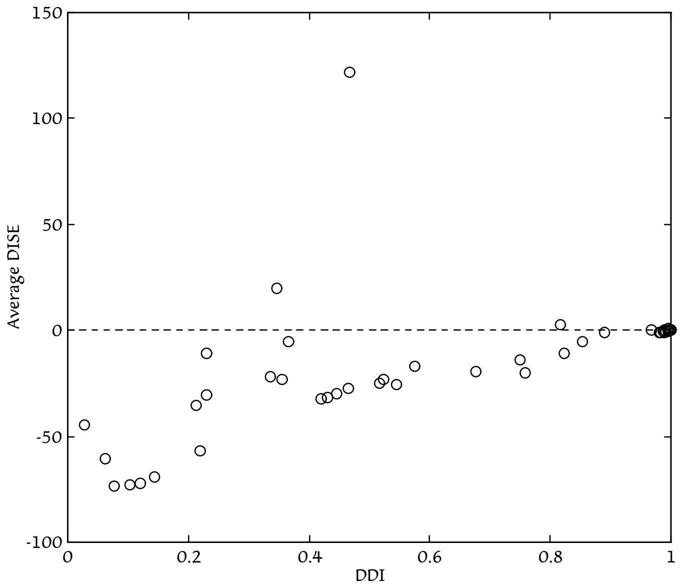

Figure 7: Plot of DISE versus DDI

5. Conclusions

From the analysis of previously defined indexes and of simulations of time responses for all the examined cases some general conclusion can be reached.

The TCR is a very effective way to identify a desired point of the Nyquist curve lying in the third quadrant, as it requires shorter times for the experimental tests and allows to achieve reasonable performance for many cases. However, in a limited number of cases the system does not develop stable oscillations and the identification can not be performed.

The ATV$^+$ requires about twice longer experimental times because more points are identified in the Nyquist plane, but, once it is coupled with a an appropriate design of the controllers which allows to exploit the improve knowledge on the process, better performance can be achieved for all the cases with PI/PID controllers and further improvement with model based controllers.

6. References

- Åström K. J., Hägglund T. (1984). Automatic Tuning of Simple Regulators with Specifications on Phase and Amplitude Margins. *Automatica*, **20**, pp. 645-651
- Friman M., Waller K. V. (1997). A Two-Channel Relay for Autotuning. *Ind. Eng. Chem. Res.* **36**, pp. 2662-2671
- Li W., Eskinat E., Luyben W. L. (1991). An Improved Autotune Identification Method. *Ind. Eng. Chem. Res.* **30**, pp.1530-1541
- Luyben W. L. (1987). Derivation of Transfer Functions for Highly Nonlinear Distillation Columns. *Ind. Eng. Chem. Res*, **26**, pp. 2490-2495
- Scali C, Marchetti G., Semino D. (1999): Identification and Autotuning of Completely Unknown Processes. *Ind. Eng. Chem. Res* (in press).
- Semino D., Scali C. (1998). Improved Identification and Autotuning of PI Controllers for MIMO Processes by Relay techniques. *Journal of Process Control,* **8** (3), pp. 219-227.
- Ziegler J. G., Nichols N. B. (1942). Optimum Settings for Automatic Controllers. *Trans ASME,* **65**, pp. 433-444