A BAYESIAN APPROACH TO CLOSED-LOOP SYSTEM IDENTIFICATION

Stéphane Thil and **Marion Gilson**

Centre de Recherche en Automatique de Nancy (CRAN),
CNRS UMR 7039, Université Henri Poincaré, Nancy 1,
F-54506 Vandœuvre-lès-Nancy Cedex, France,
{stephane.thil,marion.gilson}@cran.uhp-nancy.fr

Abstract: The challenging issue of identifying a closed-loop system from short and/or non-informative data records is addressed. A bayesian approach is developed within this framework. It is shown that accurate estimates and realistic confidence intervals can be obtained by taking into account prior knowledge on the system. The performances of the proposed method are illustrated with a simulation example.

Keywords: Closed-loop identification, Bayesian technique, MCMC methods, non-informative data records.

1. INTRODUCTION

Dynamic system identification consists in building a mathematical model from a finite amount of data. This model must be able to reproduce correctly (depending on its use) the behavior of the real system. This paper focuses on the identification of linear parametric models. More precisely, the goal is to identify a closed-loop system when the inputs are non persistently exciting of sufficient order, and to obtain realistic confidence intervals on the parameters.

Most traditional identification methods are based on prediction error and rely on asymptotic-in-data-length results to obtain confidence intervals, results which are assumed to hold for a finite amount of data available. This assumption can however turn out to be dangerous, especially if that amount is low.

It is therefore justified to look for methods that use non-asymptotic error quantifications. In this spirit, a Bayesian approach has been recently proposed to solve this problem, within the framework of open-loop system identification (Ninness and Henriksen, 2003).

The Bayesian approach relies on the posterior distribution, up to a constant equal to the product of the likelihood function and the prior distribution of the parameters to be estimated. Prior knowledge is used in order to "correct" the likelihood function when the data records are arbitrarily short or not enough informative.

Moreover, for many industrial production processes, safety and production restrictions are often strong reasons for not allowing identification experiments in open-loop. In such situations, experimental data can only be obtained under so-called closed-loop conditions. The main difficulty in closed-loop identification is due to the correlation between the disturbances and the control signal, induced by the loop. Furthermore, the data collected are less informative than those stemmed from an open-loop experiment since the goal of the controller is to minimize the sensitivity of the system to disturbances, thus making the identification more complicated (Forssell and Ljung, 1999).

Several alternatives are available to cope with this problem, broadly classified into three main categories of approaches: direct, indirect and joint input-output (Van den Hof, 1998).

A Bayesian technique is introduced to handle this closed-loop system identification problem, in the case of arbitrarily short and non-informative data records. To our best knowledge, this issue has not received appropriate attention so far.
This paper is organized as follows: after the preliminaries, the Bayesian inference is reviewed in section 3. Section 4 introduces the use of this Bayesian approach to handle the closed-loop system identification problem. The performances of the proposed approach are illustrated with a simulation example and compared to two traditional closed-loop identification methods in section 5.

2. PRELIMINARIES

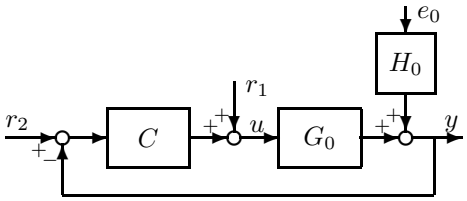

Figure 1. Closed-loop configuration.

Consider a linear SISO closed-loop system shown in Figure 1. The process is denoted by $G_0(z)$ and the controller by $C(z)$; $u(t)$ describes the process input signal, $y(t)$ the process output signal and $\{e_0(t)\}$ is a sequence of independent identically distributed (i.i.d.) random variables of variance λ_0. The external signals $r_1(t), r_2(t)$ are assumed to be uncorrelated with $e_0(t)$. For ease of notation we also introduce the signal $r(t) = r_1(t) + C(q)r_2(t)$. With this notation, the data generating system becomes

$$\mathcal{S}: \begin{cases} y(t) = G_0(q)u(t) + H_0(q)e_0(t), \\ u(t) = r(t) - C(q)y(t). \end{cases} \tag{1}$$

The real plant G_0 is considered to satisfy $G_0(q) = B_0(q^{-1})/A_0(q^{-1})$, while in these expressions q^{-1} is the delay operator, and the numerator and denominator degree is n_0. The m-th order controller is assumed to be known and specified by

$$C(q) = \frac{Q(q^{-1})}{P(q^{-1})} = \frac{q_0 + q_1 q^{-1} + \cdots + q_m q^{-m}}{1 + p_1 q^{-1} + \cdots + p_m q^{-m}}, \tag{2}$$

with the pair of polynomials (P, Q) assumed to be coprime.
The following closed-loop model is considered

$$\begin{cases} \mathcal{M}(\theta): \ y(t) = G(q,\theta)u(t) + H(q,\theta)e(t), \\ \text{with} \ \ u(t) = r(t) - C(q)y(t), \end{cases} \tag{3}$$

where a parameterized process model is used

$$\begin{aligned} \mathcal{G}: \quad G(q,\theta) &= \frac{B(q^{-1},\theta)}{A(q^{-1},\theta)} \\ &= \frac{b_1 q^{-1} + \cdots + b_n q^{-n}}{1 + a_1 q^{-1} + \cdots + a_n q^{-n}}, \end{aligned} \tag{4}$$

and the process model parameters are stacked columnwise in the parameter vector

$$\theta = \begin{bmatrix} a_1 & \cdots & a_n & b_1 & \cdots & b_n \end{bmatrix}^T \in \mathbb{R}^{2n}. \tag{5}$$

The identification consists then in finding the best (in the sense of the criterion used) model within the set $\mathcal{M}$. From equations (3), the closed-loop output can also be written in terms of $r(t)$ which is uncorrelated with the noise

$$y(t) = \frac{1}{1 + C(q)G(q,\theta)} \left[G(q,\theta)r(t) + H(q,\theta)e(t) \right]. \tag{6}$$

Define the following closed-loop transfers

$$G_{cl}(q,\theta) \triangleq \frac{G(q,\theta)}{1 + C(q)G(q,\theta)}, \tag{7}$$

$$H_{cl}(q,\theta) \triangleq \frac{H(q,\theta)}{1 + C(q)G(q,\theta)}. \tag{8}$$

Equation (6) can then be rewritten as

$$\mathcal{M}(\theta): \ y(t) = G_{cl}(q,\theta)r(t) + H_{cl}(q,\theta)e(t). \tag{9}$$

Note that the closed-loop system is parameterized in the plant parameter θ (like for example in the tailor made instrumental variable method (Gilson and Van den Hof, 2001)). Therefore, the open-loop is directly identified without a previous closed-loop estimation which would lead to an over-parameterized problem.
Most of the existing methods critically rely on the assumption of a persistently exciting of sufficiently high order input signal. Therefore, in the case where this assumption is not verified, they provide unsatisfactory results. This paper gives an attempt to fill this gap by using a Bayesian technique where prior knowledge is added to overcome the lack of information in the data.

3. BAYESIAN TECHNIQUE

3.1 Bayesian inference

Although it relies on the likelihood function as well, the Bayesian approach differs from the classical approach of the maximum likelihood in that it allows to integrate prior knowledge.
Let us consider a statistical model where the probability distribution $p(\cdot \mid \theta)$ generating the observations is given by a parametric model which depends on an unknown parameter $\theta \in \mathbb{R}^k$. The likelihood function associated with an i.i.d. sample $x = (x_1, ..., x_n)$ is given as

$$p(x \mid \theta) = \prod_{i=1}^{n} p(x_i \mid \theta). \tag{10}$$

The prior knowledge on the parameter θ can be expressed through a prior distribution $p(\theta)$. The joint distribution of (x, θ) is given as

$$p(x,\theta) = p(x \mid \theta)p(\theta) = p(\theta \mid x)p(x). \tag{11}$$

It yields the posterior density of θ given x

$$p(\theta \mid x) = \frac{p(x \mid \theta)p(\theta)}{p(x)}, \tag{12}$$

where the denominator does not depend on θ and is written as

$$p(x) = \int p(x \mid \theta)p(\theta)d\theta. \tag{13}$$

Once the posterior density is obtained, the Bayesian inference strictly saying, is over. Different estimators are obtained, depending on the cost function to minimize, like for example the maximum *a posteriori* estimator or the mean *a posteriori* estimator.

Therefore, the Bayesian approach offers a simple, natural and flexible solution. The practical problem is to specify the prior distribution $p(\theta)$ taking the available information into account, to properly characterize the likelihood $p(x \mid \theta)$ as in the frequentist approach (that is the choice and the specification of a good statistical model) and lastly to calculate the posterior distribution $p(\theta \mid x)$ with (12).

However, the numerical computation of the posterior distributions is generally complex. The integral in (13) has indeed to be solved to get the denominator of $p(\theta \mid x)$. If necessary, it is then possible to compute $\mathbb{E}(\theta \mid x)$, $Var(\theta \mid x)$ or $P(\theta \in A \mid x)$.

Furthermore, the parameter vector of interest is often multidimensional: $\theta = (\theta_1, \ldots, \theta_k) \in \mathbb{R}^k$. In this case, posterior marginal densities may be obtained from

$$p(\theta_j \mid x) = \int p(\theta \mid x)d\theta_1 ... d\theta_{j-1} d\theta_{j+1} ... d\theta_k. \tag{14}$$

The computational difficulties are therefore mainly due to the resolution of integrals. When dealing with simple problems, a wise choice of the prior density allows to analytically solve them, but the formulae obtained in this way often become disheartening because of their complexity. However, for most problems no analytical solution is available and it is thus needed to resort to numerical integration; the use of effective software programs and algorithms is then compulsory.

An idea to circumvent those difficulties is to construct a Markov chain which converges to an invariant density equal to the desired posterior density. Sampling from this chain then provides a means for computing posteriors with respect to this density via sample averages from the simulated chain (Ninness and Henriksen, 2003).

3.2 The Metropolis-Hastings algorithm

This section summarizes a specific Markov Chain Monte Carlo (MCMC) method: the Metropolis-Hastings algorithms. Their principle is briefly reviewed here. For a more rigorous and precise treatment of MCMC methods, see (Tierney, 1994) for example.

Let π denote a distribution of interest. MCMC methods rely on the use of an ergodic Markov chain which admits π for stationary distribution. Following an initial transient phase (the "burn-in") the generated chain can be considered as sampled from π.

A Markov chain is a sequence $(X_n)_{n\in\mathbb{N}}$ of random variables such that the conditional distribution of X_n given $X_{n-1}, X_{n-2}, ..., X_0$ is the same that the one of X_n given X_{n-1}. Starting from an initial value X_0, the Metropolis-Hastings algorithm associated with the target distribution π and with the proposal distribution κ produces a Markov chain (X_t) following these steps:

- if the current state of the chain is $X_n = x$, a candidate y for the next state X_{n+1} is generated from κ;
- the candidate is then accepted with a probability $\alpha(x, y)$ and then $X_{n+1} = y$; otherwise it is rejected and the chain does not move: $X_{n+1} = X_n$.

where $\alpha(x, y)$ is the acceptance probability defined as

$$\alpha(x, y) = \min\left\{\frac{\pi(y)\kappa(y, x)}{\pi(x)\kappa(x, y)}, 1\right\}. \tag{15}$$

In this algorithm, π only appears in the ratio $\pi(y)/\pi(x)$: that distribution is therefore only needed to be known up to a constant. This is very interesting, especially in Bayesian inference, because the integral (13) has not to be computed, avoiding to resort to numerical integrations.

Under soft assumptions on the proposal distribution κ (Roberts and Smith, 1994), the Metropolis-Hastings algorithm provides a Markov chain which admits π for stationary distribution. It yields a measure of its "universality": if E denotes the support of the distribution of interest π, *any* distribution whose support contains E permits to generate samples from π. This property can however be prejudicial. An unwise choice of κ can indeed lead to a very slow convergence: it is the case if κ is seldom simulating points in the support of π, for example.

The Metropolis-Hastings algorithm is quite general, admitting many special cases, depending on the choice of the distribution κ. In this paper, the random walk Metropolis-Hastings algorithm is chosen.

The random walk Metropolis-Hastings algorithm. A natural approach for the practical construction of a Metropolis-Hastings algorithm is to take into account the previous simulated value to generate the next value: this prospect is already present in the simulated annealing and stochastic gradient methods. Let f be a density with values in a space E, and suppose that the distribution κ is written as $\kappa(x, y) = f(y - x)$. Then clearly the candidate

y is obtained by $y = x + \varepsilon$ where ε is a random disturbance with distribution f: the Markov chain is a *random walk* on E. The distributions most often considered for the disturbance are uniform distributions on spheres centered in the origin, normal distributions or Student distributions.
Note that the choice of an even function f permits to obtain the original Metropolis algorithm (Metropolis *et al.*, 1953), where, as $\kappa(x, y) = \kappa(y, x)$, the acceptance probability (15) simplifies to

$$\alpha(x, y) = \min\left\{\frac{\pi(y)}{\pi(x)}, 1\right\}. \quad (16)$$

This particular algorithm is used in the simulation example in section 5.

4. CLOSED-LOOP IDENTIFICATION BY A BAYESIAN TECHNIQUE

Suppose that the available data in (9) consists of N pairs $(r(t), y(t))$. Let us define $Y_N \triangleq (y_1, ..., y_N)$. As the noise is an i.i.d sequence of random variables with distribution p_e, the independence assumption allows to write the likelihood function (10) as

$$p(Y_N \mid \theta) = \prod_{t=1}^{N} p_e(e(t)). \quad (17)$$

Hence the posterior distribution is given as

$$p(\theta \mid Y_N) = \frac{p(Y_N \mid \theta)p(\theta)}{p(Y_N)} \quad (18)$$

$$= \frac{p(\theta)}{p(Y_N)} \prod_{t=1}^{N} p_e(e(t)), \quad (19)$$

where $p(\theta)$ denotes the prior distribution of θ and $p(Y_N)$ is a constant which is not to be calculated: a Metropolis-Hastings algorithm is used to obtain samples from the posterior distribution (18).
From (2), (4) and (6) it results that

$$e(t) = (AP + BQ)y(t) - BPr(t) + (1 - APH)e(t), \quad (20)$$

where the q and θ notations have been omitted for ease of reading. However, writing the likelihood function using equation (20) yields

$$p(Y_N \mid \theta) = \prod_{t=1}^{N} p_e\Big((AP + BQ)y(t) - BPr(t) + (1 - APH)e(t)\Big). \quad (21)$$

So the noise $e(t)$ has to be known, which is obviously paradoxical. A first solution to overcome this problem is then to substitute it by a known quantity such as the prediction error. This proposition is connected to the maximum likelihood technique and pseudo-linear regressions (Johansson, 1993; Ljung, 1999). Another solution consists in modifying the considered set of models $\mathcal{M}$ so as not to be confronted with this noise problem anymore. For the sake of place, only the latter proposition is developed in this paper.
From (2), (4) and (6) the model to be identified is described by

$$\mathcal{M}(\theta) : y(t) = \frac{BP}{AP+BQ} r(t) + \frac{APH}{AP+BQ} e(t). \quad (22)$$

The set of models (9) within which the best model is sought is changed to the following ARX model

$$\bar{\mathcal{M}}(\theta) \ : \ (AP + BQ)y(t) = BPr(t) + e(t). \quad (23)$$

Therefore, the context of the identification is no longer $\mathcal{S} \in \mathcal{M}$ but is reduced to $G_0 \in \mathcal{G}$, which is more realistic (Ljung, 1999): this modification is not really restrictive since generally no information concerning the noise is available.
From equation (23) it is now possible to write

$$e(t) = (AP + BQ)y(t) - BPr(t). \quad (24)$$

Nothing prevents us now from writing the likelihood function: the noise no longer appears in the right part of the equation (compare to (20)).
Although this method may seem simple, it provides good results if the signal-to-noise ratio (SNR) is not too small. It is indeed logical to obtain biased estimates as the noise becomes increasingly important. This is however critical only for a very small SNR, for which the usual methods give erroneous results as well.

5. SIMULATION EXAMPLE

The following example is used to illustrate the performances of the proposed method. The process to be identified is described by (1), where

$$G_0(q) = \frac{q^{-1} + 0.5q^{-2}}{1 - 1.85q^{-1} + 0.525q^{-2}}, \quad (25)$$

$$C(q) = \frac{0.35 - 0.28q^{-1}}{1 - 0.8q^{-1}}. \quad (26)$$

The parameter vector to be estimated is therefore given by $\theta_0 = (-1.85,\ 0.525,\ 1,\ 0.5)$. The excitation signal $r(t)$ is a simple step function made up of only $N = 10$ points, switching from 1 to 0 at the $5th$ sample. $e_0(t)$ is a Gaussian white noise uncorrelated with $r(t)$. The signal-to-noise ratio is

$$SNR = 10 \log\left(P_{y_d}/P_e\right) = 30dB, \quad (27)$$

where P_x denotes the power of the signal x and y_d is the noise-free output signal. Figure (2) shows the excitation signal r and the response of the system y.
Figures (3) and (4) show the results obtained by the proposed algorithm: the histograms of the posterior distributions, as well as the Markov chains that have been generated are plotted. The prior distributions on the parameters are Gaussian distributions centered on the true values but with large variance (equal to 10 for each parameter).

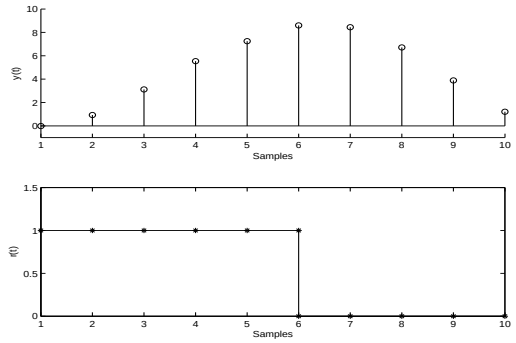

Figure 2. Input/output data.

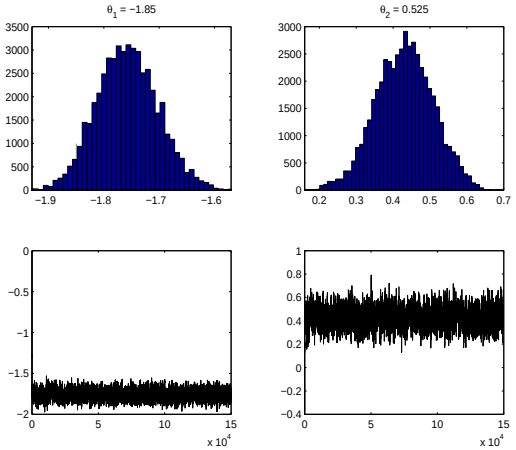

Figure 3. Histograms of the posterior distribution (on θ_1 and θ_2) and Markov chains associated, 150000 iterations.

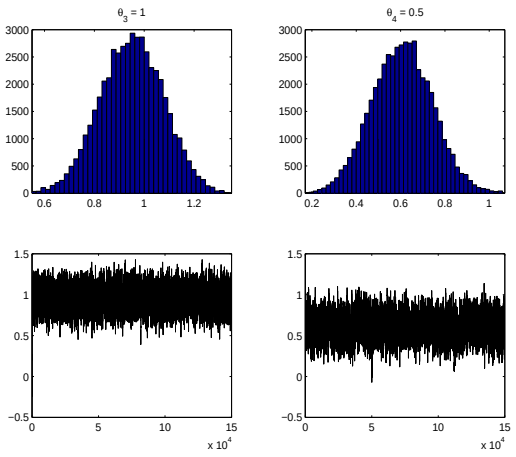

Figure 4. Histograms of the posterior distribution (on θ_3 and θ_4) and Markov chains associated, 150000 iterations.

The mean *a posteriori* estimator yields

$$\hat{\theta}_{bay} = (-1.7579 \quad 0.4310 \quad 0.9499 \quad 0.6066). \quad (28)$$

The parameter estimates obtained from the tailor-made instrumental variable (TIV) method (Gilson and Van den Hof, 2001) are

$$\hat{\theta}_{tiv} = (-2.0449 \quad 0.8274 \quad 0.9307 \quad 0.2276). \quad (29)$$

A first conclusion is that the mean *a posteriori* estimate seems to give more accurate results than the TIV method for this particular case. However, it is difficult if not impossible to drive a comparison between the proposed approach and the TIV one only from this single estimation. Therefore, the variance of the TIV estimates are computed: $\hat{\theta}_{tiv}$ has been shown to be asymptotically Gaussian distributed

$$\sqrt{N}(\hat{\theta}_{tiv} - \theta^*) \xrightarrow{\mathcal{D}} \mathcal{N}(0, P_{tiv}), \quad (30)$$

with θ^* the asymptotic estimate and P_{tiv} the covariance matrix (Gilson and Van den Hof, 2005). Then, Figure (5) represents the posterior density of the parameters estimated with the Bayesian approach along with the asymptotic results (30) used in a finite data setting in order to provide error quantification. While these quantifications are not strictly comparable to the posterior distribution, since they evaluate different quantities, it would still seem interesting to compare the two in terms of their utility for informing a user of what system information can be extracted from the available data (Ninness and Henriksen, 2003). As we can see, the estimates obtained with the

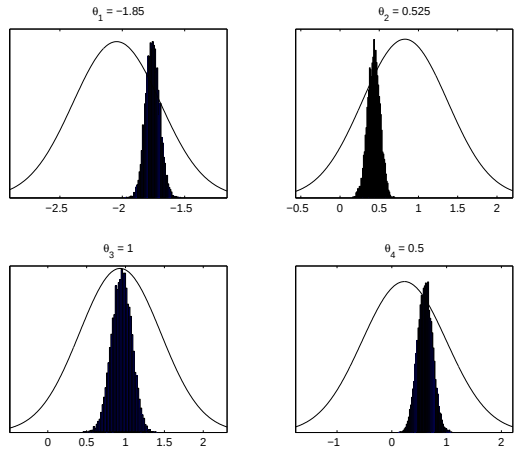

Figure 5. Posterior density of $\hat{\theta}_{bay}$ (histograms) and asymptotic distribution of the TIV estimates (solid line).

proposed method within the framework of very short data length, little informative excitation signal, and a relatively low SNR are really satisfying. The estimates are slightly biased, but the variance is significantly lower than the one obtained by the TIV method. Moreover, the methods that are said to be 'unbiased' are actually *asymptotically unbiased*: for short data length, the bias exists and can be greater than the one obtained with the Bayesian technique, especially if the assumption of a persistently exciting of sufficiently high order input signal is not fulfilled; see Figure (5).

Finally, to give an idea of the conditions under which this identification was carried out, here is given the error message obtained while identifying this system via the two-stage method (Van den Hof and Schrama, 1993), which, when using the MATLAB oe function, delivered "There are too many parameters to estimate for this amount of data".

Another experiment is performed in order to show the performance of the proposed method even in the more realistic case of non accurate *a priori* values. The new prior distributions are therefore no longer centered on the true values but on deliberately erroneous ones. An error normally distributed with zero mean and unit variance is added to the parameter real values to compute the following erroneous *a priori* parameters

$$\theta_{prior} = (-2.2826 \ -1.1406 \ 1.1253 \ 0.7877). \quad (31)$$

The prior distributions are then Gaussian distributions centered on θ_{prior} (31) with variance equal to 10 for each of them. Then, the mean *a posteriori* estimator and the TIV method yield

$$\hat{\theta}_{bay} = (-1.8077 \ \ 0.4667 \ \ 1.0596 \ \ 0.4886) \quad (32)$$

$$\hat{\theta}_{tiv} = (-2.0518 \ \ 0.8959 \ \ 0.9246 \ \ 0.0090). \quad (33)$$

An immediate conclusion is that the results are still satisfying and better than those obtained from the TIV method, even with a sometimes sizeable error on the mean of the prior distributions. Moreover, as previously, the posterior density of $\hat{\theta}_{bay}$ and the asymptotic results stemmed from the TIV method are represented in Figure 6. In conclusion, it can be seen that the proposed method gives accurate results even in the case of an erroneous prior knowledge.

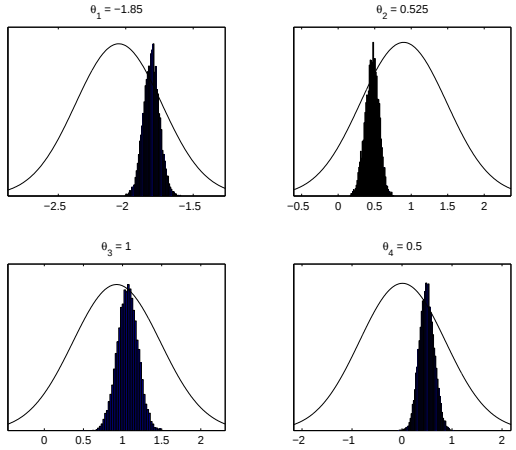

Figure 6. Posterior density of $\hat{\theta}_{bay}$ (histograms) and asymptotic distribution of the TIV estimates (solid line). Case where the prior distributions are not centered in the true values.

6. CONCLUSION AND PROSPECTS

A first attempt to use Bayesian techniques for identifying closed-loop systems has been proposed. In the case of a non persistently exciting of sufficiently high order input signal, existing methods hardly provide satisfying results (if not totally erroneous) while the proposed Bayesian technique offers an interesting and accurate solution. Moreover, contrary to the traditional methods, the proposed technique provides reliable confidence intervals even with a very short data record. Lastly, if the signals are highly noise-corrupted, the behavior of the classical methods is again unsatisfactory as these ones mainly rely on the information contained in the input/output data without taking into account a potential *a priori* information.

On the contrary, the suggested method uses prior knowledge to overcome the problem induced by the insufficiently rich data records. This work is a preliminary proposition and many issues have to be investigated. For example, further work would be to obtain realistic prior distributions by exploiting for example the stability of the system, the gain and the phase margins; it is also possible to lay down a softness constraint on the frequency response, etc.

REFERENCES

Forssell, U. and L. Ljung (1999). Closed-loop identification revisited. *Automatica* **35**(7), 1215–1241.

Gilson, M. and P. Van den Hof (2001). On the relation between a bias-eliminated least-squares (BELS) and an IV estimator in closed-loop identification. *Automatica* **37**(10), 1593–1600.

Gilson, M. and P. Van den Hof (2005). Instrumental variables methods for closed-loop system identification. *Automatica* **41**(2), 241–249.

Johansson, R. (1993). *System modeling and Identification*. Prentice Hall.

Ljung, L. (1999). *System Identification: Theory for the User*. 2nd ed.. Prentice Hall.

Metropolis, N., A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller and E. Teller (1953). Equations of state calculations by fast computing machines. *The Journal of Chemical Physics* **21**(6), 1087–1091.

Ninness, B. and S. Henriksen (2003). System identification via a computational bayesian approach. In: *13th IFAC Symposium on System Identification*. Rotterdam - Netherlands. pp. 785–790.

Roberts, G. O. and A. F. M. Smith (1994). Simple conditions for the convergence of Gibbs sampler and Metropolis-Hastings algorithms. *Stochastic Processes and their Applications* **49**, 207–216.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *The Annals of Statistics* **22**(4), 1701–1762.

Van den Hof, P. (1998). Closed-loop issues in system identification. *Annual Reviews in Control* **22**, 173–186.

Van den Hof, P. and R. Schrama (1993). An indirect method for transfer function estimation from closed-loop issues. *Automatica* **29**(6), 1751–1770.