EnviroInfo 2013: Environmental Informatics and Renewable Energies
Copyright 2013 Shaker Verlag, Aachen, ISBN: 978-3-8440-1676-5

Development of mathematical models for forecasting hydraulic loads of water and wastewater networks[1]

Jan Studzinski[2], Lidia Bartkiewicz[3], Marcin Stachura[4]

In municipal waterworks the operation of water and wastewater networks decides about the functioning of the sewage treatment plant that is the last element of the whole water and sewage system. The both networks are connected each other and the work of the water net affects the operation of the wastewater one. The parameters which are important for right leading of all waterworks objects are their hydraulic loads that have to be not exceeded. Too large loads can cause accidents in the wastewater net or the tratment plant and an early knowledge of them is of importance for undertaking some counteractions. In the paper different algorithms to model hydraulic loads of municipal water and wastewater nets are described and compared regarding their computation velocity and accuracy. Some exemplary computations have been done with some real data received from a Polish water company.

1. Introduction

The main tasks of municipal waterworks are the production and distribution of drink water and its supplying to the water net end users as well as the sewage cleaning and derivation. In order to realize these tasks effectively some ICT tools can be used that will ensure optimal operation of all waterworks objects. Since a couple of years integrated IT systems for complex management of water and wastewater networks are under development at the Systems Research Institute (IBS PAN) (Studzinski, 2013, Sluzalec et all, 2012) in which the key programs are hydraulic models of the networks and mathematical models for forecasting their hydraulic loads. With the water net load forecasted the hydraulic load of the wastewater net can be predicted that in turn can help to develop predictive control algorithms for optimal operation of the sewage treatment plant. To model the hydraulic loads of water and wastewater networks time series algorithms, neuronal nets and fuzzy sets have been used and described in the following.

2. Preparation of data

The data to determine the hydraulic load models came from the waterworks in the Polish city Rzeszow (Bartkiewicz, 2000). While the water net modeling the daily water production values (WP) from the time period of 1.242 days and while the wastewater net modeling the values of the raw sewage inflow to the sewage treatment plant in Rzeszow (SI) from the same time period have been used. In the latter case the rainfalls data for Rzeszow (RF) and the values of the water level in the river passing the city (WL) have been either taken into account. While calculating the models for water and wastewater nets the one-day-forecasts of their loads have been determined. To evaluate the models the results of their simulation runs were used that have been done each time after the

[1] The paper describes the results of the research project No. N N519 6521 40 sponsored by the Polish National Center of Science (NCN)
[2] Systems Research Institute, Polish Academy of Sciences, E-mail studzins@ibspan.waw.pl
[3] Technical University Kielce, E-mail lidiab@tu.kielce.pl
[4] Warsaw University of Technology, E-mail m.stachura@mchtr.pw.edu.pl

modeling approach and by means of additional data series. In case of neuronal nets the initial data set was divided into three subsets for learning (L), validation (V) and testing (T) runs and for the simulation runs the later two subsets have been taken into consideration. In case of the time series and fuzzy sets modeling the initial data set was divided into two subsets for learning and testing runs and the testing runs were considered as the simulation ones.

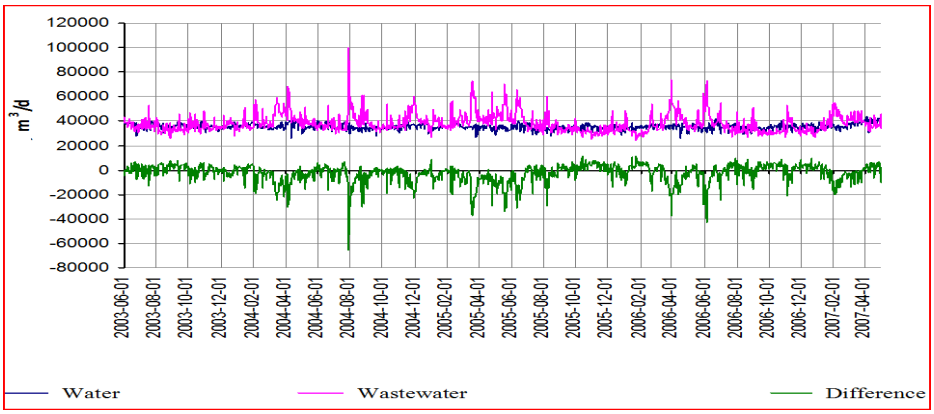

Figure 1. Water and wastewater production (WP and SI) by the water and wastewater networks in Rzeszow.

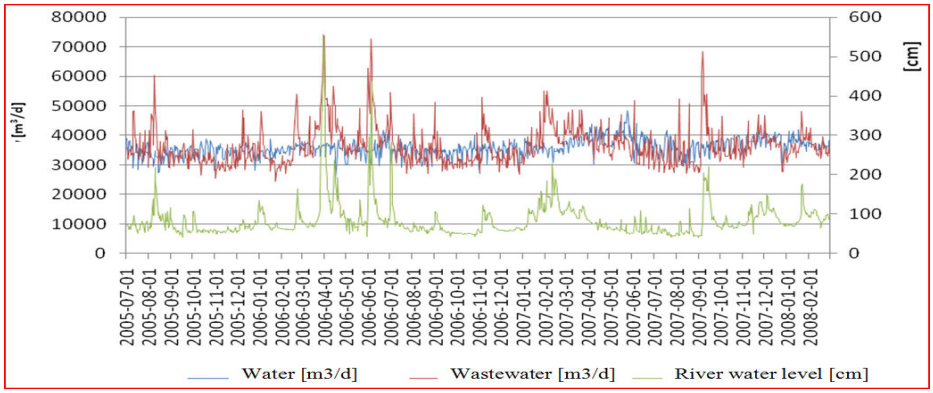

Figure 2. Data of water and wastewater production (WP and SI) and of the river water level (WL) for Rzeszow.

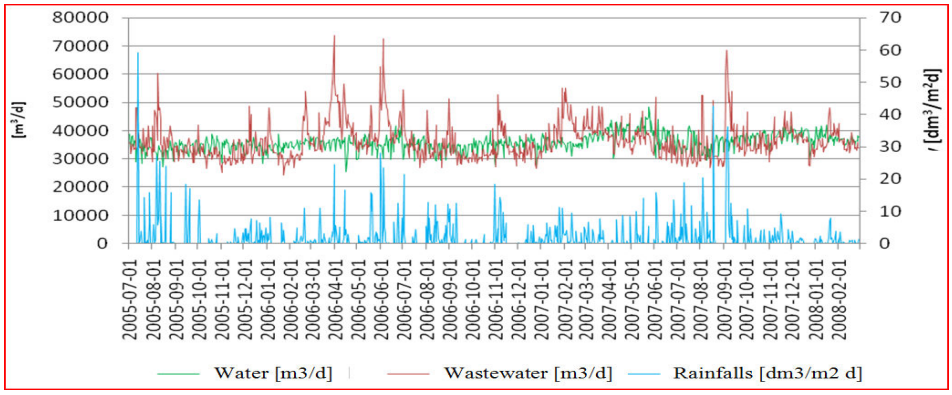

Figure 3. Data of water and wastewater production (WP and SI) and rainfalls data (RF) for Rzeszow.

3. Time series models

To model the hydraulic loads of the networks with the time series method the least squares algorithms of Kalman (K), of Clarke (Cl) and of the maximal likelihood (ML) with the following model description (Nahorski, Studzinski, 1988):

$$y_n = -A(z^{-1})y_n - \sum_{i=1}^{M} B(z^{-1})x_{in} + v_n$$

$$\hat{y}_n = -\hat{A}(z^{-1})y_n - \sum_{i=1}^{M} \hat{B}(z^{-1})x_{in}$$

have been used where $n = 1,2,...,N$, N – number of measurements data, M – number of model inputs, $A(z^{-1}), B(z^{-1})$ – difference operators for output y_n and inputs x_{in} of the process, $\hat{A}(z^{-1}), \hat{B}(z^{-1})$ – difference operators for the model signals, and v_n – correlated noise. These equations are used directly in the Kalman algorithm while in the algorithms of Clarke and of the maximal likelihood the noise v_n is additionally modeled with the following difference operators, respectively:

$$(1 + D(z^{-1}))v_n = \varepsilon_n$$

$$v_n = (1 + D(z^{-1}))\,\varepsilon_n$$

While calculating the time series models their evaluation has been done with the following synthetic criteria concerning the whole models:

$$MSE = \frac{1}{N}\sum_{1}^{N}(y_n - \hat{y}_n)^2$$

$$KOR = \frac{\sum_{n=1}^{N}(y_n - m_y)(\hat{y}_n - m_{\hat{y}})}{(\sum_{n=1}^{N}(y_n - m_y)^2)^{1/2}(\sum_{n=1}^{N}(\hat{y}_n - m_{\hat{y}})^2)^{1/2}}$$

$$RR = \frac{\sum_{n=1}^{N} (y_n - m_y)^2}{\sum_{n=1}^{N} (\hat{y}_n - m_{\hat{y}})^2}$$

where y_n – measurements data, $\hat{y}_n$ – model output, as well as with the use of standard deviations of their individual parameters σ_i. This makes possible to eliminate from the model operators the parameters whose standard deviations are essentially less than the parameter values.

Table 1. Best time series models of the water net hydraulic load.

Model	a_1/σ_1	a_2/σ_2	a_3/σ_3	a_4/σ_4	a_5/σ_5	a_6/σ_6	a_7/σ_7
K	-0,68/0,03	-0,15/0,03	-----	-0,09/0,03	-----	-----	-0,060,02
Cl	a_1/σ_1 d_1/σ_1	a_2/σ_2 d_2/σ_2	a_3/σ_3 d_3/σ_3	a_4/σ_4 d_4/σ_4	a_5/σ_5 d_5/σ_5	a_6/σ_6 d_6/σ_6	a_7/σ_7 d_7/σ_7
	-0,67/0,03	-0,15/0,03	-----	-0,08/0,03	-----	-----	-0,09/0,02
	-----	-----	-----	-----	-----	-----	0,09/0,03
ML	-0,69/0,02	-0,18/0,03	-----	-----	-----	- 0,10/0,02	-----
	-----	-----	-----	0,04/0,03	-----	-----	-----

In case of load modeling for the water net the autoregressive time series models with only one output being the water production (WP) and without any inputs, i.e. without the difference operators $A(z^{-1}), B(z^{-1})$, have been investigated. The results obtained for the water net load modeling are shown in Tables 1 and 2 and in Figure 4. As the best model turned out the Kalman model of seventh order with four parameters a_1, a_2, a_4 and a_7 for which MSE_L values for the learning runs and MSE_T values for testing runs are smallest and the respective correlation values *KOR* and the model deviation coefficients *RR* are biggest (Bartkiewicz, Studzinski, 2010).

Table 2. Evaluation results for the best time series models of the water net hydraulic load.

Model	K		Cl		ML	
	MSE_L/MSE_T	KOR_L/KOR_T	MSE_L/MSE_T	KOR_L/KOR_T	$MSE_L MSE_T$	KOR_L/KOR_T
	5322/4298	0,91/0,92	5326/4307	0,90/0,91	4480/4345	0,73/0,82

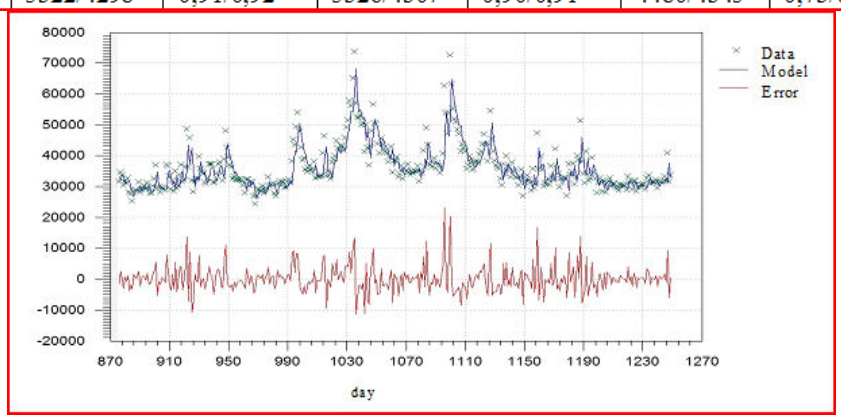

Figure 4. Modeling results for the best Kalman model of the water net hydraulic load.

Table 3. Evaluation results for the time series models of the wastewater net hydraulic load.

Model	$\sqrt{MSE_L}$	KOR_L
K/6/WP-WL-RF	100,1	0,87
Cl/6/WP-WL-RF	100,1	0,87
ML/3/WP-WL-RF	146,1	0,80
K/6/WP-WL	166,5	0,77
Cl/6/WP-WL	166,5	0,77
ML/3/WP-WL	189,6	0,74
K/6/WP-RF	110,2	0,85
Cl/6/WP-RF	110,2	0,85
ML/3/WP-RF	152,3	0,79
K/6/WP	182,5	0,74
K/6/RF	131,3	0,83
K/6/WL	187,8	0,75

In case of load modeling for the wastewater net the time series models with one output being the raw sewage inflow to the sewage treatment plant (SI) and with up to three inputs being the water production for the water net (WP), the rainfalls data for the city (RF) and the water level in the river flowing through the city (WL) have been calculated. The results obtained for the wastewater net load modeling are shown in Table 3 and in Figure 5 (an exemplary designation K/6/WP in Table 3 means the Kalman model of sixth order with 1 input WP). Also in this case as the best model turned out to be the Kalman model of sixth order and with three inputs considered (WP, WL, RF), for which the values of its statistical criteria are in general better than these ones of other models.

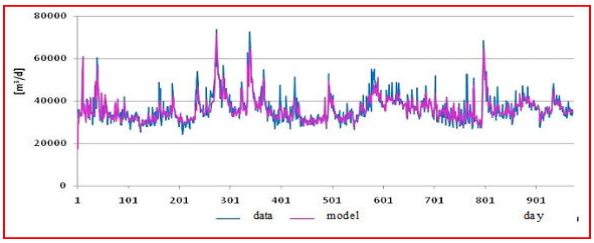

Figure. 5. Modeling results for the best Kalman model K/6/WP-WL-RF of the wastewater net hydraulic load.

4. Neuronal net models

The commonly most used neuronal nets are MLP networks which are one-directional nets with rearward error propagation (www.statsoft.pl). They have got usually a multilayer structure in which the signals pass in only one direction from the input layer to the output one and the input of each neuron on a hidden or output layer is connected with the outputs of all neurons on the preceding layer. In Fig. 6 a MLP neuronal net with 3 layers and with only 1 hidden layer is shown.

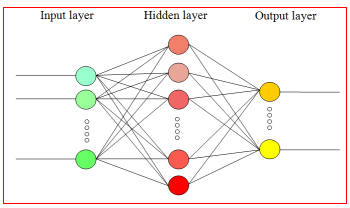

Figure. 6. MLP neuronal network with 3 layers.

At the calculations of the water net load with the MLP neuronal nets only 1 neuron was placed on the output layer being the daily water production (WP) in day *n* and on the input layer seven neurons were placed being the water production in the gradually previous days *n*-1, *n*-2,…, *n*-7 (Table 4). In this way in a neuronal model the measurements data from a whole week are simultaneously considered what corresponds to use difference operators of seventh order in an autoregressive time series model (Bogdan, Studzinski, 2010).

Table 4. Part of the file with input data for calculating MLP models.

y_n	y_{n-1}	y_{n-2}	y_{n-3}	y_{n-4}	y_{n-5}	y_{n-6}	y_{n-7}
37864	37934	38391	37336	38787	37756	37661	37748
37934	38391	37336	38787	37756	37661	37748	38365
38391	37336	38787	37756	37661	37748	38365	37589
37336	38787	37756	37661	37748	38365	37589	39682
38787	37756	37661	37748	38365	37589	39682	38360
37756	37661	37748	38365	37589	39682	38360	38805
37661	37748	38365	37589	39682	38360	38805	35670

The results of modeling with the neuronal nets of the water net load are shown in Table 5 and Figure 7 (an exemplary designation MLP/7/3/1 in Table 4 means the MLP model with 7, 3 and with 1 neuron on the input, hidden and output layer, respectively; MSE_V and KOR_V mean the values of the mean squared error and of the correlation received for the validation runs of MLP models). As the best model turned out the neuronal net with the numbers of neurons 7-10-1 on the subsequent layers. Evaluating the models with respect to the values of their statistical criteria one can see (Table 4) that all computed models are in principle similar. Moreover the model that has been taken as the best one depicts the measurements relatively well concerning its fitting to the measurement values as well as its matching with the measurement changes (see Figure 7).

Table 5. Neuronal net models of the water net hydraulic load.

Model	MSE_L	MSE_T	MSE_V	KOR_L	KOR_T	KOR_V
MLP/7/10/1	6208	6262	6063	0,58	0,63	0,59
MLP/7/7/1	6157	6311	6133	0,58	0,62	0,57
MLP/7/7/1	6244	6203	6117	0,57	0,63	0,58
MLP/7/4/1	6121	6259	6155	0,59	0,63	0,57
MLP/7/3/1	6441	6327	6284	0,56	0,63	0,56

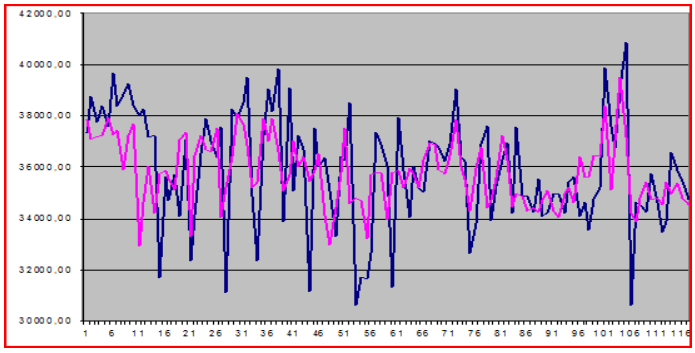

Figure 7. Modeling results for the best neuronal model MLP/7/10/1 of the water net hydraulic load (*measurements data in blue and modeling results in red color*)**.**

Table 6. The neuronal net models of the wastewater net load.

Model	KOR_L	KOR_T	KOR_V
1 input – WP			
MLP/1/1/1-2-1	0,10	0,15	0,17
1 input – WL			
MLP/2/1/2-3-1	0,67	0,45	0,61
1 input – RF			
MLP/3/1/3-3-1	0,49	0,59	0,53
2 inputs (without RF)			
MLP/1/2/2-3-1	0,76	0,73	0,78
2 inputs (without WL)			
MLP/2/2/4-3-1	0,51	0,62	0,55
2 inputs (without WP)			
MLP/2/2/4-4-1	0,68	0,46	0,64
3 inputs (WP and WL and RF)			
MLP/4/3/12-5-1	0,83	0,76	0,79

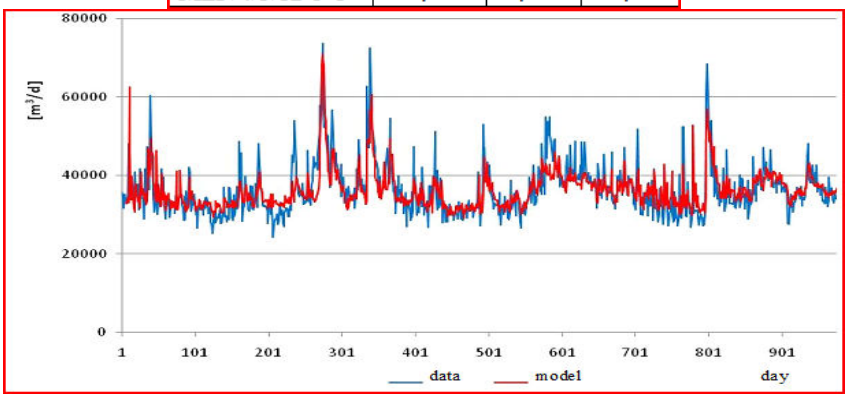

Figure 8. The best neuronal model MLP/4/3/12-5-1 of the wastewater net load.

The best results of modeling of the wastewater net load are presented in Table 6 and Figure 8 (an exemplary designation MLP/4/3/12-5-1 in Table 5 means the MLP model with the time shifting of

measurements in the input signals equal to 4 days, with 3 input signals WP, WL and RF and with the numbers of neurons 12-5-1 on the input, hidden and output layers of the network, respectively). There is to see there that model MLP/4/3/12-5-1 with three inputs turned out to be the best regarding the correlation values (*KOR*) for the learning, testing and validation runs. In this model the measurements in all input signals are shifted back of 4 days what corresponds to use difference operators of fourth order for each input in an autoregressive time series model.

5. Fuzzy set models

Inputs of a fuzzy sets model are signals that parameterize the model and they are not the physical signals which are influencing directly changes of the model outputs. For modeling the hydraulic loads of water and wastewater networks fuzzy set models of Takagi-Sugeno-Kanga structure are used in which the output element called 'successor' is a linear or nonlinear function of input signals (Takagi, Sugeno, 1985). The process of modeling with the fuzzy sets method consists in general of the following steps: definition of some affiliation functions for fuzzyfication of the input signals of the model investigated, input signals fuzzyfication with the functions defined, determination of some conclusion rules for the following concluding step and at the end of the process defuzzyfication approach follows meaning the model output computation (Fig. 9). At the beginning of the modeling the value ranges of the signals being the physical model inputs are divided into partition zones for which will be dedicated. The values range of these functions is between 0 and 1 and they are mostly in form of trapezoids or triangles (Fig. 10). Then the input signals are fuzzyficated and using the fuzzy functions resulted and the conclusion rules determined the output of the model is formed.

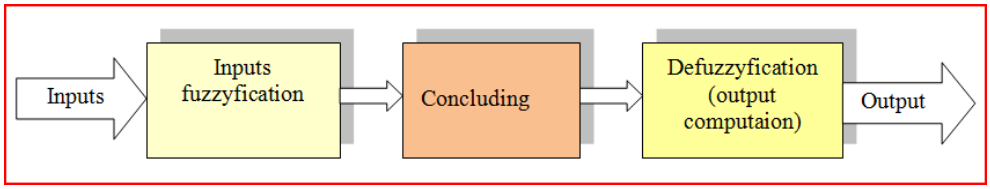

Figure 9. Diagram of modeling with TSK fuzzy set models [10].

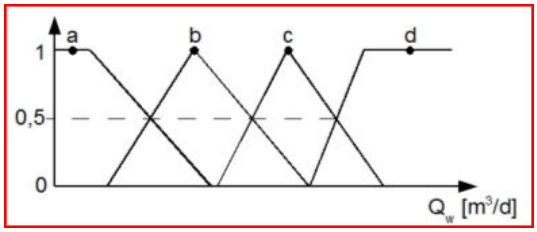

Figure 10. Affiliation functions defined for TSK models of the water net load.

In order to get a dynamic model of the process modeled a linear difference operator is applied in a TSK model as its output signal in which the transformed inputs are appearing as variables. A potential nonlinearity of the process modeled is emulated through the inputs fuzzyfication by means of nonlinear affiliation functions. Such model structure facilitates the model analysis and the comparison between the

model and the process modeled can be made relatively easy. TSK models with outputs in form of linear difference equations exert the following interesting features:

- thanks to considering the process dynamics in the successor element of TSK models they are in state to imitate the process modeled sufficiently good also by a small number of partition zones defined,
- the successor element of a TSK model is in form of a classic linear equation and this makes possible to treat a fuzzy sets model like a set of local linear models which are assigned one by one to the different partition zones,
- linear models are already known and researched very well and this facilitates making the comparisons between the main parameters of the models like delay times, time constants and model roots and the potential values of parameters of the modeled process, resulted from theoretical process analysis.

The evaluation of the model parameters included into the output difference equation occurs with a method of static optimization what is exactly the same operation as in the case of time series or neuronal models. While computing the forecasting models the data set used is divided into two subsets for making the modeling (learning) and simulation (testing) runs.

Table 7. The fuzzy sets model of the water net hydraulic load.

Model	*RR**100%	
TSK	Learning	Testing
	7,17 %	7,94 %

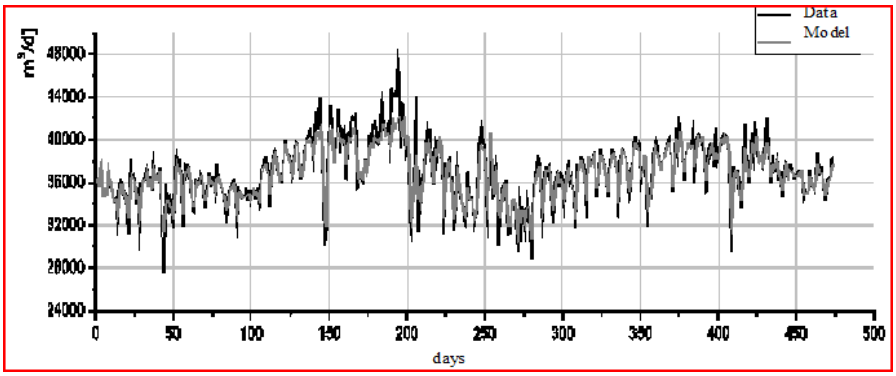

Figure 11. Testing results of the TSK model of the water net load.

The results obtained while modeling the water net load are shown in Table 7 and Fig. 11 (Studzinski, Stachura, 2010). The determined TSK model is in form of a set of autoregressive difference operators of seventh order and this result as a whole is similar to the time series model of Kalman. During the fuzzy sets modeling for each data zone a sub-model is calculated separately and the end model of the process is formed as a sum of all sub-models determined. The evaluation results of the model shown its usefulness for forecasting the water net hydraulic load with the forecast horizon of 1 day.

Table 8. The fuzzy sets models of the wastewater net load.

Model	$\sigma_{\hat{y}}$		$\sigma_{\hat{y}}/\sigma_y$	
	Learning	Testing	Learning	Testing
TSK/WP-WL-RF	4290	4034	0,671	0,631
TSK/WP-WL	4006	5253	0,626	0,821
TSK/WP-RF	4784	5827	0,748	0,911
Model	$\sigma_{\hat{y}}$ (mean value)	$\sigma_{\hat{y}}/\sigma_y$ (mean value)	*MSE* (mean value)	*KOR* (mean value)
TSK/WP-WL-RF	4162	0,65	177,7	0,60
TSK/WP-WL	4630	0,72	223,8	0,54
TSK/WP-RF	5306	0,83	303,5	0,48

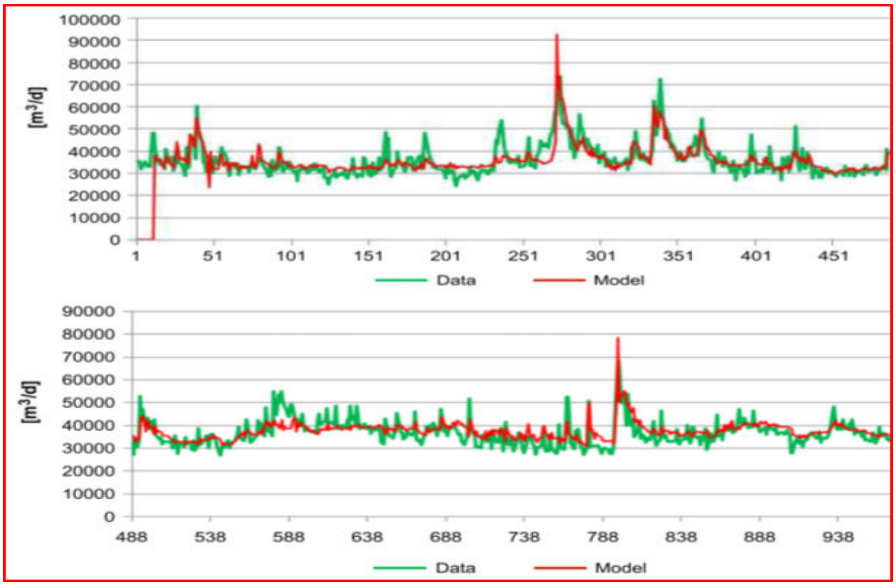

Figure 12. Calculation results of the best *fuzzy-sets*-Model TSK/WP-WL-RF for the learning (*on the top*) and testing data.

While modeling the wastewater net load only the models with two and three inputs have been investigated. The models with only 1 input have displayed a bad fitting to the data and the same conclusion has been formulated either basing on the calculation results got with the neuronal nets. The models determined have got an autoregressive structure what makes them structurally similar to the time series and neuronal models. The results of modeling are shown in Table 8 and Figure 12. There is to see from there that the models reproduce daily changes of the raw sewage inflow to the sewage treatment plant on principle well although the fitting of the models to the measurements data is generally worse as in case of the time series and neuronal models. As the best model turned out the model TSK/WP-WL-RF with three inputs whereas the water production WP with two trapezoid functions and the rain falls RF with three trapezoid functions were fuzzyficated. The third input (water level WL) as well as the model output (sewage inflow SI) have been not fuzzyficated. The best model obtained is in form of a difference equation with four difference operators of fourth order for all three inputs and for the output.

The results for the best models obtained while modeling the wastewater net load with different methods are shown in Table 9.

Table 9. Comparison of the best time series, neuronal and fuzzy sets models of the wastewater net load.

Model	$\sigma_{\hat{y}}/\sigma_y$	*KOR*
K/6/WP-WL-RF	0,47	0,87
MLP/4/3/12-5-1	0,63	0,83
TSK/WP-WL-RF	0,67	0,60

6. Additional calculations

After the MLP and TSK models of the water net load for 1-day-forecasts have been calculated also the neuronal and fuzzy set models with 30-days-forecasts were computed. The results received are shown in Figures 13, 14 and 15 and in Tables 10 and 11.

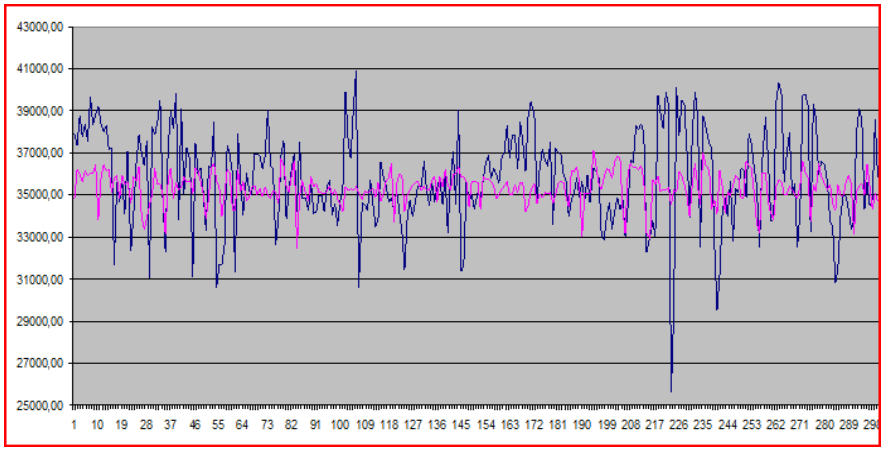

Figure 13. The best neuronal model MLP/7/12/1 of the water net load with 30-days-horizon of forecast; results for the learning run (*measurements data in blue and model results in red*).

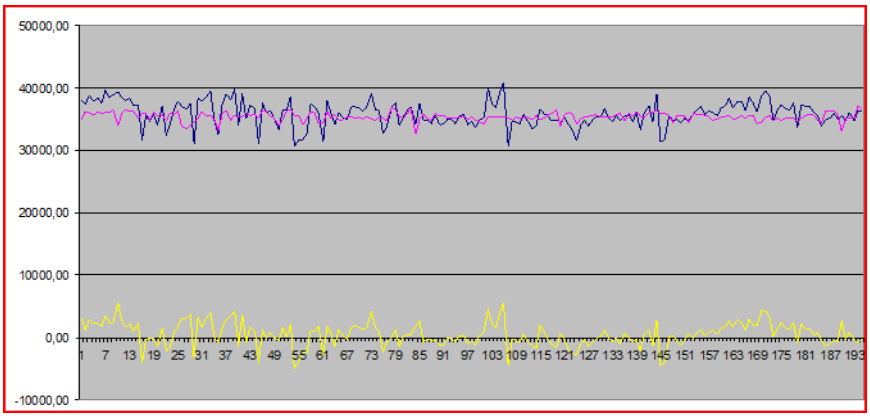

Figure 14. More detailed results of modeling with the MLP/7/12/1 model for 30-days-forecast (*error between measurements and model in yellow*).

Table 10. Some results of learning run of the MLP/7/12/1 model with the 30-days-forecast.

y_n	$\hat{y}_n$	$y_n - \hat{y}_n$	%
37934,00	34836,56	3097,44	8,17%
37336,00	36215,52	1120,48	3,00%
38787,00	36007,54	2779,46	7,17%
37756,00	35622,47	2133,53	5,65%
38365,00	36178,62	2186,38	5,70%
37589,00	35916,57	1672,43	4,45%
39682,00	36083,82	3598,18	9,07%

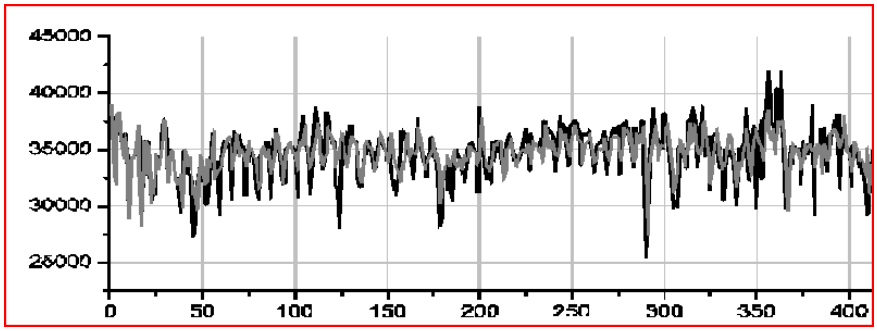

Figure 15. The best fuzzy sets model of the water net load with 30-days-horizon of forecast; results for the learning run (*measurements data in black and model results in gray*).

Table 11. Some results of learning run of the best fuzzy sets model with the 30-days-forecast.

y_n	$\hat{y}_n$	$y_n - \hat{y}_n$	%
36657	34095,71	2561,28	6,99
35947	36216,71	-269,71	-0,75
34726	36000,74	-1274,74	-3,67
36183	34352,33	1830,66	5,06
35282	35020,89	261,1	0,74
35199	34505,59	693,4	1,97
30891	33670,67	-2779,67	-9,00

There is to notice from the results computed that the neuronal nets are not good enough for modeling the water net loads with such the long forecast horizons as 30-day-periods and in the case of fuzzy set models the results received show that TSK-models could be a quite good tool for forecasting the water net loads with very different forecast horizons.

7. Conclusions

In the paper the results of water net load modeling by means of the time series, neuronal nets and fuzzy set methods are presented. There is to notice that the calculated 1-day-forecasts of the water net load are in general correct while applying each of the methods tested though the little better results are obtained with the time series method of Kalman. This is the simplest and fastest method under all methods used. An

interesting observation coming out from the calculations with neuronal nets is that all neuronal models computed are very similar each other what means that they are little sensitive regarding the choice of the number of neurons on the hidden layer of the net.

For modeling the wastewater net loads with 1-day-forecast horizon also three modeling methods have been applied. The results of modeling show that the simplest and fastest time series method of Kalman creates exacter models than other more complicated times series methods like Clarke's or maximum likelihood method. The Kalman method is in state also to determine better models than the much more complex methods of neuronal nets and fuzzy sets.

As to modeling the water net loads with the forecast horizon of 30 days and with the use of neuronal nets and fuzzy sets the following conclusions can be formulated:

The MLP models are not suitable for forecasting the water net loads with such the long forecast periods. The computed percentage errors between the model calculated and the measurements used are not big but the output signal of the best MLP model imitates in general only the moving average of measurements and it does not depict the right form of the original data.

The TSK models are more suitable for modeling the water net loads with longer forecast periods than the MLP models. Although the calculation errors of the fuzzy set models are not a lot smaller than by the MLP models but the TSK models show a quite good reproduction of measurements data in contrast to the MLP models.

Bibliography

Bartkiewicz, L. (2000) Modeling of the sewage inflow entering the sewage treatment plant. Dissertation Thesis, University of Technology Kielce, (*in Polish*).

Bartkiewicz, L., Studzinski, J. (2010) Mathematical modeling of the hydraulic load of communal wastewater networks. Conference Proceedings, ESM'2010, Hasselt.

Bogdan, L., Studzinski, J. (2010) Modeling and forecasting of the water net load using the neuronal nets. Conference Proceedings, BOS'2010, Bydgoszcz (*in Polish*).

http://www.statsoft.pl/textbook.

Nahorski, Z., Studzinski, J. (1988) IDOL software to identify dynamic linear objects. IBS PAN Report, Warsaw (*in Polish*).

Słuzalec, A., Studzinski, J., Ziolkowski, A. (2012) MOSKAN – eine WWW Anwendung in der HTML-5-Technologie zur Abbildung und Simulation von kommunalen Abwassernetzen". In: Simulation in Umwelt- und Geowissenschaften (J. Wittmann, B. Page, Hrsg) Workshop Hamburg 2012, Shaker Verlag, Aachen, 85-95.

Studzinski, J. (2013) IT system for computer aided management of communal water networks by means of GIS, SCADA, mathematical models and optimization algorithms. In: Proceedings of the ICT for Sustainability Conference: ICT4S, Zurich, February 14-16, 123-127.

Studzinski, J., Stachura, M. (2011) Fuzzysatz- und Neuronalmodelle zur Vorhersage der Wassernetzbelastung". Workshop on Simulation in den Umwelt- und Geowissenschaften, Berlin.

Takagi, T., Sugeno, M. (1985) Fuzzy identifications of system and its applications to modeling and control. IEEE Trans. on System Man and Cybernetic.