

# A Novel Approach to Model Error Modelling using the Expectation-Maximization Algorithm

Ramón A. Delgado, Graham C. Goodwin, Rodrigo Carvajal  
and Juan C. Agüero.

**Abstract**—In this paper we develop a novel approach to model error modelling. There are natural links to others recently developed ideas. However, here we make several key departures, namely (i) we focus on relative errors; (ii) we use a broad class of model error description which includes, inter alia, the earlier idea of stochastic embedding; (iii) we estimate both, the nominal model and undermodelling simultaneously using the Expectation-Maximization (EM) algorithm. Simulation studies illustrate the performance of the proposed technique.

**Index Terms**—model error modelling, system identification, EM algorithm.

## I. INTRODUCTION

The idea of modelling error has been a central theme of statistics, time series analysis, econometrics and system identification (see e.g. [1]–[3] and the reference therein). In practice, physical systems are more complex than models that describe the systems behaviour (see e.g. [2], [4]). This fact is well recognized in the areas of system and control, where robustness to model inadequacy has been a key focus.

Most methods in classical estimation theory treat modelling errors as additive noise. On the other hand, multiplicative error is recognized as an useful error model description [5]–[7].

Uncertainty modelling have attracted increasing attention in the past 20 years, were methods such as Set Membership, Stochastic Embedding and Model Error Modelling has been developed. The Set membership (SM) approach (see e.g. [8], [9]), deals with the problem of “bounded but unknown errors” in a deterministic framework. SM delivers a set of possible solutions, and turn the focus into the “worst-case” model.

On the other hand, a probabilistic framework was considered in [10]–[13], where a set of possible solutions can be found based on finite sample properties. Moreover, confidence set were found using Sub-sampling, Resampling and/or Bootstrapping techniques (see e.g. [14]–[17]).

In [18] it was shown that the prior information of the smoothness in the impulse response can be used to estimate bounds to the set of possible solutions.

The methods mentioned above make few assumptions about the model error. On the other hand, some approaches assign a model to the model error. For example, in the *Model Error Modelling* (MEM) method [19], the model error

description is obtained from the residues,  $\varepsilon = y - \hat{y}$ . In MEM, the model error description is considered as the sum of two terms. The first term is a function of the input signal, and the second term is a function of a random process.

Another method that assigns a model to the model error corresponds to the *Stochastic Embedding* (SE) approach (see e.g. [20], [21]). This approach proposes to embed the nominal model structure into a larger class of models. In this sense, a characterization of the undermodelling is included as a realization of a stochastic process. Moreover, the estimation is carried out in a two step procedure. (i) obtaining a non parametric estimation of the frequency response, then (ii) the parametric model is obtained. In SE the nominal model is obtained by using Least Squares. This limits the complexity of the models that can be handled. For a comparison between SE, SM and MEM see [22].

The key difficulty in undermodelling description is finding a satisfactory, broad, and systematic method to describe the model error. In this paper we propose a systematic methodology to describe a broad class of model uncertainties. The main idea is to treat the model uncertainty as a realization of a stochastic process. The estimation problem is solved using an Expectation-Maximization (EM) algorithm, estimating the nominal and uncertainty models simultaneously.

The layout of the remainder of the paper is as follows: In section II we review MEM and SE. In section III we present a description of the model of interest in our approach. In section IV the estimation algorithm is described, and the main result is presented. In section V we show a numerical example. Finally, we draw conclusions in section VI.

## II. MODEL ERROR MODELLING AND STOCHASTIC EMBEDDING

In this section we present a brief description of the MEM, and the SE approaches.

### A. Model Error Modelling

In MEM, the dynamic system  $G(\beta, q)$  corresponds to a model class parametrized by  $\beta \in \Omega$ , where  $\Omega$  is a constraint set in the parameter space, and  $q$  is the forward shift operator. The system output is represented by

$$y_t = G(\beta, q)u_t + v_t, \quad (1)$$

where  $u_t$  is the input signal,  $v_t$  is additive noise and  $y_t$  the measured output,  $t = 0, 1, \dots, N-1$ . By minimizing a cost function, such as the utilized in Prediction Error Method (PEM) [23], it is possible to obtain an estimate of the system

R.A. Delgado, G.C. Goodwin, R. Carvajal and J.C. Agüero are with the School of Electrical Engineering and Computer Science, The University of Newcastle, Australia.

Peer-reviewed author's copy of:

R.A. Delgado, G.C. Goodwin, R. Carvajal and J.C. Agüero. **A Novel Approach to Model Error Modelling using the Expectation-Maximization Algorithm**. In *51st IEEE Conference on Decision and Control (CDC)*, Maui, USA, 2012.

Available at <http://dx.doi.org/10.1109/CDC.2012.6426633>

©2017 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

parameters, denoted by  $\hat{\beta}$ . This estimate is then utilized to obtain the model errors (residues) of the estimates (see e.g. [19])

$$\varepsilon_t = y_t - G(\hat{\beta}, q)u_t. \quad (2)$$

Next, the following model structure is assigned to the residues

$$\varepsilon_t = F(\gamma, q)u_t + H(\gamma, q)w_t, \quad (3)$$

where  $F(\gamma, q)$ , and  $H(\gamma, q)$  are rational transfer function in the operator  $q$ , and parametrized by  $\gamma \in \Gamma$ . The parameters of the error model in (3) are then estimated using an estimation procedure such as PEM.

This procedure provides a confidence set that can be used for model validation or for an specific application, such as robust control.

### B. Stochastic Embedding

The SE approach treats both noise and undermodelling as stochastic process.

In [20] the SE approach considers additive undermodelling. However, in [21] the undermodelling is described as a multiplicative error in the form  $G = G_o(1 + G_\Delta)$  where  $G_o$  is the nominal model and  $G_\Delta$  is the model uncertainty.

When the undermodelling is considered as multiplicative noise, the estimation procedure is given as follows. Consider the input signal as a finite sum of cosine functions [21]

$$u_t = \sum_{r=1}^m a_r \cos(\omega_r t), \quad (4)$$

where  $m$  is the number of elements in the summation,  $\omega_r$  is the  $r$ -th normalized discrete frequency given by  $\omega_r = \frac{2\pi k_r}{N}$ ,  $k_r \in \{1, \dots, N\}$ , and  $m \leq N$ . The periodic input,  $u_t$  is used to obtain a point (non-parametric) estimate of the transfer function of the system,  $g(j\omega_r)$ ,  $\forall r$ , and its associated measurement errors at each  $\omega_r$  of the (not necessarily equally spaces) frequencies set  $\{\omega_1, \omega_2, \dots, \omega_m\}$ . Then, these frequency domain estimates are used to obtain a parametric model of  $g(j\omega_r)$  by estimating the parameters that define it, and the parameters of the embedded stochastic process that represent the undermodelling.

First, assuming that the measurements are obtained under steady state conditions, the sampled output response is expressed as

$$y_t = \sum_{r=1}^m a_r g^r(\omega_r) \cos(\omega_r t) - \sum_{r=1}^m a_r g^i(\omega_r) \sin(\omega_r t) + v_t, \quad (5)$$

where  $g^r(\omega_r)$  and  $g^i(\omega_r)$  are the real and imaginary parts of  $g(j\omega_r)$ , i.e.  $g(j\omega_r) = g^r(\omega_r) + jg^i(\omega_r)$ . Also,  $v_t$  is additive white noise with variance  $\sigma_v^2$ . Next, we rewrite (5) in vector form as

$$Y = \Phi G + V, \quad (6)$$

where

$$Y = [y_1 \ y_2 \ \dots \ y_N]^T, \quad (7)$$

$$G = [g^r(\omega_1) \ g^i(\omega_1) \ \dots \ g^r(\omega_m) \ g^i(\omega_m)]^T, \quad (8)$$

$$\Phi = \begin{bmatrix} \cos(\omega_1) & \sin(\omega_1) & \dots & \sin(\omega_m) \\ \cos(\omega_1 2) & \sin(\omega_1 2) & \dots & \sin(\omega_m 2) \\ \vdots & \vdots & \ddots & \vdots \\ \cos(\omega_1 N) & \sin(\omega_1 N) & \dots & \sin(\omega_m N) \end{bmatrix}, \quad (9)$$

$$\times \text{diag}[a_1, -a_1, a_2, -a_2, \dots, a_m, -a_m], \quad (10)$$

$$V = [v_1 \ v_2 \ \dots \ v_N]^T. \quad (11)$$

Then, an unbiased estimate of  $G$  is given by

$$\hat{G} = \begin{bmatrix} \hat{g}^r(\omega_1) \\ \hat{g}^i(\omega_1) \\ \vdots \\ \hat{g}^r(\omega_m) \\ \hat{g}^i(\omega_m) \end{bmatrix} = (\Phi^T \Phi)^{-1} \Phi^T Y. \quad (12)$$

The estimate  $\hat{g}(j\omega)$  is represented in terms of a given set of basis functions  $b_1(j\omega), \dots, b_p(j\omega)$ . We define the vectors  $\mathbf{B}^r(\omega) = [b_1^r(\omega), \dots, b_p^r(\omega)]$  and  $\mathbf{B}^i(\omega) = [b_1^i(\omega), \dots, b_p^i(\omega)]$  containing the real and imaginary parts of the basis function. We denote by  $\theta$  the vector containing the coefficient of the basis functions that describe the nominal model, and by  $\bar{\theta}$  the vector of coefficients of the basis functions that describe the undermodelling. Thus, we can represent our system as

$$\hat{G} = \mathcal{B}\theta + \Lambda\bar{\theta} + \tilde{G}, \quad (13)$$

where

$$\mathcal{B} = [\mathbf{B}^r(\omega_1) \ \mathbf{B}^i(\omega_1) \ \dots \ \mathbf{B}^r(\omega_m) \ \mathbf{B}^i(\omega_m)]^T \quad (14)$$

$$\tilde{G} = [\tilde{g}^r(\omega_1) \ \tilde{g}^i(\omega_1) \ \dots \ \tilde{g}^r(\omega_m) \ \tilde{g}^i(\omega_m)]^T, \quad (15)$$

$$\Lambda = \text{diag}[\lambda^r(\omega_1), \lambda^i(\omega_1), \dots, \lambda^r(\omega_m), \lambda^i(\omega_m)]. \quad (16)$$

The noise signals  $\{\tilde{g}^r\}$  and  $\{\tilde{g}^i\}$  in (15), are uncorrelated white noises having variance  $2\sigma_v^2/(a_r^2 N)$ . The random process  $\{\lambda^r\}$  and  $\{\lambda^i\}$  in (16) are two independent random walk process. A different model can be considered assuming that  $\{\lambda^r\}$  and  $\{\lambda^i\}$  are two independent integrated random walks. For more details see [21].

*Remark 1:* In order to incorporate the undermodelling as a multiplicative error,  $\theta = \bar{\theta}$  must be satisfied in (13). However, if this equality is satisfied, then the estimation problem is difficult, because the matrix of regressors  $\Lambda\mathcal{B}$  is a function of the undermodelling. For this reason, the estimation is carried out in two steps. First, an estimate  $\hat{\theta}$  of  $\theta$  is obtained. Next, the estimation of the undermodelling variance is obtained considering that  $\bar{\theta} = \hat{\theta}$ .  $\square$

### III. MODEL DESCRIPTION

Consider the following description of the model

$$G(j\omega_k) = G_o(j\omega_k)(1 + G_\Delta(j\omega_k)), \quad (17)$$

and the corresponding data generating system

$$Y(j\omega_k) = G(j\omega_k)U(j\omega_k) + V(j\omega_k), \quad (18)$$

where  $\omega_k = \frac{2\pi k}{N}$ ,  $Y(j\omega_k)$  and  $U(j\omega_k)$  are the Discrete Fourier Transform of the measurement and the input signal, respectively. Moreover,  $V(j\omega_k)$  and  $G_\Delta(j\omega_k)$  are realizations of random process, with probability distributions  $p(G_\Delta(j\omega_k))$  and  $p(V(j\omega_k))$ , respectively.

This representation of  $G_\Delta(j\omega_k)$  as a realization of a random process allow the representation of a class of possible models characterized by  $G_o(j\omega_k)$  and  $p(G_\Delta(j\omega_k))$ .

*Remark 2:* The assumption that  $G_\Delta(j\omega_k)$  is a realization of a stochastic process suggest that when collecting data multiple realizations of uncertainty must be collected.  $\square$

In order to simplify the notation, we denote  $G_{\Delta,k}^{(p)} \triangleq G_\Delta(j\omega_k)^{(p)}$ ,  $G_\Delta^{(p)} = \{G_{\Delta,k}^{(p)}\}_{k=0}^L$ ,  $Y_k^{(p)} \triangleq Y(j\omega_k)^{(p)}$  and  $U_k^{(p)} \triangleq U(j\omega_k)^{(p)}$ .

#### IV. EM-BASED ESTIMATION

In this section, we provide a description of the estimation algorithm that is based on the Expectation-Maximization (EM) algorithm [24].

The EM algorithm has been used to identify different classes of dynamic systems, such as, continuous time systems using sampled-data [25], finite impulse response systems using quantized systems [26], state-space systems using incomplete data [27], channel estimation in telecommunications [28], bilinear state-space systems [29], and non-linear state-space systems [30].

The EM algorithm is an iterative two step procedure [24] where the concept of complete data is introduced. The complete data is composed of the measured data,  $\mathcal{Y}$ , and also an unmeasured data set known as the *hidden data*,  $\mathcal{X}$ . Then, (loosely speaking) one estimates the hidden data based on the current parameter estimate (in the Expectation step, E-step) and then updates the parameters by maximizing a function that depends on the joint probability density function (pdf) of the hidden data and the measurements evaluated at the estimated hidden data (in a Maximization step, M-step).

In detail, given a current estimate  $\hat{\theta}_i \in \Omega$ , where  $\Omega$  is the constraint set in the parameter space, an iteration of the EM algorithm is defined by:

##### E-step

$$\mathcal{Q}(\theta, \hat{\theta}_i) = \mathbf{E}\{\log p(\mathcal{Y}, \mathcal{X}|\theta)|\mathcal{Y}, \hat{\theta}_i\}. \quad (19)$$

##### M-step

$$\hat{\theta}_{i+1} = \arg \max_{\theta \in \Omega} \mathcal{Q}(\theta, \hat{\theta}_i). \quad (20)$$

where  $p(\mathcal{Y}, \mathcal{X}|\theta)$  is the joint probability density function (pdf) of  $\mathcal{Y}$  and  $\mathcal{X}$  given  $\theta$ .

For the development of the algorithm we consider  $G_\Delta(j\omega_k)$  as the hidden variable.

*Lemma 1:* Consider the system given by (17)-(18), and that  $G_\Delta \sim p(G_\Delta)$  is the hidden variable in the EM algorithm. Also consider that the data is collected from  $N_{exp}$  independent experiments, i.e.  $\{G_\Delta^{(p)}, \mathcal{Y}^{(p)}\}_{p=1}^{N_{exp}}$ . Then

the auxiliary function  $\mathcal{Q}(\theta, \hat{\theta}_i)$  in the EM algorithm is given by

$$\begin{aligned} \mathcal{Q}(\theta, \hat{\theta}_i) = & \sum_{p=1}^{N_{exp}} \mathbf{E} \left\{ \log p(\mathcal{Y}^{(p)} | G_\Delta^{(p)}, \theta) | \mathcal{Y}^{(p)}, \hat{\theta}_i \right\} \\ & + \sum_{p=1}^{N_{exp}} \mathbf{E} \left\{ \log p(G_\Delta^{(p)} | \theta) | \mathcal{Y}^{(p)}, \hat{\theta}_i \right\} \end{aligned} \quad (21)$$

*Proof:* Following some ideas developed in [31], we rewrite (19) to consider all the experiments as

$$\mathcal{Q}(\theta, \hat{\theta}_i) = \mathbf{E} \left\{ \log p \left( \{\mathcal{Y}^{(p)}, G_\Delta^{(p)}\}_{p=1}^{N_{exp}} | \theta \right) \middle| \{\mathcal{Y}^{(p)}\}_{p=1}^{N_{exp}}, \hat{\theta}_i \right\}. \quad (22)$$

Given that the experiments are independent, using Bayes' theorem, we have that

$$\mathcal{Q}(\theta, \hat{\theta}_i) = \sum_{p=1}^{N_{exp}} \mathbf{E} \left\{ \log p \left( \mathcal{Y}^{(p)}, G_\Delta^{(p)} | \theta \right) \middle| \mathcal{Y}^{(p)}, \hat{\theta}_i \right\}, \quad (23)$$

and using Bayes' theorem we express  $\log p(\mathcal{Y}^{(p)}, G_\Delta^{(p)} | \theta)$  as

$$\log p(\mathcal{Y}^{(p)}, G_\Delta^{(p)} | \theta) = \log p(\mathcal{Y}^{(p)} | G_\Delta^{(p)}, \theta) + \log p(G_\Delta^{(p)} | \theta). \quad (24)$$

Finally, replacing (24) in (23), the result in (21) follows.  $\blacksquare$

In particular, when  $G_{\Delta,k+1}^{(p)}$  and  $Y_k^{(p)}$  are independent, proper and Gaussian distributed [32], then, the joint pdf of  $\mathcal{Y}^{(p)}$  and  $G_\Delta^{(p)}$  is given by

$$p(\mathcal{Y}^{(p)}, G_\Delta^{(p)} | \theta) = \prod_{k=0}^{L-1} p \left( G_{\Delta,k+1}^{(p)}, Y_k^{(p)} | G_{\Delta,k}^{(p)}, \theta \right) p(G_{\Delta,0}^{(p)}), \quad (25)$$

and the joint pdf of  $G_{\Delta,k+1}^{(p)}$  and  $Y_k^{(p)}$  is given by

$$p \left( G_{\Delta,k+1}^{(p)}, Y_k^{(p)} | G_{\Delta,k}^{(p)}, \theta \right) \sim N_p \left( \begin{bmatrix} \mu_{\Delta,k}^{(p)} \\ \mu_{y,k}^{(p)} \end{bmatrix}; \begin{bmatrix} \sigma_{\Delta,k}^2 & 0 \\ 0 & \sigma_{y,k}^2 \end{bmatrix} \right) \quad (26)$$

where  $N_p(\cdot, \cdot)$  represents a proper Gaussian distribution, and

$$G_{\Delta,0}^{(p)} \sim N_p(\mu_{\Delta,-1}^{(p)}; \sigma_{\Delta,-1}^2). \quad (27)$$

*Corollary 1:* Consider that the pdf of the complete data is given by (25)-(26), and  $G_{\Delta,0}^{(p)}$  is given by (27). Then, the auxiliary function  $\mathcal{Q}(\theta, \hat{\theta}_i)$  of the EM algorithm is given by

(excluding constant terms)

$$\begin{aligned} \mathcal{Q}(\theta, \hat{\theta}_i) = & -N_{exp} \left( \sum_{k=0}^{L-1} \log \sigma_{\Delta,k}^2 + \sum_{k=0}^{L-1} \log \sigma_{y,k}^2 + \log \sigma_{\Delta,-1}^2 \right) \\ & - \sum_{p=1}^{N_{exp}} \sum_{k=0}^{L-1} \left[ \frac{1}{\sigma_{\Delta,k}^2} \right. \\ & \cdot \mathbf{E} \left\{ (G_{\Delta,k+1}^{(p)} - \mu_{\Delta,k}^{(p)})(G_{\Delta,k}^{(p)} - \mu_{\Delta,k}^{(p)})^* \middle| \mathcal{Y}^{(p)}, \hat{\theta}_i \right\} \\ & - \sum_{p=1}^{N_{exp}} \sum_{k=0}^{L-1} \frac{1}{\sigma_{y,k}^2} \mathbf{E} \left\{ (Y_k^{(p)} - \mu_{y,k}^{(p)})(Y_k^{(p)} - \mu_{y,k}^{(p)})^* \middle| \mathcal{Y}^{(p)}, \hat{\theta}_i \right\} \\ & - \sum_{p=1}^{N_{exp}} \frac{1}{\sigma_{\Delta,-1}^2} \end{aligned} \quad (28)$$

$$\times \mathbf{E} \left\{ (G_{\Delta,0}^{(p)} - \mu_{\Delta,-1}^{(p)})(G_{\Delta,0}^{(p)} - \mu_{\Delta,-1}^{(p)})^* \middle| \mathcal{Y}^{(p)}, \hat{\theta}_i \right\}, \quad (29)$$

where  $*$  denotes the complex conjugate.

*Proof:* From (19) and (25) we have that

$$\begin{aligned} \mathcal{Q}(\theta, \hat{\theta}_i) = & \sum_{p=1}^{N_{exp}} \mathbf{E} \left\{ \sum_{k=0}^{L-1} \log p \left( G_{\Delta,k+1}^{(p)}, Y_k^{(p)} \middle| G_{\Delta,k}^{(p)} \right) \middle| \mathcal{Y}^{(p)}, \hat{\theta}_i \right\} \\ & + \sum_{p=1}^{N_{exp}} \mathbf{E} \left\{ \log p(G_{\Delta,0}^{(p)}) \middle| \mathcal{Y}^{(p)}, \hat{\theta}_i \right\}. \end{aligned} \quad (30)$$

From (26) we deduce the following

$$\begin{aligned} \log p \left( G_{\Delta,k+1}^{(p)}, Y_k^{(p)} \middle| G_{\Delta,k}^{(p)}, \theta \right) = & -\log \pi^2 \\ & - \log \det \left( \begin{bmatrix} \sigma_{\Delta,k}^2 & 0 \\ 0 & \sigma_{y,k}^2 \end{bmatrix} \right) \\ & - \left( \begin{bmatrix} G_{\Delta,k+1}^{(p)} \\ Y_k^{(p)} \end{bmatrix} - \begin{bmatrix} \mu_{\Delta,k+1}^{(p)} \\ \mu_{y,k}^{(p)} \end{bmatrix} \right)^H \begin{bmatrix} \sigma_{\Delta,k}^2 & 0 \\ 0 & \sigma_{y,k}^2 \end{bmatrix}^{-1} \\ & \times \left( \begin{bmatrix} G_{\Delta,k+1}^{(p)} \\ Y_k^{(p)} \end{bmatrix} - \begin{bmatrix} \mu_{\Delta,k+1}^{(p)} \\ \mu_{y,k}^{(p)} \end{bmatrix} \right), \\ = & -\log \pi^2 - \log \sigma_{\Delta,k}^2 - \log \det \sigma_{y,k}^2 \\ & - \frac{1}{\sigma_{\Delta,k}^2} (G_{\Delta,k+1}^{(p)} - \mu_{\Delta,k}^{(p)})^* (G_{\Delta,k+1}^{(p)} - \mu_{\Delta,k}^{(p)}) \\ & - \frac{1}{\sigma_{y,k}^2} (Y_k^{(p)} - \mu_{y,k}^{(p)})^* (Y_k^{(p)} - \mu_{y,k}^{(p)}). \end{aligned} \quad (31)$$

In a similar way, for  $G_{\Delta,0}^{(p)}$ , we have:

$$\begin{aligned} \log p(G_{\Delta,0}^{(p)}) = & -\log \pi - \log \det \sigma_{\Delta,-1}^2 \\ & - \frac{1}{\sigma_{\Delta,-1}^2} (G_{\Delta,0}^{(p)} - \mu_{\Delta,-1}^{(p)})^* (G_{\Delta,0}^{(p)} - \mu_{\Delta,-1}^{(p)}), \end{aligned} \quad (32)$$

then replacing back (31) and (32) in (30) we obtain (29). ■ Once  $\mathcal{Q}(\theta, \hat{\theta}_i)$  has been calculated, we can maximize it using standard optimization tools (e.g. Newton-Raphson method).

*Remark 3:* The expected values over the hidden variables, given that  $\mathcal{Y}$  is available, can be obtained, in general, by using the Kalman smoother algorithm (for more details, see e.g. [27], [29], [33], [34]). □

In practice, due to physical properties of real systems, typical error bounds must roughly grow with the frequency

[35]. To consider this, the stochastic process  $G_{\Delta,k}$  can be modelled as a random walk in the frequency domain [21]. In this section, we explore this idea, and consider that  $\tilde{G}_{\Delta}$  is a random walk in the frequency domain, with  $G_{\Delta}(j\omega_0) = 0$ . Then (17) and (18) can be rewritten as

$$G_{\Delta,k+1}^{(p)} = G_{\Delta,k}^{(p)} + W_k^{(p)}, \quad (33)$$

$$Y_k^{(p)} = G_k^{(p)}(\beta)(1 + G_{\Delta,k}^{(p)})U_k^{(p)} + V_k^{(p)}, \quad (34)$$

where  $\beta \in \Omega$  contains the parameters of the nominal model, and  $W_k^{(p)}$  and  $V_k^{(p)}$  are two independent proper white Gaussian noise sequences, with mean and variance given by:

$$\begin{bmatrix} W_k^{(p)} \\ V_k^{(p)} \end{bmatrix} \sim \mathcal{N}_p \left( 0; \begin{bmatrix} \sigma_w^2 & 0 \\ 0 & \sigma_v^2 \end{bmatrix} \right). \quad (35)$$

We are interested in estimating the nominal model parameters  $\beta$ , and the covariances  $\gamma = [\sigma_w^2 \ \sigma_v^2]^T$ . For simplicity, we denote the parameters to be estimated by  $\theta = [\beta^T \ \gamma^T]^T$ .

*Lemma 2:* Consider the system given by (33)-(34), where the random noise sequences are given by (35). The auxiliary function  $\mathcal{Q}^{(p)}(\theta, \hat{\theta}_i)$ , for the experiment  $p$ , of the EM algorithm is given by,

$$\begin{aligned} \mathcal{Q}^{(p)}(\theta, \hat{\theta}_i) = & -L \log \sigma_w^2 - L \log \sigma_v^2 \\ & - \frac{1}{\sigma_w^2} \sum_{k=0}^{L-1} \left[ \mathbf{E} \left\{ G_{\Delta,k+1}^{(p)} G_{\Delta,k+1}^{(p)*} \middle| \mathcal{Y}^{(p)}, \hat{\theta}_i \right\} \right. \\ & - 2\mathbf{Re} \left\{ \mathbf{E} \left\{ G_{\Delta,k+1}^{(p)} G_{\Delta,k}^{(p)*} \middle| \mathcal{Y}^{(p)}, \hat{\theta}_i \right\} \right\} \\ & + \mathbf{E} \left\{ G_{\Delta,k}^{(p)} G_{\Delta,k}^{(p)*} \middle| \mathcal{Y}^{(p)}, \hat{\theta}_i \right\} \left. \right] \\ & + \frac{1}{\sigma_v^2} \sum_{k=0}^L \left[ |Y_k^{(p)}|^2 - 2\mathbf{Re} \left\{ Y_k^{(p)} U_k^{(p)*} G_k(\beta)^* \right\} \right. \\ & - 2\mathbf{Re} \left\{ Y_k^{(p)} \mathbf{E} \left\{ G_{\Delta,k}^{(p)*} \middle| \mathcal{Y}^{(p)}, \hat{\theta}_i \right\} G_k(\beta)^* U_k^{(p)*} \right\} \\ & + |G_k(\beta) U_k^{(p)}|^2 \left( 1 + 2\mathbf{Re} \left\{ \mathbf{E} \left\{ G_{\Delta,k}^{(p)*} \middle| \mathcal{Y}^{(p)}, \hat{\theta}_i \right\} \right\} \right) \\ & + \mathbf{E} \left\{ G_{\Delta,k}^{(p)} G_{\Delta,k}^{(p)*} \middle| \mathcal{Y}^{(p)}, \hat{\theta}_i \right\} \left. \right], \end{aligned} \quad (36)$$

where  $|\cdot|$  and  $\mathbf{Re}\{\cdot\}$  denote the magnitude and the real part of a complex number, respectively.

*Proof:* In (29) consider that the term inside of the first summation is

$$\begin{aligned} \mu_{\Delta,k}^{(p)} &= G_{\Delta,k}^{(p)}, & \sigma_{\Delta,k}^2 &= \sigma_w^2, \\ \mu_{y,k}^{(p)} &= G_k(\beta)(1 + G_{\Delta,k}^{(p)})U_k^{(p)}, & \sigma_{y,k}^2 &= \sigma_v^2, \end{aligned}$$

and  $G_{\Delta,0}^{(p)} = 0$ .

Then we have that

$$\begin{aligned} \sum_{k=0}^{L-1} \log \sigma_{\Delta,k}^2 &= L \log \sigma_w^2, \\ \sum_{k=0}^{L-1} \log \sigma_{y,k}^2 &= L \log \sigma_v^2. \end{aligned}$$

The term in the first summation in (29) is, then, given by

$$\begin{aligned} & \mathbf{E} \left\{ (G_{\Delta,k+1}^{(p)} - \mu_{\Delta,k}^{(p)})(G_{\Delta,k}^{(p)} - \mu_{\Delta,k}^{(p)})^* | \mathcal{Y}^{(p)}, \hat{\theta}_i \right\} = \\ & \mathbf{E} \left\{ G_{\Delta,k+1}^{(p)} G_{\Delta,k+1}^{(p)*} | \mathcal{Y}^{(p)}, \hat{\theta}_i \right\} - \mathbf{E} \left\{ G_{\Delta,k+1}^{(p)} G_{\Delta,k}^{(p)*} | \mathcal{Y}^{(p)}, \hat{\theta}_i \right\} \\ & - \mathbf{E} \left\{ G_{\Delta,k}^{(p)} G_{\Delta,k+1}^{(p)*} | \mathcal{Y}^{(p)}, \hat{\theta}_i \right\} + \mathbf{E} \left\{ G_{\Delta,k}^{(p)} G_{\Delta,k}^{(p)*} | \mathcal{Y}^{(p)}, \hat{\theta}_i \right\}, \end{aligned}$$

which is equivalent to the term inside of the first summation in (36). On the other hand, the term in the second summation in (29) can be expanded as

$$\begin{aligned} & \mathbf{E} \left\{ (Y_k^{(p)} - \mu_{y,k}^{(p)})(Y_k^{(p)} - \mu_{y,k}^{(p)})^* | \mathcal{Y}^{(p)}, \hat{\theta}_i \right\} = |Y_k^{(p)}|^2 \\ & - Y_k^{(p)} U_k^{(p)*} G_k(\beta)^* - G_k(\beta) U_k^{(p)} Y_k^{(p)*} \\ & - Y_k^{(p)} \mathbf{E} \left\{ G_{\Delta,k}^{(p)*} | \mathcal{Y}^{(p)}, \hat{\theta}_i \right\} G_k(\beta)^* U_k^{(p)*} \\ & - G_k(\beta) \mathbf{E} \left\{ G_{\Delta,k}^{(p)} | \mathcal{Y}^{(p)}, \hat{\theta}_i \right\} U_k^{(p)} Y_k^{(p)*} \\ & + |G_k(\beta) U_k^{(p)}|^2 \left( 1 + 2\mathbf{Re} \left\{ \mathbf{E} \left\{ G_{\Delta,k}^{(p)*} | \mathcal{Y}^{(p)}, \hat{\theta}_i \right\} \right\} \right) \\ & + \mathbf{E} \left\{ G_{\Delta,k}^{(p)} G_{\Delta,k}^{(p)*} | \mathcal{Y}^{(p)}, \hat{\theta}_i \right\}. \end{aligned}$$

Which is equivalent to the term in the second summation in (36). Moreover, given that  $G_{\Delta,0}^{(p)}$  is deterministic, the last term in (29) is not included in (36). ■

The computation of  $\mathbf{E} \left\{ G_{\Delta,k} G_{\Delta,k}^* | \mathcal{Y}, \hat{\theta}_i \right\}$ ,  $\mathbf{E} \left\{ G_{\Delta,k+1} G_{\Delta,k}^* | \mathcal{Y}, \hat{\theta}_i \right\}$  and  $\mathbf{E} \left\{ G_{\Delta,k} | \mathcal{Y}, \hat{\theta}_i \right\}$  is carry out using the Kalman smoother (see Appendix).

*Remark 4:* The extension of Lemma 2, for  $V(j\omega_k)$  in (34) being coloured noise, can be done by following the ideas in [36]. □

## V. NUMERICAL EXAMPLE

Consider the system defined in (33) and (34) where

$$G(\theta, z) = \frac{0.12}{z^2 - 1.3z + 0.42}, \quad (37)$$

and  $\sigma_w^2 = 4 \cdot 10^{-4}$ ,  $\sigma_v^2 = 0.1$ . The input signal consist of  $L = 50$  complex-valued samples,  $\{U_k\}_{k=0}^{L-1}$ , generated as a proper Gaussian sequence with zero mean and variance  $\sigma_u^2 = 4$ .

We run  $N_{exp} = 20$  different experiments each with  $L = 50$  samples, which correspond to  $N_{exp}$  realizations of the noise and the uncertainty.

Figure 1 shows the magnitude and phase of the frequency response of different realizations of undermodelling,  $G_{\Delta,k}^{(i)}$ , and the frequency response of the estimated model  $G_k(\hat{\beta})$ . In Figure 1(b) the magnitude of  $G_k$  and  $G_k(\hat{\beta})$  is shown. We can see that the estimated model follows the characteristics of the realizations of the true system (37). The estimated model is given by  $G(\hat{\beta}, z) = 0.116/(z^2 - 1.307z + 0.423)$ ,  $\hat{\sigma}_w^2 = 0.0033$ , and  $\hat{\sigma}_v^2 = 0.0997$ .

## VI. CONCLUSIONS AND FUTURE WORK

In this paper we have applied a novel approach to Model Error Modelling based on the EM algorithm. The proposed method can simultaneously estimate the nominal and the

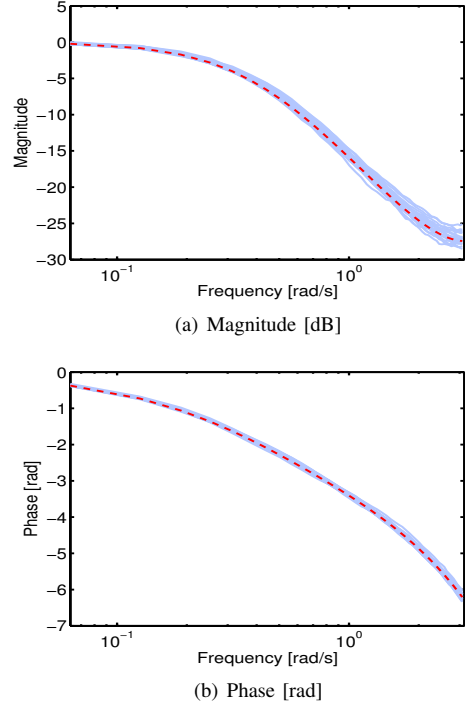


Fig. 1. Magnitude and Phase of the frequency response of different realizations of undermodelling (light blue and continuous line), and the frequency response of the estimated model (red-dashed line).

uncertainty models. Moreover, the estimation algorithm can handle general uncertainty models.

We have presented a numerical example that illustrates the benefits of the proposed approach for system identification.

Future work include a comparison of the proposed approach with other methods, and analysis of performance under several types of uncertainty.

## APPENDIX

### Kalman Smoother

In this appendix, we describe the Kalman smoother algorithm. Consider the following system

$$x_t = Ax_{t-1} + B_t + w_t, \quad (38)$$

$$y_t = C_t x_t + D_t + v_t, \quad (39)$$

where measurements of  $y_t$  are available for  $t = 1, \dots, n$ . The Kalman smoother provides estimates for  $x_t$  based on the entire data sample  $\{y_t\}_{t=1}^n$ , for  $t \leq n$ . The Kalman smoother equations are based on the quantities computed by the Kalman Filter. We first describe the Kalman filter, and later we present the Kalman smoother.

We denote by  $x_{t|s} \triangleq \mathbf{E} \{x_t | y_s, y_{s-1}, \dots, y_1\}$  and by  $P_{t|s} = \mathbf{E} \{ (x_t - x_{t|s})(x_t - x_{t|s})^H | y_s, y_{s-1}, \dots, y_1 \}$ .

We assume that the initial condition  $x_{0|0}$  has mean  $\mu$  and variance  $P_0$ . Then the Kalman Filter for  $t = 1, \dots, n$  is given by (see e.g. [37])

$$x_{t|t-1} = Ax_{t-1|t-1} + B_t, \quad (40)$$

$$P_{t|t-1} = AP_{t-1|t-1}A^T + \Sigma_w, \quad (41)$$

with

$$x_{t|t} = x_{t|t-1} + K_t(y_t - C_t x_{t|t-1} - D_t), \quad (42)$$

$$P_{t|t} = [I - K_t C_t] P_{t|t-1}, \quad (43)$$

where

$$K_t = P_{t|t-1} C_t^H [C_t P_{t|t-1} C_t^H + \Sigma_v]^{-1}. \quad (44)$$

Once the quantities of the Kalman filter has been computed. The Kalman smoother can be computed for  $t = n, n-1, \dots, 1$  by (see e.g. [37])

$$x_{t-1|n} = x_{t|t-1} + J_t(x_{t|n} - x_{t|t-1}), \quad (45)$$

$$P_{t-1|n} = P_{t|t-1} + J_{t-1}(P_{t|n} - P_{t|t-1})J_{t-1}^H, \quad (46)$$

where

$$J_{t-1} = P_{t-1|t-1} A^T [P_{t|t-1}]^{-1}. \quad (47)$$

In our case, we need to compute  $M_{t|n} = \mathbf{E} \{(x_t - x_{t|n})(x_{t-1} - x_{t-1|n})^H | y_n, y_{n-1}, \dots, y_1\}$ . This expectation can be obtained using the quantities in the Kalman smoothing, by defining

$$M_{n|n} = [I - K_n C_n] A P_{n-1|n-1}, \quad (48)$$

and for  $t = n, n-1, \dots, 2$

$$M_{t-1|n} = P_{t-1|t-1} J_{t-2}^H + J_{t-1}(M_{t|n} - A P_{t-1|t-1}) J_{t-2}^H. \quad (49)$$

For the system (33)-(34) we have:  $A = 1$ ,  $B_t = 0$ ,  $C_t = G_t(\beta)U_t$ , and  $D_t = G_t(\beta)U_t$ .

## REFERENCES

- [1] A. Onatski and N. Williams, "Modelling model uncertainty," *Journal of European Economic Association*, vol. 1, no. 5, pp. 1087–1122, Sept. 2003.
- [2] C. Chatfield, "Model uncertainty, data mining and statistical inference," *Journal of the Royal Statistical Society, Series A*, vol. 158, no. 3, pp. 419–466, 1995.
- [3] B. Ninness and G. C. Goodwin, "Estimation of model quality," *Automatica*, vol. 31, no. 12, pp. 1771–1797, 1995.
- [4] G. C. Goodwin, "Some observations on robust estimation and control," in *7th IFAC Symposium on Identification and System Parameter Estimation*, York, UK, 1985.
- [5] G. C. Goodwin, J. C. Agüero, J. S. Welsh, J. I. Yuz, G. J. Adams, and C. R. Rojas, "Robust identification of process models from plant data," *Journal of Process Control*, vol. 18, pp. 810–820, 2008.
- [6] J. Doyle and G. Stein, "Multivariable feedback design: concepts for a classical/modern synthesis," *IEEE Transactions on Automatic Control*, 1981.
- [7] J. C. Agüero, G. C. Goodwin, T. Söderström, and J. I. Yuz, "Sampled data errors in variables systems (I)," in *15th IFAC Symposium on System Identification*, Saint-Malo, France, 2009.
- [8] M. Milanese and A. Vicino, "Optimal estimation theory for dynamic system with set membership uncertainty: an overview," *Automatica*, vol. 27, pp. 997–1009, 1991.
- [9] M. Milanese, J. P. Norton, H. Piet-Lahanier, and E. Walter, Eds., *Bounding approaches to system identification*. Plenum Press, 1996.
- [10] M. C. Campi and E. Weyer, "Finite sample properties of system identification methods," *IEEE Transactions on Automatic Control*, vol. 47, pp. 1329–1334, 2002.
- [11] —, "Guaranteed non asymptotic confidence regions in system identification," *Automatica*, vol. 41, pp. 1751–1764, 2005.
- [12] S. R. Venkatesh and M. A. Dahleh, "On system identification of complex systems from finite data," *IEEE Transactions on Automatic Control*, vol. 46, pp. 235–257, 2001.
- [13] M. Vidyasagar and R. L. Karandikar, "A learning theory approach to system identification and stochastic adaptive control," *Journal of Process Control*, vol. 18, no. 3–4, pp. 421 – 430, 2008.
- [14] F. Tjörnström and L. Ljung, "Using the Bootstrap to estimate the variance in case of undermodelling," *IEEE Transactions on Automatic Control*, vol. 47, pp. 395–398, 2002.
- [15] S. Bittanti and M. Lovera, "Bootstrap-based estimates of uncertainty in subspace identification methods," *Automatica*, vol. 36, pp. 1605–1615, 2000.
- [16] W. J. Dunstan and R. R. Bitmead, "Empirical estimation of parameter distributions in system identification," in *Proc. of the 13th IFAC Symposium on system Identification*, Rotterdam, The Netherlands, 2003.
- [17] S. Garatti and R. R. Bitmead, "On resampling and uncertainty estimation in linear system identification," *Automatica*, vol. 46, no. 5, pp. 785 – 795, 2010.
- [18] T. Chen, H. Ohlsson, G. C. Goodwin, and L. Ljung, "Kernel selection in linear system identification part ii: A classical perspective," in *50th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC)*, Orlando, FL, USA, December 2011.
- [19] L. Ljung, "Model validation and model error modelling," in *The Åström symposium on control*, Lund, Sweden, 1999.
- [20] G. C. Goodwin and M. E. Salgado, "Stochastic embedding approach for quantifying uncertainty in the estimation of restricted complexity models," *International Journal of Adaptive Control and Signal Processing*, vol. 3, no. 4, pp. 333–356, 1989.
- [21] G. C. Goodwin, J. H. Braslavsky, and M. M. Seron, "Non-stationary stochastic embedding for transfer function estimation," *Automatica*, vol. 38, no. 1, pp. 47 – 62, 2002.
- [22] W. Reinelt, A. Garulli, and L. Ljung, "Comparing different approaches to model error modeling in robust identification," *Automatica*, vol. 38, no. 5, pp. 787 – 803, 2002.
- [23] L. Ljung, *System Identification: Theory for the user*, 2nd ed. Prentice Hall, 1999.
- [24] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from imcomplete data via the EM algorithm," *Journal of the Royal Statistical Society, Series B*, vol. 39, no. 1, pp. 1–38, 1977.
- [25] J. Yuz, J. Alfaro, J. Agüero, and G. Goodwin, "Identification of continuous-time state space models from nonuniform fast-sampled data," *IET Control Theory and Applications*, vol. 5, no. 7, pp. 842–855, 2011.
- [26] D. Marelli, B. Godoy, and G. Goodwin, "A scenario-based approach to parameter estimation in state-space models having quantized output data," in *Decision and Control (CDC), 2010 49th IEEE Conference on*, dec. 2010, pp. 2011 –2016.
- [27] J. C. Agüero, W. Tang, J. I. Yuz, R. A. Delgado, and G. C. Goodwin, "Dual time-frequency domain system identification," *To appear in Automatica*, 2012.
- [28] R. Carvajal, J. C. Agüero, B. I. Godoy, and G. C. Goodwin, "EM-based channel estimation in OFDM systems with phase noise," in *GLOBECOM*, 2011, pp. 1–5.
- [29] S. Gibson, A. Wills, and B. Ninness, "Maximum-likelihood parameter estimation of bilinear systems," *Automatic Control, IEEE Transactions on*, vol. 50, no. 10, pp. 1581 – 1596, oct. 2005.
- [30] T. B. Schön, A. Wills, and B. Ninness, "System identification of nonlinear state-space models," *Automatica*, vol. 47, no. 1, pp. 39–49, Jan. 2011.
- [31] J. Agüero, J. Yuz, and G. Goodwin, "Frequency domain identification of MIMO state space models using the em algorithm," in *European Control Conference - ECC'07*, Kos, Greece, 2–5 July 2007.
- [32] B. Picinbono, *Random signals and systems*. Englewood Cliffs (NJ): Prentice Hall, 1993.
- [33] R. H. Shumway and D. S. Stoffer, "An approach to time series smoothing and forecasting using the EM algorithm," *Journal of Time Series Analysis*, vol. 3, no. 4, pp. 253–264, 1982.
- [34] A. Jazwinski, *Stochastic processes and filtering theory*, ser. Mathematics in science and engineering. Academic Press, 1970, no. v. 63.
- [35] K. Zhou, J. Doyle, and K. Glover, *Robust and optimal control*. Prentice Hall, 1996.
- [36] R. Pintelon and J. Schoukens, *System Identification: A Frequency Domain Approach*. Wiley, 2012.
- [37] R. H. Shumway and D. S. Stoffer, *Time Series Analysis and Its Applications*. Springer-Verlag, 2000.