

Consistency Aspects of Wiener-Hammerstein Model Identification in Presence of Process Noise

Giuseppe Giordano and Jonas Sjöberg

Abstract—The Wiener-Hammerstein model is a block-oriented model consisting of two linear blocks and a static nonlinearity in the middle. Several identification approaches for this model structure rely on the fact that the best linear approximation of the system is a consistent estimate of the two linear parts, under the hypothesis of Gaussian excitation. But, these approaches do not consider the presence of other disturbance sources than measurement noise. In this paper we consider the presence of a disturbance entering before the nonlinearity (process noise) and we show that, also in this case, the best linear approximation is a consistent estimate of underlying linear dynamics. Furthermore, we analyse the impact of the process noise on the nonlinearity estimation, showing that a standard prediction error method approach can lead to biased results.

I. INTRODUCTION

Block-oriented models represent a common and efficient approach for nonlinear system identification. They are suitable for both actual description of some physical properties and often a good approximation of more general nonlinear behaviours. They consist of interconnection of linear time-invariant dynamic blocks and static nonlinear blocks.

One example, which we consider in this paper, is the Wiener-Hammerstein (W-H) model, consisting of two linear dynamic blocks and a static nonlinearity in the middle, see Figure 1. The signals $u(t)$ and $y(t)$ are, respectively, the measurable input and output of the system. The intermediate signal $x(t)$ is not available. In this paper we will consider the presence of the process noise $w(t)$, entering before the nonlinearity, and measurement noise $e(t)$. We will assume that these signals are independent and G_1 and G_2 are two stable, LTI systems.

In the literature, two main branches dealing with the identification of W-H systems can be found:

- approaches considering only measurement noises (input and/or output) affecting the system. For this branch, early results can be found in [1] and [2]. More recent works are [4], where the design of special excitation signals is discussed, and [13], where a nonparametric approach is introduced. In this area, several identification algorithms rely on the fact that, under some assumptions on the data, the best linear approximation (BLA) model is a consistent estimate of the concatenation of the two linear blocks [11], [5]. Hence, given the BLA, the remaining problem is to divide the dynamics into two

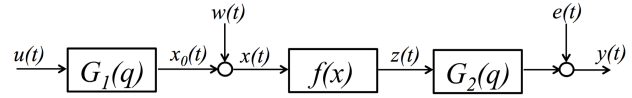


Fig. 1: The Wiener-Hammerstein model. The input $u(t)$ and the output $y(t)$ are measurable, but not the intermediate signal $x(t)$. Noise sources are represented by $w(t)$ and $e(t)$. G_1 , G_2 and f are, respectively, two linear systems and a static nonlinear function.

parts, and estimate the nonlinearity. This is done, for example, in [12], [17], [10]. In [15] an algorithm based on the best split of the BLA is proposed, while fractional approaches, where the position of poles and zeros from BLA is parametrized using fractional exponents, can be found in [16] (frequency-domain) and [6] (time-domain).

- the second branch consists of works considering process noise as well, but in this case the assumption is that right split of the linear parts is known. For example in [8] a Maximum Likelihood identification algorithm is implemented.

The main contribution of this paper is consistency results for approaches based on the BLA, in case the system is affected by process noise. In particular, after an introduction to the problem (Section II), we will show that the BLA still provides an unbiased estimate of the linear parts and, in principle, it can be used as an initial guess for the iterative search of the model parameters (Section III). Then, we will consider the problem of splitting the dynamics contained in the BLA and study how the process noise affects the estimation of the nonlinear function f (Section IV). Finally, simulation examples will be shown in order to confirm the obtained results (Section V).

II. PROBLEM SETTING

The Wiener-Hammerstein system, defined in Figure 1, can be described by the following equations:

$$y(t) = G_2(q, \beta)f(x(t), \gamma) + e(t) \quad (1a)$$

$$x(t) = x_0(t) + w(t) \quad (1b)$$

$$x_0(t) = G_1(q, \alpha)u(t) \quad (1c)$$

where G_1 and G_2 belongs to a parametrized model class such as FIR filters, rational transfer functions (OE models) or state space models. The nonlinear function f may be polynomial, spline, piecewise linear, neural network, in general a basis

G. Giordano and J. Sjöberg are with the Department of Signals and Systems, Chalmers University of Technology, 412 96 Gothenburg, Sweden giuseppe.giordano@chalmers.se, jonas.sjoberg@chalmers.se

function expansion. In this framework, the identification problem aims to estimate the parameters α , β and γ that best match the input/output behaviour described by the data.

Given a data set $\{u(t), y(t)\}_{t=1}^N$ of N input and output samples, a common approach for parameter estimation is to find the best fit of the model by minimizing the criterion

$$V_N(\alpha, \beta, \gamma) = \frac{1}{N} \sum_{t=1}^N (y(t) - G_2(q, \alpha) f(G_1(q, \beta) u(t), \gamma))^2. \quad (2)$$

In case the process noise $w(t)$ is zero, this criterion is the prediction error criterion (PEM). If the measurement noise is white and Gaussian, (2) is also the Maximum Likelihood criterion and the estimate is thus consistent, under the hypothesis of identifiability and persistent excitation, (see [9]).

If the process noise is not zero, it is still reasonable to use the error criterion (2). In [7], this criterion is used for Wiener system affected by process noise, and it is called approximative Prediction Error Method (approximative PEM), since the predictor used in the model will not lead to a criterion that can be written as a Maximum Likelihood function.

Independently on the presence of process noise, the predictor used in (2) is a nonlinear function of the parameters, hence, in general, the optimization problem is not convex. Therefore iterative search methods must be used in order to compute the estimate and generating a good initial guess is crucial in order to avoid local minima. As already mentioned in the introduction, if there is no process noise, the BLA represents a consistent estimate of the product of the two linear parts and it can be used to generate this initial guess. This is done in several identification algorithms, because the true linear systems G_1 and G_2 are actually described by one possible split of the BLA. Hence, the remaining question for approaches based on BLA splitting, is to find the right split and estimate the parameters of the nonlinearity.

In the next section we will show that, actually, also in presence of process noise, the BLA is a consistent estimate of the concatenation of the linear parts.

III. CONSISTENCY OF THE BLA

First, we define precisely the BLA. The best linear approximation of a time-invariant nonlinear system to a given class of stationary input signals \mathcal{U} , containing sequences of length N , is defined as the best linear system approximating the system's output in the mean square sense [5], [11],

$$G_{BLA}(q) = \operatorname{argmin}_{G \in \mathcal{G}} \frac{1}{N} \sum_{t=1}^N (y(t) - G(q, \theta) u(t))^2, \quad (3)$$

where $G(q)$ is a linear model belonging to the class of linear systems \mathcal{G} . Clearly, the BLA depends on the class of input signals \mathcal{U} .

We also recall Bussgang's theorem [3] about cross spectra transformation.

Theorem 1: (Bussgang) Let $m(t)$ and $n(t)$ be two real-valued, jointly Gaussian stationary process. Let f be a

nonlinear function and g a stochastic process defined by

$$g(t) = f(n(t)). \quad (4)$$

Then the cross spectrum between m and n is proportional to the cross spectrum between m and g :

$$\Phi_{mg}(\omega) = k \Phi_{mn}(\omega) \quad (5)$$

where k is areal-valued constant (that may be zero).

Now we can use the result from Bussgang's theorem to extend the consistency result of the BLA to our more general situation with process noise. We consider the Wiener-Hammerstein structure with process noise, as defined in Figure 1. The equations defining the true system are

$$y(t) = G_2^0(q) f(x(t)) + e(t) \quad (6a)$$

$$x(t) = x_0(t) + w(t) \quad (6b)$$

$$x_0(t) = G_1^0 u(t) \quad (6c)$$

where G_1^0 , G_2^0 and $f(\cdot)$ indicate the true linear parts and nonlinear function. Then we have the following theorem.

Theorem 2: If the following assumptions are satisfied,

- 1) *The input $u(t)$ and the process noise $w(t)$ are independent, Gaussian, stationary processes;*
- 2) *The measurement noise $e(t)$ is a stationary stochastic process, independent of u and w ;*
- 3) *$G(q, \theta)$ is an arbitrary transfer function parametrization with freely adjustable gain, such that $G(q, \theta_0) = G_1^0(q) G_2^0(q)$, for some parameter value θ_0 ;*
- 4) *Parameter θ is estimated from u and y using an output error method,*

$$\hat{\theta}_N = \operatorname{argmin}_{\theta} \frac{1}{N} \sum_{t=1}^N (y(t) - G(q, \theta) u(t))^2, \quad (7)$$

then

$$G(q, \hat{\theta}_N) \rightarrow k G_1^0(q) G_2^0(q) \text{ w.p. 1 as } N \rightarrow \infty \quad (8)$$

Proof: Since e and u are independent, the cross spectra between u and y , y_0 will be the same: $\Phi_{yu} = \Phi_{y_0u}$. Since u and w are Gaussian, then $x(t)$ is also Gaussian and Bussgang's theorem tells us that $\Phi_{zu} = k \Phi_{xu}$. Signals u and w are also independent, then $\Phi_{xu} = \Phi_{x_0u}$. Thus, we have the following result,

$$\Phi_{zu}(\omega) = k \Phi_{xu}(\omega) = k \Phi_{x_0u}(\omega) = k G_1^0(e^{i\omega}) \Phi_u(\omega). \quad (9)$$

Similarly, for the second linear system, we get

$$\Phi_{yu}(\omega) = G_2^0(e^{i\omega}) \Phi_{zu}(\omega) = k G_2^0(e^{i\omega}) G_1^0(e^{i\omega}) \Phi_u(\omega). \quad (10)$$

Now, we know that $\hat{\theta}_N$, asymptotically, will minimize (see e.g. Chapter 8 in [9]) the function

$$\begin{aligned} V(\theta) &= E(y(t) - G(q, \theta) u(t))^2 \\ &= \frac{1}{2\pi} \int (\Phi_y(\omega) - 2 \operatorname{Re}[G(e^{-i\omega}, \theta) \Phi_{yu}(\omega)] \\ &\quad + |G(e^{i\omega}, \theta)|^2 \Phi_u(\omega)) d\omega. \end{aligned} \quad (11)$$

In (11), we can add and subtract the quantity $k^2|G_1^0(e^{i\omega})G_2^0(e^{i\omega})|^2\Phi_u(\omega)$, obtaining

$$\begin{aligned} V(\theta) = & \frac{1}{2\pi} \int (\Phi_y(\omega) + k^2|G_1^0(e^{i\omega})G_2^0(e^{i\omega})|^2\Phi_u(\omega) \\ & - k^2|G_1^0(e^{i\omega})G_2^0(e^{i\omega})|^2\Phi_u(\omega) \\ & - 2Re[G(e^{-i\omega}, \theta)\Phi_{yu}(\omega)] \\ & + |G(e^{i\omega}, \theta)|^2\Phi_u(\omega))d\omega. \end{aligned} \quad (12)$$

Since $\Phi_y(\omega)$ and $k^2|G_1^0(e^{i\omega})G_2^0(e^{i\omega})|^2\Phi_u(\omega)$ are θ -independent terms, minimizing $V(\theta)$ is the same as minimizing

$$\begin{aligned} W(\theta) = & \frac{1}{2\pi} \int (k^2|G_1^0(e^{i\omega})G_2^0(e^{i\omega})|^2\Phi_u(\omega) \\ & - 2Re[G(e^{-i\omega}, \theta)\Phi_{yu}(\omega)] \\ & + |G(e^{i\omega}, \theta)|^2\Phi_u(\omega))d\omega. \end{aligned} \quad (13)$$

We can substitute now the relation for $\Phi_{yu}(\omega)$ from (10),

$$\begin{aligned} W(\theta) = & \frac{1}{2\pi} \int (k^2|G_1^0(e^{i\omega})G_2^0(e^{i\omega})|^2\Phi_u(\omega) \\ & - 2Re[kG(e^{-i\omega}, \theta)G_2^0(e^{i\omega})G_1^0(e^{i\omega})\Phi_u(\omega)] \\ & + |G(e^{i\omega}, \theta)|^2\Phi_u(\omega))d\omega \end{aligned} \quad (14)$$

that leads to

$$W(\theta) = \frac{1}{2\pi} \int (|kG_1^0(e^{i\omega})G_2^0(e^{i\omega}) - G(e^{i\omega}, \theta)|^2\Phi_u(\omega))d\omega, \quad (15)$$

which is minimized by $G(e^{i\omega}, \theta) = kG_1^0(e^{i\omega})G_2^0(e^{i\omega})$. ■

The result is that, up to a gain, the linear system $G(q)$ (BLA) will be a consistent estimate of the concatenation of G_1^0 and G_2^0 . Hence, in principle, we can still use the BLA to obtain the underlying dynamics of the W-H system. The next step, then, would be to decide how to split the BLA into G_1 and G_2 and estimate the nonlinearity.

In the next section we will address the problem of the estimation of the nonlinearity, in presence of process noise.

IV. IMPACT OF THE PROCESS NOISE ON THE ESTIMATION OF THE NONLINEARITY

The main result of this section is to show that, due to process noise, the estimation of the nonlinearity will be inconsistent, using a standard PEM approach. This result is important since it influences the splitting of the BLA into the two linear parts.

In general, several splits of the dynamics from the BLA are possible, and one of them will be the correct one (G_1^0 , G_2^0). The question, then, is how to find the right split when process noise is present. When there is no process noise, and under the assumption of white Gaussian measurement noise, the model containing the right split of the dynamics will also be the true predictor for the system. This means that a prediction error method used to estimate the parameters will be a Maximum Likelihood criterion, and the estimate will be, thus, consistent [15]. This is what is done, for example, in the "brute force" approach in [15]. With this approach, the dynamics contained in the BLA are split in all possible ways

between G_1 and G_2 , and for each possible split a Wiener-Hammerstein model is estimated by fitting the parameters of the nonlinearity, using PEM.

On the contrary, if process noise is present, the previous approach is not possible any more, due to the bias in the nonlinearity. This case will be considered in the next section and we will show that, when the right split is available, a standard PEM approach provides a biased estimation of the parameters of the nonlinearity.

A. Biased Estimation of the Nonlinearity

In this section we show that the estimation of the parameters of the nonlinearity of a Wiener-Hammerstein model is biased, in presence of process noise and when the true linear parts are known.

Assume that the true system is within the model class, i.e. there exist parameters $(\alpha_0, \beta_0, \gamma_0)$ such that

$$y(t) = G_2(q, \alpha_0)f(G_1(q, \beta_0)u(t) + w(t), \gamma_0) + e(t). \quad (16)$$

The approximate PEM criterion, introduced in Section II, is described by

$$V_N(\alpha, \beta, \gamma) = \frac{1}{N} \sum_{t=1}^N (y(t) - G_2(q, \beta)f(G_1(q, \alpha)u(t), \gamma))^2. \quad (17)$$

Consistency of the parameters means that

$$\hat{\alpha}, \hat{\beta}, \hat{\gamma} \rightarrow \alpha_0, \beta_0, \gamma_0 \quad \text{when } N \rightarrow \infty, \quad (18)$$

where $[\hat{\alpha}, \hat{\beta}, \hat{\gamma}] = \text{argmin} V_N(\alpha, \beta, \gamma)$. The true dynamics are given from the BLA, and they are fixed, i.e. $\alpha = \alpha_0$, $\beta = \beta_0$.

Then we have the following theorem.

Theorem 3: Let the nonlinearity $f(x(t))$ be polynomial or well approximated by a polynomial function. Under the assumption of ergodicity, the estimate of γ , obtained by minimization of the approximate PEM criterion, see (17), is biased.

Proof: By using the input-dependent transformation of the process noise after the nonlinearity, similarly to what has been done in [7], the true system can be written as

$$y(t) = G_2(q, \beta_0)[f(G_1(q, \alpha_0)u(t), \gamma_0) + \tilde{w}(t)] + e(t) \quad (19)$$

where

$$\tilde{w}(t) = f(G_1(q, \alpha_0)u(t) + w(t), \gamma_0) - f(G_1(q, \alpha_0)u(t), \gamma_0). \quad (20)$$

Statistical properties are not preserved in the transformation from $w(t)$ to $\tilde{w}(t)$. In particular, signal $\tilde{w}(t)$ is not independent of $u(t)$, as it happens for $w(t)$. This will be used later in the proof.

By using expression (20), the cost function $V_N(\alpha, \beta, \gamma)$ can be written as

$$V_N(\alpha, \beta, \gamma) = \frac{1}{N} \sum_{t=1}^N [G_2^0(f_0(t) + \tilde{w}(t)) + e(t) - G_2f(t, \gamma)]^2, \quad (21)$$

where, to simplify the notation, the following symbols are introduced,

$$G_2 = G_2^0 \triangleq G_2(q, \beta_0) \quad (22a)$$

$$f_0(t) \triangleq f(G_1(q, \alpha_0)u(t), \gamma_0) \quad (22b)$$

$$f(t, \gamma) \triangleq f(G_1(q, \alpha_0)u(t), \gamma). \quad (22c)$$

The approximate PEM criterion then, becomes

$$\begin{aligned} V_N(\gamma) &= \frac{1}{N} \sum_{t=1}^N [G_2^0(f_0(t) + \tilde{w}(t)) + e(t) - G_2^0 f(t, \gamma)]^2 \\ &= \frac{1}{N} \sum_{t=1}^N [G_2^0(f_0(t) - f(t, \gamma))]^2 \\ &\quad + \frac{1}{N} \sum_{t=1}^N [G_2^0 \tilde{w}(t) + e(t)]^2 \\ &\quad + \frac{2}{N} \sum_{t=1}^N [G_2^0(f_0(t) - f(t, \gamma))(G_2^0 \tilde{w}(t) + e(t))]. \end{aligned} \quad (23)$$

Under the ergodicity assumption, the time averages tend to the mathematical expectations as N tends to infinity. The operator E , then, denotes both mathematical expectation and averaging over time. Since the measurement noise $e(t)$ is zero mean and signals $u(t)$ and $w(t)$ are independent, as N tends to infinity, the criterion tends to

$$\begin{aligned} \bar{V}_N(\gamma) &= [G_2^0 E(f_0(t) - f(t, \gamma))]^2 + [G_2^0 E \tilde{w}(t)]^2 + E e(t)^2 \\ &\quad + 2[G_2^0]^2 [E(f_0(t) - f(t, \gamma)) \tilde{w}(t)]. \end{aligned} \quad (24)$$

This cost is quadratic except for the last term $E(f_0(t) - f(t, \gamma)) \tilde{w}(t)$, linear in $f(t, \gamma)$. We want to show that, due to the presence of this term, the estimation of γ is biased, i.e.

$$\exists \gamma^* \neq \gamma_0 : \bar{V}_N(\gamma^*) < \bar{V}_N(\gamma_0). \quad (25)$$

Let us consider an arbitrarily small ε such that

$$\gamma^* = \gamma_0 + \varepsilon. \quad (26)$$

Since ε is small, we can focus on the first order Taylor approximation of the term $E(f_0(t) - f(t, \gamma)) \tilde{w}(t)$, at γ_0 . This will be

$$-E \left[\left. \frac{df(t, \gamma)}{d\gamma} \right|_{\gamma_0} \tilde{w}(t) \right] \varepsilon. \quad (27)$$

Since ε is arbitrary, it is enough to show that

$$E \left[\left. \frac{df(t, \gamma)}{d\gamma} \right|_{\gamma_0} \tilde{w}(t) \right] \neq 0. \quad (28)$$

Let us consider a linearly parametrized nonlinearity, i.e.

$$f(x_0(t), \gamma) = \gamma^T g(x_0(t)), \quad (29)$$

with $x_0(t) = G_1(q, \alpha_0)$. Thus,

$$\left. \frac{df(t, \gamma)}{d\gamma} \right|_{\gamma_0} = g(x_0(t)). \quad (30)$$

On the other hand, $\tilde{w}(t)$ can be written, in terms of $g(x_0)$, as

$$\tilde{w}(t) = \gamma_0^T [g(x_0(t) + w(t)) - g(x_0(t))] \quad (31)$$

Since g is a polynomial nonlinearity, it exists at least one $n \geq 2$, such that $g(x_0 + w) = (x_0 + w)^n \neq 0$. This term can be expanded as

$$(x_0 + w)^n = x_0^n + n x_0^{n-1} w + \binom{n}{2} x_0^{n-2} (w)^2 + \mathcal{O}(w^3). \quad (32)$$

Therefore we have

$$\begin{aligned} g(x_0) &= x_0^n \\ \tilde{w}(t) &= \gamma_0^T [(x_0^n + n x_0^{n-1} w + \mathcal{O}(w^3)) - x_0^n], \end{aligned} \quad (33)$$

and, since w is zero mean, the expression (28) becomes

$$\gamma_0^T E [x_0^n (\binom{n}{2} x_0^{n-2} (w)^2 + \mathcal{O}(w^3))]. \quad (34)$$

Therefore, the argument of the expectation operator contains one term in the form $x_0^{2n-2} w^2$. Since, x_0 and w are independent and they are raised to positive powers, their expectation is different from 0. ■

Remark The above theorem only considers non-linear functions that can be well approximated by polynomials. For other common choices of non-linearities, such as piece-wise linear function, the proof must be developed differently, but similar arguments can be used.

The theorem tells that the estimate will not be consistent, due to the presence of process noise. Thus, the "brute force" algorithm, consisting in splitting the dynamics in all possible ways and estimating the nonlinearity, will not provide a consistent estimate.

This will make it harder to find the right split of the BLA and, then, the open question is whether we can find another approach that provides unbiased estimate of the nonlinearity in presence of process noise. One approach could be to try to extend the result in [7], where a consistent Maximum Likelihood estimator is presented for the Wiener model, to the Wiener-Hammerstein case. This will be addressed in future research.

V. SIMULATION EXAMPLES

In this section two simulation examples are shown in order to illustrate the theoretical results obtained in previous sections. Two different Wiener-Hammerstein systems with process noise are simulated to obtain data. For comparison, for each system two different data sets are generated, one with process noise (*w/ pc*) and one without (*w/out pc*). The generated data are used to estimate the best linear approximation of the system, and then, assuming the right split of the linear dynamics, to estimate the parameters of the nonlinearity. Monte-Carlo simulations are used to generate estimation distributions over 1000 sets of 1000 data points for each example.

In both cases it will be shown that the estimation of the best linear model is unbiased (see Section III), but when the parameters of the nonlinearity are identified, the process noise causes a bias in the estimation (see Section IV).

TABLE I: BLA estimation - Example 1

Poles/Zeros	True	Estimated ($\mu \pm \sigma$)
p_1	0.4	0.3988 ± 0.1000
p_2	0.8	0.7998 ± 0.0401

TABLE II: Nonlinearity estimation - Example 1

Par	True	Est w/ pc ($\mu \pm \sigma$)	Est w/out pc ($\mu \pm \sigma$)
c_0	1.0	1.4591 ± 0.3934	1.0010 ± 0.0009
c_1	1.0	2.4337 ± 0.3344	0.9987 ± 0.0008
c_2	0.01	0.0014 ± 0.0213	$0.0095 \pm 6.3 * 10^{-5}$
c_3	0.01	-0.0075 ± 0.0034	$0.0101 \pm 9.5 * 10^{-6}$

The PEM criterion for BLA and the approximative PEM criterion for nonlinearity estimation are minimized using a software package for the Mathematica platform [14]. To estimate the BLA, an Output Error model structure is used [9].

A. Example 1: Two first order systems with polynomial nonlinearity

The first example is a Wiener-Hammerstein system consisting in two first order linear systems with a polynomial nonlinearity in the middle:

$$\begin{aligned} x_0(t) &= \frac{q}{q - a_1} u(t) \\ z(t) &= f(x_0(t) + w(t)) \\ y(t) &= \frac{q}{q - a_2} z(t) + e(t) \end{aligned} \quad (35)$$

where the $f(x)$ is a third degree polynomial:

$$f(x(t)) = c_0 + c_1 x(t) + c_2 x^2(t) + c_3 x^3(t) \quad (36)$$

The model structures and the degree of the polynomial are assumed to be known. The input u is white Gaussian noise with standard deviation 1 and the process noise w is also Gaussian with standard deviation 4. The measurement noise e is Gaussian with variance 1. The signals u , w and e are mutually independent. The values a_1 and a_2 correspond to two real poles in $p_1 = 0.4$ and $p_2 = 0.8$.

First the BLA estimation is performed. In Table I, there is the comparison between the true poles p_1 and p_2 and the corresponding poles of the estimated OE model (BLA). The simulation confirms the theoretical results: the estimation is consistent. Mean (μ) and standard deviation (σ) are computed over 1000 runs.

Now the two linear parts are fixed according to the right split and the nonlinearity is estimated, using the approximative PEM (see Section IV). In Table II, the estimation of the nonlinear parameters in presence of process noise (*Est w/ pc*) is shown together with the true values of the parameters (*True*) and the estimation when $w(t) \equiv 0$ (*Est w/out pc*). While the estimation with $w(t) \equiv 0$ provides unbiased results, if process noise is affecting the data then the estimate is not consistent any more. Especially parameter c_1 shows a big bias.

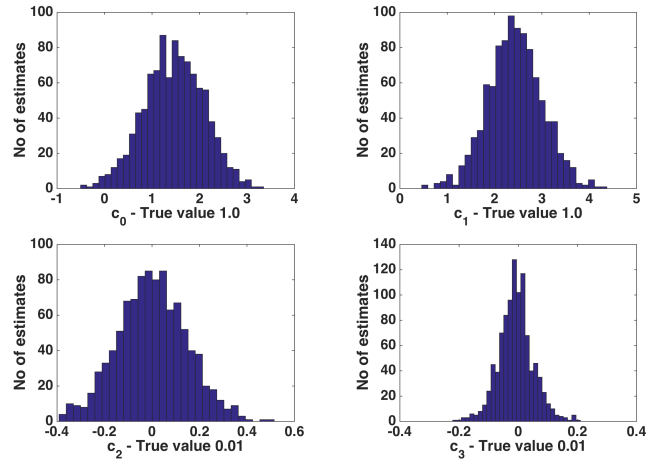


Fig. 2: Example 1. Histograms over 1000 Monte-Carlo simulations of the estimates of the parameters of the nonlinearity in presence of process noise (see also Table II). Parameters c_0 and c_1 show bias in the estimate.

TABLE III: BLA estimation - Example 2

Poles/Zeros	True	Estimated ($\mu \pm \sigma$)
$p_{1,2}$	$0.8 \pm 0.4i$	$(0.7965 \pm 0.012) \pm (0.3999 \pm 0.0154)i$
$p_{3,4}$	$0.4 \pm 0.7i$	$(0.3954 \pm 0.037) \pm (0.6997 \pm 0.0504)i$
z_1	0.6	0.5714 ± 0.1015

In Figure 2, histograms of the estimates distribution are shown for the parameters of the nonlinearity.

B. Example 2: Two second order systems with polynomial nonlinearity

The second example is a Wiener-Hammerstein system consisting in two second order linear systems with a polynomial nonlinearity in the middle:

$$\begin{aligned} x_0(t) &= \frac{q - b_1}{q^2 + a_1 q + a_2} u(t) \\ z(t) &= f(x_0(t) + w(t)) \\ y(t) &= \frac{q}{q^2 + a_3 q + a_4} z(t) + e(t) \end{aligned} \quad (37)$$

where the $f(x)$ is a third degree polynomial:

$$f(x(t)) = c_0 + c_1 x(t) + c_2 x^2(t) + c_3 x^3(t) \quad (38)$$

Signals u , w and e have same statistical properties as previous example and standard deviations, respectively, 5, 5 and 1. Linear parameters b_1 , a_1 , a_2 , a_3 , a_4 correspond, respectively, to a real zero in $z_1 = 0.6$ and two pairs of complex poles in $p_{1,2} = 0.8 \pm 0.4i$ and $p_{3,4} = 0.4 \pm 0.7i$.

Also in this case the BLA is consistent (Table III). In Table IV, the true nonlinear parameters and their estimation (with and without process noise) are compared. In this case the estimation of c_1 is clearly biased and the variance of c_0 is very high, when process noise is present. On the other hand, the estimation when $w(t) \equiv 0$ is consistent.

In Figure 3, histograms of the estimates distribution are shown for the parameters of the nonlinearity.

TABLE IV: Nonlinearity estimation - Example 2

Par	True	Est w/ pc ($\mu \pm \sigma$)	Est w/out pc ($\mu \pm \sigma$)
c_0	0.0	0.3517 ± 2.8729	0.0000 ± 0.0014
c_1	0.7	1.2484 ± 0.0812	$0.7000 \pm 1.5 * 10^{-5}$
c_2	0.01	0.0099 ± 0.0002	$0.0100 \pm 2.5 * 10^{-8}$
c_3	0.005	$0.0048 \pm 1.04 * 10^{-6}$	$0.0050 \pm 1.2 * 10^{-10}$

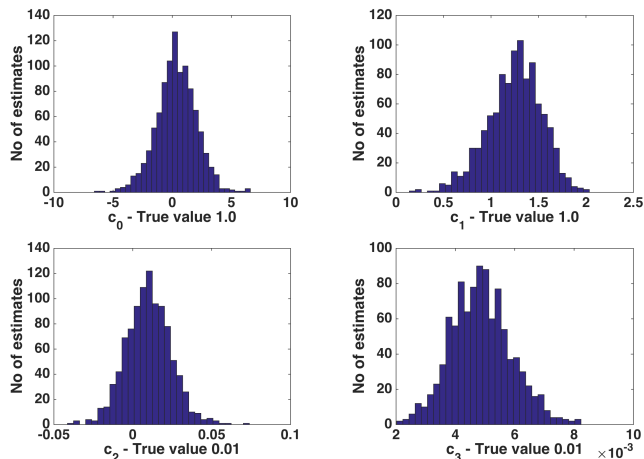


Fig. 3: Example 2. Histograms over 1000 Monte-Carlo simulations of the estimates of the parameters of the nonlinearity in presence of process noise (see also Table IV). There is a clear bias in the estimation of parameter c_1 .

VI. CONCLUSIONS AND FUTURE WORKS

A. Conclusions

In this paper we focused on the problem of identification of Wiener-Hammerstein models, when disturbances are affecting the input of the nonlinearity. Several identification algorithms are based on the split of the dynamics contained in the best linear approximation, which is a consistent estimate of the concatenation of the two linear parts, when only measurement noise is present. Here we addressed the consistency aspects of the BLA, also in presence of process noise. We obtained the following results:

- the presence of process noise does not influence the consistency of the best linear approximation (BLA) of the system; with or without process noise the BLA represents an unbiased estimate of the concatenation of the two linear parts;
- once the two linear parts are known, estimating the nonlinearity, using a standard PEM approach, leads to a biased result. This affects several identification algorithms based on the BLA, since it makes harder to split the dynamics into the correct linear blocks.

All the results have been illustrated using Monte-Carlo simulations, confirming the theoretical results.

B. Future Works

Future research will address the problem of finding an identification algorithm for Wiener-Hammerstein systems based on the split of the BLA, which also provides consistent estimate of the nonlinearity when process noise is present. In particular, it will be interesting to investigate possible splitting problems of the dynamics contained in the BLA, when disturbances are affecting the input of the nonlinearity.

REFERENCES

- [1] S. A. Billings and S. Y. Fakhouri. Identification of systems containing linear dynamic and static nonlinear elements. *Automatica*, 18(1):15–26, January 1982.
- [2] D. R. Brillinger. The identification of a particular nonlinear time series system. *Biometrika*, 64(3):509–515, 1977.
- [3] J. Bussgang. Crosscorrelation functions of amplitude-distorted gaussian signals. *RLE Technical Reports*, 216, 1952.
- [4] P. Crama and J. Schoukens. Computing an initial estimate of a wiener-hammerstein system with a random phase multisine excitation. *Instrumentation and Measurement, IEEE Transactions on*, 54(1):117–122, 2005.
- [5] M. Enqvist and L. Ljung. Linear approximations of nonlinear {FIR} systems for separable input processes. *Automatica*, 41(3):459 – 473, 2005. Data-Based Modelling and System Identification.
- [6] G. Giordano and J. Sjöberg. A time-domain fractional approach for wiener-hammerstein systems identification. *IFAC-PapersOnLine*, 48(28):1232 – 1237, 2015. 17th {IFAC} Symposium on System Identification {SYSID} 2015Beijing, China, 1921 October 2015.
- [7] A. Hagenblad, L. Ljung, and A. Wills. Maximum likelihood identification of wiener models. *Automatica*, 44(11):2697 – 2705, 2008.
- [8] A. Haryanto and K. S. Hong. Maximum likelihood identification of wienerhammerstein models. *Mechanical Systems and Signal Processing*, 41(12):54 – 70, 2013.
- [9] L. Ljung, editor. *System Identification (2Nd Ed.): Theory for the User*. Prentice Hall PTR, Upper Saddle River, NJ, USA, 1999.
- [10] R. Pintelon, Schoukens J., Vandersteen G., and K. Barb. Estimation of nonparametric noise and {FRF} models for multivariable systems part i: Theory. *Mechanical Systems and Signal Processing*, 24(3):573 – 595, 2010.
- [11] R. Pintelon and J. Schoukens. *System identification: a frequency domain approach*. IEEE press, New York, 2001.
- [12] J. Schoukens, R. Pintelon, and M. Enqvist. Study of the {LTI} relations between the outputs of two coupled wiener systems and its application to the generation of initial estimates for wienerhammerstein systems. *Automatica*, 44(7):1654 – 1665, 2008.
- [13] J. Schoukens, R. Pintelon, J. Paduart, and G. Vandersteen. Nonparametric initial estimates for wiener-hammerstein systems. In *System Identification*, volume 14, pages 778–783, 2006.
- [14] J. Sjöberg and H. Hjalmarsson. A system, signals and identification toolbox in mathematica with symbolic capabilities. In *15th IFAC Symposium on System Identification, Saint-Malo, France, 2009*.
- [15] J. Sjöberg and J. Schoukens. Initializing WienerHammerstein models based on partitioning of the best linear approximation. *Automatica*, 48(2):353–359, 2012.
- [16] L. Vanbeylen. A fractional approach to identify Wiener-Hammerstein systems. *Automatica*, 50(3):903 – 909, 2014.
- [17] A. Wills and B. Ninness. Generalised Hammerstein-Wiener system estimation and a benchmark application. *Control Engineering Practice*, 20(11):1097 – 1108, 2012. Special Section: Wiener-Hammerstein System Identification Benchmark.