Proceeding of the 2004 American Control Conference
Boston, Massachusetts June 30 - July 2, 2004

Model-based PID autotuning enhanced by neural structural identification

Alberto Leva, Luigi Piroddi

Abstract— **This paper presents an autotuning method for industrial PID controllers in the 1-d.o.f. ISA form. The major feature of the method is that the model structure employed for the process is selected on-line based on a step response record, by means of a multilayer perceptron neural network. Thanks to the exclusive use of normalized I/O data, the network can be trained off-line with simulated data, therefore simplifying the method's implementation. Once the model structure is selected and its parameters are identified, the IMC approach is used for synthesizing a regulator that is then approximated with a PID. Simulation and experimental results are reported to show the effectiveness of the proposed tuning method and its advantages with respect to IMC-based PID tuning with the model structure fixed *a priori*.**

I. Introduction and problem statement

Though model-based PID (auto)tuning can be regarded as quite a mature research domain [14], [15], several new issues have been addressed in the last years [1], [7], renewing the interest in the field also from the application point of view [8].

Model-based tuning techniques may be broadly classified in two categories. In the first one the regulator structure is fixed, and that of the model is specified *a priori*, based mainly on the type of regulator to be tuned [5]. This may appear a peculiar choice, because it is apparent that the model structure should depend more on what the process dynamics looks like than on the regulator. It is however necessary, because given a regulator type only a few model structures allow the derivation of simple and reliable tuning rules (an interesting work aimed at overcoming this problem is [7]). Typically, very low-order models are used, and their effective identification for control synthesis purposes is an issue deserving much research effort [2]. Note, incidentally, that such techniques are not difficult to automate. In the second category, the model structure is chosen so as to capture the process dynamics as precisely as possible. As a consequence, the structure of the regulator cannot be specified *a priori*. Tuning methods of this category are extremely powerful, but difficult to automate.

Despite the research interest on it, model-based PID tuning is still less widespread in applications than it could. This is witnessed by the fact that also recent professional papers on 'tuning guidelines' stick mostly to the Ziegler-Nichols method, or some of its derivates (see e.g. [4]). The loss of achievable performance produced by neglecting the possibilities of modern techniques, and specifically of model-based ones, is now recognized also by plant audit campaigns (see e.g. [3]).

A full discussion on the reasons why model-based techniques are difficult would extend beyond the scope of this paper (see e.g. [8]). A major problem, however, is surely that, to really exploit their possibilities, the structure of the process model employed must not be fixed, but rather chosen on-line based on the observed process dynamics. In other words, the above considerations suggest that a very promising way of improving industrial autotuners is to couple model-based synthesis methods with automatic model structure selection capabilities. Proposing one such method is the purpose of this paper.

II. The proposed method

Based on the preliminary results of [11], [12], where a neural technique for structural identification is proposed, and of [8], where some implementation-related issues of model-based PID tuning are discussed, this paper proposes a PID tuning method with the following characteristics:

- A neural technique is used to select the structure of the process model from measured I/O data. The data are normalized prior to the neural structure identification, to allow off-line training of the network. Off-line training is extremely important for the implementation of the proposed tuning method, as it significantly reduces the overall computational burden. Moreover, the possibility of training the network off-line avoids the necessity of many process data. It is worth noting that the difficulty of obtaining large amounts of process information is a major obstacle for the application of neural techniques in the autotuning domain.
- The parameterization of the model is subsequent to the structure selection and can be realized with standard identification algorithms. This is a peculiarity of the proposed approach. Thanks to the improved adherence of that structure to the measured data, the parameter identification algorithm is less critical than it turns out to be if the structure is chosen *a priori*. Note, by the way, that with the presented approach only the selected structure needs parameterizing, while standard structure selection methods (e.g. those that compare prediction errors) require all the candidate structures to be parameterized.

More precisely, in the proposed approach, structural information is used in two ways. First, when parameterizing the model, the number of poles and zeros selected by the structural identification is used to avoid

A. Leva and L. Piroddi are with Dipartimento di Elettronica e Informazione, Politecnico di Milano, Via Ponzio 34/5, 20133 Milano, Italy {leva,piroddi}@elet.polimi.it

0-7803-8335-4/04/$17.00 ©2004 AACC

under- or over-parameterization. In so doing, it can be reasonably expected that any parameter identification algorithm will evidence the detailed characteristics (e.g., oscillations) of the process dynamics, and will produce a smaller model error than if the structure were fixed. Then, in the regulator tuning, more detailed information, such as the presence of over- or under-damped dynamics, can be used to select a specialized tuning method [12]. In this work, however, a different approach is adopted for the tuning, as explained below.
- A unitary approach is used for tuning the regulator, so as to cast the tuning problem in a consistent and well established framework. The IMC method is firstly applied, then the obtained regulator is approximated with a PID. This is not standard practice, as it is more frequent that the model is reduced prior to the tuning. The proposed approach appears to be very promising, however, because the information loss is postponed as much as possible. Studies are underway to formalize this concept, and the results will be presented in future works. It is worth recalling that a small model error is highly beneficial both for the quality of the tuning obtained with the IMC and for the interpretability of its design parameter.

A. Determining the model structure and identifying parameters

Any control practitioner would classify typical step responses more or less as depicted in figure 1. For space limitations, in this paper we concentrate on asymptotically stable processes.

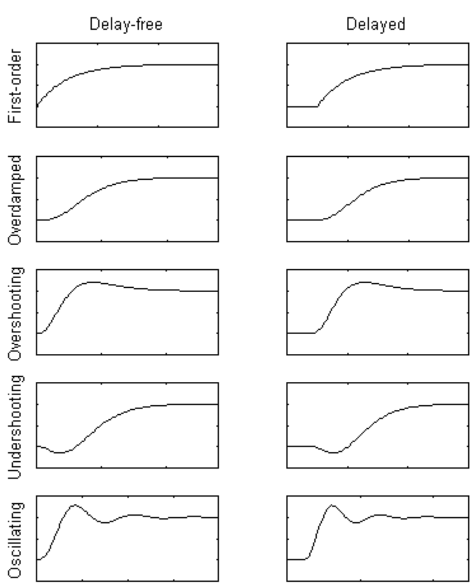

Fig. 1. Classification of open-loop step responses.

Such a classification is substantially based on the recognition of a set of *features* of the step response. This suggests that the task may be automated as a pattern recognition problem. The effectiveness of such an approach is witnessed by several works, see e.g. [6]. In [11], [12], a method is devised for this purpose based on a multilayer perceptron neural network used as a classifier. First, the unit step response of the process is recorded, then filtered with the first-order discrete lowpass transfer function $0.5/(z-0.5)$ and normalized both in time, so that the record be of 64 samples, and in amplitude, so that the steady-state value reached by the response be the unity. The filtering and normalization process is illustrated in figure 2.

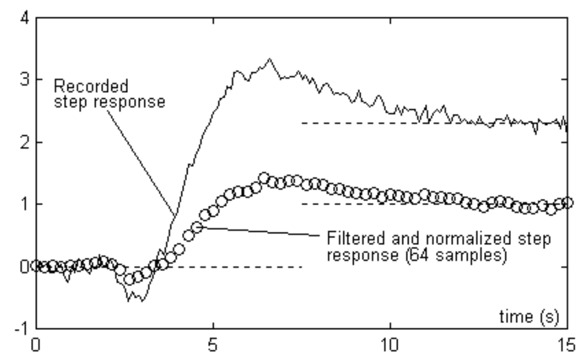

Fig. 2. Step response filtering and normalization for the neural model structure selection.

The neural network is fed with a the filtered and normalized response, and first recognizes the presence or absence of some relevant features. As discussed in [12], a sufficiently informative set of features (for asymptotically stable processes) is

- first-order behaviour (nonzero initial derivative);
- apparent delay;
- presence of an overshoot;
- presence of an undershoot;
- oscillatory behaviour.

Based on these features and on the normalized data, the network selects the most appropriate process model structure within the following set:

$$\begin{aligned}
M_1(s) &= \tfrac{\mu}{1+sT}, & M_2(s) &= e^{-sL}M_1(s), \\
M_3(s) &= \tfrac{\mu}{(1+sT_1)(1+sT_2)}, & M_4(s) &= e^{-sL}M_3(s), \\
M_5(s) &= \tfrac{\mu(1+sT_z)}{(1+sT_1)(1+sT_2)}, & M_6(s) &= e^{-sL}M_5(s), \\
M_7(s) &= \tfrac{\mu(1-sT_z)}{(1+sT_1)(1+sT_2)}, & M_8(s) &= e^{-sL}M_7(s), \\
M_9(s) &= \tfrac{\mu}{1+2\frac{\xi}{\omega_n}s+\frac{s^2}{\omega_n^2}}, & M_{10}(s) &= e^{-sL}M_9(s).
\end{aligned} \tag{1}$$

The final classification is not based on the extracted features only, but also on the data. This is necessary, since there are some important facts that are difficult to translate into response features, or that - if translated into features - may not appear with enough evidence to be recognized with certainty. A typical example is when an overdamped step response presents a slope change, typically caused by a zero whose time constant is intermediate with respect to those of the poles. If the process order is high enough, the selected model will not be first-order and, since there is no overshoot, most likely a model without zeros will be selected. In such a situation, virtually any parameter identification method will overestimate the dominant time constant, leading to a more or less sluggish tuning. This is illustrated in the experiment of section IV.

The technique proposed in [12] has two important advantages, that make it very suited for the purpose of the research presented herein. First, the generalization capabilities of the neural network have proven to select a good structural approximation also for processes not belonging to the set (1). Second, and maybe most important, since the network only employs normalized data, it can be trained off-line with responses generated in simulation from a suitable set of systems belonging to the considered classes.

The network training is made in two steps, using many random models of classes (1). First a subnetwork devoted to recognizing the features above is trained, having the presence or absence of each feature as the training signal. After training this subnetwork, its weights are fixed, and the compound network is trained having the correct class as the training signal. More details on the network structure and training are in [12].

Once the model structure is selected, plenty of methods are available for selecting its parameters. Discussing this phase in detail is not relevant in this paper. It suffices here to say that, provided the selected model structure 'fits the data', the choice of the identification method is far less critical than it would be if the structure were chosen *a priori*. In this work, identification is done by simple LS ISE minimization, considering the model's simulation (not prediction) error. Notice that selecting the model structure *a priori* allows to parameterize only the chosen structure, and not all the candidate structures. This eases the identification from the step response, and provides better tuning results.

If, on the other hand, all the candidate model structures were identified and an *a posteriori* choice were made based on classical prediction performance, overparameterized models would typically be selected, which do not necessarily guarantee optimal simulation performance. In particular, especially in noisy cases or in the presence of spurious phenomena during identification, such models are not guaranteed to capture the control relevant dynamics, and therefore are not particularly suited for structure specific tuning rules, nor for tuning approaches where the model is contained in the regulator explicitly, as is the case with the IMC.

B. Tuning the regulator

The IMC scheme, first proposed in [14], has found a number of successful applications. Its rationale is shown in the block diagram of figure 3. Here $P(s)$ is the transfer function of the process, $M(s)$ is the model of the process obtained from I/O data, $Q(s)$ and $F(s)$ are two asymptotically stable transfer functions, y°, d, and n are the set point, a load disturbance and a measurement noise, y and $\widehat{y}$ are the true and nominal controlled variables. The feedback signal is $y - \widehat{y}$, which motivates the method's name in that the regulator (the gray blocks in figure 3) contains a model of the process *explicitly*.

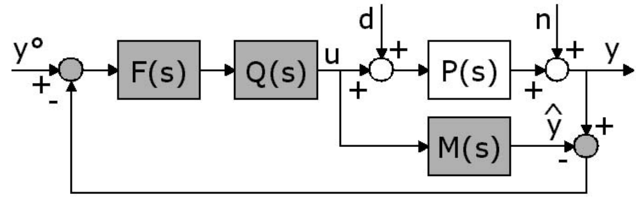

Fig. 3. The IMC control scheme.

If $M(s) = P(s)$, $d = n = 0$, and it is possible to choose $Q(s)$ as the exact inverse of $M(s)$, i.e. of $P(s)$, then the transfer function from set point to controlled variable equals $F(s)$, and can thus be chosen arbitrarily. Moreover, suppose that $d \neq 0$ while still $M(s) = P(s)$, $n = 0$ and $Q(s) = M^{-1}(s)$. In this case, if $F(s) = 1$ the disturbance d is rejected completely; otherwise, it is rejected asymptotically provided that $F(0) = 1$. Therefore, the IMC approach shows that for stable and *exactly* known processes feedback is necessary only because of disturbances, while for stable and only *partially* known processes known it is also necessary due to the model error. The IMC regulator of figure 3 is equivalent to a feedback regulator $R(s)$ given by

$$R(s) = \frac{F(s)Q(s)}{1 - F(s)Q(s)M(s)} \tag{2}$$

and, under the hypothesis $P(s) = M(s)$, the control system is internally asymptotically stable iff $M(s)$, $Q(s)$ and $F(s)$ are asymptotically stable. Hence, the IMC provides a parametrization of *all* the regulators which stabilize a control system containing a (known) asymptotically stable process.

Apparently, the IMC method has one big advantage: by itself, it is in no sense connected to a specific regulator type. In fact, not only can the scheme of figure 3 be used whatever the model $M(s)$ is, but if this is done, the most important part of the regulator is provided by the model directly. If the model is structurally good, the task of finding suitable $Q(s)$ and $F(s)$ is not difficult nor critical, if the guidelines devised in the following are adopted.

1) First, for the purpose of PID tuning, replace the possible delay of $M(s)$ with its (1,1) Padé approximation $(1 - sL/2)/(1 + sL/2)$, which provides a rational process model $M_R(s)$. This is standard practice in IMC-based tuning [14], and can be safely done provided that the requested control system's bandwidth is not excessively wide—a precaution taken implicitly in the following steps.

2) Then, figure out the upper limit ω_M of the band where the model represents the process 'precisely enough' from the regulator synthesis standpoint. If the model structure fits experimental data, as in our context, a first-cut (yet reasonable and quite conservative) value for ω_M is four times the maximum frequency of all the poles and zeros of $M_R(s)$ if $M_R(s)$ is minimum-phase, or half the lowest frequency of the RHP zeros of $M_R(s)$ in the opposite case.
3) Once ω_M is determined, compute the inverse of the minimum-phase part of $M_R(s)$. If it is not proper, augment it with poles located above (say at twice) ω_M. The number of these poles is chosen so that the result, i.e. $Q(s)$, be of relative degree -1.
4) Carrying on, set $F(s)$ to a first-order filter with a cutoff frequency below (say at 1/5 of) ω_M. In so doing, the relative degree of $Q(s)F(s)$ is zero, which is suitable for the subsequent approximation of the regulator with a real PID. Observe that, as the regulator must have integral action, it is required - see (2) - that $Q(0)F(0)M(0) = 1$, so choose the gain of $F(s)$ accordingly (normally one is correct since there is no reason for $Q(0)$ not being the reciprocal of $M(0)$). This leads to the 'full' IMC regulator $R_{IMC}(s)$ in the form (2), that in the general case is not a PID.
5) The (nominal) phase margin of the open loop transfer function $R_{IMC}(s)M(s)$ (not $R_{IMC}(s)M_R(s)$, notice) must now be checked, to see if the design is reasonably stable and robust (45 degrees is a suitable threshold value). It is also to be checked that the control system's (nominal) cutoff frequency $\hat{\omega}_c$, estimated with $M(s)$, is below ω_M. In the opposite case, the cutoff frequency of $F(s)$ has to be reduced and $R_{IMC}(s)$ to be recomputed.
6) The next step is to approximate $R_{IMC}(s)$ with a PID. In this work, we adopt the 1-d.o.f. ISA form of the PID control law, i.e.,
$$R_{PID}(s) = K\left(1 + 2\frac{1}{sT_i} + \frac{sT_d}{1 + sT_D/N}\right). \quad (3)$$
This approximation is briefly discussed below.
7) Finally, the phase margin of the open loop transfer function $R_{PID}(s)M(s)$ is checked. If it is too low, the IMC regulator (and then the PID) are recomputed decreasing the cutoff frequency of $F(s)$, as the performance achieved with the 'full' IMC cannot be preserved with the PID approximation.

The key point of the proposed procedure is that, in all the cases (i.e., model classes) considered, $R_{IMC}(s)$ can be approximated with a PID well enough. For this approximation, an *ad hoc* numeric procedure is employed. The rationale is to preserve the low- and high-frequency aspect of the regulator's frequency response, and to preserve the mid-frequency phase lead (when required) while keeping the regulator zeros' damping over a given value (typically 0.6) when they turn out to be complex. To convince the reader that this task is neither difficult nor critical, figure 4 reports the Bode diagrams of the 'full' IMC regulators in all the cases that may arise with the set of model classes adopted. Numeric details are omitted here for brevity. Note that, though the structure of $R_{IMC}(s)$ depends on that of $M(s)$, and though the frequency distribution of the model error is correspondingly heterogeneous [13], all the possible shapes of its Bode diagrams are compatible with the PID structure.

III. A SIMULATION EXAMPLE

In this example, the process is described by the transfer function
$$P(s) = \frac{1}{(1+2s)\left(1 + 2\frac{0.25}{0.5}s + \frac{1}{0.25}s^2\right)}, \quad (4)$$
that does not belong to any of the model classes $M_1 - M_{10}$. The apparent oscillatory behavior of its step response causes the selection of a model of class M_9, which is then parameterized as
$$M_9(s) = \frac{1}{1 + 2s + 4.94s^2}. \quad (5)$$
The proposed method is then employed taking
$$Q(s)F(s) = \frac{1 + 2s + 4.94s^2}{(1+0.5s)(1+5s)}, \quad (6)$$
and leading, by means of the proposed PID approximation, to the regulator
$$R(s) = 0.2\frac{1 + 2.7s + 4.9s^2}{s(1+0.8s)}. \quad (7)$$
where the zeros' damping is limited to 0.6. A model of class M_2 can also be identified, to compare the proposed method with standard IMC-PID tuning based on First-Order Plus Dead Time (FOPDT) models. Since LS ISE minimization is inadequate with such a low damping, a modified version of the method of areas is employed, and the parameterized model turns out to be
$$M_2(s) = \frac{e^{-0.2s}}{1 + 5.5s}. \quad (8)$$
In figure 5, the 'full' IMC regulator is compared with the IMC-PID obtained from model (8) with $\lambda = 1.5(L+T)$ and $\lambda = 3(L+T)$. Recall that in the IMC-PID method [14] λ is interpreted as the desired dominant closed-loop time constant. Figure 5, left column, reports the identification results with a FOPDT model and with the Second-Order (SO) model selected by the neural network; figure 5, right column, shows the performances of the proposed method and those of the FOPDT-based IMC-PID; both are evaluated with the (nominal) model and with the real process.

Note how, in the IMC-PID method, the model mismatch makes the choice of λ critical. In fact, there is a quantifiable relationship between the model error magnitude and the minimum value of λ that stabilizes the real system with

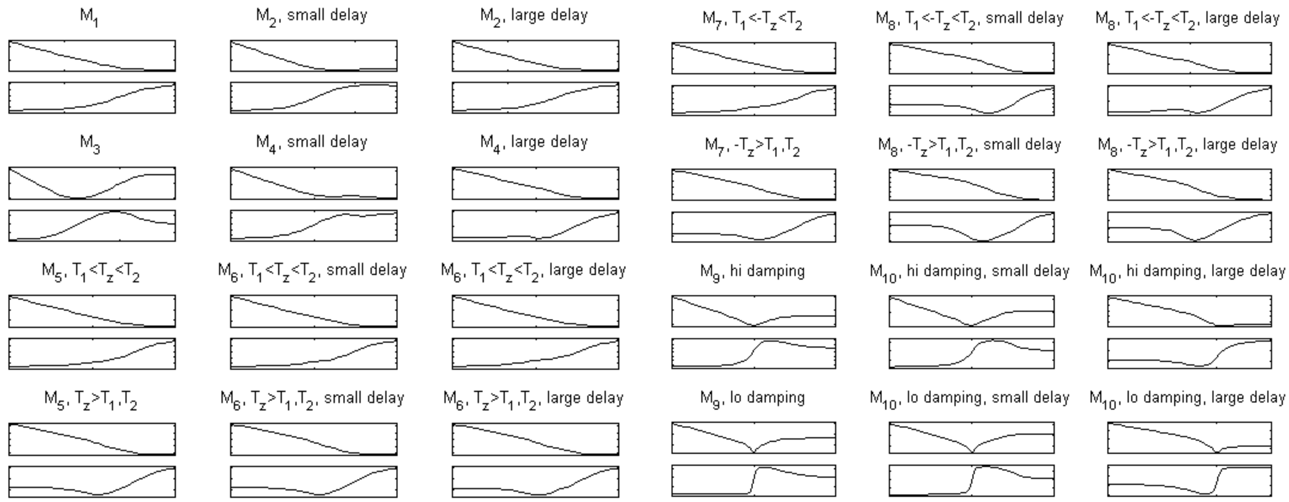

Fig. 4. Qualitative Bode diagrams (top: magnitude; bottom: phase) of the IMC regulators.

a PID tuned on the (nominal) model; this is investigated in detail in [10]. It is also important to observe that the model mismatch adversely affects the *a priori* evaluation of the control behaviour in the IMC-PID case, while with the IMC regulator (that uses a structurally correct model) the tuning results can be evaluated more consistently.

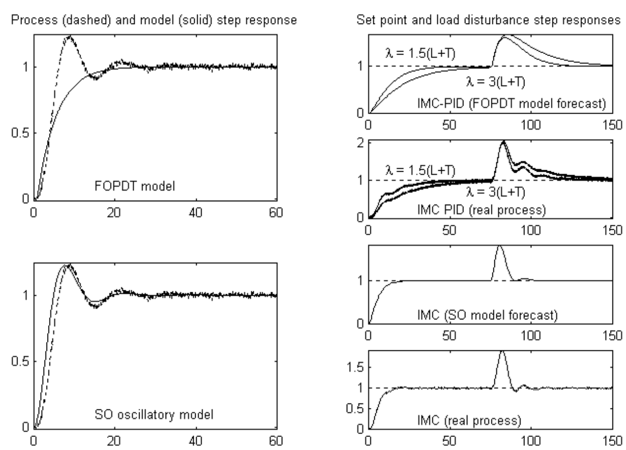

Fig. 5. Identification and tuning results (proposed method versus FOPDT-based IMC-PID).

To back up the statement that the 'full' IMC regulator can be approximated with a PID effectively in all the model classes considered, and more in general whenever the process (and model) structure is suited enough for a PID, figure 6 reports the results obtained with the regulator (7).

IV. Experimental results

To show the effectiveness of the proposed method, we now present an experimental application in a case where model structure selection is relevant. The experimental setup is a temperature control system [9], in which a metal plate is connected to two electric heaters. One heater is the

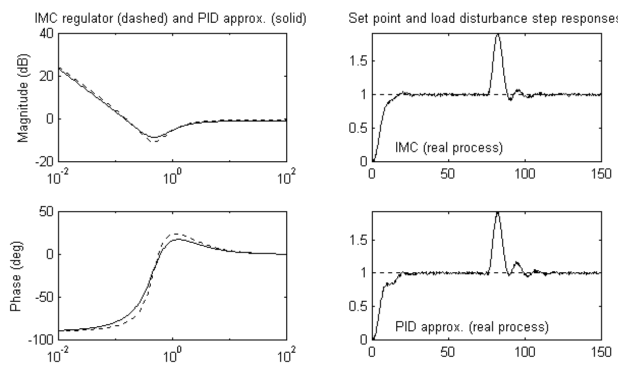

Fig. 6. IMC regulator and PID approximation.

control actuator, the other provides a load disturbance. The controlled variable is the plate temperature. Figure 7(a) shows an open-loop step experiment and two identified models: one is of the class selected by the method (M_6), the other has a FOPDT structure (i.e., it is of class M_2). The parameterized models are

$$M_6(s) = \frac{2.12(1+85s)e^{-1.8s}}{(1+165s)(1+35s)}, \quad (9)$$

$$M_2(s) = \frac{2.12e^{-2.5s}}{1+125s}. \quad (10)$$

Figure 7(b) shows the tuning results of the proposed method. Taking

$$Q(s)F(s) = \frac{(1+165s)(1+35s)}{2.12(1+85s)(1+0.9s)} \quad (11)$$

leads to the approximating PID regulator

$$R_1(s) = 85\left(1+\frac{1}{35s}+\frac{0.6s}{1+0.3s}\right). \quad (12)$$

In figure 7(c) this PID is compared to one tuned with the IMC-PID rules [14], requesting a closed-loop dominant time constant ('λ' in [14]) of 10s, i.e.,

$$R_2(s) = 100\left(1 + \frac{1}{126s} + \frac{s}{1+1.2s}\right). \qquad (13)$$

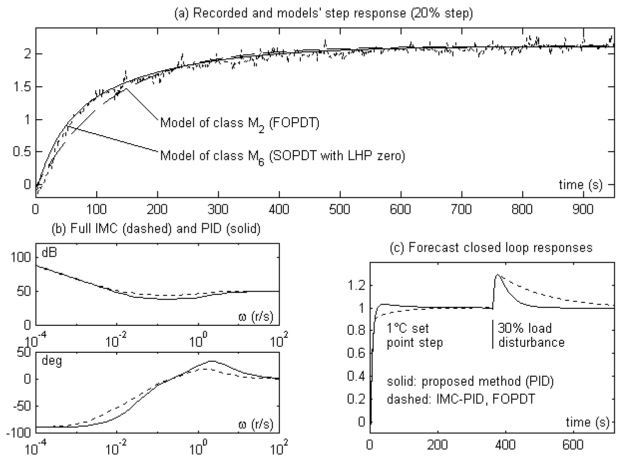

Fig. 7. Open-loop experiment and tuning.

Finally, figure 8 reports the experimental closed-loop transients obtained with the proposed method (R_1) and the IMC-PID rules with the FOPDT model (R_2). Note that the FOPDT model has a small delay and a large time constant. With the IMC approach, this means that the integral time equals that time constant. In the case at hand, this causes an apparent lack of integral action on the part of R_2, and the fact is structural, since no sensible identification algorithm for FOPDT models, starting from the measured step response of figure 7(a), would ever produce a time constant less than 100s or so. Conversely, the structure-conscious tuning of R_1 overcomes the problem. The integral time is one third of that of R_1, and the transients of figure 8 are self-explanatory.

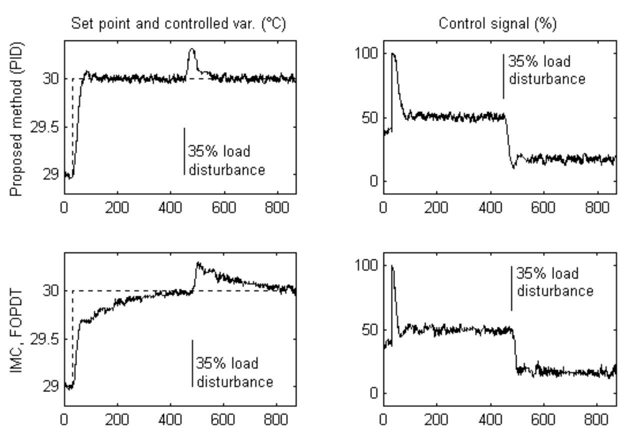

Fig. 8. Closed-loop experimental results.

V. Concluding remarks

An autotuning method for industrial PID controllers in the 1-d.o.f. ISA form has been presented. The model structure employed for the process is selected on-line based on a step response record, which enhances the proposed method's flexibility. The structural recognition task is performed by a multilayer perceptron neural network, which operates on-line based on a suitably normalized record of the process step response. It is worth stressing that the network can be trained off-line. The training of the neural network is made with an *ad hoc* procedure, organized in two steps. First some relevant features of the process step response (e.g., a delay or an overshoot) are detected; then, these features and the step response data are used to classify the process with respect to a predefined set of model structures. Once the model structure is selected and the model parameters are identified, the IMC approach is used for synthesizing a regulator that is then approximated with a PID. The structural coherence between process and model makes the parameter identification method less critical an issue with respect to standard model-based autotuning. Both simulation and experimental results have been reported, showing the proposed method's advantages with respect to IMC-based PID autotuners that select the model structure *a priori*.

References

[1] Åström, K.J and T. Hägglund (2001). *The Future of PID Control.* Control Eng. Practice, **9**, 1163-1175.

[2] Bi, Q., W.J. Cai, E.L. Lee, Q.G. Wang, C.C. Hang and Y. Zhang (1999). *Robust Identification of First-Order Plus Dead Time Model from Step Response.* Control Eng. Practice, **7**, 71-77.

[3] EnTech Control Engineeering Inc. (1993). Automatic Controller Dynamic Specification, version 1.0, 11/93.

[4] Harrold, D. (1999). *Process Controller Tuning Guidelines.* Control Engineering, August 1999.

[5] Huang, H.P., C.L. Chen, C.W. Lai and G.B. Wang (1996). *Autotuning for Model-Based PID Controllers.* AIChE Journal, **42**(9), 2687-2691.

[6] Huang, H.P., M.W. Lee and C.L. Chen (2001). *A System of Procedures for Identification of Simple Models Using Transient Step Response.* I&EC Res., **40**, 1903-1915.

[7] Isaksson, A.J. and S.F. Graebe (1999). *Analytical PID Parameter Expression for Higher Order Systems.* Automatica, **35**, 1121-1130.

[8] Leva, A. (2001). *Model-based Tuning: the Very Basics and Some Useful Techniques.* Journal-A, **42**(3), special issue on controller tuning, 14-22.

[9] Leva, A. (2002). *A Hands-On Experimental Laboratory for Undergraduate Courses in Automatic Control.* IEEE Trans. on Education (to appear).

[10] Leva, A. and A.M. Colombo (2000). *Estimating Model Mismatch Overbounds for the Robust Autotuning of Industrial Regulators.* Automatica, **36**, 1855-1861.

[11] Leva, A. and L. Piroddi (1994). *A Neural Network-Based Technique for Structural Identification of SISO Systems.* Proc. IMTC/94, Hamamatsu, 135-138.

[12] Leva, A. and L. Piroddi (1996). *Model-Specific Autotuning of Classical Regulators: a Neural Approach to Structural Identification.* Control Engineering Practice, **4**(10), 1381-1391.

[13] Maffezzoni, C. and P. Rocco (1997). *Robust Tuning of PID Regulators Based on Step Response Identification.* European J. of Control, **3**, 125-136.

[14] Morari, M. and E. Zafiriou (1989). *Robust Process Control.* Prentice-Hall.

[15] Rivera, D.E., M. Morari and S. Skogestad (1986). *Internal Model Control. 4. PID Controller Design.* Ind. Eng. Chem. Process. Des. Dev., **25**, 252-265.