Environment Protection Engineering

Vol. 37 2011 No. 2

MAŁGORZATA KABSCH-KORBUTOWICZ*, MAŁGORZATA KUTYŁOWSKA*

USE OF ARTIFICIAL INTELLIGENCE IN PREDICTING THE TURBIDITY RETENTION COEFFICIENT DURING ULTRAFILTRATION OF WATER

The artificial neural network (ANN) is widely used as new mathematical tool offering an alternative way of dealing with complex problems. To predict water quality after coagulation/ultrafiltration processes using an immersed membrane, various structures of multilayer perceptrone with one hidden layer were created. Feed water turbidity, turbidity in a tank, pH and temperature in the tank as well as transmembrane pressure and permeate flux were treated as input signals. Based on the best network chosen, prognosis of water quality was done. The created ANN was able to predict very accurately behaviour of a real system with relative errors of turbidity retention coefficient equal to 2.16×10^{-2}.

1. INTRODUCTION

According to strict regulations of water quality, in order to achieve the assumed level of purification, among water treatment technologies integrated membrane processes are applied. Hybrid processes being a combination of several physicochemical processes are believed to achieve a very high efficiency of water treatment. The processes of coagulation/ultrafiltration or ion exchange-sedimentation/ultrafiltration are state of art and nowadays are widely applied for drinking water treatment [1, 2]. Despite substantial progress made in recent years in membrane sciences, many initial problems associated with these processes such as limitations in the ability to control and predict water quality after the treatment process remain unresolved. Unfortunately, there is a lack of a traditional mathematical model solving the problem of permeate quality and no data are available on prediction permeate quality using artificial networks. Thus a suitable method of optimization should be developed to model the

*Institute of Environmental Protection Engineering, Wrocław University of Technology, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland. Corresponding author M. Kabsch-Korbutowicz, e-mail: malgorzata.kabsch-korbutowicz@pwr.wroc.pl

most important membrane parameters. The method of modelling based on an artificial neural network (ANN) is lately very popular in chemical and environmental engineering. Predicting the permeate quality using ANN will be a useful approach.

1.1. BASIC INFORMATION ABOUT ARTIFICIAL NEURAL NETWORKS

Human nervous system performance is the base of information transferring in artificial neural networks. The neurons in the network are data processors and are responsible for summarizing input signals. ANN just computes output values from input values. The sum of transferred information is weighted. These weighted connections are modified during the learning process. Nowadays, the most popular type of network is multilayer perceptrone, characterized by one input layer, one or more hidden layers and one output layer. It should be remembered that ANN modelling is like a "black box" approach and that is why it is impossible to penetrate deeply inside the way of forming the network structure, especially in the hidden layer. It is needed to create the artificial network for each problem separately. The solved problem extorts the number of neurons, the kind of activation function and the training methods. ANN has the ability to generalize and to approximate optimized matter. This issue is very important, due especially to the forecasting of membrane parameters because the changes during the separation are difficult to locate and to describe.

1.2. APPLICATION OF ANN IN MODELLING MEMBRANE PARAMETERS

Water quality variables which are complex and nonlinear need the application of alternative techniques of modelling [3]. In recent years, artificial neural networks were often used to model and forecast parameters of low-pressure membrane processes. The full literature review has been already summarized in the paper [4]. A lot of elaborations concern the modelling parameters of membrane processes in food industry (e.g., milk ultrafiltration and microfiltration or purification of juices) [5–7]. It is possible to use artificial intelligence to predict hydraulic parameters during membrane water and wastewater treatment [8]. Until now mainly the values of the quantity parameters were modelled [9, 10] (e.g., permeate flux or fouling rate). ANN were used also for predicting the quality parameters (retention coefficient of humic substances) [11] but such investigations are rather rare.

2. EXPERIMENTAL. COAGULATION/ULTRAFILTRATION OF WATER

In the experiments (done at University of Technology in Dresden, Institute of Urban Water Management, Chair of Water Supply Engineering by Professor Uhl's scientific group) the coagulation/ultrafiltration process was used to treat reservoir water

[12, 13]. The system was equipped with 3 submerged capillary modules ZeeWeed 500c produced by the Zenon company. The total surface of the membranes amounted to 60 m^2. Nominal pore size of the membrane was 0.04 μm and cut-off 200 kDa.

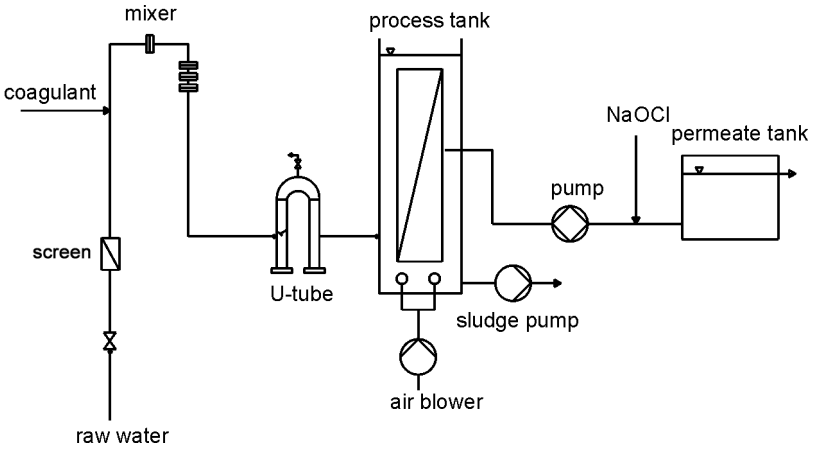

Fig. 1. Pilot plant set-up [13]

Figure 1 shows the set up of the pilot installation. Prior to coagulant ($Al_2(SO_4)_3$) dosage raw water was filtered by a screen filter in order to remove mechanical impurities. The coagulant (0.5 g $Al^{3+} \cdot m^{-3}$) was added before the tank and mixed precisely by cross-sectional changes to obtain a homogeneous solution. The first step of coagulation proceeded in a U-tube which had the diameter sixfold larger than the diameter of the pipe. This device provided the retention time of 9 min for aggregation of micro- and macrofloccule complexes. Water with flocks fed the tank with immersed membranes. During the filtration, water from the tank was sucked to the fibre using partial vacuum in the range 15–55 kPa. Filtration time was equal to 10 min. When the transmembrane pressure (partial vacuum) reached the level of approximately 55 kPa (after ca. 6–12 weeks of work), the chemical cleaning of the membrane with hydrochloric acid and/or citric acid was performed. While cleaning other equipment was disinfected with sodium hypochlorite. The total solids from the container were carried away by special sludge pump and stored. The results of experiments carried out in the pilot plant were used for creating, learning and verification of an artificial neural network.

3. ASSUMPTIONS OF ANN STRUCTURES FOR MODELLING OF THE TURBIDITY RETENTION COEFFICIENT

The approach of the artificial neural network was performed using the program Statistica 8.0. The aim of the work was to forecast, using ANN, the turbidity retention

coefficient after integrated process coagulation/ultrafiltration of reservoir water on the submerged membrane Zenon ZeeWeed 500c. Continuously (every 5 min) such parameters as transmembrane pressure (ΔP, kPa), temperature (T_t, K) and pH (pH_t) in the tank, real permeate flux in normal conditions 293.15 K (J, $m^3 \cdot m^{-2} \cdot d^{-1}$) and inlet turbidity of raw water (M_i, NTU) were measured. Turbidity in the tank (M_t, NTU) and permeate turbidity (M_p, NTU) were measured only pointwise (ten times during the whole experimental cycle). The coagulant dose (D) was constant and amounted to 0.5 g $Al^{3+} \cdot m^{-3}$. Variable D, being constant, was not included in the input signals. During the training of the network input parameters were changing in the range: 10.13–14.76 NTU for M_i; 22.60–46.10 NTU for M_t; 32.12–33.61 kPa for ΔP; 278.00––278.01 K for T_t; 7.01–7.02 for pH_t; 1.62–1.67 $m^3 \cdot m^{-2} \cdot d^{-1}$ for J. Experimental output signal (R_e – turbidity retention coefficient) varied between 0.9970 and 0.9977. During the prognosis, input parameters were changing in the range: 10.13–14.76 NTU for M_i; 22.60–46.10 NTU for M_t; 32.27–33.70 kPa for ΔP; 278.00–278.01 K for T_t; 7.02–7.05 for pH_t; 1.62–1.67 $m^3 \cdot m^{-2} \cdot d^{-1}$ for J. Experimental output signal (R_e) varied between 0.9970 and 0.9977. To obtain reliable results of forecasting, it was required to dispose the parameters with similar scopes of changes. It is worth pointing out that the practical significance of knowing the turbidity retention coefficient is really important because water turbidity allows us to classify the water quality and the quantity of natural organic matter which seems to be very crucial when technological processes of water treatment are planned.

Twenty neural network models have been created. Five models characterized by the smallest mean square error during the modelling were chosen as optimum. These models were used to make the prognosis. Then the best model (with the lowest error during the forecasting process) was chosen as most advantageous and useful in the future.

During the analysis of the network structure, there were investigated:

A. Choice of subsets. A random sampling method was used. From all 18 data used for network learning, 50% were chosen for training, 25% for testing and 25% for validation. In prognosis, the best created network used 100% of 12 values for testing.

B. Type of network. Only multilayer perceptrone (MLP I-H-O) with I number of input neurones, H number of hidden neurones and O number of output neurones was examined. It is the most popular network architecture used nowadays.

C. Minimum number of hidden neurons. To have reliable boundary conditions, the lowest number of the hidden neurons was established at the level of 2.

D. Maximum number of hidden neurons. To have reliable boundary conditions, the highest number of hidden neurons was established at the level of 30. More neurons lead to overfitting the network [14].

E. Activation functions in the hidden and output layer. Typical and simple activation functions were chosen: linear, logistic, hyperbolic tangent, exponential.

F. Methods of learning. For training the following algorithms were investigated: BFGS (quasi-Newton approach), fastest decline, conjugated gradient descent.

G. Number of learning epochs. It was reasonable to choose the range of training epochs in the range 1–65.

4. RESULTS AND DISCUSSION

20 models of neural networks were realized. The most important variables (listed in Table 1) i.e. learning errors and quality of learning, number of training epochs, activation functions, relative and absolute error, Pearson and determination coefficients were determined.

Table 1

Influence of used network architecture on the learning quality

Network type	Learning quality	Learning error	No. of training epochs	Activation function – neurones in a layer		Mean square error		Correlation coefficient	Determination coefficient
				hidden	output	absolute $\times 10^4$	relative $\times 10^{-2}$ [%]		
MLP6-10-1	0.6120	0.0000	1	log	log	2.63	2.64	0.6507	0.4234
MLP6-12-1	0.5941	0.0000	21	log	log	2.58	2.58	0.6394	0.4088
MLP6-30-1	0.9831	0.0000	9	lin	exp	2.01	2.01	0.7083	0.5017
MLP6-16-1	0.7903	0.0000	4	tan	log	1.78	1.78	0.8180	0.6691
MLP6-24-1	0.5781	0.0000	3	lin	exp	2.28	2.29	0.6116	0.3741
MLP6-27-1	0.9938	0.0000	21	tan	log	1.56	1.56	0.8496	0.7218
MLP6-12-1	**0.9916**	**0.0000**	**19**	**tan**	**exp**	**1.30**	**1.31**	**0.9216**	**0.8494**
MLP6-8-1	0.6690	0.0000	3	lin	tan	1.99	2.00	0.7183	0.5160
MLP6-15-1	0.7497	0.0000	4	lin	tan	1.70	1.71	0.8082	0.6531
MLP6-15-1	0.7861	0.0000	5	lin	tan	1.65	1.65	0.8317	0.6917
MLP6-28-1	0.7432	0.0000	4	lin	tan	1.94	1.95	0.7928	0.6285
MLP6-11-1	**0.9519**	**0.0000**	**19**	**exp**	**log**	**0.756**	**0.758**	**0.9645**	**0.9303**
MLP6-20-1	**0.9865**	**0.0000**	**33**	**exp**	**log**	**0.831**	**0.833**	**0.9682**	**0.9375**
MLP6-14-1	0.6595	0.0000	3	exp	tan	2.03	2.04	0.7116	0.5064
MLP6-20-1	0.5768	0.0000	11	log	log	2.61	2.61	0.6163	0.3798
MLP6-24-1	**0.9092**	**0.0000**	**60**	**log**	**log**	**1.20**	**1.20**	**0.9183**	**0.8432**
MLP6-14-1	0.5822	0.0000	5	lin	exp	2.35	2.35	0.6137	0.3766
MLP6-5-1	0.7741	0.0000	65	log	log	1.86	1.87	0.8136	0.6619
MLP6-2-1	**0.9114**	**0.0000**	**13**	**log**	**lin**	**1.24**	**1.24**	**0.9056**	**0.8201**
MLP6-22-1	0.7463	0.0000	4	tan	tan	1.85	1.85	0.7924	0.6278

Abbreviations: logistic – log, exponential – exp, tangential – tan, linear – lin.

Five models were chosen as the best ones. These optimum models (bold letters in Table 1) were determined by the lowest absolute mean-square errors (even very small

of the order of 10^{-5}) and the highest correlation coefficients in the range 90–-96%. As can be seen, only 13–60 epochs of learning were sufficient to obtain consistent results of modelling. These selected models were characterized by all possible (used during the training of the network) activation functions. None of the function predominates. This fact suggests that the problem is quite complicated and should be solved by the combination of various mathematical relationships.

Table 2

Results of forecasting of the turbidity retention coefficient

Model	Absolute mean square error	Relative mean square error [%]	R_p	Correlation coefficient	Determination coefficient
MLP 6-12-1	**2.16×10^{-4}**	**2.16×10^{-2}**	**0.9970–0.9978**	**0.8295**	**0.6881**
MLP 6-11-1	2.67×10^{-4}	2.68×10^{-2}	0.9970–0.9977	0.5189	0.2692
MLP 6-20-1	2.96×10^{-4}	2.97×10^{-2}	0.9970–0.9976	0.5035	0.2535
MLP 6-24-1	2.20×10^{-4}	2.20×10^{-2}	0.9970–0.9976	0.6419	0.4120
MLP 6-2-1	2.35×10^{-4}	2.35×10^{-2}	0.9967–0.9976	0.6388	0.4081

In Table 2, results of prognosis for the turbidity retention coefficient (R_p) are shown for 5 optimum network models (cf. Table 1). The optimum model (bold letters) characterized by the lowest absolute mean square error (2.16×10^{-4}) had two times more hidden neurons than numbers of inputs. This fact is directly connected with the solved problem. Probably such a great hidden vector was necessary just to train the network correctly and to generalize the matter of forecasting the turbidity retention coefficient. As can be seen, all five models were characterized by the relative error at the level of 2×10^{-2} but generally in optimizing problems we approach the best solution. That is why the selected network was described by the relative mean square error equal to 2.16×10^{-2}. The range of changes of forecasted retention coefficient (R_p) is almost the same as the experimental one (R_e), which suggests that the prognosis is very precise. It is obvious that in learning the error is lower than in forecasting because the prognosis was done on a different data set. These data were not known previously during forecasting and that is why the correlation is a little bit worse in comparison to the training. The structure of the optimum network model is presented in Fig. 2.

Absolute and relative errors have been compared in the prognosis step because only in this process it was reasonable to check if the network results were convergent with experimental values. In all network models, the method of training based on the quasi-Newton approach was used as the best in comparison to fastest decline and conjugated gradient descent algorithms. The method seems to be the most effective. Unfortunately, the number of data used for training (18) and also in forecasting (12) was small and thus the errors were not as small as expected and the correlation coefficient equalled only to 82.95% in the forecasting step.

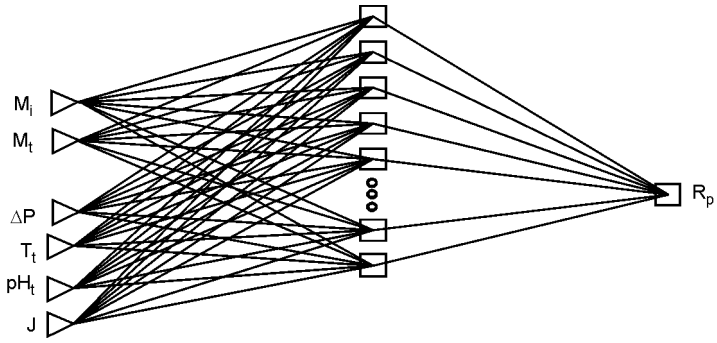

Fig. 2. Structure of the best network model

To obtain more reliable results, the network should be trained on the minimum one hundred data [15]. Only in such a case it could be possible to show the network great range of changes in the input signals. Despite usage of quite simple activation functions (hyperbolic tangent and exponential), the problem was solved with insufficient convergence (Fig. 3). Modelling using an artificial neural network has one disadvantage – the size of the training vector should be great enough to show the model the correct pattern of parameters and to obtain convergent results and small errors. Figure 3 shows only this quite great scatter of results because of an insufficient size of the learning vector. In the training process as well as in the forecasting step, the dispersion between experimental and predicted values of the retention coefficient is quite high. The conclusion drawn from Fig. 3 is obvious: the more experimental data, the greater the learning quality and the better results of forecasting are. Predicting the turbidity retention coefficient seems to be a crucial problem because knowledge about such a parameter allows us to design (having the information about the quantity of natural organic matter in water) the technological water treatment process properly. Probably the number of learning epochs should be higher than 19 to train properly in such cases when the number of variables is not sufficient. But in the case presented (the size of the learning vector is not significant) it is also a problem of overtraining: the network should be learnt to approximate but not to fit exactly the shown pattern just to predict in the future reliable results when the range of the input signals will change a little bit. The number of hidden neurons (12) seems to be sufficient because it is twofold more than the magnitude of input vector. It is not recommended to create the model which is too complicated in the structure. Nowadays the way of creating, learning and changing the model architecture is optimized. Hence the most important issue in ANN modelling is to receive the optimum (quite simple) model which could be adapted to only one problem. It is not possible to use the above created model to predict permeate turbidity after the ultrafiltration process without coagulation. It is a completely different optimizing problem due to changes in the operation conditions. The model is an assignment to only one kind of optimized question.

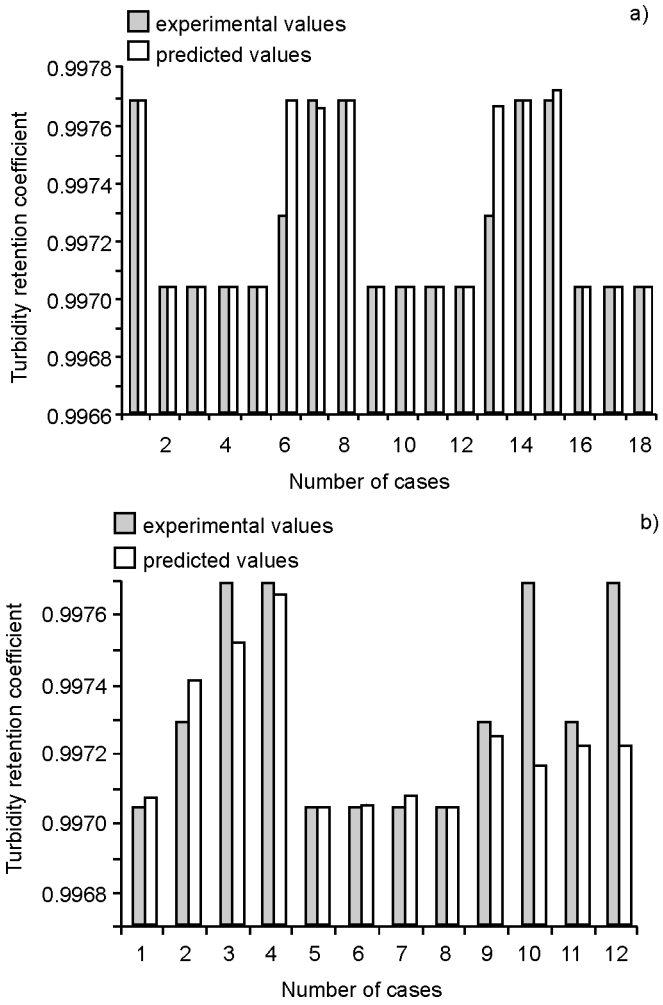

Fig. 3. Dispersion between experimental and predicted values during: a) learning, b) prognosis

It is clear that during the network learning the neurons were placed at various activation positions. Some neurons were activated more than others. The behaviour of the network is determined by changes in the activity of each single neuron. Negative values mean less active behaviour of the neuron. The higher was the value of the weight, the stronger influence and impact on the neuron. Very interesting is the fact that the highest value of weight (4.8748) was obtained between the transmembrane pressure and the first hidden neuron, while the lowest (–3.6701) was between pH in the tank and seventh hidden neuron. Probably this means that to forecast the turbidity retention coefficient data about the quality of raw water is not needed but rather the information

about the operation conditions (e.g. ΔP). As can be seen, pH has no influence (negative weight value) on satisfactory prediction of turbidity retention coefficient even when the ultrafiltration process is integrated with the coagulation.

5. CONCLUSIONS

The modelling of the turbidity retention coefficient using artificial neural networks might be summarized as follows:

- The optimum neural network model seems to be the great method of mathematical modelling due to lack of typical mathematical models solving problems of forecasting water quality after the treatment process.
- Absolute mean square error equal to 2.16×10^{-4} and the correlation coefficient at the level of 82% in the forecasting process can be acceptable from the engineering point of view, but the agreement between experimental and predicted values will be higher when the model is trained by the higher and more accurate variables.
- The most important role in predicting the turbidity retention coefficient is played by the transmembrane pressure because the value of the weight is in this case the highest.
- It should be remembered that ANN modelling is like a "black box" approach. Hence to achieve reliable solutions it is sometimes required to apply a "trial and error method" during creation of the network model.
- A created network could be used in the future for forecasting quality parameters of permeate in hybrid processes of reservoir water purification but only when the conditions of the membrane processes and the input signals are similar.

ACKNOWLEDGEMENTS

The authors express their gratitude to the Chair of Water Supply Engineering at the University of Technology in Dresden, especially to Andre Lerch, Susanne Müller and Professor Wolfgang Uhl for making available their experimental data for artificial neural network modelling.

SYMBOLS

ANN – artificial neural network
D – coagulant dose, g $Al^{3+}\cdot m^{-3}$
J – real permeate flux in normal conditions, $m^3\cdot m^{-2}\cdot d^{-1}$
H – number of hidden neurons
I – number of input neurons
M_i – inlet turbidity of raw water, NTU
MLP – multilayer perceptrone
M_p – permeate turbidity, NTU

M_t – turbidity in the tank, NTU
O – number of output neurons
ΔP – transmembrane pressure, Pa
pH_t – pH in the tank
R_e – experimental turbidity retention coefficient
R_p – predicted turbidity retention coefficient
T_t – temperature in the tank, K

REFERENCES

[1] KONIECZNY K., RAJCA M., BODZEK M., KWIECIŃSKA A., Environ. Prot. Eng., 2009, 35 (1), 5.
[2] KOWALSKA I., Environ. Prot. Eng., 2010, 36 (3), 33.
[3] MERDUN H., ÇINAR Ö., Environ. Prot. Eng., 2010, 36 (2), 95.
[4] KABSCH-KORBUTOWICZ M., KUTYŁOWSKA M., Environ. Prot. Eng., 2008, 34 (1), 15.
[5] RAZAVI S. M., MOUSAVI S. M., MORTAZAVI S. A., Chem. Eng. Sci., 2003, 58, 4185.
[6] RAI P., MAJUMDAR G. C., DASGUPTA S., DE S., J. Food Eng., 2005, 71, 273.
[7] GUADIX A., ZAPATA J., ALMECIJA C., GUADIX E., Desalination, 2010, 250, 1118.
[8] STRUGHOLTZ S., PANGLISCH S., GEBHARDT J., RIMBEL L., J. Water Supply: Res. Techn. – AQUA, 2008, 57, 23.
[9] CURCIO S., CALABRO V., IORIO G., Desalination, 2009, 236, 234.
[10] HWANG T., CHOI Y., NAM S., LEE S., OH H., HYUN K., CHOUNG Y., Des. Water Treat., 2010, 15, 134.
[11] AL-ABRI M., HILAL N., Chem. Eng. J., 2008, 141, 27.
[12] MÜLLER S., UHL W.J., Water Supply: Res. Techn. – AQUA, 2009, 58, 170.
[13] WÖLLNER A., *Untersuchungen zur Ultrafiltration von Uferfiltrat sowie deren Auswirkung auf eine nachfolgende Aktivkohlefiltration*, Master Thesis, Institute for Urban Water Management, Dresden University of Technology, 2004.
[14] KABSCH-KORBUTOWICZ M., KUTYŁOWSKA M., Ochr.Środ., 2010, 32 (2), 23.
[15] KABSCH-KORBUTOWICZ M., KUTYŁOWSKA M., Chem. Proc. Eng., 2009, 30, 307.