

Article

Kernel Identification of Non-Linear Systems with General Structure

Grzegorz Mzyk ^{*}, Zygmunt Hasiewicz  and Paweł Mielcarek 

Faculty of Electronics, Wrocław University of Science and Technology, 50-372 Wrocław, Poland;

zygmunt.hasiewicz@pwr.edu.pl (Z.H.); pawel.mielcarek@pwr.edu.pl (P.M.)

* Correspondence: grzegorz.mzyk@pwr.edu.pl

Received: 27 October 2020; Accepted: 4 December 2020; Published: 6 December 2020



Abstract: In the paper we deal with the problem of non-linear dynamic system identification in the presence of random noise. The class of considered systems is relatively general, in the sense that it is not limited to block-oriented structures such as Hammerstein or Wiener models. It is shown that the proposed algorithm can be generalized for two-stage strategy. In step 1 (non-parametric) the system is approximated by multi-dimensional regression functions for a given set of excitations, treated as representative set of points in multi-dimensional space. ‘Curse of dimensionality problem’ is solved by using specific (quantized or periodic) input sequences. Next, in step 2, non-parametric estimates can be plugged into least squares criterion and support model selection and estimation of system parameters. The proposed strategy allows decomposition of the identification problem, which can be of crucial meaning from the numerical point of view. The “estimation points” in step 1 are selected to ensure good task conditioning in step 2. Moreover, non-parametric procedure plays the role of data compression. We discuss the problem of selection of the scale of non-parametric model, and analyze asymptotic properties of the method. Also, the results of simple simulation are presented, to illustrate functioning of the method. Finally, the proposed method is successfully applied in Differential Scanning Calorimeter (DSC) to analyze aging processes in chalcogenide glasses.

Keywords: system identification; hammerstein system; wiener system; non-parametric methods; kernel regression

1. Introduction

The problem of non-linear system modeling has been intensively examined over the past four decades. Owing to many potential applications (see [1]) and interdisciplinary scope of the topic, both scientists and engineers look for more precise and numerically efficient identification algorithms. First attempts at generalization of linear system identification theory for non-linear models were based on Volterra series representation ([2]). Traditional Volterra series-based approach leads to relatively high numerical complexity, which is often not acceptable from practical point of view. To cope with this problem regularization or tensor network techniques have been proposed recently ([3,4]). However, strong restrictions are imposed on the system characteristics (e.g., smoothness of nonlinearity, and short memory of the dynamics). Alternatively, the concept of block-oriented models was introduced ([5]). It was assumed that the system can be represented, or approximated with satisfactory accuracy, with the use of structural models including interconnected elementary blocks of two types—linear dynamics and static nonlinearities ([6]). The most popular structures in this class are Hammerstein system (see e.g., [7–12]) with static nonlinearity followed by linear dynamics, and the Wiener system [13–18] including the same blocks connected reversely.

Traditionally, identification method needs two kinds of knowledge—set of input–output measurements and a priori parametric formula describing system characteristics and including finite

number of unknown parameters ([6,19]). Usually, the polynomial model of static characteristics and the difference equation with known orders are assumed. This convention leads to relatively fast convergence of parameter estimates, but the risk of systematic approximation error appears, if the assumed model is not correct. As an alternative, the theory of non-parametric system identification ([20–23]) was proposed to solve this problem. The algorithms work under mild prior restrictions, such as stability of linear dynamic block and local continuity of static non-linear characteristics. Although the estimates converge to the true characteristics, the rate of convergence is relatively slow in practice, as a consequence of assumptions relaxation.

This paper represents the idea of combined, i.e., parametric-non-parametric, approach (see e.g., [24–30]), in which both parametric and non-parametric algorithms support each other to achieve the best possible results of identification for moderate number of measurements and to guarantee the asymptotic consistency (i.e., convergence to the true system characteristics), when the number of observations tends to infinity. Since the preliminary step of structure selection is treated rather cursorily in the literature, generalizations of the approach towards wider classes of systems seem to be of high importance from practical point of view. The paper was also motivated by the project of Differential Scanning Calorimeter ([31]) developed in the team, and particularly, modeling of heating process for examination the properties of chalcogenide glasses.

Main contribution of the paper lays in the following aspects:

- proposed identification algorithm is run without any prior knowledge about the system structure and parametric representation of nonlinearity,
- non-parametric multi-dimensional kernel regression estimate was generalized for modeling of non-linear dynamic systems, and the dimensionality problem was solved by using special input sequences,
- the scheme elaborated in the paper was successfully applied in Differential Scanning Calorimeter for testing parameters of chalcogenide glasses.

The paper is organized as follows. In Section 2 the general class of considered systems and the identification problem is formulated in detail. Next, in Section 3.1, purely non-parametric, regression-based approach is presented, and its disadvantages are discussed. Then, to cope with dimensionality problem, the idea of some specific input sequences is presented in Section 3.2. Owing to that, the system characteristics are identified only for some selected points, but the convergence is much faster. The idea of combined, two-stage strategy is introduced in Section 4. It allows use of prior knowledge to expand the model on the whole input space. Also, the results of simple simulation are included in Section 5 to illustrate and discuss some practical aspects of the approach. Finally, in Section 6, the algorithm is successfully applied in Differential Scanning Calorimeter to model aging properties of modern materials (chalcogenide glasses).

2. Problem Statement

2.1. Class of Systems

We consider discrete-time non-linear dynamic system with general representation

$$y_k = \mathcal{F}(\{u_{k-i}\}_{i=0}^{\infty}) + z_k, \quad (1)$$

where $\{u_{k-i}\}_{i=0}^{\infty}$ is bounded random input sequence ($|u_k| < u_{\max}$), z_k is zero mean random disturbance. The transformation $\mathcal{F}(\cdot)$ is Lipschitz with respect to all arguments, and has the property of exponential forgetting (fading memory), i.e., if we put $u_{k-i} = 0$ for $i \geq s$ and define the cut-off sequence as

$$\bar{u}_{k-i} \triangleq \begin{cases} u_{k-i}, & \text{as } i = 0, 1, \dots, s-1 \\ 0, & \text{as } i \geq s \end{cases}, \quad (2)$$

We assume that it holds that

$$\Delta(s) \triangleq |\mathcal{F}(\{u_{k-i}\}_{i=0}^{\infty}) - \mathcal{F}(\{\bar{u}_{k-i}\}_{i=0}^{\infty})| \leq c\lambda^s, \tag{3}$$

with some unknown $c = \text{const}$, and $0 < \lambda < 1$. Similar class of fading memory systems, in which the output depends less and less on historical inputs, was considered in [32]. The goal is to identify the system (build the model $\hat{\mathcal{F}}(\cdot)$) using the sequence of N input–output measurements $\{(u_k, y_k)\}_{k=1}^N$.

In considered system class, hysteresis is not admitted.

2.2. Examples

In this section, we show that some systems (popular in applications) fall into above description as special cases.

2.2.1. Hammerstein System

For Hammerstein system (see Figure 1), described by the equation

$$y_k = \sum_{i=0}^{\infty} \gamma_i \mu(u_{k-i}) + z_k \tag{4}$$

with the Lipschitz non-linear characteristic, i.e., such that

$$|\mu(u_a) - \mu(u_b)| \leq l|u_a - u_b|, \tag{5}$$

and asymptotically stable dynamics, i.e., $|\gamma_i| \leq c_H \lambda^i$, we get

$$\Delta(s) = \left| \sum_{i=s}^{\infty} \gamma_i (\mu(u_{k-i}) - \mu(0)) \right| \leq l u_{\max} \left| \sum_{i=s}^{\infty} \gamma_i \right| \leq c\lambda^s. \tag{6}$$

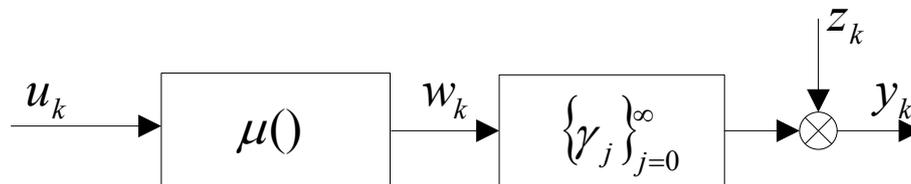


Figure 1. Hammerstein system.

2.2.2. Wiener System

Analogously, for Wiener system (Figure 2), where the stable linear dynamics is followed by the Lipschitz static non-linear block

$$y_k = \mu \left(\sum_{i=0}^{\infty} \lambda_i u_{k-i} \right) + z_k, \tag{7}$$

We get

$$\Delta(s) = \left| \mu \left(\sum_{i=0}^{\infty} \gamma_i u_{k-i} \right) - \mu \left(\sum_{i=0}^{s-1} \gamma_i u_{k-i} \right) \right| \tag{8}$$

$$\leq l \left| \sum_{i=s}^{\infty} \gamma_i u_{k-i} \right| < c\lambda^s. \tag{9}$$

Remark 1. Analogously, it can be simply shown that also Wiener–Hammerstein (L–N–L) and Hammerstein–Wiener (N–L–N) sandwich systems belong to the assumed class.

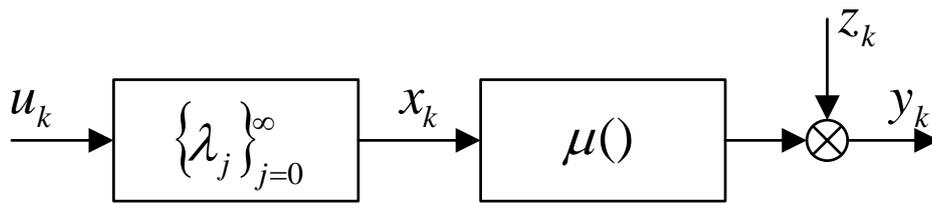


Figure 2. Wiener System.

2.2.3. Finite Memory Bilinear System

Another important and often met in application special case is the bilinear system with finite order m . It is described by the formula

$$y_k = \sum_{l_0, l_1, \dots, l_{m-1} : l_i \geq 0 \cap \sum_i l_i \leq m} c_{l_0, l_1, \dots, l_{m-1}} \prod_{i=0}^{m-1} u_{k-i}^{l_i} + z_k \tag{10}$$

i.e., for $m = 1$ we get

$$y_k = c_{0,0} + c_{1,0}u_k + c_{0,1}u_{k-1}, \tag{11}$$

for $m = 2$ we have that

$$y_k = c_{0,0} + c_{1,0}u_k + c_{0,1}u_{k-1} + c_{1,1}u_k u_{k-1} + c_{2,0}u_k^2 + c_{0,2}u_{k-1}^2. \tag{12}$$

Since y_k does not depend on $u_{k-m}, u_{k-m-1}, \dots$, for $s \geq m$ we get $\mathcal{F}(\{u_{k-i}\}_{i=0}^\infty) = \mathcal{F}(\{\bar{u}_{k-i}\}_{i=0}^\infty)$, and obviously $\Delta(s) = 0$.

Considered example falls into the more general class of Volterra representation. Presented approach works without the knowledge of parametric representation. As regards applicability to the class (1), for $m < \infty$ we have finite memory system, i.e., $\Delta(s) = 0$, as $s > m$. Moreover, since the input is assumed to be bounded ($|u_k| < u_{\max}$), resulting mapping $\mathcal{F}()$ fulfills Lipschitz condition (as ordinary polynomial on compact support).

3. Non-Parametric Regression

3.1. General Overview

Let us introduce the s -dimensional input regressor

$$\mathbf{u}_k^{(s)} = (u_k, u_{k-1}, \dots, u_{k-(s-1)})^T, \tag{13}$$

and the regression function

$$R_s(\mathbf{x}^{(s)}) = E(y_k | \mathbf{u}_k^{(s)} = \mathbf{x}^{(s)}), \tag{14}$$

with the argument vector

$$\mathbf{x}^{(s)} = (x_0, x_1, \dots, x_{s-1})^T. \tag{15}$$

In particular, for $s = 1$ we get

$$R_1(x_0) = E(y_k | u_k = x_0), \tag{16}$$

and for $s = 2$

$$R_2(x_0, x_1) = E(y_k | u_k = x_0 \cap u_{k-1} = x_1). \tag{17}$$

The non-parametric kernel estimate ([20–23,33]) of $R_s(\mathbf{x}^{(s)})$ has the following form

$$\widehat{R}_s(\mathbf{x}^{(s)}) = \frac{\sum_{k=1}^N y_k K\left(\frac{\|\mathbf{u}_k^{(s)} - \mathbf{x}^{(s)}\|}{h}\right)}{\sum_{k=1}^N K\left(\frac{\|\mathbf{u}_k^{(s)} - \mathbf{x}^{(s)}\|}{h}\right)}, \tag{18}$$

where $\|\cdot\|$ denotes Euclidean norm, $K(\cdot)$ plays the role of kernel function, e.g.,

$$K(v) = \begin{cases} 1, & \text{as } |v| \leq 1 \\ 0, & \text{as } |v| > 1 \end{cases}, \tag{19}$$

and h is a bandwidth parameter, responsible for the balance between bias and variance of the estimate. The class of possible kernels can obviously be generalized. Nevertheless, previous experiences shows that the kind of kernel function used for estimation is of secondary importance, whereas behavior of $h(N)$ with respect to N is fundamental. We limited the presentation to Parzen (window) kernel for clarity of exposition. It fulfills all general assumptions made for kernels, i.e., it is even, non-negative and square integrable. The system (1) is thus approximated by the model

$$\widehat{R}_s(u_k, u_{k-1}, \dots, u_{k-(s-1)}), \tag{20}$$

and s can be interpreted as its complexity. Obviously, both $h = h(N)$ and $s = s(N)$ need to be set depending on the number of measurements N . Observing that

$$\mathcal{F}(\{u_{k-i}\}_{i=0}^\infty) = R_\infty(\mathbf{u}_k^{(\infty)}), \tag{21}$$

the mean squared error of the model $\widehat{R}_s(\mathbf{u}_k^{(s)})$ can be expressed as follows

$$MSE(\widehat{R}_s(\mathbf{u}_k^{(s)})) = E(\widehat{R}_s(\mathbf{u}_k^{(s)}) - \mathcal{F}(\{u_{k-i}\}_{i=0}^\infty))^2 = E(\widehat{R}_s(\mathbf{u}_k^{(s)}) - R_\infty(\mathbf{u}_k^{(\infty)}))^2,$$

and introducing the true finite-dimensional regression function $R_s(\mathbf{u}_k^{(s)})$ we get

$$\begin{aligned} MSE(\widehat{R}_s(\mathbf{u}_k^{(s)})) &= E\left\{\left(\widehat{R}_s(\mathbf{u}_k^{(s)}) - R_s(\mathbf{u}_k^{(s)})\right) + \left(R_s(\mathbf{u}_k^{(s)}) - R_\infty(\mathbf{u}_k^{(\infty)})\right)\right\}^2 = \\ &= E\left\{\widehat{R}_s(\mathbf{u}_k^{(s)}) - R_s(\mathbf{u}_k^{(s)})\right\}^2 + E\left\{R_s(\mathbf{u}_k^{(s)}) - R_\infty(\mathbf{u}_k^{(\infty)})\right\}^2 + \\ &+ 2E\left\{\left(\widehat{R}_s(\mathbf{u}_k^{(s)}) - R_s(\mathbf{u}_k^{(s)})\right)\left(R_s(\mathbf{u}_k^{(s)}) - R_\infty(\mathbf{u}_k^{(\infty)})\right)\right\}. \end{aligned}$$

Since both $E\left\{R_s(\mathbf{u}_k^{(s)}) - R_\infty(\mathbf{u}_k^{(\infty)})\right\}^2 \rightarrow 0$ as $N \rightarrow \infty$ and $\left(R_s(\mathbf{u}_k^{(s)}) - R_\infty(\mathbf{u}_k^{(\infty)})\right) \rightarrow 0$ as $N \rightarrow \infty$, these components can be set arbitrarily small by using appropriate scale s . Owing to above, for fixed s we focus on the first component of the MSE error of the form

$$ERR = E\left\{\widehat{R}_s(\mathbf{u}_k^{(s)}) - R_s(\mathbf{u}_k^{(s)})\right\}^2 = \text{bias}^2 \widehat{R}_s(\mathbf{u}_k^{(s)}) + \text{var} \widehat{R}_s(\mathbf{u}_k^{(s)}), \tag{22}$$

where

$$\text{bias} \widehat{R}_s(\mathbf{u}_k^{(s)}) \triangleq E \widehat{R}_s(\mathbf{u}_k^{(s)}) - R_s(\mathbf{u}_k^{(s)}), \tag{23}$$

$$\text{var} \widehat{R}_s(\mathbf{u}_k^{(s)}) \triangleq E\left\{\widehat{R}_s(\mathbf{u}_k^{(s)}) - E \widehat{R}_s(\mathbf{u}_k^{(s)})\right\}^2. \tag{24}$$

It can simply be shown that

$$\text{bias}\widehat{R}_s(\mathbf{x}^{(s)}) = o(h(N)), \tag{25}$$

$$\text{bias}^2\widehat{R}_s(\mathbf{x}^{(s)}) = o(h^2(N)), \tag{26}$$

The bias order follows directly from Lipschitz condition, and the fact that $\|\mathbf{u}_k^{(s)} - \mathbf{x}_{[p]}^{(s)}\| \leq h$ for all k 's selected by kernel. Moreover,

$$\text{var}\widehat{R}_s(\mathbf{x}^{(s)}) = o\left(\frac{1}{Nh^s(N)}\right). \tag{27}$$

For window kernel, Lipschitz function \mathcal{F} , and strictly positive input probability density function around the estimation point, probability of selection in s -dimensional space can be obviously evaluated from below by ch^s , where c is some constant. Hence, expected number of successes is proportional to Nch^s (not less than). The variance order is thus a simple consequence of Wald's identity. Hence, to assure the convergence $\widehat{R}_s(\mathbf{x}^{(s)}) \rightarrow R_s(\mathbf{x}^{(s)})$, as $N \rightarrow \infty$ in the mean square sense, the following conditions must be fulfilled

$$h \rightarrow 0 \text{ and } Nh^s \rightarrow \infty, \text{ as } N \rightarrow \infty, \tag{28}$$

which leads to typical setting

$$h(N) \sim N^{-\alpha}, \text{ where } \alpha \in \left(0, \frac{1}{s}\right). \tag{29}$$

To obtain the best asymptotic trade-off between squared bias and variance and comparing its orders we get

$$h^2(N) = \frac{1}{Nh^s(N)}, \tag{30}$$

$$h^{s+2}(N) = \frac{1}{N}, \tag{31}$$

$$h_{opt}(N) \sim N^{-\frac{1}{s+2}}, \tag{32}$$

$$ERR \sim N^{-\frac{2}{s+2}}. \tag{33}$$

Moreover, to assure the balance between the estimation error and approximation error of order $o(\lambda^s)$, connected with neglecting the tail $\{u_{k-i}\}_{i=s}^\infty$ we get

$$N^{-\frac{2}{s(N)+2}} = \lambda^{s(N)}, \tag{34}$$

$$-\frac{2}{s(N)+2} \log N = s(N) \log \lambda, \tag{35}$$

$$\frac{-2}{s(N)(s(N)+2)} = \frac{\log \lambda}{\log N}, \tag{36}$$

$$s^2(N) + 2s(N) = \frac{2}{\log \frac{1}{\lambda}} \log N, \tag{37}$$

where $\frac{2}{\log \frac{1}{\lambda}} = \text{const}$. Owing to above, the scale $s(N)$ must not be faster than $\sqrt{\log N}$, i.e.,

$$s_{opt}(N) = \mathcal{O}\left(\sqrt{\log N}\right). \tag{38}$$

which illustrates slowness of admissible model complexity increasing in general case. The property (38) commonly known as "curse of dimensionality problem" illustrates the main drawback of multi-dimensional

non-parametric regression approach to system modeling in traditional form. The reason is that probability of kernel selection

$$P \left\{ K \left(\frac{\|\mathbf{u}_k^{(s)} - \mathbf{x}^{(s)}\|}{h} \right) = 1 \right\} = P \left\{ \|\mathbf{u}_k^{(s)} - \mathbf{x}^{(s)}\| \leq h \right\} \sim h^s \tag{39}$$

decreases rapidly when s grows large. We also refer the reader to the proof of Theorem 3 in [26], where a detailed discussion concerning an analogous problem can be found.

3.2. Dimension Reduction

To cope with the problem shown in (39) we consider two cases of some specific input excitation processes to speed up the rate of convergence.

3.2.1. Discrete Input

In case 1 we assume that in each s -element input sub-sequence $u_k, u_{k+1}, \dots, u_{k+s-1}$, there exists d inputs with discrete distribution on finite set of possible realizations. Consequently, all the points $\mathbf{u}_k^{(s)} \in R^s$ lay on a finite number of separable and compact subspaces with the internal dimension

$$s^* = s - d, \tag{40}$$

and for $\mathbf{x}^{(s)} = \mathbf{u}_k^{(s)}$ we have

$$ERR \sim N^{-\frac{2}{s^*+2}}. \tag{41}$$

For each measurement point probability of kernel selection behaves like ch^{s^*} , where s^* , ($s^* < s$), is internal dimension of this subspace. In particular, for $d = s$ (all input variables quantized) we get $ERR \sim N^{-1}$. The sets of possible input sequences for $s = 2$ are illustrated in Figure 3 and 4.

3.2.2. Periodic Input

In case 2 we assume the input is periodic with the period N_0 , i.e., $u_k = u_{k+N_0}$ for each $k = 1, 2, \dots, N - N_0$. Then, the value of the regressor $\mathbf{u}_k^{(s)}$ (see (13)) evaluates to one of N_0 distinct points in R^s , $n \in Z$,

$$\mathbf{x}_{[1]}^{(s)} = \begin{bmatrix} u_s \\ u_{s-1} \\ \vdots \\ u_1 \end{bmatrix}, \dots, \mathbf{x}_{[N_0]}^{(s)} = \begin{bmatrix} u_{N_0+s-1} \\ u_{N_0+s-2} \\ \vdots \\ u_{N_0} \end{bmatrix} \tag{42}$$

with probabilities

$$P \left\{ \mathbf{u}_k^{(s)} = \mathbf{x}_{[p]}^{(s)} \right\} = \frac{1}{N_0}, p = 1, 2, \dots, N_0 \tag{43}$$

Measurements are uniformly distributed on the finite set of N_0 distinct points

$$\mathbf{x}_{[1]}^{(s)}, \dots, \mathbf{x}_{[p]}^{(s)}, \dots, \mathbf{x}_{[N_0]}^{(s)} \tag{44}$$

Narrowing of $h(N)$ does not affect the kernel estimator asymptotically (i.e., $s^* = 0$). Consequently, we get the best possible convergence rate $ERR \sim N^{-1}$. However, it must be admitted that estimators are calculated only for finite number of points, and, increasing N_0 causes increase of variance of the regression estimator for particular points $\mathbf{x}_{[p]}^{(s)}$ (as the number of selected data is of order N/N_0).

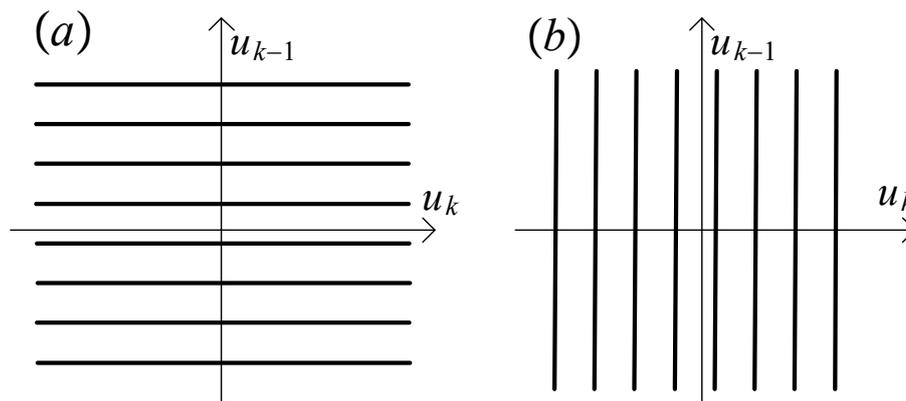


Figure 3. Input space for $s = 2$ and $d = 1$, (a) k odd, (b) k even.

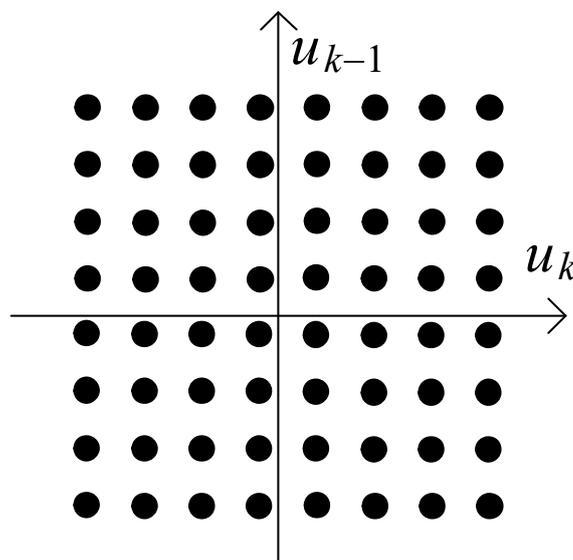


Figure 4. Input space for $s = 2$ and $d = 2$.

4. Hybrid/Combined Parametric-Non-Parametric Approach

Since the special input excitations allows for fast recovering the system characteristics only in some points, additional prior knowledge is needed to extend the model for arbitrary process $\{u_k\}$. In this section we assume that the transformation $\mathcal{F}(\{u_{k-i}\}_{i=0}^{\infty})$ belongs to the given (a priori known), finite dimensional class of systems $F(\mathbf{u}_k^{(s)}, \theta)$

$$\mathcal{F}(\{u_{k-i}\}_{i=0}^{\infty}) \subset F(\mathbf{u}_k^{(s)}, \theta) \tag{45}$$

with unknown parameter vector θ . In the proposed methodology, one of the input excitations described in Section 4 is applied. The system is identified on the finite set of N_0 representative points $\mathbf{x}_{[1]}^{(s)}, \mathbf{x}_{[2]}^{(s)}, \dots, \mathbf{x}_{[p]}^{(s)}, \dots, \mathbf{x}_{[N_0]}^{(s)}$, where $\mathbf{x}_{[p]}^{(s)} \in R^s$, and $p = 1, 2, \dots, N_0$. Let us denote by θ^* —the true and unknown vector of system parameters. We assume that θ^* is identifiable, i.e., the following property holds

$$\mathcal{F}(\{u_{k-i}\}_{i=0}^{\infty}) = F(\mathbf{u}_k^{(s)}, \theta) \iff \theta = \theta^*. \tag{46}$$

Moreover, let the quality index

$$Q(\theta) = E [y_k - F(\mathbf{u}_k^{(s)}, \theta)]^2 \tag{47}$$

be convex for $\theta \in \Xi$, where Ξ is some neighborhood of the true θ^*

$$\theta^* = \arg \min_{\theta} Q(\theta). \tag{48}$$

The following two-step algorithm is proposed.

Step 1. (non-parametric) Using the input–output observations $\{(u_k, y_k)\}_{k=1}^N$ for $p = 1, 2, \dots, N_0$ compute the estimates

$$\widehat{R}_s(\mathbf{x}_{[1]}^{(s)}), \dots, \widehat{R}_s(\mathbf{x}_{[p]}^{(s)}), \dots, \widehat{R}_s(\mathbf{x}_{[N_0]}^{(s)}), \tag{49}$$

Step 2. (parametric) Minimize the empirical version of the least squares criterion (47)

$$\widehat{\theta} = \arg \min_{\theta} \frac{1}{N_0} \sum_{p=1}^{N_0} \left(\widehat{R}_s(\mathbf{x}_{[p]}^{(s)}) - F(\mathbf{x}_{[p]}^{(s)}, \theta) \right)^2. \tag{50}$$

Lemma 1. Let $F(\mathbf{u}_k^{(s)}, \theta)$ be Lipschitz with respect to all u_{k-l} 's included in $\mathbf{u}_k^{(s)}$ and all parameters in θ . If the error of non-parametric estimate behaves like $\left\| \widehat{R}_s(\mathbf{x}_{[p]}^{(s)}) - F(\mathbf{x}_{[p]}^{(s)}, \theta^*) \right\| = O(N^{-\tau})$ in the mean square sense, for all $p = 1, 2, \dots, N_0$, then

$$\left\| \widehat{\theta} - \theta^* \right\| = O(N^{-\tau}) \tag{51}$$

in the parametric step 2.

Proof. The property (51) can be proven following the lines of the proof of Theorem 1 in [25]. \square

Remark 2. Fulfillment of (46) and the method of non-linear optimization in (50) are strictly dependent on the specifics of the problem. In active experiment, when the input can be generated arbitrarily, appropriate selection of the points $\{\mathbf{x}_{[p]}^{(s)}\}_{p=1}^{N_0}$ can significantly simplify operations in Step 2 (see example below).

Example 1. For the system

$$y_k = e^{\theta_1 u_k} + \theta_2 u_k u_{k-1} + z_k, \tag{52}$$

in step 1 we can estimate two-dimensional ($s = 2$) regression function $\widehat{R}_2(\mathbf{x}^{(2)})$ in $N_0 = 2$ representative points $\mathbf{x}_{[1]}^{(2)} = (1, 0)^T$, and $\mathbf{x}_{[2]}^{(2)} = (1, 1)^T$, i.e., compute the pattern

$$\widehat{R}_2(\mathbf{x}_{[1]}^{(2)}), \quad \widehat{R}_2(\mathbf{x}_{[2]}^{(2)}). \tag{53}$$

Since the true values of the regression function are respectively $R_2(\mathbf{x}_{[1]}^{(2)}) = e^{\theta_1}$ and $R_2(\mathbf{x}_{[2]}^{(2)}) = e^{\theta_1} + \theta_2$, in step 2 we get trivial estimates of parameters

$$\widehat{\theta}_1 = \log \widehat{R}_2(\mathbf{x}_{[1]}^{(2)}), \tag{54}$$

$$\widehat{\theta}_2 = \widehat{R}_2(\mathbf{x}_{[2]}^{(2)}) - \widehat{R}_2(\mathbf{x}_{[1]}^{(2)}). \tag{55}$$

5. Simulation Example

To illustrate the proposed method, we simulated simple Wiener system (see Figure 2) with

$$x_k = 0.5x_{k-1} + u_k \tag{56}$$

$$v_k = \arctg(x_k) \tag{57}$$

$$y_k = v_k + z_k \tag{58}$$

excited by random process uniformly distributed on equidistant set of points

$$u_k \sim \{-1, -0.75, -0.5, -0.25, 0, 0.25, 0.5, 0.75, 1\}, \tag{59}$$

and uniformly distributed output disturbance

$$z_k \sim U[-0.1, 0.1]. \tag{60}$$

For $N = 10^4$ simulated input–output pairs $\{(u_k, y_k)\}_{k=1}^N$, the non-parametric models $\widehat{R}_s(\mathbf{u}_k)$ were computed for $s = 1, 2, 3$ and compared with respect to the following empirical error

$$\delta_s = \frac{1}{N - s + 1} \sum_{k=s}^N \left(y_k - \widehat{R}_s(\mathbf{u}_k^{(s)}) \right)^2. \tag{61}$$

The results are presented in Table 1 and Figures 5 and 6. Explicit derivation of the true finite order (2D or 3D) regression function is problematic, owing to the fact that the neglected part of input signal, i.e., the ‘tail’ connected with terms $\{u_{k-\tau}\}_{\tau=s+1}^{\infty}$, is transferred through the non-linear characteristic $\arctan(\cdot)$. Figures 5 and 6 illustrate non-parametric character and non-linear properties of the model, and give a general view on the shape of input–output relationship, which can be helpful for eventual parametrization. Quantized input (59) can be a good choice, when $s = \text{const} < \infty$, and the non-parametric estimate $\widehat{R}_s(\mathbf{u}_k^{(s)})$ plays supporting role for non-linear least squares-based parameter estimation in step 2. Nevertheless, in purely non-parametric approach, i.e., for $s(N) \rightarrow \infty$, the number of possible realizations of $\mathbf{u}_k^{(s)}$ increases exponentially. In the considered example, for 9 points in (59), probability of kernel selection for each $\mathbf{x}^{(s)} = \mathbf{u}_k^{(s)}$ behaves like $\sim 9^{-s}$ and the estimate becomes sensitive on the noise z_k .

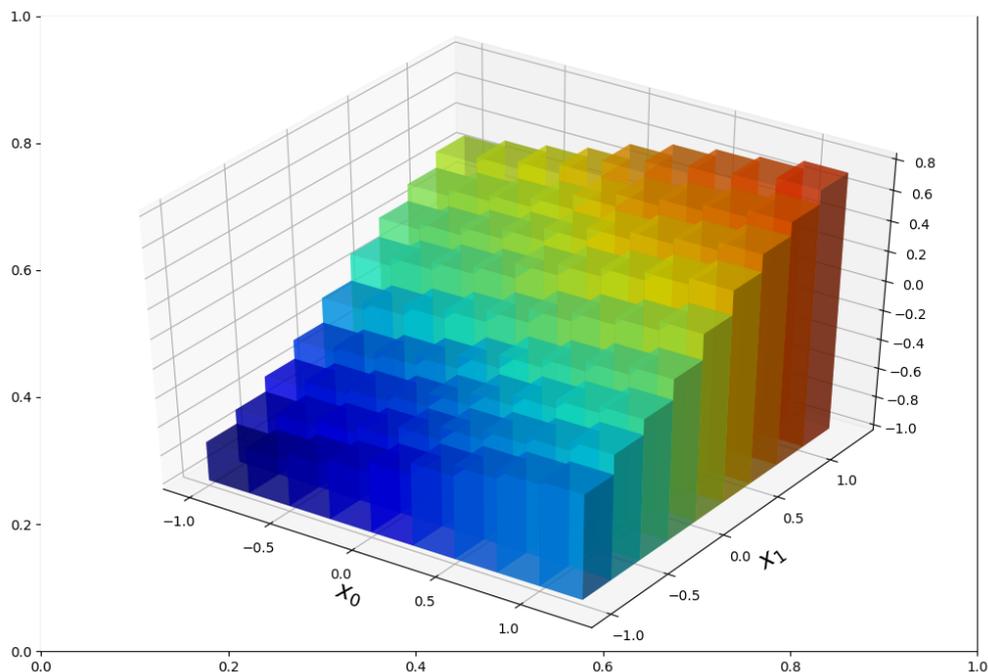


Figure 5. Two-dimensional regression-based model $\widehat{R}_2(x_0, x_1)$.

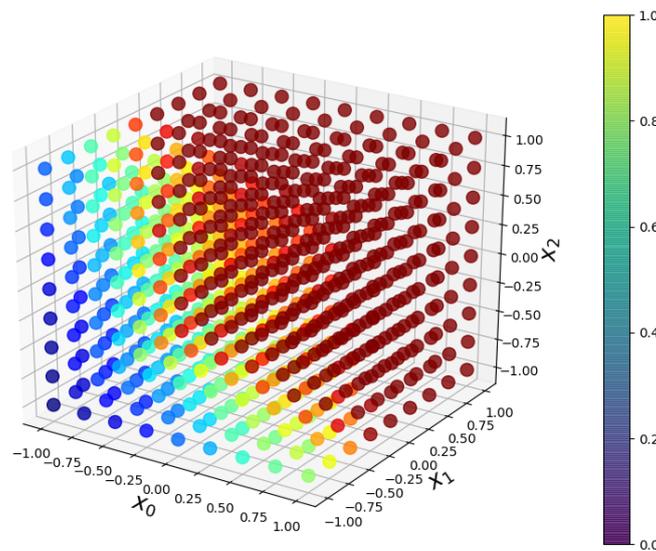


Figure 6. Three-dimensional regression-based model $\hat{R}_3(x_0, x_1, x_2)$.

Table 1. Mean squared errors δ_s of model outputs for $s = 1, 2, 3$, compared to best linear approximation (BLA).

	δ_1	δ_2	δ_3
\hat{R}_s	0.0656	0.0279	0.0137
BLA	0.0660	0.0304	0.0176

Table 1 illustrates reduction of error with respect to scale of the regression. The results are also compared with the best linear approximations $FIR(s)$ of Wiener system. We emphasize that improvement is achieved under mild prior restrictions, i.e., the non-linear model is built based on measurements knowledge only.

6. Application in Testing of Chalcogenide Glasses with the Use of DSC Method

In this section, we apply the proposed algorithm for identification of heating process in Differential Scanning Calorimeter [31].

6.1. Chalcogenide Glasses

Materials with non-linear optical properties play a key role in frequency conversion and optical switching. One of the most promising materials in this area are chalcogenide glasses, because of good non-linear, passive and active properties. They are considered to be optical medium for the fibers of the 21st century. Chalcogenide glass fibers transmit into the IR, hence they can have numerous potential applications in the civil, medical and military areas. The IR light sources, lasers and amplifiers developed using these phenomena will be very useful in many areas of industry. High-speed optical communication requires ultra-fast all-optical processing and switching capabilities. In DSC experiment energy (heat flow) in function of time or temperature could be established. The energy from an external source is required to set to zero the difference in temperature of the tested and reference samples. Both samples are heated or cooled in a controlled mode and both techniques enable the detection of thermal events observed in the physical or chemical transformation under the influence of the changing temperature in a specific manner. Owing to that, many thermodynamically important parameters can be established, e.g., a glass transition or softening temperature, melting temperature, and melting enthalpy. The results also allow observation of physical aging processes. The goal is to control temperature of heating module precisely and ensure linearity of it. It is planned to design Model Following Control (MFC) structure of system to optimize quality indexes of temperature controlling. Below we present the results of identification experiment.

6.2. Results of Experiment

Treating the sample temperature as system output y_k , and the power of the heating element as input u_k , the non-parametric multi-dimension regression model $\hat{R}_s(\mathbf{u}_k^{(s)})$ was computed for $s = 1, 2, 3, 4$. The results are shown in Figures 7–10.

Differential Scanning Calorimeter for chalcogenide glasses (built by members of the team), was first approximated by the linear model, and the results were not satisfying. To improve accuracy, Hammerstein model was applied, and the decision of model structure was made arbitrarily. To avoid the risk of bad parameterization the general approach presented in the paper was applied. The results are comparable, although obtained without making any restrictive assumptions about the block-oriented structure of the model.

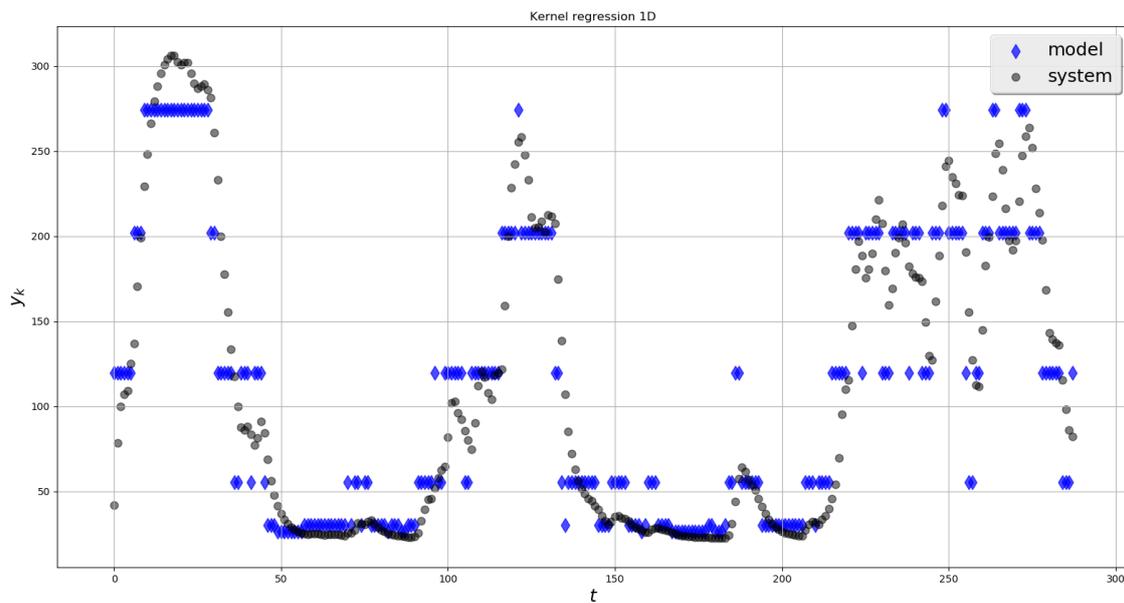


Figure 7. Experimental data—1-dimensional regression.

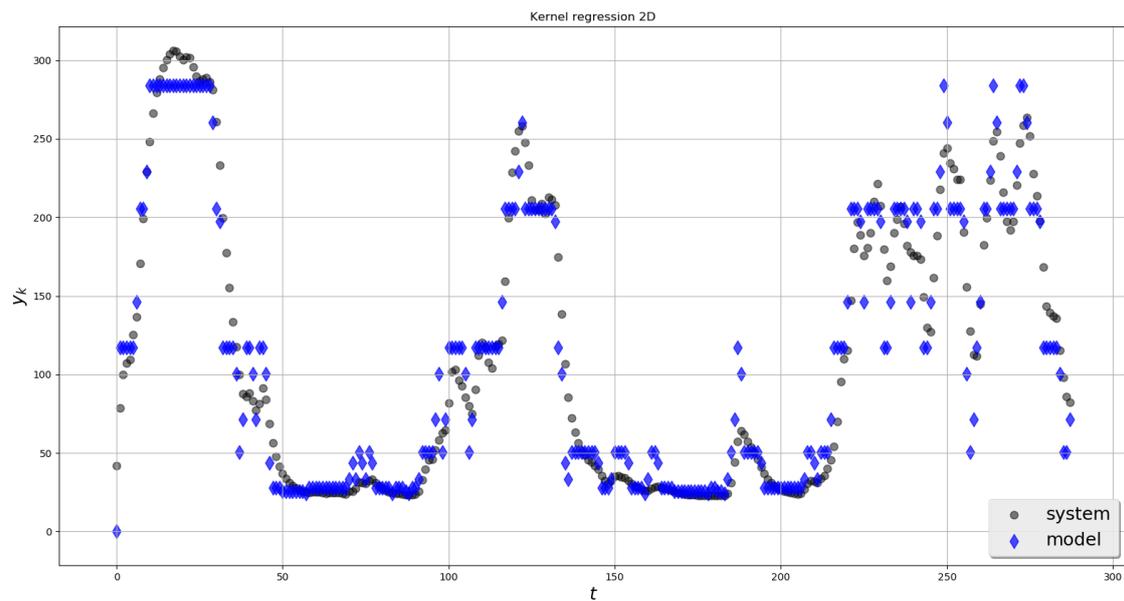


Figure 8. Experimental data—2-dimensional regression.

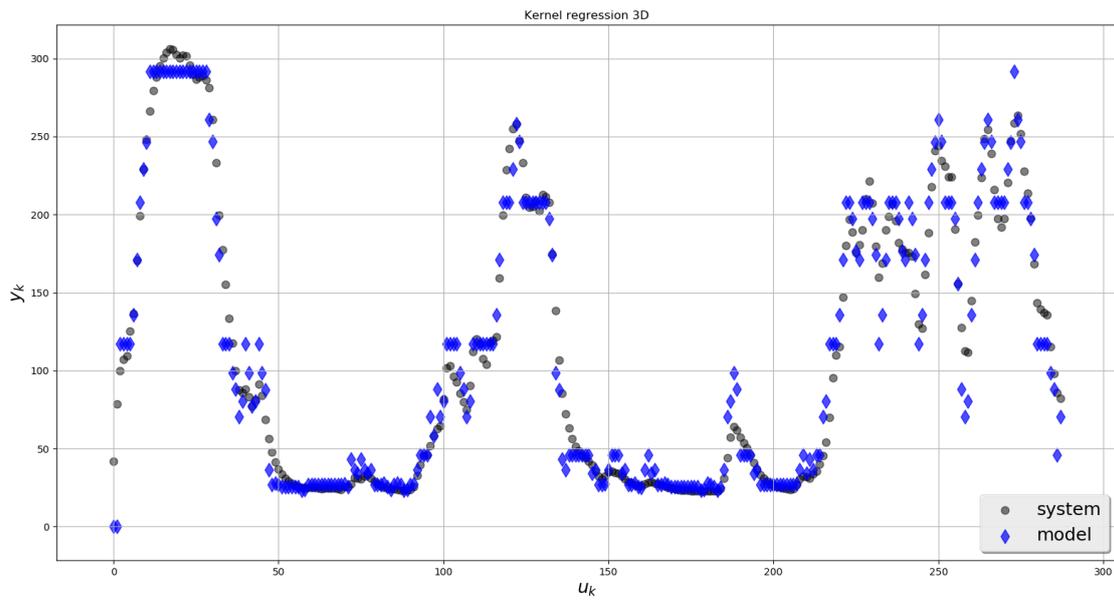


Figure 9. Experimental data—3-dimensional regression.

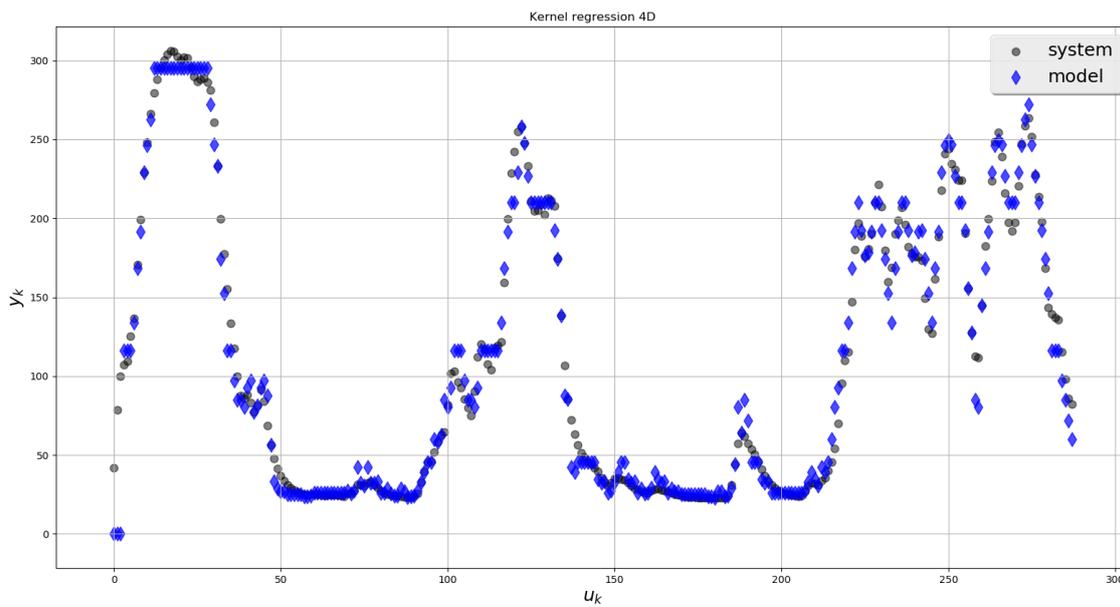


Figure 10. Experimental data—4-dimensional regression.

In Table 2 resulting mean squared errors for various scales $s = 1, 2, 3, 4$ are shown. The strong point of the method is that asymptotically, as $s(N) \rightarrow \infty$, the model becomes free of approximation error, on the contrary to linear or block-oriented representation.

Table 2. Mean squared errors δ_s of model outputs for $s = 1, 2, 3, 4$.

	δ_1	δ_2	δ_3	δ_4
\hat{R}_s (presented method)	1017	477	232	169
BLA (Linear FIR(s))	1710	1331	1165	1114
Hammerstein polynomial (3rd order + FIR(s))	1102	553	296	202

The results have been compared to FIR(s) linear model and parametric Hammerstein model. Regarding non-parametric modeling of Hammerstein systems, proposed by Greblicki and Pawlak in the 1980s [20], their algorithms suffer from correlation of input, and they are not applicable here. On the other hand, for parametric Hammerstein model, the results are strongly dependent on the arbitrarily selected basis functions of nonlinearity. In our experiment we applied 3rd order polynomial function $\mu(\cdot)$ connected in cascade with FIR(s) linear dynamic filter. Table 2 shows that our method is more accurate, emphasizing that it works under mild prior restrictions.

7. Conclusions

The main contribution of the work lays in the fact that the model is built with lack of prior knowledge about the structure of the system and its characteristics. No decision of using particular Hammerstein or Wiener model is needed at the beginning to start the procedure. Obtained non-parametric estimators $\widehat{R}_s(\mathbf{x}_{[p]}^{(s)})$ can be eventually plugged into the least squares optimization criterion in step 2, to provide parametric representation of the relationship. Both parametric and non-parametric methods can be combined to design strategy, which includes advantages of both approaches. Step 1 (non-parametric) is run to estimate selected points of system characteristic. It is done effectively thanks to generation of specific input excitation (discrete or periodic), which allows to avoid the problem of high dimensionality. Moreover, appropriate selection of estimation points can significantly decrease level of difficulty of the non-linear optimization task in step 2. The rate of convergence of parameter estimate is the same as for non-parametric ones.

The scheme presented in the paper is universal for a broad class of systems including Hammerstein and Wiener structures, and their interconnections. Non-parametric data pre-filtering plays also the role of compression algorithm and the result of step 1 can be treated as the simplified pattern of system for eventual structure detection and selection of its best parametric model. Regression-based non-parametric model can be computed only for the set of selected points, and the resulting pairs $\left\{(\mathbf{x}_{[p]}^{(s)}, \widehat{R}_s(\mathbf{x}_{[p]}^{(s)}))\right\}_{p=1}^{N_0}$ can be used as compressed pattern (as $N_0 \ll N$) of the system, instead of N data points $\left\{(\mathbf{u}_k^{(s)}, \widehat{R}_s(\mathbf{u}_k^{(s)}))\right\}_{k=1}^N$.

Non-parametric pattern $\widehat{R}_s(\mathbf{x}_{[1]}^{(s)}), \dots, \widehat{R}_s(\mathbf{x}_{[N_0]}^{(s)})$ can help to support decision of model selection from the list of potential candidates, and model competitions can be performed in the user-defined regions of interests, e.g., in the working points.

Author Contributions: Formal analysis, G.M.; methodology, Z.H.; software, P.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research (including APC) was funded and supported by the National Science Centre, Poland, Grant No. 2016/21/B/ST7/02284.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Giannakis, G.; Serpedin, E. A bibliography on nonlinear system identification. *Signal Process.* **2001**, *81*, 533–580.
2. Schetzen, M. *The Volterra and Wiener Theories of Nonlinear Systems*; John Wiley & Sons: Hoboken, NJ, USA, 1980.
3. Batselier, K.; Chen, Z.; Wong, N. Tensor Network alternating linear scheme for MIMO Volterra system identification. *Automatica* **2017**, *84*, 26–35.
4. Birpoutsoukis, G.; Marconato, A.; Lataire, J.; Schoukens, J. Regularized non-parametric Volterra kernel estimation. *Automatica* **2017**, *82*, 324–327.
5. Narendra, K.; Gallman, P.G. An iterative method for the identification of nonlinear systems using the Hammerstein model. *IEEE Trans. Autom. Control* **1966**, *11*, 546–550.

6. Giri, F.; Bai, E.W. *Block-Oriented Nonlinear System Identification*; Lecture Notes in Control and Information Sciences; Springer: Berlin/Heidelberg, Germany, 2010.
7. Bai, E.; Li, D. Convergence of the iterative Hammerstein system identification algorithm. *IEEE Trans. Autom. Control* **2004**, *49*, 1929–1940.
8. Billings, S.; Fakhouri, S. Identification of systems containing linear dynamic and static nonlinear elements. *Automatica* **1982**, *18*, 15–26.
9. Chang, F.; Luus, R. A noniterative method for identification using Hammerstein model. *IEEE Trans. Autom. Control* **1971**, *16*, 464–468.
10. Giri, F.; Rochdi, Y.; Chaoui, F. Parameter identification of Hammerstein systems containing backlash operators with arbitrary-shape parametric borders. *Automatica* **2011**, *47*, 1827–1833.
11. Śliwiński, P. Lecture Notes in Statistics. In *Nonlinear System Identification by Haar Wavelets*; Springer: Berlin/Heidelberg, Germany, 2013.
12. Stoica, P.; Söderström, T. Instrumental-variable methods for identification of Hammerstein systems. *Int. J. Control* **1982**, *35*, 459–476.
13. Bershad, N.; Celka, P.; Vesin, J. Analysis of stochastic gradient tracking of time-varying polynomial Wiener systems. *IEEE Trans. Signal Process.* **2000**, *48*, 1676–1686.
14. Chen, H.F. Recursive identification for Wiener model with discontinuous piece-wise linear function. *IEEE Trans. Autom. Control* **2006**, *51*, 390–400.
15. Hagenblad, A.; Ljung, L.; Wills, A. Maximum likelihood identification of Wiener models. *Automatica* **2008**, *44*, 2697–2705.
16. Lacy, S.; Bernstein, D. Identification of FIR Wiener systems with unknown, non-invertible, polynomial non-linearities. *Int. J. Control* **2003**, *76*, 1500–1507.
17. Vörös, J. Parameter identification of Wiener systems with multisegment piecewise-linear nonlinearities. *Syst. Control Lett.* **2007**, *56*, 99–105.
18. Wigren, T. Convergence analysis of recursive identification algorithms based on the nonlinear Wiener model. *IEEE Trans. Autom. Control* **1994**, *39*, 2191–2206.
19. Pintelon, R.; Schoukens, J. *System Identification: A Frequency Domain Approach*; Wiley-IEEE Press: Hoboken, NJ, USA, 2004.
20. Greblicki, W.; Pawlak, M. *Nonparametric System Identification*; Cambridge University Press: Cambridge, UK, 2008.
21. Györfi, L.; Kohler, M.; Krzyżak, A.; Walk, H. *A Distribution-Free Theory of Nonparametric Regression*; Springer: New York, NY, USA, 2002.
22. Härdle, W. *Applied Nonparametric Regression*; Cambridge University Press: Cambridge, UK, 1990.
23. Wand, M.; Jones, H. *Kernel Smoothing*; Chapman and Hall: London, UK, 1995.
24. Hasiewicz, Z.; Mzyk, G. Combined parametric-nonparametric identification of Hammerstein systems. *IEEE Trans. Autom. Control* **2004**, *48*, 1370–1376.
25. Hasiewicz, Z.; Mzyk, G. Hammerstein system identification by non-parametric instrumental variables. *Int. J. Control* **2009**, *82*, 440–455.
26. Mzyk, G. A censored sample mean approach to nonparametric identification of nonlinearities in Wiener systems. *IEEE Trans. Circuits Syst.* **2007**, *54*, 897–901.
27. Mzyk, G. *Combined Parametric-Nonparametric Identification of Block-Oriented Systems*; Lecture Notes in Control and Information Sciences; Springer: Berlin/Heidelberg, Germany, 2014.
28. Mzyk, G.; Wachel, P. Kernel-based identification of Wiener-Hammerstein system. *Automatica* **2017**, *83*, 275–281.
29. Mzyk, G.; Wachel, P. Wiener system identification by input injection method. *Int. J. Adapt. Control Signal Process.* **2020**, *34*, 1105–1119.
30. Wachel, P.; Mzyk, G. Direct identification of the linear block in Wiener system. *Int. J. Adapt. Control Signal Process.* **2016**, *30*, 93–105.
31. Kozdraś, B.; Mzyk, G.; Mielcarek, P. Identification of the heating process in Differential Scanning Calorimetry with the use of Hammerstein model. In Proceedings of the 2020 21st International Carpathian Control Conference (ICCC), High Tatras, Slovakia, 27–29 October 2020.

32. Boyd, S.; Chua, L. Fading memory and the problem of approximating nonlinear operators with Volterra series. *IEEE Trans. Circuits Syst.* **1985**, *32*, 1150–1161.
33. Van der Vaart, A.W. *Asymptotic Statistics*; Cambridge University Press: Cambridge, UK, 2000.

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).