Multilayer Feed-Forward Neural Networks in Prediction and Predictive Control of Semi-Batch Reactor

Lubomir Macku and David Samek

***Abstract*—**The contribution studies prediction of the given semi-batch reactor using multilayer feed-forward neural networks. The two prediction approaches are tested – signal prediction approach and system prediction methodology. The first approach is commonly applied in time series prediction, while the input-output models in the second methodology are used for example in the control tasks. Furthermore, the resulting predictor is used for the model predictive control of the reactor in order to test performance of the developed method.

***Keywords*—** Artificial neural network, chemical reactor, prediction, predictive control.

I. Introduction

PREDICTION of nonlinear and complex systems can be performed by various methods. One of them is simplification and linearization that leads to linear models [1, 2]. Some authors use wavelet filtering in order to divide stochastic and deterministic parts that are modeled separately [4, 5]. Javadi et al. and Phaiboon present successful application of fuzzy logic to prediction [6, 7]. However, probably the most popular approach in these cases is based on artificial neural networks.

Artificial neural networks are commonly used in various fields, for example weather forecasting [8], time series prediction of financial data [9, 10], biology and medicine [11, 12], power engineering [13] and process control [14, 15]. There is lot of types of artificial neural networks, but not all of them are usable for prediction. The most common are multilayer feed-forward neural networks [8, 11, 16, 17]. Fairly wide group of artificial neural networks belongs to recurrent neural networks [18, 19]. Very popular due to their fast training are radial basis function neural networks [18, 19]. Al-Shayea and El-Refae use generalized regression neural network for prediction of Spanish banks data [16]. There are also other less frequent methods such as fuzzy-neural networks [20], adaptive linear networks [21], unsupervised Kohonen neural networks [22, 23]. Interesting generalization of neural networks, which is achieved by using multi-argument and learnable functions, bring functional networks [24].

Generally, it can be distinguished between two main approaches in the prediction. The first approach deals with the predicted signal only. It means that the predicted value must be predicted from known older values of the same signal. This approach is very common in financial engineering and it is usually denoted as a signal prediction or time series prediction (in case of time series).

The second approach uses other signals (usually system inputs) for the prediction and it is commonly called system or process prediction. This approach is widely used in the process industry, especially in the control (predictive control).

In this contribution both approaches to prediction are studied and the multilayer feed-forward neural networks are tested on the prediction of temperature in the chemical semi-batch reactor. The motivation of this work comes from our long time research focused on predictive control of chemical reactors. Usually, the predictor must be adapted during the function, because of its inaccuracy or changing parameters of the system to be controlled. The goal of the presented work is to design and test the proper predictor based on multilayer feed-forward neural network that would not require adaptation during control of the given reactor.

The paper is structured along these lines: First, the chemical semi-batch reactor is introduced. After that, artificial multilayer feed-forward neural networks are explained briefly. The following part describes the predictors design, their testing and validation. Then, the resulting predictors are compared and discussed. This chapter is followed by the explanation of the model predictive control of the given semi-batch reactor. The final part of the paper contains few concluding remarks.

II. Chemical Semi-Batch Reactor

In this paper, a semi-batch reactor model is used to study the prediction abilities of multilayer feed-forward neural networks and their applicability in the model predictive control. The model input data comes from a real process - the chromium waste recycling process [15]. The task of the predictor is to

Manuscript received July 30, 2012: Revised version received August 16, 2012.

L. Macku is with the Department of Electronics and Measurements, Faculty of Applied Informatics, Tomas Bata University in Zlin, nam. T. G. Masaryka 5555, 760 01 Zlin Czech Republic (e-mail: macku@fai.utb.cz).

D. Samek is with the Department of Production Engineering, Faculty of Technology, Tomas Bata University in Zlin, nam. T. G. Masaryka 5555, 760 01 Zlin Czech Republic (phone: +420-576-035-157; fax: +420-576-035-176; e-mail: samek@ft.utb.cz).

predict the in-reactor temperature.

Let us consider a single input – single output (SISO) system of the chemical exothermic semi-batch reactor, which is described by the following mathematical model:

$$\frac{\mathrm{d}\,m(t)}{\mathrm{d}\,t} = F_I \tag{1}$$

$$\frac{\mathrm{d}\,a(t)}{\mathrm{d}\,t} = \frac{F_I}{m(t)} - A \cdot e^{-\frac{E}{R \cdot T(t)}} \cdot a(t) \tag{2}$$

$$\begin{aligned} \frac{\mathrm{d}T(t)}{\mathrm{d}t} &= \frac{F_I \cdot c_I \cdot T_I}{m(t) \cdot c} + \frac{A \cdot e^{-\frac{E}{R \cdot T(t)}} \cdot \Delta H_r \cdot a(t)}{c} - \\ &- \frac{K \cdot S \cdot T(t)}{m(t) \cdot c} + \frac{K \cdot S \cdot T_C(t)}{m(t) \cdot c} \end{aligned} \tag{3}$$

$$\begin{aligned} \frac{\mathrm{d}\,T_C(t)}{\mathrm{d}\,t} &= \frac{F_C \cdot T_{CI}}{m_C} + \frac{K \cdot S \cdot T(t)}{m_C \cdot c_C} - \\ &- \frac{K \cdot S \cdot T_C(t)}{m_C \cdot c_C} - \frac{F_C \cdot T_C(t)}{m_C} \end{aligned} \tag{4}$$

where m is the total weight of the reaction components in the reactor, a is the mass concentration of the reaction component in the reactor, $c = 4500\ \mathrm{J \cdot kg \cdot K^{-1}}$ is the specific heat capacity of the reactor content, T is the temperature of the reactor content. F_I, $T_I = 293.15$ K and $c_I = 4400\ \mathrm{J \cdot kg \cdot K^{-1}}$ is the reaction component input mass flow rate, the temperature and the specific heat capacity. $F_C = 1\ \mathrm{kg \cdot s^{-1}}$, $T_{CI} = 288.15$ K, T_C, $c_C = 4118\ \mathrm{J \cdot kg \cdot K^{-1}}$ and $m_C = 220$ kg is the cooling water mass flow rate, the input temperature, the output temperature, the specific heat capacity and the weight of the cooling water in the cooling system of the reactor, respectively.

Other constants: $A = 219.588\ \mathrm{s^{-1}}$, $E = 29967.5087\ \mathrm{J \cdot mol^{-1}}$, $R = 8.314\ \mathrm{J \cdot mol^{-1} \cdot K^{-1}}$, $\Delta H_r = 1392350\ \mathrm{J \cdot kg^{-1}}$, $K = 200\ \mathrm{kg \cdot s^{-3} \cdot K^{-1}}$ and $S = 7.36\ \mathrm{m^2}$ is the effective heat-transfer area [15].

III. Multilayer Feed-Forward Neural Networks

Multilayer feed-forward neural networks (MFFNNs) are very often called backpropagation networks because of the typical training algorithm. Some authors prefer name multilayer perceptrons (MLP) [25, 26], because MFFNNs have been developed by generalization from Rosenblatt's perceptron with binary transfer function. This structure is depicted in the Fig. 1. The u_{in} stands for input value, the N is number of inputs and y is the output of the Rosenblatt's perceptron.

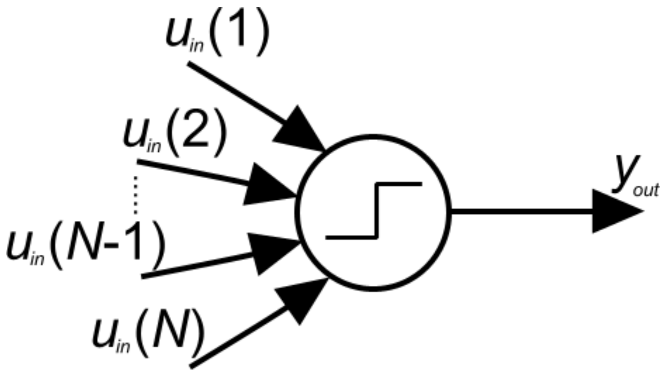

Fig. 1 Rosenblatt's perceptron

In MFFNN the signals flow only in the one direction (from the input to the output). All neurons are structured into layers and, typically, all neurons in the specific layer use same transfer function. In contrast to Rosenblatt's perceptron multilayer feed-forward neural networks use various transfer functions, usually continuous (e.g. linear, hyperbolic tangent, sigmoidal functions, etc.). Example of the two-layer feed-forward neural network is shown in the Fig. 2. The $\boldsymbol{u_{in}}$ is the input data vector, $\boldsymbol{W}$ is weighting matrix, $\boldsymbol{b}$ stands for bias vector, S is transfer function and $\boldsymbol{x}$ is output of internal layer.

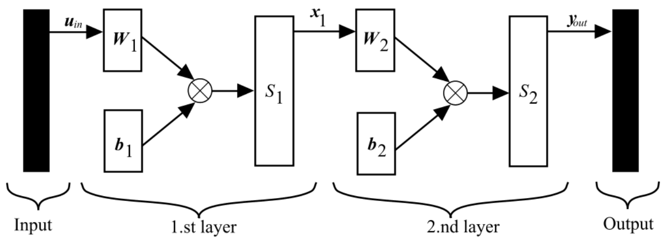

Fig. 2 Example of the two-layer feed-forward neural network

By applying of Kolmogorov theorem it has been proved [27] that two-layer MFFNN (with one hidden layer) can approximate any function with certain accuracy while non-polynomial transfer function in the hidden layer is used. This methodology is adopted in the paper.

IV. Prediction

As was already mention hereinbefore, there were studied two approaches of the prediction: signal prediction and system prediction. For all simulations in this paper MATLAB with Neural Network Toolbox, Optimization Toolbox and Simulink were used.

In the first case the artificial neural network used five past values of the predicted signal as the input vector and predicted only one step ahead. In other words, when it was needed the MFFNN repeatedly used its own predictions as inputs. Thus, in the input (zero) layer of the artificial network were five neurons and the output layer consisted one neuron.

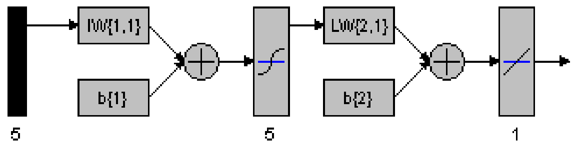

Fig. 3 Scheme of MFFNN with 5 input neurons (*net1*)

As is depicted in the Fig. 3, the network had five neurons in the hidden layer. This number was obtained by many experiments as sufficient for this case. The hidden layer used hyperbolic tangent sigmoid as the transfer function, while the output layer employed linear transfer function.

Three simulation results from our previous work [15] were chosen as the training and testing data. Only one simulation data were used for training (furthermore they will be symbolized as *data1*), whilst the rest two simulation results served as testing group (*data2* and *data3*). The training was performed using built-in Matlab/Neural Network Toolbox Levenberg-Marquardt backpropagation algorithm. This artificial neural network will be in the following text denoted as *net1*.

After the network training, the obtained predictor was tested to all three signals (Fig. 5, 7, 9). For better comparability the prediction error for all testing signals was computed (Fig 6, 8, 10).

In order to compare influence of the input signal knowledge to the prediction quality, the second predictor was also tested. This predictor used not only the predicted signal old values but also old values of the system input. In this case it was the chromium sludge dosing speed. Thus the multilayer feed-forward neural network involved ten neurons in the input layer (see Fig. 4). As well as in the previous approach, the hidden layer used hyperbolic tangent sigmoid as the transfer function, while the output layer employed linear transfer function. This artificial neural network will be in the following text denoted as *net2*.

With the intention of obtaining comparable results the same methodology as in case of *net1* was used. It means that same training data, the same training algorithm and same testing data were applied. Results are depicted in Fig. 11-16.

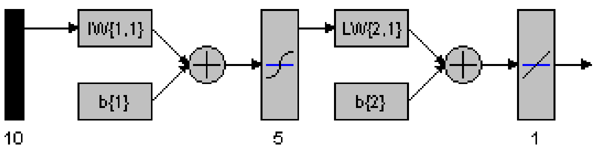

Fig. 4 Scheme of MFFNN with 10 input neurons (*net2*)

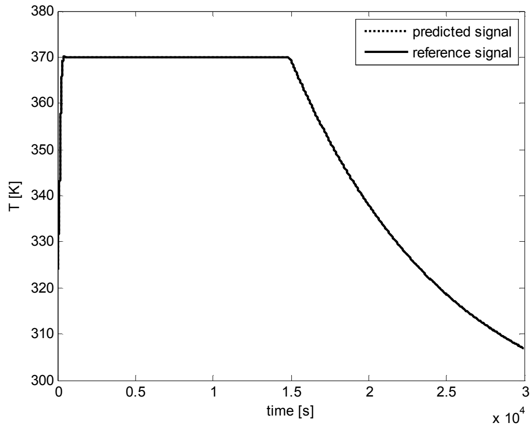

Fig. 5 *Net1* test on the training data (*data1*)

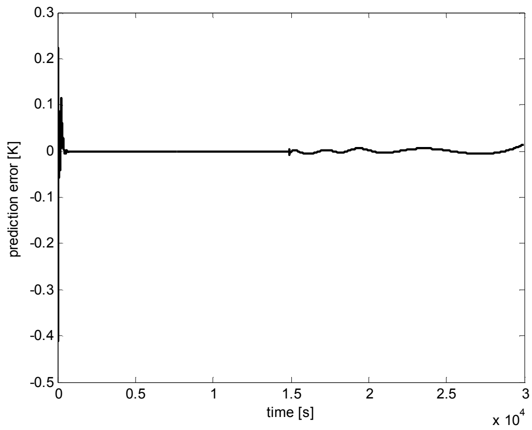

Fig. 6 Prediction error of *net1* for the training data (*data1*)

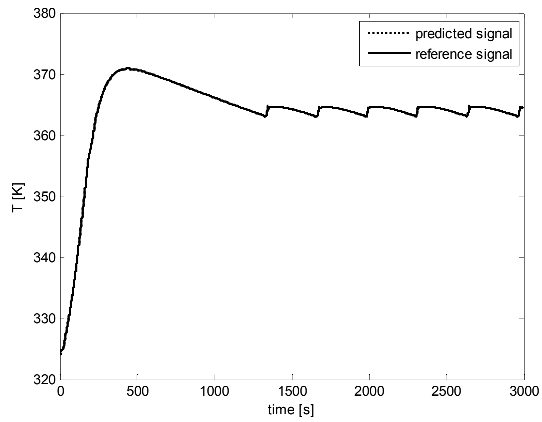

Fig. 7 *Net1* test on the testing data (*data2*)

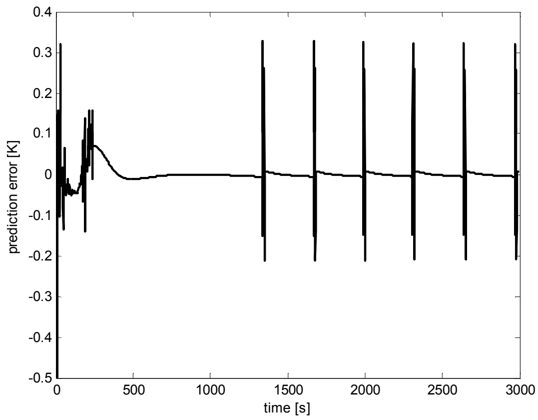

Fig. 8 Prediction error of *net1* for the testing data (*data2*)

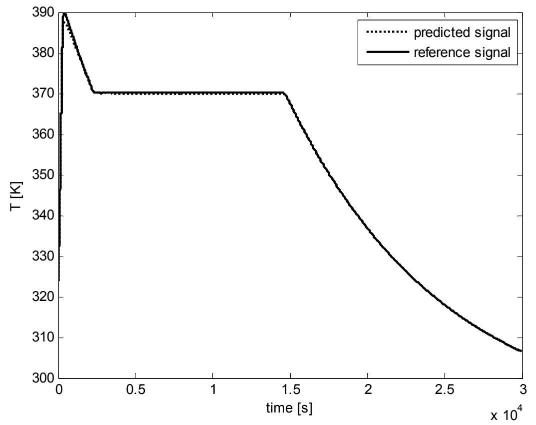

Fig. 9 *Net1* test on the testing data (*data3*)

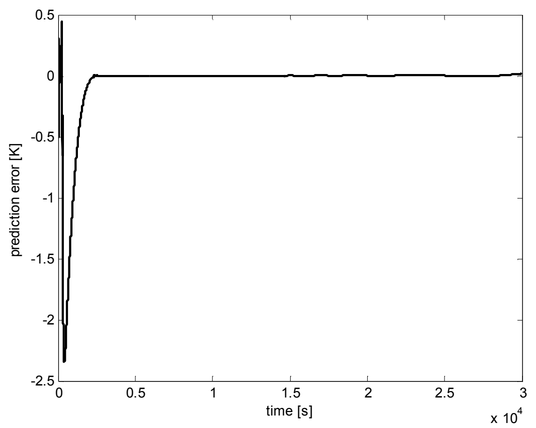

Fig. 10 Prediction error of *net1* for the testing data (*data3*)

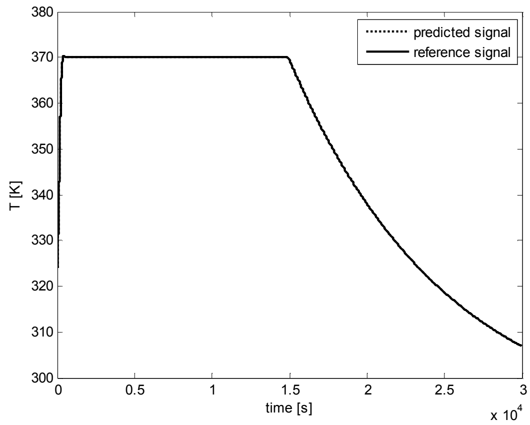

Fig. 11 *Net2* test on the training data (*data1*)

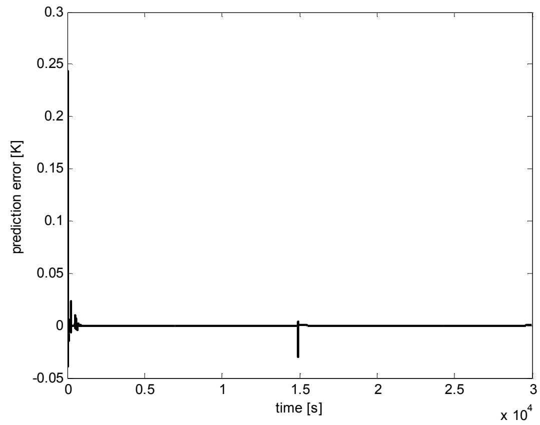

Fig. 12 Prediction error of *net2* for the training data (*data1*)

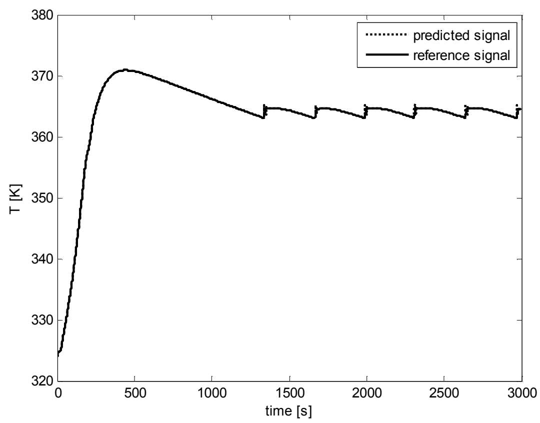

Fig. 13 *Net2* test on the testing data (*data2*)

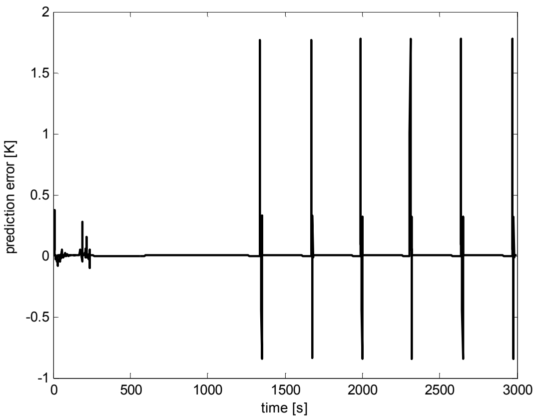

Fig. 14 Prediction error of *net2* for the testing data (*data2*)

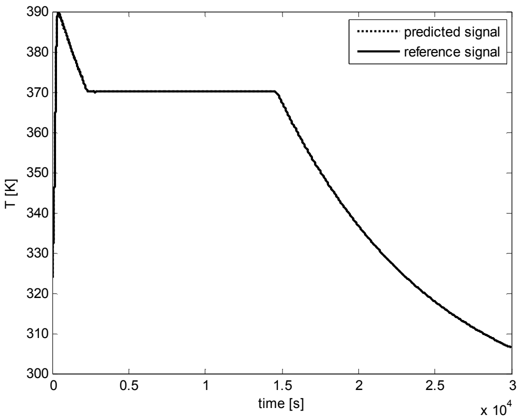

Fig. 15 *Net2* test on the testing data (*data3*)

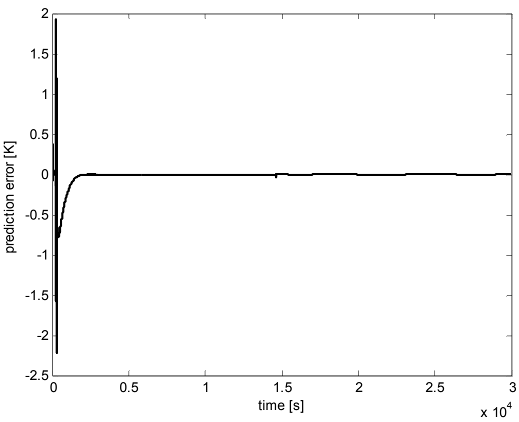

Fig. 16 Prediction error of *net2* for the testing data (*data3*)

V. Comparison of Predictors

For better comparison two prediction quality criterions were defined. The first criterion C_1 describes total sum of absolute values of prediction errors relative to number predictions, whilst the second criterion function C_2 characterizes total sum of squares of prediction errors relative to number predictions. The C_1 criterion gives same importance to all errors. On the other hand C_2 emphasizes higher errors and lower prediction errors are suppressed.

$$C_1 = \frac{\sum_{i=1}^{N} \left| t(i) - p(i) \right|}{N} \tag{5}$$

$$C_2 = \frac{\sum_{i=1}^{N} \left(t(i) - p(i) \right)^2}{N} \tag{6}$$

N stands for number of predictions (length of the predicted signal), t is target (original) signal, p denotes predicted signal and i is number of the prediction.

Table I. Prediction quality criterions

	net1		*net2*	
	C_1	C_2	C_1	C_2
data1	0.0024	4.529e-05	9.828e-05	2.218e-06
data2	0.0133	0.0013	0.0115	0.0107
data3	0.0565	0.0806	0.0155	0.0080

As can be seen from the results, the prediction is in both cases very good. However, if the best method should be judged, the second approach (*net2*) must be chosen, because it provided much better results with the one and only exception (criterion C_2 for *data2*). Therefore, this predictor will be tested in the model predictive controller of the given reactor in the following text.

VI. Model Predictive Control

The selected predictor *net2* was applied into model predictive controller (MPC) introduced in our previous paper [15]. The controller uses explicit predictor and optimization box based on the Levenberg-Marquart method. The criterion function J to be optimized is defined along these lines:

$$\begin{aligned} J = {} & \lambda \sum_{j=N_1}^{N_2} \left[y_r(k+j) - \hat{y}(k+j) \right]^2 + \\ & + \rho \sum_{j=1}^{N_u} \left[u_t(k+j-1) - u_t(k+j-2) \right]^2 + \\ & + \gamma(k) \sum_{j=1}^{N_u} u_t(k+j) \end{aligned} \tag{7}$$

$$\gamma(k) = \gamma(k-1) - \gamma_c \tag{8}$$

where the λ, ρ and γ parameters determine the contribution that the particular sum has on the performance index. The γ parameter is decreasing during the control. The speed of the change is defined by the (8) using parameter γ_c. In other words, the third sum in the beginning of the control has the maximum value, and after initial phase it equals to zero. The N_1, N_2 and N_u define horizons over which the tracking error and the control increments are evaluated. The u_t variable is the tentative control signal; y_r is the desired response and $\hat{y}$ is the predicted value of the temperature. Index k symbolizes step of the control (sample time of the control was 60s).

The settings of the controller are adopted from the [15] in order to obtain comparable result (see Table II). These settings will be in the following text denoted as *controller1*. However, with the intention of revealing the *net2* predictor abilities the other controller settings were tested. The all controllers used same horizons $N_2 = N_u = 8$ and parameter $N_1, = 1$.

As can be seen from Fig. 17, the *net2* predictor based MPC controller successfully controls temperature in the given chemical reactor. The dosing time was 3060 s and maximum in-reactor temperature reached 370.5 K. The decreasing temperature of the reaction mixture after the time 3060 s is caused by the fact that the reactor works in semi-batch cycle. In other words, after consumption of the all dosing batch (chromium sludge), the control is finished and the operator must wait until the reactor content cools down. The Fig. 18 shows course of the control signal (the dosing speed of the chromium sludge). The Fig. 19 depicts the development of the total mass of the in-reactor mixture. The speed of the reaction is illustrated in the Fig. 20, where can be seen steep changes of the chromium sludge concentration. The temperature of the cooling water in the cooling system was also observed in our simulations (Fig. 21).

Table II. MPC controller settings

	λ	ρ	γ	γ_c
controller1	1000	100000	10000	200
controller2	1000	100000	10000	100
controller3	1000	1	600000	10000
controller4	1000	100000	600000	10000
controller5	1000	100000	600000	18000

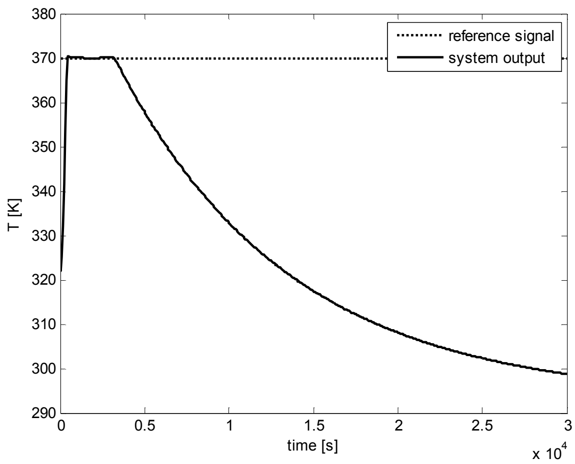

Fig. 17 The in-reactor temperature obtained using *controller1*

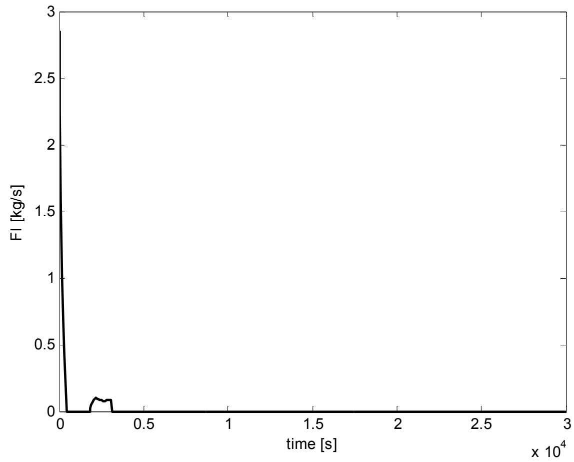

Fig. 18 The chromium-sludge dosing obtained using *controller1*

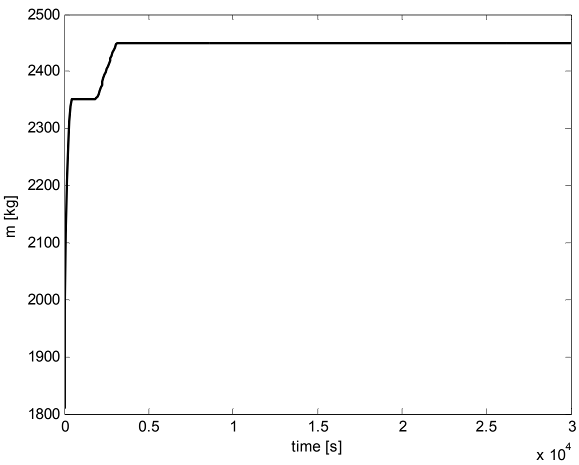

Fig. 19 The mass of the reaction mixture obtained using *controller1*

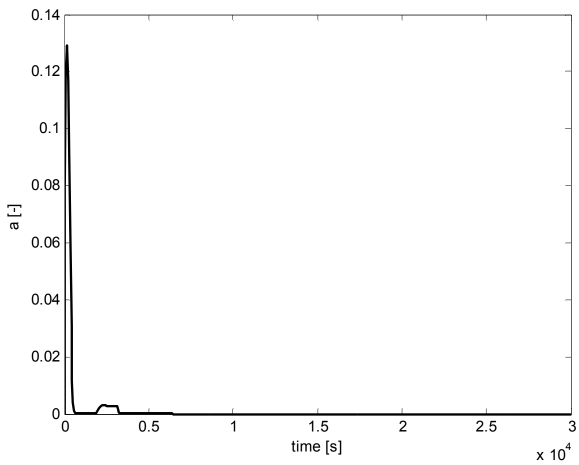

Fig. 20 The in-reactor chromium sludge concentration obtained using *controller1*

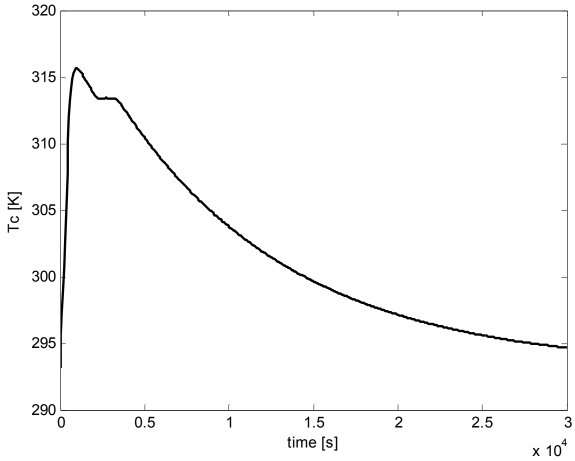

Fig. 21 The temperature in the cooling system obtained using *controller1*

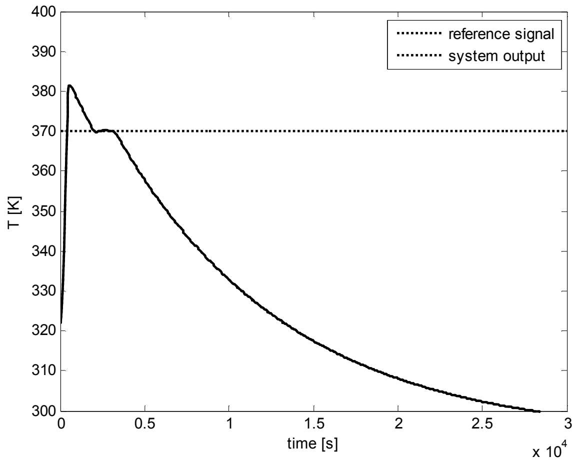

Fig. 22 The in-reactor temperature obtained using *controller2*

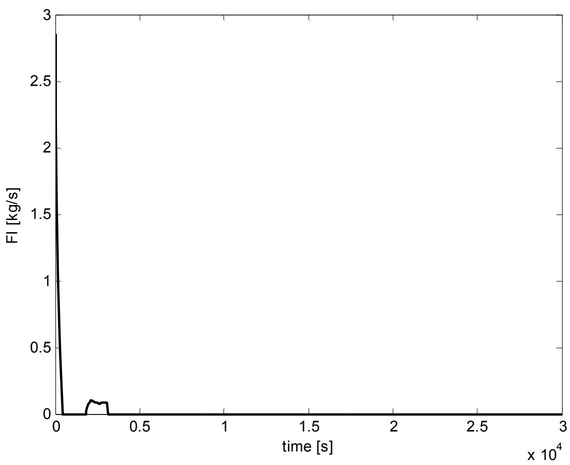

Fig. 23 The chromium-sludge dosing obtained using *controller2*

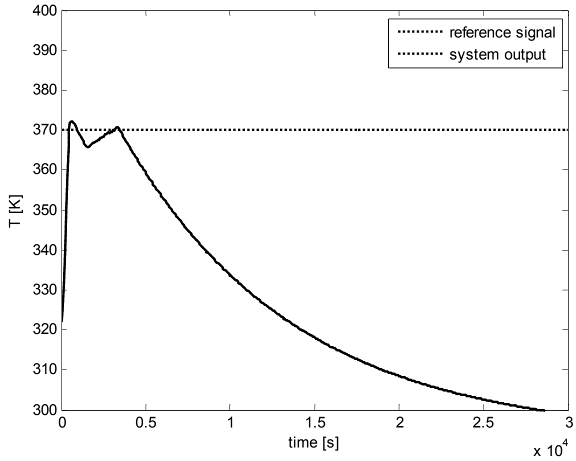

Fig. 24 The in-reactor temperature obtained using *controller3*

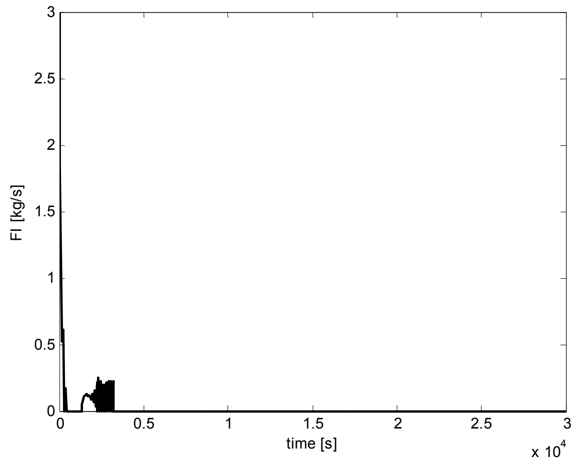

Fig. 25 The chromium-sludge dosing obtained using *controller3*

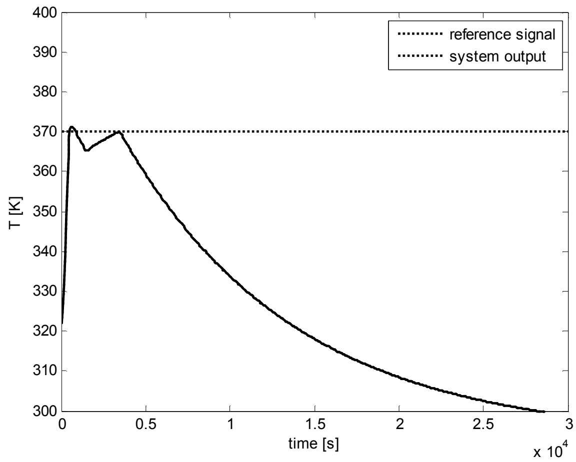

Fig. 26 The in-reactor temperature obtained using *controller4*

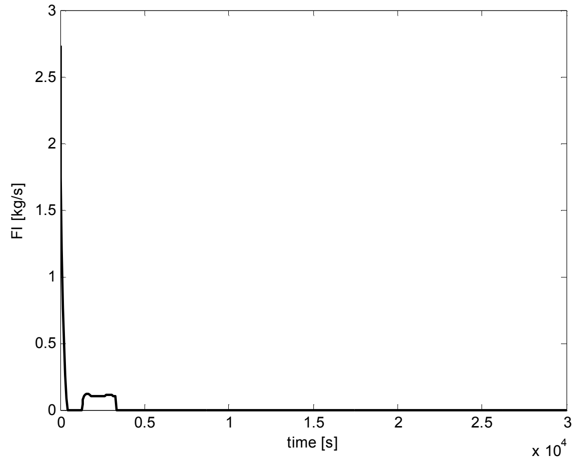

Fig. 27 The chromium-sludge dosing obtained using *controller4*

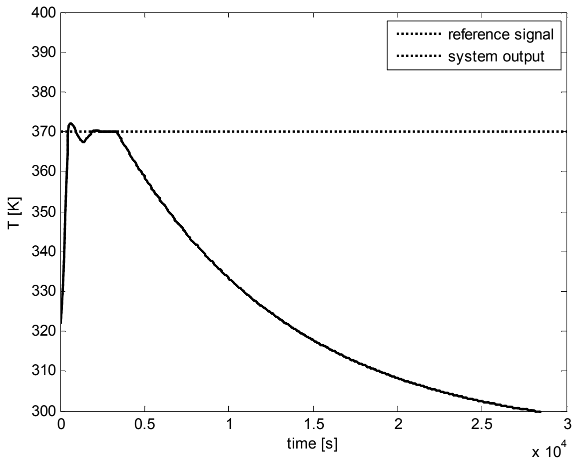

Fig. 28 The in-reactor temperature using *controller5*

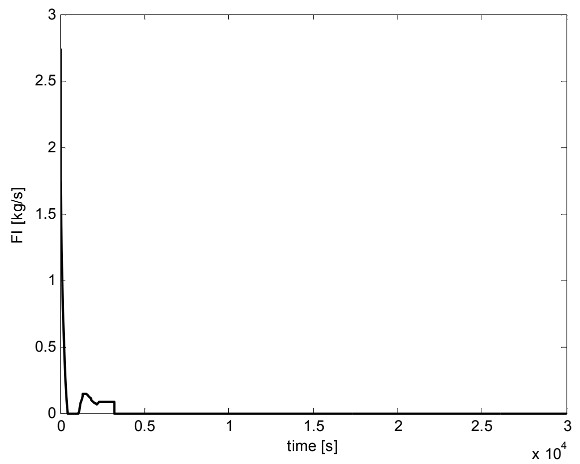

Fig. 29 The chromium-sludge dosing obtained using *controller5*

The Fig. 22 – 29 show the influence of various settings of the controller to the control result. Fig. 22 - 23 demonstrate that even small change of one parameter leads to unsatisfactory overshoot of the temperature. This behavior needs to be removed, but it leads to longer dosing times (Fig. 24 - 29). The *controller3* had the dosing time 3180 s and oscillating control signal F_I (Fig. 24 - 25). By increasing the ρ parameter of the controller the control signal becomes smoother, however the temperature is still fluctuating around the reference value and the dosing times are still longer (Fig. 26 and 29).

VII. Conclusion

This paper continues and advances our previous work published in [28] and [15]. The artificial neural network based predictor developed in this paper offers comparable results as three-layer feed forward neural network used earlier. The proposed predictor is simpler and faster with equivalent accuracy. The simulations of the various settings of the model predictive controller showed that even good and exact predictor may suffer when the controller is not properly set.

References

[1] E. F. Camacho and C. Bordons, "*Model Predictive Control in the Process Industry*." Springer - Verlag, 2004.

[2] P. Chalupa and V. Bobal, "Modelling and Predictive Control of Inverted Pendulum," In *22nd European Conference on Modelling and Simulation, European Council for Modelling and Simulation*, 2008, pp. 531-537.

[3] T. Minerva, "Wavelet Filtering for Prediction in Time Series Analysis," In: *Non-Linear Systems and Wavelet Analysis*, WSEAS Press, 2010, pp. 89-94.

[4] M. Dong, D. Yang, Y. Kuang, D. He, S. Erdal and D. Kenski, "PM2.5 concentration prediction using hidden semi-Markov model-based times series data mining," *Expert Systems with Apps*, Vol. 36, No. 5, 2009, pp. 9046-9055.

[5] J. Yan, C. Guo and X. Wang, "A dynamic multi-scale Markov model based methodology for remaining life prediction," *Mechanical Systems and Signal Processing*, Vol. 25, No. 4, 2011, pp. 1364-1376.

[6] S. Javadi and Z. Hojjatinia, "Wind Speed Modeling and Prediction in Wind Farms Using Fuzzy Logic," In *Applied Mathematics in Electrical and Computer Engineering*, WSEAS, 2012, pp. 98-103.

[7] S. Phaiboon, "RF Macro-cell Prediction Using Fuzzy Logic: Case Study in Bangkok City –Thailand," In *Latest Trends on Communications and Inf. Tech.*, WSEAS, 2010, pp. 105-109.

[8] I. Maqsood, M. R. Khan and A. Abraham, "An ensemble of neural networks for weather forecasting," *Neural Computing & Applications*, Vol.13, No.2, 2004, pp. 112-122.

[9] E. M. Azoff, "*Neural network Time Series Forecasting of Financial Markets*," John Wiley & Sons, 1994.

[10] E. Diaconescu, "The use of NARX Neural Networks to Predict Chaotic Time Series," *WSEAS Transactions on Computer Research*, Vol.3, No.3, 2008, pp. 182-191.

[11] G. Brion, Ch. Viswanathan, T. R. Neelakantan, S. Ligireddy, R. Girones, D. Lees, A. Allard and A. Vantarakis, "Artificial Neural Network Prediction of Viruses in Shellfish," *Applied and Enviromental Microbiology*, Vol.71, No.9, 2005, pp. 5244-5253.

[12] E. Pasomsub, C. Sukasem, S. Sungkanuparph, B. Kijsirikul and W. Chantratita, "The application of artificial neural networks for phenotypic drug resistance prediction: evaluation and comparison with other interpretation systems," *Japanese journal of infectious diseases*, Vol.63, No.2, 2010, pp. 87-94.

[13] A. D. Papalexopoulos, H. Shangyou and P. Tie-Mao, "An implementation of a neural network based load forecasting model for the EMS," *IEEE Transactions on Power Systems*, Vol.9, No.4, 1994, pp. 1956-1962.

[14] D. Samek and P. Dostal, "Artificial Neural Networks in Prediction and Predictive Control." In *Proceedings of the 22nd European Conference on Modelling and Simulation ECMS 2008*. European Council for Modelling and Simulation, 2008. pp. 525-530.

[15] L. Macku and D. Samek, "Two step, PID and model predictive control using artificial neural network applied on semi-batch reactor." *WSEAS TRANSACTIONS on SYSTEMS*. Vol.9, No.10, 2010, pp. 1039-1049.

[16] Q. K. Al-Shayea and G. A. El-Refae, "Evaluation of Banks Insolvency Using Artificial Neural Networks," In *Recent Researches in Artificial Intelligence and Database Management*, WSEAS, 2012, pp. 113-118.

[17] M. Oprea and A. Matei, "Applying Artificial Neural Networks in Environmental Prediction Systems," In *RECENT ADVANCES in AUTOMATION & INFORMATION*, WSEAS, 2010, pp. 110-115.

[18] R. Zemouri and P. C. Patic, "Prediction Error Feedback for Time Series Prediction: a way to improve the accuracy of predictions," In *Proceedings of the 4th EUROPEAN COMPUTING CONFERENCE*, WSEAS, 2010, pp. 58-62.

[19] M. H. B. Karami, "Application of Neural Networks In Short-Term Load Forecasting," In *7th WSEAS Int. Conf. on Mathematical Methods and Computational Techniques In Electrical Eng*., WSEAS, 2005, pp. 37-41.

[20] P. M. Patil, S. N. Kulkarni, D. D. Doye and U. V. Kulkarni, "Modular General Fuzzy Hypersphere Neural Network," In *Proceedings of the 4th WSEAS/ IASME International Conference on Systems Science and Simulation in Engineering*, WSEAS, 2005, pp. 211-216.

[21] D. Samek and D. Manas, "Artificial neural networks in artificial time series prediction benchmark," *International Journal of Mathematical Models and Methods in Applied Sciences*, Vol. 5, No. 6, 2011, pp. 1085-1093.

[22] F. Ortega, et al. *An Hybrid Approach to Prediction of Electric Load with MARS and Kohonen Maps*, [online] Available: http://neuron.tuke.sk/competition/reports/ FranciscoOrtega.pdf

[23] F. J. Marin, F. Garcia-Lagos, G. Joya and F. Sandoval, *Peak Load Forecasting using Kohonen Classification and Intervention Analysis*, [online] Available: http://neuron.tuke.sk/competition/reports/JavierMarin.pdf

[24] E. Castillo, B. Guijarro and A. Alonso, *Electricity load forecast using functional networks*, [online], Available: http://neuron.tuke.sk/ competition/reports/ BerthaGuijarro.pdf

[25] S. F. Crone and N. Kourentzes, "Feature selection for time series prediction – A combined filter and wrapper approach for neural networks," *Neurocomputing*, Vol. 73, 2010, pp. 1923-1936.

[26] E. Guresen, G. Kayakutlu and T. U. Daim, "Using artificial neural network models in stock market index prediction," *Expert Systems with Apps*, Vol. 38, No. 8, 2011, pp. 10389-10397

[27] M. Leshno, V. Y. Lin, A. Pinkus and S. Schocken, "Multilayer feedforward networks with a non-polynomial activation can approximate any function." *Neural networks*. 1993, Vol. 6, No. 6, pp. 861-867.

[28] L. Macku and D. Samek, "Prediction of Semi-Batch Reactor Using Multilayered Feed-Forward Neural Networks," in *Recent Advances in Circuits & Systems*, Kos, Greece, 2012, pp. 453 – 458.