Nonparametric Identification of a Particular Nonlinear Time Series System

Włodzimierz Greblicki and Mirosław Pawlak

Abstract— **In this correspondence a nonlinear time series system is identified. The system has a cascade structure, that is, it consists of a nonlinear memoryless element followed by a dynamic linear system.**

Given a Gaussian input, the Hermite series based method for the recovering of the system nonlinearity is presented. The proposed identification procedure is nonparametric since it is able to be consistent for a broad class of non-polynomial characteristics. The consistency and rate of convergence of the procedure are established. Also, some data driven methods for selecting the optimal number of terms in the procedure are proposed. The results are illustrated by a numerical example.

I. Introduction

Identification of a system is the problem of complete determnation of its characteristics from corresponding values of input and output data.

A vast number of techniques exist for the identification of linear systems [10]. Rarely, however, is an actual physical process of the linear form.

For general nonlinear systems there are no universal identification techniques. All of them largely depend on the prior knowledge about the system, i.e., about its mathematical representation. The nonparametric nonlinear system identification methods mostly rely on the Volterra series or Wiener expansions [11]. Another approach is based on the assumption that the system structure is known but no conditions on the structure descriptor are made. Thus, the problem we face is a nonparametric one.

Some authors consider the cascade system described by the Hammerstein model shown in Fig. 1, see [6], [8], and the references cited therein. The system consists of two subsystems connected in cascade. The first of them is a memoryless nonlinearity with a transfer characteristic $m(.)$ and is followed by a linear asymptotically stable system having the impulse response sequence $\{g_i\}$. The output of the whole system is distributed by a white noise Z_i.

The purpose of this correspondence is to identify the system nonlinearity from the finite record n of input-output data. The signal V_i, interconnecting both subsystems is inaccessible to measurements. Such a problem has been investigated by a number of authors, see [1], [8].

All these authors presented algorithms for identification of both the subsystems, assuming that the nonlinear element $m(.)$ is not only continuous, but also has a polynomial form. The problem investigated by them is parametric since they estimate only a finite number of coefficients. It is clear that if $m(.)$ is not a polynomial, their identification algorithms do not converge to $m(x)$. Furthermore, they do not provide any rigorous convergence analysis of the proposed procedures.

W. Greblicki is with the Institute of Engineering Cybernetics, Technical University of Wrocław, 50-370 Wrocław, Poland

M. Pawlak is with the Institute of Engineering Cybernetics, Technical University of Wrocław, 50-370 Wrocław, Poland

Nonparametric identification of the Hammerstein system has been studied in [4]-[6]. The proposed method (nonparametric kernel regression estimate) cannot have a polynomial form and it exhibits a slower rate of convergence than an estimate studied in this correspondence.

In this correspondence we propose, assuming that the input is a white Gaussian signal, an estimate which has a polynomial form and yet is consistent for any Borel measurable odd nonlinear characteristics that do not increase faster than $e^{-x^2/4}$. No assumptions concerning continuity of $m(.)$ are made. The estimate employs the Hermite polynomial system with appropriate selected number of terms. It is shown that the algorithm converges to the characteristic in the mean integrated square error sense as the number of observations tends to infinity. The rate of convergence is also evaluated.

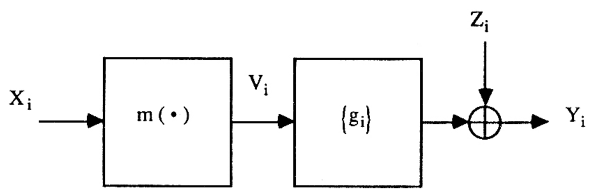

Fig. 1. The identified nonlinear system.

The estimate construction requires selection of an integer-valued parameter-the truncation point. A data-driven method for selecting that value is proposed.

Finally, let us note that the linear part of the system can be identified independently from the technique used for the nonlinear characteristic [1], [6].

II. Identification Problem

In this section we propose an estimate of the nonlinear characteristic $m(\cdot)$. We assume the input $\{X_i, i = ... - 1, 0, 1, ...\}$ is a stationary white random process with Gaussian $N(0,1)$ density

$$f(x) = (2\pi)^{-1/2} e^{-x^2/2}. \qquad (1)$$

Moreover, $m(\cdot)$ is a Borel measurable function such that

$$\int m^2(x) e^{-x^2/2} dx < \infty. \qquad (2)$$

The function is assumed to be odd, i.e..

$$m(-x) = -m(x), \text{ all } x. \qquad (3)$$

Nevertheless, some further extensions to weaker assumptions are discussed later in Section V.

The linear subsystem is described by the following state-space equation:

$$\begin{cases} \xi_{i+1} & = A\xi_i + bV_i \\ Y_i & = c^T \xi_i + Z_i \end{cases}$$

where ξ_i is the state vector and A is an asymptotically stable matrix. Clearly, $g_0 = 0$ and $g_i = c^T A^{i-1}$ for $i \geq 1$. Let also $g_1 = 1$.

The noise $\{Z_i\}$ is a stationary white process with zero mean and finite variance σ^2. The noise is independent of the input. Owing to these assumptions $\text{var}[Y_n] = Em^2(X_0)\sum_{i=1}^{\infty} g_i^2 + \sigma^2$ and $\text{cov}[Y_n, Y_{n+s}] = Em^2(X_0)\sum_{i=1}^{\infty} g_i g_{i+s} < \infty$, $s \neq 0$, i.e., the output process $\{Yn\}$ is stationary. Thus, the restriction (2) is totally ndependent of the identification method. The characteristic is identified with the help of the Hermite polynomial system $\{h_k(x); k = 0, 1, 2, ...\}$, where

$$h_k(x) = (2^k k! \pi^{1/2})^{-1/2} H_k(x),$$

and

$$H_k(x) = (-1)^k e^{x^2}(d^k/dx^k)e^{-x^2}$$

is the kth Hermite polynomial [12], i.e., $H_0(x) = 1$, $H_1(x) = 2x$, $H_k(x) = 2xH_{k-1}(x) - 2(k-1)H_{k-2}(x)$ for $k = 2, 3, \ldots$. It is well known [12] that $\{h_k(x)\}$ constitutes an orthogonal complete system with respect to the weighting function e^{-x^2}.

Clearly

$$m(x) \sim \sum_{k=0}^{\infty} a_k h_k(x) \tag{4}$$

where

$$\begin{aligned} a_k &= \int m(x) h_k(x) e^{-x^2} dx \\ &= (2\pi)^{1/2} E\{m(X) h_k(X) e^{-X^2/2}\}. \end{aligned}$$

Here and throughout the correspondence X as well as ξ, Y, and V stand for X_0, ξ_0, Y_0, and V_0, respectively. Owing to the fact that the input density is even and (3), we get $EV = 0$ and $E\xi = 0$. Thus

$$E\{Y_{i+1}|X_i = x\} = g_1 m(x) = m(x).$$

Thus, the nonlinearity $m(x)$ can be derived from the regression function $E\{Y_{i+1}|X_i = x\}$. Owing to this we introduce the following estimate of the nonlinear characteristic computed from (X_0, Y_0),

$$m_n(x) = \sum_{k=0}^{N(n)} a_{k,n} h_k(x) \tag{5}$$

where

$$a_{k,n} = (2\pi)^{1/2} n^{-1} \sum_{i=0}^{n-1} Y_{i+1} h_k(X_i) e^{-X_i^2/2} \tag{6}$$

is an unbiased estimate of a_k and where $\{N(n)\}$ is an integer sequence. That sequence plays the role of the truncation point and the problem of its selection is deferred to Section IV. The estimate (5), (6) is closely related with the orthogonal series estimate of a density function studied in the literature, see, for example, [2], [9], and the references cited therein. The Hermite series estimate of regression function, in the case of independent data, has been examined in [7], see also [3].

Let us observe that for $m(\cdot)$ having a finite representation of type (4), i.e., for $m(x) = \sum_{k=0}^{M} a_k h_k(x)$, M known, i.e., for $m(\cdot)$ being a polynomial of a finite and known order, our estimate is unbiased if only $N(n) = M$.

In general, however, the form of the characteristic is unknown and the estimate is biased, which is a rule with nonparametric estimating.

Let the mean integrated square error be chosen as a discrepancy measure between m_n and m, i.e.,

$$\text{MISE}(m_n) = E\int (m_n(x) - m(x))^2 e^{-x^2} dx.$$

This is a very commonly used measure in nonparametric curve estimation [3].

We give conditions under which the error converges to zero as the number of observations increases to infinity. Moreover, the asymptotic rate of convergence is established. Those results reveal that there is an optimal value of the truncation point $N(n)$ minimizing $\text{MISE}(m_n)$. A data-driven method of estimating that value is proposed.

III. Integrated Error

Clearly, by virtue of Parseval's formula and unbiasedness of $a_{k,n}$,

$$\begin{aligned} \text{MISE}(m_n) &= \sum_{k=0}^{N(n)} E(a_{k,n} - a_k)^2 + \sum_{k=N(n)+1}^{\infty} a_k^2 \\ &= \sum_{k=0}^{N(n)} \text{var}[a_{k,n}] + \sum_{k=N(n)+1}^{\infty} a_k^2 \end{aligned} \tag{7}$$

By virtue of (2), $\sum_{k=0}^{\infty} a_k^2 < \infty$, and the condition

$$N(n) \to \infty \tag{8}$$

implies that the second term in (7) vanishes as n tends to infinity. In order to verify a similar property for the first term in (7) we have the following lemma.

Lemma 1: For $i > 0$ we have

$$\begin{aligned} Q_{i,k} &= \text{cov}[\xi_1 h_k(X_0) e^{-X_0^2/2}, \xi_{i+1} h_k(X_i) e^{-X_i^2/2}] \\ &= R_{i,k} E\{h_k(X) e^{-X^2/2}\}, \end{aligned}$$

where

$$R_{i,k} = \begin{cases} A^i P E\{h_k(X) e^{-X^2/2}\} \\ \quad + A^i b b^T E\{m^2(X) h_k(X) e^{-X^2/2}\}, \text{for } i > 1 \\ A P E\{h_k(X) e^{-X^2/2}\} \\ \quad + A b b^T E\{m(X) h_k(X) e^{-X^2/2}\}, \ \text{for } i = 1, \end{cases}$$

and where $P = AE\{\xi\xi^T\}A^T$.

The lemma can be proved similarly as in [4].

Corollary 1: Recalling the fact that $E\{h_k(X)e^{-X^2/2}\}$ is equal to zero or $(2\pi^{1/2})^{-1/2}$ as $k>0$ or $k=0$, respectively, and $E\{m(X)h_k(X)e^{-X^2/2}\}=0$, we obtain $Q_{i,k}=0$ for $k>0$ and

$$Q_{i,0}=\begin{cases}(2\pi^{1/2})^{-1}A^iP \\ +(2\pi^{1/2})^{-1}A^ibb^TE\{m^2(X)h_k(X)e^{-X^2/2}\}, \\ \qquad\qquad\text{for } i>1 \\ (2\pi^{1/2})^{-1}AP(2\pi^{1/2})^{-1}, \text{ for } i=1.\end{cases}$$

We are now able to verify the following theorem.

Theorem 1: Let matrix A be asymptotically stable and let (1)–(9) hold. If (8) and

$$N^{5/6}(n)/n\to\infty \tag{9}$$

are satisfied then

$$\mathrm{MISE}(m_n)\to 0 \text{ as } n\to\infty.$$

Proof: Due to the stationarity, we get

$$\begin{aligned}\mathrm{var}[a_{k,n}]&=2\pi n^{-1}\mathrm{var}[Y_1h_k(X_0)e^{-X_0^2/2}]\\&+4\pi n^{-2}\sum_{i=1}^{n-1}(n-i)\\&\times\mathrm{cov}[Y_1h_k(X_0)e^{X_0^2/2},Y_{i+1}h_k(X_i)e^{X_i^2/2}].\end{aligned} \tag{10}$$

Clearly, the second term on the right-hand side of (10) equals $4\pi n^{-2}\sum_{i=1}^{n-1}(n-i)c^TQ_{i,k}c$. This, in turn, by virtue of the lemma and Corollary 1 is equal to 0 for $k>0$ and

$$2\pi^{1/2}n^{-2}(n-1)c^TAPc+2\pi^{1/2}n^{-1}\sum_{i=2}^{n-1}\left(1-\frac{i}{n}\right)c^TA^iPc$$

$$+(8\pi^{1/2})^{1/2}n^{-1}\sum_{i=2}^{n-1}\left(1-\frac{i}{n}\right)c^TA^ibb^TcE\{m^2(X)e^{-X^2/2}\}$$

for $k=0$. The asymptotic stability of A implies that the second term on the right-hand side of (10) is of the order $O(n^{-1})$ for $k=0$ and 0 otherwise. Let us now take the first term in (10) into account. Clearly,

$$\begin{aligned}&\mathrm{var}[Y_1h_k(X_0)e^{-X_0^2/2}]\\&=c^T\mathrm{cov}[\xi_1h_k(X_0)e^{-X_0^2/2},\xi_1h_k(X_0)e^{-X_0^2/2}]c\\&+\sigma^2E\{h_k^2(X)e^{-X^2}\}.\end{aligned} \tag{11}$$

The first term on the right-hand side in (11) is equal to $\mathrm{var}[m(X)h_k(X)e^{-X^2/2}]$. Since $|h_k(x)e^{-x^2/2}|\le c(k+1)^{-1/12}$, [12, p. 242], therefore the expression in (11) is not greater than

$$c^2(k+1)^{-1/6}(Em^2(X)+\sigma^2).$$

Thus

$$\sum_{k=0}^{N(n)}\mathrm{var}[a_{k,n}]\le c_1N^{5/6}(n)/n+c_2/n,$$

where $c_1=(24\pi/5)c^2(E\{m^2(X)\}+\sigma^2)$ and c_2 is independent on n. The proof of Theorem 1 has been completed. □

Remark 1: Since $|h_k(x)e^{-x^2/2}|\le c(k+1)^{-1/4}$ for every x in the finite interval [12], then the condition in (9) can be sharpened to $N^{1/2}(n)/n\to 0$ for the characteristics with compact support. That is, if $N(n)=[n^\alpha]$ then only $0<\alpha<6/5$ is allowed to get the consistency, or $0<\alpha<6/5$ for functions with compact support.

Let us now evaluate the rate of convergence of the procedure (5). To do this we need to impose some smoothness conditions on the nonlinear characteristic. Suppose that $\varphi(x)=m^{(1)}(x)-2xm(x)$ is square integrable with weight e^{-x^2}. Then, using integration by parts and formula $h_{k+1}^{(1)}(x)=-(2(k+1))^{1/2}h_k(x)$ we have

$$b_k=(2k)^{1/2}a_{k-1} \tag{12}$$

where $b_k=\int\varphi(x)h_k(x)e^{-x^2}dx$ is the kth coefficient of the expansion of $\varphi(x)$. Owing to (12) one can easily evaluate the second term in (7). Hence

$$\begin{aligned}\sum_{k=N(n)+1}^{\infty}a_k^2&=\sum_{k=N(n)+2}^{\infty}b_k^2/2k\\&\le\left(\int\varphi^2(x)e^{-x^2/2}dx\right)/2N(n).\end{aligned}$$

This, along with the fact that the first term in (7) is of order $N^{5/6}(n)/n$ yield

Theorem 2: Let all the assumptions of Theorem 1 be satisfied. Let, moreover,

$$\int\varphi^2(x)e^{-x^2/2}dx<\infty.$$

If $N(n)\sim n^{6/11}$ then

$$\mathrm{MISE}(m_n)=O(n^{-6/11}).$$

Using Remark 1 and under conditions of Theorem 2 it is seen that for nonlinear characteristics with compact support we have the rate $O(n^{-2/3})$ for the $N(n)$ selected as $[n^{2/3}]$.

The above result can be generalized for smoother $m(\cdot)$. That is, if $m(\cdot)$ has p square integrable (with weight $e^{-x^2/2}$) derivatives and if $N(n)\sim n^{6/(5+6p)}$ then $\mathrm{MISE}(m_n)=O(n^{-6p/(5+6p)})$. Hence, the more regular the characteristic, the fewer the coefficients that need be stored in computer memory. Such a property has not been observed for other types of nonparametric estimates of $m(\cdot)$.

IV. Selecting $\{N(n)\}$

The results of the last section reveal that there exists an optimal value of the truncation point, i.e., the value of $N(n)$ for which $\mathrm{MISE}(m_n)$ is minimal.

Theorem 2 states that such $N(n)$ can be of the form $\beta n^{-\alpha}$, where α depends merely on the degree of smoothness of $m(\cdot)$, whereas β might depend on unknown $m(\cdot)$, $\{g_i\}$, and σ^2. Nevertheless, the result is asymptotical, i.e., it says how to select $N(n)$ (yet we have to estimate β and α) for large n. The practical problem arises of how to choose $N(n)$ for a finite data size.

Let us denote I_N =MISE(m_n) for any fixed $n \geq 1$. Owing to (7) we have

$$\Delta_N = I_N - I_{N-1} = \text{var}[a_{N,n}] - a_N^2. \qquad (13)$$

Under the condition of Theorem 2 it is easy to show that there is N satisfying $\Delta_N \geq 0$, i.e., there exists an optimal N. Thus, one can include the term $a_{N,k}h_N(x)$ to estimate (5) if $\Delta_N < 0$. This procedure is repeated until $\Delta_M > 0$ for some M, then the optimal $N^* = M - 1$. Let us note, however, that the above inclusion rule is not feasible since we do not know var$[a_{N,n}]$ and a_N. From the proof of Theorem 1 we have

$$\text{var}[a_{N,n}] = 2\pi n^{-1} - \text{var}[Y_1 h_N(X_0)e^{-X_0^2/2}] + I(N=0)d/n, \qquad (14)$$

where d does not depend on N and $I(N=0)$ equals one for $N = 0$ and zero otherwise. Thus, for $N \geq 1$ one can ignore the second term in (15).The variance in the first term of (14) is estimated by

$$\eta_N = (n-1+\alpha)^{-1} \sum_{i=0}^{n-1} \left(Y_{i+1} h_N(X_i) e^{-X_i^2/2} - \mu_n \right) \qquad (15)$$

where $\mu_n = (2\pi)^{1/2} a_{N,n}$. The constant α reflects the correlation of the data and can be chosen by the user. It should be within the interval $(-(n-1), 1)$. The value $\alpha = 0$ corresponds to the classical variance estimate which is unbiased, assuming that observations are independent. Thus, owing to (13)–(15) one can propose the following empirical version of Δ_N:

$$\hat{\Delta} N = 4\pi \eta_N / n - a_{N,n}^2,$$

where η_N is defined in expression (15). Now, we can consider an adaptive version of the inclusion rule: Include all succesive terms which pass the test

$$\hat{\Delta} N < 0. \qquad (16)$$

Sometimes it is required to truncate the estimate whenever the rule fails for *a priori* determined number of consecutive terms. This prevents the algorithm from getting stuck in a local minimum. Such a scheme, in the context of nonparametric density estimation, was proposed in [9], see also [2], [3].

To illustrate the above concepts let us consider the system

$$\begin{cases} \xi_{i+1} = & \begin{bmatrix} 0 & 1 \\ -0.03 & -0.4 \end{bmatrix} \xi_i + \begin{bmatrix} 1 \\ 1 \end{bmatrix} m(X_i) \\ Y_i = & [0.5, 0.5]\xi_i + Z_i \end{cases}$$

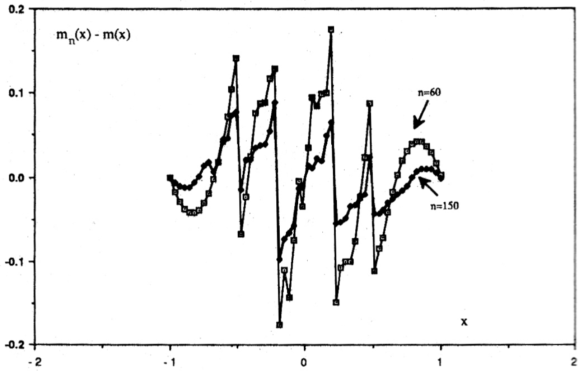

Fig. 2. Values of $m_n(x) - m(x)$, -1$\leq x \leq 1$, $n = 60$, 150.

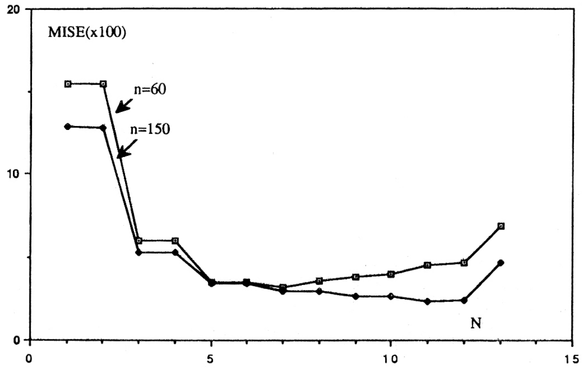

Fig. 3. MISE versus N, $n = 60$, 150.

with the nonlinearity $m(x)$ defined as

$$m(x) = \begin{cases} 0, & \text{for } |x| < 0.2 \\ 0.75, & \text{for } 0.2 \leq x \leq 0.5 \\ 1, & \text{for } x > 0.5 \\ -0.75, & \text{for } -0.2 \leq x \leq -0.5 \\ -1 & \text{for } x < -0.5. \end{cases}$$

The noise Z_i is $N(0, 0.01)$.

In Fig. 2 the pointwise error $m_n(x) - m(x)$, $-1 \leq x \leq 1$, is depicted for $N = 7$, $n = 60, 150$. The loss of the estimate efficiency in the vicinity of discontinuity points is observed. This is not surprising since the Hermite expansion, similarly as the Fourier series, undergoes the so-called Gibbs phenomenon. Fig. 3 shows the MISE as a function of N, for $n = 60, 150$. The optimal truncation points are equal to 7 and 11, respectively. The inclusion rule (16) (with $a = -n/2$) has also been applied (100 realizations of the input-output data have been generated) yielding over the 90% correct decisions concerning the optimal N's.

V. Concluding Remarks

Our assumptions concerning $m(\cdot)$ are rather weak. In particular, (2) is satisfied by all $m(\cdot)$ increasing not faster than x^p, p finite. Restriction (3) is also often met. If not,

$$E\{Y_{i+1}|X_i = x\} = m(x) + \gamma,$$

where $y = E\{m(X)\} \sum_{j=2}^{\infty} g_j$ and we are able to estimate $m(\cdot)$ up to an additive constant.

The difficulty which now arises is not a disadvantage of our identification method. It cannot be got around by any other technique. Nevertheless, if the value of $m(\cdot)$ is known at one point, say $x = \eta$, then one can define a modified estimate

$$\tilde{m}_n(x) = m_n(x) - m_n(\eta) + m(\eta).$$

Then, assuming that $m_n(x)$ is continuous at $x = \eta$ we can show that MISE$(\tilde{m}_n) \to 0$ as $n \to \infty$.

The estimate defined in (5) is tailored to the case of the $N(0, 1)$ distributed input signal. If the input is $N(0, \sigma^2)$, with unknown σ^2, one can estimate $m(x)$ by (5) with $a_{k,n}$ replaced by

$$\bar{a}_{k,n} = (2\pi)^{-1}\sigma_n n^{-1} \sum_{i=0}^{n-1} Y_{i+1} h_k(X_i) e^{-X_i^2(1-1/2\sigma_n^2)},$$

where σ_n is an estimate of σ. This results from the fact that $a_k = (2\pi)^{1/2}\sigma E\{Y_{i+1}h_k(X_i)e^{-X_i^2(1-1/2\sigma^2)}\}$. Next, observing that $(\tilde{a}_{k,n} - \bar{a}_{k,n})^2$ is of the order $O(n^{-2})$ in probability, one can establish the consistency of this estimate under assumptions analogical as in Theorem 1. Here $\bar{a}_{k,n}$ is defined as $\tilde{a}_{k,n}$ with σ_n replaced by σ.

If density $f(x)$ of $\{X_n\}$ is completely unknown, then the following estimate of $m(x)$ can be proposed:

$$\hat{m}_n(x) = \sum_{k=0}^{N(n)} \hat{a}_{k,n} h_k(x) \Bigg/ \sum_{k=0}^{M(n)} \hat{b}_{k,n} h_k(x)$$

where $\hat{a}_{k,n} = n^{-1}\sum_{i=0}^{n-1} Y_{i+1}h_k(X_i)e^{X_i^2}$ and $\hat{b}_{k,n} = n^{-1}\sum_{i=0}^{n-1} h_k(X_i)e^{X_i^2}$ and $M(n)$ is another truncation parameter, see [7]. The estimator $\hat{m}_n(x)$ is defined as the ratio of two random variables, and computation of the MISE$(\hat{m}_n)$ is not a straightforward matter. Some consistency results in the case of independent data have been reported in [7].

Acknowledgement

The authors thank one of the referees for his careful reading of the manuscript.

References

[1] D. R. Brillinger, "The identification of a particular nonlinear time series system," *Biomerrika*, vol. 64, pp. 509-515, 1977.

[2] P. Diggle and P. Hall, "The selection of terms in an orthogonal series density estimator," *J. Amer. Stat. Ass.*, vol. 81, pp. 230-233, 1986.

[3] L. Eubank, *Spline Smoothing and Nonparamerric Regression.* New York: Marcel Dekker, 1988.

[4] W. Greblicki and M. Pawlak, "Identification of discrete Hammerstein systems using kemel regression estimates," *IEEE Trans. Auromat. Contr.*, vol. AC-31, pp. 74-77, 1986.

[5] W. Greblicki and M. Pawlak, "Hammerstein system identification by nonparametric regression estimation," *Int. J. Contr.*, vol. 45, pp. 343-354, 1987.

[6] W. Greblicki and M. Pawlak, "Nonparametric identification of Hammerstein systems," *IEEE Trans. Inform. Theory*, vol. IT-35, pp. 409-418, 1989.

[7] W. Greblicki and M. Pawlak, "Fourier and Hermite series estimates of regression functions," *Ann. Inst. Stat. Math.*, vol. 31, pp. 443-459, 1985.

[8] I. W. Hunter and M. J. Korenberg, "The identification of nonlinear biological systems: Wiener and Hammerstein cascade models," *Biol. Cybern.* vol. 55, pp. 135-144, 1986.

[9] R. Kronmal and M. Tarter, "The estimation of probability densities and cumulatives by Fourier series methods," *J. Amer. Stat. Ass.*, vol. 925-952, 1968.

[10] L. Ljung, *System Idenrificufion: Theory for the User.* Englewood Cliffs, NJ: Prentice-Hall, 1987.

[11] W. J. Rugh, *Nonlinear System Theory-The Volrerra –Wiener Approach.* Baltimore, MD: Johns Hopkins University Press, 1981

[12] G. Szegö, *Orthogonal Polynomials*, Amer. Math. Soc. Coll. Pub., 1959