ON THE ROLE OF THE PROCESS MODEL IN MODEL-BASED AUTOTUNING

Alberto Leva [1], **Francesco Schiavo** [2]

Dipartimento di Elettronica e Informazione, Politecnico di Milano
Via Ponzio, 34/5 - 20133 Milano (Italy)

Abstract: Some aspects of model-based (auto)tuning are investigated, taking the IMC rules for the PID regulator as a representative and illustrative example, but without loss of generality. The presented study leads to conclude that it would be very beneficial to reconsider in depth the role of the process model used for the tuning. A new perspective is proposed in this respect, namely the adoption od a 'dual model approach' in model-based tuning, and the possible advantages of it are evidenced. The matter is still open for future research.

Keywords: Autotuners, model-based control, adaptive control, PID control, process control.

1. INTRODUCTION

Model-based tuning methods for industrial regulators have started appearing in the literature about thirty years ago. In the last decade they have been encountering an encouraging success also in the applications, as an alternative to more traditional (and less performing) tuning recipes such as the Ziegler-Nichols rules and the great number of their derivates (the quarter-damping, the Cohen-Coon method and so on). Nevertheless, the use of such modern techniques is less widespread than it could, and also quite recent professional papers on tuning guidelines (see e.g. (Harrold, August 1999)) stick mostly to the Ziegler-Nichols method. The consequent loss of achievable performance (not to say the inability of tuning satisfactorily) is nowadays recognised also on the basis of thorough plant audit campaigns (see e.g. (EnTech Control Engineeering Inc., 1992)).

The application of model-based tuning techniques is hindered by some facts that, curiously enough, are seldom addressed in the literature (Leva and Colombo, 2001*b*). In synthesis, all those problems can be thought as deriving from the interplay between the particular identification method used to obtain the process model and the subsequent choice of the tuning parameter(s). Note that in many works on model-based tuning the identification method is not discussed, and in some of them is not even specified, see e.g. the review given in (O'Dwyer, 2003) with reference to PI-PID tuning.

The (preliminary) study presented herein leads to propose a 'dual model' approach as a systematic way to circumvent those problems, thus increasing the possibilities, and the potential acceptance, of model-based tuning methods.

2. BACKGROUND

The ideas and findings of this work apply to the general model-based tuning method that synthesises a LTI (Linear, Time-Invariant) regulator based on a LTI process model (termed from now

[1] Corresponding author. Voice (39) 02 2399 3410, Fax (39) 02 2399 3412, e-mail leva@elet.polimi.it
[2] PhD student at the Dipartimento di Elettronica e Informazione

on the 'tuning model') drawn from I/O data, and some design parameter(s). To present some figures, we shall refer to the IMC (Internal Model Control) PID tuning method as a representative example, but without loss of conceptual generality. It is also necessary to introduce the technical hypothesis that the I/O data be generated by an arbitrarily complex LTI system, termed the 'real process', otherwise the model error definition that will be used has no significance. The presence of noise in the data is admitted, of course.

Denoting by $P(s)$ and $M_t(s)$ the transfer functions of the real process and the tuning model, the IMC method determines a feedback regulator $R(s) = Q(s)F(s)/(1 - Q(s)F(s)M_t(s))$, where $Q(s)$ is an (approximate) inverse of $M_t(s)$, and $F(s)$ is a lowpass filter aimed at trading control bandwidth against robustness, see e.g. (Morari and Zafiriou, 1989) for full details. These characteristics are shared by any model-based tuning method.

In the IMC PID procedure $M_t(s)$ is a FOPDT (First-Order Plus Dead Time) model and $F(s)$ is first-order with unity gain. Omitting inessential details, the parameters of the ISA PID (Åström and Hägglund, 1995)

$$R(s) = K\left(1 + \frac{1}{sT_i} + \frac{sT_d}{1 + sT_d/N}\right) \quad (1)$$

are determined as

$$\begin{aligned} T_i &= T + \frac{L^2}{2(\lambda + L)}, \\ K &= \frac{T_i}{\mu(\lambda + L)}, \\ N &= \frac{T(\lambda + L)}{\lambda T_i} - 1, \\ T_d &= \frac{\lambda L N}{2(\lambda + L)}. \end{aligned} \quad (2)$$

where μ, T and L are the gain, time constant and delay of $M_t(s)$, respectively, and λ is the time constant of $F(s)$(Leva and Colombo, 2001*b*).

3. SOME RELEVANT FACTS

We shall now evidence some facts that concern the IMC PID and, with the convenient adaptation, any model-based tuning method. First, the use of model error information to select the design parameter(s) is mandatory. To evidence this, we observe that in the IMC PID the closed-loop control system containing $M_t(s)$ (termed from now on the'tuning system') is asymptotically stable whatever value is selected for λ, and only the

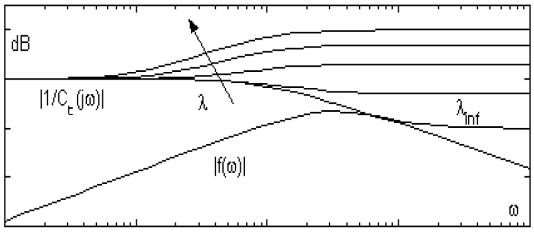

Fig. 1. Lower bound for λ.

degree of stability robustness is influenced by that parameter (Leva and Colombo, 2001*b*). The control sensitivity function $C_t(s)$ of the tuning system takes the form $(1 + sT)/(\mu(1 + s\lambda))$ and, denoting by $W(s) = P(s) - M_t(s)$ the additive model error, the well known criterion introduced in (Doyle *et al.*, 1992) states that the stability of the tuning system carries over to all the systems where $||W(j\omega)C_t(j\omega)||_\infty < 1$. Therefore, given $W(j\omega)$ (or, more realistically, a function $f(\omega) \geq |W(j\omega)|\ \forall\omega$) one can find a lower bound λ_{inf} for λ as suggested by figure 1.

In general, either a method stabilises the tuning system for any value of its design parameters (as is the case with the IMC PID), or the set of stabilising parameters can be computed based on $M_t(s)$. The design parameters always have an influence on robustness, as they determine (in various senses depending on the particular method) 'how much the method trusts the model'. Since assuming the presence of model error is merely a matter of realism, in model-based tuning design parameters cannot be selected effectively without model error information.

Curiously enough, and coming back to the IMC example, a number of rules have been introduced to select λ based on $M_t(s)$, and one might wonder how these rules can work. To sketch out an answer, we take a batch of non-FOPDT processes, identify a FOPDT model for each of them, compute λ_{inf} exactly (in this ideal case the model error is known), and then compare λ_{inf} to the values of λ provided by rules based on the FOPDT models. Given the two processes with unity gain

$$\begin{aligned} P_1(s) &= \frac{1 + 1.6s}{(1 + s)^3}, \\ P_2(s) &= \frac{1 + 3s}{(1 + 5s)(1 + s)^2}, \end{aligned} \quad (3)$$

we compute their open-loop unit step responses and, denoting by t_{perc} the time when the response reaches $perc/100$ of its final value, we parameterise six FOPDT models for each process by setting (a) $\mu = 1$, $T = t_{63} - t_{05}$, $L = t_{05}$, (b) $\mu = 1$, $T = t_{63} - t_{10}$, $L = t_{10}$, (c) $\mu = 1$, $T = (t_{95} - t_{05})/2.5$, $L = t_{05}$, (d) $\mu = 1$, $T = (t_{95} - t_{10})/2.5$, $L = t_{10}$, (e) by using the method of areas,

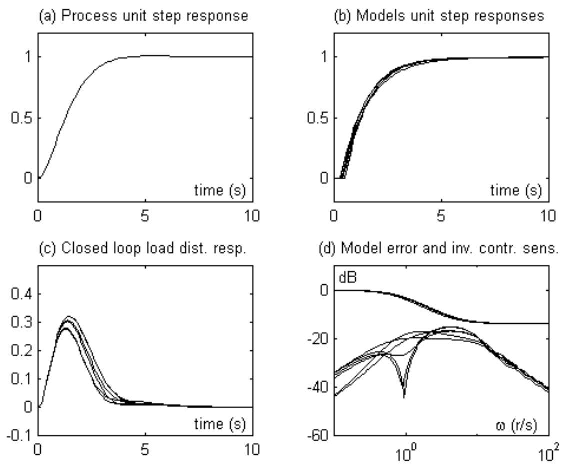

Fig. 2. Different FOPDT models and PIDs for $P_1(s)$.

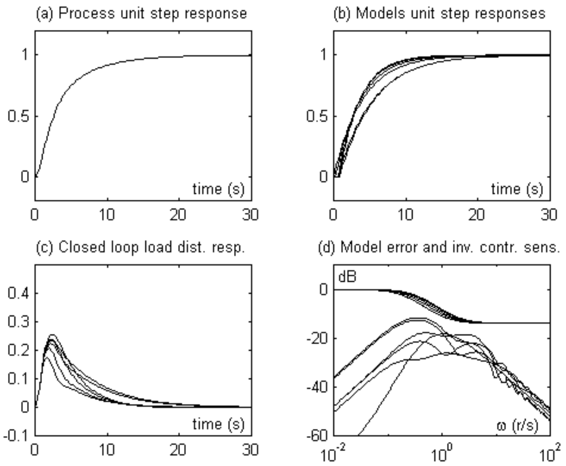

Fig. 3. Different FOPDT models and PIDs for $P_2(s)$.

and (f) by minimising the ISE numerically. Then, for each model, we compute λ with the widely used rule $\lambda = max(0.25L, 0.2T)$, and we tune a PID with the IMC formulæ. The results are illustrated in figures 2 and 3, that refer to $P_1(s)$ and $P_2(s)$, respectively. The figures show (a) the unit step response of the real process, (b) those of the six FOPDT models, (c) the load disturbance responses of the control systems made of the six PIDs and the real process, and (d) the magnitude of the (exact) additive model errors (lower curves) and of the inverse of the nominal control sensitivity functions (upper curves), so that the minimum distance between each couple of curves measures the degree of stability robustness.

The response of $P_1(s)$, see figure 2, is reproduced very well by the FOPDT models. Both the models and the disturbance rejection characteristics are almost equivalent. However, this is not true for the degree of stability robustness, since the sensitivities are almost equal, but the model errors are not. Moreover, even the most demanding regulator could have the gain almost doubled safely: the regulators are similar because they are all very conservative, as $\lambda \gg \lambda_{inf}$. In the case of $P_2(s)$, see figure 3, the FOPDT model is less fit to the response. The models, the regulators and the disturbance rejection characteristics are less similar, and the differences in terms of stability robustness are very significant. The interested reader can easily produce several examples like these, and also include phenomena (like noise) that we have not considered here for brevity and clarity. Quite intuitively, the problems evidenced would stem even more apparently than in the ideal situation considered herein.

This example should convince that an efficient choice of λ is only a matter of robustness, and therefore *cannot be done based only on the nominal process model*. Rigorously speaking, any rule that operates in that way is nonsense. Such rules often lead to excessive conservatism, and therefore hinder the acceptance of potentially powerful tuning methods in the applications.

Strictly connected to the problem of model error, and its use, is that of characterising the role of the identification method used to obtain $M_t(s)$. The use of model error is mandatory to determine λ, but the identification method can make that task more or less critical. To illustrate the problem, we look at another example, where the mere inspection of the step response may lead to erratic conclusions on the tuning results unless model error is estimated and used. Consider the two processes

$$P_1(s) = \frac{32(s+1)}{(s+0.5)(s+2)(s+4)(s+8)}, \qquad P_2(s) = \frac{3.591}{s^2+6.537s+3.591} \tag{4}$$

whose step responses differ by less than 4% at any time instant. The structure of their dynamics is very different, but this is not revealed immediately by inspecting the step response. To show the consequences of this, it is interesting to identify a FOPDT model and estimate the model error on the step response of one process, tune a PID, and then apply it to both processes. The result of this comparison with the processes above are shown. Two tuning models were identified: $M_{t1}(s)$ was obtained with the method of areas applied to the step response of $P_1(s)$, while $M_{t2}(s)$ came from the step response of $P_2(s)$ taking as delay the time to reach 10% of the final value, and as time constant the time from 10% to 63% of the final value. Notice, however, that in real-life cases such estimates can be affected by the time quantisation up to a significant extent. The two models are

$$M_{t1}(s) = \frac{e^{-0.15s}}{1+1.8s}$$
$$M_{t2}(s) = \frac{e^{-0.35s}}{1+1.5s} \qquad (5)$$

Then, the four values of λ_{inf} corresponding to the four model/response couples were computed with the method proposed in (Leva and Colombo, 2001*a*), and four PIDs were tuned with $\lambda = 4\lambda_{inf}$. The so obtained closed-loop responses are shown in figure 4, and illustrate how choosing λ based on the model error produces satisfactory (and reasonably uniform) results in both cases.

In synthesis, then, we can state that the use of model error information allows to choose design parameters objectively and may reduce the criticality of the identification method. This has been shown for the IMC method, but the concept is general. However, none of the considerations made so far allows to tackle the inherent tendency of of model-based tuning to deal with the *tuning model* dynamics, irrespective of their ability to represent the process dynamics relevant for the problem at hand. Addressing this aspect leads to an interesting perspective, as shown in the next section.

4. A NEW AND INTERESTING PERSPECTIVE

Among the major drawbacks of model-based synthesis is that, in various senses, it tends to operate by cancellation (the IMC is a good example). This means, for example, that load disturbance rejection transients are sluggish if the desired closed-loop dynamics are significantly faster than the open-loop ones. Problems like this have been addressed in several ways. For example, in the IMC case, an interesting idea is to use filters $F(s)$ of different structures: this is done e.g. in (Horn *et al.*, 1996), but the filter is selected based on the structure of the tuning model.

More precisely, however, the IMC approach inherently implies that the poles of the *tuning model* are canceled, so that the response is poor if those poles are slow in comparison to the dynamics of the *process* that are relevant for the specific problem (and not always coincide with the dominant ones). As long as the entire synthesis process, including the choice of the IMC filter, is centred on the tuning model, the fidelity of that model to the measured I/O data is of value *per se*; the type of problem that model is meant for is not considered.

We believe that it is very beneficial, not to say necessary, to abandon this attitude. If a complex model of the process is available, it is not necessary that the FOPDT model employed reproduce the I/O data precisely, but rather that it be chosen so as to produce a regulator with the desired characteristics, and to be ‘similar enough’ to the complex model to preserve stability—a property that is easily verified based on the model error. In other words, given the simple structure of the tuning models, we believe not only that the core of the problem lies in the model error, but also that this reasoning should be followed up to the conclusion that the tuning model itself, in some sense, has to be considered as a parameter in the overall design. This is a significant shift with respect to classical model-based tuning, but we believe it is a step forward in that it gives the tuning model only the role it can reasonably play, that is, the role of an ‘intermediate product’ useful to select the regulator parameters with viable rules.

5. THE PROPOSED APPROACH

Based on the considerations above, we propose a ‘dual model approach’ to model-based tuning: the idea is to identify a ‘precise’ process model $M_p(s)$, derive from it one or more tuning models $M_{t,i}(s)$ and the corresponding ‘estimated model errors’ $W_{e,i}(s) := M_p(s) - M_{t,i}(s)$ depending on the particular control problem, tune the regulator with a satisfactory tuning model and assess its robustness with respect to the corresponding model error. In this section we briefly sketch out, by means of an example, how this general idea can be applied to the IMC tuning method. Suppose that the goal is to achieve good load disturbance suppression when controlling the complex model

$$P(s) = \frac{1+10s}{(1+s)(1+100s)}. \qquad (6)$$

First, identify a FOPDT model $M_1(s)$ with gain μ_1, delay L_1 and time constant T_1, by fitting its step response, compute the corresponding λ_{inf1}, and then the quantity $r_1 = \lambda_{inf1}/T_1$. Then, since the problem requires a control band as wide as possible, start decreasing the model time constant T, computing the corresponding λ_{inf}, until the quantity λ_{inf}/T either becomes less than $r_1/5$, or stops decreasing (this is a first-cut criterion but works quite well). Denoting by T_2 the so obtained time constant, and by λ_{inf2} the corresponding λ_{inf}, apply the IMC rules with $M_2(s) = \mu_1 e^{-sL_1}/(1 + sT_2)$, and $\lambda = 4\lambda_{inf2}$. In figure 5, this regulator is compared to that obtained with $M_1(s)$, and $\lambda = 4\lambda_{inf1}$; numbers are omitted for brevity. Apparently $M_2(s)$ does

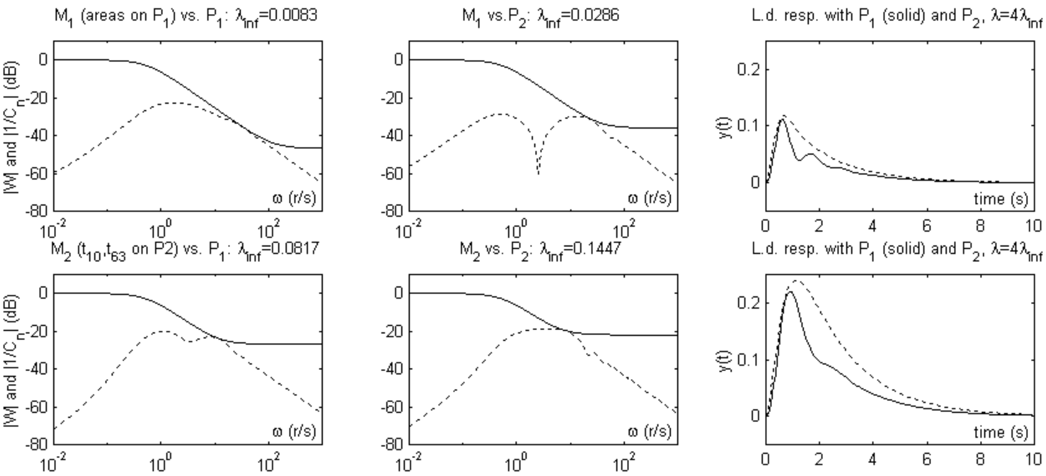

Fig. 4. Usefulness of model error information to reduce the influence of the identification method on the tuning results.

not fit the data as $M_1(s)$, but it is far more fit to the problem.

In synthesis, we can state the following. First, in model-based tuning, the identification method (though seldom discussed) plays a crucial role. There is simply no point in discussing a tuning method without considering the identification phase. Another important fact is the strong interplay between the choice of the tuning model and the significance of the tuning parameters. The latter relies on the correctness of the former, in that if the model is 'unfit' for the problem, sometimes large changes of the design parameters have almost no effect, while in some other cases the same parameters are very critical. When the model fits the problem, conversely, there is continuity and proportionality between the change of a tuning parameter and its effect. This is very important to make a method acceptable in the application domain. Second, allowing for a selectable tuning model *structure* would reduce the criticality of the identification method. The difficulty lies in the derivation of tuning rules for complex models(Horn *et al.*, 1996; Isaksson and Graebe, 1999). Third, the use of model error information is mandatory to limit conservatism to a reasonable amount. A nonstructured approach would be preferable, as proven in (Leva and Colombo, 2001*a*), but also identifying a complex model and deriving the tuning model(s) from it, estimating the model error with respect to the complex model as if it were the process, can help a lot.

The proposed 'dual model' approach comes from these considerations, and appears very promising. There are some problems still open, however, and these are briefly discussed in the following.

6. OPEN PROBLEMS

One could think that the main problem is choosing the structure of the complex model. Curiously, experience shows that this is not true. In practice, there is no need to allow for arbitrarily complex models: an order of 3 or 4 is enough to represent virtually any process one may encounter (and is also enough to prove the usefulness of the proposed approach, as for such dynamics a FOPDT tuning model chosen 'unconsciously' can be *very* inadequate).

A really difficult question is how to employ model error information effectively. The performance improvement illustrated by figures 2 and 3 is limited by the use of robustness criteria based on magnitude relationships, as those criteria guarantee that the regulator stabilises the tuning model, the complex model, and as a consequence a number of other systems it will never have to deal with. Apparently this leaves room for a further conservatism reduction, but the task is far from trivial.

In addition, the choice of the tuning model is quite easy to do for an expert human, but it is not easy to imagine how that choice could be automated. This problem does not appear as difficult as the previous one, but surely deserves further attention. Some help in this respect may come from pattern recognition techniques based on soft computing.

7. CONCLUSIONS

After discussing some seldom addressed aspects of model-based (auto)tuning, the conclusion was reached that it would be useful to reconsider in depth the role of the process model used for the

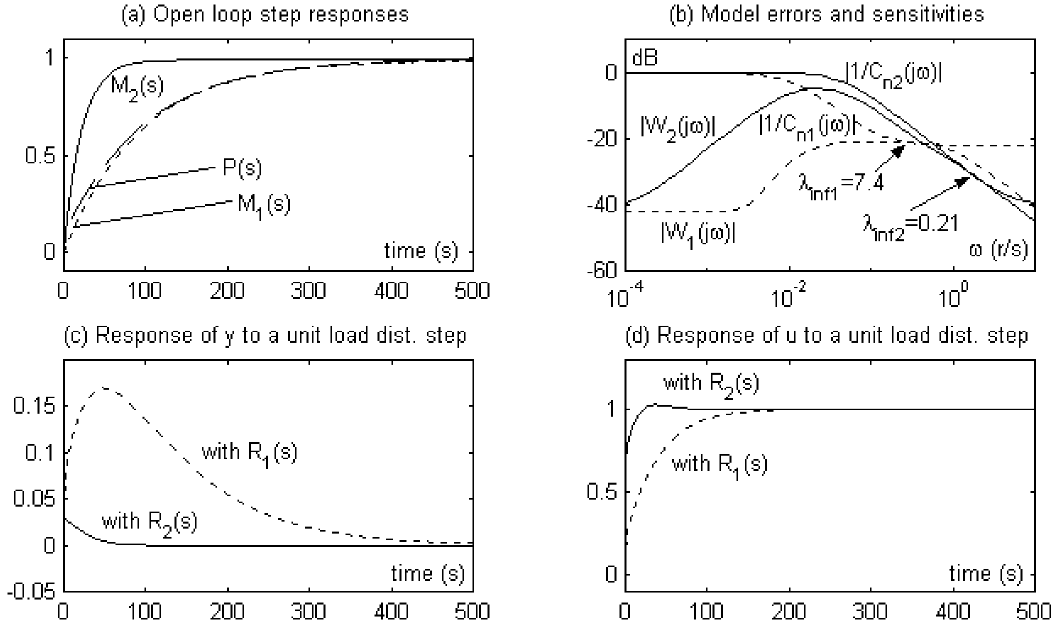

Fig. 5. Reasoned choice among different FOPDT models for the same process; in plot (b) $W_{e,i}(j\omega) = P(j\omega) - M_{t,i}(j\omega)$, $C_{t,i}(j\omega) = \frac{1+sT_i}{\mu_1(1+s\lambda_{infi})}$, $i = 1, 2$.

tuning. Based on the analysis of some conveniently chosen illustrative examples, a new perspective in this respect was proposed, consisting essentially in the adoption od a 'dual model approach' where a complex model is identified and one or mode simpler tuning models are derived from it. The possible advantages of the proposed approach were evidenced, together with some problems that are still open for future research.

8. REFERENCES

Åström, K.J. and T. Hägglund (1995). *PID controllers: theory, design and tuning - second edition.* Instrument Society of America. Research Triangle Park, NY.

Doyle, J.C., B.A. Francis and A.R. Tannenbaum (1992). *Feedback control theory.* MacMillan. Basingstoke, UK.

EnTech Control Engineeering Inc. (1992). Automatic controller dynamic specification, v. 1.0.

Harrold, D. (August 1999). Process controller tuning guidelines. *Control Engineering.*

Horn, C.C., J.R. Arulandu, C.J. Gombas, J.G. VanAntwerp and R.D. Braatz (1996). Improved filter design in internal model control. *Ind. Eng. Chem. Res.* **35**, 3437–3441.

Isaksson, A.J. and S.F. Graebe (1999). Analytical PID parameter expression for higher order systems. *Automatica* **35**, 1121–1130.

Leva, A. and A.M. Colombo (2001*a*). Estimating model mismatch overbounds for the robust autotuning of industrial regulators. *Automatica* **36**, 1855–1861.

Leva, A. and A.M. Colombo (2001*b*). IMC-based synthesis of the feedback block of ISA-PID regulators. *Proc. ECC 2001* pp. 125–247.

Morari, M. and E. Zafiriou (1989). *Robust process control.* Prentice-Hall. Upper Saddle River, NJ.

O'Dwyer, A. (2003). *Handbook of PI and PID controller tuning rules.* World Scientific Publishing. Singapore.