15th Triennial World Congress, Barcelona, Spain

BIAS DISTRIBUTION IN MPC RELEVANT IDENTIFICATION

Ratna Bhushan Gopaluni [*] **Rohit S. Patwardhan** [**]
Sirish L. Shah [*]

[*] *Department of Chemical & Materials Engineering, University of Alberta, Edmonton, AB, CANADA - T6G 2G6*
[**] *Applications Group, Matrikon Consulting Inc., 1800 10405 Jasper Av, Edmonton, AB, CANADA T5J 3N4*

Abstract: In this paper a **M**odel Predictive Control **R**elevant **I**dentification (MRI) method is applied to a general class of PEM models and the effect of bias distribution on the multi step ahead predictions is studied. Good multi step ahead predictions are essential for model predictive controllers. Therefore it is important to distribute the bias in such a way that it is compatible with the predictive control objective. This paper deals with the analysis of use of MRI methods on the bias distribution and its effect on the control loop performance.

Keywords: identification, bias distribution, model predictive controllers

1. INTRODUCTION

It is well known that Model Predictive Controllers (MPCs) require models which can provide good multi step ahead predictions, as opposed to, one step ahead predictions (Shook et al. (1992); Huang and Wang (1999); Rossiter and Kouvaritakis (2001)). If the structures of the true process model and the noise model are known *a priori* then the maximum likelyhood estimate which gives optimal multistep ahead predictions is the true process and noise models. However, in practice it is not possible to know the model structures accurately and moreover most of the real processes are nonlinear. Therefore models tuned for multi step ahead predictions are vital for good closed loop performance when using predictive controllers. It is important to emphasize that MPC relevant identification (MRI) methods are useful only when certain amount of bias either in the process model or the noise model is expected. Hence the MRI problem reduces to that of distributing the bias in process and noise models accordingly.

Huang and Wang (1999) discuss this method in the context of data prefiltering. They show that the MRI algorithm can be reduced to that of one step ahead prediction error method by filtering the inputs and outputs with a filter that depends on the noise model. Expressions for closed loop multi step ahead predictors are also derived in this paper. They also provide simulation examples to show that a predictive controller under no constraints would perform better with a model developed for multistep ahead predictions. Though there is no clear proof, as yet, to show that MRI models indeed perform better under closed loop, certain heuristic and quantitative arguments are presented in Shook et al. (1992) and Gopaluni et al. (2001).

Rossiter and Kouvaritakis (2001), on the other hand, use multiple models for multi step ahead predictions. They develop an optimal k-step ahead prediction model for each $k \in \{1, 2, \cdots, P\}$ where P is the controller prediction horizon. Then all the P models are simultaneously used for predictions. Even though this method provides “optimal" multi step ahead predictions for ARX type models, the number of parameters involved in estimating the models can be quite large. In fact, it is well known that the asymptotic parameter

variance is proportional to the ratio between the order of the process model and the data length (Ljung (1999)). Since, ARX type models are used in their paper the parameter variance is proportional to the number of parameters and therefore a large data set is required to get reasonably good parameter estimates with acceptable uncertainty. Moreover, for multi-input and multi-output processes the number of parameters can be large.

This paper discusses the MRI method from the point of view of distributing the bias in the process model. It is common knowledge that the noise model and the input spectrum determine the bias distribution. Hence, a "good" noise model is essential for good predictions. In fact, if there is no noise model (*i.e.,* if the noise is white) then there is no need for MRI. It is known that the MRI methods differ from the traditional one step ahead prediction error methods in the choice of a filter that depends on the noise model. An analysis of this filter is presented in this paper.

This paper is organized as follows: Section 2 presents the general setting and in Section 3 the factors affecting the bias distribution under open loop conditions are discussed. In Section 4, certain issues related to bias distribution under closed loop conditions are elucidated. Simulation examples illustrating the results discussed in the previous sections are presented where appropriate, followed by concluding remarks in Section 5.

2. PRELIMINARIES

In order to assess the accuracy of an identified model and to evaluate its statistical properties it is necessary to assume certain properties of the true model. In this paper we are going to assume that the true process is

$$\mathbb{S} : y(t) = G(q)u(t) + H(q)e(t) \tag{1}$$

where $e(t)$ is white noise with variance σ_e^2. $u(t)$ represents the input and $y(t)$ is the output. It is assumed that the input is persistently exciting. For the sake of simplicity the process is assumed to be SISO. q represents the forward shift operator. In the rest of this paper we refer to $G_0(q)$ as the true process model and to $H_0(q)$ as the true noise model. The estimated model consists of an estimated process model represented by $\hat{G}(q,\theta)$ and an estimated noise model, $\hat{H}(q,\theta)$ where θ is the parameter vector. It is also assumed that $\hat{H}$ and its inverse are both stable.

Then the k-step ahead optimal predictor is given by (Ljung (1999))

$$\hat{y}(t+k|t) = \hat{W}_k \hat{G} u(t+k) + (1-\hat{W}_k) y(t+k)$$

where

$$\hat{W}_k = \hat{F}_k \hat{H}^{-1}$$
$$\hat{F}_k = \sum_{i=0}^{k-1} \hat{h}(i) q^{-i} \; ; \quad \hat{F}_1 = \hat{h}(0) = 1$$

and $\hat{h}(i)$ are the impulse response coefficients of $\hat{H}$. It is assumed that all the inputs and the outputs are quasi-stationary signals.The following standard notation is adopted through out this article. Given a stochastic function $f(t)$, we define

$$\bar{E} f(t) \triangleq \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} E f(t) \tag{2}$$

where E is the expectation operator (expectation is taken only with respect to the noise). It is assumed that the above limit exists wherever the operator $\bar{E}$ is used. This limit must exist for the signals considered in this paper due to the assumption that they are all quasi-stationary Ljung (1999).

We also assume that $\hat{G}$ and $\hat{H}$ are independent of the data length by defining the identified models as

$$\hat{G} = \lim_{N\to\infty} \hat{G}_N \tag{3}$$
$$\hat{H} = \lim_{N\to\infty} \hat{H}_N \tag{4}$$

where $\hat{G}_N$ and $\hat{H}_N$ are models identified from a data of length N. In the rest of this paper all transfer functions with a subscript 'N' denote the identified models from a data set of length 'N'. Transfer function estimates without the superscript are the model estimates as $N \to \infty$.

The following additional notation is introduced for convenience.

$$\tilde{G}(q) := G - \hat{G} \quad \text{and} \quad \tilde{H} := H - \hat{H} \tag{5}$$

The data collected from identification experiments on the real process are denoted by

$$Z^N = \{u(1), y(1), \cdots, u(N), y(N)\} \tag{6}$$

3. BIAS IN OPEN LOOP

Under open loop conditions, all the traditional identification methods try to minimize either one step ahead prediction errors or k-step ahead prediction errors *i.e.,* an objective function of the form

$$V_N^k(\theta, Z^N) = \frac{1}{N} \sum_{t=1}^{N} \varepsilon_k^2(t,\theta) \tag{7}$$

where the k-step ahead prediction errors are given by

$$\varepsilon_k(t,\theta) = y(t+k) - \hat{y}(t+k|t,\theta). \tag{8}$$

It is straightforward to show that

$$\bar{E}\left[V_N^k(\theta,Z^N)\right] = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\hat{W}_k\tilde{G}\right|^2\Phi_u(\omega) + \left|\hat{W}_kH\right|^2\Phi_e(\omega)d\omega \quad (9)$$

where $\Phi_u(\omega)$ and $\Phi_e(\omega)$ are input and noise spectra respectively. In this case the filter on the bias is $|\hat{W}_k|^2\Phi_u(\omega) := \hat{L}_k(e^{i\omega})\Phi_u(\omega)$. However, predictive controllers minimize an objective function of the form [1] (see Fig.6 for notation)

$$J_{mpc} = \sum_{j=1}^{P}\left[r(t)-\hat{y}(t+j|t)\right]^2. \quad (10)$$

Hence, intuitively speaking, an identification algorithm that can minimize the sum of multi step ahead prediction errors should give better models for predictive controllers (see Shook et al. (1992) and Gopaluni et al. (2001) for rigorous arguments). The MRI objective function is then defined as

$$V_N(\theta,Z^N,P) = \frac{1}{(N-P)P}\sum_{t=1}^{N-P}\sum_{k=1}^{P}\varepsilon_k^2(t,\theta) \quad . (11)$$

Note that the objective function depends explicitly on the controller prediction horizon, P. It is easy to show that (using Parseval's theorem)

$$\bar{E}\left[V_N(\theta,Z^N,P)\right] \triangleq \bar{V}(\theta,P) = \sum_{k=1}^{P}\frac{1}{2\pi}\int_{-\pi}^{\pi}\left[|\hat{W}_k\tilde{G}|^2\Phi_u(\omega) + |\hat{W}_kH|^2\Phi_e(\omega)\right]d\omega \quad (12)$$

Hence under open loop conditions the optimal weighting on the model bias, $\tilde{G}$, must be

$$\mathbf{W_b} \triangleq \hat{L}(e^{i\omega})\Phi_u(\omega) \triangleq \sum_{k=1}^{P}|\hat{W}_k|^2\Phi_u(\omega). \quad (13)$$

Clearly, the bias distribution depends to the noise filter and the input spectrum. Notice that it is only the product between the filter, $L(e^{i\omega})$, and the input spectrum, Φ_u, that determines the bias distribution. Their individual values are immaterial. Only the properties of the filter $\hat{L}(e^{i\omega})$ are studied.

As mentioned earlier, a k-step ahead predictor has a similar filter (see (9)). In this case as the prediction horizon increases (*i.e.,* as k increases) the filter tends to unity. Mathematically,

$$\lim_{k\to\infty}\hat{W}_k = \lim_{k\to\infty}\frac{\hat{F}_k}{\hat{H}} \to 1. \quad (14)$$

or equivalently, for any given $\delta > 0$, there exists an $N(\delta) \in \mathbb{N}$ such that for all $k \geq N(\delta)$

$$|\hat{W}_k(e^{i\omega}) - 1| < \delta \quad \forall\ \omega. \quad (15)$$

Consequently, for any $k < P$ and a corresponding δ and for all ω,

$$(1-\delta)^2 \leq \frac{\hat{L}(e^{i\omega}) - \sum_{j=1}^{k}\hat{L}_j(e^{i\omega})}{P-k} \leq (1+\delta)^2 \quad (16)$$

Notice that the upper and lower bounds of the above quantity are independent of the prediction horizon, P and the frequency, ω. Hence, the difference between the lower and upper bounds of $\hat{L}(e^{i\omega})$ is $(P-k)(4\delta)$. Now for any $\varepsilon > 0$, it is possible to choose a P and a δ satisfying

$$k < P \leq \left\lfloor k + \frac{\varepsilon}{4\delta}\right\rfloor \quad (17)$$

$$\delta \leq \frac{\varepsilon}{4} \quad (18)$$

such that $(P-k)(4\delta) < \varepsilon$. Where $\lfloor . \rfloor$ denotes the largest integer less than its argument. Therefore, for any arbitrarily small ε there exists a prediction horizon such that the magnitude of the filter $\hat{L}(e^{i\omega})$ is uniform up to a constant ε.

For small values of k, the filters, $\hat{L}_k$, generally tend to give more weighting on high frequencies and as k increases, weighting in the low frequency range improves (Ljung (1999)).

On the other hand, MPC relevant filter, $\hat{L}$, provides a weighting that is "optimal" in some sense for multi step ahead predictions by suitably increasing the weighting in the low frequency range depending on the prediction horizon (see Fig.1). As the prediction horizon tends to infinity, the weighting tends to be approximately uniform at all frequencies. Typically, the effect of using the MRI filter in identification is to increase the weighting on the bias term in the low frequency range compared to one step ahead prediction models [2] . Naturally, this would result in a better model in the low frequency range. In fact, a good model in the low and mid frequency ranges is essential for good control performance.

Generally, the drawback of large weighting in the high frequency region and small weighting in the low frequency region when using one step ahead prediction error methods is overcome by sufficient input excitation in the low frequency region. MRI methods, in theory should perform better even if the input excitation in the low frequency region is not "sufficient".

3.1 Fixed/known noise model

In case the noise model is known *a priori* (which is rarely the case) then the MRI problem reduces to

[1] For the sake of simplicity we have assumed that there are no time delays.

[2] Note that this is not true for all types of noise models.

that of minimizing a modified form of the objective function

$$\bar{V}_m(\theta,P) := \sum_{k=1}^{P} \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[|\hat{W}_k \tilde{G}|^2 \Phi_u(\omega)\right] d\omega. \quad (19)$$

Traditional identification methods minimize only the first term in the above equation but on the other hand (11) shows that it is important to minimize other terms as well to reduce variance in the prediction errors. If G lies in the set, $\mathscr{G}$, of models over which the optimization is done, then the estimated model is going to be same as the true model and therefore there are no incentives in using MRI models. For example, if FIR models are used, $G \in \mathscr{G}$ and therefore the noise model plays a vital role in determining the performance of the predictor and therefore the controller (Gopaluni et al. (2001)). Nevertheless, it is impractical to assume that $G \in \mathscr{G}$ since reduced complexity models are of interest and that is when MPC relevant identification method helps us in distributing the bias. An example showing the effect of the MRI filter when the noise model is known is presented below.

***Example** 1.* The following process and noise models are used in the simulations (from Huang and Wang (1999)).

$$G = \frac{0.0077q^{-1} + 0.0212q^{-2} + 0.0036q^{-3}}{1 - 1.9031q^{-1} + 1.1514q^{-2} - 0.2158q^{-3}}$$
$$H = \frac{1}{(1 - 0.8q^{-1})} \quad (20)$$

The noise variance is 0.1 and the input signal is a white RBS signal with variance 1. Also, note that the settling time of this process model is about 40 seconds. An estimated model with the following structure is assumed

$$\hat{G} = \frac{b_1 q^{-1} + b_2 q^{-2}}{1 - a_1 q^{-1}}. \quad (21)$$

As explained in the previous section, it is easy to

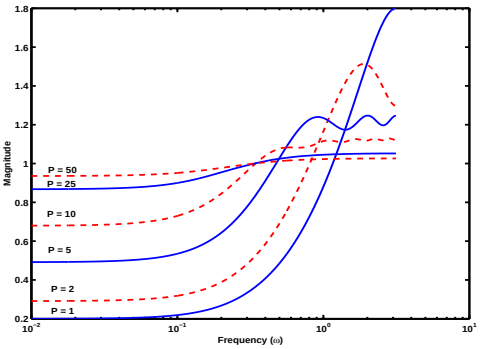

Fig. 1. Magnitude plot of the bias filter for multi step ahead predictions - $\frac{L(e^{i\omega})}{P}$

see from the Fig.1 that the filter $L(e^{i\omega})$ improves the weighting in the low frequency region with increasing prediction horizon. Note that $P = 1$ corresponds to the traditional one step ahead prediction error based identification. The Bode plots of the estimated models for various prediction horizons are shown in Fig.2. It is clear that there is a marked improvement of the fit in the low frequency region. Huang and Wang (1999) used an approximation for the MRI filter based on spectral factorization. No approximations for the filter are used in this paper. The original objective function is minimized directly, instead of filtering the data. The question that comes to one's mind is: Why use MRI method? One can instead use OE structure ($\hat{H} = 1$) in identification which naturally is going to result in uniform weighting on the bias term and therefore provides a good fit in the low frequency region. The MRI method "optimally" adjusts the weighting for multi step ahead predictions. In other words, it provides a good balance between the low and high frequency ranges for better multi step ahead predictions.

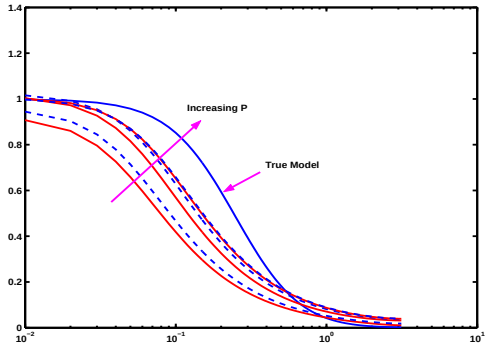

Fig. 2. Example 1 - Magnitude plot of the estimated first order model for $P = \{1,2,5,10,25,50\}$. Magnitude plot of the true process is also shown

3.2 Unknown noise model

If the noise model is not known, and if $G \in \mathscr{G}$, the set of models over which the optimization is performed, then the optimization problem can be solved in two steps : 1) Choose an arbitrary noise *e.g..*, $\hat{H} = 1$ and identify an unbiassed estimate of G. 2) Then identify a noise model by minimizing

$$\sum_{k=1}^{P} \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[|\hat{W}_k H|^2 \Phi_e(\omega)\right] d\omega. \quad (22)$$

In any practical situation, it is impossible to know the true structures of the process and noise models. In such a case, the MRI estimator has to be minimized with respect to both the noise and process models. Clearly, the structure of the process and noise models chosen are going to influence the bias distribution of the estimated process model.

***Example** 2.* The aim of this example is to show that a bias distribution that is suitable for multi step ahead

predictions is going to be different from the distribution suitable for one step ahead predictions.

Consider the process model used in the previous example and a new noise model

$$H = \frac{1}{(1-0.8q^{-1})^2} \tag{23}$$

But a noise model with the following structure is assumed to show that the structure of the noise model determines the bias distribution in the process model. All the other parameters are same as those in the previous example.

$$\hat{H} = \frac{1}{1-a_2 q^{-1}} \tag{24}$$

A plot showing the bias distribution is shown in Fig.3 and in Fig.4 the optimal MRI noise pole is plotted against the prediction horizon. Notice that as in the previous example there is a significant improvement in the process model accuracy around low and mid frequency regions at the expense of poor modelling in the high frequency region. The cost function in (11) is evaluated for various values of the prediction horizon using the models obtained from the one step ahead prediction error method and the MRI method. Their ratio is plotted in Fig.5. There is a significant improvement in the accuracy of the multi step ahead predictions obtained using the MRI method when compared with the predictions obtained using the one step ahead prediction error model.

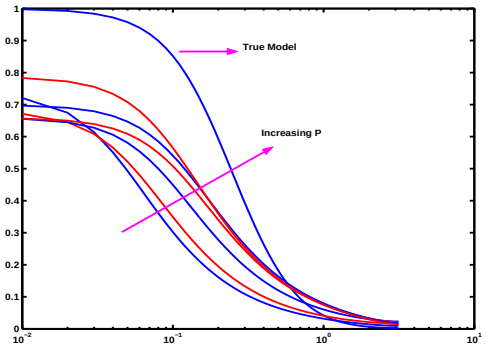

Fig. 3. Example 2 - The bode magnitude plot of the estimated and true process models for $P \in \{1,2,5,10,25,50\}$

4. BIAS IN CLOSED LOOP

There is extensive literature on closed loop identification (Forssell and Ljung (1999); MacGregor and Fogal (1995); Huang and Shah (1997)) based on minimizing one step ahead prediction errors. In this paper closed loop identification based on minimizing multi step ahead prediction errors is of concern. The expressions for the objective function have to be modified under closed loop to account for the correlation between the

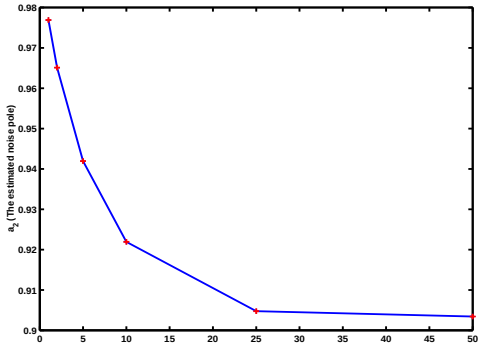

Fig. 4. The estimated noise pole as a function of prediction horizon

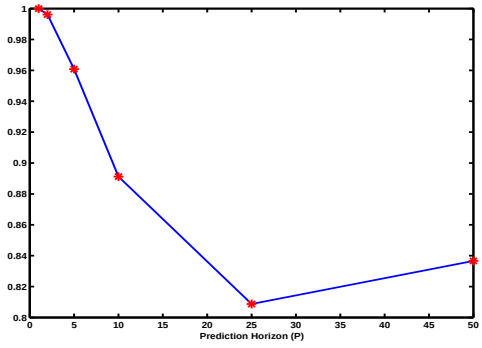

Fig. 5. The ratio between values of the cost function (in (11)) evaluated using the MRI model and the one step ahead prediction error model

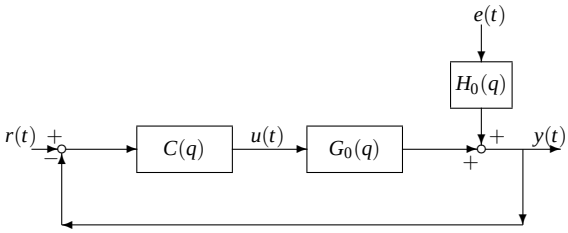

Fig. 6. Block diagram of a Closed loop system

input and the noise. The MRI objective function under closed loop is

$$\bar{V}_C(\theta,P) = \frac{1}{2\pi}\sum_{k=1}^{P}\int_{-\pi}^{\pi} \left|\hat{W}_k \tilde{G}_p\right|^2 |CS|^2 \Phi_r(\omega) + \left|\hat{W}_k G_L\right|^2 \frac{|S|^2}{|\hat{S}|^2}\Phi_e(\omega)d\omega \tag{25}$$

where S and $\hat{S}$ are the true and designed sensitivity functions defined by

$$S \triangleq \frac{1}{1+CG} \qquad \hat{S} \triangleq \frac{1}{1+C\hat{G}}. \tag{26}$$

and Φ_r is the set point spectrum.

As in the open loop case if the noise model is fixed, in this case to $\hat{S}$ (*i.e.*, $\hat{W}_k = \hat{S}$),[3] and if $G \in \mathscr{G}$, then the optimal solution of this objective function is the true model (this is equivalent to the two step closed loop identification (Forssell and Ljung (1999))). Therefore, one step ahead prediction error methods and the MRI method result in the same optimal solution. On the other hand, if $G \notin \mathscr{G}$ then the model bias has to be distributed in such a way that the above objective function is minimum.

The weighting on the bias term under closed loop is :

$$\mathbf{W}_{\mathbf{b}}^{\mathbf{C}} \triangleq \hat{L}_{\mathrm{C}}(e^{i\omega})\Phi_r(\omega) \triangleq \sum_{j=1}^{P} |\hat{W}_j|^2 |CS|^2 \Phi_r(\omega). \quad (27)$$

Since experiment design is not dealt with in this paper, it is assumed that the set point spectrum is uniform at all frequencies. The nature of the filter $\hat{L}_{\mathrm{C}}$ is then going to determine the bias in the process model.

Most of the commonly used controllers have an integrator and as a result the weighting on the bias term in the low frequency region is, typically, close to the inverse of the plant gain (for direct closed loop identification). Therefore, the effect of MRI method on the bias distribution of the process model in the low frequency region may not be significantly different from that in the open loop. On the other hand, if the excitation signal is introduced at the input (also called dither signal) then the bias filter is going to be :

$$\mathbf{W}_{\mathbf{b}}^{\mathbf{C}} \triangleq \hat{L}_{\mathrm{C}}(e^{i\omega})\Phi_r(\omega) \triangleq \sum_{j=1}^{P} |\hat{W}_j|^2 |S|^2 \Phi_d(\omega). \quad (28)$$

where $\Phi_d(\omega)$ is the spectrum of the dither signal. Clearly, in this case the MRI filter results in a significant improvement in the low and mid frequency weighting.

***Example** 3.* An IMC controller is designed based on the first order approximation of the original process model in (20). A plot of the filter in (27) is shown in Fig.7 for various prediction horizons. This filter distributes the bias under direct closed loop identification. For other types of closed loop identification *viz.*, indirect and join input-output methods, the bias distribution may depend on different types of filters.

5. CONCLUDING REMARKS

The nature of the filter obtained by minimizing multi step ahead predictions, as opposed to, one step ahead predictions is discussed in detail. The difference between the two filters lies in improving the weighting

[3] Strictly speaking $\hat{W}_k \neq \hat{S}$ for all j but this substitution helps in decoupling the noise model from the process model in the identification step.

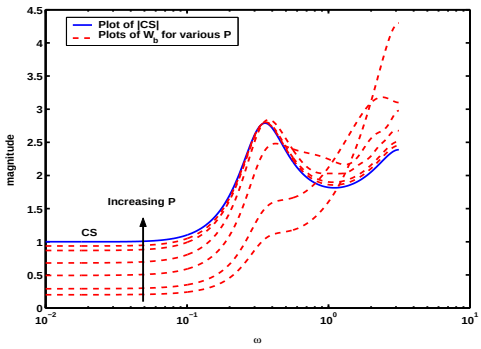

Fig. 7. The filter, $\mathbf{W}_{\mathbf{b}}^{\mathbf{C}}$, in (27) is shown for $P \in \{1,2,5,10,25,50\}$ with $\Phi_r(\omega) = 1 \quad \forall \ \omega$

on the bias term in the low and mid frequency ranges. MRI method improves the bias weighting in the low frequency region in an "optimal" way taking into account the prediction horizon of the controller. MRI methods showed a significant improvement in the accuracy of multi step ahead predictions compared to the traditional methods. MRI methods, depending on the noise model chosen, can do away with the need for exclusively low frequency excitation. Certain analysis of closed loop systems is also presented in this article. Further investigation into the effect of MRI filter on the bias distribution under closed loop is needed. The effect of MRI models on the closed loop performance also needs to be studied further. There is also a need to study the effect of input spectrum on the bias distribution.

REFERENCES

U. Forssell and L. Ljung. Closed-loop identification revisited. *Automatica*, 35:1215–1241, 1999.

R.B. Gopaluni, R.S. Patwardhan, and S.L. Shah. MPC relevant identification - tuning the noise model. *submitted to Journal of Process Control*, 2001.

B Huang and S.L Shah. Closed loop identification: A two step approach. *Journal of Process Control*, 7 (6):425–438, 1997.

B Huang and Z Wang. The role of data prefiltering for integrated identification and model predictive control. *Proceedings of IFAC World Congress*, 1999.

L. Ljung. *System Identification: Theory for the user*. Prentice Hall, 1999.

J.F MacGregor and D.T Fogal. Closed-loop identification: the role of the noise model and prefilters. *Journal of Process Control*, 5(3):163–171, 1995.

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

D.S. Shook, S. Mohtadi, and S.L. Shah. A control - relevant identification strategy for gpc. *IEEE Transactions on Automatic Control*, 37(7):975–980, July 1992.