Application of a Linear PEM Estimator to a Stochastic Wiener-Hammerstein Benchmark Problem ⋆

Mohamed Rasheed Abdalmoaty   Håkan Hjalmarsson

Automatic Control Lab and ACCESS Linnaeus Center,
KTH Royal institute of Technology, Stockholm, Sweden
(e-mail: {mohamed.abdalmoaty, hakan.hjalmarsson}@ee.kth.se).

Abstract: The estimation problem of stochastic Wiener-Hammerstein models is recognized to be challenging, mainly due to the analytical intractability of the likelihood function. In this contribution, we apply a computationally attractive prediction error method estimator to a real-data stochastic Wiener-Hammerstein benchmark problem. The estimator is defined using a deterministic predictor that is nonlinear in the input. The prediction error method results in tractable expressions, and Monte Carlo approximations are not necessary. This allows us to tackle several issues considered challenging from the perspective of the current mainstream approach. Under mild conditions, the estimator can be shown to be consistent and asymptotically normal. The results of the method applied to the benchmark data are presented and discussed.

Keywords: System identification, Nonlinear systems, Stochastic systems, Wiener-Hammerstein, Benchmark problem.

1. INTRODUCTION

System identification of stochastic linear models is a well developed and understood subject; algorithms based on the Maximum Likelihood (ML) Method, Prediction Error Methods (PEM), Subspace Methods, Instrumental Variables Methods, etc. have been studied and used for many decades. The availability of excellent textbooks, such as Hannan and Deistler [1988], Söderström and Stoica [1989], Ljung [1999], Pintelon and Schoukens [2012], is a clear indication of the maturity of the subject. Nevertheless, systems in real-life exhibit a multitude of nonlinear behaviors; and the assumptions underlying the theoretical framework of linear system identification are often not met in practice. Still, in many cases, depending on the intended use of the model, linear system identification can give satisfactory results; see Schoukens et al. [2016]. Otherwise, nonlinear models have to be identified and system identification of stochastic nonlinear models is a field undergoing rapid development.

Historically, most of the literature on nonlinear system identification dealt with cases where the only source of uncertainty is at the systems' output, e.g., due to the imperfections of the measurement sensors. Under that assumption, the formulation of the estimation problem is straightforward and the focus has been on two main issues: (*i*) model set selection and parameterization, and (*ii*) optimization and initialization methods; see the books Nelles [2001], Giri and Bai [2010], Billings [2013], Mzyk [2013] and the articles Billings [1980], Haber and Unbehauen [1990], Juditsky et al. [1995], Sjöberg et al. [1995].

⋆ This work was supported by the Swedish Research Council under contracts 2015-05285, 2016-06079 and by the European Research Council under the advanced grant LEARN, contract 267381.

A popular model set in this context is the set of block-oriented models where models are constructed by connecting blocks of two types: Linear Time-Invariant (LTI) dynamical blocks, and static nonlinearity blocks. These models are usually easier to understand compared to general nonlinear models, since intuition and prior knowledge about the system may help in selecting the model structure. Moreover, it is usually possible to separate the estimation of the linear and the nonlinear blocks; see Sjöberg and Schoukens [2012], Sjöberg et al. [2012], Schoukens [2015], Schoukens and Tiels [2016]. The Best Linear Approximation (BLA) [Ljung 2001, Enqvist 2005, Schoukens et al. 2014], with respect to a certain class of inputs, is known to be related to the LTI blocks of the model, and may be used to initialize the estimation problem.

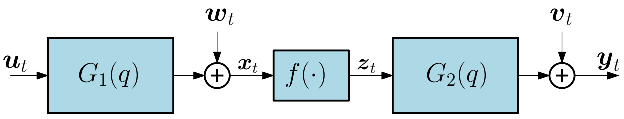

Fig. 1. A Wiener-Hammerstein model with process disturbance. The blocks G_1, G_2 are LTI dynamical models, and the block f represents a general static nonlinearity. The signals $\boldsymbol{u}_t$ and $\boldsymbol{y}_t$ are the input and output respectively, while $\boldsymbol{w}_t$ and $\boldsymbol{v}_t$ are unobserved stochastic processes that may be colored.

A relatively simple block-oriented model is the Wiener-Hammerstein (WH) model shown in Figure 1. It is constructed by connecting three blocks in series: a static nonlinearity block in between two LTI blocks. Several methods have been developed for the estimation of WH models when the process disturbance, $\boldsymbol{w}_t$, is not present (or enters at the output of the nonlinearity); see, for example, the preface of the Control Engineering Practice special issue on

Wiener-Hammerstein system identification [Hjalmarsson et al. 2012]. Nevertheless, it has been shown in Hagenblad et al. [2008] that ignoring an existing process disturbance leads to a biased estimate.

From the parameter estimation perspective, the main difficulty with the existence of a process disturbance before the nonlinearity is the analytical intractability of the likelihood function. In recent years, there have been a growing interest in this problem. Methods based on Monte Carlo importance sampling [Durbin and Koopman 1997, Abdalmoaty and Hjalmarsson 2016, for example], as well as on sequential Monte Carlo (SMC, a.k.a. particle filters) algorithms and (particle) Markov Chain Monte Carlo ((p)MCMC) algorithms have been proposed; see Ninness [2009], Ninness et al. [2010], Schön et al. [2011], Lindsten and Schön [2013], Schön et al. [2015]. These methods are asymptotically efficient and have been shown to provide acceptable results on several academic examples; however, their application is so far limited to cases where fundamental problems—such as particle degeneracy of SMC methods—can be avoided.

In this contribution, we apply a computationally attractive PEM estimator [Abdalmoaty 2017] to a stochastic WH benchmark problem using the real-data sets provided by Schoukens and Noël [2017]. The used estimator relies only on the first moment of the assumed model and, under some conditions, can be shown to be consistent. A detailed asymptotic analysis of the method will be provided in a future contribution by the authors. We start in Section 2 by formulating the problem and setting the main assumptions. In Section 3, we introduce an Output-Error-type (OE-type) predictor which is then used in Section 4 to construct a PEM. The performance of the suggested estimator is demonstrated on a couple of numerical simulation examples. In Section 5, we apply the proposed estimation method to the benchmark data and discuss the achieved performance. The paper is concluded in Section 6.

2. PROBLEM FORMULATION

Consider a Single-Input Single-Output (SISO) stochastic WH model defined by the relations

$$\begin{aligned} \boldsymbol{y}_t &= G_2(q;\theta^\circ)\boldsymbol{z}_t + \boldsymbol{v}_t, \quad t = 1,2,3\ldots \\ \boldsymbol{z}_t &= f(\boldsymbol{x}_t;\theta^\circ) \\ \boldsymbol{x}_t &= G_1(q;\theta^\circ)u_t + \boldsymbol{w}_t, \\ \boldsymbol{w}_t &= H(q)\boldsymbol{\varepsilon}_t, \end{aligned} \tag{1}$$

in which θ° is a finite-dimensional real parameter, $G_1(q,\theta^\circ)$ and $G_2(q,\theta^\circ)$ are stable rational functions with known order, and q is the shift operator. The function f is a parameterized nonlinear function. The input and the output are denoted by u_t and $\boldsymbol{y}_t$ respectively [1]. It will be assumed that the input u_t is known $\forall t$, $\boldsymbol{v}_t$ and $\boldsymbol{w}_t$ are zero mean stationary stochastic processes, and $\mathbb{E}[\boldsymbol{w}_t^r] = m_r$ is finite for some finite $r > 1$. Notice that the values m_r can be part of the parameter vector θ° to be estimated. Moreover, we will assume zero initial conditions. Let a data set

$$D_N := \{(y_t, u_t) : t = 1, \ldots, N\},$$

corresponding to a realization of the model output, be given. Our objective is to use D_N to estimate θ°.

[1] A bold font is used to denote random variables and a regular font is used to denote realizations thereof.

The ML estimator is usually the favored choice due to its optimal asymptotic properties; however, it requires the evaluation of the likelihood function given by the integral

$$\int_{\mathbb{R}^N} p_v(y_t - G_2(q;\theta)f(x_t;\theta))p_\varepsilon(H^{-1}(q)(x_t - G_1(q;\theta)u_t)\prod_{t=1}^{N} \mathrm{d}x_t$$

(it is assumed that $\boldsymbol{v}_t$ and $\boldsymbol{\varepsilon}_t$ are independent processes with PDFs p_v and p_ε respectively). This integral is analytically intractable in general, and can be time-consuming to approximate using numerical integration. Unfortunately, computing the PEM estimator based on the optimal mean-square error predictor is as hard as computing the ML estimator. However, it is possible to use the PEM to construct suboptimal but consistent estimators.

3. AN OUTPUT-ERROR-TYPE PREDICTOR

Notice that the PEM does not require an "optimal" predictor in order to construct consistent estimators. To see this, assume for the moment that

$$f(x;\theta^\circ) = x \;\forall x \in \mathbb{R} \quad \text{and} \quad G_2(q;\theta^\circ) = 1.$$

Then,

$$\boldsymbol{y}_t = G_1(q;\theta^\circ)u_t + H(q)\boldsymbol{\varepsilon}_t + \boldsymbol{v}_t, \quad \forall t,$$

i.e., $\boldsymbol{y}_t$ is a linear stationary process. In this case, it is well known that a misspecified noise model does not affect the convergence of the PEM estimator, and under mild conditions [Ljung 1999, Theorem 8.4, pp. 259],

$$\hat{\boldsymbol{\theta}}_N = \arg\min_\theta \sum_{t=1}^{N} (\boldsymbol{y}_t - G_1(q;\theta)u_t)^2 \xrightarrow{\text{a.s.}} \theta^\circ \quad \text{as} \quad N \to \infty.$$

where the symbol $\xrightarrow{\text{a.s.}}$ denotes almost sure convergence. Observe that the only used information regarding the probabilistic structure of $\boldsymbol{y}_t$ is its mean value, and the computation of the likelihood function is not required.

In a straightforward manner, we extend the above observation to stochastic WH models and define what we call an *Output-Error-type (OE-type) predictor* .

Definition 1. (OE-type Predictor). The Output-Error-type predictor of $\boldsymbol{y}_t$ is defined as the deterministic quantity

$$\hat{y}_t(\theta) = \mathbb{E}[\boldsymbol{y}_t|U_t;\theta], \quad t = 1,2,3,\ldots$$

in which $U_t := [u_1 \ldots u_t]^\top$.

The OE-type predictor is very simple to compute: it is given in terms of a closed-form expression in several relevant cases. Consider, for example, the WH model in (1), and let

$$f(\boldsymbol{x}_t;\theta^\circ) = \sum_{k=0}^{d_f} f_k \boldsymbol{x}_t^k, \quad f_k \in \mathbb{R} \;\; \forall k = 0,\ldots,d_f \in \mathbb{N}.$$

Then,

$$\mathbb{E}[\boldsymbol{y}_t|U_t;\theta] = G_2(q;\theta)\mathbb{E}[\boldsymbol{z}_t|U_t;\theta],$$

$$\mathbb{E}[\boldsymbol{z}_t|U_t;\theta] = \sum_{k=0}^{d_f}\sum_{l=0}^{k}\binom{k}{l} f_k m_{k-l}[G_1(q;\theta)u_t]^l$$

in which m_r is the r^{th} moment of $\boldsymbol{w}_t$, and therefore, the OE-type predictor is given by the closed-form expression

$$\hat{y}_{t|t-1}(\theta) = \sum_{k=0}^{d_f}\sum_{l=0}^{k}\binom{k}{l} f_k m_{k-l} G_2(q;\theta)[G_1(q;\theta)u_t]^l,$$

for $t = 1,\ldots,N$. For general nonlinearities, numerical integration may be used to evaluate the required expectation.

4. CONSISTENT PREDICTION ERROR METHOD

A PEM estimator based on the OE-Type predictor can now be defined; we call it the *OE-PEM estimator.*

***Definition* 2.** (OE-PEM estimator). The OE-type PEM estimator of θ° is defined as

$$\hat{\boldsymbol{\theta}} := \arg\min_{\theta} \sum_{t=1}^{N} (\boldsymbol{y}_t - \hat{y}_t(\theta))^2,$$

in which $\hat{y}_t(\theta) = \mathbb{E}[\boldsymbol{y}_t | U_t; \theta]$ is the OE-type predictor given in Definition 1.

Under some conditions on the data and the parameterization of the model, the OE-PEM can be shown to converge almost surely to θ° as $N \to \infty$. We note here that, similar to the LTI case [Ljung 1999], erroneous initial conditions will not affect the asymptotic properties of the method.

In general, the resulting OE-PEM estimator has no closed-form expression, and numerical optimization routines, such as Levenberg-Marquardt algorithm, are necessary; and a good initial guess of θ° is required. It has been shown in Giordano and Sjöberg [2016] that the Best Linear Approximation (BLA) converges to the concatenation of the two LTI blocks of the WH model, modulo a scalar parameter, even when $\boldsymbol{w}_t$ is present. Hence, an algorithm based on the best split of the BLA model [Sjöberg and Schoukens 2012] may be used to initialize the LTI blocks. For each possible split, a WH model is estimated by estimating the nonlinearity using the OE-PEM estimator; and the split with the minimum OE-PEM cost is selected. The effectiveness of this method is illustrated in the following two subsections using simulation examples.

4.1 Consistent estimation of the nonlinearity

We first illustrate the consistency of the OE-PEM estimator, in comparison to the asymptotically biased estimator ignoring $\boldsymbol{w}_t$. Here, only the parameters of the nonlinearity are estimated and the two LTI models are considered fixed and known. This is equivalent to a single step in the best split algorithm of the BLA model.

Consider a SISO WH model where

$$G_1(q) = \frac{1}{1 - 0.7q^{-1}}, \quad G_2(q) = \frac{1}{1 + 0.5q^{-1}},$$
$$H(q) = \frac{q^{-1}(1 + 0.8q^{-1})(1 - 0.6q^{-1})(1 - 0.02q^{-1})}{(1 + 0.5q^{-1})(1 - 0.5q^{-1})(1 - 0.6q^{-1} + 0.34q^{-2})},$$
$$\boldsymbol{\varepsilon}_t \sim \mathcal{N}(0, 0.5)\ \forall t, \ \ \boldsymbol{\varepsilon}_t \perp \boldsymbol{\varepsilon}_s\ \forall t \neq s.$$

and let f be a third order polynomial such that

$$\boldsymbol{z}_t = \theta_1^\circ \boldsymbol{x}_t + \theta_2^\circ \boldsymbol{x}_t^2 + \theta_3^\circ \boldsymbol{x}_t^3,$$
$$\theta^\circ = [1\ -0.2\ -0.1]^\top.$$

The objective is to estimate θ°. We ran 1000 independent Monte Carlo simulations over $\boldsymbol{u_t} \sim \mathcal{N}(0, 5)$, $\boldsymbol{v}_t \sim \mathcal{N}(0, 0.1)$, and $\boldsymbol{w}_t$ where we assumed that $N = 5000$. The data sets were then used to evaluate the PEM estimator ignoring $\boldsymbol{w}_t$ as well as the OE-PEM estimator in two cases: (i) known process disturbance variance $\lambda_w = \mathbb{E}[\boldsymbol{w}_t^2]$, (ii) unknown process disturbance variance (jointly estimated with θ) .

The results are shown in Figure 2 and Table 1. It is clear that ignoring the process disturbance leads to an asymptotically biased estimator; however, the simple OE-PEM estimator given in Definition 2 is consistent.

***Remark* 3.** Notice that knowing λ_w is advantageous: firstly, with known λ_w, the optimization problem is linear in the parameters and linear least-squares can be used to obtain closed-form solutions, secondly, the accuracy of the estimator is improved; for the above example, knowing λ_w decreases the standard deviation of θ_1 by factor of 2.

When λ_w is unknown, the resulting optimization problem involves products of the parameters of the nonlinearity and monomials of λ_w. In this case, a PEM ignoring $\boldsymbol{w}_t$ may be used to initialize the parameters; however, it is not yet clear how an initial estimate of the process disturbance variance can be obtained. In the above example, the `covar` function of MATLAB was used to find that $\lambda_w = 0.8666$, and the optimization problem was initialized at $\lambda_w = 1$. In the next example, the method is used to estimate all the blocks of a WH model simultaneously.

4.2 A complete simulation example

In this subsection, we would like to estimate the linear blocks as well as the nonlinearity of a WH model using the OE-PEM estimator.

Consider a SISO WH model where

$$G_1(q;\theta^\circ) = \frac{(1 - 0.7q^{-1})(1 + 0.4q^{-1} + 0.29q^{-1})}{(1 + 0.8q^{-1})(1 - q^{-1} + 0.8125q^{-1})},$$
$$G_2(q;\theta^\circ) = \frac{(1 + 0.5q^{-1})(1 - 0.8q^{-1} + 0.2q^{-1})}{(1 - 0.7q^{-1})(1 + 0.8q^{-1} + 0.65q^{-1})},$$

and let the true nonlinearity be a sigmoid function (representing a symmetric (odd) saturation nonlinearity) as follows

$$\boldsymbol{z}_t = \frac{0.5}{1 + \exp(-0.75\boldsymbol{x}_t)} - \frac{1}{4}, \tag{2}$$

and assume that $\boldsymbol{w}_t$ is an independent process such that $\boldsymbol{w}_t \sim \mathcal{N}(0, 1)$, and let $\boldsymbol{v}_t \sim \mathcal{N}(0, 0.01)$. The input is a known realization of $\boldsymbol{u_t} \sim \mathcal{N}(0, 4)$ and $N = 8000$. Figure 3 shows the output of G_1, the realization of $\boldsymbol{w}_t$ and the output of the sigmoid function; it is obvious that a saturation is active.

Our objective is to estimate the parameters of the linear models. To do so, f is approximated (according to the prior knowledge) with a fifth order odd polynomial: $f(x;\theta) = f_1 x + f_2 x^3 + f_3 x^5$, and the OE-PEM estimator based on the best split of the BLA model is realized (the knowledge of the models orders was used). The obtained estimates:

$$G_1(q;\hat{\theta}) = \frac{(1 - 0.756q^{-1})(1 + 0.434q^{-1} + 0.322q^{-1})}{(1 + 0.796q^{-1})(1 - 1.002q^{-1} + 0.811q^{-1})},$$
$$G_2(q;\hat{\theta}) = \frac{(1 + 0.49q^{-1})(1 - 0.75q^{-1} + 0.206q^{-1})}{(1 - 0.65q^{-1})(1 + 0.805q^{-1} + 0.668q^{-1})},$$

are close to the true models. The estimated polynomial nonlinearity approximates well the true sigmoid function, as shown in Figure 4. In Figure 5, the true poles and zeros of G_1 and G_2 are shown against those of the best initial split of the BLA model and the final OE-PEM estimate; observe how the OE-PEM estimator moved the poles and zeros of the initial BLA model closer to the true ones.

5. THE BENCHMARK PROBLEM

We now turn to the WH benchmark problem [Schoukens and Noël 2017]. The benchmark data [2] is generated using

Table 1. The mean and the standard deviation of the estimated coefficients for the estimators

	θ_1	θ_2	θ_3
True values	1	-0.2	-0.1
classical PEM	0.737 ± 0.0039	-0.203 ± 0.014	-0.099 ± 0.004
OE-PEM (known λ_w)	0.997 ± 0.103	-0.199 ± 0.014	-0.099 ± 0.004
OE-PEM (estimated λ_w)	1.021 ± 0.207	-0.199 ± 0.016	-0.099 ± 0.004

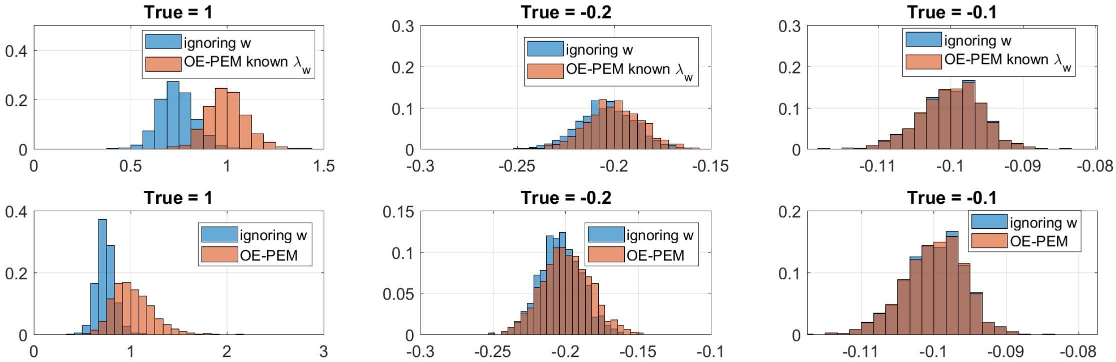

Fig. 2. (Results of the example in Section 4.1) The Histograms of 1000 independent realizations of a PEM estimator ignoring $\boldsymbol{w}_t$ and the OE-PEM estimator in two cases: known process disturbance variance (the top panels), estimated process disturbance variance (the bottom panels).

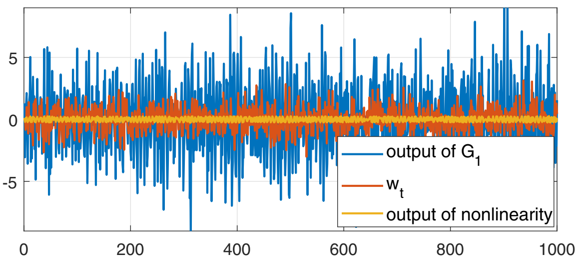

Fig. 3. The output of G_1 in blue, the realization of $\boldsymbol{w}_t$ in red, and the output of the saturation nonlinearity in orange. It is clear that the saturation is active.

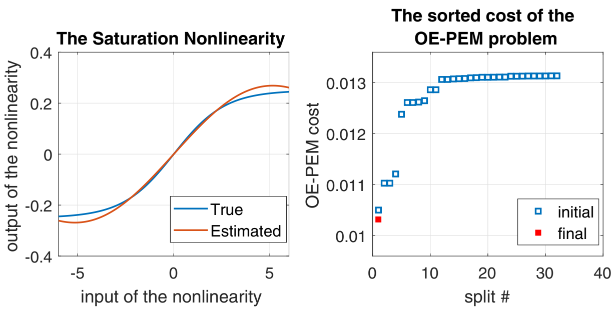

Fig. 4. Results of the example in Section 4.2. On the left: the sigmoid function in (2), and the estimated fifth order odd polynomial approximation. On the right: the sorted OE-PEM cost for the initial 32 splits; the red square shows the final cost of the best initial split.

an electronic circuit, see Figure 6, modeling a stochastic WH model with a saturation nonlinearity. The problem has two main challenges: (i) the contribution of the process disturbance is large, (ii) the LTI model succeeding the saturation is not stably invertible; therefore the nonlinearity is not directly accessible using the available data. Both the input and the output signals are measured; however, the measurement noise is relatively small compared to the process disturbance and may be ignored. The objective is

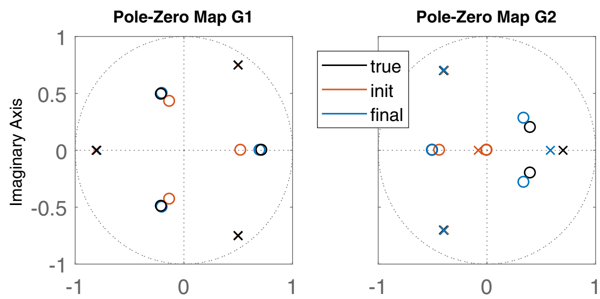

Fig. 5. Results of the example in Section 4.2. The poles (x) and zeros (o) of the true model (in black), the best initial split of the BLA model (in red), and the final OE-PEM estimate (in blue)

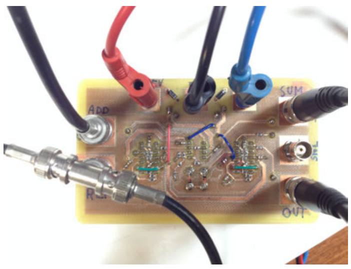

Fig. 6. The electronic circuit used to generate the benchmark data (source: see footnote 2)

to use the measured data to identify a model of the circuit.

To test the performance of the estimated models, two *test data sets measured with* $\boldsymbol{w}_t = 0$ are provided: the outputs due to a random multisine input, and the outputs due to a sine-sweep input. According to the benchmark problem instructions, and to allow for comparison of different methods, the following performance measure is used

$$\text{RMSE}_N = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(y_t - \bar{y}_t)^2},$$

[2] publicly available at http://www.nonlinearbenchmark.org/

in which y_t is the estimated model simulated output due to the inputs of the test data, and $\bar{y}_t$ is the output of the test data due to the same input.

To estimate a model, we used the multisine experiments [3]. The data contains 10 independent experiments, each of which corresponds to two steady-state periods with $N = 8192$. The first experiment was used to identify a BLA model in the time-domain using the `oe` function in Matlab; then, the best split of the estimated BLA was found using the OE-PEM estimator and the same data. The nonlinearity is modeled by a sigmoid function of the form

$$f(x;\theta) = \frac{L}{1+\exp(-kx)} - \frac{L}{2}.$$

Observe that its expectation can be well approximated using the normal cumulative distribution function; and we did that in our solution. The following initial values were used in the splitting algorithm: $L = 0.05$, $k = 1$, and $\lambda_w = 0.25$. The number of poles of the LTI models was fixed in the algorithm to 3, and only causal splits were allowed. The obtained estimate is then used to initialize the OE-PEM problem.

To solve the OE-PEM we used the Levenberg-Marquardt algorithm in Matlab, and concatenated the 10 multisine experiments to form one long data set of length $N = 81920$. The time for the joint optimization was about 35 seconds on a personal Dell laptop. The final estimated process disturbance variance is $\lambda_w = 3.19$. Figures 7 and 8 show the simulated and measured output of the test data. The estimated nonlinearity is shown in Figure 9. The RMSE of the estimated model is given in the last column of Table 2. The first column shows the RMSE of the BLA model, and the second shows the RMSE after splitting the BLA and estimating the nonlinearity. Both are obtained using the first multisine experiment. The third column shows the RMSE of the final model when estimated using the 7^{th} mutlisine experiment [4].

Furthermore, using the same initial values from above, we estimated models using the first 10 experiments of the modulated multisine data sets [5]. Here, $N = 65536$ for each experiment; the obtained RMSEs are given in Table 3. The best result was achieved using the second experiment.

Table 2. The RMSE of the model estimated using the multisine input[3]. The result in the third column was obtained using the 7^{th} multisine experiment (see the text).

	BLA	BLA+NL	OE-PEM *(exp. 7)*	OE-PEM *(all data)*
Swept-sine	0.0281	0.0127	0.0116	**0.0091**
Multisine	0.0339	0.0261	0.0171	**0.0148**

These results are quite encouraging. They are almost the same as the results obtained using the state-of-the-art ML method; however, while the computational time of OE-PEM is in seconds, the computational time of the ML method in this case is a few hours [Svensson et al. 2018].

[3] Stored in `WH_EstimationExample.mat`.
[4] This is the smallest RMSE over the 10 multisine experiments; the estimation time is about 10 seconds.
[5] Stored in `WH_Triangle2_meas.mat`. It contains 50 independent experiments of modulated multisine inputs with a triangular envelope.

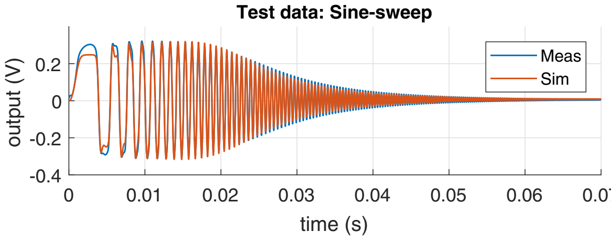

Fig. 7. The first part of the the swept-sine test data: the measured is in blue, and the simulated is in red. Observe how the simulated output follows closely the test data.

Fig. 8. A part of the the multisine test signal: the measured is in blue, and the simulated is in red. Observe how the simulated output follows closely the test data.

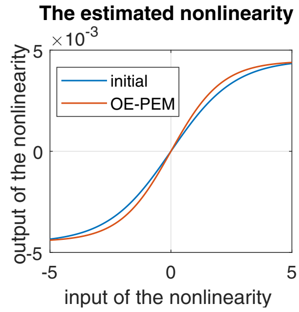

Fig. 9. In blue: the estimate obtained using the best split of the BLA. In red: the OE-PEM estimate obtained using a concatenation of the multisine data.

6. CONCLUSIONS

In this contribution, we applied a linear PEM estimator to a real-data stochastic Wiener-Hammerstein benchmark problem. The estimator is constructed based on an OE-type predictor which is linear in the input; and therefore, the estimator is computationally attractive: the estimation time for $N = 81920$ is about 30 seconds on a personal laptop (compared to a few hours for the alternative (SMC) methods [Svensson et al. 2018]). As with basically all alternative methods, a good starting value for the parameter optimization is necessary. In the WH case, we showed that we could use the BLA to find a good initial guess for the parameters of the model based on the OE-PEM estimator.

REFERENCES

Abdalmoaty, M.R. (2017). *Learning Stochastic Nonlinear Dynamical Systems Using Non-stationary Linear Predictors*. Licentiate thesis, KTH Royal Institute of Technology, Automatic Control Department.

Abdalmoaty, M.R. and Hjalmarsson, H. (2016). A simulated maximum likelihood method for estimation of

Table 3. The RMSE of the models estimated using the OE-PEM estimator and a triangularly modulated multisine input[5]. Here, $N = 65536$ and the estimation time is about 30 seconds for a single experiment. The used initial value is the same as the one used for the results in Table 2: a BLA model is estimated using the first experiment in the multisine data set.

	exp 1	**exp 2**	exp 3	exp 4	exp 5	exp 6	exp 7	exp 8	exp 9	exp 10
Swept-sine	0.0074	**0.0072**	0.0085	0.0074	0.0084	0.0097	0.0089	0.0073	0.0082	0.0086
Multisine	0.0155	**0.0143**	0.0262	0.0190	0.0243	0.0264	0.0237	0.0185	0.0150	0.0252

stochastic Wiener systems. In *2016 IEEE 55th Conference on Decision and Control (CDC)*, 3060–3065.

Billings, S.A. (1980). Identification of nonlinear systems-a survey. *IEE Proceedings D - Control Theory and Applications*, 127(6), 272–285.

Billings, S.A. (2013). *Nonlinear System Identification: NARMAX Methods in the Time, Frequency, and Spatio-Temporal Domains*. Wiley.

Durbin, J. and Koopman, S.J. (1997). Monte carlo maximum likelihood estimation for non-gaussian state space models. *Biometrika*, 84(3), 669–684.

Enqvist, M. (2005). *Linear Models of Nonlinear Systems*. Linköping Studies in Science and Technology. Dissertations No. 985, Institutionen för systemteknik, Linköping University, Sweden.

Giordano, G. and Sjöberg, J. (2016). Consistency aspects of Wiener-Hammerstein model identification in presence of process noise. In *2016 IEEE 55th Conference on Decision and Control (CDC)*, 3042–3047. Institute of Electrical and Electronics Engineers (IEEE).

Giri, F. and Bai, E. (2010). *Block-oriented Nonlinear System Identification*. Springer.

Haber, R. and Unbehauen, H. (1990). Structure identification of nonlinear dynamic systems - a survey on input/output approaches. *Automatica*, 26(4), 651 – 677.

Hagenblad, A., Ljung, L., and Wills, A. (2008). Maximum likelihood identification of Wiener models. *Automatica*, 44(11), 2697 – 2705.

Hannan, E. and Deistler, M. (1988). *The statistical theory of linear systems*. Wiley.

Hjalmarsson, H., Rojas, C.R., and Rivera, D.E. (2012). System identification: A Wiener-Hammerstein benchmark. *Control Engineering Practice*, 20(11), 1095 – 1096.

Juditsky, A., Hjalmarsson, H., Benveniste, A., Delyon, B., Ljung, L., Sjöberg, J., and Zhang, Q. (1995). Nonlinear black-box models in system identification: Mathematical foundations. *Automatica*, 31(12), 1725 – 1750.

Lindsten, F. and Schön, T.B. (2013). Backward simulation methods for monte carlo statistical inference. *Foundations and Trends in Machine Learning*, 6(1), 1–143.

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

Ljung, L. (2001). Estimating linear time-invariant models of nonlinear time-varying systems. *European Journal of Control*, 7(2), 203 – 219.

Mzyk, G. (2013). *Combined Parametric-Nonparametric Identification of Block-Oriented Systems*. Springer.

Nelles, O. (2001). *Nonlinear System Identification: From Classical Approaches to Neural Networks and Fuzzy Models*. Springer.

Ninness, B. (2009). Some system identification challenges and approaches. *IFAC Proceedings Volumes*, 42(10), 1 – 20. 15th IFAC Symposium on System Identification.

Ninness, B., Wills, A., and Schön, T. (2010). Estimation of general nonlinear state-space systems. In *49th IEEE Conference on Decision and Control, Atlanta, Georgia, USA*, 1–6.

Pintelon, R. and Schoukens, J. (2012). *System Identification: A Frequency Domain Approach*. Wiley, 2nd edition.

Schön, T.B., Lindsten, F., Dahlin, J., Wågberg, J., Naesseth, C.A., Svensson, A., and Dai, L. (2015). Sequential Monte Carlo methods for system identification. *IFAC-PapersOnLine*, 48(28), 775 – 786. 17th IFAC Symposium on System Identification SYSID 2015, Beijing.

Schön, T.B., Wills, A., and Ninness, B. (2011). System identification of nonlinear state-space models. *Automatica*, 47(1), 39 – 49.

Schoukens, J., Marconato, A., Pintelon, R., Rolain, Y., Schoukens, M., Tiels, K., Vanbeylen, L., Vandersteen, G., and Mulders, A.V. (2014). System identification in a real world. In *2014 IEEE 13th International Workshop on Advanced Motion Control (AMC)*, 1–9.

Schoukens, J., Vaes, M., and Pintelon, R. (2016). Linear system identification in a nonlinear setting: Nonparametric analysis of the nonlinear distortions and their impact on the best linear approximation. *IEEE Control Systems*, 36(3), 38–69.

Schoukens, M. and Noël, J. (2017). Three benchmarks addressing open challenges in nonlinear system identification. *IFAC-PapersOnLine*, 50(1), 446 – 451. 20th IFAC World Congress.

Schoukens, M. (2015). *Identification of parallel block-oriented models starting from the best linear approximation*. Ph.D. thesis, Vrije Universiteit Brussel.

Schoukens, M. and Tiels, K. (2016). Identification of nonlinear block-oriented systems starting from linear approximations: A survey. *CoRR*, abs/1607.01217.

Sjöberg, J., Lauwers, L., and Schoukens, J. (2012). Identification of Wiener-Hammerstein models: Two algorithms based on the best split of a linear model applied to the SYSID'09 benchmark problem. *Control Engineering Practice*, 20(11), 1119 – 1125.

Sjöberg, J. and Schoukens, J. (2012). Initializing Wiener-Hammerstein models based on partitioning of the best linear approximation. *Automatica*, 48(2), 353 – 359.

Sjöberg, J., Zhang, Q., Ljung, L., Benveniste, A., Delyon, B., Glorennec, P.Y., Hjalmarsson, H., and Juditsky, A. (1995). Nonlinear black-box modeling in system identification: a unified overview. *Automatica*, 31(12), 1691 – 1724.

Söderström, T. and Stoica, P. (1989). *System Identification*. Prentice Hall. Prentice Hall.

Svensson, A., Schön, T.B., and Lindsten, F. (2018). Learning of state-space models with highly informative observations: A tempered sequential Monte Carlo solution. *Mechanical Systems and Signal Processing*, 104, 915 – 928.