Hammerstein System Identification using LS-SVM and Steady State Time Response

Ricardo Castro-Garcia, Oscar Mauricio Agudelo, Koen Tiels, Johan A. K. Suykens

***Abstract*— In this paper a new system identification approach for Hammerstein systems is proposed. A straightforward estimation of the nonlinear block through the use of LS-SVM is done by making use of the behavior of Hammerstein systems in steady state. Using the estimated nonlinear block, the intermediate variable is calculated. Using the latter and the known output, the linear block can be estimated. The results indicate that the method can effectively identify Hammerstein systems also in the presence of a considerable amount of noise. The well-known capabilities of LS-SVM for the representation of nonlinear functions play an important role in the generalization capabilities of the method allowing to work with a wide range of model classes. The proposed method's main strength lies precisely in the identification of the nonlinear block of the Hammerstein system. The relevance of these findings resides in the fact that a very good estimation of the inner workings of a Hammerstein system can be achieved.**

I. INTRODUCTION

In the field of system identification, several block structured models have been introduced [1]. Even simple nonlinear models can often provide much better approximations to process dynamics than linear ones. Hammerstein models [2] are nonlinear models often used. They have been employed to model heat exchangers [3], sticky control valves [4], electrical drives [5] and physiological systems [6]. Hammerstein systems consist of a static part $f(\cdot)$ containing the nonlinearity, followed by a linear part $G_0(q)$ containing the dynamics of the process (see Fig. 1).

In this paper the q-notation will be used. The operator q is a time shift operator of the form $q^{-1}x(t) = x(t-1)$.

In the literature, several identification methods for Hammerstein systems have been presented. An overview of previous works can be found in [7]. Different classifications of these methods can be found in [8], [9] and [10].

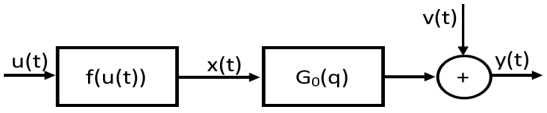

Fig. 1. A Hammerstein system. $G_0(q)$ is a linear dynamical system and $f(u(t))$ is a static nonlinearity. $v(t)$ is the measurement noise.

The idea in this paper is to use Least Squares Support Vector Machines (LS-SVM) [11] while making use of the characteristic behavior of Hammerstein systems under steady state. The resulting methodology turns out to be easily implementable while giving good results. Also, it allows to separate the identification of the linear and nonlinear parts.

Previous works in the system identification literature have used LS-SVM. Some authors have applied them specifically to Hammerstein systems [12], [13] and others have attempted to include information about the structure of the system into the LS-SVM models [14], [15]. However, none of those techniques have attempted a straightforward calculation of the nonlinear block using LS-SVM.

The proposed method is based on applying a multilevel input signal in which the duration of the steps is longer than the settling time of the system. It uses a forward approach as defined in [16] where the nonlinear block is identified first, and the linear block is modeled afterwards. More precisely, the method consists of the following steps:

- The system's settling time is estimated through the application of a step signal.
- A multilevel input signal is created based on the calculated settling time.
- An LS-SVM model is trained using the levels of the multilevel signal as inputs and their corresponding output values in steady state as outputs.
- An additional experiment is carried out in order to identify the linear block. Here the applied input is evaluated using the obtained nonlinearity in order to estimate the intermediate variable. With the intermediate variable and the known output, the linear block is estimated through least squares.

A somewhat similar approach was proposed in [17]. However, there it is assumed that the nonlinearity is a linear combination of known functions and that it is locally invertible. In this work, those assumptions are not necessary. Additionally, in this paper a way for identifying Hammerstein

The research leading to these results has received funding from: European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC AdG A-DATADRIVE-B (290923). This paper reflects only the authors' views, the Union is not liable for any use that may be made of the contained information. Research Council KUL: GOA/10/09 MaNet, CoE PFV/10/002 (OPTEC), BIL12/11T; PhD/Postdoc grants. Flemish Government: FWO: projects: G.0377.12 (Structured systems), G.088114N (Tensor based data similarity); PhD/Postdoc grants. IWT: projects: SBO POM (100031); PhD/Postdoc grants. iMinds Medical Information Technologies SBO 2014. Belgian Federal Science Policy Office: IUAP P7/19 (DYSCO, Dynamical systems, control and optimization, 2012-2017). Fund for Scientic Research (FWO-Vlaanderen), by the Flemish Government (Methusalem), the Belgian Government through the Inter university Poles of Attraction (IAP VII) Program, and by the ERC advanced grant SNLSID, under contract 320378.

Ricardo Castro-Garcia, Oscar Mauricio Agudelo and Johan A.K. Suykens are with KU Leuven, ESAT-STADIUS, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium. (ricardo.castro@esat.kuleuven.be, mauricio.agudelo@esat.kuleuven.be, johan.suykens@esat.kuleuven.be)

Koen Tiels is with the Vrije Universiteit Brussel, Faculty of Engineering, Department of Fundamental Electricity and Instrumentation, Pleinlaan 2 1050 Brussels, Belgium (koen.tiels@vub.ac.be)

systems for which the linear block is a high pass filter is offered, which is not possible in [17].

The proposed method provides an easy way to directly use standard LS-SVM for the identification of Hammerstein systems while bearing in mind the structure of such systems. It allows to estimate the nonlinear block in a straightforward manner independently of the linear block and does not require any particularly complex set of inputs-outputs. This is important as it implies that the method can be applied to a wide set of problems. Also, given the way it works, it can give very good approximations to the intermediate variable (up to a scaling factor) even in the presence of heavy white Gaussian zero mean noise.

In this work scalars are represented in lower case, lower case followed by (t) is used for signals in the time domain, vectors are represented with bold lower case and matrices with bold upper case. E.g. x is a scalar, $x(t)$ is a signal in the time domain, $\boldsymbol{x}$ is a vector and $\boldsymbol{X}$ is a matrix.

The paper is organized as follows: In Section II, the proposed methodology is presented. Section III illustrates the results found when applying the described methodology on two simulation examples. Finally, in Section IV, the conclusions are presented.

II. Proposed Method

In this method, the first step is to construct a data set where the input $u_1(t)$ is a multilevel signal in which each step lasts a constant amount of time T_C defined as:

$$T_C = T_S + \Delta_T, \tag{1}$$

where T_S is the settling time of the system and Δ_T is an arbitrary additional time. This way of constructing $u_1(t)$ guarantees that during each step of the input signal some samples will be taken after the system has reached steady state (i.e. those taken during Δ_T after T_S). The input signal $u_1(t)$ can then be described as:

$$u_1(t) = r_k, \text{ for } kT_C \leq t < (k+1)T_C. \tag{2}$$

For each of the steps $k \in \mathbb{N}$, $u_1(t)$ has a constant value r_k.

The settling time of the system T_S is estimated by applying a step signal to the system and determining the time it takes for the corresponding output to stay within a certain range.

It is assumed that the linear block is stable (i.e. all the poles are inside the unit circle). Also, it is assumed for now that the step response of the system does not tend to zero as time tends to infinity, that is:

$$\lim_{t\to\infty} y(t) \neq 0, \tag{3}$$

for

$$x(t) = \begin{cases} 0, & t < 0 \\ r, & 0 \leq t < \infty. \end{cases} \quad \text{with } r \neq 0 \tag{4}$$

In Section III-E a way for overcoming this limitation is presented.

The samples of the output $y_1(t)$ taken during $kT_C + T_S \leq t < (k+1)T_C$ are averaged for each k in order to minimize the effect of the measurement noise during each step.

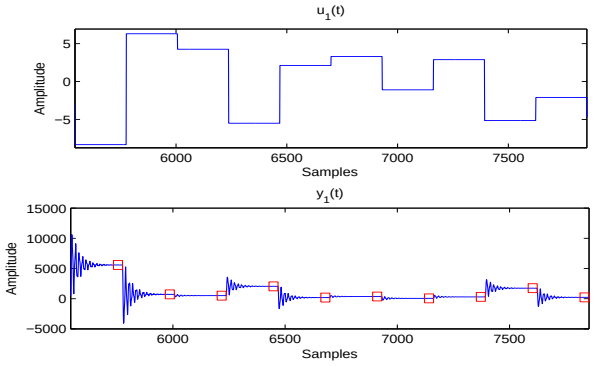

Fig. 2. Example of a training signal. (Top) Input signal $u_1(t)$. (Bottom) Output signal $y_1(t)$.

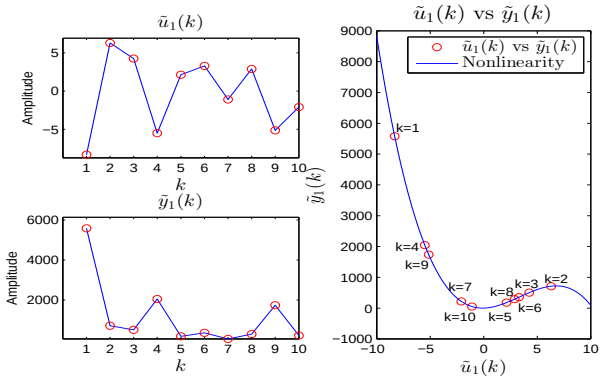

Fig. 3. Corresponding training points for the example of Fig. 2. (Top-Left) r_k values. (Bottom-Left) Averaged $y_1(t)$ values. (Right) $\tilde{u}(k)$ vs $\tilde{y}(k)$ and the rescaled nonlinearity.

In Fig. 2 an excerpt of a training signal is shown to illustrate the samples taken after the settling time at each step of the signal. The red boxes indicate the values of the output signal that are averaged for each step.

Let us define $\tilde{u}_1(k) = r_k$, a signal containing the amplitude level of each step of the input signal. Also, let us define $\tilde{y}(k)$, a signal containing the output averages corresponding to the inputs during $kT_C + T_S \leq t < (k+1)T_C$. Using $\tilde{u}(k)$ as input and $\tilde{y}(k)$ as output, an LS-SVM model can be trained. For the example shown in Fig. 2, the corresponding extracted values $\tilde{u}(k)$ and $\tilde{y}(k)$ are presented in Fig. 3.

In this paper, LS-SVM under a 10-fold cross validation setting is used to obtain the estimation of the nonlinear block. Once this is done, another experiment is carried out, where a new input signal $u_2(t)$ is generated and its corresponding output $y_2(t)$ is obtained. This input signal is then evaluated using the estimated nonlinearity to obtain an approximation to the intermediate variable $x_2(t)$ (i.e. $\hat{x}_2(t)$).

The linear block is a discrete-time rational transfer function of the form

$$\hat{y}(t) = \sum_{j=0}^{m} b_j x(t-j) - \sum_{i=1}^{n} a_i y(t-i), \tag{5}$$

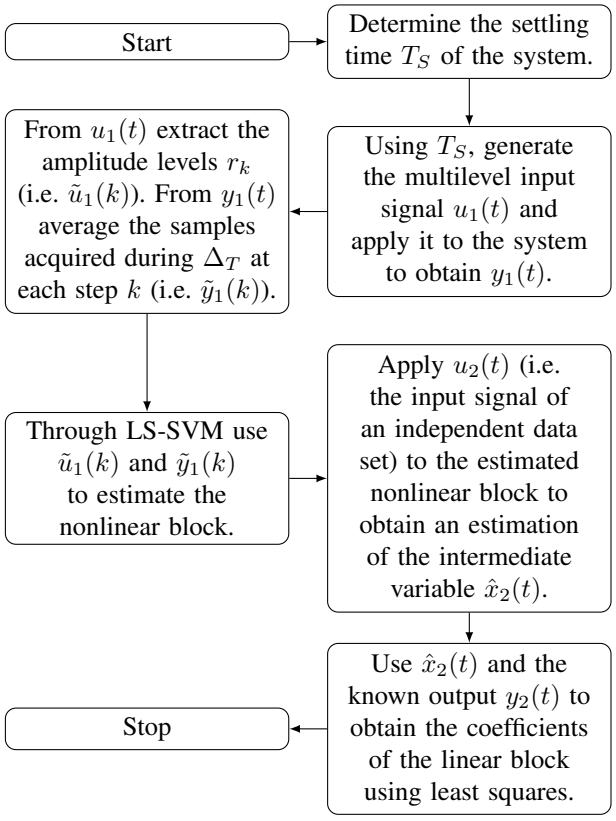

Fig. 4. Summary of the method.

and so, $y_2(t) = \sum_{j=0}^{m} \hat{b}_j \hat{x}_2(t-j) - \sum_{i=1}^{n} \hat{a}_i y_2(t-i)$. The coefficients $\hat{b}_j$ and $\hat{a}_i$ are estimated here using standard least squares to find an approximation of the linear block. This is done using $\hat{x}_2(t)$ and the known output $y_2(t)$. During this step, several orders for the numerator and denominator are tried out.

In Fig. 4, a simplified summary of the method is presented.

III. Results

A. Example

The proposed methodology was applied to a system in the discrete time domain. The system was generated through a nonlinear block:

$$x(t) = \frac{u(t)^2 \sin(\pi u(t))}{(\pi u(t))} \tag{6}$$

and a linear block:

$$y(t) = \frac{B_1(q)}{A_1(q)} x(t) \tag{7}$$

where

$$B_1(q) = 0.1129q^4 - 0.2128q^3 + 0.283q^2 - 0.2128q + 0.1129 \tag{8}$$

$$A_1(q) = q^4 - 2.485q^3 + 2.528q^2 - 1.184q + 0.2245.$$

The system is shown in Fig. 5.

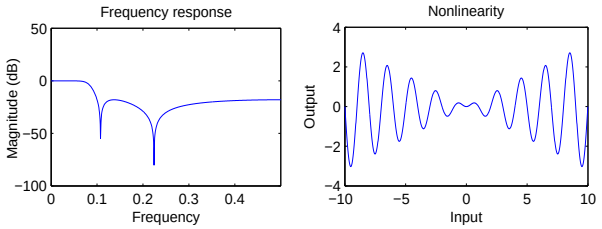

Fig. 5. (Left) Linear block representation in the frequency domain (normalized frequency). (Right) Nonlinear block representation in the time domain.

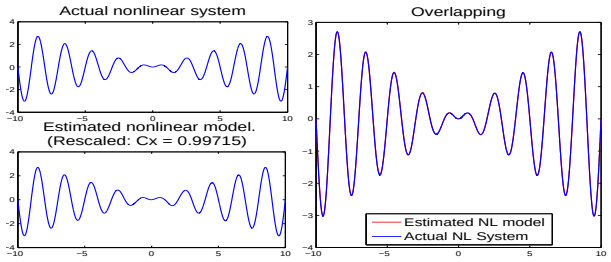

Fig. 6. In all the plots, the vertical axes represent the output value for the corresponding values in the horizontal axes..

B. Signals description

In order to be able to make a comparison between results, let us have the Normalized MAE defined as shown in (9) for a signal with N measurements. Note that the Normalized MAE uses the noise free signal $y_{test}(t)$.

$$\%\text{MAE} = \frac{100}{N} \frac{\sum_{t=1}^{N} |y_{test}(t) - \hat{y}_{test}(t)|}{|\max(y_{test}(t)) - \min(y_{test}(t))|} \tag{9}$$

To construct $u_1(t)$, the settling time T_S was established first by exciting the system with a step of amplitude 10. In this example $T_S = 191$ samples. Afterward, the signal was constructed by adding 40 extra samples at each step (i.e. Δ_T) to those required to achieve steady state. The amplitudes of the steps in this signal (i.e. r_k) were randomly drawn from a uniform distribution ranging between -10 and 10.

From the resulting $y_1(t)$ the values corresponding to the output of the last Δ_T samples at each step were retrieved and averaged (i.e. $\tilde{y}(k)$). In order to estimate the nonlinear block, 500 input-output pairs were used.

In Fig. 6 the resulting nonlinear block of the example is compared with the actual one for a run with a Signal to Noise Ratio (SNR) of 40dB. Note that a rescaling constant is present there. If both, the linear and nonlinear, blocks are considered, there will be a gain factor of the combined blocks. However, this gain could be distributed in any way between the two blocks [18]. The actual difference in scaling has no effect on the input-output behavior of the Hammerstein system (i.e. any pair of $\{f(u(t))/\eta,\ \eta G(q)\}$ with $\eta \neq 0$ would yield identical input and output measurements). Up to this scaling factor, it is clear that the estimated nonlinear block is a good representation of the actual one.

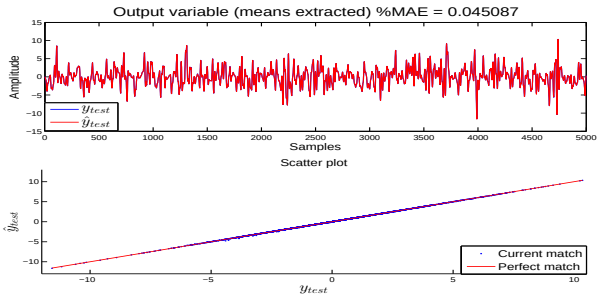

Fig. 7. (Top) Overlapping of the actual and estimated output variables. (Bottom) Scatter plot illustrating the behavior of the overlapped plots.

To estimate the linear block, a new data set of 5000 points was generated. $u_2(t)$, the input to generate this data set, is a multilevel signal where each step has a duration $T_C = 10$ samples. The amplitudes at each level were drawn from a uniform distribution ranging between -10 and 10.

Using $u_2(t)$ and the estimated nonlinear block, an estimation of the intermediate variable $\hat{x}_2(t)$ is calculated. Using $\hat{x}_2(t)$ and the known output $y_2(t)$, the linear block is estimated through least squares. Orders ranging between 1 and 10 were tried out for numerator and denominator. Note that in order to fulfill the made assumptions, given a linear block defined as in (5), only cases where $m \leq n$ can be considered.

Finally, the system was tested in a third data set. The input for generating this set, $u_{test}(t)$, is a multilevel signal where each step has a duration $T_C = 10$ samples. The amplitudes at each level were drawn from a uniform distribution ranging between -10 and 10. This data set consists of 5000 points.

In Fig. 7 the estimated output is compared with the actual one for the same run used in Fig. 6.

Note that white Gaussian noise with zero mean was applied to the output of each data set. In Section III-C the effect of noise in the method is explained.

C. Noise effect analysis

In order to evaluate how the noise affects the performance of the proposed method, 100 Monte Carlo simulations were carried out for each of four different SNRs varying between 10dB and 80dB.

In Fig. 8 the results of the Monte Carlo simulations are presented. As can be seen, the performance of the proposed method dramatically changes as the level of the noise varies.

It is important to highlight that the impact of noise can be further reduced if more points are considered in the data set employed for estimating the nonlinear block. To illustrate this, in Fig 9 it is shown how the performance of the method changes for the example when using a SNR of 10dB.

D. Methods comparison

The proposed method was compared with:

- A NARX LS-SVM [11] with 10 lags of input and 10 lags of output.

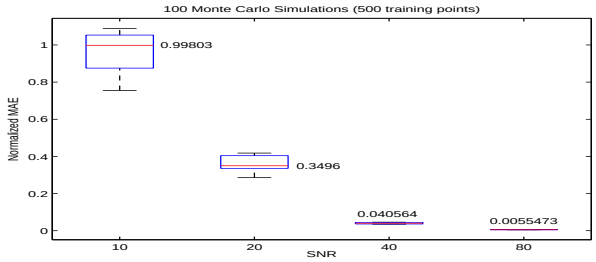

Fig. 8. Evolution of the normalized MAE of the output of the model as the SNR changes. The corresponding median values appear next to each box.

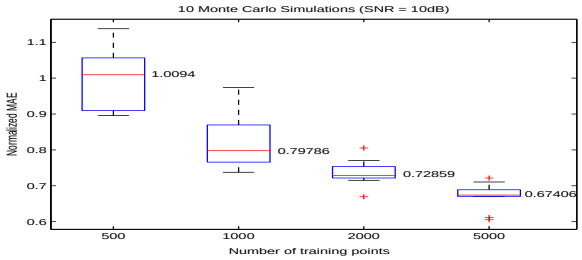

Fig. 9. Evolution of the normalized MAE of the output of the model as the number of training points changes. The corresponding median values appear next to each box.

- The Hammerstein and Wiener Identification procedure (in this paper denoted by WHIP) presented in [19].
- The iterative method (in this paper denoted by IM) presented in [20].

The proposed method was implemented using a RBF kernel for the LS-SVM part. This kernel requires the tuning of a kernel parameter σ and a regularization parameter γ [11].

For the IM method a Gaussian noise input was used. This signal had a standard deviation as large as the standard deviation of the concatenation of the input signals $u_1(t)$ and $u_2(t)$ as described in Section III-B. The models were estimated using 115500 samples, while 5000 samples were used to look for the best model order (i.e. scan over orders 2, 3, 4, 5, and 6). To model the nonlinearity a piecewise linear function with 50 breakpoints was used. It is important to note that the choice for the input signal of the IM method is such that as many samples and as much total energy is used for the identification of the system as for the proposed method.

For the WHIP method a random-phase multisine was employed. Again, this signal had the same standard deviation as the concatenation of signals $u_1(t)$ and $u_2(t)$. Seven phase realizations and 2 periods (plus an additional period to reduce the effect of transients) of the multisine with 5000 samples per period were used to estimate the models. One period (no transient removal) of an additional realization was used to look for the best model order with the same order scanning used for the IM method.

In Table I the results of the comparison in Normalized MAE form are presented. Each of the presented results corresponds to an average over 10 runs.

TABLE I
RESULTS COMPARISON IN NORMALIZED MAE ON TEST DATA.

	SNR (dB)	
	10	**20**
Proposed method	1.5259	0.6925
NARX LS-SVM	9.9266	9.9314
IM	8.8621	9.3660
WHIP	5.6204	8.3097

The results indicate that the proposed method obtains better results as the noise is reduced. For the NARX LS-SVM the results seem to stay almost the same as the noise is increased. For WHIP and IM, the results are better when the noise is increased. This result is explained by the presence of outliers in the results, which indicates that these methods are sensitive to local minima.

The IM and WHIP methods assume that the nonlinearity can be represented in a basis function expansion form with known basis functions. Note that the nonlinearity in (6) is hard to model by a polynomial of reasonable degree and in consequence, piecewise linear basis functions were used. Since a finite number of breakpoints is used, the true nonlinearity is not in the model class. This can be an explanation for the poor results of the last two methods in the example.

As can be seen, the proposed method performs very well in the example. This behavior suggests that it is robust against the amount of noise used.

E. High pass filter case

The proposed methodology gives good results in the established framework. However, as it is presented, the method is unable to deal with situations where the assumption introduced in Eqs. (3) and (4) is violated. A clear illustration of this occurs when the linear block is a high pass filter. In this particular situation:

$$\lim_{t\to\infty} y(t) = 0, \tag{10}$$

for

$$x(t) = \begin{cases} 0, & t < 0 \\ r, & 0 \leq t < \infty. \end{cases} \quad \text{with } r \neq 0. \tag{11}$$

In this case, for training the LS-SVM model, the corresponding output points would always be zero or very close to zero:

$$\tilde{y}(k) = 0 \ \forall k. \tag{12}$$

In order for the method to be able to work in these situations, the addition of one or several integrators to the output signal is proposed, this is represented in Fig. 10. Note that this has to be done only in the first stage of the method, that is, for the estimation of the nonlinear block. The number of integrators required depends directly on the linear block. However, it can be easily established through direct observation. If more integrators than needed are added, the system will become unstable.

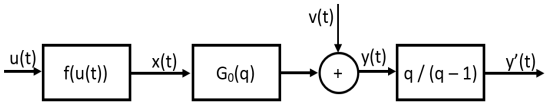

Fig. 10. Hammerstein system with an added integrator at the output for estimation of the nonlinear block.

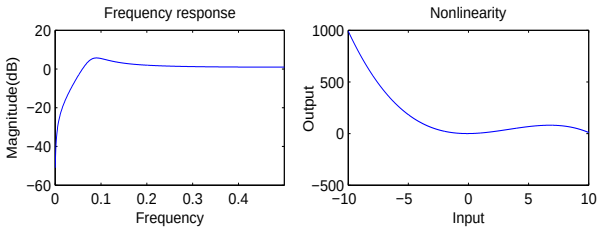

Fig. 11. High pass filter example: (Left) Linear block representation in the frequency domain (normalized frequency). (Right) Nonlinear block representation in the time domain.

To illustrate the high pass filter case, an example is presented where the nonlinear block has the form

$$x(t) = u(t) + 5u(t)^2 - \frac{u(t)^3}{2} \tag{13}$$

and the linear block is given by:

$$y(t) = \frac{B_2(q)}{A_2(q)} x(t) \tag{14}$$

where

$$\begin{aligned} B_2(q) &= q^2 - 1.8q + 0.8 \\ A_2(q) &= q^2 - 1.5q + 0.7225. \end{aligned} \tag{15}$$

This system is illustrated in Fig. 11. In this example, the signals used are very similar to those described in Section III-B, however, 100 $\{\tilde{u}(k), \tilde{y}(k)\}$ pairs were used instead of 500. Also, the second data set (i.e. $\{u_2(t), y_2(t)\}$) and the test set consisted of 1000 samples.

Once the linear block is estimated as explained in Section II, the model of the system is tested with an independent data set. The resulting output variable behavior is presented in Fig. 12.

In Fig. 13 the results of a Monte Carlo simulation of 10 runs for different levels of noise is shown. It shows how the

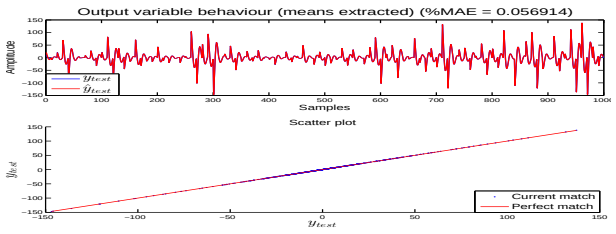

Fig. 12. High pass filter example: (Top) Overlapping of the actual and estimated output variables. (Bottom) Scatter plot illustrating the behavior of the overlapped plots.

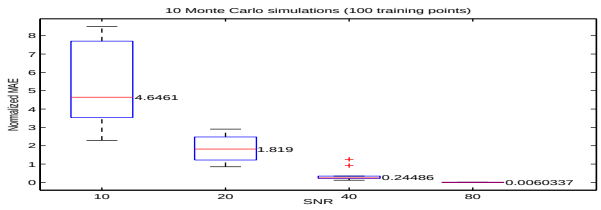

Fig. 13. High pass filter example: Normalized MAE for different levels of noise. The corresponding median values appear next to each box.

normalized MAE evolves as the level of noise changes in the example represented by Eqs. (13) and (14).

Note that this approach can be sensitive to cases with zeros very close to 1 but not exactly at 1 in the unit circle. In those cases, using the proposed method with both the non-integrated or the integrated output might yield unsatisfactory results.

IV. CONCLUSIONS AND FUTURE WORKS

A. Conclusions

The presented method offers a simple way for accurate Hammerstein system identification. This is done mainly by making use of the behavior of the system in steady state. In this work, this was done through LS-SVM which allows a good generalization capability when using different model classes.

The main strength of the proposed method lies in the identification of the nonlinear block of Hammerstein systems. The presented results indicate that the method is very effective in the presence of zero mean, white Gaussian noise.

Once the nonlinear model is learned, it can be easily applied. It is shown that even with a small amount of training points, the results are already quite accurate. In practice, this means that the calculation of the model can also be done very quickly. It is also possible to improve the performance of the method by using more training points for modeling the nonlinearity.

The estimated nonlinear model is very close to the original one (up to a scaling factor). This allows insight into the behavior of the studied system as it is possible to visualize the way the nonlinear block will respond to the inputs. Naturally, this allows as well an estimation of the intermediate variable.

The way $u_1(t)$ is constructed is quite simple and given its shape, it allows the application of the method in many fields.

A possible drawback of the methodology lies in the fact that depending on the evaluated system, constructing the initial input signal $u_1(t)$ could require a long time.

B. Future Works

The central idea of this work can be used in the identification of Wiener and Wiener-Hammerstein systems as the working concepts would be basically the same. Though not as straightforward as in the Hammerstein case, full models of these structures could be estimated after the nonlinearity is modeled.

More complex cases like MIMO Hammerstein, Wiener and Wiener-Hammerstein can also be considered though they will not be as easily adapted.

REFERENCES

[1] S. Billings and S. Fakhouri, "Identification of systems containing linear dynamic and static nonlinear elements," *Automatica*, vol. 18, no. 1, pp. 15–26, 1982.

[2] K. Narendra and P. Gallman, "An iterative method for the identification of nonlinear systems using a Hammerstein model," *IEEE Transactions on Automatic Control*, vol. 11, no. 3, pp. 546–550, 1966.

[3] E. Eskinat, S. H. Johnson, and W. L. Luyben, "Use of Hammerstein models in identification of nonlinear systems," *AIChE Journal*, vol. 37, no. 2, pp. 255–268, 1991.

[4] R. Srinivasan, R. Rengaswamy, S. Narasimhan, and R. Miller, "Control loop performance assessment. 2. Hammerstein model approach for stiction diagnosis," *Industrial & engineering chemistry research*, vol. 44, no. 17, pp. 6719–6728, 2005.

[5] A. Balestrino, A. Landi, M. Ould-Zmirli, and L. Sani, "Automatic nonlinear auto-tuning method for Hammerstein modeling of electrical drives," *IEEE Transactions on Industrial Electronics*, vol. 48, no. 3, pp. 645–655, 2001.

[6] E. J. Dempsey and D. T. Westwick, "Identification of Hammerstein models with cubic spline nonlinearities," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 2, pp. 237–245, 2004.

[7] F. Giri and E.-W. Bai, Eds., *Block-oriented nonlinear system identification*. Springer, 2010, vol. 1.

[8] E. W. Bai, "Frequency domain identification of Hammerstein models," *IEEE Transactions on Automatic Control*, vol. 48, no. 4, pp. 530–542, 2003.

[9] R. Haber and L. Keviczky, *Nonlinear System Identification Input-Output Modeling Approach*. Springer The Netherlands, 1999, vol. 1.

[10] A. Janczak, *Identification of Nonlinear Systems Using Neural Networks and Polynomial Models: A Block-Oriented Approach*. Springer-Verlag Berlin Heidelberg, 2005, vol. 310.

[11] J. A. K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, and J. Vandewalle, *Least Squares Support Vector Machines*. World Scientific, 2002.

[12] I. Goethals, K. Pelckmans, J. A. K. Suykens, and B. De Moor, "Identification of MIMO Hammerstein models using Least-Squares Support Vector Machines," *Automatica*, vol. 41, no. 7, pp. 1263–1272, 2005.

[13] R. Castro-Garcia, K. Tiels, J. Schoukens, and J. A. K. Suykens, "Incorporating Best Linear Approximation within LS-SVM-Based Hammerstein System Identification," in *Accepted for presentation at the 54th IEEE Conference on Decision and Control (CDC 2015)*. IEEE, 2015.

[14] T. Falck, K. Pelckmans, J. A. K. Suykens, and B. De Moor, "Identification of Wiener-Hammerstein systems using LS-SVMs," in *Proceedings of the 15th IFAC symposium on system identification (SYSID 2009)*, 2009, pp. 820–825.

[15] T. Falck, P. Dreesen, K. De Brabanter, K. Pelckmans, B. De Moor, and J. A. K. Suykens, "Least-Squares Support Vector Machines for the identification of Wiener-Hammerstein systems," *Control Engineering Practice*, vol. 20, no. 11, pp. 1165–1174, 2012.

[16] L. Sun, W. Liu, and A. Sano, "Identification of a dynamical system with input nonlinearity," in *Control Theory and Applications, IEE Proceedings-*, vol. 146, no. 1. IET, 1999, pp. 41–51.

[17] F. Ikhouane and F. Giri, "A unified approach for the parametric identification of SISO/MIMO Wiener and Hammerstein systems," *Journal of the Franklin Institute*, vol. 351, no. 3, pp. 1717–1727, 2014.

[18] S. Boyd and L. O. Chua, "Uniqueness of a basic nonlinear structure," *IEEE Transactions on Circuits and Systems*, vol. 30, no. 9, pp. 648–651, 1983.

[19] M. Schoukens, "Identification of Parallel Block-Oriented Models starting from the Best Linear Approximation," Ph.D. dissertation, Vrije Universiteit Brussel (VUB), 2015.

[20] E. W. Bai and K. Li, "Convergence of the iterative algorithm for a general Hammerstein system identification," *Automatica*, vol. 46, no. 11, pp. 1891–1896, 2010.