ON THE ROLE OF PRE-FILTERING IN NONLINEAR SYSTEM IDENTIFICATION

W. Spinelli, L. Piroddi, M. Lovera

Dipartimento di Elettronica e Informazione,
Politecnico di Milano
P.za Leonardo da Vinci 32, 20133 Milano, Italy
Phone: + 39 02 23993592; Fax: +39 0223993412
Email: {wspinell, piroddi, lovera}@elet.polimi.it

Abstract: Data pre-filtering is often used in linear identification to increase model accuracy in a specified frequency band, as pre-filtering is equivalent to a frequency weighting on the prediction error function. However, this interpretation applies only to a strictly linear setting of the identification problem. In this paper, the role of data and error pre-filtering in nonlinear system identification is analyzed and a frequency domain interpretation is provided, based on the use of Volterra series representations for the true system and the considered model structure. Simulation results illustrate the conclusions of the analysis.

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

1. INTRODUCTION

An increasing attention is being dedicated in the system identification literature to the problem of identifying suitable dynamic models for nonlinear systems: many different model classes and algorithms have been proposed in recent years (Sjoberg *et al.*, 1995; Piroddi and Spinelli, 2003; Previdi and Lovera, 2003; Schoukens *et al.*, 2003; Murray-Smith and Johansen, 1997), some of which are devised as extensions of well-known linear models and procedures to the nonlinear case. However, there are some aspects of linear identification that have become common practice also in nonlinear identification, but that should be carefully reexamined to be correctly applied. One of these is data pre-filtering, which is systematically employed in linear identification, mostly for anti-aliasing reasons and to reduce the effect of high frequency disturbances in order to increase the signal-to-noise ratio. Moreover, pre-filtering may be used to improve model accuracy in a specified frequency band, *e.g.*, as a result of control specifications which narrow the frequency region over which accurate models are actually needed (Ljung, 1999). Typically, input and output data are filtered with suitable low-pass or band-pass filters and the filtered data are used for identification: in the linear case this is equivalent to a filtering of the prediction error which determines a weight on the cost function in prediction error estimation, and is thus one of the most important design variables for selectively emphasizing the fit of the linear model.

In principle, the same considerations on imposing a frequency weighting on the cost function may be extended also to the nonlinear case. However, it will be shown that a naive data pre-filtering may yield unwanted results (parameter estimation is biased even in the ideal case where the true system belongs to the considered model set), while direct error filtering can be effective, though *ad hoc* parameter estimation algorithms have to be devised.

To extend the interpretation of pre-filtering to the nonlinear case, a nonlinear frequency domain analy-

sis is necessary: to this end the well-known Volterra series representation of nonlinear systems (Schetzen, 1980) will be used, for which frequency analysis tools have been developed, based on the Generalized Frequency Response Functions (GFRFs). For identification purposes, the class of polynomial Nonlinear Auto-Regressive Moving Average with eXogenous inputs (NARMAX) models (Leontaritis and Billings, 1985) will be used instead. These models are linear-in-the-parameters, so that straightforward parameter estimation methods can be employed for model identification (Billings *et al.*, 1989). Also, procedures are available for the computation of the Volterra kernels of these nonlinear models (Billings and Tsang, 1989).

2. PRE-FILTERING IN LINEAR IDENTIFICATION

This Section briefly summarizes the role of pre-filtering in linear identification, with specific reference to Output Error (OE) models (see (Ljung, 1999) for a more general treatment of this topic).

Consider a linear identification problem where the input-output data is generated by the 'true system'

$$\mathcal{S} \; : \; y(k) = H^o(q)u(k) + e(k) \tag{1}$$

where the noise source e is a zero mean, white gaussian noise with variance λ^2 and $H^o(q)$ is a linear transfer function. Given a model class $\mathcal{M}(\vartheta)$ parameterized by the vector ϑ and the optimal one step ahead predictor

$$\mathcal{M}(\vartheta) \; : \; y(k) = H(q;\vartheta)u(k) + \xi(k) \tag{2}$$

$$\hat{\mathcal{M}}(\vartheta) \; : \; \hat{y}(k|k-1;\vartheta) = H(q;\vartheta)u(k) \tag{3}$$

the prediction error is given by

$$\begin{aligned} \varepsilon(k;\vartheta) &= y(k) - \hat{y}(k|k-1;\vartheta) = \\ &= \Delta H(q;\vartheta)u(k) + e(k) \end{aligned} \tag{4}$$

where $\Delta H(q;\vartheta) = H^o(q) - H(q;\vartheta)$ is the estimation error on the system's transfer function.

The Prediction Error Minimization (PEM) approach to system identification is typically based on the minimization of a quadratic norm of the prediction error filtered through a stable linear filter $\varepsilon_L(k;\vartheta) = L(q)\varepsilon(k;\vartheta)$. It is well known that filtering the prediction error is equivalent to performing the identification on a pre-filtered input-output data set.

Filter $L(q)$ allows extra freedom during identification, that can be used to remove the effects of high frequency disturbances or slow drift terms not essential to the modelling problem. Moreover, some properties of the model can be enforced or suppressed with a proper choice of the filter. Writing the prediction error with the input-dependent part of the filtered error expressed as an inverse Fourier transform gives

$$\begin{aligned} \varepsilon_L(k;\vartheta) = \frac{1}{2\pi}\int_{-\pi}^{\pi} & L(e^{j\omega})\Delta H(e^{j\omega};\vartheta) \\ & U(\omega)e^{j\omega k}\mathrm{d}\omega + L(q)e(k) \end{aligned} \tag{5}$$

In general, if the considered model class is such that $\Delta H(e^{j\omega};\vartheta^\star)$ is identically zero for some $\vartheta^\star$, *i.e.*, if $\mathcal{S} \in \mathcal{M}$, then the PEM estimate will be consistent. If, on the other hand, it is not possible to assume that $\mathcal{S} \in \mathcal{M}$, then the identified model will be biased and the filter $L(q)$ may be chosen to affect the bias distribution: for example, if $L(q)$ is a low-pass filter, estimation errors in the filter frequency band are relatively more weighted so that the prediction error identification algorithm will yield a more accurate model in the same band, with respect to the unfiltered case.

3. PRE-FILTERING IN NONLINEAR IDENTIFICATION

3.1 Volterra series representation of nonlinear systems

The Volterra series (Schetzen, 1980) generalizes the convolution integral expression to represent the input-output relationship for a discrete-time deterministic causal SISO nonlinear system. The output of the system is given by $y(k) = \sum_{n=1}^{\infty} y_n(k)$, where each signal $y_n(k)$ can be interpreted as the output of the n-th order subsystem defined by

$$y_n(k) = \sum_{k_1=0}^{\infty} ... \sum_{k_n=0}^{\infty} h_n(k_1,...,k_n) \prod_{l=1}^{n} u(k-k_l) \tag{6}$$

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

An equivalent representation of the nonlinear system in the frequency domain (Schetzen, 1980) may be obtained considering the *Generalized Frequency Response Function* (*GFRF*) of order n, which is defined as the Fourier transform of the n-th order kernel. Conversely, the inverse Fourier transform of the n-th order GFRF gives the n-th order kernel of the system. The output of the n-th order kernel as a function of the corresponding GFRF is given by

$$\begin{aligned} y_n(k) = \left(\frac{1}{2\pi}\right)^n \int ... \int_{-\pi}^{\pi} & H_n(e^{j\omega_1},...,e^{j\omega_n}) \\ & \prod_{l=1}^{n} U(\omega_l)e^{j\omega_l k}\mathrm{d}\omega_l. \end{aligned} \tag{7}$$

3.2 Formulation of the nonlinear identification problem

We will assume that the identification data is generated by a 'true system' $\mathcal{S}$ in (8) where the noise source e is a zero mean, white gaussian noise with variance λ^2, and we will consider a generic model class $\mathcal{M}(\vartheta)$ which, for analysis purposes only, will be represented in Volterra series form. The general expression based on Volterra kernels for a NOE model parameterized by the vector ϑ and the optimal one step ahead predictor are given by (9) and (10). Therefore, the prediction error for a NOE model can be expressed as a function of the differences between the actual and estimated GFRFs (11).

$$\mathcal{S} \; : \; y(k) = \sum_{n=1}^{\infty} \left(\frac{1}{2\pi}\right)^n \int ... \int_{-\pi}^{\pi} H_n^o(e^{j\omega_1},..., e^{j\omega_n}) \prod_{l=1}^{n} U(\omega_l) e^{j\omega_l k} \mathrm{d}\omega_l + e(k) \tag{8}$$

$$\mathcal{M}(\vartheta) \; : \; y(k) = \sum_{n=1}^{\infty} \left(\frac{1}{2\pi}\right)^n \int ... \int_{-\pi}^{\pi} H_n(e^{j\omega_1},..., e^{j\omega_n}; \vartheta) \prod_{l=1}^{n} U(\omega_l) e^{j\omega_l k} \mathrm{d}\omega_l + \xi(k) \tag{9}$$

$$\hat{\mathcal{M}}(\vartheta) \; : \; \hat{y}(k|k-1;\vartheta) = \sum_{n=1}^{\infty} \left(\frac{1}{2\pi}\right)^n \int ... \int_{-\pi}^{\pi} H_n(e^{j\omega_1},..., e^{j\omega_n}; \vartheta) \prod_{l=1}^{n} U(\omega_l) e^{j\omega_l k} \mathrm{d}\omega_l \tag{10}$$

$$\begin{aligned} \varepsilon(k;\vartheta) = y(k) - \hat{y}(k|k-1;\vartheta) &= \sum_{n=1}^{\infty} \left(\frac{1}{2\pi}\right)^n \int ... \int_{-\pi}^{\pi} \Delta H_n(e^{j\omega_1},..., e^{j\omega_n}; \vartheta) \prod_{l=1}^{n} U(\omega_l) e^{j\omega_l k} \mathrm{d}\omega_l + e(k) \\ \Delta H_n(e^{j\omega_1},..., e^{j\omega_n}; \vartheta) &= H_n^o(e^{j\omega_1},..., e^{j\omega_n}) - H_n(e^{j\omega_1},..., e^{j\omega_n}; \vartheta) \end{aligned} \tag{11}$$

$$\varepsilon_L(k;\vartheta) = L(q)\varepsilon(k;\vartheta) = \sum_{n=1}^{\infty} \left(\frac{1}{2\pi}\right)^n \int ... \int_{-\pi}^{\pi} L(e^{j\omega_1+...+j\omega_n}) \Delta H_n(e^{j\omega_1}, ..., e^{j\omega_n}; \vartheta) \prod_{l=1}^{n} U(\omega_l) e^{j\omega_l k} \mathrm{d}\omega_l + L(q)e(k) \tag{12}$$

$$\begin{aligned} \varepsilon^{DF}(k;\vartheta) = L(q)y(k) - \hat{y}^{DF}(k|k-1;\vartheta) &= \sum_{n=1}^{\infty} \left(\frac{1}{2\pi}\right)^n \int ... \int_{-\pi}^{\pi} \Delta H_n^{DF}(e^{j\omega_1},...e^{j\omega_n}; \vartheta) \prod_{l=1}^{n} U(\omega_l) e^{j\omega_l k} \mathrm{d}\omega_l + L(q)e(k) \\ \Delta H_n^{DF}(e^{j\omega_1}, ..., e^{j\omega_n}; \vartheta) &= L(e^{j(\omega_1+...+\omega_n)}) H_n^o(e^{j\omega_1},..., e^{j\omega_n}) - L(e^{j\omega_1})...L(e^{j\omega_n}) H_n(e^{j\omega_1},..., e^{j\omega_n}; \vartheta) \end{aligned} \tag{13}$$

3.3 Effects of pre-filtering

In the nonlinear case, error filtering and data pre-filtering are obviously not equivalent. The following two Lemmas (which can be readily derived from Theorems 3.2.1 (Composition Theorem) and 4.2.1 in (Boyd *et al.*, 1984)) describe the effect of linear filtering at the input or the output of a nonlinear system and will be useful later in the paper.

Lemma 1. Consider a series of a linear filter with transfer function $L(q)$ and a nonlinear system described in terms of a Volterra series with GFRFs $H_n(e^{j\omega_1}, ..., e^{j\omega_n})$. Then, the GFRFs of the overall system are $L(e^{j\omega_1})...L(e^{j\omega_n})H_n(e^{j\omega_1}, ..., e^{j\omega_n})$.

Lemma 2. Consider a series of a nonlinear system described in terms of a Volterra series with GFRFs $H_n(e^{j\omega_1}, ..., e^{j\omega_n})$ and a linear filter with transfer function $L(q)$. Then, the GFRFs of the overall system are $L(e^{j(\omega_1+...+\omega_n)})H_n(e^{j\omega_1}, ..., e^{j\omega_n})$.

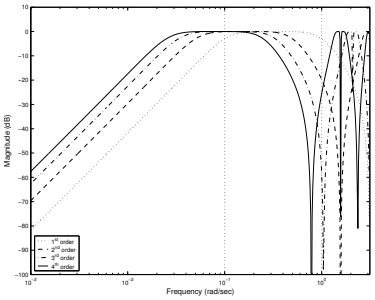
3.3.1. Error filtering Recalling (11) and applying Lemma 2, the filtered prediction error can be written as in (12). The above expression for the filtered prediction error lends itself to considerations not unlike the ones made for the linear case with respect to equation (5): if the considered model class is such that $\Delta H_n(e^{j\omega_1}, ..., e^{j\omega_n}; \vartheta^\star)$ is identically zero for some $\vartheta^\star$, then the estimate will be consistent, *i.e.*, the minimization of the prediction error norm and of the filtered prediction error norm lead to the same model, at least asymptotically, as in the linear case. If, on the other hand, it cannot be assumed that $\mathcal{S} \in \mathcal{M}$, then the identified model will be biased and the pre-filter $L(q)$ may be chosen to affect the frequency distribution of the estimation error. Note, however, that the frequency weightings $L(e^{j\omega_1+...+j\omega_n})$ acting on the kernel estimation errors of increasing order have significantly different shapes.

Further insight in the shape of the weighting factors can be obtained by inspection of their behaviour along the main diagonal in the frequency domain, *i.e.*, setting $\omega_l = \omega, \; l = 1, ..., n$. As an example, Figure 1(a) illustrates the Bode plot of the magnitude of the weighting factor $L(e^{jn\omega})$ for the case of a second order digital band-pass Butterworth filter with bandwidth $[0.1, \; 1]$. It is apparent from Figure 1(a) that, while in the linear case the effect of pre-filtering is to 'cut out' all the frequencies external to the bandwidth over which the fit is required, in the nonlinear case the bandwidth of the weighting factor is a function of the kernel order and the fit is also important in different bands, due to the nonlinear effect that scatters the spectrum.

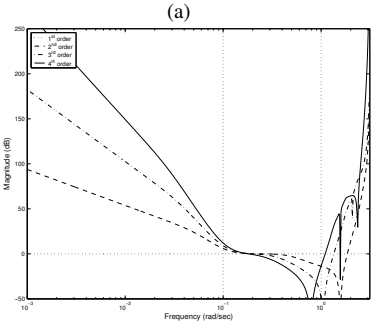
3.3.2. Data pre-filtering If we denote $\varepsilon^{DF}(k;\vartheta)$ the prediction error sequence obtained after filtering the identification data, we can write (13), in view of Lemmas 1-2. In this case, even assuming that the considered model structure is sufficiently complex so as to have a value $\vartheta^\star$ such that $\Delta H_n^{DF}(e^{j\omega_1}, ..., e^{j\omega_n}; \vartheta^\star)$ is identically zero for all n, the input dependent part of the prediction error will be (ideally) eliminated for a choice of $\vartheta^\star$ such that

$$\begin{aligned} &H_n(e^{j\omega_1}, ..., e^{j\omega_n}; \vartheta^\star) = \\ &\quad = \frac{L(e^{j(\omega_1+...+\omega_n)})}{L(e^{j\omega_1})...L(e^{j\omega_n})} H_n^o(e^{j\omega_1}, ..., e^{j\omega_n}). \end{aligned} \tag{14}$$

It is apparent from the above expression that, apart from the case $n = 1$, $H_n(\cdot;\vartheta^\star)$ will correspond to a weighted version of $H_n^o(\cdot)$, where the weighting factor depends both on the characteristics of filter $L(q)$ and on the order of the considered kernel. In particular, $H_n(\cdot;\vartheta^\star) \approx H_n^o(\cdot)$ in the frequency regions where



(a)



(b)

Fig. 1. Bode diagram of weighting factor for a second order Butterworth filter with bandwidth $[0.1,\ 1]$: (a) $\left|L(e^{jn\omega})\right|$, (b) $\left|L(e^{jn\omega})/L(e^{j\omega})^n\right|$.

$\frac{L(e^{j(\omega_1+\ldots+\omega_n)})}{L(e^{j\omega_1})\ldots L(e^{j\omega_n})} \approx 1$, while the approximation degrades where this ratio differs from 1 significantly.

This approximation depends on the shape of the filter and, as in the case of error filtering. Along the diagonal, (14) reduces to

$$H_n(e^{j\omega},...,e^{j\omega};\vartheta)=\frac{L(e^{jn\omega})}{L(e^{j\omega})^n}H_n^o(e^{j\omega},...,e^{j\omega}). \quad (15)$$

It is clear from expression (15) that for rational filters $L(q)$ with positive relative degree (such as, *e.g.*, conventional low-pass or band-pass filters) the frequency weighting factor $L(e^{jn\omega})/L(e^{j\omega})^n$ will be of high-pass type and will take abnormally high values outside the filter bandwidth for higher order kernels: this implies, in the first place, that the identification algorithm will try to fit a severely distorted kernel. More disturbingly, though this distortion takes place outside the filter bandwidth, in some cases the shape of the weighting factor also affects the model accuracy inside the filter bandwidth. In fact, the minimization procedure in the identification algorithm will be enticed to consider as numerically significant only the portion of the identification data with frequency content *outside* the filter bandwidth, thus producing an unwanted bias also inside the filter bandwidth. Notice, finally, that even in the ideal case when the system belongs to the model family, the presence of the weighting factor in (15) actually configures an under-parameterized identification problem.

More specifically, Figure 1(b) considers a second order digital band-pass Butterworth filter with bandwidth $[0.1,\ 1]$. In the optimal model, the accuracy of higher order kernels would be reasonable only within a fraction of the filter bandwidth, which gets smaller as the kernel order increases. Also, a severe distortion would result both outside and *inside* the filter bandwidth.

All these considerations clearly suggest that prefiltering data is not a good practice when a nonlinear model has to be identified: the resulting model will always be biased, even if $\mathcal{S} \in \mathcal{M}$. On the other hand, including error filtering in the identification algorithm could still be valuable, in order to "shape" the bias distribution of the identified models. Note also, that the above analysis holds regardless of the specific parameterization for the model class and is also therefore largely independent of the considered parameter estimation algorithm.

3.4 Error filtering for nonlinear polynomial models

The inclusion of an error filtering mechanism in a general algorithm for nonlinear model identification is quite straightforward for a model class which is linear in the parameter vector ϑ. Consider the general black-box nonlinear model structure (Sjoberg *et al.*, 1995)

$$\hat{\mathcal{M}}(\vartheta)\ :\ \hat{y}(k;\vartheta) = g(\phi(k),\vartheta), \quad (16)$$

where g is a nonlinear function parameterized in ϑ and $\phi(\cdot)$ is a vector of regressors, defined as past input $u(k-\kappa)$, output $y(k-\kappa)$ and estimated output $\hat{y}(k-\kappa;\vartheta)$ values. Such model structure includes NFIR, NARX and NOE models along with the classification of (Sjoberg *et al.*, 1995). Now assume that g depends linearly on ϑ, as happens *e.g.*, for polynomial input/output recursive black-box models

$$\hat{\mathcal{M}}(\vartheta)\ :\ \hat{y}(k;\vartheta) = \psi(k)^T\vartheta, \quad (17)$$

where the elements of $\psi(k)$ are linear and nonlinear monomials of $u(k-\kappa)$, $y(k-\kappa)$ and $\hat{y}(k-\kappa;\vartheta)$. For such models, the filtered prediction error results

$$\begin{aligned}\varepsilon_L(k;\vartheta) &= L(q)y(k) - L(q)\hat{y}(k;\vartheta) = \\ &= y_L(k) - \psi_L(k)^T\vartheta,\end{aligned} \quad (18)$$

where $\psi_L(\cdot)$ denotes the vector of filtered monomials. Filtered versions of monomials including $u(k-\kappa)$ and $y(k-\kappa)$ terms only can be computed at the onset of the minimization procedure, while monomials containing also terms of the type $\hat{y}(k-\kappa;\vartheta)$ must be filtered at each algorithm iteration after $\hat{y}(\cdot;\vartheta)$ has been recalculated given the current parameterization ϑ.

4. SIMULATION EXAMPLES

This Section reports two simulation examples which illustrate the results discussed in the paper. For simplicity reasons, only NFIR and NARX models will be considered, both when the system belongs to the model family and when the model family is under-parameterized (see (Spinelli *et al.*, 2004) for additional details).

Table 1. Estimated parameters for the model in Example 1

Regressors	true	LS	DFLS	EFLS
$u(k-1)$	1.0	1.00278	1.01590	1.00426
$u(k-2)$	0.7	0.70071	0.74577	0.70254
$u(k-1)^2$	2.0	1.99979	34.01886	1.99569
$u(k-2)^2$	1.0	0.99996	33.43990	1.00531
$u(k-1)u(k-2)$	-0.5	-0.49918	-61.36556	-0.50142

Example 1. Consider the system

$$\begin{aligned}\mathcal{S}_1 : y(k) &= u(k-1) + 0.7u(k-2) + \\ &+ 2u(k-1)^2 + u(k-2)^2 \\ &- 0.5u(k-1)u(k-2) + e(k)\end{aligned} \tag{19}$$

and the NFIR model structure

$$\begin{aligned}\mathcal{M}_1(\vartheta) : y(k) &= a_1 u(k-1) + a_2 u(k-2) + \\ &+ b_1 u(k-1)^2 + b_2 u(k-2)^2 + \\ &+ b_3 u(k-1)u(k-2) + \xi(k),\end{aligned} \tag{20}$$

where clearly $\mathcal{S}_1 \in \mathcal{M}_1$. The identification is performed over a data set of 10000 samples generated with both input and noise signals, $u(\cdot)$ and $e(\cdot)$, selected as white gaussian noises, with variances 1 and λ^2 respectively, the latter value being chosen in order to obtain $SNR = 20$ dB. The parameters of the model have been estimated using basic least squares (LS), as well as least squares with data pre-filtering (DFLS) and error filtering (EFLS). In the last two cases a second order low-pass digital Butterworth filter with bandwidth $[0,\ 0.5]$ has been considered. Figures 2-3 show the Bode diagrams of the identified GFRFs and the corresponding bias distribution, respectively. While with both LS and EFLS the estimated model converges to the true system, the DFLS algorithm estimates the second order frequency response badly in all the frequency range (notice, in particular, the increasing amplitude at high frequency, in agreement with equation (15)). Figure 3 shows that LS and EFLS obtain comparable accuracies, whereas DFLS has significantly more bias in all the frequency range. Evaluation in simulation gives yet another confirmation that the model estimated with DFLS is inaccurate. This is reflected also by the numerical values of the estimated parameters (see Table 1). While the estimates computed using LS and EFLS are consistent, DFLS leads to abnormally high values for the parameters associated with the quadratic regressors.

The correct estimation of the first order GFRF is a consequence of the fact that for a second order nonlinear system with normally distributed white gaussian input, the spectrum of the error can be written as the sum of two independent contributions, related to the first order and second order bias, respectively, while the cross spectrum is asymptotically zero. In the general case, however, the estimation of the various GFRFs is not independent, since the quadratic prediction error norm also contains cross terms, which account for the interactions between different GFRFs.

Example 2. Consider the NARX system

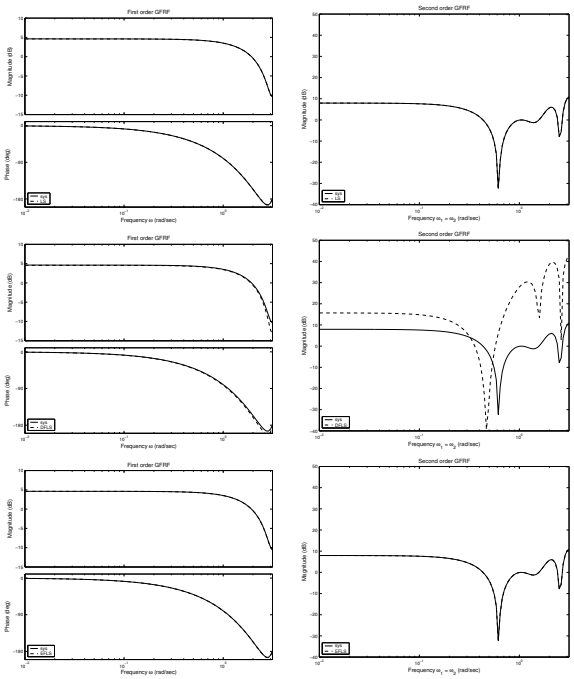


Fig. 2. Actual (solid) and estimated (dashed) GFRFs for Example 1: LS (upper), DFLS (middle), EFLS (lower).

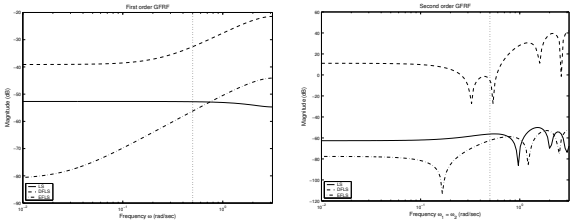


Fig. 3. Bias distribution for the estimated GFRFs of Example 1: LS (solid), DFLS (dashed), EFLS (dash-dot).

$$\begin{aligned}\mathcal{S}_2 : y(k) &= 0.5y(k-1) + 0.75u(k-1) + \\ &+ 0.25u(k-2) + 0.15u(k-1)^2 + \\ &+ 0.35u(k-2)^2 + e(k)\end{aligned} \tag{21}$$

and the (under-parameterized) NARX model structure

$$\begin{aligned}\mathcal{M}_2(\vartheta) : y(k) &= ay(k-1) + bu(k-1) + \\ &+ cu(k-1)^2 + \xi(k)\end{aligned} \tag{22}$$

so that $\mathcal{S}_2 \notin \mathcal{M}_2$. The identification is performed over a data set of 1000 samples generated with white gaussian input $u(\cdot)$ (variance 1) and noise $e(\cdot)$ (variance λ^2, chosen in order to obtain $SNR = 20$ dB), using all the considered identification methods. The filter used in DFLS and EFLS is the same used in the previous example. The parameters of the identified models are reported in Table 2, while the frequency domain characteristics of the identified models obtained using LS, DFLS and EFLS are compared in Figures 4-5.

Consistently with Example 1, (low-pass) error filtering can be used to reduce the bias both in the first and second order kernel estimates in the frequency band of interest. Unlike the previous example, however, the

Table 2. Estimated parameters for the models in Example 2

Regressors	LS	DFLS	EFLS
$y(k-1)$	0.71893	0.85327	0.63673
$u(k-1)$	0.74604	0.50235	0.79515
$u(k-1)^2$	0.19342	0.48249	0.37221

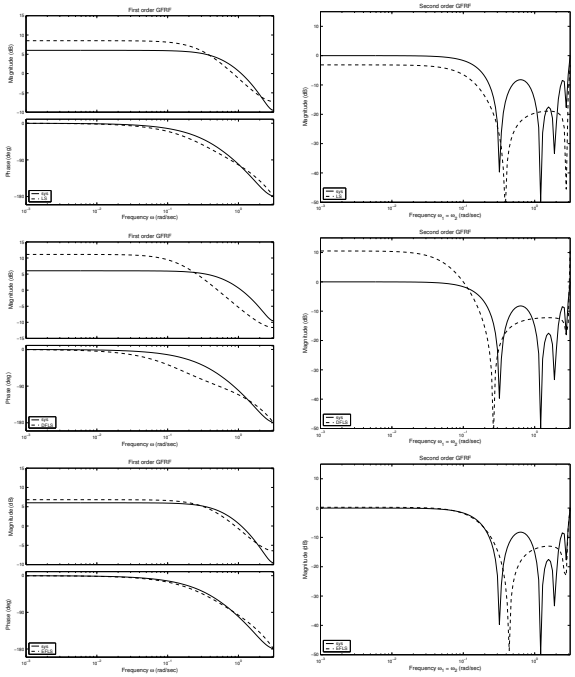


Fig. 4. Estimated GFRFs for Example 2: LS (upper), DFLS (middle), EFLS (lower).

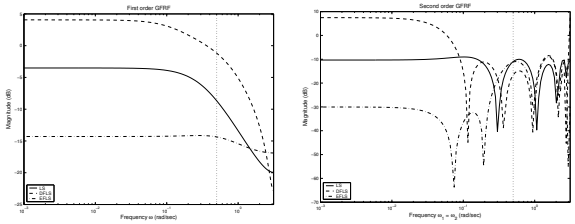


Fig. 5. Bias distribution for the GFRFs of Example 2.

estimate of the second order kernel obtained with data pre-filtering does not exhibit the high frequency amplification which was apparent in the previous cases. This is due to the presence of the autoregressive term in the model class $\mathcal{M}_2$, which structurally prevents this kind of behavior. However, the undesirable effect of the weighting function (15) is still visible in terms of the increased low frequency bias in the kernel estimates obtained via DFLS with respect to the corresponding LS estimates.

5. CONCLUDING REMARKS

The role of pre-filtering in nonlinear system identification has been investigated within the framework of Volterra series representation of nonlinear systems. The classical results on the frequency domain interpretation of prediction error filtering available in the linear system identification literature have been extended to the nonlinear case and the pitfalls associated with the naive application to nonlinear problems of the practice of data pre-filtering have been analyzed and illustrated via simulation examples of NFIR and NARX model. Future developments include the investigation of nonlinear data and error pre-filters.

ACKNOWLEDGEMENTS

Paper supported by the MIUR project "Innovative Techniques for Identification and Adaptive Control of Industrial Systems".

REFERENCES

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

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

Boyd, S., L.O. Chua and C.A. Desoer (1984). Analytical foundations of Volterra series. *IMA Journal of Mathematical Control and Information* **1**, 243–282.

Leontaritis, I.J. and S.A. Billings (1985). Input output parametric models for non linear systems. *International Journal of Control* **41**, 303–344.

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

Murray-Smith, R. and T. A. Johansen (1997). *Multiple Model Approaches to Modeling and Control*. Taylor and Francis, London.

Piroddi, L. and W. Spinelli (2003). An identification algorithm for polynomial NARX models based on simulation error minimization. *International Journal of Control* **76**(17), 283–295.

Previdi, F. and M. Lovera (2003). Identification of a class of nonlinear parametrically varying models. *International Journal of Adaptive Control and Signal Processing* **17**, 33–50.

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

Schoukens, J., J. G. Nemeth, P. Crama, Y. Rolain and R. Pintelon (2003). Nonlinear black-box modeling in system identification: a unified overview. *Automatica* **39**(7), 1267–1274.

Sjoberg, J., Q. Zhang, L. Ljung, A. Benveniste, B. Delyon, P. Glorennec, H. Hjalmarsson and A. Juditsky (1995). Nonlinear Black-box Modeling in System Identification: a Unified Overview. *Automatica* **31**, 1691–1724.

Spinelli, W., L. Piroddi and M. Lovera (2004). On the role of pre-filtering in nonlinear system identification. Technical Report 2004-20. Dipartimento di Elettronica e Informazione, Politecnico di Milano.