XIII International SAUM Conference
on Systems, Automatic Control and Measurements
Niš, Serbia, November 09th-11th, 2016

Outlier Robust One-Step-ahead Adaptive Predictor for Hammerstein Models

Lj. M. Dubonjić +V. Ž. Filipović ++ and V. S. Đorđević +++

[+] Department of Automatic Control, Robotics and Fluid Technique, University of Kragujevac, Faculty of Mechanical and Civil Engineering, Dositejeva 19, 36000 Kraljevo, Serbia, Phone: (381) 036-383-269, E-Mail: dubonjic.lj@mfkv.kg.ac.rs

[++] Department of Automatic Control, Robotics and Fluid Technique, University of Kragujevac, Faculty of Mechanical and Civil Engineering, Dositejeva 19, 36000 Kraljevo, Serbia, Phone: (381) 036-383-269, E-Mail: v.filipovic@mts.rs

[+++] Department of Automatic Control, Robotics and Fluid Technique, University of Kragujevac, Faculty of Mechanical and Civil Engineering, Dositejeva 19, 36000 Kraljevo, Serbia, Phone: (381) 036-383-269, E-Mail: djordjevic.v@mfkv.kg.ac.rs

Abstract *– In paper considered outliers-robust recursive stochastic approximation algorithm for adaptive prediction of MIMO (multiple-input multiple output) Hammerstein models. The static nonlinear block has polynomial form and linear block is output-error model. It is supposed that a priori known class of distributions to which belongs the real disturbance. In that situation we can use Huber`s methodology for design of robust algorithm which introduces nonlinear transformation of prediction error. Model transformation allows representation of unknown matrix parameters in the form of vector. The problem is not considered before in the field of adaptive prediction. Simulation study presents the practical behaviour of algorithm.*

Key words: *Hammerstein model, outliers, prediction, stochastic approximation*

I. INTRODUCTION

Output error (OE) models are frequently used for identification, prediction and adaptive control system. This paper focuses on the synthesis of one-step-ahead predictors, for MIMO non-linear systems. The non-linear model discussed in the paper belongs to class of block-oriented models. The example is a Hammerstein's model where static non-linear block is given in polynomial form while linear model is OE model.

Adaptive prediction focuses on recursive estimation of future values of system outputs based on past and present values of system inputs and outputs [1]. Applications of prediction algorithms are numerous [2]-[3]. Adaptive prediction in this work is based on algorithm of stochastic approximation. Standard algorithms of adaptive prediction assume that distribution of stochastic disturbance is exactly known (usually it is Gaussian distribution). However, analysis of practical measurements has shown that, in population of observations, exists rare large observations (outliers) [4]. Direct implication is that assumption on exact knowledge of disturbance distribution should be replaced by the assumption on a priori knowledge of the class of distribution to which the relevant disturbance belongs. The theory of robust statistics is built on this basis [5]-[6]. Based on this theory it is possible to get, in the statistical sense, robust recursive algorithms (reduced sensitivity to change of the disturbance distribution) for estimation of the parameters of dynamic phenomena. The application of the above mentioned ideas in the problems of identification and predictions was demonstrated in references [7]-[11] and [17]-[22]. Simulations have shown superiority of robust algorithms in relation to classical (linear) algorithms. In regards to algorithms based on stochastic approximation it is necessary to invest additional effort for increase of speed of algorithm convergence (especially on initial iterations).

As far as authors are informed, new robust recursive algorithm for prediction of non-linear MIMO OE models is proposed in this paper. Namely, algorithm is modified so that the unknown parameters are given in the form of vector. For approximately normal distributions non-linear transformation of prediction error is Huber's function and algorithm gain is exactly defined for this case (matrix M in algorithm) which, as well, represents new result. This was achieved by the application of Laplace functions. [12].

The algorithm developed in this paper is generalized form of algorithm for identification of MIMO systems when unknown parameter has vector form [14].

II. ROBUST RECURSIVE STOCHASTIC GRADIENT ALGORITHM

Suppose that the considered system is described by the nonlinear multivariable output-error (OE) model with r-dimensional input and p-dimensional output:

$$\boldsymbol{y}_k = q^{-1}\boldsymbol{F}^{-1}\left(q^{-1}\right)\boldsymbol{B}\left(q^{-1}\right) f\left(\boldsymbol{u}_k\right) + \boldsymbol{w}_k \tag{1}$$

where $\boldsymbol{B}\left(q^{-1}\right)$ and $\boldsymbol{F}^{-1}\left(q^{-1}\right)$ are matrix polynomials and q^{-1} denotes the shift-back operator $\left(q^{-1}\boldsymbol{x}_k = \boldsymbol{x}_{k-1}\right)$. Orders of polynomials $\boldsymbol{B}\left(q^{-1}\right)$ and $\boldsymbol{F}^{-1}\left(q^{-1}\right)$ are *m* and *n*, respectively

$$\boldsymbol{B}\left(q^{-1}\right) = \boldsymbol{B}_0 + \boldsymbol{B}_1 q^{-1} + ... + \boldsymbol{B}_m q^{-m} \tag{2}$$

$$\boldsymbol{F}\left(q^{-1}\right) = \boldsymbol{I} + \boldsymbol{F}_1 q^{-1} + ... + \boldsymbol{F}_n q^{-n} \tag{3}$$

where $\boldsymbol{B}_i\left(i = 1,2,...,m\right)$ are p x r matrices, and $\boldsymbol{F}_i\left(i = 1,2,...,n\right)$ are p x p matrices. The stochastic disturbance $\left\{\boldsymbol{w}_k\right\}$ is a martingale-difference in relation to the nondecreasing family of σ-algebras $\left\{F_k\right\}$. The Hammerstein model is given on the Figure 1.

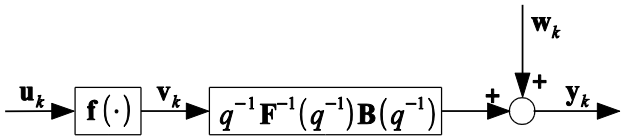

Fig.1. Hammerstein model

The methodology for adaptive prediction, in this paper, is based on the model reference control paradigm [15]-[16].

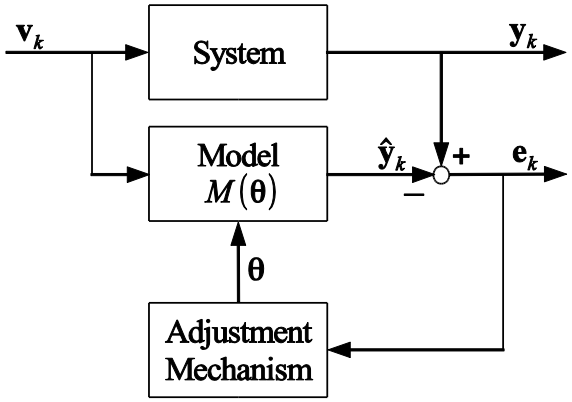

Fig. 2. The model reference concept

Output of the system is compared to that of adjustable model. The model parameters are updated until the difference cannot be further improved. The procedure is presented in Figure 2 where the model $M(\theta)$, for known parameters, is

$$y_k^0 = q^{-1}\boldsymbol{F}\left(q^{-1}\right)\boldsymbol{B}\left(q^{-1}\right)\boldsymbol{v}_k = q^{-1}\boldsymbol{F}\left(q^{-1}\right)\boldsymbol{B}\left(q^{-1}\right) f\left(\boldsymbol{u}_k\right) \quad (4)$$

It is supposed that $\{\boldsymbol{w}_k\}$ and $\{\boldsymbol{u}_k\}$ are independent. The optimal one-step-ahead predictor can be given by minimization of the following criterion

$$J_p = E\left\{\left\|\boldsymbol{e}_k^0\right\|^2 | F_{k-1}\right\} \quad , \quad \boldsymbol{e}_k^0 = \boldsymbol{y}_k - \boldsymbol{y}_k^0 \quad (5)$$

where $\boldsymbol{y}_k^0$ is an optimal prediction for one step. Introducing (1) in (5), one gets

$$\begin{aligned} & E\left\{\left\|\boldsymbol{y}_k - \boldsymbol{y}_k^0\right\|^2 | F_{k-1}\right\} = \\ & = E\left\{\left\|q^{-1}\boldsymbol{F}^{-1}\left(q^{-1}\right)\boldsymbol{B}\left(q^{-1}\right)\boldsymbol{f}\left(\boldsymbol{u}_k\right) - \boldsymbol{y}_k^0\right\|^2 | F_{k-1}\right\} + \\ & + E\left\{\left\|\boldsymbol{w}_k\right\|^2 | F_{k-1}\right\} + \\ & \quad + 2E\left\{\boldsymbol{w}_k^T\left(q^{-1}\boldsymbol{F}^{-1}\left(q^{-1}\right)\boldsymbol{B}\left(q^{-1}\right)\boldsymbol{f}\left(\boldsymbol{u}_k\right) - \boldsymbol{y}_k^0\right) | F_{k-1}\right\} \end{aligned} \quad (6)$$

The third term of (6) is zero owing the property of $\boldsymbol{w}_k$ that is independent with respect to $\boldsymbol{u}_k, \boldsymbol{u}_{k-1,\ldots}$. The second term does not depend upon the choice of $\boldsymbol{y}_k^0$ and the criterion will be minimized if the first term becomes zero. This leads to nonlinear optimal predictor

$$\boldsymbol{y}_k^0 = q^{-1}\boldsymbol{F}^{-1}\left(q^{-1}\right)\boldsymbol{B}\left(q^{-1}\right)\boldsymbol{f}\left(\boldsymbol{u}_k\right) \quad (7)$$

Model (1) belongs to the class of block-oriented models (Hammerstein model) [13]. Function $\boldsymbol{f}(\boldsymbol{u}_k)$ is nonlinear vector function

$$\boldsymbol{f}\left(\boldsymbol{u}_k\right) = \left[f_1\left(u_k^1\right), f_2\left(u_k^2\right) \ldots f_r\left(u_k^r\right)\right]^T, \quad \boldsymbol{f}\left(\boldsymbol{u}_k\right) \in R^r \quad (8)$$
$$\boldsymbol{u}_k = \left[u_k^1, u_k^2, \ldots, u_k^r\right]^T$$

where $f_i\left(u_k^i\right)$ $(i = 1,2,\ldots,\mathrm{r})$ is a nonlinear function of a known basis $(\gamma_1, \gamma_2, \ldots, \gamma_s)$ and unknown parameters $d_i\,(i = 1,2,\ldots,\mathrm{s})$ [13]

$$f_i\left(u_k^i\right) = d_1\gamma_1\left(u_k^i\right) + d_2\gamma_2\left(u_k^i\right) + \ldots + d_s\gamma_s\left(u_k^i\right) \quad , \; i = 1,2,\ldots,\mathrm{r} \quad (9)$$

Remark 1. Special form of the nonlinear part (relation (9)) in the Hammerstein model is a polynomial of a known order in the input

$$f_i\left(u_k^i\right) = d_1 u_k^i + d_2\left(u_k^i\right)^2 + \ldots + d_s\left(u_k^i\right)^s$$

It is, also, possible to represent nonlinear part of model in the form of cubic splines [18]-[19].

From relation (9) we have

$$\begin{aligned} \boldsymbol{f}\left(\boldsymbol{u}_k\right) &= d_1\boldsymbol{\gamma}_1\left(\boldsymbol{u}_k\right) + d_2\boldsymbol{\gamma}_2\left(\boldsymbol{u}_k\right) + \ldots + d_s\boldsymbol{\gamma}_s\left(\boldsymbol{u}_k\right) \\ \boldsymbol{\gamma}_i^T\left(\boldsymbol{u}_k\right) &= \left[\gamma_i\left(u_k^1\right)\gamma_i\left(u_k^2\right)\ldots\gamma_i\left(u_k^r\right)\right] \end{aligned} \quad (10)$$

Let us notice that

$$\begin{aligned} & q^{-1}\boldsymbol{B}\left(q^{-1}\right)\boldsymbol{f}\left(\boldsymbol{u}_k\right) = \\ & = d_1\left[\boldsymbol{B}_0\boldsymbol{\gamma}_1\left(\boldsymbol{u}_{k-1}\right) + \boldsymbol{B}_1\boldsymbol{\gamma}_2\left(\boldsymbol{u}_{k-2}\right) + \ldots + \boldsymbol{B}_m\boldsymbol{\gamma}_1\left(\boldsymbol{u}_{k-m-1}\right)\right] + \\ & + d_2\left[\boldsymbol{B}_0\boldsymbol{\gamma}_2\left(\boldsymbol{u}_{k-1}\right) + \boldsymbol{B}_1\boldsymbol{\gamma}_2\left(\boldsymbol{u}_{k-2}\right) + \ldots + \boldsymbol{B}_m\boldsymbol{\gamma}_2\left(\boldsymbol{u}_{k-m-1}\right)\right] + \\ & \vdots \\ & + d_s\left[\boldsymbol{B}_0\boldsymbol{\gamma}_s\left(\boldsymbol{u}_{k-1}\right) + \boldsymbol{B}_1\boldsymbol{\gamma}_s\left(\boldsymbol{u}_{k-2}\right) + \ldots + \boldsymbol{B}_m\boldsymbol{\gamma}_s\left(\boldsymbol{u}_{k-m-1}\right)\right] \end{aligned} \quad (11)$$

Remark 2. The static nonlinear block is described as

$$\boldsymbol{v}_{k-i} = \boldsymbol{f}\left(\boldsymbol{u}_{k-i}\right) \quad , \quad i = 0,1,2,\ldots. \quad (12)$$

whereby

$$\boldsymbol{f}\left(\boldsymbol{u}_{k-i}\right) = d_1\boldsymbol{\gamma}_1\left(\boldsymbol{u}_{k-i}\right) + \ldots + d_s\boldsymbol{\gamma}_s\left(\boldsymbol{u}_{k-i}\right), \; i = 0,1,2,\ldots. \quad (13)$$

The shift-back operator q^{-i} is applied to $\boldsymbol{v}_k$, i.e.

$$q^{-i}\boldsymbol{v}_k = \boldsymbol{v}_{k-i}$$

From (7) it follows that

$$\boldsymbol{y}_k^0 = \left(\boldsymbol{\theta}^M\right)^T \boldsymbol{X}_k^0 \quad (14)$$

where

$$\begin{aligned} & \left(\boldsymbol{X}_k^0\right)^T = \left[-\left(\boldsymbol{y}_{k-1}^0\right)^T, -\left(\boldsymbol{y}_{k-2}^0\right)^T, \ldots, \right. \\ & \left(\boldsymbol{y}_{k-n}^0\right)^T \boldsymbol{\gamma}_1^T\left(\boldsymbol{u}_{k-1}\right), \boldsymbol{\gamma}_1^T\left(\boldsymbol{u}_{k-2}\right), \ldots, \boldsymbol{\gamma}_1^T\left(\boldsymbol{u}_{k-m-1}\right), \\ & \boldsymbol{\gamma}_2^T\left(\boldsymbol{u}_{k-1}\right), \boldsymbol{\gamma}_2^T\left(\boldsymbol{u}_{k-2}\right), \ldots, \boldsymbol{\gamma}_2^T\left(\boldsymbol{u}_{k-m-1}\right), \ldots, \\ & \left.\boldsymbol{\gamma}_s^T\left(\boldsymbol{u}_{k-1}\right), \boldsymbol{\gamma}_s^T\left(\boldsymbol{u}_{k-2}\right), \ldots, \boldsymbol{\gamma}_s^T\left(\boldsymbol{u}_{k-m-1}\right)\right] \end{aligned} \quad (15)$$

$$\left(\boldsymbol{\theta}^M\right)^T = \left[\left[\boldsymbol{F}_1, \boldsymbol{F}_2, ..., \boldsymbol{F}_n, d_1\boldsymbol{B}_0, d_1\boldsymbol{B}_1, ..., d_1\boldsymbol{B}_m, \right.\right.$$
$$\left.\left. d_2\boldsymbol{B}_0, d_2\boldsymbol{B}_1, ..., d_2\boldsymbol{B}_m, ..., d_s\boldsymbol{B}_0, d_s\boldsymbol{B}_1, ..., d_s\boldsymbol{B}_m\right]\right. \quad (16)$$

We, also, have from (1), (7) and (15)

$$\boldsymbol{y}_k = \left(\boldsymbol{\theta}^M\right)^T \boldsymbol{X}_k^0 + \boldsymbol{w}_k \quad (17)$$

Note that in relation (17) the quantities $\boldsymbol{y}_{k-i}^0 (i = 1,2,\ldots,n)$ are unknown. We will replace them with

$$\hat{\boldsymbol{y}}_k^0 = \left(\boldsymbol{\theta}_k^M\right)^T \boldsymbol{X}_k \quad (18)$$

where $\boldsymbol{\theta}_k^M$ is estimate of unknown parameter and

$$\boldsymbol{X}_k^T = \left[\left(\hat{\boldsymbol{y}}_{k-1}^0\right)^T, -\left(\hat{\boldsymbol{y}}_{k-2}^0\right)^T, ..., \right.$$
$$\left(\hat{\boldsymbol{y}}_{k-n}^0\right)^T \boldsymbol{\gamma}_1^T\left(\boldsymbol{u}_{k-1}\right), \boldsymbol{\gamma}_1^T\left(\boldsymbol{u}_{k-2}\right), ..., \boldsymbol{\gamma}_1^T\left(\boldsymbol{u}_{k-m-1}\right), \quad (19)$$
$$\boldsymbol{\gamma}_2^T\left(\boldsymbol{u}_{k-1}\right), \boldsymbol{\gamma}_2^T\left(\boldsymbol{u}_{k-2}\right), ..., \boldsymbol{\gamma}_2^T\left(\boldsymbol{u}_{k-m-1}\right), ...,$$
$$\left.\boldsymbol{\gamma}_s^T\left(\boldsymbol{u}_{k-1}\right), \boldsymbol{\gamma}_s^T\left(\boldsymbol{u}_{k-2}\right), ..., \boldsymbol{\gamma}_s^T\left(\boldsymbol{u}_{k-m-1}\right)\right]$$

Let us introduce the matrix

$$\boldsymbol{\varphi}_k = \begin{bmatrix} \boldsymbol{X}_k^T & & 0 \\ & \ddots & \\ 0 & & \boldsymbol{X}_k^T \end{bmatrix} = \boldsymbol{I} \otimes \boldsymbol{X}_k^T \quad (20)$$

where the symbol $\otimes$ denotes the Kronecker product.

For the p x r matrix $\boldsymbol{U} = \left[u_{ij}\right]$ the operator "vec" generates a column vector by setting the columns of the matrix U one below the other.

$$\text{vec } \boldsymbol{U} = \begin{bmatrix} \boldsymbol{U}^1 \\ \boldsymbol{U}^2 \\ \vdots \\ \boldsymbol{U}^r \end{bmatrix}, \text{vec } \mathbf{U} \in R^{(pr)x1}$$

Based on (19) let us introduce the vector:

$$\boldsymbol{\theta} = vec\ \boldsymbol{\theta}^M, \quad \dim\boldsymbol{\theta} = np^2 + m(2pr + s - 1)$$

The sequence of estimate $\left\{\boldsymbol{\theta}_k^M\right\}$ will be defined recursively. At time k , before the estimate $\boldsymbol{\theta}_k^M$ is available, output of the model $M(\boldsymbol{\theta})$ is (according with the output error methodology)

$$\hat{\boldsymbol{y}}_k = \left(\boldsymbol{\theta}_{k-1}^M\right)^T X_k \quad (21)$$

According with above considerations we have

$$\left(\boldsymbol{\theta}^M\right)^T X_k = \boldsymbol{\varphi}_k\boldsymbol{\theta} \quad (22)$$

and the prediction error

$$\boldsymbol{e}_k = \boldsymbol{y}_k - \hat{\boldsymbol{y}}_k = \boldsymbol{y}_k - \boldsymbol{\varphi}_k\boldsymbol{\theta}_{k-1} \quad (23)$$

will be used for definition of identification functional.

The dominant assumption in the literature is that the probability density for stochastic disturbance $\boldsymbol{w}_k$ is exactly known (usually normal). However, numerous researches have shown that such assumption is not justified in practice. Because, in this paper it is assumed that the only distribution class to which the disturbance belongs is known. We will consider next distribution class

$$P_\varepsilon = \left\{P : P = (1-\varepsilon)N + \varepsilon G \ ,\ G \text{ is symmetric}\right\} \quad (24)$$

where $\varepsilon \in [0,1)$ is the contamination degree and $N\left(0,\sigma^2\right)$ denotes a zero-mean Gaussian distribution with a variance σ^2. This distribution is suited to describe the presence of outliers in observations. The distribution G is an arbitrary symmetric distribution. For the Huber's theory is important the restriction of the form (24)

$$P = \{P : P \geq (1-\varepsilon)N\}$$

which is more convenient for solving variational problem whose solution is the least favorable probability density (relation (25)).

We will further consider random vector $\boldsymbol{w}_k^T = \left[w_k^1 w_k^2 .... w_k^p\right]$. It is assumed that component of the vector $\boldsymbol{w}_k$ are independent. Each component has the distribution (24). Applying Huber`s methodology, the least favorable probability density on a class (24) is obtained

$$p_i\left(w_k^i\right) = \begin{cases} \dfrac{1-\varepsilon}{\sqrt{2\pi}\sigma_i}\exp\left\{-\dfrac{\left(w_k^i\right)^2}{2\sigma_i^2}\right\}, & \left|w_k^i\right| < k_\varepsilon^i \\ \dfrac{1-\varepsilon}{\sqrt{2\pi}\sigma_i}\exp\left\{-\dfrac{k_\varepsilon}{\sigma_i^2}\left(\left|w_k^i\right| - \dfrac{k_\varepsilon}{2}\right)\right\}, & \left|w_k^i\right| > k_\varepsilon^i \end{cases} \quad (25)$$

where the relationship between the contamination degree ε and the parameter k_ε of Huber's function is given by the following relation

$$\frac{2\Phi_N\left(k_\varepsilon^i\right)}{k_\varepsilon^i} - 2\Phi_N\left(-k_\varepsilon^i\right) = \frac{\varepsilon}{1-\varepsilon} \ ,$$
$$\Phi_N(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{y^2}{2}}dy \quad (26)$$

Due to the independence of the components of the vector $\boldsymbol{w}_k$, the least favorable probability density of the vector w_k is:

$$p^*\left(\boldsymbol{w}_k\right) = \prod_{i=1}^{p} p_i^*\left(w_k^i\right) \quad (27)$$

On the basis of (27), by using the maximum likelihood methodology, it can be defined the loss function:

$$\Phi(\boldsymbol{x}) = -\log p^*(\boldsymbol{x}) \ , \ \Phi(\cdot): R^p \to R^1 \quad (28)$$

Using (22) and (27) the identification criterion can be defined

$$J(\boldsymbol{\theta}) = E\left\{\Phi\left(\boldsymbol{e}_k\right)\right\} \quad (29)$$

The recursive minimization of the criterion (29) can be realized by applying the Newton-Raphson algorithm

$$\boldsymbol{\theta}_k = \boldsymbol{\theta}_{k-1} - \left[k\nabla_{\theta}^2 J_k\left(\boldsymbol{\theta}_{k-1}\right)\right]^{-1}\left[k\nabla_{\theta} J_k\left(\boldsymbol{\theta}_{k-1}\right)\right] \tag{30}$$

where

$$J_k\left(\boldsymbol{\theta}\right) = \frac{1}{k}\sum_{i=1}^{k}\Phi\left(\boldsymbol{e}_i\right) \tag{31}$$

is empirical functional.

The resulting form of stochastic gradient algorithm is (it follows from relation (30))

$$\boldsymbol{\theta}_k = \boldsymbol{\theta}_{k-1} + \frac{1}{r_k}\boldsymbol{\varphi}_k^T\boldsymbol{\psi}\left(\boldsymbol{e}_k\right) \quad , \quad \hat{\boldsymbol{\theta}}_0 = \boldsymbol{0} \tag{32}$$

$$r_k = r_{k-1} + tr\left\{\boldsymbol{\varphi}_k^T\boldsymbol{M}\boldsymbol{\varphi}_k\right\} \quad , \quad r_0 = 1 \tag{33}$$

where $r_k = tr\left[k\nabla_{\theta}^2 J_k\left(\boldsymbol{\theta}_{k-1}\right)\right]^{-1}$, $\boldsymbol{\theta} = vec\boldsymbol{\theta}^M$ ($\boldsymbol{\theta}^M$ is defined with relation (16)) and $\boldsymbol{\psi}\left(\cdot\right)$ is defined as $\boldsymbol{\psi}\left(\cdot\right) = \Phi'\left(\cdot\right)$ The matrix **M** has the form

$$\boldsymbol{M} = \begin{bmatrix} m_1 & & 0 \\ & \ddots & \\ 0 & & m_p \end{bmatrix} \tag{34}$$

where

$$m_i = E\left\{\psi_i'\left(w_k^i\right)\right\} \tag{35}$$

The predictor is, according with relations (21) and (22)

$$\hat{\boldsymbol{y}}_k = \boldsymbol{\varphi}_k\boldsymbol{\theta}_{k-1} \tag{36}$$

The adaptive predictor is described with relations (32), (33) and (36).

From last fact and relations (25) and (35) it follows that

$$\begin{aligned} m_i &= E\left\{\psi_i'\left(w_k^i\right)\right\} = \int_{-\infty}^{\infty}\psi_i'\left(w_k^i\right)p^*\left(w_k^i\right)dw_k^i = \\ &= \frac{2\left(1-\varepsilon\right)}{\sqrt{2\pi}}\int_0^{k_{\varepsilon}^i} e^{-\frac{1}{2}\left(\frac{w_k^i}{\sigma_{1i}}\right)^2} d\left(\frac{w_k^i}{\sigma_{1i}}\right) = 2\left(1-\varepsilon\right)\phi_L^i\left(\frac{k_{\varepsilon}^i}{\sigma_{1i}}\right) \end{aligned} \tag{37}$$

where

$$\phi_L\left(x\right) = \frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{y^2}{2}}dy$$

The algorithm (32) and (33) is the most useful form of the recursive algorithm when unknown parameter is in vector form [14].

V. CONCLUSION

In this paper is proposed outlier robust adaptive predictor for MIMO Hammerstein OE model. Linear part of model is MIMO OE model. The impact of outliers are reduced by introduction of non-linear prediction error transformation. The transformation depends on the a priori information about the stochastic disturbance. In this paper it is supposed that disturbance belongs to the a priori known class of distributions (not to the exact known distribution).

ACKNOWLEDGMENT

This paper is part of the projects TR33026 and TR33027 at the University of Kragujevac, Faculty of Mechanical and Civil Engineering in Kraljevo, and was supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia.

REFERENCES

[1] Goodwin G, Ramadge P, Caines P. A global convergent adaptive predictor. *Automatica* 1981; **17**(1): 135-140.

[2] Jamal W, Das S, Oprescu I, Maharatna K. Prediction of synchrostate transmission in EEG signals using Markov chain model. *IEEE Trans. Signal Process. Letters* 2015; **22**(2): 149-152.

[3] Manoj C, Jagannatham A. Optimal prediction likelihood tree based source-channel ML decoder for wireless sensor networks. *IEEE Trans. Signal Process. Letters* 2014; **21**(2): 135-139.

[4] Pearson R. Exploring Data in Engineering the Sciences and Medicine. Oxford University Press: Oxford, 2011.

[5] Huber P-J. Robust estimation of a location parameter. *Annals of Mathematical Statistics* 1964;**35**(1):73-101.

[6] Huber P, Ronchetti E. *Robust statistics*. Wiley: New York, 2009.

[7] Filipovic V. Consistency of the robust recursive Hammerstein model identification algorithm. *Journal of the Franklin Institute* 2015;**352** (5):1932-1945.

[8] Filipovic V. Recursive identification of multivariable ARX models in the presence of a priori information: Robustness and regularization. *Signal Processing* 2015;**116** (1): 68-77.

[9] Filipovic V, Djordjevic V. Design of robust recursive identification algorithms for large- scale stochastic systems. *Facta Universitates, Series: Automatic Control and Robotics* 2015;**14** (1): 43-54.

[10] Filipovic V. Robust adaptive one-step-ahead predictor. *IMA Journal of Math. Control and Information* 2001;**18** (3): 491-501.

[11] Asato M, Poor V. Robust system-parameter identification: a polynomial approach. *International Journal of Robust and Nonlinear Control* 1992; **2**(2): 223-237.

[12] Papoulis A. Probability, Random Variables and Stochastic Processes. Mc Graw-Hill: New York, 1965.

[13] Giri E, Bai V (Eds.). *Block-oriented nonlinear system identification*. Springer: Berlin, 2010.

[14] Caines P-E. *Linear Stochastic Systems*. Wiley: New York, 1988.

[15] Landau I. A*daptive control*. *The Model Reference Approach*. Dekker: New York, 1979.

[16] Dugard L, Landau I. Recursive output-error identification algorithms. *Automatica* 1980; **16** (5): 443-462.

[17] Cerone V, Regruto D. Parameters bounds for discrete-time Hammerstein models with bounded output errors. *IEEE Transactions on Automatic Control* 2003; **48**(10): 1855-1860.

[18] Dempsey E, Westwick D. Identification of Hammerstein models with cubic spline nonlinearities. *IEEE Trans. Biomedical Engineering* 2004; **51**(2): 237-245.

[19] Chan K, Bao J, Whiten W. Identification of MIMO Hammerstein systems using cardinal spline functions. *Journal of Process Control* 2006; **16**(3): 659-670.

[20] Verhaegen M. Westwick D. Identifying MIMO Hammerstein systems in the context of subspace model identification methods. *International Journal of Control* 1996; **63**(2): 331-349.

[21] Goethals I, Pelckmans K, Suykens J, De Moor B. Identification of MIMO Hammerstein models using least squares support vector machines. *Automatica* 2005; **41**(7): 1263-1272.

[22] Louhichi B, Toumi A. Identification of MIMO Hammerstein models using singular value decomposition approach. *International Journal of Computer Applications* 2012; **42**(9): 29-37.