# Semi-Parametric Identification of Time-Varying Block-Oriented Systems

Grzegorz Mzyk

**Abstract**

In the paper, we identify nonlinear block-oriented dynamic systems with time-varying characteristics. Nonparametric kernel regression estimate for nonstationary regression estimation is introduced and applied to identification of Hammerstein and Narmax systems. For the systems with periodical changes of parameters the consistent estimate is proposed. Also the issue of the best parametric model selection of time-varying systems is discussed.

**Keywords**

System identification, time-varying systems, tracking, nonparametric methods, Hammerstein system, Wiener system.

## I. Introduction

A large number of physical systems are nonstationary. Identification of nonstationary processes have been widely studied in the literature for linear systems. Traditional techniques for identifying linear time-varying (LTV) systems are based on the recursive weighted least squares methods (see [17], [2], [9], [3]). The weights are dependent on time, in the sense that the most recent measurements are privileged, while the oldest have the smallest influence on the estimate. If the time horizon is too long, i.e. the weights decrease too slow, we obtain the bias connected with parameter changes. On the other hand, if the horizon is short, the estimate becomes sensitive on the noise and the variance error appears. The goal is thus to design a good compromise between

The author works for the Institute of Computer Engineering, Control and Robotics, Wrocław University of Technology, Janiszewskiego 11/17, 50-372 Wrocław, Poland, tel. +48 71 320 25 49, fax. +48 71 321 26 77, e-mail: Grzegorz.Mzyk@pwr.wroc.pl

bias and variance, i.e., we look for a good trade-off between tracking ability and noise rejection [10], [15], [8], [16]. Some of methods proposed in the literature use the Kalman filter approach [1], or expand changes of the coefficients in the wavelet series [21]. As regards the identification of time-varying nonlinear block-oriented (Hammerstein and Wiener) systems, comparatively little attention has been paid in the literature. It is commonly assumed that the nonlinear regression is quasi-stationary, i.e., that the system becomes stationary as the measurement index tends to infinity (see [6], [18], [19], [20]).

In this chapter we adopt some results of adaptive modeling theory of linear system for the nonlinear dynamic systems. In Section II we formulate the problem of nonparametric kernel regression estimation in nonstationary case. Next, in Section III we generalize the 3-stage algorithm, presented in [14], for parameter-varying NARMAX systems. In Section IV we consider a special case, when the system characteristics change periodically with the known time-period $T$. Finally, in Section V we show the application of the model detection method, proposed in [13], for time-varying systems.

## II. Kernel estimate for time-varying regression

Let us first consider, for simplicity of presentation, the problem of estimation of constant value $\theta^*$, observed in the presence of additive, zero-mean random noise $z_k$ with finite variance $\sigma_z^2$

$$y_k = \theta^* + z_k,$$

i.e., we recover $\theta^*$ from the observations $\{y_k\}_{k=1}^N$. In general, the measurements $\{y_k\}_{k=1}^N$ can be taken into account with different levels of significancy, i.e., we want to find the weighted least squares minimum

$$\widehat{\theta} = \arg\min_{\theta} \sum_{k=1}^{N} \alpha_k \left(y_k - \theta\right)^2, \tag{1}$$

where $\alpha_k$'s are some weights representing the priorities of respective measurements $y_k$'s. Since

$$\frac{\partial}{\partial\theta} \sum_{k=1}^{N} \alpha_k \left(y_k - \theta\right)^2 = 2 \sum_{k=1}^{N} \left(\theta\alpha_k - y_k\alpha_k\right),$$

the minimum in (1) is the solution of the equation

$$\theta \sum_{k=1}^{N} \alpha_k = \sum_{k=1}^{N} y_k\alpha_k$$

which leads to

$$\widehat{\theta} = \frac{\sum_{k=1}^{N} y_k \alpha_k}{\sum_{k=1}^{N} \alpha_k}, \tag{2}$$

where $\sum_{k=1}^{N} \alpha_k$ can be understand as effective number of measurements. In particular, if $\alpha_k$'s are the same for each $k$, i.e. $\alpha_k = \alpha = \mathrm{const}$, then we obtain the standard least squares solution

$$\widehat{\theta} = \frac{\alpha \sum_{k=1}^{N} y_k}{N\alpha} = \frac{1}{N}\sum_{k=1}^{N} y_k.$$

Non-uniform $\alpha_k$'s are commonly applied due to two main reasons: (i) time-varying estimated value $\theta^*(N) = \theta_N^*$, and (ii) error-in variable problem in local regression estimation.

(i) If the estimated value $\theta_N^*$ is changing, the popular choice is to take

$$\alpha_k = \lambda^{N-k}, \text{ with } 0 < \lambda < 1 \tag{3}$$

which damps old measurements and increase the influence of recent observations on the resulting estimate. Moreover, such a choice is very convenient for designing the recursive versions of identification algorithms, since

$$\alpha_{k+1} = \frac{1}{\lambda}\alpha_k.$$

On the other hand, since the effective number of measurements is finite, i.e. $\sum_{k=1}^{\infty} \alpha_k = \sum_{k=1}^{\infty} \lambda^{N-k} < \infty$, the variance of the estimate (2) does not tend to zero. It is a cost paid for good tracking abilities of the algorithm. The most popular models of $\theta_N^*$ variations are random walk, random walk with trends, jump changes, Markov chains and knwoledge-based descriptions (for details see [10] and the references cited therein).

Also the restrictive (hard) selection of the fixed number $n$ of the last measurements can be applied, i.e.,

$$\alpha_k = \begin{cases} 1, \text{ as } N-n < k \leq N \\ 0, \text{ as } k \leq N-n \end{cases}.$$

The estimate $\widehat{\theta}_n^{(N)}$, selecting $n$ last observations from the $N$ element data set, is then as follows

$$\widehat{\theta}_n^{(N)} = \frac{1}{n}\sum_{k=N-n+1}^{N} y_k$$

Let us assume that the true value of estimated parameter jumps, at the time instant $N$, from $\theta^*$ to $\theta^* + \Delta$, and the moments of the measurement noise $z_k$ remain the same. Let us also introduce the following quality index of the tracking procedure

$$Q(n) = \sum_{i=N+1}^{N+H} \left( \operatorname{var} \widehat{\theta}_n^{(i)} + \operatorname{bias}^2 \widehat{\theta}_n^{(i)} \right),$$

where $H$ denotes horizon of the tracking. We simply get

$$\begin{aligned} Q(n) &= \frac{H\sigma_z^2}{n} + \frac{\Delta^2}{n^2} \sum_{j=1}^{n-1} j^2 = \frac{H\sigma_z^2}{n} + \frac{\Delta^2}{n^2} \left( \frac{n(n+1)(2n+1)}{6} - n^2 \right) \\ &\simeq \frac{H\sigma_z^2}{n} + \frac{\Delta^2}{3} n, \end{aligned}$$

where the cummulated variance component $\frac{H\sigma_z^2}{n}$ dominates for small $n$, while the cummulated bias error $\frac{\Delta^2}{3} n$ inreases for large $n$ (see Fig. 1). The optimal value of $n$

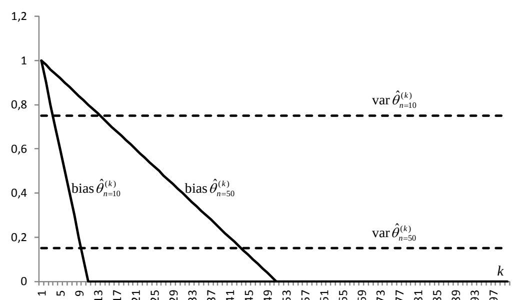


Figure 1. Sample behaviour of bias and variance after parameter jump for $n = 10$ and $n = 50$.

$$n_{opt.} = \arg\min_n Q(n)$$

obviously depends on the relation between horizon $H$, height of jump $\Delta$, and the variance of the noise $\sigma_z^2$, i.e. $n_{opt.} = n_{opt.}(H, \Delta, \sigma_z^2)$. Nevertheless, this dependence is weak, in the sense that in typical cases ($\Delta$ comparable with $\sigma_z^2$) $n_{opt}$ usually lays between $10$ and $20$ (see Example in Fig. 2). Let us also emphasize that since $\Delta$ and $\sigma_z^2$ are unknown, computing $n_{opt.}$ is not possible. Also the asymptotic convergence of the estimate cannot be achieved, because of finite number of effective measurements.

(ii) In the traditional kernel regression function estimation we recover the value of time-invariant function $\theta^* = \mu(u)$ for a given point $u$, using the pairs of observations $\{(u_k, y_k)\}_{k=1}^N$, where $u_k$'s are random (not necessary equal to $u$) and

$$y_k = \mu(u_k) + z_k.$$

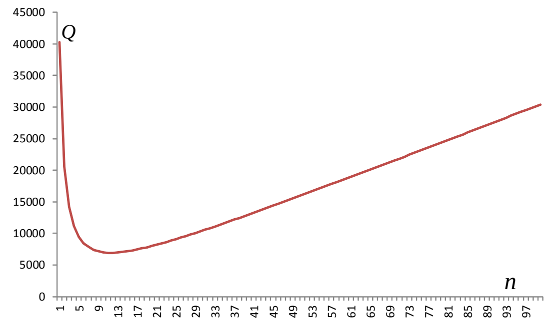


Figure 2. Dependence between tracking error $Q$ and the window length $n$, for $H = 100$, $\sigma_z = 20$, and $\Delta = 30$.

Assuming that $\mu()$ is a Lipschitz function, the pairs $(u_k, y_k)$ are taken into account with different level of significancy, and the following minimization is perfored

$$\widehat{\mu}(u) = \arg\min_{\theta} \sum_{k=1}^{N} \alpha_k \left(y_k - \theta\right)^2, \tag{4}$$

where $\alpha_k$'s are some weights representing the priorities of respective measurements $(u_k, y_k)$'s. The weights $\alpha_k$'s are dependent on the distance between $k$th input observation $u_k$, and the given estimation point $u$. For example we can set

$$\alpha_k = K\left(\frac{u_k - u}{h}\right), \tag{5}$$

where $K()$ is a kernel function, and $h$ – a bandwidth parameter.

For a *non-stationary and nonlinear block* $\theta_N^* = \mu_N(u)$, when the estimated value $\theta^*$ changes in time ($N$ is the time instant in which we want to estimate $\theta^*$), we propose to combine (3) with (5) in the following way

$$\alpha_k = \lambda^{N-k} K\left(\frac{u_k - u}{h}\right).$$

It leads to the following, nonparametric tracking procedure

$$\widehat{\theta}_N = \widehat{\mu}_N(u) = \frac{\sum_{k=1}^{N} y_k \lambda^{N-k} K\left(\frac{u_k - u}{h}\right)}{\sum_{k=1}^{N} \lambda^{N-k} K\left(\frac{u_k - u}{h}\right)}. \tag{6}$$

The generalized kernel estimate (6) with the forgetting factor $\lambda$ can be applied in the regression estimation in Hammerstein systems (see [5]), inverse regression estimation in Wiener systems (see [4]), and in the censored sample mean approach to Wiener-Hammerstein systems (see

[11] and [12]). It can also support generation of instrumental variables for the IIR Hammerstein/NARMAX systems (see [14]), and compress the data in the 3-step procedure of model recognition (see [7] and [13]).

Below, we present the results of simple experiment, in which the time-varying static characteristic $\mu_N(u) = u^2 + c_N$, with jumping offset $c_N = 1_{N-30}$ was recovered in the point $u = 0$, under random input $u_k \sim U[-1, 1]$ and in the presence of random output noise $z_k \sim U[-0.1, 0.1]$. As can be seen in Fig. 3, for small value of $\lambda$ ($\lambda = 0.6$) we observe rapid reaction with huge

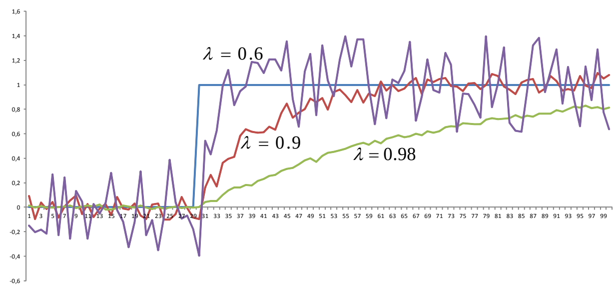


Figure 3. Tracking results for various values of the forgetting factor $\lambda$.

variance of the results. For $\lambda$ close to $1$ the variance is reduced, but it is done at the expense of inertia.

## III. Weighted least squares for NARMAX systems

In this point we analyze applicability of the recursive weighted least squares and instrumental variables algorithms and the modified kernel regression method (6) in the 3-Stage identification of additive NARMAX system (see [14]) described as follows

$$y_k = \sum_{j=1}^{p} \lambda_j \eta(y_{k-j}) + \sum_{i=0}^{n} \gamma_i \mu(u_{k-i}) + z_k \tag{7}$$

where

$$\mu(u) = \sum_{t=1}^{m} c_t f_t(u) \tag{8}$$

$$\eta(y) = \sum_{l=1}^{q} d_l g_l(y)$$

and $f_1(u),...,f_m(u),g_1(y),...,g_q(y)$ are linearly independent basis functions of known form. Below, we present the stages of the procedure for identification of parameters $c_t$ and $d_l$ of time-varying NARMAX systems.

*Stage 1 (nonparametric).* Generate empirical matrix of instruments $\widehat{\Psi}^*_{N,M} = (\widehat{\psi}^*_{1,M}, \widehat{\psi}^*_{2,M}, ...\widehat{\psi}^*_{N,M})^T$, where

$$\begin{aligned} \widehat{\psi}^*_{k,M} &= (f_1(u_k),...,f_m(u_k),...,f_1(u_{k-n}),...,f_m(u_{k-n}), \\ &\quad \widehat{R}_{1,M}(u_{k-1}),......,\widehat{R}_{q,M}(u_{k-1}),...,\widehat{R}_{1,M}(u_{k-p}),...,\widehat{R}_{q,M}(u_{k-p}))^T, \end{aligned} \tag{9}$$

and $\widehat{R}_{l,M}(u_j) = \sum_{i=1}^{j}\left(g_l(y_i)\lambda^{j-i}K(\frac{u-u_i}{h(M)})\right) / \sum_{i=1}^{j}\lambda^{j-i}K(\frac{u-u_i}{h(M)})$, with $0<\lambda<1$.

*Stage 2 (parametric).* Estimate the aggregated parameter vector

$$\begin{aligned} \theta &= (\gamma_0 c_1,...,\gamma_0 c_m,...,\gamma_n c_1,...,\gamma_n c_m,\lambda_1 d_1,...,\lambda_1 d_q,...,\lambda_p d_1,...,\lambda_p d_q)^T \\ &= (\theta_1,...,\theta_{(n+1)m},\theta_{(n+1)m+1},...,\theta_{(n+1)m+pq})^T \end{aligned} \tag{10}$$

using the recursive least squares or instrumental variables method

$$\begin{aligned} \widehat{\theta}_k^{(LS)} &= \widehat{\theta}_{k-1}^{(LS)} + P_k^{(LS)}\phi_k(y_k - \phi_k^T\widehat{\theta}_{k-1}^{(LS)}), \\ \widehat{\theta}_k^{(IV)} &= \widehat{\theta}_{k-1}^{(IV)} + P_k^{(IV)}\psi_k(y_k - \phi_k^T\widehat{\theta}_{k-1}^{(IV)}), \end{aligned} \tag{11}$$

with

$$\begin{aligned} P_k^{(LS)} &= P_{k-1}^{(LS)} - \frac{1}{1+\phi_k^T P_{k-1}^{(LS)}\phi_k} P_{k-1}^{(LS)}\phi_k\phi_k^T P_{k-1}^{(LS)}, \\ P_k^{(IV)} &= P_{k-1}^{(IV)} - \frac{1}{1+\phi_k^T P_{k-1}^{(IV)}\psi_k} P_{k-1}^{(IV)}\psi_k\phi_k^T P_{k-1}^{(IV)}, \end{aligned} \tag{12}$$

or minimize the weighted criterion (see [10])

$$\sum_{k=1}^{N}\alpha_k\left(y_k-\phi_k^T\theta\right)^2 \to \min_{\theta},$$

in the following way

$$\begin{aligned} \widehat{\theta}_k^{(LS)} &= \widehat{\theta}_{k-1}^{(LS)} + L_k^{(LS)}(y_k - \phi_k^T\widehat{\theta}_{k-1}^{(LS)}), \\ \widehat{\theta}_k^{(IV)} &= \widehat{\theta}_{k-1}^{(IV)} + L_k^{(IV)}(y_k - \phi_k^T\widehat{\theta}_{k-1}^{(IV)}), \end{aligned} \tag{13}$$

where

$$L_k^{(LS)} = \frac{P_{k-1}^{(LS)}\phi_k}{\alpha_k + \phi_k^T P_{k-1}^{(LS)}\phi_k}, \qquad L_k^{(IV)} = \frac{P_{k-1}^{(IV)}\phi_k}{\alpha_k + \phi_k^T P_{k-1}^{(IV)}\phi_k},$$

and

$$
\begin{aligned}
P_k^{(LS)} &= \frac{1}{\alpha_k}\left[P_{k-1}^{(LS)} - \frac{P_{k-1}^{(LS)}\phi_k\phi_k^T P_{k-1}^{(LS)}}{\alpha_k + \phi_k^T P_{k-1}^{(LS)}\phi_k}\right], \\
P_k^{(IV)} &= \frac{1}{\alpha_k}\left[P_{k-1}^{(IV)} - \frac{P_{k-1}^{(IV)}\phi_k\phi_k^T P_{k-1}^{(IV)}}{\alpha_k + \phi_k^T P_{k-1}^{(LS)}\phi_k}\right].
\end{aligned}
$$

*Stage 3 (decomposition).* Similarly as for time-invariant system compute the SVD (singular value decomposition) of the matrices $\widehat{\Theta}_{\lambda d}^{(IV)}$ and $\widehat{\Theta}_{\gamma c}^{(IV)}$, i.e., $\widehat{\Theta}_{\gamma c}^{(IV)} = \sum_{i=1}^{\min(n,m)} \sigma_i \widehat{\mu}_i \widehat{\nu}_i^T$, $\widehat{\Theta}_{\lambda d}^{(IV)} = \sum_{i=1}^{\min(p,q)} \delta_i \widehat{\xi}_i \widehat{\zeta}_i^T$ to obtain the estimates of changing parameters.

## IV. Nonparametric identification of periodically varying systems

Let us consider the continuous-time Hammerstein system with periodically varying nonlinear static characteristic with a priori known period $T$. The goal is to estimate periodic regression

$$R(u,t) = R(u,t+T)$$

for each $t \in [0,T]$, from the randomly sampled measurements $\{(t_k, u_k, y_k)\}$, where $t_k$ is the time instance of the collected pair $(u_k, y_k)$.

$$
\widehat{R}_N(u,t) = \frac{\sum_{k=1}^N y_k K\left(\frac{1}{h_N}\left\|\begin{bmatrix} u_k \\ t_k^* \end{bmatrix} - \begin{bmatrix} u \\ t \end{bmatrix}\right\|\right)}{\sum_{k=1}^N K\left(\frac{1}{h_N}\left\|\begin{bmatrix} u_k \\ t_k^* \end{bmatrix} - \begin{bmatrix} u \\ t \end{bmatrix}\right\|\right)}
$$

where

$$t_k^* = t_k - n_k T, \text{ and } n_k = \left\lfloor \frac{t_k}{T} \right\rfloor.$$

Let $f(u,t)$ be joint probability density function of the input $u$ and the time $t$. The following theorem holds.

*Theorem 1:* If $h_N \to 0$ and $Nh_N^2 \to \infty$, as $N \to \infty$, and moreover both $R()$ and $f()$ are continuous in the point $(u,t)$, then

$$\widehat{R}_N(u,t) \to R(u,t)$$

provided that $f(u,t) > 0$.

## V. Detection of structure changes

The modified kernel estimate for time-varying regression (see (6)) can be also applied in the first step of the procedure of model selection, described in detail in [7] and [13]. The recognition procedure is as follows (see Fig. 4).

*Step 1. Nonparametric smoothing/denoising/data-compression*

For the grid of $N_0$ selected fixed points $\overline{u}_1,\overline{u}_2,...,\overline{u}_{N_0}$, compute nonparametric estimates $\widehat{R}_N(\overline{u}_1), \widehat{R}_N(\overline{u}_2), ..., \widehat{R}_N(\overline{u}_{N_0})$ of the regression function, playing the role of the true characteristic $\mu()$.

*Step 2. Least squares approximation*

For each class $l = 1, 2, ..., M$ minimize the loss function

$$\widehat{Q}_l(\theta_l) = \frac{1}{N_0}\sum_{i=1}^{N_0}\left(\widehat{R}_N(\overline{u}_i) - \overline{R}^{(l)}\left(\overline{u}_i, \theta_l\right)\right)^2, \tag{14}$$

being the empirical counterpart of the least squares criterion

$$Q_l(\theta_l) = \frac{1}{N_0}\sum_{i=1}^{N_0}\left(R(\overline{u}_i) - \overline{R}^{(l)}\left(\overline{u}_i, \theta_l\right)\right)^2,$$

with respect to $\theta_l$, getting the estimate

$$\widehat{\theta}_l^* = \arg\min_{\theta_l} \widehat{Q}_l(\theta_l) \tag{15}$$

of the best parameters $\theta_l^*$ in the $l$-th class.

*Step 3. Nearest neighbour model selection*

Select the 'nearest neighbour' model, i.e.

$$\left\{\overline{R}^{(l_0)}\left(u, \widehat{\theta}_{l_0}^*\right)\right\}, \tag{16}$$

where

$$l_0 = \arg\min_l [\varkappa_l] \Delta_l(\widehat{\theta}_l^*). \tag{17}$$

Obviously, when the true system characteristic changes in time, then, the index of optimal approximating class is also function of time. The only needed modification is to replace the standard kernel estimate in Step 1, with the weighted approach, given in (6). Steps 2 and 3 of the procedure remain unchanged. Such a strattegy seems to be an interesting proposition in switched system control and fault detection.

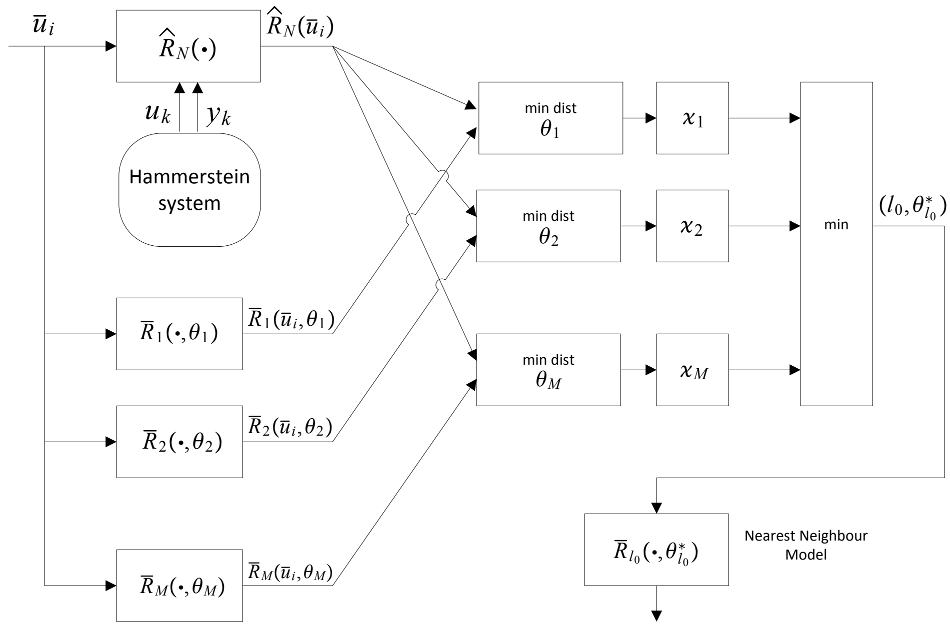


Figure 4. Scheme of the 3-step method of model selection

# References


[1] R.G. Brown and P. Hwang. *Introduction to random signals and applied Kalman filtering*. New York, John Wiley & Sons, 1992.

[2] S. Chen, S.A. Billings, and W. Luo. Orthogonal least squares methods and their application to non-linear system identification. *International Journal of Control*, 50(5):1873–1896, 1989.

[3] W.J. Fitzgerald, R.L. Smith, A.T. Walden, and P.C. Young. *Nonlinear and Nonstationary Signal Processing*. Cambridge University Press, 2001.

[4] W. Greblicki. Nonparametric identification of Wiener systems. *Information Theory, IEEE Transactions on*, 38(5):1487–1493, 1992.

[5] W. Greblicki and M. Pawlak. *Nonparametric System Identification*. Cambridge University Press, 2008.

[6] W. Greblicki, D. Rutkowska, and L. Rutkowski. An orthogonal series estimate of time-varying regression. *Annals of the Institute of Statistical Mathematics*, 35(1):215–228, 1983.

[7] Z. Hasiewicz, G. Mzyk, and P. Śliwiński. Quasi-parametric recovery of Hammerstein system nonlinearity by smart model selection. *Artifical Intelligence and Soft Computing*, pages 34–41, 2010.

[8] L. Ljung. *System identification*. Wiley Online Library, 1999.

[9] L. Ljung and S. Gunnarsson. Adaptation and tracking in system identification - a survey. *Automatica*, 26(1):7–21, 1990.

[10] L. Ljung and T. Söderström. *Theory and Practice of Recursive Identification*. MIT Press, Cambridge, MA, 1983.

[11] G. Mzyk. *Parametric versus nonparametric approach to Wiener systems identification*, volume 404 of *Lecture Notes in Control and Information Sciences*, chapter 8, pages 111–125. 2010.

[12] G. Mzyk. Wiener-Hammerstein system identification with non-Gaussian input. In *IFAC International Workshop on Adaptation and Learning in Control and Signal Processing, Antalya, Turkey*, 2010.

[13] G. Mzyk. Fast model order selection of the nonlinearity in Hammerstein systems. In *Carpathian Control Conference (ICCC), 2012 13th International*, pages 507–510. IEEE, 2012.

[14] G. Mzyk. Nonparametric instrumental variables for Narmax system identification. *International Journal of Applied Mathematics and Computer Science (submitted to)*, 2012.

[15] M. Niedzwiecki. First-order tracking properties of weighted least squares estimators. *Automatic Control, IEEE Transactions on*, 33(1):94–96, 1988.

[16] M. Niedzwiecki. *Identification of Time-Varying Processes*. Wiley, 2000.

[17] J.P. Norton. Optimal smoothing in the identification of linear time-varying systems. *Electrical Engineers, Proceedings of the Institution of*, 122(6):663–668, 1975.

[18] L. Rutkowski. Nonparametric identification of quasi-stationary systems. *Systems & control letters*, 6(1):33–35, 1985.

[19] L. Rutkowski. Real-time identification of time-varying systems by non-parametric algorithms based on Parzen kernels. *International Journal of Systems Science*, 16(9):1123–1130, 1985.

[20] L. Rutkowski. Application of multiple Fourier series to identification of multivariable non-stationary systems. *International Journal of Systems Science*, 20(10):1993–2002, 1989.

[21] M.K. Tsatsanis and G.B. Giannakis. Time-varying system identification and model validation using wavelets. *Signal Processing, IEEE Transactions on*, 41(12):3512–3523, 1993.