Int. J. Mech. Eng. Autom.
Volume 2, Number 4, 2015, pp. 159-165
Received: January 25, 2015; Published: April 25, 2015

International Journal of
Mechanical Engineering
and Automation

Model Order Selection Based on Robust Akaike's Criterion

Vladimir Stojanovic and Novak Nedic

Department of Automatic Control, Robotics and Fluid Technique, Faculty of Mechanical and Civil Engineering, University of Kragujevac, Kraljevo 36000, Serbia

Corresponding author: Vladimir Stojanovic (vladostojanovic@open.telekom.rs)

Abstract: This paper considers the model order selection due to process of identification of OE (output error) models with constant parameters in the presence of measurements with non-Gaussian noise distributions. In practical conditions, in measurements there are rare, inconsistent observations with the largest part of population of observations. Therefore, synthesis of robust algorithms is of primary interest. The presence of outliers can considerably degrade the performance of linearly recursive algorithms based on the assumptions that measurements have Gaussian distributions. In this paper, the robust parameter estimation algorithm is proposed which is based on Huber's theory of robust statistics. On the other side, ad hoc selection of model orders leads to overparametrization or parsimony problem. The natural frame to avoid these problems is AIC (Akaike's information criterion) for model order selection, which is obtained by minimization of the Kullback-Leibler information distance. The originally proposed Akaike's criterion cannot be applied since stochastic disturbance in the model belongs to the class of ε-contaminated distributions. By determining the least favourable probability density for a given class of probability distribution represents a base for design of the RAIC (robust version of Akaike's information criterion). The benefits of RAIC for robust parameter estimation procedure is illustrated through intensive simulations which demonstrate the superiority of the proposed robust procedure in relation to the linear algorithms (derived under the assumption that the stochastic disturbance has a Gaussian distribution).

Keywords: Model order selection, output error model, ε-contaminated distributions, robust Akaike's criterion.

Nomenclature

ε	Degree of contamination
$\Psi(\cdot)$	Vector influence function
$\varphi(k)$	Regression vector
$\theta(k)$	Vector of true parameters
$\hat{\theta}(k)$	Vector of parameter estimates
a_j, b_j	System parameters
$E(\cdot)$	Mathematical expectation
$J(\theta)$	Identification criterion
$\Phi(\cdot)$	Robust loss function
$p^*(.)$	The least favourable distribution of probability
$\mathcal{P}_\varepsilon$	Approximately normal distribution class
$p(e)$	Probability density function
σ^2	Variance
$P(k\|k)$	A posteriori covariance matrix
$P(k\|k-1)$	A priori covariance matrix
$e(k)$	Stochastic disturbance
$u(k)$	Input signal
$y(k)$	Output signal
$\varepsilon(k)$	Prediction error
$y_M(k)$	Output of an auxiliary model

1. Introduction

It is well known that obtaining models of physical systems based on the fundamental laws of physics is a difficult problem. The system identification is an alternative approach, which ensures obtaining the mathematical models based on input/output measurements [1, 2]. Most identification algorithms assume that the model structure is a priori known. As is well known, a fundamental difficulty in statistical analysis is the choice of an appropriate model and determining the order of a model. In recent years, the necessity of introducing the concept of model has been recognized and the problem is posed how to choose the "best approximating" model among a class of competing models with different numbers of parameters by a suitable model selection criterion given a data set. Also, there is presently a great deal of interest in simple criteria represented by parsimony of parameters for choosing one of a set of competing models to describe a given data set. The selection of a parsimonious model, in general, is a nontrivial problem without the aid of model selection criteria.

Several information criterions have been proposed for structure selection in linear dynamic input/output models. The model which minimizes the criterion is then chosen as the best model from the available set. Examples of the classical criterions are the FPE (final prediction error), AIC (Akaike's information criterion) and BIC (Bayesian information criterion). These techniques find a tradeoff between goodness of fit and model complexity. The performance of an order-selection criterion is optimal if the model of the selected order is the most accurate model in the considered set of estimated models.

Used way for deriving model selection criteria is based on the quantification of "how close are" the probability density of the generating model and the probability density of the fitted approximating model. Several coefficients or "measures" have been introduced in the literature for this quantification. The Kullback-Leibler information distance is the most frequently used information theoretic coefficient for measuring divergence or separation between two probability densities [3]. The AIC is a commonly used tool for choosing between alternative models [4].

Here, those results are extended on the case when the measurement noise is a non-Gaussian. Justification of this approach was confirmed in practice [5, 6]. Namely, in measurements there are rare, inconsistent observations with the largest part of population of observations (outliers). The presence of outliers can considerably degrade the performance of linearly recursive algorithms based on the assumptions that measurements have a Gaussian distribution.

The synthesis of robust algorithms is of primary interest. The synthesis is based on Huber's theory of robust statistics [6]. As a generator of a recursive algorithm, according Huber's theory, it is defined the functional based on the least favourable probability distribution for a given class of probability distribution.

This paper considers the model order selection using robust Akaike's criterion. The recursive algorithm for the OE (output error) model with time invariant parameters has also been discussed. Robustness of the used robust OE parameter estimation algorithm is accomplished by introducing the nonlinear transformation of prediction error (Huber's function). Robust recursive algorithms based on this idea in identification of linear dynamical systems are discussed in Refs. [7, 8] while in an area of nonlinear filtering are discussed in Ref. [9].

The performances of the algorithm are described through simulation results that demonstrate the superiority of the proposed algorithm in relation to the linear algorithm (derived under the assumption that the stochastic noise has a Gaussian distribution).

The rest of the paper is organized as follows: Section 2 considers the robust identification algorithm for OE models; Section 3 presents the robust version of Akaike's information criterion; illustration of the algorithm behavior based on simulations is given in Section 4 and concluding remarks are given in Section 5.

2. Robust Parameter Estimation for OE Model

The general form of the OE model is

$$y(k) = \frac{B(q^{-1})}{A(q^{-1})} u(k) + e(k) \tag{1}$$

where $u(k)$, $y(k)$ and $e(k)$ are input, output and stochastic noise, respectively. Polynomials $A(q^{-1})$ and $B(q^{-1})$ have the form:

$$\begin{aligned} A(q^{-1}) &= 1 + a_1 q^{-1} + \ldots + a_n q^{-n} \\ B(q^{-1}) &= b_1 q^{-1} + \ldots + b_m q^{-m} \end{aligned} \tag{2}$$

The block structure of OE model is given in Fig. 1.

Practical and theoretical studies have shown that in a stochastic model of the system there are some observations that are inconsistent with the largest part of the population (outliers) [5], and that is why the disturbance (measurement noise) $e(k)$ in the model (1) is a non-Gaussian. Hence, the probability density function of the disturbance belongs to approximately normal distribution class:

$$\mathcal{P}_\varepsilon = \{p(e) : p(e) = (1-\varepsilon) p_1(e) + \varepsilon p_2(e)\} \tag{3}$$

which $p_1(e) : \mathcal{N}(0, \sigma_1^2)$, $p_2(e) : \mathcal{N}(0, \sigma_2^2)$, $\sigma_2^2 >> \sigma_1^2$.

In other words, the probability density function $p(e)$ represents a mixture of normal (Gaussian) distributions where σ_1^2 and σ_2^2 denote variances. The parameter $0 \le \varepsilon < 1$ is called the degree of contamination.

Let us introduce an auxiliary model:

$$y_M(k) = \frac{B(q^{-1})}{A(q^{-1})} u(k) \tag{4}$$

Since the parameters a_i (i = 1, …, n) and b_i (j = 1, …, m) are unknown, their estimates are used, so the output of the auxiliary model is calculated as

$$\begin{aligned} \hat{y}_M(k) &= -\hat{a}_1 \hat{y}_M(k-1) - \ldots - \hat{a}_n \hat{y}_M(k-n) + \\ &+ \hat{b}_1 u(k-1) + \ldots + \hat{b}_m u(k-m) \end{aligned} \tag{5}$$

Let $\hat{\theta}$ is the estimated vector of OE parameters, and $\varphi(k)$ is the observation vector of OE parameters:

$$\begin{aligned} \hat{\theta} &= [\hat{a}_1, \ldots, \hat{a}_n, \hat{b}_1, \ldots, \hat{b}_m]^T \\ \varphi(k) &= [-\hat{y}_M(k-1) \ldots -\hat{y}_M(k-n), u(k-1) \ldots u(k-m)]^T \end{aligned} \tag{6}$$

At the moment k, before the estimate $\hat{\theta}(k)$ is known, the prediction of the model is [8]

$$\hat{y}_M(k) = \hat{\theta}^T(k-1)\varphi(k) \tag{7}$$

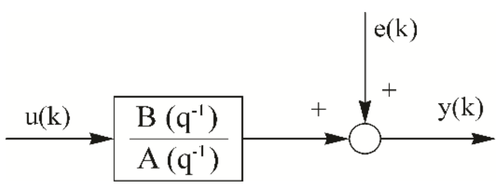

Fig. 1 Output error model.

The natural definition of the prediction error (residual) is

$$\varepsilon(k) = y(k) - \hat{y}_M(k) \tag{8}$$

The identification criterion (a generator of recursive parameter estimation procedure) is based, according to OE methodology, on the prediction error and has a mathematical form, for systems with constant parameters:

$$J(\theta) = E\{\Phi(\varepsilon(k))\} \tag{9}$$

in which $\Phi(\cdot)$ represents a robust loss function, which would suppress undesirable observations:

$$\Phi(\cdot) = -\log p^*(\cdot) \tag{10}$$

In the last relation, $p^*(.)$ represents the least favourable distribution of probability for a given class of probability distribution (3). This distribution is obtained by using the mathematical machinery of robust statistics [6].

The empirical functional for systems with time-invariant parameters has the form (obtained from Eq. (9) for sufficiently large k):

$$J_k(\theta) = \frac{1}{k} \sum_{t=1}^{k} \{\Phi(\varepsilon(i))\} \tag{11}$$

Expanding $J_k(\theta)$ in the vicinity of the preceding estimate $\hat{\theta}(\mathrm{k}-1)$ in Taylor series, one obtains:

$$\begin{aligned} J_k(\theta) &= J_k\left(\hat{\theta}(\mathrm{k}-1)\right) + \nabla_\theta J_k\left(\hat{\theta}(\mathrm{k}-1)\right)\left[\theta - \hat{\theta}(\mathrm{k}-1)\right] + \\ &+ O\left(\left\|\theta - \hat{\theta}(\mathrm{k}-1)\right\|^2\right) \end{aligned} \tag{12}$$

where

$$\lim_{\|x\| \to \infty} \frac{O(\|x\|)}{\|x\|} = 0 \tag{13}$$

and $\|\cdot\|$ denotes the Euclidean norm. The desired value $\hat{\theta}(k)$ can be obtained by solving the equation:

$$\nabla_\theta J_k\left(\hat{\theta}(\mathrm{k})\right) = 0 \tag{14}$$

Based on Eq. (12) after differentiating, twice one can obtain:

$$k\nabla_\theta^2 J_k(\theta) = (k-1)\nabla_\theta^2 J_{k-1}(\theta) + \Psi'(\varepsilon(k))\varphi(k)\varphi^T(k) \tag{15}$$

A derivative of the loss function $\Psi(\cdot) = \Phi'(\cdot)$, for the class of ε-contaminated distributions of probabilities is Huber's function, and it is given by

$$\Psi(\varepsilon(k)) = \min\{|\varepsilon(k), k_\varepsilon|\}\operatorname{sgn}(\varepsilon(k)) \tag{16}$$

where k_ε is a suitable chosen constant [6]. Huber's function and its derivative are shown in Fig. 2.

Let us assume further that the following assumptions are satisfied:

(a) The estimate $\hat{\theta}(k)$ is in the vicinity of the estimate $\hat{\theta}(k-1)$;

(b) The estimate $\hat{\theta}(k-1)$ is optimal at the instant k-1.

Taking into account $\theta = \hat{\theta}(k-1)$ in Eq. (15), and bearing in mind the assumption a), one can obtain

$$\begin{aligned} &k\nabla_\theta^2 J_k(\hat{\theta}(\mathrm{k}-1)) = (k-1)\nabla_\theta^2 J_{k-1}(\hat{\theta}(k-2)) + \\ &+\Psi'(\varepsilon(k))\varphi(k)\varphi^T(k) \end{aligned} \tag{17}$$

Based on the assumption (a) it also follows that $O(\|\cdot\|)$ in Taylor's series (12) becomes 0.

By introducing the notation $\overline{R}(k) = k\nabla_\theta^2 J_k(\hat{\theta}(k-1))$ from the solution of Eqs. (14) and (17), one can obtain

$$\hat{\theta}(k) = \hat{\theta}(k-1) - \overline{R}^{-1}(k)\left[k\nabla_\theta J_k\left(\hat{\theta}(k-1)\right)\right] \tag{18}$$

$$\overline{R}(k) = \overline{R}(k-1) + \Psi'(\varepsilon(k))\varphi(k)\varphi^T(k) \tag{19}$$

From the assumption (b), it follows $\nabla_\theta J_{k-1}(\hat{\theta}(k-1)) = 0$. Based on this condition, and if $\theta = \hat{\theta}(k-1)$ is put in the first derivative of the solution of Eq. (14), a recursive robust algorithm, from Eqs. (18)-(19), is finally obtained as

$$\hat{\theta}(k) = \hat{\theta}(k-1) + \overline{R}^{-1}(k)\varphi(k)\Psi(\varepsilon(k)) \tag{20}$$

$$\overline{R}(k) = \overline{R}(k-1) + \Psi'(\varepsilon(k))\varphi(k)\varphi^T(k) \tag{21}$$

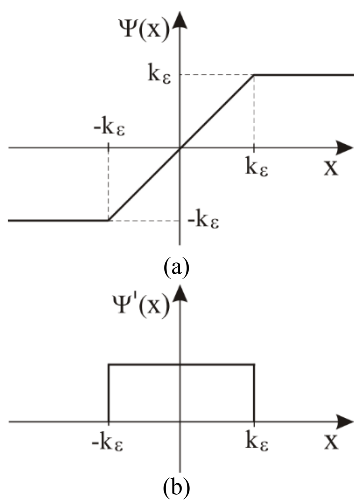

Fig. 2 Nonlinear function of residuals (a) Huber's function; (b) derivative of Huber's function.

To avoid computing the inverse matrix $\overline{R}^{-1}(k)$ in each iteration, let us introduce the matrix $P(k) = \overline{R}^{-1}(k)$. Using this notation and applying the matrix inversion lemma [1], from Eqs. (20) and (21), one can obtain the definitive form of a recursive algorithm for identification of dynamic systems with time-invariant parameters:

$$\hat{\theta}(k) = \hat{\theta}(k-1) + P(k)\varphi(k)\Psi(\varepsilon(k)) \tag{22}$$

$$P(k) = P(k-1) - \frac{P(k-1)\varphi(k)\varphi^T(k)P(k-1)}{\left[\Psi'(\varepsilon(k))\right]^{-1} + \varphi^T(k)P(k-1)\varphi(k)} \tag{23}$$

3. Robust Akaike's Criterion

In a general case, a model of system can be described by an assumed probability density function of measurements. This probability density is put in correspondence with the exact probability density measurements. The consistency between two probability densities describes the Kullback-Leibler information distance. By minimization of the information distance it is obtained the criterion for determining the model order [1]. For the given model order this criterion is identical to the maximum likelihood criterion. If it is assumed that the model (1) has constant parameters and a stochastic noise $e(k)$ has a Gaussian distribution, Akaike's criterion has the form:

$$W_A(k) = \sum_{i=1}^{k} \varepsilon^2(i) + p, \quad p = n + m \tag{24}$$

in which k represents a number of measurements and p is a number of parameters. In this paper, it is necessary to define Akaike's criterion for a general case:

(1) The system parameters are time-invariant;

(2) The stochastic noise has a non-Gaussian distribution described by Eq. (3).

Based on Eq. (10) and the least favourable distribution of probability for a given class of probability distribution (3), it is obtained:

$$\Phi(\varepsilon(k)) = \begin{cases} \dfrac{\varepsilon^2(k)}{2\sigma_1^2} + \ln\dfrac{\sqrt{2\pi}\sigma_1}{1-\varepsilon} & |\varepsilon(k)| \le k_\varepsilon \\ \dfrac{k_\varepsilon}{\sigma_1^2}\left(|\varepsilon(k)| - \dfrac{k_\varepsilon}{2}\right) + \ln\dfrac{\sqrt{2\pi}\sigma_1}{1-\varepsilon} & |\varepsilon(k)| > k_\varepsilon \end{cases} \tag{25}$$

Since in the paper estimation algorithm is based on robust statistics [2], the criterion for the selection of the model structure will be called robust Akaike's criterion. Taking into account conditions (1) and (2) this criterion has the form:

$$W_{RA}(k) = \sum_{i=1}^{k} \Phi(\varepsilon(k)) + p, \quad p = n + m \tag{26}$$

Based on the point of criterion minimum (26), polynomial orders $A(\cdot,\cdot)$ and $B(\cdot,\cdot)$ are determined.

Remark 1: The criterion (26) determine models collection because when p is determined from minimum of the criterion there are multiple combinations of polynomial orders m and n which satisfy the condition. Because, it is adopted:

$$n = m, \quad p = 2n \tag{27}$$

4. Simulation Results

The proposed robust Akaike's criterion has been tested on the following OE model:

$$y(k) = \frac{0.5q^{-1} + 0.3q^{-2}}{1 - 0.7q^{-1} + 0.5q^{-2}} u(k) + e(k) \tag{28}$$

The system identification example, is based on measured 1,000 input-output data points obtained during the experiments.

During the simulations, it is assumed that measured noise has non-Gaussian distribution:

$$\mathcal{P}_{\varepsilon} = \left\{ p(e) = (1-\varepsilon)\cdot \mathcal{N}(0;0.1) + \varepsilon \cdot \mathcal{N}(0;10) \right\} \tag{29}$$

PRBS signal is used for input signal. Figs. 3-5 show noise signal, system input and corresponding system output, respectively.

Based on the point of criterion minimum (26), for nine different model orders, it is shown that the observed system can be best described by a second order model (Fig. 6).

To demonstrate the superiority of the proposed robust OE identification algorithm, a comparison with linear OE identification algorithm [8], when input signal is PRBS signal, is made.

The simulation results are compared in terms of mean square error (MSE), defined by

$$MSE = \ln\left(\mathrm{E}\left\| \hat{\theta}(k) - \theta(k) \right\|^2 \right) \tag{30}$$

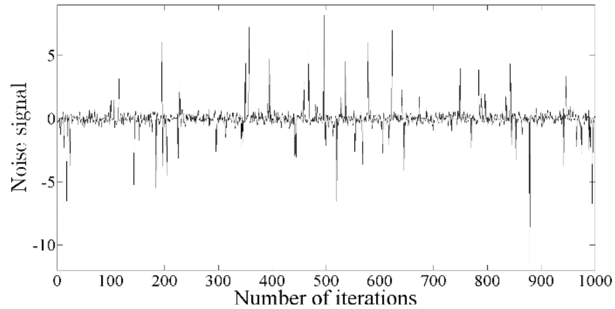

Fig. 3 A non-Gaussian noise sequence, $\varepsilon = 0.1$.

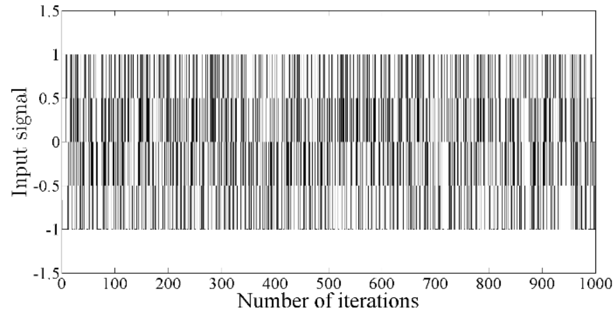

Fig. 4 PRBS excitation signal.

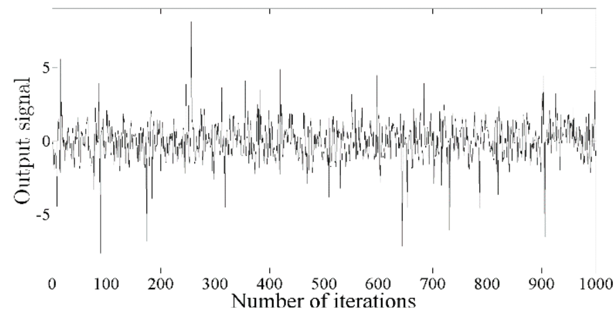

Fig. 5 Measured output signal of the system with contamination $\varepsilon = 0.1$.

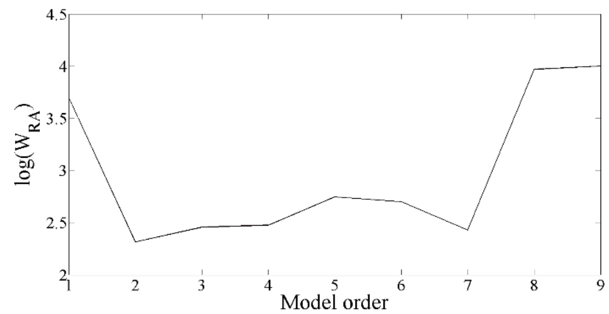

Fig. 6 RAIC criterion for selection of model order.

Figs. 7-9 show parameter estimates, and mean square errors.

Figs. 10 and 11 show noise signal and system output respectively, the contamination $\varepsilon = 0.1$.

Figs. 12-14 show parameter estimates, and mean square errors.

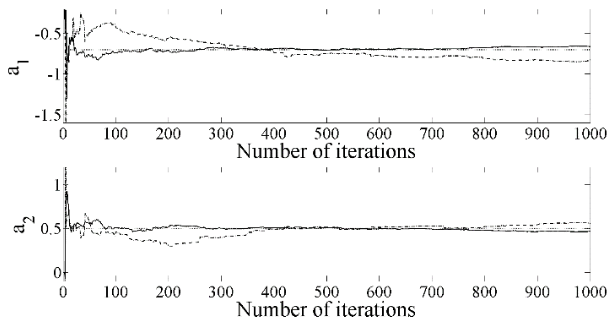

Fig. 7 Estimates of parameters a_1 and a_2 obtained in non-Gaussian noise environment with contamination $\varepsilon = 0.1$ (solid line: parameter estimates robust OE, dash-dot: parameter estimates using linear OE algorithm, dotted line: true parameter values).

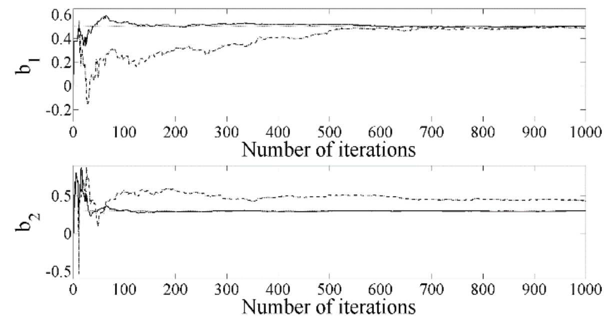

Fig. 8 Estimates of parameters b_1 and b_2 obtained in non-Gaussian noise environment with contamination $\varepsilon = 0.1$ (solid line: parameter estimates robust OE, dash-dot: parameter estimates using linear OE algorithm, dotted line: true parameter values).

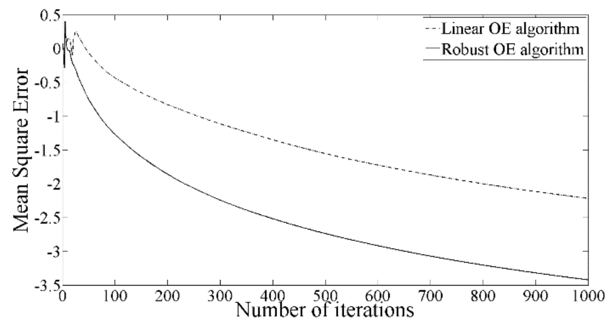

Fig. 9 Mean square error, obtained in non-Gaussian noise environment with contamination $\varepsilon = 0.1$.

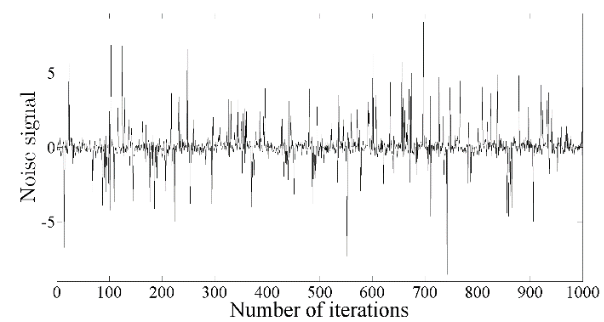

Fig. 10 A non-Gaussian noise sequence $\varepsilon = 0.2$.

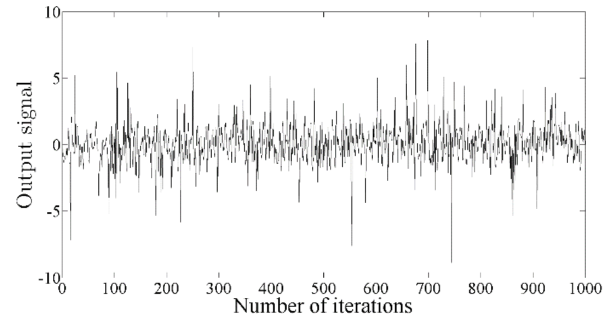

Fig. 11 Measured output signal of the system with contamination $\varepsilon = 0.2$.

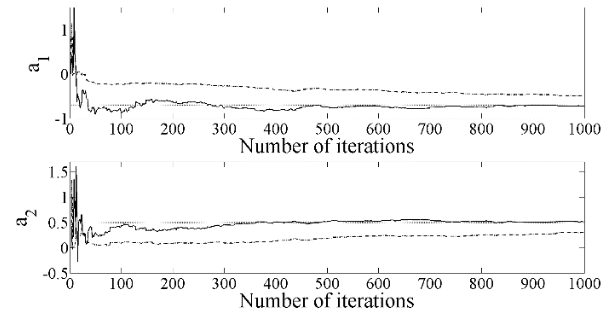

Fig. 12 Estimates of parameters a_1 and a_2 obtained in non-Gaussian noise environment with contamination $\varepsilon = 0.2$ (solid line: parameter estimates robust OE, dash-dot: parameter estimates using linear OE algorithm, dotted line: true parameter values).

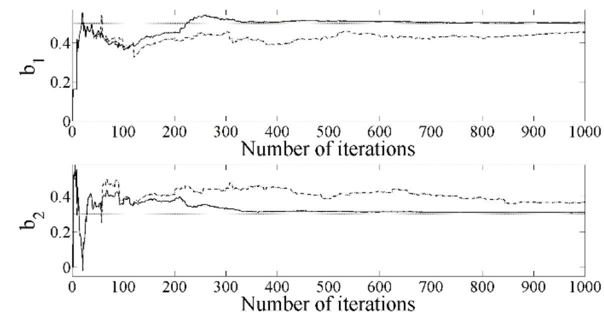

Fig. 13 Estimates of parameters b_1 and b_2 obtained in non-Gaussian noise environment with contamination $\varepsilon = 0.2$ (solid line: parameter estimates robust OE, dash-dot: parameter estimates using linear OE algorithm, dotted line: true parameter values).

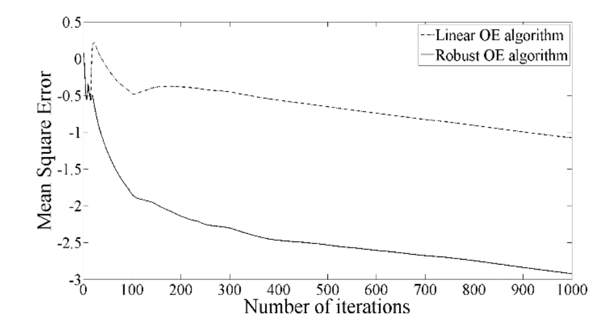

Fig. 14 Mean square error, obtained in non-Gaussian noise environment with contamination $\varepsilon = 0.2$.

The simulation results have shown that the classical linear algorithm is very sensitive to the presence of non-Gaussian noise, as opposed to the proposed robust algorithm. Comparing Figs. 9 and 14, it can be clearly seen that the superiority of the proposed robust algorithm is greater in higher degrees of contamination.

5. Conclusions

The paper considers robust identification of OE models with time invariant parameters where observations are disturbed by non-Gaussian noise. On the other hand, it is well known that ad hoc selection of model orders leads to overparametrization or parsimony problem. The natural frame to avoid these problems was Akaike's criterion for model order selection. The synthesis of robust algorithms is of primary interest because the presence of non-Gaussian noises can considerably degrade the performance of linearly algorithms. Based on Huber's theory of robust statistics, the robust versions of identification algorithm and Akaike's criterion for model order selection were proposed.

Simulation results have illustrated significant increasing of accuracy in parameter estimates of OE model by using the proposed robust identification algorithm with robust proposed robust version of Akaike's criterion for model order selection, in relation to the traditional linear identification algorithm.

Acknowledgments

The authors would like to express their gratitude to the Serbian Ministry of Education, Science and Technological Development for supporting this paper through projects TR33026 and TR33027.

References

[1] L. Ljung, System Identification: Theory for the User, 2nd ed., Prentice Hall, New Jersey, USA, 1999.

[2] T. Söderström, P. Stoica, System Identification, Prentice Hall, London, England, 1989.

[3] S. Kullback, R.A. Leibler, On information and sufficiency, The Annals of Mathematics Statistics 22 (1) (1951) 76-86.

[4] H. Akaike, A new look at the statistical model identification, IEEE Transactions on Automatic Control 19 (6) (1974) 716-723.

[5] V. Barnett, T. Lewis, Outliers in Statistical Data, 3rd ed., Wiley-Blackwell, New York, USA, 1994.

[6] P.J. Huber, E.M. Ronchetti, Robust Statistics, 2nd ed., Wiley, New Jersey, USA, 2009.

[7] V. Filipovic, N. Nedic, V. Stojanovic, Robust identification of pneumatic servo actuators in the real situations, Forschung im Ingenieurwesen 75 (4) (2011) 183-196.

[8] V. Stojanovic, V. Filipovic, Adaptive input design for identification of output error model with constrained output, Circuits, Systems, and Signal Processing 33 (1) (2014) 97-113.

[9] V. Stojanovic, N. Nedic, Robust Kalman filtering for nonlinear multivariable stochastic systems in the presence of non-Gaussian noise, International Journal of Robust and Nonlinear Control (2015) 1-16.

[10] I.D. Landau, R. Lozano, M. M'Saad, Adaptive Control, 1st ed., Springer, Berlin, Germany, 1998.