IFAC

NARX model identification with error filtering [1]

Luigi Piroddi * Marco Lovera *

* Dipartimento di Elettronica e Informazione
Politecnico di Milano
P.za Leonardo da Vinci 32, 20133 Milano, Italy
Phone: + 39 02 23993592; Fax: +39 02 23993412
Email: {piroddi, lovera}@elet.polimi.it

Abstract: Model identification of polynomial NARX models involves a lengthy and computationally intensive procedure for selecting the model structure among a possibly large set of candidate regressors. If the model structure is under-parameterized to reduce the burden of the model selection phase, unsatisfactory results are generally obtained. This inaccuracy problem can be somewhat circumvented by focusing the identification process on the obtainment of an accurate local model over a specific frequency range. Such frequency tailoring is achieved in the nonlinear modeling framework by direct error filtering, as opposed to the data pre-filtering practice adopted in the linear context. This work discusses the application of error filtering to classical NARX model identification methods. A simulation example is provided to show the performance of the proposed approach in deriving accurate local models, despite an under-parameterized model structure. It is also shown that a proper error filtering may increase the model accuracy in simulation with respect to available identification techniques.

Keywords: System identification, Nonlinear systems, Frequency domain, Volterra series, Filtering techniques, Data processing.

1. INTRODUCTION

A widely used class of models in black-box nonlinear system identification is the Nonlinear AutoRegressive Moving Average model with eXogenous inputs (NARMAX) [Leontaritis and Billings, 1985], in view of its representation capabilities and the flexibility of the model structure. The NARMAX model is an input-output recursive model where the current output depends on lagged inputs, outputs and noise terms through a suitable (typically polynomial) nonlinear function. The simpler NARX model is often preferred, although the absence of a disturbance model may result in bias problems. If a polynomial functional expansion is employed, a linear-in-the-parameters (or at least pseudo-linear, in the NARMAX case) structure results, which lends itself to easy interpretation and to the application of well-known identification algorithms of the Least Squares (LS) family. Another advantage of the polynomial NARX/NARMAX model class is the availability of nonlinear frequency analysis tools based on high order frequency response functions [Billings and Tsang, 1989, Peyton Jones and Billings, 1989].

However, a full polynomial NARMAX model can easily involve a great number of terms, thus exceeding acceptable model complexity, not to mention the numerical ill-conditioning problems that may arise due to over-parameterization [Ljung, 1999, Aguirre and Billings, 1995b]. Structure selection is thus a critical issue in

NARX/NARMAX identification, and several approaches have been proposed in the literature. Iterative algorithms are typically used to build up the model using regression and orthogonalization techniques [Haber and Unbehauen, 1990]. An example of such methods is the forward-regression orthogonal estimator (FROE) [Korenberg et al., 1988, Billings et al., 1989]. Other works suggest the use of pruning techniques to reduce model complexity and of simulation error minimization criteria to improve model robustness [Piroddi and Spinelli, 2001, 2003]. The interested reader is also directed to Mendes and Billings [2001], Kukreja et al. [2006], Peng et al. [2006] for some recently proposed alternative approaches.

In any case, the dimension of the set of candidate regressors clearly impacts on the computational load of the selection procedure, so that one has to trade model flexibility for algorithm efficiency. Various heuristic techniques can be used to reduce such set. For example, Spinelli et al. [2006] proposes a cluster based technique with this objective. This work, in a different perspective, aims at concentrating the model accuracy on a specified frequency band, on the grounds that a simplified model structure may be capable of achieving sufficient 'local' accuracy and therefore the selection procedure can narrow the search over a restricted regressor set. Notice that this 'local' accuracy requirement may also be the result of control specifications oriented at operating in that frequency band in closed-loop [Ljung, 1999]. Model accuracy in a specific frequency band can be obtained by weighting the prediction error minimization cost function with a suitable filter

---

(error filtering). In the linear systems framework this is equivalent to simply pre-filtering the input/output data used for identification with the same filter. In nonlinear modeling, direct error filtering must be used instead [Spinelli et al., 2005]. This work addresses the problem of model structure selection with error filtering, suitably extending the FROE algorithm and discusses, by way of a simple example, the possible implications of using error filtering in nonlinear identification. In order to deal with frequency-related issues in the nonlinear modeling context, Generalized Frequency Response Functions (GFRFs) will be employed, which are based on the frequency domain Volterra series expansion [Schetzen, 1980].

2. POLYNOMIAL NARX MODELS

Consider the Nonlinear AutoRegressive model with eXogenous inputs (NARX) [Leontaritis and Billings, 1985]

$$y(k) = f\left(y(k-1), \ldots, y(k-n_y), u(k-1), \ldots, \right.$$
$$\left. u(k-n_u)\right)) + \xi(k), \quad (1)$$

where $y(\cdot)$, $u(\cdot)$, $\xi(\cdot)$ are the output, input and noise signals, respectively, and n_y, n_u are the associated input and output maximum lags. The sequence $\xi(\cdot)$ is assumed to be a white noise, while $f(\cdot)$ is a suitable nonlinear function. In polynomial NARX models, $f(\cdot)$ is assumed to be a polynomial function. As a result, model (1) can be reformulated as a linear regression:

$$y(k) = \varphi^T(k)\vartheta + \xi(k), \quad (2)$$

where the elements of vector $\varphi(k)$ are the regressors, that contain linear and nonlinear combinations of the delayed input and output signals, and ϑ is the parameter vector containing the coefficients of the polynomial expansion.

A classical algorithm for the identification of NARX (and NARMAX) models is the FROE [Billings et al., 1989], which iteratively increments the model structure until a specified prediction accuracy is obtained, starting from an empty model and adding a new regressor at each iteration. The Orthogonal Least Squares (OLS) approach is employed to decouple the estimation of new parameters from that of the parameters already included in the model. In this way, at each step the significance of each candidate regressor can be separately evaluated by computing the Error Reduction Ratio (ERR) criterion, which can be shown to be proportional to the improvement in the Mean Squared Prediction Error (MSPE) that the candidate m-th regressor would achieve if added to the model:

$$\text{ERR}_m = \frac{\text{MSPE}(\mathcal{M}_i) - \text{MSPE}(\mathcal{M}_{i+1})}{\frac{1}{N}\sum_{k=1}^{N} y^2(k)} \quad (3)$$

where $\mathcal{M}_i$ is the model obtained at the i-th iteration and $\mathcal{M}_{i+1}$ is the candidate model at the subsequent iteration, with the inclusion of the m-th regressor. At each iteration, the regressor with the highest ERR value is added to the model.

3. DATA PRE-FILTERING AND ERROR FILTERING IN NONLINEAR IDENTIFICATION

This Section briefly summarizes the analysis provided in Spinelli et al. [2005], concerning the differences between data pre-filtering and error filtering in nonlinear identification. The results are based on the Volterra series representation, which is also briefly recalled for clarity.

3.1 Volterra series representation of nonlinear systems

As linear systems may be formulated using a convolution integral, so nonlinear systems may be represented by means of the Volterra series expansion [Schetzen, 1980]

$$y(k) = \sum_{n=1}^{\infty} y_n(k),$$

$$y_n(k) = \sum_{k_1=0}^{\infty}\ldots\sum_{k_n=0}^{\infty} h_n(k_1, \ldots, k_n)\prod_{l=1}^{n} u(k-k_l),$$

where $h_n(k_1, \ldots, k_n)$ is the n-th order Volterra kernel of the nonlinear system.

An equivalent representation of the nonlinear system in the frequency domain may be obtained by an extended Fourier transform. More precisely, the output of the n-th order kernel can be expressed as

$$y_n(k) = \left(\frac{1}{2\pi}\right)^n \int \ldots \int_{-\pi}^{\pi} H_n(e^{j\omega_1}, \ldots, e^{j\omega_n})$$
$$U(\omega_1)\ldots U(\omega_n)e^{j(\omega_1+\ldots+\omega_n)k}d\omega_1\ldots d\omega_n$$

where $H_n(e^{j\omega_1}, \ldots, e^{j\omega_n})$ denotes the GFRF of order n, which is defined as the Fourier transform of the n-th order kernel [Schetzen, 1980]. The GFRFs of polynomial NARMAX models can be computed by means of the harmonic (or probing) method [Billings and Tsang, 1989, Peyton Jones and Billings, 1989].

3.2 Data pre-filtering vs. error filtering

In the following, we will assume that the identification data is generated by a 'true system' expressed as a Volterra series expansion

$$\mathcal{S} : y(k) = \sum_{n=1}^{\infty}\left(\frac{1}{2\pi}\right)^n \int \ldots \int_{-\pi}^{\pi} H_n^\circ(e^{j\omega_1}, \ldots, e^{j\omega_n})$$
$$\prod_{l=1}^{n} U(\omega_l)e^{j\omega_l k}d\omega_l + e(k), \quad (4)$$

where the noise source e is a zero mean, white gaussian noise with variance λ^2. Let also the model class $\mathcal{M}(\vartheta)$ be correspondingly represented as

$$\mathcal{M}(\vartheta) : y(k) = \sum_{n=1}^{\infty}\left(\frac{1}{2\pi}\right)^n \int \ldots \int_{-\pi}^{\pi} H_n(e^{j\omega_1}, \ldots, e^{j\omega_n}; \vartheta)$$
$$\prod_{l=1}^{n} U(\omega_l)e^{j\omega_l k}d\omega_l + \xi(k). \quad (5)$$

The corresponding optimal one step ahead predictor is given by

$$\hat{\mathcal{M}}(\vartheta) : \hat{y}(k|k-1; \vartheta) =$$
$$\sum_{n=1}^{\infty}\left(\frac{1}{2\pi}\right)^n \int \ldots \int_{-\pi}^{\pi} H_n(e^{j\omega_1}, \ldots, e^{j\omega_n}; \vartheta)\prod_{l=1}^{n} U(\omega_l)e^{j\omega_l k}d\omega_l$$

therefore, the prediction error can be expressed as a function of the differences between the actual and estimated GFRFs

$$\varepsilon(k;\vartheta) = y(k) - \hat{y}(k|k-1;\vartheta) =$$

$$\sum_{n=1}^{\infty} \left(\frac{1}{2\pi}\right)^n \int ... \int_{-\pi}^{\pi} \Delta H_n(e^{j\omega_1},...,e^{j\omega_n};\vartheta)$$

$$\prod_{l=1}^{n} U(\omega_l)e^{j\omega_l k}\mathrm{d}\omega_l + e(k) \quad (6)$$

where

$$\Delta H_n(e^{j\omega_1},...,e^{j\omega_n};\vartheta) =$$
$$H_n^{\circ}(e^{j\omega_1},...,e^{j\omega_n}) - H_n(e^{j\omega_1},...,e^{j\omega_n};\vartheta).$$

The filtered prediction error (see [Spinelli et al., 2005] for further details) can be written as

$$\varepsilon_L(k;\vartheta) = L(q)\varepsilon(k;\vartheta) =$$

$$\sum_{n=1}^{\infty} \left(\frac{1}{2\pi}\right)^n \int ... \int_{-\pi}^{\pi} L(e^{j\omega_1+...+j\omega_n})\Delta H_n(e^{j\omega_1},...,e^{j\omega_n};\vartheta)$$

$$\prod_{l=1}^{n} U(\omega_l)e^{j\omega_l k}\mathrm{d}\omega_l + L(q)e(k). \quad (7)$$

Assuming that the considered model class is such that $\Delta H_n(e^{j\omega_1},...,e^{j\omega_n};\vartheta^{\star})$ is identically zero for some $\vartheta^{\star}$, the estimate will be consistent, i.e., the minimization of the prediction error norm and of the filtered prediction error norm will lead to the same model, at least asymptotically, as in the linear case. However, if $\mathcal{S} \notin \mathcal{M}$, then the identified model will be biased and the filter $L(q)$ may be chosen to affect the frequency distribution of the estimation error. While in the linear case the error filtering effectively focuses the identification algorithm on the assigned frequency band, it can be shown that the weighting factor operating on each kernel of a nonlinear model is actually a function of the kernel order, so that different kernels are fitted accurately on different bands.

Not surprisingly, error filtering and data pre-filtering are not interchangeable in the nonlinear modeling context. The prediction error sequence obtained after data pre-filtering can be expressed as [Spinelli et al., 2005]

$$\varepsilon^{DF}(k;\vartheta) = L(q)y(k) - \hat{y}^{DF}(k|k-1;\vartheta) =$$

$$\sum_{n=1}^{\infty} \left(\frac{1}{2\pi}\right)^n \int ... \int_{-\pi}^{\pi} \Delta H_n^{DF}(e^{j\omega_1},...e^{j\omega_n};\vartheta)$$

$$\prod_{l=1}^{n} U(\omega_l)e^{j\omega_l k}\mathrm{d}\omega_l + L(q)e(k) \quad (8)$$

where

$$\Delta H_n^{DF}(e^{j\omega_1},...,e^{j\omega_n};\vartheta) =$$
$$L(e^{j(\omega_1+...+\omega_n)})H_n^{\circ}(e^{j\omega_1},...,e^{j\omega_n})$$
$$- L(e^{j\omega_1})...L(e^{j\omega_n})H_n(e^{j\omega_1},...,e^{j\omega_n};\vartheta).$$

Clearly, minimizing the error in this case would lead to biased solutions where each kernel would tend to a weighted version of the correct one

$$H_n(e^{j\omega_1},...,e^{j\omega_n};\vartheta^{\star}) =$$
$$\frac{L(e^{j(\omega_1+...+\omega_n)})}{L(e^{j\omega_1})...L(e^{j\omega_n})}H_n^{\circ}(e^{j\omega_1},...,e^{j\omega_n}).$$

Notice also that such weighting factor would depend both on the characteristics of the filter $L(q)$ and on the order of the considered kernel.

In view of the above discussion, data pre-filtering should be avoided in nonlinear model identification, whereas direct error filtering can still be applied to influence the accuracy bands of the kernels of the identified model.

4. NARX IDENTIFICATION WITH ERROR FILTERING

Recalling (2), we can express the filtered prediction error for NARX models as

$$\varepsilon_L(k;\vartheta) = L(q)y(k) - L(q)\hat{y}(k;\vartheta) = y_L(k) - \varphi_L^T(k)\vartheta, \quad (9)$$

where $y_L(\cdot)$ is the filtered output signal and $\varphi_L(\cdot)$ denotes the vector of filtered regressors. Assume that N data samples $\{u(k),y(k)\}_{k=1}^{N}$ are used for model identification. Then, the objective of the identification procedure is to find ϑ that minimizes the cost functional $J = \frac{1}{N}\sum_{k=1}^{N}\varepsilon_L^2(k;\vartheta)$.

In the general NARMAX case some regressors may depend on past *estimated* output $\hat{y}(k-\kappa;\vartheta)$ values, so that the minimization of J cannot be addressed with plain Least Squares, and an iterative algorithm must be employed. Most conveniently, in the NARX case all the regressors depend on past input $u(k-\kappa)$ and output $y(k-\kappa)$ values only, so that the filtered regressors can be computed at the onset of the minimization procedure. In other words, for polynomial NARX models, error filtering simply amounts to pre-filtering the output and the regressors constructed from the original input/output data and solving the linear regression:

$$y_L(k) = \phi_L^T(k)\theta + \xi_L(k), \quad (10)$$

where subscript L denotes filtering through function $L(q)$.

Exploiting the least squares formulation (10), structure selection can also be enacted by adopting the OLS approach. A filtered version of the ERR criterion is actually employed in this case, denoted Filtered ERR (FERR) in the sequel. The FERR can be interpreted as the reduction in the Mean Squared Filtered Prediction Error (MSFPE) due to the inclusion of the m-th term as a fraction of the maximum MSFPE (variance of the filtered output)

$$[\text{FERR}]_m = \frac{\text{MSFPE}(\mathcal{M}_i) - \text{MSFPE}(\mathcal{M}_{i+1})}{\frac{1}{N}\sum_{k=1}^{N}y_L^2(k)}. \quad (11)$$

This selection criterion can be used to iteratively build up the model structure in a forward-regression fashion, similarly to the FROE (i.e., starting from an empty model and adding at each step the regressor with the highest value of the $[\text{FERR}]_m$) or in a mixed forward- and backward regression type of procedure. A termination condition can be formulated as the obtainment of a sufficient accuracy threshold, such as $1 - \sum_{i=1}^{m}[\text{FERR}]_i < \delta$. Notice that $1 - \sum_{i=1}^{m}[\text{FERR}]_i$ is precisely the portion of the unexplained variance of the filtered output.

In the following, we will denote EF-FROE the extension of the FROE algorithm that includes error filtering. As already mentioned, for NARX models the EF-FROE is equivalent to the FROE applied on the filtered regressors instead of the original ones. The reader is referred to Billings et al. [1989] for a description of the implementation of the FROE algorithm in explicit form, which may be easily adapted to the EF-FROE as well.

5. IDENTIFICATION OF UNDER-PARAMETERIZED MODELS USING ERROR FILTERING

This Section illustrates the differences between error filtering and data pre-filtering approaches for nonlinear model structure identification on a simple simulation example. The aim is to show that the error filtering approach is actually capable of achieving at least 'local' accuracy when the model structure is under-parameterized, where approaches based on data pre-filtering or 'global' criteria may fail. Notice that in realistic experimental settings model under-parameterization is the norm.

Consider as system generating the identification data the following (noiseless) NARX system

$$\mathcal{S} : y(k) = a_1^o y(k-1) + a_2^o y(k-2) + b_1^o u(k-1) + \\ + b_2^o u(k-2) + c_{uu,1}^o u^2(k-1) + c_{uu,2}^o u^2(k-2) \quad (12)$$

where $a_1^o = 1.09$, $a_2^o = -0.099$, $b_1^o = 1$, $b_2^o = -0.9$, $c_{uu,1}^o = 0.15$, $c_{uu,2}^o = 0.35$, and the input signal is assumed to be a non-zero mean gaussian white noise, $u(\cdot) \sim WN(0.01, 0.01)$. A data set of 5000 samples is generated for identification purposes. Suppose that, in order to ease the computational burden of the FROE, we want to restrict the set of candidate regressors to $[y(k-1)\ u(k-1)\ u^2(k-1)\ u(k-1)y(k-1)\ y^2(k-1)]$. Clearly, there is little chance to obtain global model accuracy given the lack of some 2nd order terms, but local model accuracy can still be achieved in specific frequency bands. For this purpose, a second order low-pass digital Butterworth filter with bandwidth $[0, 0.05]$ and a high-pass filter with bandwidth $[0.01, 1]$ have been considered in the following.

Model accuracy has been evaluated on both the 1st and 2nd order GFRFs, computed as follows. The 1st order frequency response function $H_1(\omega)$ coincides with the ordinary response function of the linear part of the system (obtained by canceling out all terms of order greater than 1). The 2nd order frequency response function $H_2(\omega_1, \omega_2)$ is obtained by means of the harmonic or probing method [Billings and Tsang, 1989]. More precisely, consider the full model structure implied by the assigned set of candidate regressors

$$\mathcal{M}(\vartheta) : y(k) = ay(k-1) + bu(k-1) + c_{yy}y^2(k-1) + \\ + c_{yu}y(k-1)u(k-1) + c_{uu}u^2(k-1), \quad (13)$$

and let the input signal be defined as the sum of two harmonic functions at different frequencies

$$u(k) = e^{j\omega_1 k} + e^{j\omega_2 k}. \quad (14)$$

Omitting the terms of order higher than 2, which are inessential in the computation, the corresponding output has the form

$$y(k) = H_1(\omega_1)e^{j\omega_1 k} + H_1(\omega_2)e^{j\omega_2 k} + H_2(\omega_1, \omega_1)e^{j2\omega_1 k} \\ + 2H_2(\omega_1, \omega_2)e^{j(\omega_1 k + \omega_2 k)} + H_2(\omega_2, \omega_2)e^{j2\omega_2 k}. \quad (15)$$

Substituting $u(k)$ and $y(k)$ in equation (13), and equating only the terms in $e^{j(\omega_1 k + \omega_2 k)}$, one obtains

$$2H_2(\omega_1, \omega_2) = 2aH_2(\omega_1, \omega_2)e^{-j(\omega_1 + \omega_2)} \\ + 2c_{yy}H_1(\omega_1)H_1(\omega_2)e^{-j(\omega_1 + \omega_2)} + c_{yu}(H_1(\omega_1) \\ + H_1(\omega_2))e^{-j(\omega_1 + \omega_2)} + 2c_{uu}e^{-j(\omega_1 + \omega_2)}, \quad (16)$$

which in turn gives the following 2nd order frequency function

$$H_2(\omega_1, \omega_2) = [c_{yy}H_1(\omega_1)H_1(\omega_2) + 0.5c_{yu}(H_1(\omega_1) \\ + H_1(\omega_2)) + c_{uu}] \times \frac{e^{-j(\omega_1 + \omega_2)}}{1 - ae^{-j(\omega_1 + \omega_2)}}. \quad (17)$$

The 2nd order frequency response function of system $\mathcal{S}$ can be obtained with analogous reasoning

$$H_2^o(\omega_1, \omega_2) = \frac{c_{uu,1}^o e^{-j(\omega_1 + \omega_2)} + c_{uu,2}^o e^{-j2(\omega_1 + \omega_2)}}{1 - a_1^o e^{-j(\omega_1 + \omega_2)} - a_2^o e^{-j2(\omega_1 + \omega_2)}}. \quad (18)$$

For ease of graphical representation, the latter two functions will only be compared on the line $\omega_1 = \omega_2$.

The EF-FROE method has been applied to the given data set, using both the defined filters. The estimated models (and the corresponding performance indexes) are reported in Table 1. Notice that a 3-parameter model was selected in the low-pass filtering case, while all linear and 2 quadratic terms are included in the model obtained with high-pass filtering. Comparing the performance indexes, it appears - not surprisingly - that the prediction task is much more difficult in the high frequency range, while the simulation performance turns out to be comparable. The plain FROE has also been employed on the same data set and the estimated model reported in Table 1 for reference. Interestingly enough, the model structure is different from both the previously identified ones.

Figures 1-2 explain more clearly the performance differences in the three cases by comparing the true and estimated frequency response functions of 1st and 2nd order. Apparently, the FROE cannot manage to model all the dynamics of the system throughout the whole frequency range and performs a sort of average, resulting in significant bias in both frequency response functions. On the contrary, the EF-FROE with low-pass filtering achieves a remarkable accuracy in the low frequency range, as can be seen by inspecting the magnitude and phase plots for both frequency responses: the estimated 1st order response matches the true one - both in magnitude and phase - up to 10^{-2} rad/s, while the estimated 2nd order response provides a very satisfactory fit of the true one even up to 2×10^{-1} rad/s. The performance of the high-pass filtering version of the algorithm is clearly affected by the lack of information about the system gain conveyed by the filtered regressors. Still, the estimated model provides a very good fit of the high-frequency portion of the 1st order response. As far as the 2nd order response is concerned, the magnitude is captured less accurately by the EF-FROE than the FROE, but the phase is matched almost perfectly, while the model estimated by the FROE provides both quantitatively and qualitatively incorrect information. This seems to indicate that the EF-FROE is more successful in capturing the actual high frequency singularities of the 2nd order response.

Notice finally that if the y^2 term is not included in the candidate regressor set, on the grounds that both the system and the candidate model structure should have the same number of fixed-points (one, in this case), the model performance degrades significantly regardless of which algorithm is employed, since the model structure is not capable of compensating for the missing quadratic terms in the system.

Table 1. Estimated parameters depending on the structure selection method.

Method	a	b	c_{yy}	c_{yu}	c_{uu}	MSPE	MSSE
True terms	1.09,-0.099	1,-0.9	-	-	0.15,0.35	-	-
FROE	1.3945	0.98226	-0.6043	-	0.27308	3.2456E-003	∞
EF-FROE (LP)	0.98866	0.12676	-	-	0.61873	1.1076E-005 (*)	7.6660E-003
EF-FROE (HP)	0.33577	0.93007	-0.013074	0.067878	-	1.1205E-003 (*)	4.2160E-001
DF-FROE (LP)	0.99791	0.11781	-	-	-	1.5935E-005 (**)	3.6666E-002 (**)
DF-FROE (HP)	0.31885	0.97556	0.23063	0.07.2715	-	1.1010E-003 (**)	2.0925E-003 (**)
SEMP	0.958	-	-	-	1.7262	1.5067E-002	8.2971E-002

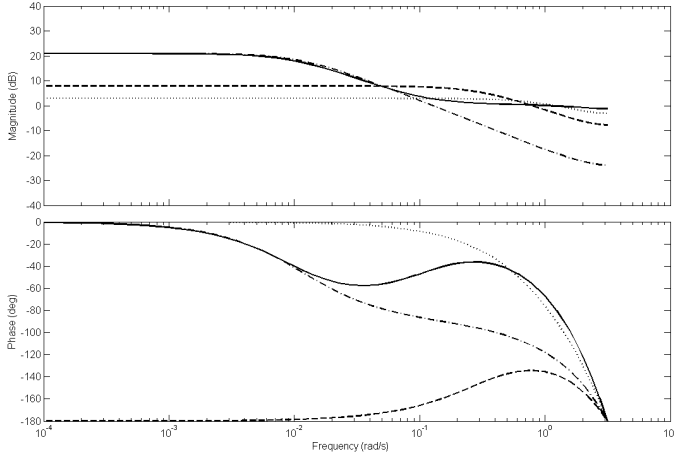(*) Criterion computed on the filtered error. (**) Criterion computed on the filtered data.

Fig. 1. Estimation of the 1st order GFRF: true function (solid line), FROE (dashed line), EF-FROE with low- and high-pass filtering (dash-dot and dotted line, respectively).

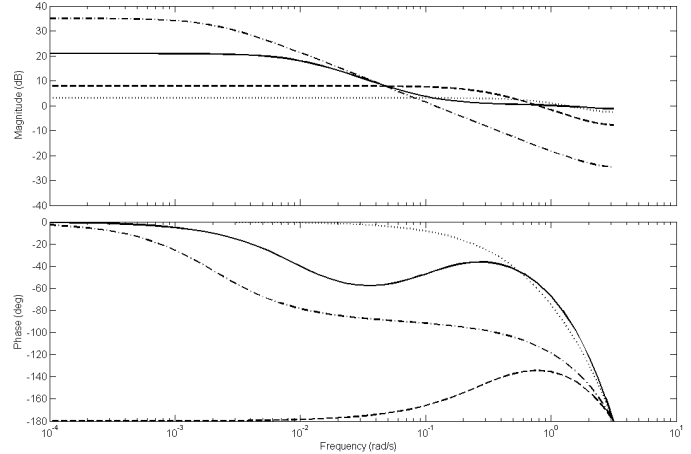
Fig. 3. Estimation of the 1st order GFRF: true function (solid line), FROE (dashed line), DF-FROE with low- and high-pass data pre-filtering (dash-dot and dotted line, respectively).
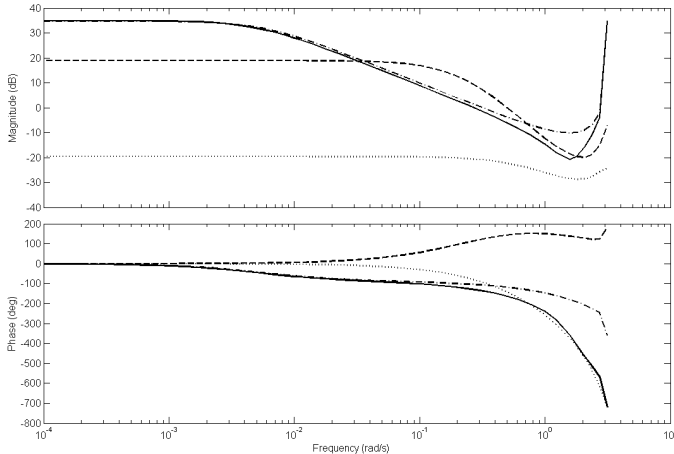

Fig. 2. Estimation of the 2nd order GFRF: true function (solid line), FROE (dashed line), EF-FROE with low- and high-pass filtering (dash-dot and dotted line, respectively).
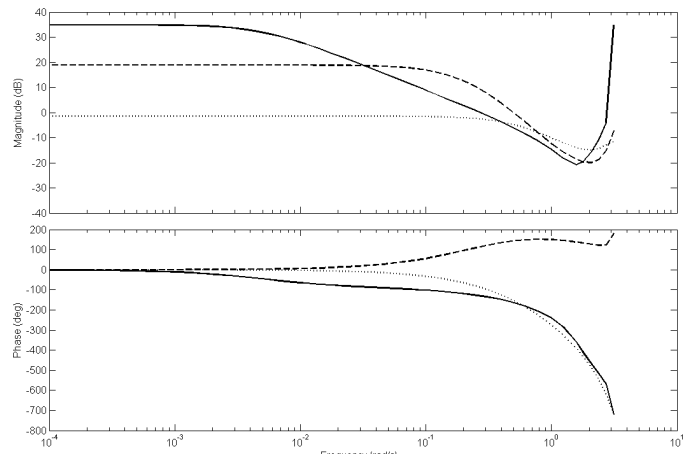

Fig. 4. Estimation of the 2nd order GFRF: true function (solid line), FROE (dashed line), DF-FROE with high-pass data pre-filtering (dotted line).

The inadequacy of the conventional data pre-filtering practice in combination with the FROE (briefly denoted DF-FROE for Data Filtering FROE), is witnessed by Figures 3-4, particularly in the low-pass filtering case. Notice that Figure 4 reports only the high-pass data pre-filtering case, due to the purely linear structure of the identified model (see Table 1).

Finally, Figure 5 shows the performance obtained on the same example with the SEMP (Simulation Error Minimization with Pruning) algorithm [Piroddi and Spinelli,

2001]. Only the 2nd order frequency response function is here reported since the obtained model structure does not allow the computation of the 1st order one (see Table 1). The SEMP is a computationally much more demanding algorithm with respect to the FROE, but usually guarantees a more robust and parsimonious modeling, thanks to the extended prediction horizon inherent in the simulation approach. Apparently, a much better performance with respect to that of the plain FROE is achieved, albeit with only two parameters. However, the EF-FROE in the low-pass filtering case obtains an even better simulation model,
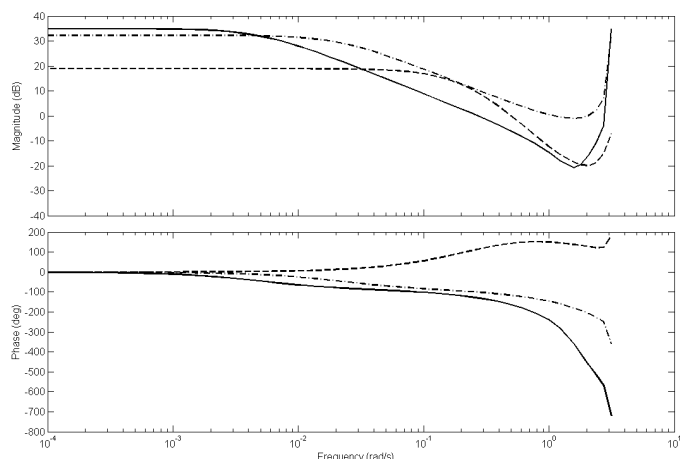
Fig. 5. Estimation of the 2nd order GFRF: true function (solid line), FROE (dashed line), SEMP (dash-dot line).

which is also the most similar to the true one, since only terms of the correct clusters are selected and the estimated parameters approach the cluster coefficients (a cluster groups all terms with the same type of nonlinearity, see [Aguirre and Billings, 1995a] for the related definitions). In other words, the EF-FROE employs more efficiently the under-parameterized model structure thanks to the frequency range restriction. This implies that low frequency error filtering, by concentrating on the dominant low frequency dynamics, may also provide an efficient method for deriving simulation models.

6. CONCLUDING REMARKS

A forward-regression orthogonal estimator for the identification of polynomial NARX models has been developed that incorporates error filtering. Similarly to data pre-filtering in the linear framework, error filtering may be employed as a means to weigh the error function of the identification algorithm in order to focus the model accuracy on a specified frequency band. In addition, thanks to this selectivity in the frequency domain, the algorithm can provide satisfactory results even with an under-parameterized model structure. This can be employed either to increase accuracy (in a selected frequency range) given the candidate regressor set, or to obtain a comparable accuracy (again, in a selected frequency range) with a reduced candidate regressor set. This second usage seems particularly promising in view of the computational load inherent in model structure selection algorithms, which increases with the size of the candidate regressor set. Finally, the low frequency error filtering approach can also be used to derive simulation models, since it disregards high frequency non-dominant dynamics. A simple simulation example has been discussed to illustrate the proposed approach.

REFERENCES

L.A. Aguirre and S.A. Billings. Improved structure selection for nonlinear models based on term clustering. *International Journal of Control*, 62:569–587, 1995a.

L.A. Aguirre and S.A. Billings. Dynamical effects of overparametrization in nonlinear models. *Physica D*, 80:26–40, 1995b.

S.A. Billings and K.M. Tsang. Spectral analysis for nonlinear systems, part I: Parametric nonlinear spectral analysis. *Mechanical Systems and Signal Processing*, 3 (4):319–339, 1989.

S.A. Billings, S. Chen, and M.J. Korenberg. Identification of MIMO nonlinear systems using a forward-regression orthogonal estimator. *International Journal of Control*, 49:2157–2189, 1989.

R. Haber and H. Unbehauen. Structure Identification of Nonlinear Dynamic Systems - A Survey on Input/Output Approaches. *Automatica*, 26(4):651–677, 1990.

M. Korenberg, S.A. Billings, Y.P. Liu, and P.J. McIlroy. Orthogonal parameter estimation algorithm for nonlinear stochastic systems. *International Journal of Control*, 48:193–210, 1988.

S.L. Kukreja, J. Löfbergand, and M.J. Brenner. A least absolute shrinkage and selection operator (lasso) for nonlinear system identification. In 14th *IFAC Symposium on System Identification*, pages 814–819, Newcastle, Australia, 2006.

I.J. Leontaritis and S.A. Billings. Input-output parametric models for non-linear systems - Part II: stochastic nonlinear systems. *International Journal of Control*, 41: 329–344, 1985.

L. Ljung. *System Identification: Theory for the User*. Prentice-Hall, Englewood Cliffs (NJ), USA, 1999.

E.M.A.M. Mendes and S.A. Billings. An alternative solution to the model structure selection problem. *IEEE Transactions on Systems, Man, and Cybernetics, part A: Systems and Humans*, 31(6):597–608, 2001.

J.X. Peng, K. Li, and E.W. Bai. A two-stage algorithm for identification of nonlinear dynamic systems. In 14th *IFAC Symposium on System Identification*, pages 570–575, Newcastle, Australia, 2006.

J.C. Peyton Jones and S.A. Billings. Recursive algorithm for computing the frequency response of a class of nonlinear difference equation model. *International Journal of Control*, 50:1925–1940, 1989.

L. Piroddi and W. Spinelli. An identification algorithm for polynomial NARX models based on simulation error minimization. *International Journal of Control*, 76(17): 283–295, 2001.

L. Piroddi and W. Spinelli. A pruning method for the identification of polynomial NARMAX models. In 13th *IFAC Symposium on System Identification*, pages 1108–1113, Rotterdam, Netherlands, 2003.

M. Schetzen. *The Volterra and Wiener Theories of Nonlinear Systems*. Wiley, New York, USA, 1980.

W. Spinelli, L. Piroddi, and M. Lovera. On the role of pre-filtering in nonlinear system identification. *IEEE Transactions on Automatic Control*, 50(10):1597–1602, 2005.

W. Spinelli, L. Piroddi, and M. Lovera. A two-stage algorithm for structure identification of polynomial NARX models. In *American Control Conference*, pages 2387–2392, Minneapolis (Minnesota), USA, 2006.