Science Arts & Métiers (SAM)
is an open access repository that collects the work of Arts et Métiers Institute of Technology researchers and makes it freely available over the web where possible.

This is an author-deposited version published in: https://sam.ensam.eu
Handle ID: .http://hdl.handle.net/10985/12468

To cite this version :

Maarten SCHOUKENS, Marc RÉBILLAT - Comparison of least squares and exponential sine sweep methods for Parallel Hammerstein Models estimation - Comparison of least squares and exponential sine sweep methods for Parallel Hammerstein Models estimation - Vol. 104, p.851-865 - 2017

Any correspondence concerning this service should be sent to the repository Administrator : scienceouverte@ensam.eu

Comparison of least squares and exponential sine sweep methods for Parallel Hammerstein Models estimation

Marc Rebillat [a,*], Maarten Schoukens [b]

[a] PIMM, Arts et Métiers/CNRS/CNAM, Paris, France
[b] Control Systems, Eindhoven University of Technology, Eindhoven, The Netherlands

Keywords:
Nonlinear system identification
Least-square method Exponential sine sweep

A B S T R A C T

Linearity is a common assumption for many real-life systems, but in many cases the non-linear behavior of systems cannot be ignored and must be modeled and estimated. Among the various existing classes of nonlinear models, Parallel Hammerstein Models (PHM) are interesting as they are at the same time easy to interpret as well as to estimate. One way to estimate PHM relies on the fact that the estimation problem is linear in the parameters and thus that classical least squares (LS) estimation algorithms can be used. In that area, this article introduces a regularized LS estimation algorithm inspired on some of the recently developed regularized impulse response estimation techniques. Another mean to estimate PHM consists in using parametric or non-parametric exponential sine sweeps (ESS) based methods. These methods (LS and ESS) are founded on radically different mathematical backgrounds but are expected to tackle the same issue. A methodology is proposed here to compare them with respect to (i) their accuracy, (ii) their computational cost, and (iii) their robustness to noise. Tests are performed on simulated systems for several values of methods respective parameters and of signal to noise ratio. Results show that, for a given set of data points, the ESS method is less demanding in computational resources than the LS method but that it is also less accurate. Furthermore, the LS method needs parameters to be set in advance whereas the ESS method is not subject to conditioning issues and can be fully non-parametric. In summary, for a given set of data points, ESS method can provide a first, automatic, and quick overview of a nonlinear system than can guide more computa-tionally demanding and precise methods, such as the regularized LS one proposed here.

1. Introduction

Systems are generally assumed to behave linearly and in a noise-free environment. This is in practice never perfectly the case. First, nonlinear dynamic behaviors are very common in real life systems [12,15]. Second, the presence of noise is a natural phenomenon that is unavoidable for all experimental measurements. In order to perform reliable model estimation of such systems, one should thus keep in mind these two issues and take care about them. Indeed, all the noise that is not correctly removed from the measurements could be misinterpreted as nonlinearities, thus polluting measurements. And if nonlinearities are not accurately estimated, they will end up within the noise signal and information about the system under study will be lost.

* Corresponding author.

E-mail address: marc.rebillat@ensam.eu (M. Rebillat).

The problem addressed here is the estimation of nonlinear models of real life systems [20,10,5]. In order to build a nonlinear model, some approaches are based on a physical modeling of the structure whereas some perform without any physical assumption (black-box models). As nonlinear mechanisms in structures are complex and various and as it is not intended here to build a model for each case, it is chosen to rely on black-box models. Among these black-box approaches, some assume a given form for the selected model (block-oriented models [2,10,5,28]) whereas some do not put constraints on the model structure. Because block-oriented models can be interpreted easily, this class of models has been retained. A class of block-oriented models that is particularly interesting is the class of Parallel Hammerstein Models (PHM, see Fig. 1). It belongs to the class of "*Sandwich models*" and is shown to possess a reasonable degree of generality [20,5].

In a PHM, each branch is composed of a nonlinear static polynomial element followed by a linear one, $h_n[k]$, as shown in Fig. 1 where k denotes the discrete-time in samples. The relation between the input $e[k]$ and the output $s[k]$ of such a system is given by Eq. (1), where $*$ denotes the convolution product:

$$y[k] = h_0 + \sum_{n=1}^{N} h_n * u^n[k]. \tag{1}$$

In this model, each impulse response $h_n[k]$ is convolved with the input signal raised to its nth power and the output $s[k]$ is the sum of these convolutions. h_0 stands for the constant offset. The first impulse response $h_1[k]$ represents the linear response of the system. The other impulse responses $\{h_n[k]\}_{n\in\{2\cdots N\}}$ model the nonlinearities. The family $\{h_n[k]\}_{n\in\{1\cdots N\}}$ will be referred to as the kernels of the model. The focus is put here on the estimation of the kernels of PHM. In terms of real life applications, PHMs have already proven their usefulness in modeling power amplifiers [11,8], loudspeakers [24], damaged structures [17,23], or digital audio effects [1].

One way to estimate PHM relies on the fact that the estimation problem is linear in the parameters and thus that classical least squares (LS) estimation algorithms can be used. A popular approach is thus to pass the input through a parallel connection of some nonlinear basis functions that are each followed by a finite impulse response (FIR) filter [8]. The corresponding multiple-input single-output problem is then solved using standard FIR identification method. For example, Gallman [9] used Hermite polynomials as orthogonal basis functions with Gaussian inputs. The main drawbacks of these approaches is that the user gets no physical insight in the number of parallel branches in the system under test, and that for systems with long memories a large number of parameters is needed due to the FIR-nature of the model. To limit these drawbacks Schoukens et al. [30,29] proposed to couple a best linear approximation [22] at different excitation levels with a singular value decomposition and a rational transfer function parametrization LTI subblocks. An alternative method relies on Volterra series analytical method and wavelet balance method under multilevel excitations [3]. In that area, this article introduces a regularized LS identification algorithm inspired on the recently developed regularized impulse response estimation techniques [21,14].

Another means to estimate PHM consists in using exponential sine sweeps (ESS) based methods. This idea for the analysis of nonlinear system has been first presented by Farina [6] and is sometime referred to as the "*nonlinear convolution*" [7]. This method has been later formally demonstrated [19,18,24] and it is now well established that almost nonparametric versions of PHM can be very easily and rapidly estimated using it. However, one drawback of these ESS based methods is that the number of parallel branches is chosen arbitrarily. Nevertheless, this drawback has been overcome by Rébillat et al. [25] making this estimation method now fully nonparametric.

These two classes of methods (LS and ESS) are founded on radically different mathematical backgrounds but have been developed to solve the same issue. A methodology is thus proposed here to compare them with respect to (i) their accuracy, (ii) their computational cost, and (iii) their robustness to noise. Tests are performed on simulated systems for several values of methods respective parameters and of signal to noise ratio. The paper is organized as follows: the regularized least squares method and the exponential sine sweep method are introduced in Sections 2 and 3. The systems under study and the methodology proposed to compare both methods are presented in Section 4. Results are then provided in Section 5 before concluding in Section 6.

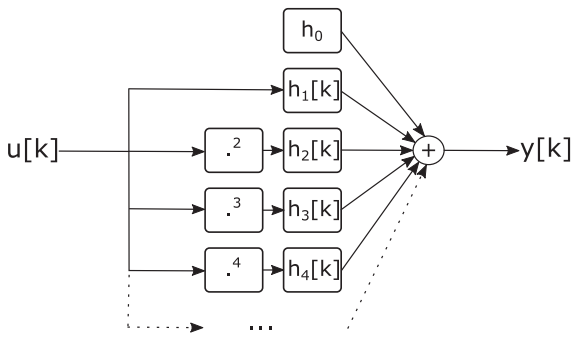

Fig. 1. Representation of Parallel Hammerstein Models (PHM).

2. Regularized Least Squares method (RLS)

2.1. Linear least squares formulation

The problem of identifying a PHM can be formulated as a linear-in-the-parameters problem when FIR representations, or other basis function expansion LTI representations, are used to model the LTI blocks $\{h_n[k]\}_{n\in\{1\cdots N\}}$ [8]. Indeed when Eq. (1) is elaborated on further, one obtains:

$$y[k] = h_0 + \sum_{n=1}^{N}\sum_{\tau=0}^{n_h} h_n[\tau]u^n[k-\tau], \tag{2}$$

where n_h is the length of the impulse response of the system under test. It can be observed that Eq. (2) can be written under matrix notation as:

$$y = K\theta, \tag{3}$$

where y is the vector containing all time measured time instances of $y[k]$, K is the regressor matrix depending on the applied input signal, and θ is the parameter vector containing the impulse responses to be estimated:

$$y = \begin{bmatrix} y[0] \\ y[1] \\ \vdots \\ y[T-1] \end{bmatrix} \in \mathbb{R}^{T\times 1} \tag{4}$$

$$K = \begin{bmatrix} 1 & K_h(u) & K_h(u^2) & \dots & K_h(u^N) \end{bmatrix} \in \mathbb{R}^{T\times N(n_h+1)+1} \tag{5}$$

$$K_h(x) = \begin{bmatrix} x[0] & x[-1] & \dots & x[-n_h] \\ x[1] & x[0] & \dots & x[1-n_h] \\ \vdots & \vdots & \ddots & \vdots \\ x[T-1] & x[T-2] & \dots & x[T-n_h-1] \end{bmatrix} \tag{6}$$

$$\theta = [\,h_0 \quad h_1[0] \quad h_1[1] \quad \dots \quad h_1[n_h] \quad \dots \quad h_N[0] \quad h_N[1] \quad \dots \quad h_N[n_h]\,]^T \in \mathbb{R}^{N(n_h+1)+1\times 1}, \tag{7}$$

where $.^T$ stands the transpose operator. It is straightforward to observe that the least squares estimate of the parameters $\hat{\theta}_{ls}$ can be obtained by:

$$\hat{\theta}_{ls} = \left(K^T K\right)^{-1} K^T y. \tag{8}$$

2.2. Introducing prior knowledge through regularization

In the previous section, the FIR are estimated using a linear least-squares approach. However, this approach does not take the prior knowledge available on the system into account. For instance, it is in many cases known that the system impulse response is exponentially decaying (stability), and highly correlated (smoothness). This type of prior knowledge can be imposed using a regularized linear least-squares approach. This approach has been successfully developed and applied in the linear time-invariant case [21,14], and has already been extended towards the identification of Hammerstein systems [26].

The regularized least squares approach introduces an extra term in the cost function that penalizes the parameters such that smoothness and exponential decay is obtained. The regularized cost function is given by:

$$V_r = ||y - K\theta||^2 + \theta^T R\theta. \tag{9}$$

The regularization matrix R is implemented here as a block-diagonal matrix using the TC kernel [21,14]:

$$R = \begin{bmatrix} 0 & & & 0 \\ & R_h & & \\ & & \ddots & \\ 0 & & & R_h \end{bmatrix} \tag{10}$$

where $R_h \in \mathbb{R}^{N(n_h+1)+1\times N(n_h+1)+1}$:

$$R_h(i,j) = \beta \frac{a}{\alpha^{(i+j)/2}(1-\alpha)} \tag{11}$$

$$a = \begin{cases} 1, & \text{if } i = j = 1 \\ 1+\alpha & \text{if } i = j \neq 1 \\ -\sqrt{\alpha} & \text{if } |i-j| = 1 \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

This regularization approach relies on two hyperparameters α and β that respectively tune the smoothness and decay, and the importance of the prior knowledge in the cost function. The optimal selection of these hyperparameters is performed here using a grid search and cross-validation. The estimated parameters $\hat{\theta}_{reg}$ are now obtained as:

$$\hat{\theta}_{reg} = \left(K^T K + R\right)^{-1} K^T y. \tag{13}$$

2.3. Drawbacks and advantages

2.3.1. Linear least-squares approach

A critical step in the use of the linear least-squares approach is the model order selection: what is the optimal length n_h of the impulse response models and how many parallel branches need to be considered? The choice of the impulse response length is typically made by performing a validation step to avoid overfitting, or by using the Akaike Information Criteria (AIC) or Minimum Description Length (MDL) indicator [22]. The number of parallel branches (the highest degree of the nonlinearity N) can be set by the user, or obtained by performing a validation step.

Another major drawback of this approach is the high number of parameters that needs to be estimated. When a system with a fairly long memory needs to be modeled, a high number of taps n_h needs to be considered in the FIR model of each branch. This results in a high number of parameters and, as a consequence, a fairly high variance on the estimated parameters. Furthermore, the large nature of the regressor matrix, constructed using similar basis functions, makes it prone to numerical conditioning problems. Note that Eqs. (8) and (13) are implemented in practice using a QR decomposition in order to obtain a more robust estimator.

2.3.2. Regularized least-squares approach

The regularization approach is much less prone to overfitting: the regularization pushes the end of the estimated impulse towards zero. Hence, the selection of the optimal FIR length n_h is not that critical, it should be selected long enough to capture a couple of time constants of the system dynamics. The number of parallel branches however still need to be set by the user, or obtained by performing a validation step.

Although it seems at first sight that the regularized least-squares approach still results in a high number of parameters, and hence a fairly high variance on the estimated parameters. This is not the case. The variance on the parameters is reduced significantly due to the regularization. The effective number of parameters is much lower than $N(n_h+1)+1$ [13]. This comes at a cost of a small bias introduced by the regularization [21].

The regularization approach introduces two hyperparameters α and β tuning the smoothness and decay, and the importance of the prior knowledge in the cost function. The optimal selection of these hyperparameters is performed here using a grid search and cross-validation. The search for the optimal hyperparameters increases the computational cost of the regularized least squares approach significantly compared to the linear least squares approach. Note that the same regularization matrix R_h is used for each branch of the parallel Hammerstein structure to keep the number of hyperparameters as low as possible.

3. Exponential Sine Sweep method (ESS)

The exponential sine sweep method developed to estimate PHMs is briefly recalled here [19,18,24,25].

3.1. Exponential sine sweeps

To experimentally cover the frequency range over which the system under study has to be identified, cosines with time-varying frequencies are commonly used. When the instantaneous frequency of $u[k] = \cos[\phi[k]]$ is increasing exponentially from f_1 to f_2 in a time interval T sampled with the sampling frequency f_s, this signal is called an "Exponential Sine Sweep". It can be shown [24,25] that by choosing $T_m = (2m\pi - \pi/2)ln(\frac{f_2}{f_1})/2\pi f_1$ with $m \in \mathbb{N}^*$ one obtains the following property:

$$\forall n \in \mathbb{N}^*, \quad \cos[n\phi[k]] = \cos[\phi[k+\Delta_n]] \quad \text{with} \quad \Delta_n = \frac{f_s T_m \ln(n)}{\ln\left(\frac{f_2}{f_1}\right)}. \tag{14}$$

Eq. (14) states that for any exponential sine sweep of duration T_m, multiplying the phase by a factor n yields the same signal, advanced in time by Δ_n samples.

3.2. Kernel recovery in the time domain

If an exponential sine sweep is presented at the input of a PHM, by combining Eqs. (1) and (14) and by using properties of Chebyshev polynomials [19,18,24,25], one obtains the following relation:

$$y[k] = \sum_{n=1}^{N} (g_n * u)[k + \Delta_n] \quad \text{with} \quad g_n[k] = \sum_{i=1}^{N} \tilde{c}_{i,n} h_i[k] \tag{15}$$

where $g_n[k]$ can be interpreted as the contribution of the different kernels to the nth harmonic and $\tilde{c}_{i,n}$ is the coefficient (n, i) of the matrix $\widetilde{\mathbf{C}}$. Details of the computation of the matrix $\widetilde{\mathbf{C}}$ are provided for example in Rébillat et al. [24].

In order to identify each kernel $h_n[k]$ separately, a signal $u^{-1}[k]$ operating as the inverse of the input signal $u[k]$ in the convolution sense can be built. After convolving the output of PHM $y[k]$ given in Eq. (15) with $u^{-1}[k]$, one obtains Eq. (16):

$$g[k] = (y * u^{-1})[k] = \sum_{n=1}^{N} g_n[k + \Delta_n]. \tag{16}$$

Because $\Delta t_n \propto \ln(n)$ and $f_2 > f_1$, the higher the order of nonlinearity n, the more advanced is the corresponding $g_n[k]$. Thus, if T_m is chosen long enough, the different $g_n[k]$ do not overlap in time and can be separated by simply windowing them in the time domain. Using Eq. (15), the family $\{h_n[k]\}_{n \in \{1 \cdots N\}}$ of the kernels of the PHM under study can then be fully extracted as

$$[h_1[k] \cdots h_N[k]]^T = \widetilde{\mathbf{C}} [g_1[k] \cdots g_N[k]]^T. \tag{17}$$

3.3. Autonomous determination of the number of kernels

An autonomous procedure has been proposed that determines the effective number of kernels which can be distinguished from noise [25]. Let's consider $g[k]$ as defined in Eq. (16) in a short time interval $\mathcal{I}_n$ centered around the expected occurrence of the n-th harmonic contribution, i.e. $k \in \mathcal{I}_n = [k_{n,1}, \ldots, k_{n,2}]$ with $\Delta_n \leqslant k_{n,1}$ and $k_{n,2} < \Delta_{n-1}$. Let's consider $n[k]$ a noise record in a short time interval than can pratically be extracted from $g[k]$. The question is whether $g[k]$ significantly protrudes from background noise $n[k]$ in that interval – and therefore originates from a nth nonlinearity – or whether it is mainly noise. This corresponds to two alternative hypotheses:

- H0: $g[k], k \in \mathcal{I}_n$ is noise only
- H1: $g[k], k \in \mathcal{I}_n$ contains an nth order harmonic contribution $g_n[k]$.

An estimate of the variance in interval $\mathcal{I}_n$ under H0 is returned by

$$\hat{\sigma}^2_{\mathcal{I}_n|H0} = \frac{1}{k_{n,2} - k_{n,1} + 1} \sum_{k=k_{n,1}}^{k_{n,2}} n^2[k] \tag{18}$$

Therefore, assumption H0 is to be rejected if the current "variance"

$$\hat{\sigma}^2_{\mathcal{I}_n} = \frac{1}{k_{n,2} - k_{n,1} + 1} \sum_{k=k_{n,1}}^{k_{n,2}} g^2[k] \tag{19}$$

is found statistically greater than $\hat{\sigma}^2_{\mathcal{I}_n|H0}$. This can be easily formalized with a F-test for the equality of two variances: H0 is rejected if the ratio $\hat{\sigma}^2_{\mathcal{I}_n} / \hat{\sigma}^2_{\mathcal{I}_n|H0}$ is greater than $F(\alpha)_{K-1,K-1}$, the critical value of the F distribution with $K - 1$ and $K - 1$ degrees of freedom and a significance level of α. K is here defined as the number of samples contained within the time interval $\mathcal{I}_n = [k_{n,1}, \ldots, k_{n,2}]$. The highest-order harmonic contribution $g_N[k]$ that is found significant through this procedure then provides the model order N. It is important to note that Gaussian distributions are assumed here and that may not be exactly the case in practice.

3.4. Advantages and drawbacks

The method described here easily provides a direct mathematical access to all the kernels of a PHM. The main advantage of this method is to be fast and simple: using only one exponential sine sweep one can have direct access to the kernels of an arbitrary vibrating device without the need of any complicated signal processing procedure (convolution, time windowing

and matrix multiplication). However, the input sine sweep used here has to satisfy some specific requirements. First of all, the input sine sweep must have its frequency varying exponentially with time. It is important to keep in mind that at the cost of computational time, solutions exist that allow to relax this assumption [27]. Secondly, in the framework presented here [24], the length of the exponential sine sweep cannot be set arbitrarily and must be chosen among a given set of admissible values. Again, it is important to notice that this assumption is not necessary and can be very easily relaxed by choosing differently the initial and final phases of the exponential sine sweep [18]. A drawback of this method is thus the relatively low flexibility of its input signal. Another limitation of this method lies in its unability to distinguish between nonlinearities and experimental noise. This could result in an overestimation of the nonlinear behavior as estimated kernels also include some plant noise. Finally, recent studies [31,4] have shown that distortion artifacts can appear in the causal part of the impulse responses estimated by means of exponential sine sweep when the system under study is not perfectly modeled by a PHM. However, as mentioned by references [31,4] those artifacts remain of very small amplitude.

4. Comparison methodology

The two methods (LS and ESS) described in Sections 2 and 3 are founded on radically different mathematical backgrounds but are expected to solve the same issue. A methodology is proposed here to compare them with respect to (i) their accuracy, (ii) their computational cost, and (iii) their robustness to noise for different systems presented in the following section.

4.1. Systems under study

This section introduces the three simulated systems that have been chosen to compare the two methods presented in Sections 2 and 3. The first one corresponds to a classical PHM system with low time constants, the second one is a PHM system with a delay and longer time constants, and the third one is a hysteretic system with a dynamic nonlinearity that cannot be mathematically represented as a PHM but only approximated by this structure.

4.1.1. System #1: PHM system with low time constants

The first system considered here constitutes an academical case study over which the methods will be compared. The parameters used to define it are provided in Table 1 and the impulse responses of each branch are represented in Fig. 2. This system of order $N = 4$ has been chosen as a basic example aiming at demonstrating the global performances of both methods.

4.1.2. System #2: PHM system with delay and longer time constants

The second system considered here is different from the first one as here two important points are investigated. The first point is related to the presence of delays in the impulse responses of each branch, this is something relatively common in practice and this can affect the performance of both methods. The second point is the fact that systems under study can have longer response times, which means in practice that more samples have to be estimated by both methods. A cascade of Hammerstein models of order $N = 4$ has again been chosen. For its Kernels, impulse responses of rational filters (2 poles and 2 zeros) have been chosen. All the parameters are provided in Table 2 and the impulse responses of each branch are represented in Fig. 3. This system has been designed in order to have a longer response time than the system #1 as well as to present a delay.

4.1.3. System #3: Hysteretic system with a dynamic nonlinearity

Hysteresis is a phenomenology commonly encountered in very diverse engineering and science disciplines. A system is said to be hysteretic when it exhibits an input–output loop as the input frequency approaches zero. This kind of system is inherently nonlinear and cannot be represented exactly by a PHM. This system has been chosen here as a mean to assess the performances of both methods for cases where the system to be identified does not correspond exactly to PHM. This is always the case in practical applications as PHMs only constitute an idealization of reality. The implementation chosen here belongs to the benchmarks proposed at the "*Workshop on Nonlinear System Identification Benchmarks*" held in Brussels in 2016. The implementation details of this benchmark are provided in Noël and Schoukens [16] and are briefly recalled here.

The vibrations of the Bouc-Wen oscillator with a single mass simulated here are governed by Newton's law of dynamics written as follows in continuous time domain:

Table 1
Parameters of the simulated PHM system #1. Matlab R2016b functions used to generate the filter coefficients are reported in the "*Parameters*" column. The sampling frequency is chosen as $f_s = 10^5$ Hz.

Order	Gain	Filter type	Filter order	Parameters
1	1	Lowpass Chebychev	3	$R = 3, W_p = 0.2$ (*cheby1*)
2	0.5	Lowpass Butterworth	2	$W_n = 0.15$ (*butter*)
3	0.25	Lowpass Chebychev	3	$R = 30, W_{st} = 0.3$ (*cheby2*)
4	0.15	Lowpass Butterworth	4	$W_n = 0.05$ (*butter*)

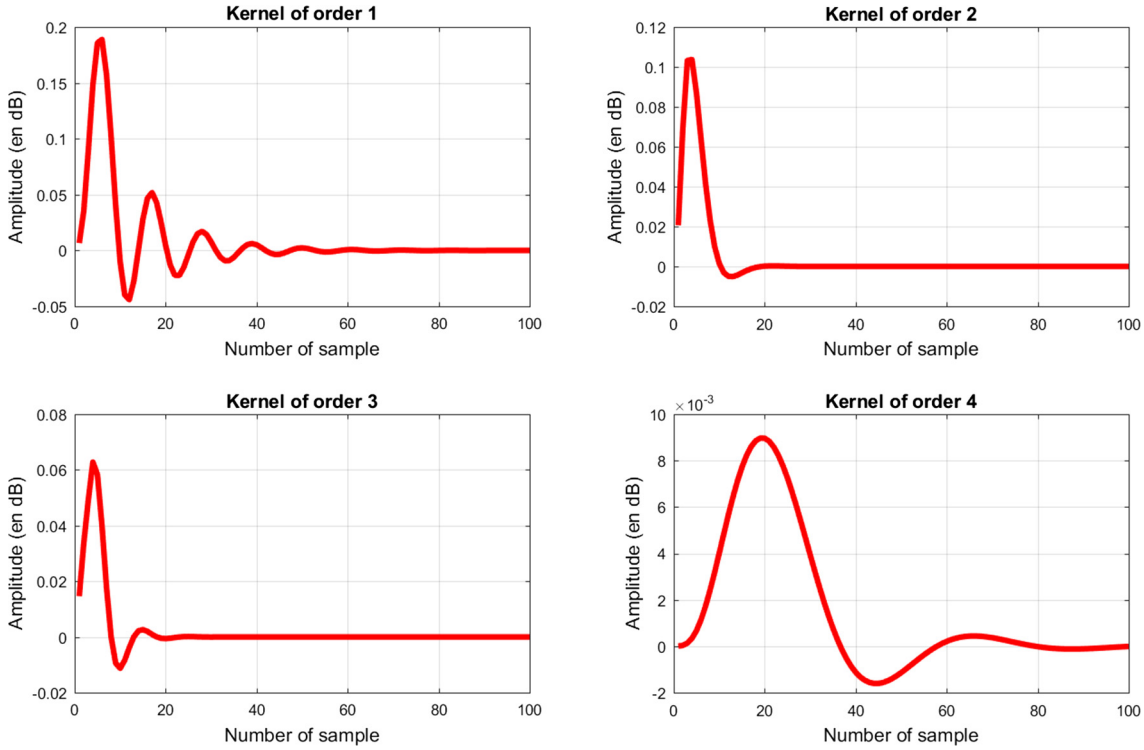

Fig. 2. Impulse responses of the four branches of the PHM system #1.

Table 2
Parameters of the simulated PHM system #2. The sampling frequency is chosen as $f_s = 10^5$ Hz. A fixed delay of 50 samples is imposed on each impulse response.

n	f_{zeros} (kHz)	$\lvert p_{\text{zeros}}\rvert$	f_{poles} (kHz)	$\lvert p_{\text{poles}}\rvert$	Gains
1	0.15	0.95	1.5	0.97	1
2	0.4	0.92	4	0.95	0.5
3	1	0.93	0.1	0.96	0.25
4	2.2	0.9	0.5	0.97	0.3

$$m\ddot{y} + c\dot{y} + ky + z(y,\dot{y}) = u \tag{20}$$

where m is the mass constant, k and c the linear stiffness and viscous damping coefficients, y the displacement, u the external force, and an over-dot indicates a derivative with respect to time. The dynamic, i.e. history-dependent, nonlinear term $z(y,\dot{y})$ encodes the hysteretic memory of the system and obeys the following first-order differential equation:

$$\dot{z}(y,\dot{y}) = \alpha\dot{y} - \beta\ \ \gamma|\dot{y}||z|^{\nu-1}z + \delta\dot{y}|z|^{\nu} \tag{21}$$

The five Bouc-Wen parameters $\alpha, \beta, \gamma, \nu$ and δ are used to tune the shape and the smoothness of the system hysteresis loop. All the parameters values are provided in Table 3. These equations are integrated in time using a Newmark method.

The sampling frequency is chosen as $f_s = 2500$ Hz for this system.

4.2. Performances indexes

In order to compare both methods several performance indexes (PI) have been considered and are presented below.

4.2.1. PI_1: Relative difference between the reconstructed signal and the output signal for an exponential sine sweep input

The first performance index considered here aims at assessing the ability of the estimated PHM model to reconstruct an output signal that is in agreement with the actual output signal for the case of an exponential sine sweep input. Let $y_{\text{ESS}}[k]$ and $\hat{y}_{\text{ESS}}[k]$ be the actual and estimated output signals over K samples. This performance index is defined as:

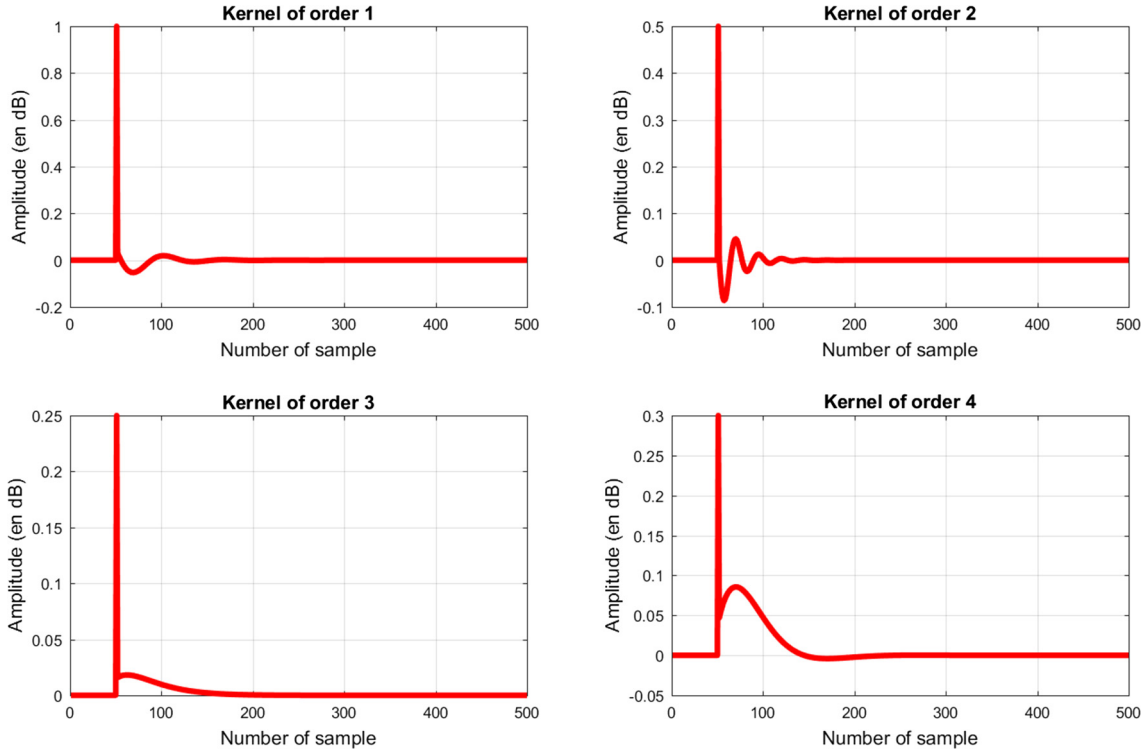

Fig. 3. Impulse responses of the four branches of the PHM system #2.

Table 3
Parameters of the simulated hysteretic Bouc-Wen system.

Parameter	m	c	k	α	β	γ	δ	ν
Value (units in SI)	2	10	5×10^4	5×10^4	1×10^3	0.8	−1.1	1

$$PI_1 = \frac{\sum_{k=0}^{K}[\hat{y}_{\mathrm{ESS}}[k] - y_{\mathrm{ESS}}[k]]^2}{\sum_{k=0}^{K}[y_{\mathrm{ESS}}[k]]^2}$$

This performance index can be computed for systems #1, #2, and #3.

4.2.2. PI_2: Relative difference between the reconstructed signal and the output signal for a noise input

The second performance index considered here aims at assessing the ability of the estimated PHM model to reconstruct an output signal that is in agreement with the actual output signal for the case of a noise input. Indeed, an exponential sine sweep is an input signal that is favorable for the ESS method but less for the RLS one, whereas the opposite is true for the noise input signal. It is thus important to consider the abilities of both methods with respect to that kind of inputs too. Let $y_{\mathrm{NOISE}}[k]$ and $\hat{y}_{\mathrm{NOISE}}[k]$ be the actual and estimated output signals over K samples. This performance index is defined as:

$$PI_2 = \frac{\sum_{k=0}^{K}[\hat{y}_{\mathrm{NOISE}}[k] - y_{\mathrm{NOISE}}[k]]^2}{\sum_{k=0}^{K}[y_{\mathrm{NOISE}}[k]]^2}$$

This performance index can be computed for systems #1, #2, and #3.

4.2.3. PI_3: Relative difference between the estimated and original PHM kernels

The third performance index deals with the ability of both estimation methods to estimate PHM kernels that are close to the original ones, when available. For system of order N, given estimated kernels $\{\hat{h}_n[k]\}_{n\in[1:N]}$ and original kernels $\{h_n[k]\}_{n\in[1:N]}$, this performance index is defined as:

$$PI_3 = \frac{1}{N}\sum_{n=1}^{N}\left(\frac{\sum_{k=0}^{K}\left[\hat{h}_n[k] - h_n[k]\right]^2}{\sum_{k=0}^{K}[h_n[k]]^2}\right)$$

This performance index can be computed only for systems #1 and #2. As the system #3 may not be represented exactly as a PHM, the notion of "*original kernels*" makes no sense for it.

4.3. Experimental plan

The aim is to compare the two methods with respect to their accuracy, their computational cost, and their robustness to noise for the different systems presented in the following section. There exists thus two types of free parameters within the experimental plan. The first type of free parameters will be linked with the methods. The second type of free parameters is linked to the input signal. All the free parameters are listed in Table 4. The different methods that will be compared here and their associated free parameters are listed in Table 5.

Among the available parameters, some are not free within the chosen experimental plan. Excepted for the nonparametric exponential sine sweep method where it is found automatically, the order of nonlinearity of the estimated PHM is set to $N = 4$ for all methods. This ensures optimal results for system #1 and #2 as it effectively corresponds to the model structure. For the input signals, care has been taken to ensure that the exponential sine sweep and the white gaussian noise have the same number of samples and the same RMS value. For the exponential sine sweep, its start and stop frequencies are always $0.001f_s$ and $0.4f_s$ and its amplitude is unitary for all systems. The parameter $N^{\max}$ corresponds to the maximum model order reachable by the NP-ESS when estimating N. This parameter has been chosen as $N^{\max} = 24$. The parameter $N_{\text{FIR}}^{\max}$ corresponds to the maximum FIR length reachable by the RLS method when estimating N_{FIR}. This parameter has been chosen as equal to the maximum FIR length tested for the DLS method for each system. For all the sets of free parameters $N_{\text{rep}} = 10$ repetitions have been performed. Computations have been run on a computing station equipped with 128 Go of random access memory and with 2 Intel Xeon ES 2630 processors running at 2.6 GHz, each with 6 cores and with 15 MB of cache memory.

5. Results

This section presents the results comparing the regularized least squares method and the exponential sine sweep method presented in Sections 2 and 3 applied to the systems under study using the methodology detailed in Section 4.

5.1. System #1

As stated before, the first system considered here constitutes an academical case study over which the methods will be compared (see Section 4.1.1). Before comparing both methods, it is interesting to assess the influence of the various free parameters linked with the methods on the computation time and on the performance indexes. From Table 4, the free parameters related to the methods are: T the duration of the input signal and N_{FIR} the number of samples retained in each FIR for the DLS method only.

With respect to computational time, it is straightforward that the larger T or N_{FIR}, the larger the computation time. Indeed it is well known that the linear least-squares algorithm has a computational complexity of which grows in TN_{FIR}^2. This computational complexity is further increased for the regularized least squares case since the hyperparameters need to be estimated too. Regarding the ESS-based methods, the most demanding computational step is a convolution that has a complexity in $O(T\log(T))$ when using FFT-based algorithms.

Results highlighting the influence of T and N_{FIR} on the various performance indexes (PIs) are provided in Fig. 4. From the left column of this figure, it can be observed that whatever the considered method and the considered PIs, the increase of the input signal duration leads to a decrease of the PI. This is rather logical that the methods perform better when more points are available. However, it can be seen that for the DLS method making use of the ESS as input signal, this is not true: PIs exhibit an increasing tendency with increasing input length. This could be attributed to numerical conditioning issues,

Table 4
List of the free parameters. ESS: Exponential sine Sweep. WGN: White Gaussian Noise.

Name	Description	Values
T	Duration of the input	0.05–5 s
N	Order of nonlinearity	4
N_{FIR}	FIR length	#1: 10 to 100 samples #2: 120–210 samples #3: 250–400 samples
$u[k]$	Input signal	ESS or WGN
SNR	Signal to noise ratio	0 dB and −60 dB

Table 5
Compared methods with their associated free parameters.

1	Non parametric exponential sine sweep method	NP-ESS	$T, N^{\max}$
2	Parametric exponential sine sweep method	ESS	T, N
3	Direct least-squares method	DLS	$T, N, N_{\text{FIR}}, u[k]$
4	Regularized least-squares method	RLS	$T, N, N_{\text{FIR}}^{\max}, u[k]$

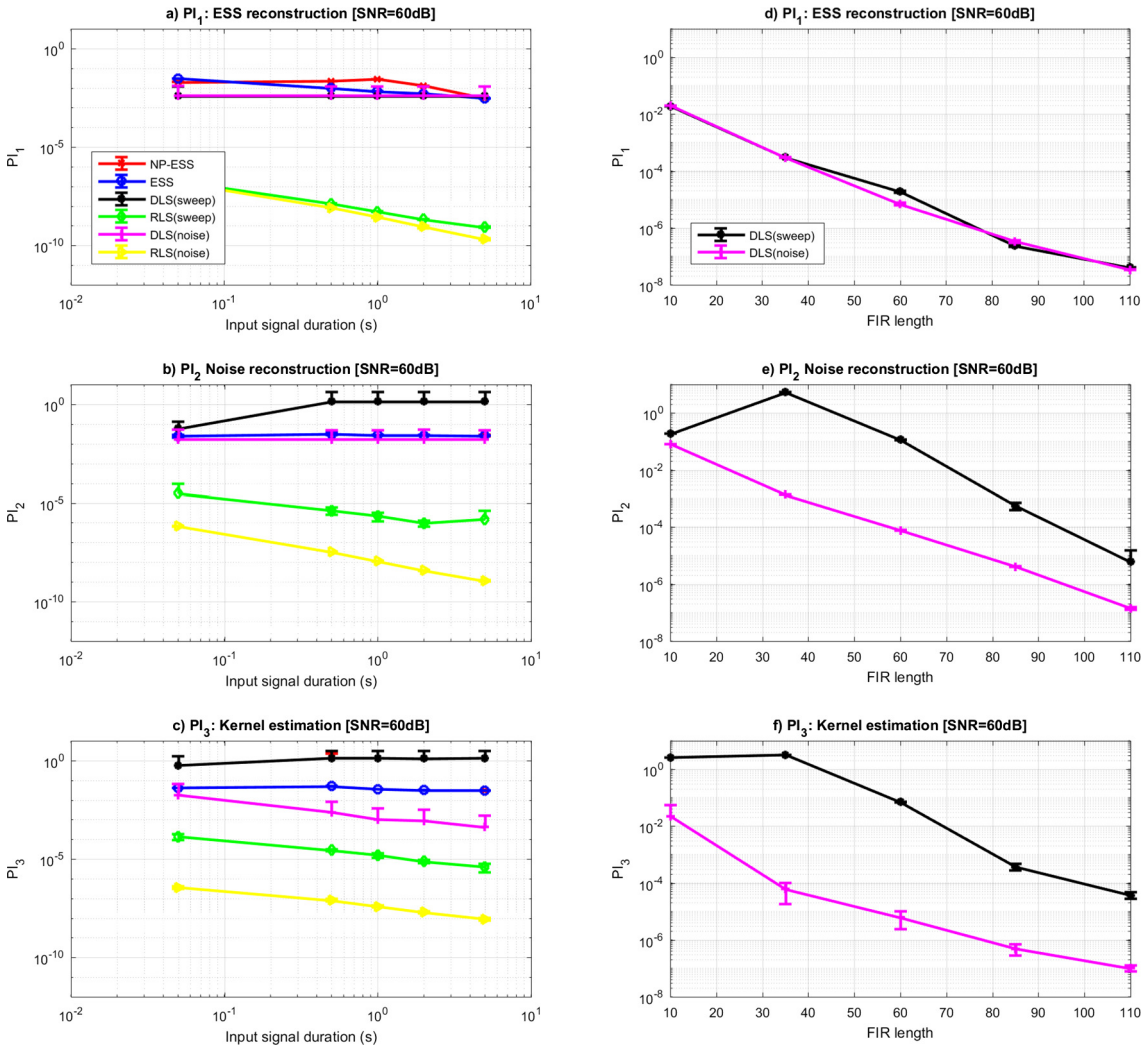

Fig. 4. Influence of the input signal duration T (left column: a,b,c) and of the FIR length N_{FIR} (right column: c,d,e) on the various performance indexes for system #1 at SNR = 60 dB. The average values of the PIs ± 1 standard deviation are plotted. See Table 5 for the method name abbreviations.

keeping in mind the fact that the ESS is not the optimal input for the DLS method. For the DLS methods, the influence of N_{FIR} on the various PIs is shown on the right column of Fig. 4. It can be observed that for each PIs, increasing the FIR length leads to a decrease of the PI, which was also expected.

The methods can also be compared in a more general manner as presented in Fig. 5. On this figure, one point presents the result obtained by one method, for one run and for a given combination of free parameter without any additional post-processing. Results are presented for SNR= 0 dB on the left column and for SNR = 60 dB on the right column. Each line corresponds to a given PI as defined in Section 4.2.

Non-parametric exponential sine sweep method. The points corresponding to the *non-parametric exponential sine sweep method (NP-ESS)* are shown as red crosses on Fig. 5. From this figure, it can be seen that regarding PI_1, i.e. the ability of the method to reconstruct an ESS, ESS and NP-ESS methods have similar performances in terms of computation time and

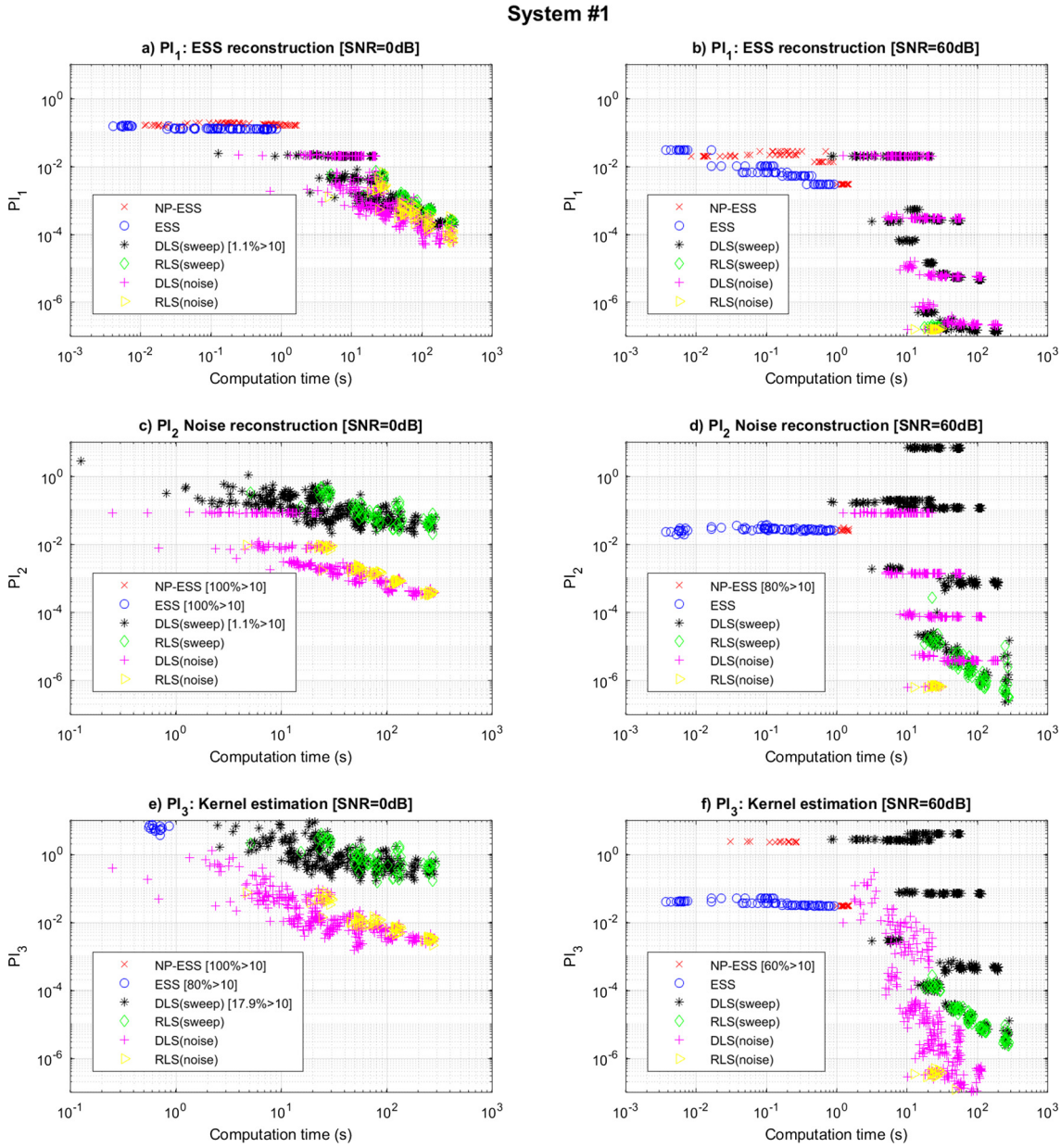

Fig. 5. Comparison results for system #1. On this figure, one point presents the result obtained by one method, for one run and for a given combination of free parameter without any additional post-processing. The DLS results are shown for all FIR lengths considered. Results are presented for SNR = 0 dB on the left column and for SNR = 60 dB on the right column. Each line corresponds to a given PIs as defined in Section 4.2. See Table 5 for the method name abbreviations. The indications $[X\% > \alpha]$ in the legend indicate the percentage of point that lie above the maximum plotted value α, *i.e.* out of the figure.

of precision. An improvement of the estimation quality with the SNR is also clearly visible. However, this is not true for PI_2 which is related to noise reconstruction. For both SNR values, the NP-ESS is not able to provide acceptable reconstruction results (acceptable results only for SNR = 60 dB and $T = 5$ s). Regarding the ability of the NP-ESS to estimate the Kernels of the system, it can be observed that in the low SNR condition (0 dB) the method fails but in the high SNR condition (60 dB) it provides fair results, specially for longer input signal length. Furthermore one can also notice that, for a given computation time, the NP-ESS method provides results of equal quality than the DLS method for a SNR = 60 dB.

Exponential sine sweep. The points corresponding to the *exponential sine sweep method (ESS)* are shown as blue circles on Fig. 5. From this figure, it can be observed that many of the commentaries done for the NP-ESS method are also true for the ESS method. The main difference between the two methods can be seen for PI_3. For this PI, the ESS method is shown to

perform better than the NP-ESS one. This could be attributed to the fact that for the ESS method, the order of the system $N = 4$ is set up in advance whereas for the NP-ESS, this parameter is estimated. Thus, if this parameter is not correctly estimated (which is the case for SNR = 0 dB), the method fails in reconstructing the system's Kernels. Again, we can emphasize the fact that for a given computational time the ESS method provide results of equal quality to the ones provided by the DLS one for ESS and WGN reconstruction. Finally, the cost of the estimation of N can be seen as negligible making the ESS and NP-ESS method comparable in terms of computation times.

Direct least-square. The points corresponding to the *direct least-square method (DLS)* are shown as black stars for an ESS as an input signal and as pink plus for a WGN as input on Fig. 5. From this figure, it can be observed that the DLS has sometimes problems with conditioning using ESS as inputs (1.1% of the trials provide bad results). Furthermore the input signal appears to have a lower influence for the quality of the reconstructed ESS than for the quality of the reconstructed WGN. For the ESS and WGN reconstruction, a clear effect of the SNR can also be observed. As stated for the ESS and NP-ESS methods, the Kernel estimation is of low quality for SNR = 0 dB but becomes fair when SNR drops to 60 dB. Finally, noise seems to have an influence on computation time as it exhibits some variability with noise realization.

Regularized least-square. The points corresponding to the *regularized least-square method (RLS)* are shown as green diamonds for an ESS as an input signal and as yellow triangle for a WGN as input on Fig. 5. From this figure, it can be seen that the RLS method is the slowest one. This is due to the hyper-parameters selection that necessitates computational resources. Regarding quality, the RLS method is the one that provides the most precise and less variable results with respect to ESS and WGN reconstruction for both SNR cases. For the present case, the input signal appears to have a very low influence for ESS reconstruction but not for the WGN reconstruction. Again, Kernel estimation in the low SNR condition is not really successful for the RLS method but excellent for the high SNR condition.

5.2. System #2

The second system considered here is different from the first one as here two important point are investigated: the presence of delays in the impulse responses of each branches, and longer response times (see Section 4.1.2). The influence of the various free parameters linked with the methods on the computation time and on the performance indexes for this system are very similar to the one observed for system #1 in the previous section and thus will not be presented here. Furthermore, many general comments made for system #1 can also be made for system #2 and will thus not be repeated. The methods can be compared in a general manner for this system as presented in Fig. 6.

The main difference between the results obtained for the system #1 and the system #2 lies in the gap in computational time that appears with system #2. The main consequence of having longer time constants is that longer FIR filters are needed. This has a direct impact on the computation time of DLS and RLS methods but not on the computation time of the NP-ESS and ESS method that do not depend on this parameter. Another difference with the results obtained for system #1 is that there is more difference between the results obtained with different input signals for the DLS and RLS methods. Indeed for system #2, the results obtained using the WGN as input signal are much more precise than the ones obtained using the ESS. This is again a consequence of the fact that the ESS signal is not sufficiently exciting the system for the least-square based methods that are prone to numerical conditioning issues. These numerical issues are here particularly enhanced as the number of FIR coefficient to be estimated is larger than in the previous example.

5.3. System #3

The system #3 is a hysteretic system with a dynamic nonlinearity and has been chosen here as a mean to assess the performances of both methods for cases where the system to be identified does not correspond exactly to PHM (see Section 4.1.3). The influence of the various free parameters linked with the methods on the computation time and on the performance indexes for this system are very similar to the one observed for system #1 in previous sections and thus will not be presented here. Furthermore, many general comments made for system #1 and #2 can also be made for system #3 and will thus not be repeated. The methods can be compared in a general manner for this system as presented in Fig. 7.

The main observation on the results obtained for the system #3 lies in the gap in computational time that really increases. In order for PHM to approximate correctly system #3 longer FIR lengths are needed which leads to longer computational time. Again, the ESS and NP-ESS methods are exempt from that side effect. Another remark of interest is linked with the obtained reconstruction quality for ESS and WGN signals which is lower for the present system than for the other ones. This lower reconstruction quality can be explained by the fact that as PHM cannot capture the whole dynamical behavior of system #3 there will unavoidably remains some modeling error that cannot be lowered by any of the estimation methods. It is interesting to observe that for this system the four methods provide ESS reconstruction results having very similar precisions but at a much lower cost for ESS based methods. It is also important to realize that the filters used in the simulated systems are globally low-pass filters. With that in mind, one important point with respect to the chosen input signals is that the ESS has a limited bandwidth whereas the WGN potentially covers the full-spectrum of interest. In that sense, models estimated using WGN should perform better in a wider frequency range than when using ESS. Indeed, the ESS does not excite the system under its start frequency f_1 and above its stop frequency f_2 and consequently, the estimated impulse responses will miss

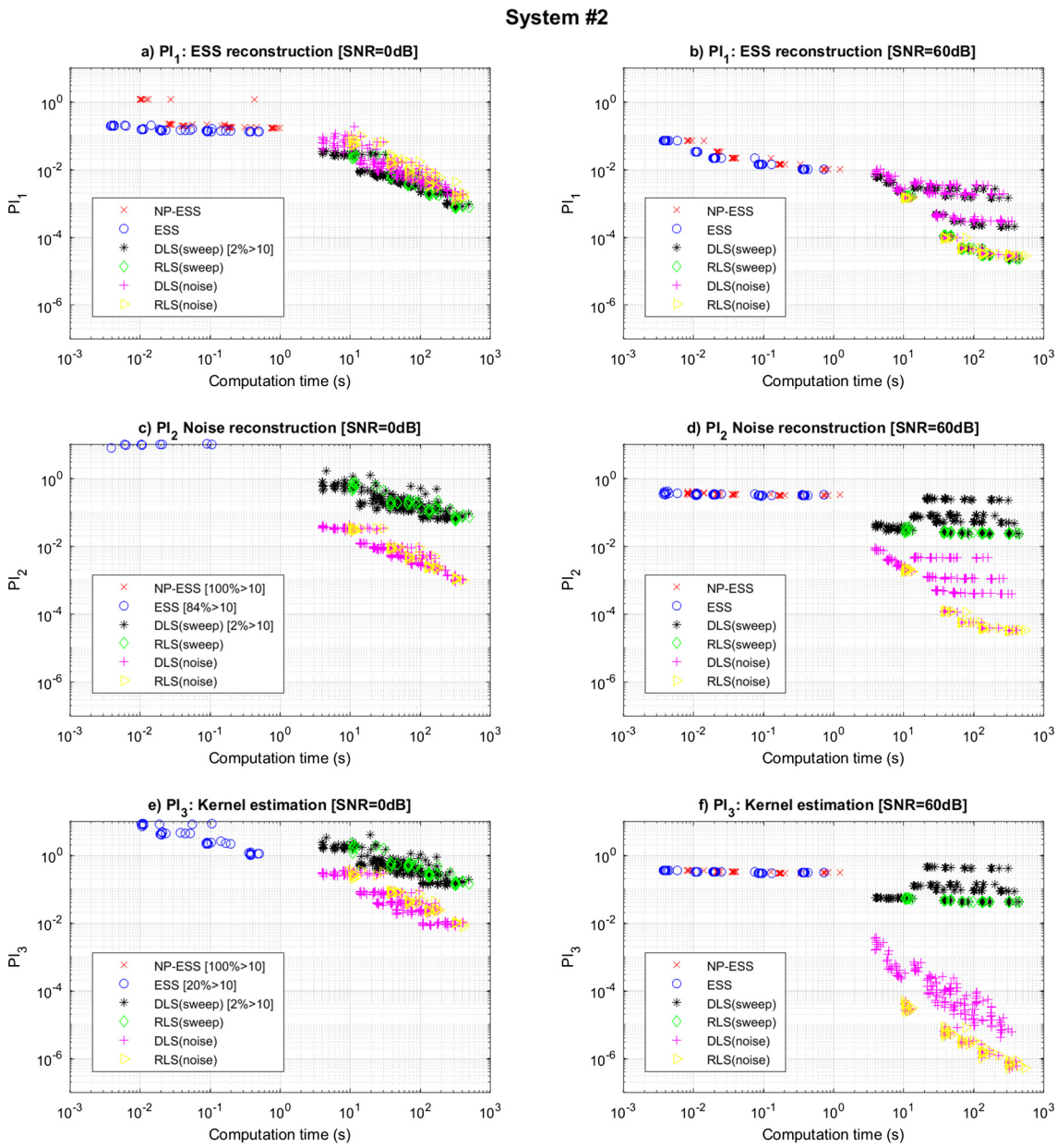

Fig. 6. Comparison results for system #2. On this figure, one point presents the result obtained by one method, for one run and for a given combination of free parameter without any additional post-processing. The DLS results are shown for all FIR lengths considered. Results are presented for SNR = 0 dB on the left column and for SNR = 60 dB on the right column. Each line corresponds to a given PIs as defined in Section 4.2. See Table 5 for the method name abbreviations. The indications $[X\% > \alpha]$ in the legend indicate the percentage of point that lie above the maximum plotted value α, *i.e.* out of the figure.

the information at frequencies lower than f_1 and higher than f_2. Considering that argument, it is thus expected that the results for the WGN are quantitatively in average better than the results for the ESS. The last comment is related to the impact of the input signal on the performances of the DLS and RLS methods. Previous observations were that both methods gave better results when using WGN as input signal as it is better suited to least-squares than the ESS. This is not true anymore for the present system. Indeed, ESS reconstruction results are here better when an ESS is used as input, and WGN reconstruction results are better when WGN is used as input. This highlights the fact that when trying to approximate a nonlinear system by an approximate model, the estimated model is valid only for a certain "*class*" of inputs that do not depart to much from the input used for the estimation.

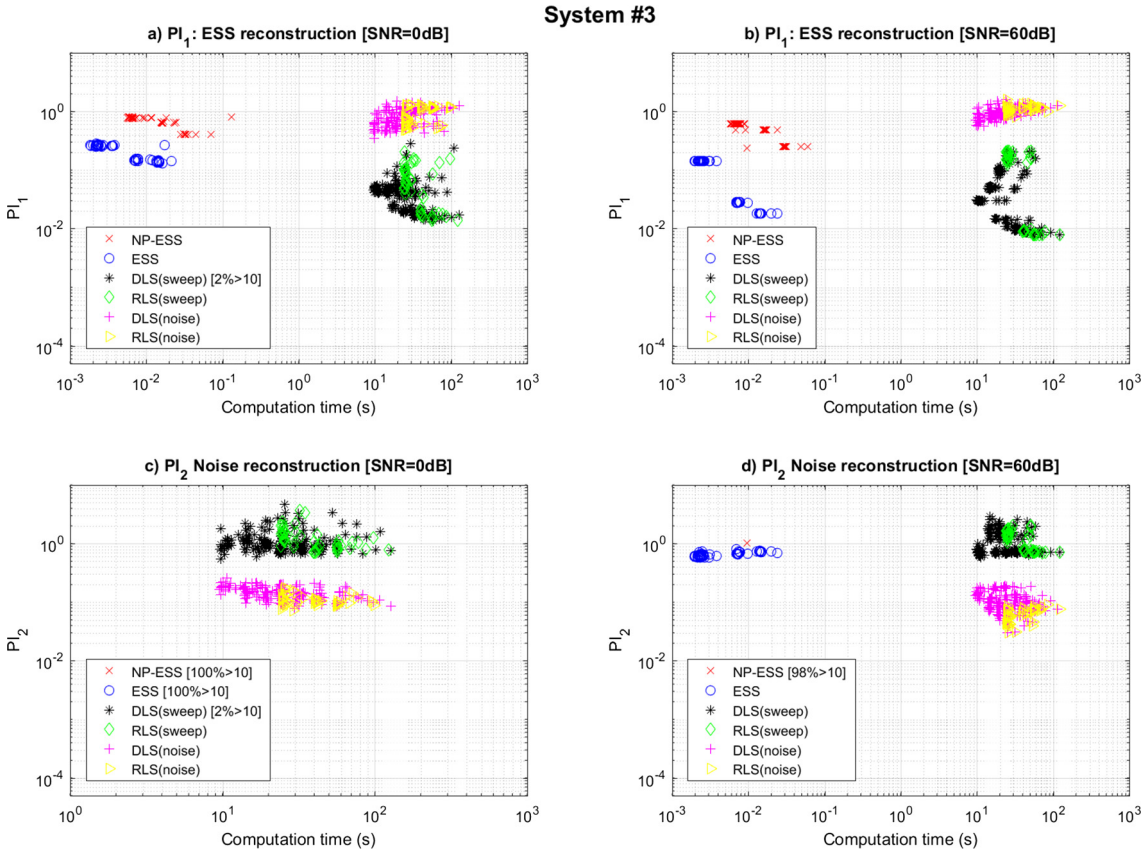

Fig. 7. Comparison results for system #3. On this figure, one point presents the result obtained by one method, for one run and for a given combination of free parameter without any additional post-processing. The DLS results are shown for all FIR lengths considered. Results are presented for SNR = 0 dB on the left column and for SNR = 60 dB on the right column. Each line corresponds to a given PIs as defined in Section 4.2. See Table 5 for the method name abbreviations. The indications $[X\% > \alpha]$ in the legend indicate the percentage of point that lie above the maximum plotted value α, *i.e.* out of the figure.

6. Conclusion

The objective of the present study is to compare ESS-based methods and LS-based methods, amongst which the regularized LS Parallel Hammerstein identification method which is introduced in this paper, in terms of computational cost, precision, robustness to noise with respect to their ability to estimate PHMs. The study has been conducted on three different systems and some general comments comparing the two type of methods can be drawn:

1. **The ESS-based methods are much less demanding than the LS-based ones in terms of computational resources.** Furthermore, their computational cost is independent of the system time constants. This comment is obviously dependent of the current method's implementations that are certainly perfectible. But even a more efficient implementation will not be able to bridge the computational gap existing between the two classes of methods.
2. **The LS-based methods, and more specifically the RLS one, provide results of higher quality than the ESS-based ones for a given realization.** However, for a given computational budget ESS-based methods and DLS methods can provide results of similar quality. The RLS based method overcome by far all the other methods in terms of quality.
3. **The LS-based methods seem more robust to noise than the ESS-based ones.** In presence of a large amount of noise, the DLS method can be subject to numerical conditioning issue and the NP-ESS method may be unable to estimate correctly the model order, thus leading to bad results. ESS-based methods and, more specifically the NP-ESS method appear as less robust to noise than the LS-based ones.
4. **The ESS-based methods need less parameters to be tuned than the LS-based ones.** For the NP-ESS, no parameters have to be given and for the ESS-one, only the model order N has to be set up in advance. For the RLS method, the model order N has to by provided by the user, while the FIR length needs to be selected long enough to capture the unknown system dynamics. These parameters cannot be set at an arbitrary high value as they impair directly the computational cost.
5. **The LS-based methods are less restrictive in terms of input signals than the ESS-based ones.** ESS-based methods rely exclusively on an ESS signal, which can be restrictive for certain applications. LS-based methods do not put constraints on the input signal. However it is observed that using WGN signals allows to get more precise results in cases where the

system under study is a PHM.This is probably due to the fact that WGN excites the system in a wider spectral range than the ESS. Despite of that, performances obtained using the two input signals are similar for the non-PHM nonlinear system under study here.

The generality of these comments has to be mitigated with the fact that only three systems have been studied here. However, one general comment that can be drawn from the previous analysis is that by relying on the fact that ESS-based methods are rapid, non-parametric and that their computation time is system independent, they can be used to get a first estimate of the system under study. From this estimate the various parameters needed to tune more precise methods (order of nonlinearity N and maximum FIR length for the RLS method for example) can be afterward deduced. On the basis of this set of parameters, these more precise methods can then be used to perform a system estimation of quality. As the computation cost associated with ESS-based methods is negligible in comparison with the one of LS-based methods, this automated pre-parametrization step do not add much constraints from this point of view. The only drawback of this approach maybe that in order to ensure an optimal result in terms of quality, the ESS pre-parametrization step must be run using an ESS input and the precise RLS estimation using a WGN input. Thus, two excitations of the system may be needed instead of one. However, one can still use only an ESS input but at the possible detriment of quality.

References

[1] R. Cauduro Dias de Paiva, J. Pakarinen, V. Välimäki. Reduced-complexity modeling of high-order nonlinear audio systems using swept-sine and principal component analysis, in: Audio Engineering Society Conference: 45th International Conference: Applications of Time-Frequency Processing in Audio, 2012.
[2] H.W. Chen, Modeling and identification of parallel nonlinear systems: structural classification and parameter estimation methods, Proc. IEEE 83 (1) (1995) 39–66.
[3] C.M. Cheng, Z.K. Peng, W.M. Zhang, G. Meng, A novel approach for identification of cascade of Hammerstein model, Nonlin. Dyn. 86 (1) (2016) 513–522.
[4] D.G. Ciric, M. Markovic, M. Mijic, D. Sumarac-Pavlovic, On the effects of nonlinearities in room impulse response measurements with exponential sweeps, Appl. Acoust. 74 (3) (2013) 375–382.
[5] F.I. Doyle, R.K. Pearson, B.A. Ogunnaike, Identification and Control using Volterra Models, Springer, 2012.
[6] A. Farina, Simultaneous measurement of impulse response and distortion with a swept-sine technique, in: 108th Convention of the Audio Engineering Society, Paper 5093, February 2000.
[7] A. Farina, A. Bellini, E. Armelloni. Non-linear convolution: a new approach for the auralization of distorting systems, in: 110th Convention of the Audio Engineering Society, Paper 5359, May 2001.
[8] M.E. Gadringer, D. Silveira, G. Magerl, Efficient power amplifier identification using modified parallel cascade Hammerstein models, in: 2007 IEEE Radio and Wireless Symposium, IEEE, 2007, pp. 305–308.
[9] P.G. Gallman, Iterative method for identification of nonlinear-systems using a Uryson model, IEEE Trans. Autom. Contr. 20 (6) (1975) 771–775.
[10] F. Giri, E.-W. Bai, Block-Oriented Nonlinear System Identification, Springer, 2010.
[11] M. Isaksson, D. Wisell, D. Ronnow, A comparative analysis of behavioral models for RF power amplifiers, IEEE Trans. Microw. Theory Tech. 54 (1) (2006) 348–359.
[12] G. Kerschen, K. Worden, A.F. Vakakis, J.C. Golinval, Past, present and future of nonlinear system identification in structural dynamics, Mech. Syst. Sign. Process. 20 (3) (2006) 505–592.
[13] A. Marconato, M. Schoukens, Y. Rolain, J. Schoukens, Study of the effective number of parameters in nonlinear identification benchmarks, in: 52nd IEEE Conference on Decision and Control, Florence, Italy, 2013, pp. 4308–4313.
[14] A. Marconato, M. Schoukens, J. Schoukens, Filter-based regularisation for impulse response modelling, IET Contr. Theory Appl. 11 (2) (2017) 194–204.
[15] J. Noël, G. Kerschen, Nonlinear system identification in structural dynamics: 10 more years of progress, Mech. Syst. Sign. Process. 83 (2017) 2–35.
[16] J.P. Noël, M. Schoukens, Hysteretic benchmark with a dynamic nonlinearity, in: Workshop on Nonlinear System Identification Benchmarks, 2016, pp. 7–14. <http://www.nonlinearbenchmark.org>(checked 23/11/2017).
[17] A. Novak, M. Bentahar, V. Tournat, R.E. Guerjouma, L. Simon, Nonlinear acoustic characterization of micro-damaged materials through higher harmonic resonance analysis, NDT&E Int. 45 (1) (2012) 1–8.
[18] A. Novak, P. Lotton, L. Simon, Synchronized swept-sine: theory, application, and implementation, J. Audio Eng. Soc. 63 (10) (2015) 786–798.
[19] A. Novak, L. Simon, F. Kadlec, P. Lotton, Nonlinear system identification using exponential swept-sine signal, IEEE Trans. Instrum. Meas. 59 (8) (2010) 2220–2229.
[20] R.K. Pearson, Discrete-Time Dynamic Models, Oxford University Press, 1999.
[21] G. Pillonetto, F. Dinuzzo, T. Chen, G.D. Nicolao, L. Ljung, Kernel methods in system identification, machine learning and function estimation: a survey, Automatica 50 (3) (2014) 657–682.
[22] R. Pintelon, J. Schoukens, System Identification: A Frequency Domain Approach, John Wiley & Sons, 2012.
[23] M. Rébillat, R. Hajrya, N. Mechbal, Nonlinear structural damage detection based on cascade of Hammerstein models, Mech. Syst. Sign. Process. 48 (1) (2014) 247–259.
[24] M. Rébillat, R. Hennequin, E. Corteel, B.F. Katz, Identification of cascade of Hammerstein models for the description of nonlinearities in vibrating devices, J. Sound Vib. 330 (2011) 1018–1038.
[25] M. Rébillat, K. Ege, M. Gallo, J. Antoni, Repeated exponential sine sweeps for the autonomous estimation of nonlinearities and bootstrap assessment of uncertainties, J. Mech. Eng. Sci. 230 (6) (2016) :1007–1018.
[26] R. Risuleo, G. Bottegal, H. Hjalmarsson, A new kernel-based approach for overparameterized Hammerstein system identification, in: 54th IEEE Conference on Decision and Control (CDC), 2015.
[27] V. Roggerone, M. Rébillat, E. Corteel, X. Boutillon, Parallel Hammerstein models identification using sine sweeps and the welch method, in: World Congress of the International Federation of Automatic Control (IFAC WC), 2017.
[28] J. Schoukens, R. Pintelon, Y. Rolain, M. Schoukens, K. Tiels, L. Vanbeylen, G. Vandersteen, Structure discrimination in block-oriented models using linear approximations: a theoretic framework, Automatica 53 (2015) 225–234.
[29] M. Schoukens, R. Pintelon, Y. Rolain, Parametric identification of parallel Hammerstein systems, IEEE Trans. Instrum. Meas. 60 (12) (2011) 3931–3938.
[30] M. Schoukens, Y. Rolain, R. Pintelon, J. Schoukens, Semi-parametric identification of parallel Hammerstein systems, in: UKACC International Conference on Control 2010, 2010, pp. 1–6.
[31] A. Torras-Rosell, F. Jacobsen, A new interpretation of distortion artifacts in sweep measurements, J. Audio Eng. Soc. 59 (5) (2011) 283–289.