Environmental Engineering and Management Journal *March 2020, Vol. 19, No. 3, 391-400*
http://www.eemj.icpm.tuiasi.ro/; http://www.eemj.eu

"Gheorghe Asachi" Technical University of Iasi, Romania

OUTLIER ROBUST IDENTIFICATION OF THE THERMAL POWER PLANT: COMBUSTION CONTROL AND VIBRATORY TRANSPORT

Vojislav Filipovic[1], Milan Matijevic[2], Petar Misljen[2*], Zeljko Despotovic[3]

[1]Faculty of Mechanical and Civil Engineering, University of Kragujevac, Kraljevo, Serbia
[2]Faculty of Engineering, University of Kragujevac, Kragujevac, Serbia
[3]"Mihailo Pupin" Institute, University of Belgrade, Belgrade, Serbia

Abstract

The goal of this paper is to contribute to environmental improvement. This goal has been achieved by the development of an algorithm that allows the optimization of fuel consumption and transport of combustion products. The paper considers the outlier – robust recursive algorithm. This algorithm is used for identification of multivariable ARMAX (Autoregressive Moving Average with Exogenous Variables) models. In this paper, we introduce pseudo-Huber loss function which is a continuous nonlinear approximation of Huber loss function and which has derivatives of all degrees. The structure of the recursive algorithm is: the relation for parameter estimation is based on Huber function and the relation for matrix gain is based on pseudo-Huber function. The novelty of this work is a new robust algorithm for recursive identification of MIMO ARMAX models. The main contributions of the paper are: (i) the new form of the extended least squares algorithm based on mixed Huber and pseudo-Huber functions; (ii) approximation of the second derivative of pseudo-Huber loss function and exact determination of matrix gain of the algorithm using the Laplace function. A special case of the algorithm, in this paper, is standard linear extended least squares algorithm. The results can be extended to an area of adaptive control, prediction, and filtering. They are suitable for practical processes such as thermal processes, vibration transport of bulk materials and other processes with multiple input variables. Practical behaviour of robust recursive procedure is illustrated by simulations.

Keywords: Huber function, recursive identification, thermal power plant, vibratory conveyor

Received: April, 2019; Revised final: July, 2019; Accepted: October, 2019; Published in final edited form: March, 2020

1. Introduction

The concept of eco-efficiency, in the case of thermal power plants is achieved through lower costs and efficient use of energy, water and fuel (Grigore et al., 2016). It was estimated that coal-fired steam power plants, with coal consumption of 40.000 tons/day, can transmit 150 tons/hour of SO_2 into the environment (Bangviwat and Sittikruear, 2018).

The presence of dust and ash fine particles from coal-fired power plants in the working environment causes cytotoxic effects in employees (Raducanu et al., 2010). Higher efficiency of thermal power plants, i.e. the reduction of fuel consumption and pollution of the environment is achieved by improving control structures in thermal power plants (Hu et al., 2018). The emission of pollutant particles into the working space decreases by improving the transport of combustion products.

The goal of this paper is to contribute to the improvement of the environment. This goal has been achieved by the development of an algorithm that allows the optimization of fuel consumption and transport of combustion products.

* Author to whom all correspondence should be addressed: e-mail: petarmisljen@gmail.com; Phone: +381603712452; Fax: +381112508474

2. Control of power plants

Thermal power plants and vibratory conveyors belong to the class of practical systems which can be represented as the MIMO (multiple-input-multiple-output) systems (Skogestad and Postlethwaite, 2005). Control of the traditional thermal power plant can be divided into two parts: combustion control and boiler-turbine control (Fig. 1). There are many variables, in traditional power plants, which must be an object of observation and control. The steam flow is a variable of particular importance. For the power plants optimal functionality, it is necessary to produce the required quantity and quality of the steam. The measure of the steam quantity is its flow rate. The flow rate directly determines the power output. Another important characteristic for the steam is its quality. The steam quality is measured by its temperature and pressure. Thus, there are two control objective of a traditional thermal power plant. The first is to maintain the steam quality, and the second is to meet the steam quantity. The steam quality is required by efficiency and safety, and the steam quality is required by the power output.

The master control signal in the boiler-turbine control is the steam demand. It is possible to configure the boiler control, depending on the power plant operation concept. There are several concepts of the master control signal behaving: the turbine control mode, the pressure control mode and the boiler control mode. There are numerous control subsystems in the power plants. The control subjects of these subsystems are fuel delivery, drum level and steam pressure (Zheng et al., 2011).

Toropov et al. (2019) studied the boiler as a separate system consisting of several mutually independent subsystems. It is one of many control examples where the robust recursive algorithm for identification of multivariable ARMAX can be applied. A block diagram of a vibrating conveyor with an electromagnetic actuator is shown in Fig. 2.

Vibratory conveyors with an electromagnetic actuator are widely used in the process industry. The conveyors are used there for the manipulation of bulk materials (Misljen et al., 2016). In thermal power plants, vibratory conveyors are used for transport of the solid fuel combustion products. The task of the vibrating conveyor is to carry out the transport of bulk material from the hopper to the load cell (Fig. 2). The heart of the vibrating conveyor shown in Fig. 2 is an electromagnetic vibration actuator (EVA).

EVA performs the transformation of current impulses from the energy converter into the impulses of the vibratory trough exciting force (f). Due to the vibrations of the trough, the bulk material flows through the vibratory trough. The energy converter is based on the h-bridge configuration with Insulated Gate Bipolar Transistor (IGBT) switches (Despotović et al., 2012). Input variables of vibratory conveyors are the parameter A_P, which represents the measure of energy to be transmitted to the actuator, and the frequency of the vibratory trough oscillations (f_r).The central processing unit (CPU) based on the set values of the parameter and, based on the reading of the value from the EVA current sensor and from the vibration sensor, controls the operation of the energy converter.

In addition to reading the values from these sensors, the CPU also reads the values from the load cell. Based on the data from the load cell, the CPU counts the flow rate of the bulk material at the output from the vibratory trough (Q). The maximum material flow rate Q is equal to the flow of material from the hopper (Q_0). The Human Machine Interface (HMI) enables the setting and visualization of input variables (A_P and f_r) and visualization of the output variables. Depending on the configuration, the output variables can be the amplitude of the oscillation of the trough (p), the flow Q, the amount of material in the trough or the amount of the transported material.

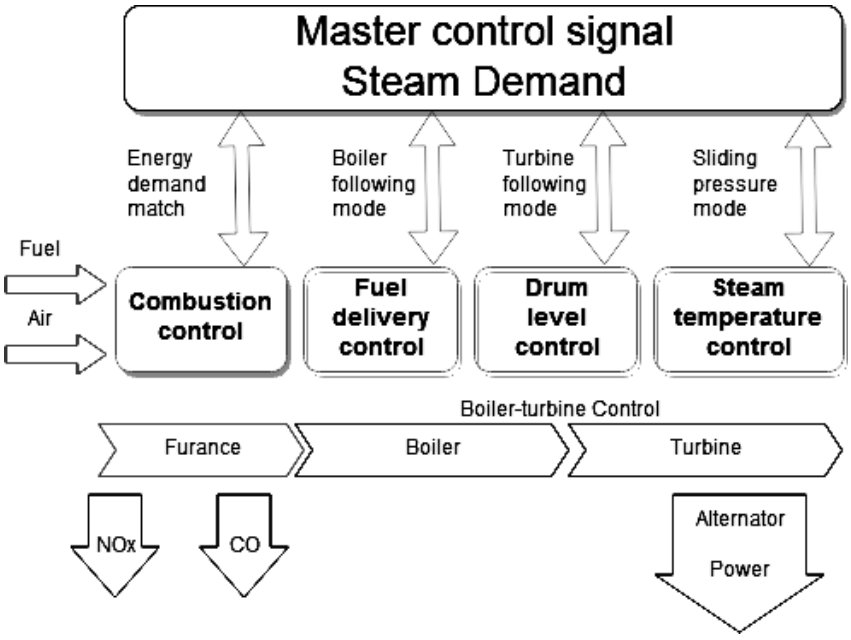

Fig. 1. Control of the traditional thermal power plant

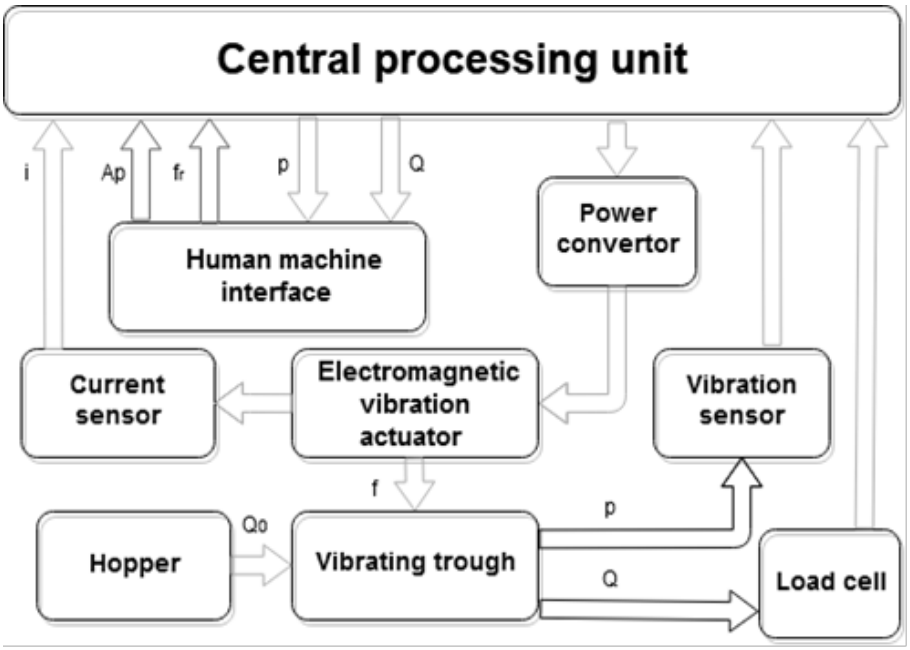

Fig. 2. Control of the vibratory conveyor with an electromagnetic actuator

The goal of system identification is to find a mathematical equation that gives an approximation to the actual behaviour of a real system (Mujahed et al., 2017). Designing of regulators for the above processes requires their mathematical models. A great deal of attention is dedicated to the identification of MIMO stochastic systems. It is usually assumed that distribution of the stochastic disturbance probability is exactly known (most frequently it is Gaussian distribution). Practical research showed that this assumption is not justified (Barnett and Lewis, 1994) since in observation populations there are rare large observations (outliers).

The consequence of this fact is that the distribution of the stochastic disturbance probability is non-Gaussian. That is why a significant effort has been invested to design recursive algorithms of identification showing minor sensitivity to the presence of outliers. In this sense, work published by Huber and Ronchetti (2009), which served as a basis for the constitution of robust statistics, represent a fundamental contribution. Thus, it is possible to replace restrictive assumption on exact knowledge of the disturbance distribution with the relaxed assumption on knowledge of the class of distribution to which the respective disturbance belongs. The last assumption is the basis for the robust, in the statistical sense, the theory of dynamic systems identification.

The Huber's function (first derivative of the Huber's loss function) is not differentiable in point ($+k_\varepsilon$) and in point ($-k_\varepsilon$), where k_ε is a Huber's parameter. To overcome the problem, it is introduced pseudo - Huber's loss function. This function has behaviour like Huber's function (Haltey, 2004). In this paper, pseudo – Huber's function is used to obtain a relation for matrix gain in a recursive algorithm. In the relation for parameter estimate, the Huber's function is retained. To simplify the realization of the algorithm, the approximation of the second derivative of pseudo - Huber's loss function is performed. Then, matrix gain is exactly determined by using the Laplace function. Robust procedures were compared to linear algorithms. In the case of outliers' presence, the simulation results indicate the superiority of robust procedures.

In order to increase the accuracy of soft sensors, new methods are introduced in engineering practice. Xiaofeng et al. (2018a, 2019a) presented a hybrid variable-wise weighted stacked autoencoder. This method can ensure that the learned features contain more information for quality prediction and overcome the limitations of traditional deep learning algorithms. ARMAX is a basic linear model to deal with dynamic data relationships.

They are other techniques that can deal with the nonlinear dynamics in data series, like linear dynamic systems and long short-term memory network. Xiaofeng et al. (2018b) proposed two kinds of weights for local linearization of the nonlinear state evolution and state emission relationships. Xiaofeng et al. (2019b) proposed a supervised long short-term memory network to model the dynamic and nonlinear behaviours of process sequential data.

3. Case studies

3.1. Process identification

The problem of regulating combustion in boiler plants is very important for the environment. The description of the combustion process based on fundamental physical laws is very complex. That's why the black box methodology has been used. In this case, the model is obtained using the theory of identification and requires only input and output measurements. The combustion process has a

multivariable character and its block diagram is shown in Fig. 3.

In a boiler, fuel and air are added in a certain relationship and there they burn. The quality of combustion is measured by the oxygen content of the exhaust gas (flue gas). Fig. 3 presents the model of the process that is used for the control structure design. Below, a mathematical model and its recursive identification procedure will be proposed.

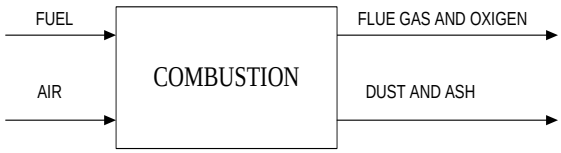

Fig. 3. Combustion in power plant

3.2. Multivariable ARMAX model

It is supposed that the considered systems described by the linear multivariable ARMAX model with p-dimensional output and r-dimensional input (Eq. 1).

$$A(q^{-1})y_k = B(q^{-1})u_k + C(q^{-1})w_k \tag{1}$$

In Eq. (1), $A(q^{-1})$, $B(q^{-1})$ and $C(q^{-1})$ are matrix polynomials, y_k is response vector, u_k is excitation vector and w_k is vector of innovations, which is an unobservable white noise process with a covariance matrix σ (Koulocheris et al., 2005). In these polynomials q^{-1} denotes the shift-back operator, so $q^{-1}y_k = y_{k-1}$. Orders of polynomials $A(q^{-1})$, $B(q^{-1})$ and $C(q^{-1})$ are *n*, *m*, and *l*, respectively.

In Eqs. (2-4), $A_i(i=1,2,\ldots,n)$ are *pxp* matrices, $B_i(i=1,2,\ldots,m)$ are *pxr* matrices, $C_i(i=1,2,\ldots,l)$ are *pxp* matrices, and *I* is identity matrix. The stochastic disturbance $\{w_k\}$ is a martingale-difference in relation to the non-decreasing family of σ-algebras $\{F_k\}$. Unknown matrix coefficients are presented in Eq. (5).

$$A(q^{-1}) = I + A_1q^{-1} + ... + A_nq^{-n} \tag{2}$$

$$B(q^{-1}) = B_1q^{-1} + ... + B_mq^{-m} \tag{3}$$

$$C(q^{-1}) = I + C_1q^{-1} + ... + C_lq^{-l} \tag{4}$$

$$\left(\boldsymbol{\theta}^M\right)^T = [A_1, A_2, ..., A_nB_1, B_2, ..., B_mC_1, ..., C_l] \tag{5}$$

Now the model (Eq. 1) can be written in the form

$$y_k = \left(\boldsymbol{\theta}^M\right)^T X_k^0 + w_k,$$

where:

$$\left(X_k^0\right)^T = [-y_{k-1}^T \ldots - y_{k-n}^T u_{k-1}^Y \ldots u_{k-m}^T w_{k-1}^T \ldots w_{k-l}^T]$$

$$, \left(X_k^0\right)^T \in R^{1\times(np+mr+lp)}.$$

Let introduce the matrix $\boldsymbol{\varphi}_k^0$ (Eq. 6).

$$\boldsymbol{\varphi}_k^0 = \begin{bmatrix} \left(X_k^0\right)^T & & 0 \\ & \ddots & \\ 0 & & \left(X_k^0\right)^T \end{bmatrix} = I \otimes \left(X_k^0\right)^T \tag{6}$$

The standard procedure in identification is to replace w_k with the estimate (e_i is the prediction error). From that fact it follows Eqs. (7-9).

$$X_k^T = \left[- y_{k-1}^T \ldots - y_{k-n}^T u_{k-1}^T \ldots u_{k-m}^T e_{k-1}^T \ldots e_{k-l}^T\right] \tag{7}$$

$$\boldsymbol{\varphi}_k = \begin{bmatrix} X_k^T & & 0 \\ & \ddots & \\ 0 & & X_k^T \end{bmatrix} = I \otimes X_k^T \tag{8}$$

$$e_k = y_k - \boldsymbol{\varphi}_k \hat{\boldsymbol{\theta}}_{k-1} \tag{9}$$

Eq. (9) is a key ingredient for the formulation of identification criterion.

Remark 1. Model (1) is a general MIMO ARMAX model. The characteristic of the model (1) for a stochastic disturbance, due to the presence of outliers, has a non - Gaussian distribution. In practical situations, the presence of outliers is inevitable (Pearson, 2011). In this case, the standard algorithms, designed in the case of Gaussian disturbance, may have an unacceptably low performance. For this case, a robust statistics methodology is applied (Huber and Rouchetti, 2009). From this comes the class of robust algorithms that are slightly sensitive to the change in the distribution of the probability of the disturbance.

Remark 2. When synthesizing recursive algorithms, the most common assumption about the nature of a stochastic disturbance is that the disturbance has a Gaussian distribution. The problem with this selection is described in Remark 1. It is also, more rarely, assumed that the disturbance is uniformly limited in size without entering the structure of the disorder (Bai et al., 1996).

Remark 3. In this paper, it is assumed that the model of the combustion process is linear. In the first step, nonlinear processes can be approximated by a linear time-invariant model (Euqvist and Ljung, 2005, Schoukens and Tiels, 2017). In this way, we can get an insight into the behavior of the system. In the next step, the process, which is nonlinear, is described by the Hammerstein model (nonlinear ARMAX model) (Filipovic, 2017) and this is a problem for further research.

3.3. Robust recursive algorithm

In the literature, it is usually assumed that the probability density of stochastic disturbance is normal. Barnett and Lewis (1994) have shown that this assumption is not justified, since in observation populations there are outliers. In that case, it is

assumed that the distribution class to which the disturbance belongs is known. The distribution (gross error model) class presented by Eq. (10) is practically important.

$$P_{1\varepsilon}=\left\{P: P=(1-\varepsilon)N_D+\varepsilon G,\ \ G\ \ is\ \ symmetric\right\} \tag{10}$$

In Eq. (10), $\varepsilon\in[0,1)$ is the contamination degree, G is symmetric distribution and N_D is a normal distribution, where G may be arbitrary distribution (Eq. 11).

$$P_{2\varepsilon}=\left\{P: P=(1-\varepsilon)N_D+\varepsilon G\right\} \tag{11}$$

Very popular contamination model in engineering literature, where σ_{1i}^2 is much less than $\sigma_{2i}^2(i=1,2,...,p)$, is a Tukey`s model (Eq. 12).

$$p_{T\varepsilon}=\left\{p: p=(1-\varepsilon)N\left(0,\sigma_{1i}^2\right)+\varepsilon N\left(0,\sigma_{2i}^2\right)\right\} \tag{12}$$

The model presented by Eq. (12) is given in the form of probability densities where $N\left(0,\sigma^2\right)$ denotes a zero-mean Gaussian probability density with a variance σ^2.

General contaminated model is presented by Eq. (13).

$$p_{\varepsilon}=\left\{p: p(x)=(1-\varepsilon)N\left(0,\sigma^2\right)+\varepsilon g(x)\right\} \tag{13}$$

In Eq. (13), $g(\cdot)$ is an arbitrary probability density. For Huber's theory is important restriction presented in Eq. (14).

$$p_{\varepsilon}=\left\{p: p(x)\geq(1-\varepsilon)N\left(0,\sigma^2\right),\ 0\leq\varepsilon<1\right\} \tag{14}$$

If Hubert's methodology is applied, the least favourable probability density on a class of approximately normal distributions is obtained (Eq. 15).

$$p_i^*\left(w_k^i\right)=\begin{cases}\dfrac{1-\varepsilon}{\sqrt{2\pi}\sigma_i}\exp\left\{-\dfrac{\left(w_k^i\right)^2}{2\sigma_i^2}\right\}, & \left|w_k^i\right|<k_{\varepsilon}^i\\ \dfrac{1-\varepsilon}{\sqrt{2\pi}\sigma_i}\exp\left\{-\dfrac{k_{\varepsilon}^2}{\sigma_i^2}\left(\left|w_k^i\right|-\dfrac{k_{\varepsilon}^i}{2}\right)\right\}, & \left|w_k^i\right|>k_{\varepsilon}^i\end{cases} \tag{15}$$

In Eq. (15), the relationship between the contamination degree ε and the parameter k_{ε}^i of Huber's function is given, for the scalar location parameter, by Eq. (16) (Huber and Ronchetti, 2009).

$$\frac{2\Phi_N\left(k_{\varepsilon}^i\right)}{k_{\varepsilon}^i}-2\Phi_N\left(-k_{\varepsilon}^i\right)=\frac{\varepsilon}{1-\varepsilon},\ \ \Phi_N(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{y^2}{2}}dy \tag{16}$$

Eq. (16) depends on variables ε and k_{ε}^i. In practice, the contamination degree ε is unknown. Intensive research has shown that the value of ε moves up to 0.2. Because of this, based on simulations, it is necessary to find k_{ε}^i for which the robust algorithms have a minimum estimation error of system parameters over the whole range of values of ε. Earlier intense simulations show that good performance of robust algorithms is provided for $k_{\varepsilon}^i\in[2,4]$. The best performance is accomplished for $k_{\varepsilon}^1=....=k_{\varepsilon}^p=3$ (Filipovic, 2005).

The components of the vector w_k are independent. Owing that, the least favourable probability density of the vector w_k can be presented as in Eq. (17).

$$p^*(w_k)=\prod_{i=1}^{p}p_i^*\left(w_k^i\right) \tag{17}$$

Using Eq. (9) and Eq. (17), it can get $\Phi(e)=-\log p^*(w)\big|_{w=\varepsilon}$ or explicitly like in Eq. (18).

$$\Phi(e)=\begin{cases}\dfrac{e^2}{2\sigma^2}+\ln\dfrac{\sqrt{2\pi}\sigma}{1-\varepsilon}, & |e|\leq k_{\varepsilon}\\ \dfrac{k_{\varepsilon}}{\sigma^2}\left(|e|-\dfrac{k_{\varepsilon}}{2}\right)+\ln\dfrac{\sqrt{2\pi}\sigma}{1-\varepsilon}, & |e|>k_e\end{cases} \tag{18}$$

The function $\Phi(\cdot)$ is Huber's loss function and derivative $\psi(\cdot)=\Phi'(\cdot)$ is Huber's function (Eq. 19).

$$\psi(e)=\begin{cases}e, & |e|\leq k_e\\ k_{\varepsilon}, & |e|>k_e\end{cases} \tag{19}$$

The Huber's function $\phi(\cdot)$ is not differentiable in point $(+k_{\varepsilon})$ and in point $(-k_{\varepsilon})$. Owing to the fact that Huber's loss function $\phi(\cdot)$ is only first-order differentiable, it follows that it is not applicable to second-order methods (for example Newton-Raphson algorithm or Levenburg-Marquardt algorithm). The pseudo-Huber's function is a smooth version of the Huber's loss function. In that case, the pseudo-Huber's function has derivatives of all degrees.

We will consider next pseudo-Huber's loss function (Eq. 20):

$$\Phi_p(e)=k_{\varepsilon}\left(\sqrt{(k_{\varepsilon})^2+e^2}-k_{\varepsilon}\right)+\ln\frac{\sqrt{2\pi}\sigma}{1-\varepsilon} \tag{20}$$

The derivatives of loss function $\Phi_p(\cdot)$ (first and second) are presented in Eqs. (21-22), respectively.

$$\psi_p(e)=\Phi_p'(e)=\frac{k_{\varepsilon}e}{\sqrt{(k_{\varepsilon})^2+e^2}} \tag{21}$$

$$\psi_p'(e)=\Phi_p''(e)=\frac{(k_{\varepsilon})^3}{\left((k_{\varepsilon})^2+e^2\right)^{3/2}} \tag{22}$$

The derivatives $\psi_p(e)$ (that is pseudo-Huber's function) and $\psi'_p(e)$ are bounded and Lipschitz continuous. To simplify the numerical aspect for recursive algorithms we introduce an approximation $E\{\boldsymbol{\psi}_a(e)\} \cong E\{\boldsymbol{\psi}'_p(e)\}$ where $\boldsymbol{\psi}_a(e)$ is defined in Eq. (23).

$$\boldsymbol{\psi}_a(e) = \begin{cases} 1 & , \ |e| \le k_{\boldsymbol{\varepsilon}} \\ 0 & , \ |e| > k_{\boldsymbol{\varepsilon}} \end{cases} \tag{23}$$

This approximation has a small influence on the behaviour of gain of the recursive algorithm. Tsypkin (1984) uses approximation $\boldsymbol{\psi}_a(\cdot)$ for the first derivative of Huber's loss function.

In this paper, a hybrid approach is proposed. For calculation of $\nabla_{\boldsymbol{\theta}} J_k(\boldsymbol{\theta}_{k-1})$ we use the Huber's loss function, where $\boldsymbol{J}_k$ refers to Jacobian matrix. The calculation of $\nabla^2_{\boldsymbol{\theta}} J_k(\boldsymbol{\theta}_{k-1})$ is based on the pseudo-Huber's loss function and approximation in the form of function $\boldsymbol{\psi}_a(\cdot)$. Based on this approach, a key part of Huber's theory is preserved (Eq. 39).

The identification criterion can be defined as in Eq. (24).

$$J_1(\boldsymbol{\theta}) = E\{\Phi(e_k)\} \tag{24}$$

In Eq. (24) $E\{\cdot\}$ represents the mathematical expectation operator. This criterion generates robust (optimal on the class) algorithms.

The Newton-Raphson algorithm (Eq. 25) can be applied for the recursive minimization of the criterion, where Eq.(26) is empirical functional.

$$\boldsymbol{\theta}_k = \boldsymbol{\theta}_{k-1} - \left[k\nabla^2_{\boldsymbol{\theta}} J_k(\boldsymbol{\theta}_{k-1})\right]^{-1}\left[k\nabla_{\boldsymbol{\theta}} J_k(\boldsymbol{\theta}_{k-1})\right] \tag{25}$$

$$J_k(\boldsymbol{\theta}) = \frac{1}{k}\sum_{i=1}^{k} \Phi(e_i) \tag{26}$$

Similarly, as defined by Filipovic (2015) it is possible to get Eqs. (27-31).

$$k\nabla_{\boldsymbol{\theta}} J_k(\boldsymbol{\theta}_{k-1}) = -\boldsymbol{\varphi}_k^T \boldsymbol{\psi}(e_k), \ \ \boldsymbol{\psi}(\cdot) = \Phi'(\cdot) \tag{27}$$

$$k\nabla^2_{\boldsymbol{\theta}} J_k(\boldsymbol{\theta}_{k-1}) = \sum_{j=1}^{k} \boldsymbol{\varphi}_j^T M \boldsymbol{\varphi}_j \tag{28}$$

$$M = E\{\boldsymbol{\psi}_a(w_k)\} \tag{29}$$

$$\boldsymbol{\psi}(e_k) = \left[\psi_1(e_k^1), \psi_2(e_k^2), ..., \psi_p(e_k^p)\right]^T \tag{30}$$

$$\boldsymbol{\psi}_i(e_k^i) = \max\left\{-k_{\boldsymbol{\varepsilon}}^i, \min\left(k_{\boldsymbol{\varepsilon}}^i, e_k^i\right)\right\} \quad , \ i = 1,2,..., p \tag{31}$$

The graphical representations of the Huber's function *i-th* component ($\psi_i(e_k^i)$) (nonlinear transformation of prediction error) and the approximation of the first derivative of pseudo-Huber's function ($\boldsymbol{\psi}_{ai}(e_k^i)$) are presented in Fig. 4.

Now, the matrix *M* can be determined based on Eqs. (32-33). We have:

$$\boldsymbol{\psi}_a(w_k) = \begin{bmatrix} \boldsymbol{\psi}_{a1}(w_k^1) & & 0 \\ & \ddots & \\ 0 & & \boldsymbol{\psi}_{ap}(w_k^p) \end{bmatrix} \tag{32}$$

$$m_i = E\{\boldsymbol{\psi}_{ai}(w_k^i)\} \tag{33}$$

$$\boldsymbol{\psi}_{ai}(w_k^i) = \begin{cases} 1 & , \ |w_k^i \le k_{\boldsymbol{\varepsilon}}^i| \\ 0 & , \ otherwise \end{cases} \tag{34}$$

From Eq. (15) and Eq. (34) it follows Eqs. (35-36).

$$m_i = E\{\boldsymbol{\psi}_{ai}(w_k^i)\} = 2(1-\boldsymbol{\varepsilon})\boldsymbol{\phi}_L^i\left(\frac{k_{\boldsymbol{\varepsilon}}^i}{\boldsymbol{\sigma}_{1i}}\right) \tag{35}$$

$$\boldsymbol{\phi}_L(x) = \frac{1}{\sqrt{2\pi}}\int_0^x e^{\frac{y^2}{2}} dy \tag{36}$$

Eq. (36) is a Laplace function. Dong et al (2017) have presented Table of values for which Eq. (36) exists.

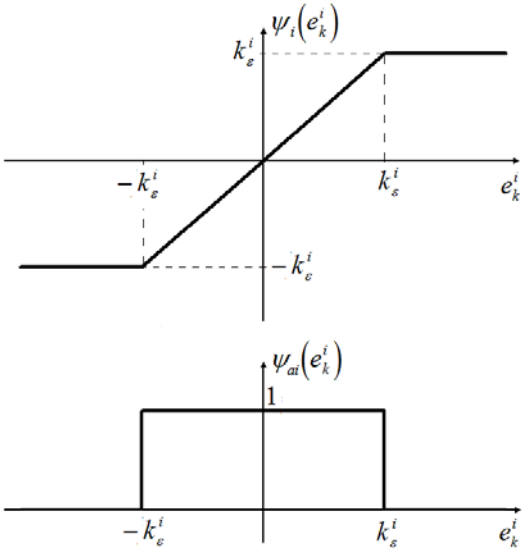

Fig. 4. Huber's function and approximation of the first derivative of the pseudo-Huber's function

Based on the above, matrix *M* can be presented by Eq. (37).

$$M = \begin{bmatrix} m_1 & & 0 \\ & \ddots & \\ 0 & & m_p \end{bmatrix} \tag{37}$$

Let introduce Eq. (38):

$$P_k = \left[k\nabla^2_{\boldsymbol{\theta}} J_k(\boldsymbol{\theta}_{k-1})\right]^{-1} \tag{38}$$

From Eq. (25), Eq. (27) and Eq. (38) it follows Eqs. (39-41).

$$\theta_k = \theta_{k-1} + P_k\ {}_k^T\psi(e_k) \quad , \theta_0 = 0 \tag{39}$$

$$P_k = P_{k-1} - P_{k-1}\ {}_k^T\left[\ {}_kP_{k-1}\ {}_k^T + M^{-1}\right]\ {}_kP_k \quad , P_0 = \gamma I \tag{40}$$

$$e_k = y_k - \boldsymbol{\varphi}_k\boldsymbol{\theta}_{k-1} \tag{41}$$

The relation for matrix gain is presented by Eq. (40). In Eq. (40), γ is much greater than 1, the dimension of the matrix P_k is *(np+mr+lp)x(np+mr+lp)* and matrix *I* is the identity matrix. The robustness notion is used in statistics, in general sense, as insensitivity against small deviation from the assumptions. In this paper, it is studied the distribution robustness. The algorithm presented by Eqs. (39-41), together with the algorithm written by Filipovic (2015), in where ARX MIMO systems are considered, as far as authors know is novel.

Remark 4. Implementation of the recursive evaluation algorithm (Eqs. 39-41) is based on the concept of the most unfavourable density of probability (Eq. 15) (min-max principle). The correct behaviour of the algorithm for the distribution case (Eq. 15) ensures a correct behaviour of the algorithm for any distribution of the distribution class (Eq. 10).

Remark 5. The important role in the algorithm (Eqs. 39-41) has the function *ψ()*. This function eliminates the effect of observing the oversized high level (outliers).

4. Results and discussion

The robust recursive algorithm for identification of multivariable ARMAX systems has been considered in previous chapters. The effectiveness of that algorithm has been considered on the simulation level. We have considered the system described by Eqs. (1-4), where:

$$A_1 = \begin{bmatrix} 0 & 0.5 \\ 1 & 0 \end{bmatrix},\ A_2 = \begin{bmatrix} 1.2 & 0 \\ 0 & 0.5 \end{bmatrix},\ B_1 = \begin{bmatrix} 0 & 0.5 \\ 1 & 0.7 \end{bmatrix},$$

$$B_2 = \begin{bmatrix} 2 & 1 \\ 3 & 1.2 \end{bmatrix},\ C_1 = \begin{bmatrix} 1.2 & 0 \\ 0 & 0.6 \end{bmatrix},\ C_2 = \begin{bmatrix} 0.4 & 0 \\ 0 & 0 \end{bmatrix}.$$

The components of input signals

$$\boldsymbol{u}_k = \left[\boldsymbol{u}_k^1 \boldsymbol{u}_k^2\right]^T$$

are shown in Fig. 5. Input signals have the next structure:

- Input $\boldsymbol{u}_k^1$ is a Gaussian sequence with zero mean and variance 1
- Input $\boldsymbol{u}_k^2$ is set to be an independent sequence of uniform distribution with zero mean and variance 1.

We have considered two different situations.

A) The stochastic disturbance has Gaussian distribution

It is supposed that both components $\left(w_k^1, w_k^2\right)$ have same Gaussian distribution *N* (0,1) (mean is equal to zero and variance is equal to 1). The disturbance is presented in Fig. 6a. For that type of disturbance, the outputs of systems are presented in Fig. 7.

B) The stochastic disturbance has Non-Gaussian distribution

We supposed that stochastic disturbance w_k^i has a Tukey distribution (Non-Gaussian distribution)

$$(1-\varepsilon)N\left(0,\sigma_{1i}^2\right) + \varepsilon N\left(0,\sigma_{2i}^2\right),\ i = 1,2,$$

where $N\left(m,\sigma^2\right)$ is Gaussian distribution with mean m and variance σ^2.

In simulations, it will be taken $\sigma_{1i}^2 = 1$, $\sigma_{2i}^2 = 100$, i = 1, 2. The stochastic disturbance is presented in Fig. 6b (for contamination degree $\varepsilon = 0.15$). In that case, the outputs of the system are presented in Fig. 8.

The comparison of algorithms is carried out for contamination degree $\varepsilon = 0.15$. In all cases, it is assumed that the parameter of Huber's function is k_ε^i, $i = 1,2,...,p$. For such value of k_ε^i we have, according to Haltey (2004), $\Phi_L^i(3) = 0.49865$.

The error of estimation is defined by Eq. (42).

$$\delta = \ln\frac{\|\boldsymbol{\theta}_k - \boldsymbol{\theta}\|}{\|\boldsymbol{\theta}\|} \tag{42}$$

In Eq. (42) the vector θ is a vector of true parameters and θ_k is a vector of estimated parameters. Comparison between robust and linear algorithms is performed for initial value for algorithms ($\theta_0 = 0, P_0 = 10^4 I$).

Fig. 9 presents the behaviour of extended least squares (ELS) and robust extended least squares (RELS). It is shown that the behaviour of RELS is superior in comparison to ELS.

Acknowledgements
The research in this paper is supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia.

5. Conclusions

This paper proposes a recursive algorithm for estimation of parameters of multivariable ARMAX model based on mixed Huber and pseudo-Huber functions. The impact of outliers is reduced by the introduction of non-linear prediction error transformation depending on a priori information on the specific class of distribution to which the relevant disturbance belongs.

Matrix gain depends on the second derivative of the pseudo-Huber loss function. Model transformation results in unknown parameters being reduced to vector form.

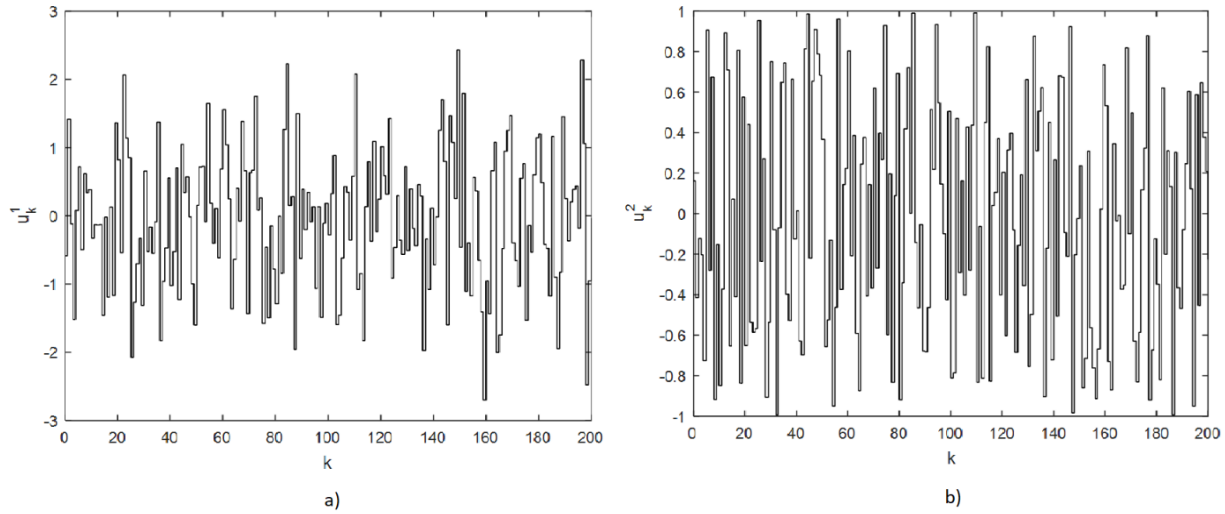

Fig. 5. Input signals: a) u_k^1 and b) u_k^2

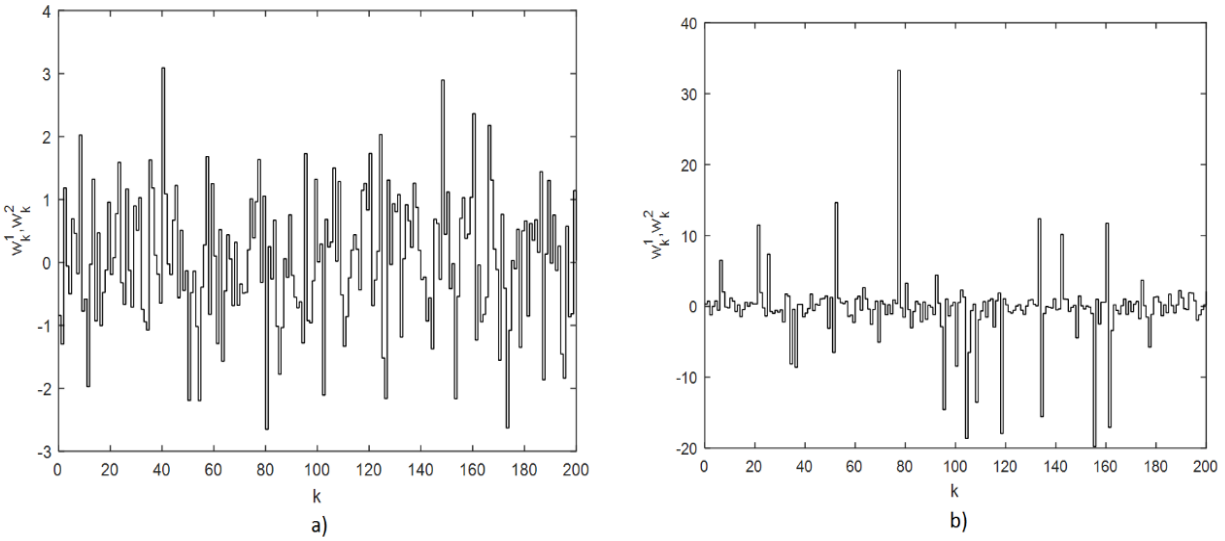

Fig. 6. Disturbance: a) Disturbance with Gaussian distribution and b) Non-Gaussian disturbance

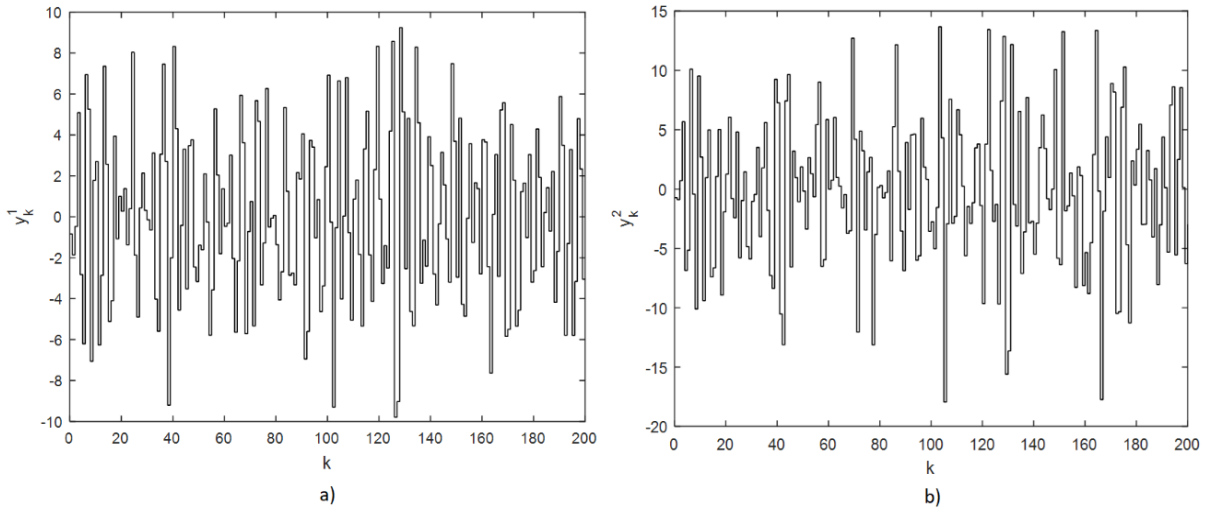

Fig. 7. System outputs for Gaussian disturbance: a) y_k^1 and b) y_k^2

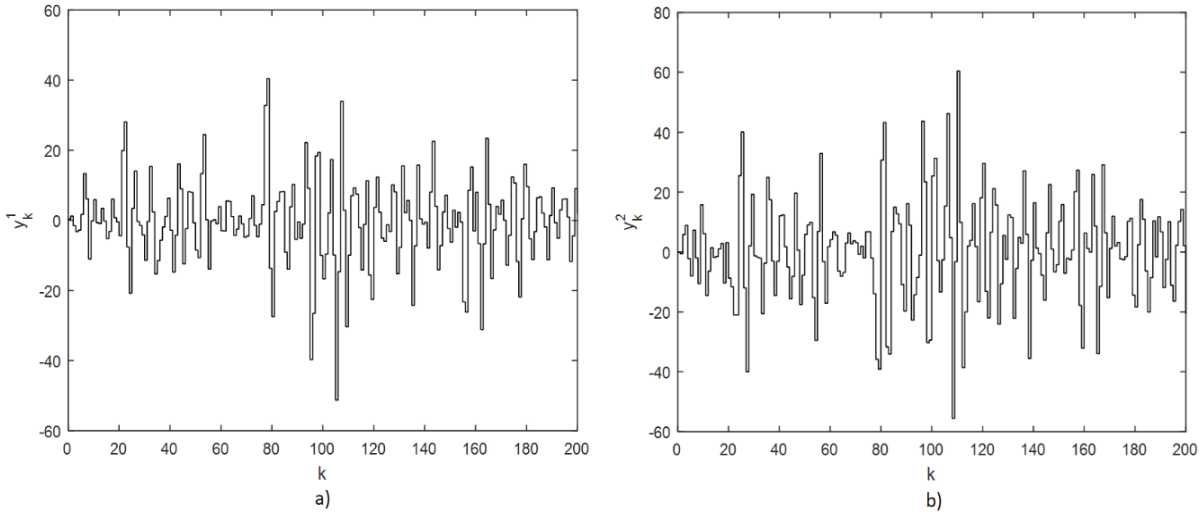

Fig. 8. System outputs for non-Gaussian disturbance ($k_\varepsilon^i = 3$): a) y_k^1 and b) y_k^2

Robust procedures were compared to linear algorithms. In the case of outliers' presence, the simulation results indicate the superiority of robust procedures. Further research is related to the model of the process in the Hammerstein form and the unknown process parameters in the matrix form.

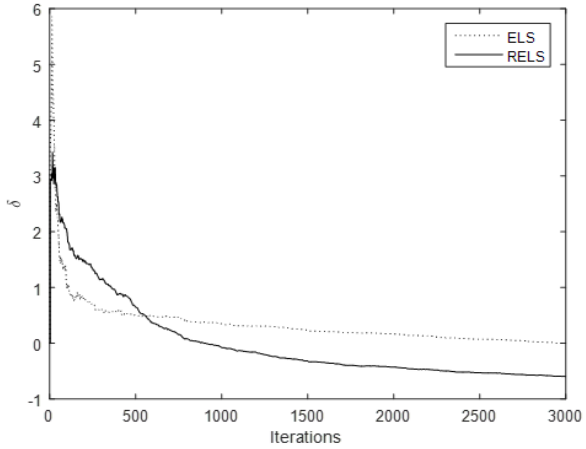

Fig. 9. Comparison of extended least squares (ELS) and robust extended least squares (RELS) for $\varepsilon = 0.15$

References

Bai E.W., Nagpal K.M., Tempo R., (1996), Bounded-error parameter estimation: Noise models and recursive algorithms, *Automatica*, **32**, 985-999.

Barnett V., Lewis T., (1994), *Outliers in Statistical Data*, 3rd Edition, Wiley, New York.

Bangviwat A., Sittikruear S., (2018), Application of analytical hierarchy process (AHP) in evaluating measures for mitigation of emission impact on communities near lignite-fired power plant at mae moh, Thailand, *Environmental Engineering and Management Journal*, **17**, 1457-1466.

Despotović Ž., Ribić A., Sinik V., (2012), Power current control of a resonant vibratory conveyor having electromagnetic drive, *Journal of Power Electronics*, **12**, 678-689.

Dong S., Liu T., Wang Q., (2017), Identification of dual-rate sampled-systems with time delay subject to load disturbance, *IET Control Theory and Applications*, **11**, 1404-1413.

Filipovic V., (2005), Stochastic multivariable self-tuning tracker for non-Gaussian systems, *International Journal of Applied Mathematics and Computer Science*, **15**, 351-357.

Filipovic V., (2015), Recursive identification of multivariable ARX models in the presence of a priori information: Robustness and regularization, *Signal Processing*, **116**, 68-77.

Filipovic V., (2017), Outlier robust stochastic approximation algorithm for identification of MIMO Hammerstein models, *Nonlinear Dynamics*, **90**, 1427-1441.

Grigore R., Capat C., Hazi A., Hazi G., (2016), Eco-efficiency indicators in the evaluation of environmental performance of thermal power plants, *Environmental Engineering and Management Journa*l, **15**, 143-149.

Haltey R., (2004), *Multiple View Geometry in Computer Vision*, Cambridge University Press, Cambridge.

Hsu C.S., Perng Y.S., Chen H.T., Chen Y.C., (2018), Planning and decision-making models in ecological engineering applied to Taiwan mid-west coast, *Environmental Engineering and Management Journal*, **17**, 2031-2039.

Huber P., Ronchetti E., (2009), *Robust Statistics*, Wiley, New York.

Koulocheris D., Spentzas K., Dertimanis V., (2005), *Vehicle System Identification Using Mimo-Armax Models*, 10th EAEC European Automotive Congress, Belgrade, vol. 1, 427-438.

Misljen P., Despotovic Z., Matijevic M., (2016), Modeling and control of bulk material flow on the electromagnetic vibratory feeder, *Automatika – Journal for Control, Measurement, Electronics, Computing and Communications*, **57**, 936-947.

Mujahed Al-D., Kottakkaran S.N., Praveen A., Alaa E., (2017), Modeling and identification of heat exchanger process using least squares support vector machines, *Thermal Science*, **21**, 2859-2869.

Pearson R.K., (2011), *Exploring Data in Engineering, the Science and Medicine*, Oxford University Press, Oxford.

Raducanu A., Suverge A., Darie G., Rau I., Grigoriu C., Viespe C., Vasiu T., Popescu I. D., Codorean E., (2010), Structural characterization and in vitro cytotoxic potential of coal dust in a Romanian power plant, *Environmental Engineering and Management Journal*, **9**, 1297-1304.

Schoukens M., Tiels K., (2017), Identification of block-oriented nonlinear systems starting from linear approximations: A survey, *Automatica*, **85**, 272-292.

Skogestad S., Postlethwaite I., (2005), *Multivariable Feedback Control. Analysis and Design*, 2nd Edition, Wiley, New York.

Toropov E., Osintsev K., Aliukov S., (2019), New theoretical and methodological approaches to the study of heat transfer in coal dust combustion, *Energies*, **12**, 1-14.

Tsypkin Y.Z., (1984), *Foundation of Information Identification Theory*, (in Russian), Nauka, Moscow, SSSR.

Xiaofeng Y., Biao H., Yalin W., Chunhua Y., Weihua G., (2018a), Deep learning-based feature representation and its application for soft sensor modeling with variable-wise weighted SAE, *IEEE Transactions on Industrial Informatics*, **14**, 3235-3243.

Xiaofeng Y., Yalin W., Chunhua Y., Zhiqiang G., Zhihuan S., Weihua G., (2018b), Weighted linear dynamic system for feature representation and soft sensor application in nonlinear dynamic industrial processes, *IEEE Transactions on Industrial Electronics*, **65**, 1508-1517.

Xiaofeng Y., Chen O., Yalin W., Chunhua Y., Weihua G., (2019a), Deep quality-related feature extraction for soft sensing modeling: A deep learning approach with hybrid VW-SAE, *Neurocomputing*, http://doi.org/10.1016/j.neucom.2018.11.107.

Xiaofeng Y., Lin L., Yalin W., (2019b), Nonlinear dynamic soft sensor modeling with supervised long short-term memory network, *IEEE Transactions on Industrial Informatics*, http://doi.org/10.1109/TII.2019.2902129.

Zheng Q., Gao Z., Tan W., (2011), *Disturbance Rejection in Thermal Power Plant*s, Proc. of the 30th Chinese Control Conf., Yantai, China, 6350-6355.