We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS BOOK CITATION INDEX INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

11

Model Predictive Control Relevant Identification

Rodrigo Alvite Romano[1], Alain Segundo Potts[2] and Claudio Garcia[2]
[1]*Instituto Mauá de Tecnologia - IMT*
[2]*Escola Politécnica da Universidade de São Paulo - EPUSP*
Brazil

1. Introduction

Model predictive control (MPC) is a multivariable feedback control technique used in a wide range of practical settings, such as industrial process control, stochastic control in economics, automotive and aerospace applications. As they are able to handle hard input and output constraints, a system can be controlled near its physical limits, which frequently results in performance superior to linear controllers (Maciejowski, 2002), specially for multivariable systems. At each sampling instant, predictive controllers solve an optimization problem to compute the control action over a finite time horizon. Then, the first of the control actions from that horizon is applied to the system. In the next sample time, this policy is repeated, with the time horizon shifted one sample forward. The optimization problem takes into account estimates of the system output, which are computed with the input-output data up to that instant, through a mathematical model. Hence, in MPC applications, a suitable model to generate accurate output predictions in a specific horizon is crucial, so that high performance closed-loop control is achieved. Actually, model development is considered to be, by far, the most expensive and time-consuming task in implementing a model predictive controller (Zhu & Butoyi, 2002).

This chapter aims at discussing parameter estimation techniques to generate suitable models for predictive controllers. Such a discussion is based on the most noticeable approaches in MPC relevant identification literature. The first contribution to be emphasized is that these methods are described in a multivariable context. Furthermore, the comparisons performed between the presented techniques are pointed as another main contribution, since they provide insights into numerical issues and the exactness of each parameter estimation approach for predictive control.

2. System identification for model predictive control

The dominating approach of the system identification techniques is based on the classical prediction error method (PEM) (Ljung, 1999), which is based on one-step ahead predictors. Predictive control applications demand models that generate reliable predictions over an entire prediction horizon. Therefore, parameters estimated from objective functions based on multi-step ahead predictors, generally result in better models for MPC applications (see Shook et al. (1991) and Gopaluni et al. (2004) for rigorous arguments). Since the last decade, an intense research has been done in order to develop system identification methods focused on providing appropriate models for model predictive control. Such methods are denoted

as model relevant identification (MRI) in the literature. Strictly speaking, MRI algorithms deal with the problem of estimating model parameters by minimizing multi-step objective functions.

Theoretically, if the model structure exactly matches the structure of the actual system, then the model estimated from a one-step ahead predictor is equivalent to the maximum likelihood estimate, which also provides optimal multi-step ahead predictions. However, in practice, even around an operating point, it is not possible to propose a linear model structure that exactly matches the system to be identified. Consequently, any estimated model has modeling errors associated with the identification algorithm. In these circumstances, models tuned for multi-step ahead predictions are more adequate for high closed-loop performance when using predictive controllers (Huang & Wang, 1999). In other words, when there is a certain amount of bias due to under-modeling (which is the more typical case), the MRI may be considered a way of distributing this bias in a frequency range that is less important for control purposes (Gopaluni et al., 2003).

Before formulating the parameter estimation problem in the MRI context, the discrete-time linear model structures to be used are specified.

2.1 Model parameterization

Consider a linear discrete-time system $\mathcal{S}$ with m inputs and p outputs

$$y(t) = G_0(q)u(t) + H_0(q)e(t)\,, \tag{1}$$

where $y(t)$ is the p-dimensional output column vector at sampling instant t, $u(t)$ is the m-dimensional input column vector and $e(t)$ is a p-dimensional zero-mean white noise column vector with a $p \times p$ diagonal covariance matrix R. The system $\mathcal{S}$ is characterized by the filter matrices $G_0(q)$ and $H_0(q)$. The process[1] and the noise models of $\mathcal{S}$ are denoted by $G(q,\theta)$ and $H(q,\theta)$, respectively. In this work, the system model is represented using matrix fraction descriptions (MFD) of the form

$$G(q,\theta) = F^{-1}(q)B(q) \tag{2}$$

$$H(q,\theta) = D^{-1}(q)C(q)\,. \tag{3}$$

where $B(q)$, $C(q)$, $D(q)$ and $F(q)$ are matrices of polynomials in the shift operator q with dimensions $p \times m$, $p \times p$, $p \times p$ and $p \times p$, respectively. The parameter vector θ is composed of the coefficients of the polynomials in such matrices. Thus, in order to determine θ, one needs to further specify the polynomial matrices in (2) and (3). The matrix $B(q)$ takes the form

$$B(q) = \begin{bmatrix} B_{11}(q) & \cdots & B_{1m}(q) \\ \vdots & \ddots & \vdots \\ B_{p1}(q) & \cdots & B_{pm}(q) \end{bmatrix}\,, \tag{4}$$

whose entries are $\mu_{ij} - 1$ degree polynomials

$$B_{ij}(q) = b_{ij}^{(1)}q^{-1} + \ldots + b_{ij}^{(\mu_{ij})}q^{-\mu_{ij}}\,,$$

[1] Sometimes (Ljung, 1999; Zhu, 2001, e.g.), the process model is referred to as transfer function.

for $i \in \{1,\ldots,p\}$ and $j \in \{1,\ldots,m\}$. One of the simplest choice to parameterize the other matrices is through the diagonal form MFD, in which $C(q)$, $D(q)$ and $F(q)$ are diagonal polynomial matrices and their nonzero polynomials are all monic, e.g.,

$$F(q) = \begin{bmatrix} F_{11}(q) & 0 & \cdots & 0 \\ 0 & F_{22}(q) & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & F_{pp}(q) \end{bmatrix}, \tag{5}$$

where the entries of $F(q)$ are ν_i degree polynomials of the form

$$F_{ii}(q) = 1 + f_{ii}^{(1)} q^{-1} + \ldots + f_{ii}^{(\nu_i)} q^{-\nu_i},$$

for each $i \in \{1,2,\ldots,p\}$. The diagonal matrices $C(q)$ and $D(q)$, as well as their respective entries, are defined analogously.

When the diagonal form is adopted, it is possible to decouple the multi-input multi-output model into a set of p multi-input single-output (MISO) models in the form

$$\begin{aligned} y_1(t) &= F_{11}^{-1}(q) \sum_{j=1}^{m} B_{1j}(q) u_j(t) + \frac{C_{11}(q)}{D_{11}(q)} e_1(t) \\ &\vdots = \vdots \\ y_p(t) &= F_{pp}^{-1}(q) \sum_{j=1}^{m} B_{pj}(q) u_j(t) + \frac{C_{pp}(q)}{D_{pp}(q)} e_p(t), \end{aligned} \tag{6}$$

in which y_i and u_j denote the i^{th} output and the j^{th} input, respectively.

Unless otherwise stated, it is assumed that all the nonzero polynomials of the matrices have the same degree n, that is to say $\mu_{ij} = \nu_i = n$, for $i \in \{1,\ldots,p\}$ and $j \in \{1,\ldots,m\}$. Although this degree is in general not the same as the McMillan degree, this choice considerably simplifies the order selection problem and, consequently, makes the model structure more suitable for applications in large scale processes.

Besides being simple to understand, the diagonal form has some relevant properties for applications in system identification (Zhu, 2001). The main of them is that algorithms developed for the SISO (single-input single-output) processes can be directly generalized for the multivariable case. Nevertheless, if there are dynamic iterations between different outputs, the estimated model based on the diagonal form can present a larger bias error (Laurí et al., 2010). Alternatively, one can add elements outside the diagonal of $F(q)$, not necessarily monic polynomials, with the purpose of incorporating the dynamic iteration between the process outputs. This approach gives rise to another MFD named "full polynomial form" (Ljung, 1999), in which any $F(q)$ entry may be nonzero. This parameterization is also employed in one of the identification methods described in Section 3.

Next, the multi-step objective function used as the basis for the development of the MRI algorithms is presented.

2.2 The model relevant identification cost function

Firstly, let us define the $p \times p$ filter matrix

$$W_k(q,\theta) \triangleq \left(\sum_{l=0}^{k-1} h(l) q^{-l} \right) H^{-1}(q,\theta) \,, \tag{7}$$

where $h(l)$ is the l^{th} impulse response coefficient of $H(q,\theta)$.

Thus, the k-step ahead predictor of the output vector (i.e., the output prediction equation at $t+k$ with data available up to instant t) may be expressed as (Ljung, 1999)

$$\hat{y}(t+k|t,\theta) = W_k(q,\theta) G(q,\theta) u(t+k) + \left(I - W_k(q,\theta)\right) y(t+k) \,. \tag{8}$$

According to (8), the k-step ahead prediction error is

$$\begin{aligned} \varepsilon(t+k|t,\theta) &= y(t+k) - \hat{y}(t+k|t,\theta) \\ &= W_k(q,\theta) \left(y(t+k) - G(q,\theta) u(t+k) \right) \,. \end{aligned} \tag{9}$$

From (7)-(9), note that the k-step prediction error is related to the one-step through the filter matrix

$$L_k(q,\theta) \triangleq \sum_{i=0}^{k-1} h(i) q^{-i} \,, \tag{10}$$

such that

$$\varepsilon(t+k|t) = L_k(q,\theta) \varepsilon(t+k|t+k-1) \,. \tag{11}$$

As argued previously, the main objective of the MRI methods is to provide models that are optimized for the generation of predictions over an entire prediction horizon. So, a natural choice for the criterion of the parameter estimation problem is the cost function

$$J_{\text{multi}}(P,\theta) = \sum_{k=1}^{P} \sum_{t=0}^{N-k} \| \varepsilon(t+k|t,\theta) \|_2^2 \,, \tag{12}$$

where $\| \cdot \|_2$ denotes the ℓ_2 norm. Hence, $J_{\text{multi}}(P,\theta)$ quantifies the mean-square error, based on predictions ranging from 1 to P steps ahead in a dataset of length N.

The challenge in estimating the model parameters by minimizing (12) is that such a criterion is highly nonlinear in the model parameters. Therefore, suitable optimization algorithms are necessary, so that local minima or convergence problems are avoided. Strictly speaking, the identification methods to be presented aims at estimating the model parameters based on J_{multi}.

3. Model parameter estimation methods

In recent years, distinct MRI techniques were proposed based on different principles. One of them, conceived by Rossiter & Kouvaritakis (2001), differs from the others since it proposes the use of multiple models to generate the predictions. Thus, an optimized model is estimated for each k-step ahead prediction. In spite of providing "optimal" predictions for the entire horizon, the number of parameters involved can be quite large, specially for multi-input and multi-output processes. It is known that the variance of the parameter estimates is

proportional to the ratio between the number of parameters and the dataset length (Ljung, 1999). Hence, the main drawback of the multi-model approach is the amount of data required to estimate a reasonable model set. Such amount of data may be prohibitive in practical situations (Gopaluni et al., 2003). Moreover, most of the MPC algorithms are based on a single model. For these reasons, the multi-model method is not considered in further analysis.

In the pioneering work by Shook et al. (1991), the MRI is performed in the context of data prefiltering using SISO ARX (Auto Regressive with eXternal input) type models. Huang & Wang (1999) extended the previous method, so that a general model structure (e.g., Box-Jenkins) could be employed. Some authors, such as (Gopaluni et al., 2003; Laurí et al., 2010), deal with the parameter estimation problem directly minimizing the MRI cost function, using nonlinear optimization techniques. In another approach, proposed by Gopaluni et al. (2004), the focus is given to the noise model parameter estimation. In this approach, a non-parsimonious process model is estimated, in order to eliminate bias errors (which are caused by under-modeling). Then, with a fixed process model, the parameters of the noise model are obtained by minimizing the cost function (12).

In the following subsections, the main MRI techniques are described in more details.

3.1 The prefiltering approach

3.1.1 The basic idea

For the sake of simplicity, the basic idea behind the prefiltering approach is shown using the SISO case ($m = p = 1$). Nevertheless, its worth mentioning that the conclusions directly apply to MIMO models represented in the diagonal form MFD.

In this case, based on predictor (9), the MRI cost function (12) can be rewritten as

$$J_{\text{multi}}(P,\theta) = \sum_{k=1}^{P} \sum_{t=0}^{N-k} \left(\frac{L_k(q,\theta)}{H(q,\theta)} \left(y(t+k) - G(q,\theta)u(t+k) \right) \right)^2 . \tag{13}$$

If we introduce an auxiliary variable $\tilde{G}(q,\theta)$ that takes into account the deterministic model mismatch, that is

$$\tilde{G}(q,\theta) \triangleq G_0(q) - G(q,\theta) ,$$

then, substituting (1) into (13) gives

$$\begin{aligned} J_{\text{multi}}(P,\theta) &= \sum_{k=1}^{P} \sum_{t=0}^{N-k} \left(\frac{L_k(q,\theta)}{H(q,\theta)} \left(\tilde{G}(q,\theta)u(t+k) + H_0(q)e(t+k) \right) \right)^2 \\ &= \sum_{k=1}^{P} \sum_{t=0}^{N-k} \left(\frac{L_k(q,\theta)}{H(q,\theta)} \left(\begin{bmatrix} \tilde{G}(q,\theta) & H_0(q) \end{bmatrix} \begin{bmatrix} u(t+k) \\ e(t+k) \end{bmatrix} \right) \right)^2 . \end{aligned} \tag{14}$$

Supposing $N \to \infty$ and applying Parseval's relationship to (14) yields

$$\begin{aligned} J_{\text{multi}}(P,\theta) = \sum_{k=1}^{P} \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \frac{L_k(e^{j\omega},\theta)}{H(e^{j\omega},\theta)} \right|^2 \begin{bmatrix} \tilde{G}(e^{j\omega},\theta) & H_0(e^{j\omega}) \end{bmatrix} \times \\ \begin{bmatrix} \Phi_u(\omega) & \Phi_{eu}(\omega) \\ \Phi_{ue}(\omega) & R \end{bmatrix} \begin{bmatrix} \tilde{G}(e^{-j\omega},\theta) \\ H_0(e^{-j\omega}) \end{bmatrix} d\omega , \end{aligned}$$

where $\Phi_u(\omega)$ is the power spectrum of $u(t)$ and $\Phi_{eu}(\omega)$ is the cross-spectrum between $e(t)$ and $u(t)$. Now, moving the summation to the inside of the integral, it follows that

$$J_{\text{multi}}(P,\theta) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\sum_{k=1}^{P}\left|L_k(e^{j\omega},\theta)\right|^2}{\left|H(e^{j\omega},\theta)\right|^2} \begin{bmatrix} \tilde{G}(e^{j\omega},\theta) & H_0(e^{j\omega}) \end{bmatrix} \times \begin{bmatrix} \Phi_u(\omega) & \Phi_{eu}(\omega) \\ \Phi_{ue}(\omega) & R \end{bmatrix} \begin{bmatrix} \tilde{G}(e^{-j\omega},\theta) \\ H_0(e^{-j\omega}) \end{bmatrix} d\omega \,. \tag{15}$$

From (15) one can see that the deterministic model mismatch is weighted by the input spectrum, while the filter

$$W_{\text{multi}}(e^{j\omega},\theta) = \sum_{k=1}^{P}\left|W_k(e^{j\omega},\theta)\right|^2 = \frac{\sum_{k=1}^{P}\left|L_k(e^{j\omega},\theta)\right|^2}{\left|H(e^{j\omega},\theta)\right|^2} \tag{16}$$

weights the whole expression. But, if P is limited to 1, which implies considering only one-step ahead predictions, we obtain

$$J_{\text{multi}}(P,\theta)\Big|_{P=1} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{1}{\left|H(e^{j\omega},\theta)\right|^2} \begin{bmatrix} \tilde{G}(e^{j\omega},\theta) & H_0(e^{j\omega}) \end{bmatrix} \times \begin{bmatrix} \Phi_u(\omega) & \Phi_{eu}(\omega) \\ \Phi_{ue}(\omega) & R \end{bmatrix} \begin{bmatrix} \tilde{G}(e^{-j\omega},\theta) \\ H_0(e^{-j\omega}) \end{bmatrix} d\omega \,. \tag{17}$$

Comparing (17) with (15), it is observed that the latter is identical to the first weighted by the frequency function

$$L_{\text{multi}}(e^{j\omega},\theta) = \sum_{k=1}^{P}\left|L_k(e^{j\omega},\theta)\right|^2 \,. \tag{18}$$

Hence, the estimation of the model parameters by minimizing the MRI cost function (15) is equivalent to using standard one-step ahead prediction error estimation algorithms (available in software packages, such as Ljung (2007)) after prefiltering the data with (18). As the prefiltering affects the model bias distribution and may also remove disturbances of frequency ranges that one does not want to include in the modeling, the role of the prefilter may be interpreted as a frequency weighting optimized for providing models suitable for multi-step ahead predictions.

3.1.2 Algorithms and implementation issues

Although the prefiltering artifice is an alternative to solve the problem of parameter estimation in the context of MRI, there is a point to be emphasized: the prefilter $L_{\text{multi}}(q,\theta)$ in (18) depends on the noise model $H(q,\theta)$, which is obviously unknown.

An iterative procedure called LRPI (Long Range Predictive Identification) to deal with the unknown noise model was proposed by (Shook et al., 1991). As mentioned previously, in the original formulation only the SISO case based on the ARX structure was concerned. Next, the LRPI algorithm is extended to the multivariable case. To this end, the following is adopted

$$G(q,\theta) = A^{-1}(q)B(q) \tag{19}$$

$$H(q,\theta) = A^{-1}(q)\,, \tag{20}$$

where the polynomial matrix $A(q)$, as well as its entries, are defined analogously to (5). According to (19)-(20), the i^{th} output equation may be expressed by

$$A_{ii}(q)y_i(t) = \sum_{j=1}^{m} B_{1j}(q)u_j(t) + e_i(t) \; . \tag{21}$$

Consider the regression $\varphi_i(t) \in \mathbb{R}^{n(m+1)}$ and the parameter $\theta_i \in \mathbb{R}^{n(m+1)}$, relative to the i^{th} system output

$$\begin{aligned} \varphi_i(t) = [&-y_i(t-1), \cdots, -y_i(t-n), u_1(t-1), \cdots, u_m(t-1), \\ &\cdots, u_1(t-n), \cdots, u_m(t-n)]^T \end{aligned} \tag{22}$$

$$\theta_i = \left[a_{ii}^{(1)}, \cdots, a_{ii}^{(n)}, b_{i1}^{(1)}, \cdots, b_{im}^{(1)}, \cdots, b_{i1}^{(n)}, \cdots, b_{im}^{(n)}\right]^T \; . \tag{23}$$

From (22) and (23), the one-step ahead prediction of $y_i(t)$ may be expressed as

$$\hat{y}_i(t+1|t, \theta_i) = \varphi_i^T(t)\theta_i \; . \tag{24}$$

Algorithm 1: Extension of the LRPI algorithm to the multivariable case

Step 1. Set $i = 1$ (that is, only the first output is considered).

Step 2. Initialize $L_{\text{multi},i}(q)$ to 1.

Step 3. Filter $y_i(t)$ and each input $u_j(t)$ for $j \in \{1, \ldots, m\}$ with $L_{\text{multi},i}(q)$, i.e.

$$y_i^f(t) \triangleq L_{\text{multi},i}(q)y_i(t) \tag{25}$$

$$u^f(t) \triangleq \begin{bmatrix} L_{\text{multi},i}(q) & 0 & \cdots & 0 \\ 0 & L_{\text{multi},i}(q) & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & L_{\text{multi},i}(q) \end{bmatrix} u(t) \; . \tag{26}$$

Step 4. Based on (25)-(26), construct the regression vector analogously to (22), so that

$$\varphi_i^f(t) = \left[-y_i^f(t-1), \cdots, -y_i^f(t-n), u^{f^T}(t-1), \cdots, u^{f^T}(t-1)\right]^T \; . \tag{27}$$

Step 5. Estimate the parameter vector θ_i by solving the linear least-squares problem

$$\hat{\theta}_i = \arg\min_{\theta_i} \sum_t \left(y_i(t) - \varphi_i^{f^T}(t)\theta_i\right)^2 \; . \tag{28}$$

Step 6. Update $L_{\text{multi},i}(q)$ through (10) and (18), based on the noise model $A_{ii}^{-1}(q)$ estimated in the previous step.

Step 7. Continue if convergence of θ_i occurs, otherwise go back to Step 3.

Step 8. If $i \neq p$, go back to Step 2, with $i = i + 1$. Otherwise, concatenate the estimated models into a MIMO representation.

Remarks:

- For the multi-output case, there are p different filters $L_{\text{multi}}(q)$, each one associated with the i^{th} output and denoted by $L_{\text{multi},i}(q)$.
- With respect to Step 6, as $L_{\text{multi}}(q)$ is a spectral factor of $L_{\text{multi}}(e^{j\omega})$, spectral factorization routines, such as the one proposed in Ježek & Kučera (1985), can be used for solving (18).
- A natural choice to determine the convergence of the algorithm is to check if the ℓ_2 norm of the difference between the parameter estimates in two consecutive iterations is less than δ. Experience has shown that a reasonable value for δ is 10^{-5}.

Alternatively, instead of using an iterative procedure as previously, in the method proposed by Huang & Wang (1999) named MPEM (Multi-step Prediction Error Method), a fixed noise model estimate is employed in order to get $L_{\text{multi}}(q)$. In what follows, the multi-step prediction error algorithm is described, based on the MFD parameterized by (2)-(5).

Algorithm 2: MPEM algorithm based on the diagonal form matrix fraction description

Step 1. Set $i = 1$.

Step 2. Get initial estimates of $C_{ii}(q)$, $D_{ii}(q)$, $F_{ii}(q)$ and, for $j \in \{1, \ldots, m\}$, $B_{ij}(q)$, using standard prediction error methods, namely, based on a one-step ahead cost function (17).

Step 3. Use a spectral factorization routine to solve (18), in which the filters defined in (10) are calculated through the impulse response of the estimated noise model $\hat{D}_{ii}^{-1}(q)\hat{C}_{ii}(q)$.

Step 4. Filter $y_i(t)$ and each input $u_j(t)$, $j \in \{1, \ldots, m\}$, with $\hat{C}_{ii}^{-1}(q)\hat{D}_{ii}(q)L_{\text{multi},i}(q)$.

Step 5. With the fixed noise model $\hat{D}_{ii}^{-1}(q)\hat{C}_{ii}(q)$, calculated in Step 2, estimate $B_{i1}(q)$,..., $B_{im}(q)$, $F_{ii}(q)$ by minimizing the output-error cost function

$$V_{\text{oe},i}\left(B_{i1}(q), \ldots, B_{im}(q), F_{ii}(q)\right) = \sum_{t} \left(y_i^f(t) - F_{pp}^{-1}(q) \sum_{j=1}^{m} B_{1j}(q) u_j^f(t) \right)^2 . \qquad (29)$$

Step 6. If $i \neq p$, go back to Step 2, with $i = i + 1$. Otherwise, concatenate the estimated models into a multi-output representation.

Remarks:

- Once more the diagonal form MFD property, which allows the independent treatment of each model output, is applied to extend the parameter estimation algorithm to the multivariable framework.
- The prefilters of Step 2 differ from the ones used in the LRPI algorithm by the additional terms $\hat{C}_{ii}^{-1}(q)\hat{D}_{ii}(q)$, each one for $i \in \{1, \ldots, p\}$, which represents the inverse of the i^{th} output noise model. Hence, while the filters $L_{\text{multi},i}(q)$ aim at providing optimal weighting for multi-step predictions, the additional terms intend to remove the noise influence for models represented as (6).

- The minimization of (12) is replaced by two nonlinear optimization problems in the MPEM algorithm. At first, it might seem that there is no relevant advantage in such an approach. Nevertheless, it is important to say that the MISO Box-Jenkins identification from Step 2, as well as the minimization of the output-error cost function in (29), can be performed using available software packages (Ljung, 2007, e.g.). Moreover, for models parameterized as (2)-(5), the numerical complexity of these problems are considered to be lower than the one of minimizing J_{multi} directly.

The LRPI algorithm involves only linear least-squares problems, which have many advantages. The most important one being that (28) can be solved efficiently and unambiguously (Ljung, 1999). The price paid for a simple parameter estimation algorithm is the adoption of a limited noise model structure. Consequently, the estimate of the $H(q,\theta)$ entries may be inaccurate, which affects the calculation of each filter $L_{\text{multi},i}(q)$. In turn, MPEM considers a more flexible noise model structure. However, local minima or convergence issues due to nonlinear optimization methods in Steps 2 and 5 may degrade the quality of the estimates. Therefore, the MPEM should outperform the LRPI algorithm, provided that the global minimum is achieved in the estimation steps. Anyway, it is suggested that models are estimated using more than one method and select the one which yields the best multi-step ahead predictions.

3.2 Direct optimization of the cost function

In the prefilter approach described previously, the filters $L_{\text{multi},i}(q)$ are calculated using any spectral factorization routine. Hence, as these filters are approximations of (18), the identified model ability to generate multi-step ahead predictions depends on the degree of the approximation and on the accuracy of the disturbance model estimate. But there is no need to worry about these aspects if the MRI cost function (12) is minimized directly. On the other hand, the model parameterization should be chosen carefully, to minimize numerical problems in the nonlinear optimization algorithm. In Laurí et al. (2010) a "full-polynomial[2] form" ARX model

$$A(q)y(t) = B(q)u(t) + e(t)\,, \tag{30}$$

with

$$A(q) = \begin{bmatrix} A_{11}(q) & \cdots & A_{1p}(q) \\ \vdots & \ddots & \vdots \\ A_{p1}(q) & \cdots & A_{pp}(q) \end{bmatrix} = I + A^{(1)}q^{-1} + \ldots + A^{(n)}q^{-n}\,, \tag{31}$$

whose entries are

$$A_{ij}(q) = \begin{cases} 1 + a_{ij}^{(1)}q^{-1} + \ldots + a_{ij}^{(n)}q^{-n} & \text{, for } i = j \\ a_{ij}^{(1)}q^{-1} + \ldots + a_{ij}^{(n)}q^{-n} & \text{, otherwise} \end{cases}$$

and the polynomial matrix $B(q)$ is defined as in (4).

[2] Note that, in order to consider output interaction, the polynomial matrix $A(q)$ is not restricted to being diagonal, as in the LRPI algorithm.

For this model structure, let us introduce the parameter matrix

$$\Theta = \begin{bmatrix} a_{11}^{(1)}, \cdots, a_{1p}^{(1)}, \cdots, a_{11}^{(n)}, \cdots, a_{1p}^{(n)}, b_{11}^{(1)}, \cdots, b_{1m}^{(1)}, \cdots, b_{11}^{(n)}, \cdots, b_{1m}^{(n)} \\ \vdots \\ a_{p1}^{(1)}, \cdots, a_{pp}^{(1)}, \cdots, a_{p1}^{(n)}, \cdots, a_{pp}^{(n)}, b_{p1}^{(1)}, \cdots, b_{pm}^{(1)}, \cdots, b_{p1}^{(n)}, \cdots, b_{pm}^{(n)} \end{bmatrix}^T \in \mathbb{R}^{n(m+p)\times p} \tag{32}$$

and a particular regression vector denoted by $\check{\varphi}(t+k|t,\Theta) \in \mathbb{R}^{n(m+p)}$, which is composed of inputs up to instant $t+k$, output data up to t and output estimates from $t+1$ to $t+k-1$, for instance

$$\check{\varphi}(t+2|t,\Theta) = \left[-\hat{y}^T(t+1|t,\Theta), -y^T(t), \cdots, -y^T(t-n+2), u^T(t+1), \cdots, u^T(t-n+2)\right]^T$$

and for an arbitrary k

$$\check{\varphi}(t+k|t,\Theta) = \left[-\check{y}^T(t+k-1|t), \cdots, -\check{y}^T(t+k-n|t), u^T(t+k-1), \cdots, u^T(t+k-n)\right]^T, \tag{33}$$

where

$$\check{y}^T(s|t) \triangleq \begin{cases} \hat{y}(s|t,\Theta) & \text{, for } s > t \\ y(s) & \text{, otherwise.} \end{cases}$$

From (32) and (33), the k-step ahead prediction of $y(t)$ is given by

$$\hat{y}(t+k|t,\Theta) = \Theta^T \check{\varphi}(t+k|t,\Theta)\,. \tag{34}$$

Although the predictor $\hat{y}(t+k|t,\Theta)$ is nonlinear in the parameters, it is important to notice that it can be calculated recursively, from $\hat{y}(t+1|t)$ for $k \in \{2,\ldots,P\}$ using (34). This is the main reason why the ARX structure was adopted. For another thing, if a more flexible model structure is adopted, the k-step ahead predictor equation would be much more complex.

Thus, based on the MRI cost function (12), the parameter estimation can be stated as a nonlinear least-squares problem

$$\hat{\Theta} = \arg\min_{\Theta} \sum_{k=1}^{P} \sum_{t=0}^{N-k} \|y(t) - \Theta^T \check{\varphi}(t+k|t,\Theta)\|_2^2\,, \tag{35}$$

which must be solved numerically. The Levenberg-Marquart algorithm is used in Laurí et al. (2010) in order to minimize (35).

3.3 Optimization of the noise model

In Gopaluni et al. (2004) it is shown that, in the absence of a noise model, there is no significant difference between MRI and one-step ahead prediction error methods. On the other hand, when the signal to noise ratio is small, the one-step ahead predictors yield worse results for P-step ahead predictions than MRI methods. Thus, in these circumstances, a suitable disturbance model is crucial to generate accurate multi-step ahead predictions.

Any identified model has bias and variance errors associated with the identification algorithm. While the former is typically associated to model mismatch (such a mismatch can be either in

the process model or in the noise model) and the second is due to the effect of unmeasured disturbances. If that there is no significant cross-correlation between the noise and system input (in open-loop, e.g.), the bias errors in the process model may be eliminated by using high order FIR (Finite Impulse Response) models (Zhu, 2001). Under that assumption, the modeling errors are restricted to the noise model.

With this in mind, in Gopaluni et al. (2004) the authors propose a two-step MRI algorithm in which the process is represented by a FIR structure, with sufficiently high order so that bias errors due to the process model can be disregarded. Then, the noise model parameters are estimated using a multi-step cost function.

Consider the multivariable FIR model

$$G_{\mathrm{FIR}}(q,\theta) = B(q) \tag{36}$$

where the polynomial matrix $B(q)$ is defined as in (4). The noise model $H(q,\eta)$ is parameterized using diagonal MFD. These choices are equivalent to (6) with $F(q) = I$. As the estimation of $G(q)$ and $H(q)$ are performed separately, in this subsection, the parameter vector is split into two parts, such that the noise model parameter vector is explicitly referred to as η. So, the i^{th} output noise model structure is

$$H_{ii}(q,\eta_i) \triangleq \frac{C_{ii}(q)}{D_{ii}(q)} = \frac{1 + c_{ii}^{(1)} q^{-1} + \ldots + c_{ii}^{(\alpha_i)} q^{-\alpha_i}}{1 + d_{ii}^{(1)} q^{-1} + \ldots + d_{ii}^{(\beta_i)} q^{-\beta_i}} \,. \tag{37}$$

Let us introduce the residual of the process model relative to the i^{th} output

$$v_i(t) \triangleq y_i(t) - \sum_{j=1}^{m} B_{ij}(q) u_j(t) \,. \tag{38}$$

Then, based on (7)-(9), the k-step ahead prediction error of the i^{th} output[3] can be written as

$$\begin{aligned} \varepsilon_i(t+k|t,\theta) &= y_i(t+k) - \hat{y}_i(t+k|t,\theta) \\ &= \left(\sum_{l=0}^{k-1} h_i(l) q^{-l} \right) \frac{D_{ii}(q)}{C_{ii}(q)} v_i(t) \,. \end{aligned} \tag{39}$$

As $C_{ii}(q)$ and $D_{ii}(q)$ are monic polynomials, the impulse response leading coefficient $h_i(0)$ is always 1. With this, expanding (39) yields

$$\begin{aligned} \varepsilon_i(t+k|t,\theta) = {} & v_i(t+k) + h_i(1) v_i(t+k-1) + \ldots + h_i(k-1) v_i(t+1) \\ & - c_{ii}^{(1)} \varepsilon_i(t+k-1|t,\theta) - \ldots - c_{ii}^{(\alpha_i)} \varepsilon_i(t+k-\alpha_i|t,\theta) \\ & + d_{ii}^{(1)} L_{k,i}(q) v_i(t+k-1) + \ldots + d_{ii}^{(\beta_i)} L_{k,i}(q) v_i(t+k-\beta_i) \,. \end{aligned} \tag{40}$$

[3] Part of the notation introduced in Section 2.2 is particularized here to the single-output context. For instance, $h_i(l)$ and $L_{k,i}(q)$ are the equivalent to the ones defined in (10), but related to the i^{th} output.

For a compact notation, we define

$$\eta_{k,i} \triangleq \left[h_i(1), \ldots, h_i(k-1), 1, c_{ii}^{(1)}, \ldots, c_{ii}^{(\alpha_i)}, d_{ii}^{(1)}, \ldots, d_{ii}^{(\beta_i)}\right]^T \tag{41}$$

$$\varphi_{k,i}(t, \eta_{k,i}) \triangleq \Big[-v_i(t+k-1), \ldots, -v_i(t+1), y_i(t+k) - v_i(t+k), \varepsilon_i(t+k-1|t,\theta), \ldots, \varepsilon_i(t+k-\alpha_i|t,\theta), -\xi_i(t+k-1), \cdots, -\xi_i(t+k-\beta_i)\Big]^T \tag{42}$$

where

$$\xi_i(t+k) \triangleq F_{k,i}(q) v_i(t+k) \, .$$

Then, we can rewrite (39) as

$$\varepsilon_i(t+k|t) = y_i(t+k|t) - \varphi_{k,i}^T(t, \eta_{k,i}) \eta_{k,i} \, . \tag{43}$$

In light of the aforementioned paragraphs, the MRI algorithm that optimizes the noise model is summarized as follows.

Algorithm 3: MRI with optimized noise model

Step 1. Set $i = 1$.

Step 2. Fix an a priori noise model to the i^{th} output, for instance

$$\frac{C_{ii}(q)}{D_{ii}(q)} = 1$$

and estimate a multi-input single-output high order FIR model using standard PEM.

Step 3. With the estimate $\hat{G}_{\text{FIR}}(q)$, from the previous step, solve the optimization problem

$$\hat{\eta}_{P,i} = \arg\min_{\eta_{P,i}} \sum_{t=1}^{N-P} \sum_{k=1}^{P} \left(y_i(t+k|t) - \varphi_{k,i}^T(t, \eta_{k,i}) \eta_{k,i} \right)^2 \tag{44}$$

subject to

$$h_i(l) = \breve{h}_i(l)(\eta_i) \text{ , for any } l = \{1, 2, \ldots, P-1\} \tag{45}$$

where $\breve{h}_i(l)(\eta_i)$ indicates the l^{th} impulse response coefficient of (37), which is obtained by polynomial long division of $C_{ii}(q)$ by $D_{ii}(q)$.

Step 4. If $i \neq p$, go back to Step 2, with $i = i + 1$. Otherwise concatenate the estimated models into a single MIMO representation.

Remarks:

- Besides providing unbiased estimates under open-loop conditions, FIR models are suitable in this case because the parameters of G_{FIR} can be efficiently estimated using linear least-squares.
- A numerical optimization method is required to solve the parameter estimation problem. Nevertheless, the Levenberg-Marquart algorithm mentioned in the previous subsection can not deal with constraints. One of the nonlinear optimization algorithm possibilities

is the Sequential Quadratic Programming (SQP), which can handle nonlinear constraints such as (45). In Gopaluni et al. (2004) it is shown that if a noise model of the form

$$H_{ii}(q,\eta_i) = \frac{1 + c_{ii}^{(1)} q^{-1} + \ldots + c_{ii}^{(\alpha_i)} q^{-\alpha_i}}{1 - q^{-1}}$$

is adopted, then the constraint (45) may be expressed through a linear in the parameters equation. In this case, (44) can be solved using the standard Quadratic Programming (QP) method.

4. Simulations

The main features of the aforementioned MRI techniques are analyzed using two simulated examples. At first, a SISO process is considered in order to illustrate the influence of the prediction horizon length P in the modeling errors presented by the identified models. Moreover, the performance of each technique is evaluated based on datasets with distinct signal-to-noise ratios (SNR). After that, the closed-loop performance provided by the estimated models is assessed. To this end, the Quadratic Dynamic Matrix Controller (QDMC) (Camacho & Bordons, 2004) and a multivariable distillation column benchmark (Cott, 1995a;b) are employed.

4.1 SISO process example

Consider the third-order overdamped system proposed in Clarke et al. (1987)

$$G_0(q) = \frac{0.00768q^{-1} + 0.02123q^{-2} + 0.00357q^{-3}}{1 - 1.9031q^{-1} + 1.1514q^{-2} - 0.2158q^{-3}}\,, \tag{46}$$

with a random-walk disturbance, that is

$$H_0(q) = \frac{1}{1 - q^{-1}}\,. \tag{47}$$

The process is excited in open-loop by a Pseudo Random Binary Sequence (PRBS) switching between $[-0.1, 0.1]$ with a clock period of 5 times the sampling interval. The noise variance is adjusted such that the signal-to-noise ratio (SNR) is 3 (in variance). A record of 1200 samples is collected, which is shown in Fig. 1. The dataset is split into two halves: the first is used for estimation and the second one for validation purposes.

The following reduced-complexity model structure is assumed[4]

$$G(q,\theta) = \frac{b_1 q^{-1} + b_2 q^{-2}}{1 + a_1 q^{-1}} \tag{48}$$

$$H(q,\theta) = \frac{1 + c_{11}^{(1)} q^{-1}}{1 + d_{11}^{(1)} q^{-1}}. \tag{49}$$

[4] Except for the noise model optimization method (Subsection 3.3), in which $d_{11}^{(1)}$ is fixed to -1, so that parameter estimation can be handled using standard quadratic programming.

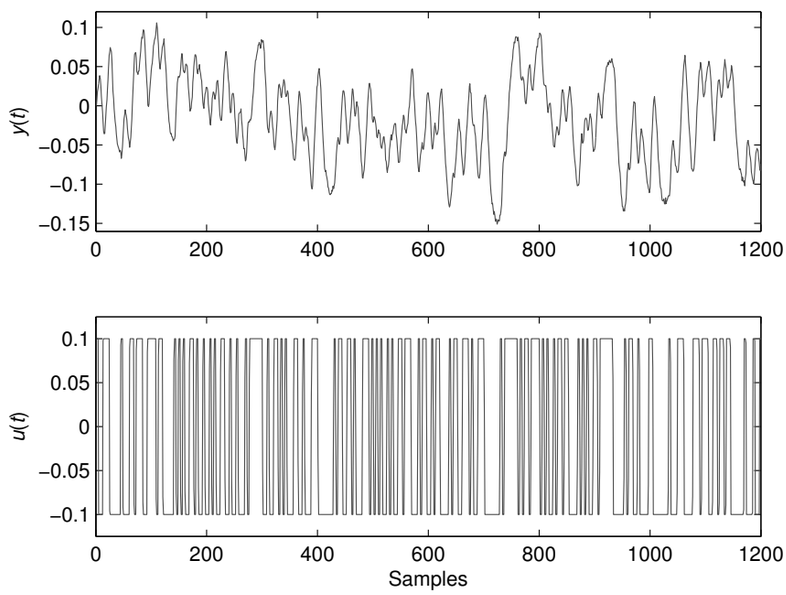

Fig. 1. The dataset from the third-order process (46)-(47).

Before analyzing the capacity of the estimated models to generate accurate multi-step ahead predictions, it is worth noting the influence of the prediction horizon length. The magnitudes of $L_{\text{multi}}(e^{j\omega})$, for $P = \{1, 2, 5, 10, 15\}$, are shown in Fig. 2. As can be seen, $L_{\text{multi}}(q)$ is a low-pass filter, whose cut-off frequency decreases as P increases. Such behavior occurs whenever the disturbance spectrum is concentrated on low frequencies (Gopaluni et al., 2003). Hence, according to (15), the higher the prediction horizon length, the narrower the error weighting.

As a consequence, an increase in P leads to lower modeling errors in low frequencies, but the frequency response of the estimated models are away from the actual one at high frequencies. This behavior is depicted in Fig. 3, which presents the absolute value of the difference between the actual and the estimated (from models obtained using the MPEM algorithm) frequency responses. One can also notice that the effect of increasing P is more prominent in the range $[1, 5]$ than between $[5, 15]$. Furthermore, as shown in Farina & Piroddi (2011), for sufficiently high values of the prediction horizon length, models estimated based on multi-step prediction errors converge to the output (simulation) error estimate.

The cost function J_{multi}, defined in (12), is applied to quantify the model accuracy in terms of multi-step ahead predictions. It is emphasized that such accuracy is quantified using fresh data, that is to say, a distinct dataset from the one used for estimation purposes. In what follows, the performance of the MRI techniques are investigated using two sets of Monte Carlo simulations, each one with 100 distinct white-noise realizations. In order to visualize the SNR effect on different parameter estimation methods, in the first simulation set, the SNR is maintained in 3 and in the other one it is increased to 10. The histograms of J_{multi} for the methods described in Section 3 are depicted in the rows of Fig. 4, for $P = 8$. The left and the right columns present the results for the signal-to-noise ratios of 3 and 10, respectively. The main Monte Carlo simulation results are summarized in Table 1, which reports the mean and

the standard deviation of J_{multi}. For comparison, the last column of Table 1 also presents the results produced by the standard (one-step ahead) PEM, based on a Box-Jenkins structure.

The histograms in Fig. 4, as well as Table 1, show that the MPEM and the noise model optimization algorithms presented the smallest J_{multi} (that is, the most accurate multi-step

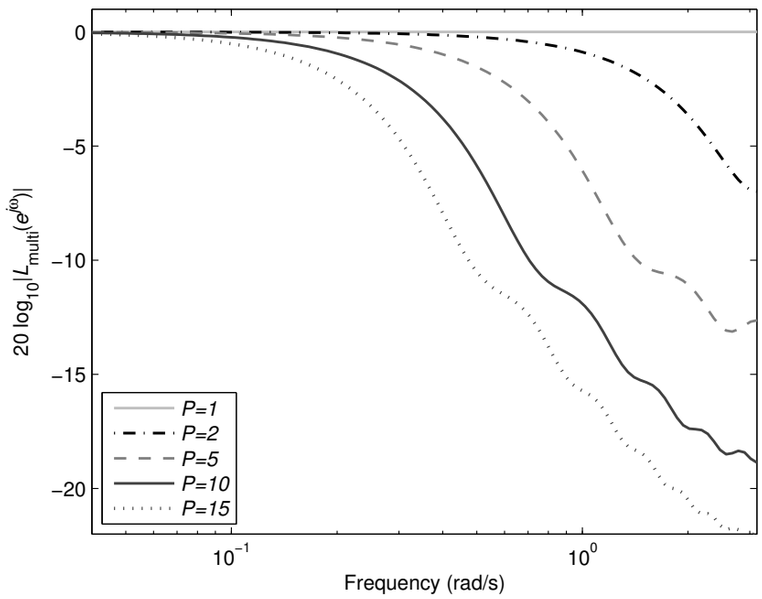

Fig. 2. Magnitude frequency response of $L_{\text{multi}}(e^{j\omega})$ for increasing prediction horizon length.

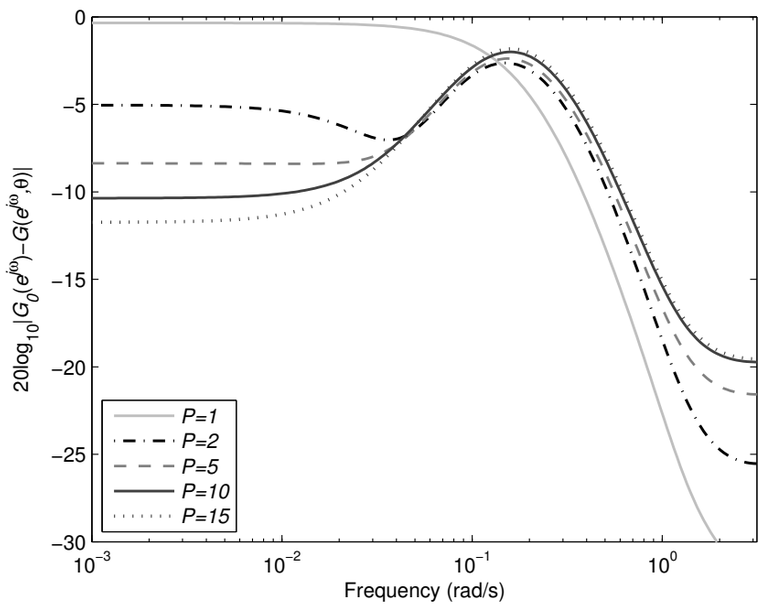

Fig. 3. Absolute value of the modeling error in the frequency domain, for models estimated with $P = \{1, 2, 5, 10, 15\}$.

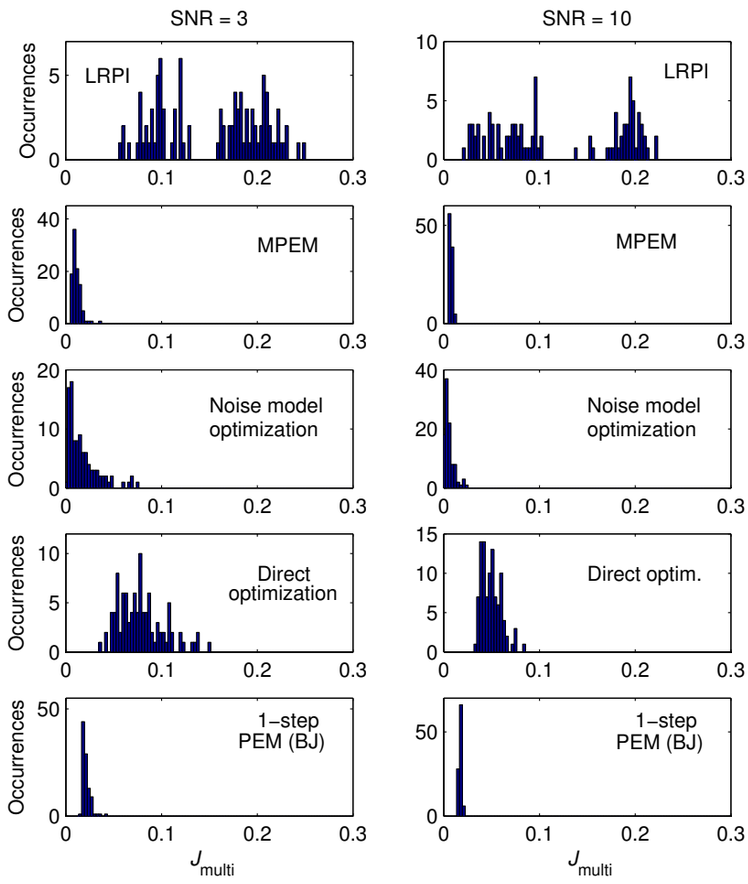

Fig. 4. Histograms of J_{multi} for each MRI method.

predictions) with the lowest variance, which means that these methods are less sensitive to a particular realization. On the other hand, LRPI and direct optimization showed worse performances because these methods are based on the ARX model structure, which is quite different from the process (46)-(47). Another aspect that may be noticed is that, as expected, a higher SNR leads to a smaller J_{multi} mean (more accurate models are expected) and lower deviations of the estimates.

Actually, the performances of the methods based on ARX structure may be interpreted in a broader sense. Although in MRI the effect of bias due to model mismatch is reduced in the

	SNR	LRPI	MPEM	Direct optim.	Noise model optimization	Standard PEM (Box-Jenkins)
mean(J_{multi})	3	0.1526	0.0111	0.0786	0.0178	0.0209
	10	0.1218	0.0074	0.0496	0.0056	0.0172
std(J_{multi})	3	0.0536	0.0045	0.0239	0.0163	0.0042
	10	0.0668	0.0015	0.0104	0.0049	0.0014

Table 1. Mean and standard deviation of the cost function.

parameter estimation step, the task of selecting a suitable model structure is still crucial to the success of a system identification procedure. This statement is also supported by the fact that, according to Table 1, considering a more favorable structure and the one-step ahead PEM is more effective than an inadequate structure whose parameters are estimated based on a multi-step cost function.

4.2 Multivariable system example

The Shell benchmark process is a model of a two-input two-output distillation column (Cott, 1995a;b). The inputs are overhead vapour flow and reboiler duty, denoted here as u_1 and u_2, respectively. The outputs are the column pressure

$$\begin{aligned}\Delta y_1(t) =& \frac{-0.6096 + 0.4022q^{-1}}{1 - 1.5298q^{-1} + 0.574q^{-2}}\Delta u_1(t) + \frac{0.1055 - 0.0918q^{-1}}{1 - 1.5298q^{-1} + 0.574q^{-2}}\Delta u_2(t) \\ &+ \frac{\lambda}{1 - 1.5945q^{-1} + 0.5945q^{-2}}e_1(t)\end{aligned} \tag{50}$$

and the product impurity

$$y_2(t) = 0.0765\frac{5 \times 10^5}{u_2(t-7) - 1500} + 0.9235y_2(t-1) + \frac{\lambda}{1 - 1.6595q^{-1} + 0.6595q^{-2}}e_2(t) \tag{51}$$

where Δy_1, Δu_1 and Δu_2 are deviation variables around the nominal operating point[5] (specified in Table 2), that is

$$\begin{aligned}\Delta y_1(t) &= y_1(t) - \bar{y}_1 \\ \Delta u_1(t) &= u_1(t) - \bar{u}_1 \\ \Delta u_2(t) &= u_2(t) - \bar{u}_2\ .\end{aligned}$$

Variable	Nominal setpoints	Normal operation
Pressure (y_1)	2800	$2700 < y_1 < 2900$
Composition (y_2)	500	$250 < y_2 < 1000$
Overhead vapour flow (u_1)	20	$10 < u_1 < 30$
Reboiler flow (u_2)	2500	$2000 < u_2 < 3000$

Table 2. Summary of distillation column operating conditions.

The disturbances are generated using uncorrelated zero-mean white noises e_1 and e_2, such that $\text{std}(e_1) = 1.231$ and $\text{std}(e_2) = 0.667$. The parameter λ is set to 0.2. The Shell benchmark is widely used to evaluate multivariable system identification or model predictive control strategies (Amjad & Al-Duwaish, 2003; Cott, 1995b; Zhu, 1998, e.g.). Besides being multivariable, the model (50)-(51) offers additional complications: as the disturbances are nonstationary, one of the outputs (product impurity) is slightly nonlinear and the overhead flow (u_1) does not affect the impurity level (y_2).

[5] For more details about the simulator operating conditions, the reader is referred to (Cott, 1995b).

The process is excited in open-loop using two uncorrelated random binary sequences (RBS), with u_1 varying from $[15, 25]$ and u_2 from $[2400, 2600]$. The minimum switching time of u_1 and u_2 is 12 and 6, respectively. The dataset is comprised of 1600 samples, where the first half is used for estimation (see Fig. 5) and the rest for validation.

The elements of the transfer function matrix $G(q, \theta)$ and of the noise models are first order. Initially, the input delay matrix

$$n_k = \begin{bmatrix} 0 & 0 \\ 37 & 7 \end{bmatrix} \tag{52}$$

was estimated applying the function `delayest` of the Matlab™System Identification toolbox (Ljung, 2007). Notice that except for the entry in which there is no coupling ($u_1 \rightarrow y_2$), the values in n_k coincide with the actual input delays. Thus, before proceeding to the parameter estimation, the input sequences are shifted according to n_k.

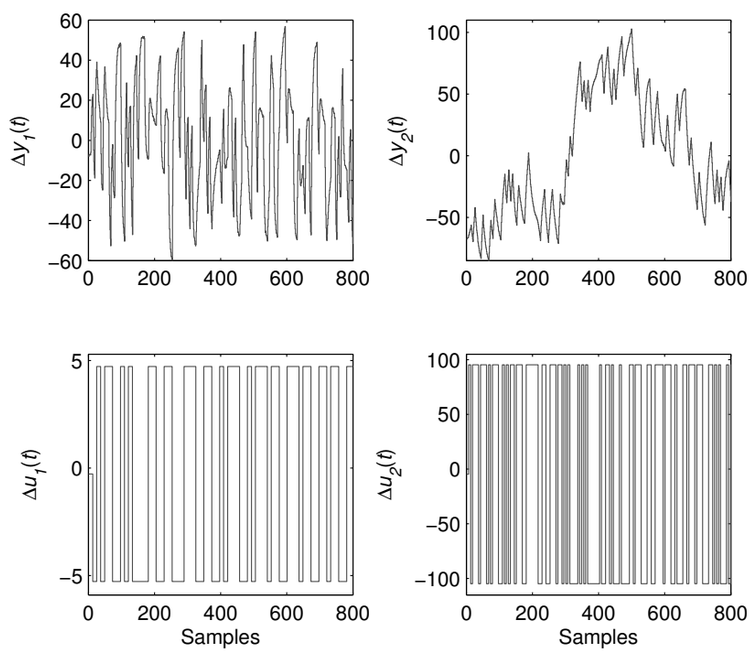

Fig. 5. Estimation dataset from the Shell benchmark simulation.

The models estimated with $P = 40$ are evaluated based on the multi-step prediction errors (12) using the validation dataset, which are presented in Table 3. The most accurate multi-step predictions are generated by the MPEM and the 1-step ahead PEM. This is because, as in the SISO example, the Box-Jenkins structure employed by both methods best suits the process dynamic behavior. Another relevant point is that the noise model optimization yields unstable

	Output	LRPI	MPEM	Direct optim.	Noise model optimization	1-step PEM (Box-Jenkins)
$J_{\text{multi}} \times 10^4$	1	0.1154	0.0328	0.1475	∞	0.0322
	2	3.2887	2.5072	3.6831	∞	2.5294

Table 3. Multi-step prediction error.

predictors (due to zeros outside the unitary circle). Consequently, the sum of the prediction errors tends to infinity.

The standard PEM provided multi-step predictions as accurate as the MPEM, even for a sub-parameterized model, which is the case of this example. This result suggests that the under-modeling issue is not the most prominent for this situation. In addition, the fact that the disturbance intensity is very high, besides being concentrated on the low frequencies (the same frequency range that should be weighted to attain improved multi-step ahead predictions) disfavors the MRI approach.

In order to test the robustness of the methods to input-output delay estimation errors, a new estimation is carried out with a modified delay matrix n_k^*, in which the dead-time from u_2 to y_2 is changed from 7 to 8 samples. As shown in Table 4, the MRI methods are less sensitive to this parameter than the 1-step ahead PEM.

	Output	LRPI	MPEM	Direct optim.	Noise model optimization	1-step PEM (Box-Jenkins)
$J_{\text{multi}} \times 10^4$	2	3.1669	2.4854	3.8126	∞	3.0794

Table 4. Multi-step prediction error of the 2nd output when there is a slight mismatch in one of the input delay matrix element.

At this point, the performance of the estimated models is investigated when they are employed in a QDMC controller (Camacho & Bordons, 2004) with a prediction and control horizons of 40 and 5, respectively. The output $\mathcal{Q}$ and the manipulated $\mathcal{R}$ weighting matrices are (Amjad & Al-Duwaish, 2003)

$$\mathcal{Q} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \text{ and } \mathcal{R} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}.$$

The closed-loop responses using the QDMC controller when each set-point is excited with a step of amplitude 1% of the nominal output values are presented in Fig. 6 and 7, where the first one is related to the input delay matrix n_k in (52) and the other refers to n_k^*. The results of the closed-loop validation are also summarized in Table 5, which shows the integrated square error (ISE) for each controlled variable: y_1 and y_2.

In a general way, the first output is closer to the set-point than y_2. This may be explained by the intensity of the disturbance introduced in each output, by the fact that the plant is non-linear whereas the identified models are linear and, finally, due to the presence of a zero in the transfer matrix which consequently affects the quality of the estimated model.

From Fig. 6, one can notice that all the controllers achieved similar responses for the column pressure (y_1). Concerning the other output (product purity), the closed-loop behavior provided by the standard PEM and the MPEM are very close (accordingly to multi-step prediction errors depicted in Table 3). Analogously, the LRPI method yielded a better performance than the direct optimization. Besides, as these two methods showed a worse multi-step prediction accuracy, it reflected in the MPC performance.

As shown in Fig. 7 and according to Table 5, the prediction capacity deterioration of the one-step ahead PEM, due to the delay matrix modification from n_k to n_k^* also leads to a worse closed-loop response. On the other hand, the closed-loop performances provided by

the models estimated through MRI algorithms are less sensitive to errors in the time delay determination.

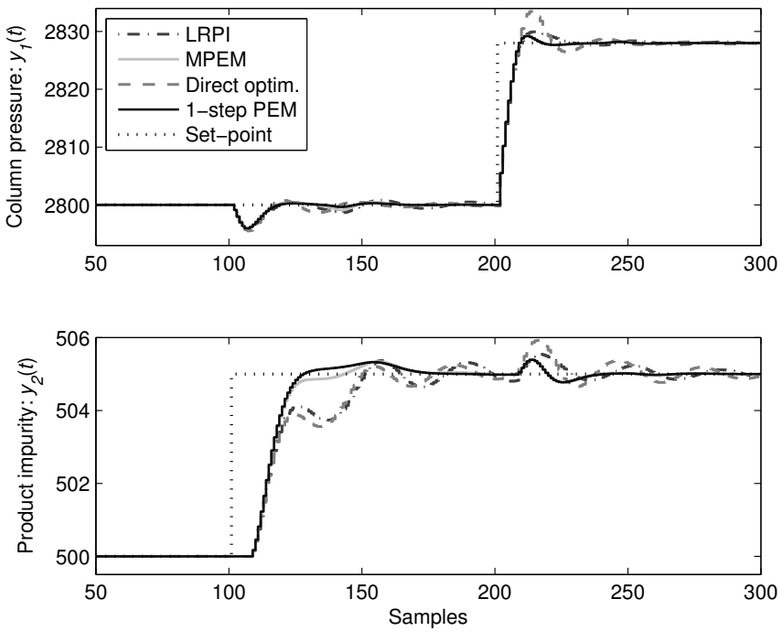

Fig. 6. Closed-loop response based on an accurate input delay estimation.

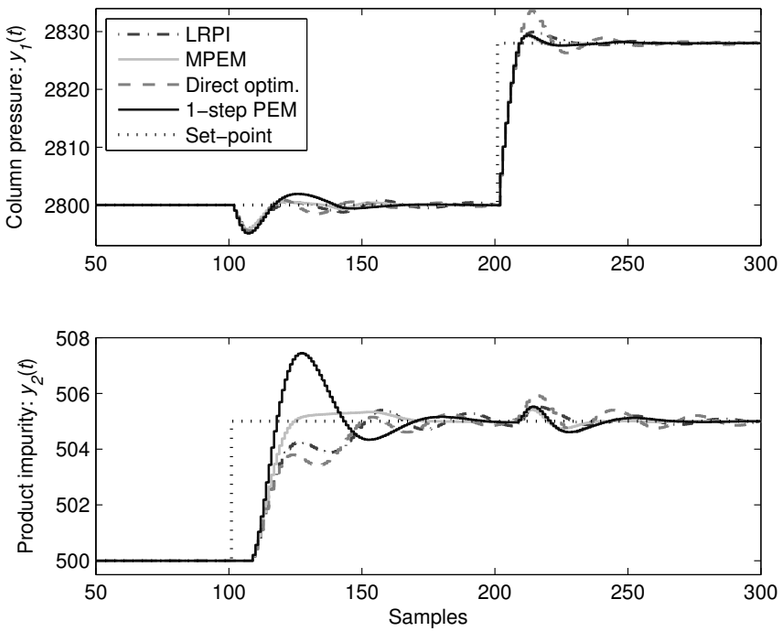

Fig. 7. Closed-loop response for a mismatch in one of the input-output delay matrix entry.

	Output	LRPI	MPEM	Direct optim.	1-step PEM (Box-Jenkins)
ISE $\times 10^3$: n_k	1	2.1921	2.0777	2.4091	2.0711
	2	0.3618	0.3182	0.3747	0.3171
ISE $\times 10^3$: n_k^*	1	2.1689	2.0883	2.4081	2.2016
	2	0.3519	0.3119	0.3802	0.3667

Table 5. Integrated square error (ISE) of the controlled variables.

5. Conclusions

This chapter focused on parameter estimation algorithms to generate suitable models for predictive controllers. The branch of identification known as MRI was studied and several different ways to obtain models were presented. They must be estimated having in mind that they must be accurate to predict multi-step ahead. Some of these techniques were published considering just the single-input single-output case and in this work they were extended to the multivariable framework. In order to compare the different algorithms, they were implemented and tested, employing a SISO and a MIMO plant. In the comparisons, the standard PEM (built to provide optimal one-step ahead predictions) was also included.

In the analysis with the SISO process, the long range prediction capacity of some of the MRI methods (MPEM and noise model optimization) was superior to the results generated by the standard PEM, based on a Box-Jenkins structure. In addition, the influence of the model structure was also highlighted in a model relevant identification context, since the standard PEM (with a Box-Jenkins) produced more accurate multi-step ahead predictions than the LRPI and the direct optimization algorithms, which are based on a less flexible model structure.

The tests performed with the multivariable plant were more concerned about the use of the MRI and PEM models, when applied to a predictive controller. The results obtained were not so convincing about the advantages of using multi-step prediction based methods in the predictive controller design, since the one-step PEM (with a Box-Jenkins model), even with structure mismatch, provided results that were comparable to the best ones obtained with the model relevant identification methods. However, it was also shown that when there was a slight error in the evaluation of the time delay of one of the input-output pairs, the advantage of the MRI approach became evident.

Although the excitation signal design and the model structure selection are beyond the scope of this work, the examples presented the complete system identification procedure, from the input signal generation, going through the use of different algorithms to estimate the model parameters up to the validation of the models through the verification of their prediction capacity. Besides, the obtained models were applied to a predictive controller to evaluate their performance in controlling a multivariable process.

The system identification for MPC is a subject prone to further research. The effect of multi-step prediction error methods on the closed-loop performance needs to be further investigated. Another important theme to be studied is in which situations the use of MRI methods for developing models for predictive controllers is in fact advantageous as compared to classical prediction error methods.

6. Acknowledgments

The authors gratefully acknowledge the support of Instituto Mauá de Tecnologia (IMT). They also thank the support provided by Petrobras.

7. References

Amjad, S. & Al-Duwaish, H. N. (2003). Model predictive controle of Shell benchmark process, *Proceedings of the 10th IEEE International Conference on Electronics, Circuits and Systems*, Vol. 2, pp. 655–658.

Camacho, E. & Bordons, C. (2004). *Model predictive control*, 2nd edn, Springer-Verlag, London.

Clarke, D. W., Mohtadi, C. & Tuffs, P. S. (1987). Generalized predictive control - part 1 and 2, *Automatica* 23(2): 137–160.

Cott, B. J. (1995a). Introduction to the process identification workshop at the 1992 canadian chemical engineering conference, *Journal of Process Control* 5(2): 67–69.

Cott, B. J. (1995b). Summary of the process identification workshop at the 1992 canadian chemical engineering conference, *Journal of Process Control* 5(2): 109–113.

Farina, M. & Piroddi, L. (2011). Simulation error minimization identification based on multi-stage prediction, *International Journal of Adaptive Control and Signal Processing* 25: 389–406.

Gopaluni, R. B., Patwardhan, R. S. & Shah, S. L. (2003). The nature of data pre-filters in MPC relevant identification – open- and closed-loop issues, *Automatica* 39: 1617–1626.

Gopaluni, R. B., Patwardhan, R. S. & Shah, S. L. (2004). MPC relevant identification – tuning the noise model, *Journal of Process Control* 14: 699–714.

Huang, B. & Wang, Z. (1999). The role of data prefiltering for integrated identification and model predictive control, *Proceedings of the 14th IFAC World Congress*, Beijing, China, pp. 151–156.

Ježek, J. & Kučera, V. (1985). Efficient algorithm for matrix spectral factorization, *Automatica* 21(6): 663–669.

Laurí, D., Salcedo, J. V., García-Nieto, S. & Martínez, M. (2010). Model predictive control relevant identification: multiple input multiple output against multiple input single output, *IET Control Theory and Applications* 4(9): 1756–1766.

Ljung, L. (1999). *System Identification: theory for the user*, 2nd edn, Prentice Hall, Upper Saddle River, NJ.

Ljung, L. (2007). *The system identification toolbox: The manual*, 7 edn, The MathWorks, Inc., MA, USA: Natick.

Maciejowski, M. (2002). *Predictive Control with Constraints*, Prentice Hall, Englewood Cliffs, NJ.

Rossiter, J. A. & Kouvaritakis, B. (2001). Modelling and implicit modelling for predictive control, *International Journal of Control* 74(11): 1085–1095.

Shook, D. S., Mohtadi, C. & Shah, S. L. (1991). Identification for long-range predictive control, *IEE Proceedings-D* 138(1): 75–84.

Zhu, Y. (1998). Multivariable process identification for MPC: the asymptotic method and its applications, *Journal of Process Control* 8(2): 101–115.

Zhu, Y. (2001). *Multivariable System Identification for Process Control*, Elsevier Science, Oxford.

Zhu, Y. & Butoyi, F. (2002). Case studies on closed-loop identification for MPC, *Control Engineering Practice* 10: 403–417.

IntechOpen