Metrol. Meas. Syst., Vol. XIX (2012), No. 1, pp. 123-132.

METROLOGY AND MEASUREMENT SYSTEMS

Index 330930, ISSN 0860-8229

www.metrology.pg.gda.pl

INFLUENCE OF THE GMDH NEURAL NETWORK DATA PREPARATION METHOD ON UTC(PL) CORRECTION PREDICTION RESULTS

Wiesław Miczulski, Łukasz Sobolewski

University of Zielona Góra, Faculty of Electrical Engineering, Computer Science and Telecommunications, Institute of Electrical Metrology, ul. Podgórna 50, 65-246 Zielona Góra, Poland (✉w.miczulski@ime.uz.zgora.pl, +48 68 328 2390*; l.sobolewski@weit.uz.zgora.pl,* +48 68 328 2390*)*

Abstract

The article presents results of the influence of the GMDH (Group Method of Data Handling) neural network input data preparation method on the results of predicting corrections for the Polish timescale UTC(PL). Prediction of corrections was carried out using two methods, time series analysis and regression. As appropriate to these methods, the input data was prepared based on two time series, ts1 and ts2. The implemented research concerned the designation of the prediction errors on certain days of the forecast and the influence of the quantity of data on the prediction error. The obtained results indicate that in the case of the GMDH neural network the best quality of forecasting for UTC(PL) can be obtained using the time-series analysis method. The prediction errors obtained did not exceed the value of ± 8 ns, which confirms the possibility of maintaining the Polish timescale at a high level of compliance with the UTC.

Keywords: GMDH neural network, national timescale, atomic clock, time series analysis.

1. Introduction

The Polish timescale UTC(PL) is a local realization of the Universal Coordinated Time (UTC). It forms the basis for propagation of the standard frequency and time signals in Poland. The UTC(PL) is implemented by the Central Office of Measures (Główny Urząd Miar, GUM) with a caesium-beam atomic clock Cs2 (internal acronym of one of 5071A type clocks working in GUM) and a control device (Microstepper Austron 2055). The control device allows corrections to be made which ensure the maximum compatibility of UTC(PL) with UTC (Fig. 1, station 1).

The quality of national timescales UTC(k), including UTC(PL), is assessed by the BIPM (Bureau International des Poids et Mesures). Each month, for the individual UTC(k), BIPM designate the corrections defining the divergence of timescales in relation to the UTC. These corrections are published by the BIPM in the "Circular T" bulletin about the 10th day of the next month. UTC(k) for countries whose corrections do not exceed the value of ±10 ns are the best group of timescales. Other groups of timescales are defined with correction values of ±20 ns and ±50 ns. The task of the GUM is to ensure that, when using a commercial atomic clock, the level of divergence of the UTC(PL) in relation to UTC is not greater than ±10 ns. Given the long delay of between 8 to 12 days in publishing the "Circular T" maintenance bulletin the best compatibility of the UTC(PL) with UTC can only be solved by predicting the corrections. Only a few national metrology laboratories are predicting the corrections for the UTC(k).

Article history: received on Sept. 16, 2011; accepted on Dec. 8, 2011; available online on March 12, 2012;
DOI: 10.2478/v10178-012-0011-1.

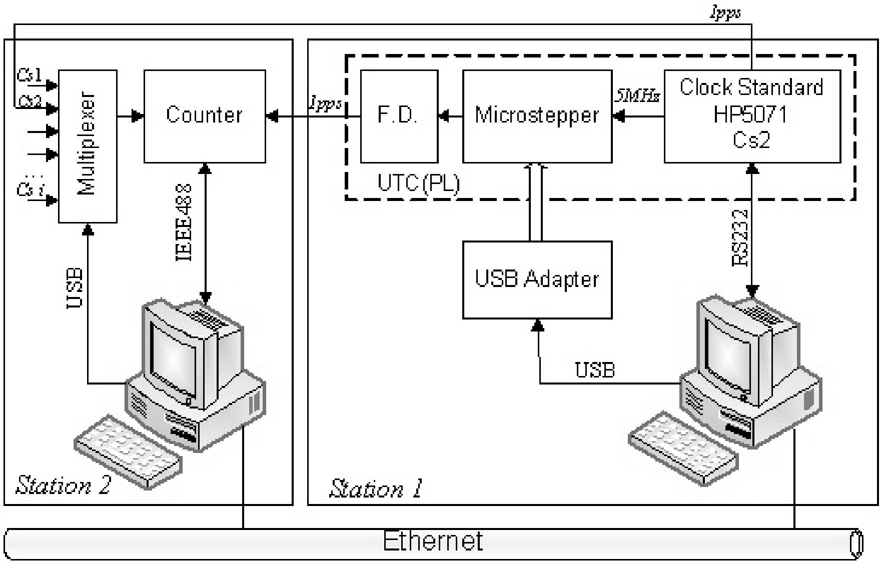

Fig. 1. Realization of the UTC(PL) scale and its comparison to the 1pps signal generated by the atomic clock.

The GUM uses the laborious procedure of predicting the corrections based on the analytical linear regression method [1]. It requires a lot of experience from the person who performs the analysis and information about the previous work of each clock available in the laboratory. When making predictions an arbitrary choice is necessary in choosing which of the predictions for the UTC(PL) can be considered as most reliable, and which should be rejected in a given month. The literature on the subject also gives methods for predicting corrections for the UTC(k) based on linear regression with stochastic differential equations [2] and using Allan deviations [3].

In collaboration with the GUM, the Institute of Electrical Metrology of the University of Zielona Góra has been running research on the application of neural networks in predicting the corrections for the UTC(PL). This type of solution, in relation to predicting the corrections for the UTC(k), is not known to the authors of the world literature. The thus-far-obtained positive results of the research on the application of the MLP and RBF networks [4] have confirmed the correctness of the selected neural model of prediction. A disadvantage of this type of neural networks is the long duration in obtaining a result for the prediction. This stems from the need to match the appropriate network structure and number of neurons to the nature of the data provided on its entry in the learning process. GMDH neural networks, which belong to the group of self-organizing networks [5], eliminate this problem. The aim of this study was to examine whether GMDH neural networks provide the level of divergence of the UTC(PL) in relation to UTC not greater than ±10 ns.

2. Data preparation for GMDH neural networks

2.1. General characteristics of the data

Predicting the corrections for the UTC(PL) based on neural networks requires a process of learning, the quality of which depends on the number of learning data inputs and the method of data preparation [6]. Data preparation for the GMDH neural network was based on the historical results of measurements of the phase time between UTC(PL) and a Cs2 clock (Fig. 1, station 2), defined for each day according to the relation

$$x_a(t) = \mathrm{UTC(PL)} - \mathrm{clock}\,. \tag{1}$$

The data was for the period from 53736 MJD (Modified Julian Date) (January 1st, 2006) to 54617 MJD (end of May 2008).

For the same period of time, the values of corrections for the UTC(PL) relative to UTC were available, with the relation

$$x_b(t) = \mathrm{UTC} - \mathrm{UTC(PL)}, \tag{2}$$

defined at the interval of five days and published in the „*Circular T”* bulletin.

Using the polynomial interpolation method for the data set from the BIPM, a mathematical model was determined, which permitted an extension of the learning data set by calculating the values of corrections for the UTC(PL) relative to UTC(PL) for each day of the analyzed period of time.

The final set of input data for the GMDH neural network was calculated from the relation

$$x(t) = x_a(t) + x_b(t) = \mathrm{UTC} - \mathrm{clock}\,. \tag{3}$$

2.2. Input data with and without trend elimination

The resulting set of data $x(t)$ is the time series (ts1), which characterizes the time instability of the Cs2 clock for each day with reference to the UTC. In time series ts1 there occurs a linear trend component $x_r(t)$ and a variable component (Fig. 2). Time series ts1 was the first set of data for which the process of GMDH neural network learning and prediction of correction values for the UTC(PL) was performed.

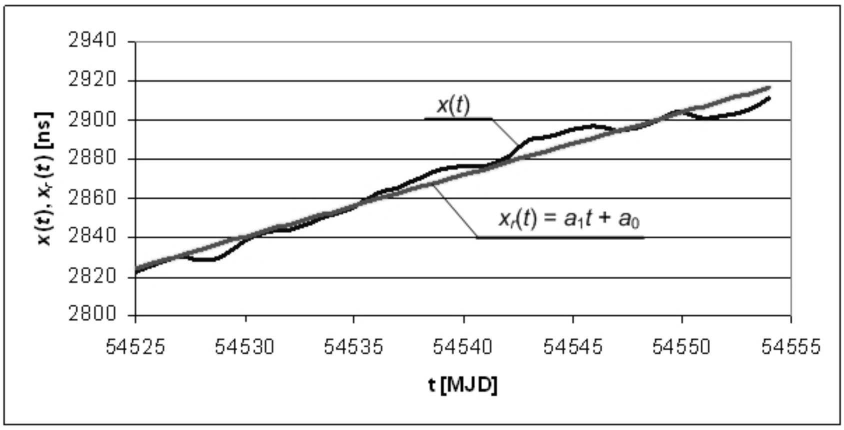

Fig. 2. Sample set of phase time $x(t)$ and trend $x_r(t)$ for a one month period.

For the GMDH neural network learning process and prediction of corrections for the UTC(PL) the time series ts2 was prepared. Its final form was obtained via elimination of the long-term trend of the phase-time changes $x_r(t)$ described with the linear regression equation (Fig. 2). The various components of time series ts2, formed by the values of the deviations from the trend $xd(t)$, were calculated from the relation

$$xd(t) = x(t) - x_r(t)\,. \tag{4}$$

Such an approach was dictated by the small values of deviations from trend $xd(t)$, which achieve values up to ±9 ns (constituting about 0.5% of $x_r(t)$). Such a situation may cause that the neural network will adopt the trend as an important piece of information in the learning process, and this may affect the deterioration of the prediction results of the UTC(PL).

However, inputting data represented by time series ts2 to the GMDH neural network will cause the network learning process to adopt a model that describes the deviation from the trend. This method of preparation of input data for neural networks is recommended in the literature [5]. It may lead to better prediction of results for the UTC(PL).

2.3. Data input sets for determining prediction based on time series analysis and regression

Prediction using a GMDH neural network was carried out based on time series analysis and the regression method. In the first case, the input used was a data vector which was represented by time series ts1 or ts2 written in the form of a k-elements vector. It was assumed that the predictions for the UTC(PL) were determined for the next five months of 2008 (from January to May). For this period a Cs2 clock compared with UTC showed the greatest deviations from the trend, which could affect the values of predicted corrections. The sample set of input data for predicting the corrections for the month of January (Table 1) contained a single vector of 730 elements.

Table 1. The sample of input data for the GMDH neural network to predict the corrections based on time series analysis for time series ts1 and ts2, for 54479 MJD.

MJD	$x(t)$ for ts1 [ns]
53736	663.021
53737	664.269
53738	664.992
53739	665.726
53740	667.304
53741	669.971
53742	673.638
53743	677.093
53744	679.485
⋮	⋮
54464	2632.183

MJD	$xd(t)$ for ts2 [ns]
53736	57.267
53737	55.958
53738	54.124
53739	52.301
53740	51.322
53741	51.432
53742	52.542
53743	53.440
53744	53.274
⋮	⋮
54464	164.837

In the second case, the input to the GMDH neural network were vectors containing successive data from time series ts1 or ts2 for the period of 30 days (t_0 – 29 ÷ t_0) and a prediction value for the 15th day (t_0 + 15). In case of time series ts2 regression trend component $x_r(t)$, coefficients (a_0 and a_1) were also added for a period of 30 days (Table 2).

The sample set of input data (Table 2), obtained on the basis of time series ts2 consisted of 685 vectors, containing 33 elements each.

Due to the way of preparing data for both methods of prediction there is a difference in the number of data references available for the GMDH neural network learning process. In predicting with the regression method, the set of input data ends at 15 days before the end of the month preceding the delivery of the forecast. However, in the case of time series analysis the set of input data ends on the last day of the month preceding the forecast, so it is longer by 15 days compared to the set of input data for the regression method.

Table 2. The sample set of input data for the GMDH neural network to predict the corrections using regression method for 54479 MJD, time series ts2.

MJD for t_0+15	$xd(t_0-29)$ [ns]	$xd(t_0-28)$ [ns]	...	$xd(t_0-1)$ [ns]	$xd(t_0)$ [ns]	a_1	a_0	$xd(t_0+15)$ [ns]
53780	3.259	1.997	...	-4.603	-4.123	2.51127	-134286	-1.511
53781	2.263	0.470	...	-3.981	-2.658	2.51572	-134525	-2.528
53782	0.698	-1.092	...	-2.646	-1.219	2.52343	-134940	-3.773
53783	-0.970	-1.923	...	-1.318	0.371	2.53133	-135365	-5.083
53784	-1.980	-1.8456	...	0.285	1.113	2.53239	-135422	-6.166
53785	-2.018	-0.880	...	1.049	1.451	2.52849	-135212	-7.182
53786	-1.041	-0.112	...	1.358	1.688	2.52604	-135080	-8.192
53787	-0.166	-0.303	...	1.580	1.303	2.52798	-135184	-9.167
53788	-0.172	-1.048	...	1.007	2.230	2.54322	-136004	-10.297
⋮	⋮	⋮	...	⋮	⋮	⋮	⋮	⋮
54464	-1.223	-2.477	...	-1.454	-0.045	3.38277	-181614	7.139

3. GMDH neural networks

GMDH neural networks were used for prediction of corrections. The networks use the Group Method of Data Handling. The networks apply the feedforward learning technique. In the course of the learning process, the network expands and evolves until enhanced efficiency of its operation has been achieved [6]. Until a new layer of neurons is attached to the current structure of the network, elements of the new layer are subjected to selection for accuracy of processing. Neurons that do not meet the criterion for assessing the condition imposed, i.e. the processing error is too large, are eliminated from the structure of the network (in Fig. 3 these neurons are marked with a bright colour).

The relationship between groups of input and output signals in the GMDH neural networks describes the transition function, which usually takes the form of a polynomial.

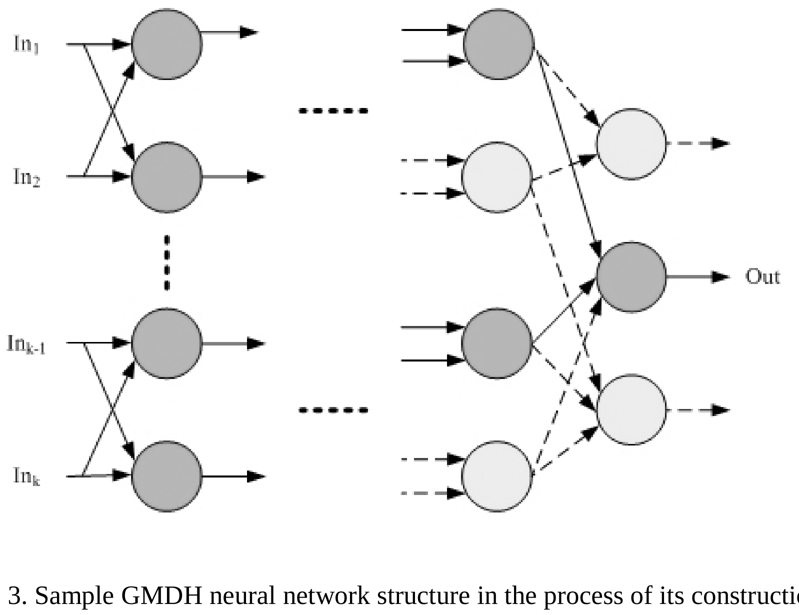

Fig. 3. Sample GMDH neural network structure in the process of its construction.

Automatic adjustment of the structure and number of neurons in the GMDH neural network to the nature of the data supplied at its input in the learning process, in contrast to the MLP and RBF neural networks, decided the selection of this type of network for testing prediction values for the UTC(PL). The process of learning in the GMDH neural network and prediction of corrections was carried out using the commercial tool GMDH Shell 1.7.

4. Research findings

4.1. Method of carrying out research

The use of the GMDH neural network for prediction was based on two methods: time series analysis and regression. The input of the GMDH neural network were vectors prepared from the data of time series ts1 and ts2, as described in subsection 2.3. The output generated from the neural network was prediction per day $(t_0 + m)$, which allows the calculation of the correction value for the UTC(PL) (t_0 – the last day of the month preceding the forecast). The predicting process designated the prediction value for the time series ts1 $x_{pred}(t_0 + m) = (\mathrm{UTC} - \mathrm{clock})_{pred}$, and on this basis in the next step the correction value $(\mathrm{UTC(PL)} - \mathrm{UTC})_{pred}$ was calculated. This prediction represents the value of correction needed to correct the UTC(PL) in order to ensure the best compatibility of UTC(PL) with UTC.

The first step for time series ts2 was to designate the prediction value of the deviation from the trend $xd_{pred}(t_0 + m) = (x(t_0 + m) - x_r(t_0 + m))_{pred}$, which was next added to the prediction value of the trend calculated using a regression equation. The result was the sought-after value of prediction $(\mathrm{UTC} - \mathrm{clock})_{pred}$. Similarly, further calculations aimed at achieving the prediction correction value were carried out on the time series ts1.

In order to compare the values of predicted corrections for the UTC(PL) obtained for time series ts1 and ts2 (GMDH neural networks) and the predictions made by GUM using the analytical linear regression method, the prediction error (Δ_{pred}) was designated according to equation (5). It defines the difference between the predicted value $(\mathrm{UTC(PL)} - \mathrm{UTC})_{pred}$ and the value $(\mathrm{UTC(PL)} - \mathrm{UTC})_{\mathrm{CIRT}}$ read from the „*Circular T*" bulletin for the same day of forecast.

$$\Delta_{pred} = (\mathrm{UTC(PL)} - \mathrm{UTC})_{pred} - (\mathrm{UTC(PL)} - \mathrm{UTC})_{\mathrm{CIRT}}. \tag{5}$$

4.2. Predicting the UTC(PL) *corrections for the fifteenth day of the month*

First of all, research was carried out in relation to predicting the corrections for the UTC(PL) on the 15$^{\text{th}}$ day $(m = 15)$ for each of five consecutive months. The results of the obtained prediction errors are shown in Fig. 4.

In the case of determining the prediction of corrections for the UTC(PL) using time series analysis (Fig. 4a), it can be noted that the preparation of both time series (ts1 and ts2) to a small extent affects the quality of prediction. For time series ts1 the obtained prediction errors in the analyzed period of five months does not exceed the value of ±7 ns, and for time series ts2, ±8 ns. Prediction errors obtained for the same period by GUM, using the analytical linear regression method to predict the corrections for the UTC(PL), exceed the value of ±12 ns. In the case of the regression method (Fig. 4b) the obtained prediction errors for time series ts1 are in the range of ±8 ns. For time series ts2 and predictions made by GUM, prediction errors exceed the value of ±12 ns.

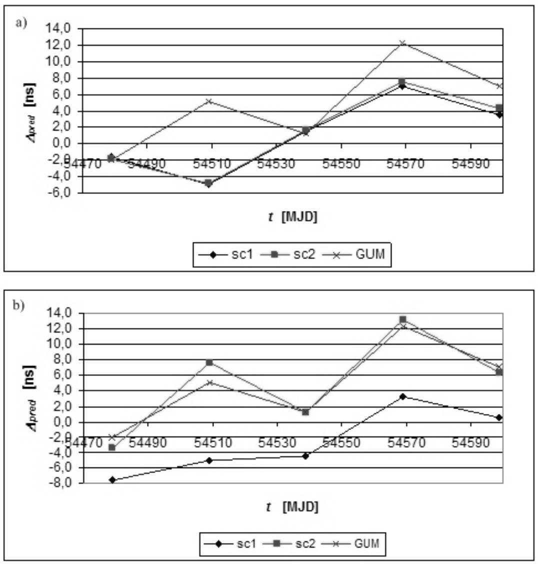

Fig. 4. Prediction errors on the 15th day of five consecutive months using: a) time series analysis method, b) regression method.

4.3. The influence of the number of elements of a time series on prediction error

The purpose of implementing subsequent research on the influence of the number of time series elements on the prediction error was to evaluate the influence of learning data vector length (k), input into the GMDH neural network, on prediction results of time series analysis. The length of learning vector (k) was changed from the largest possible value for a given prediction day (for example: $k = 720$ as a prediction for 54479 MJD, $k = 840$ as a prediction for 54599 MJD) to $k = 30$.

Fig. 5a shows the research results for time series ts1, which show that for $k \geq 570$ prediction errors are within ±10 ns. In the same way, research was conducted for time series ts2 (Fig. 5b). The results show that for $k \geq 420$ prediction errors are also within ±10 ns.

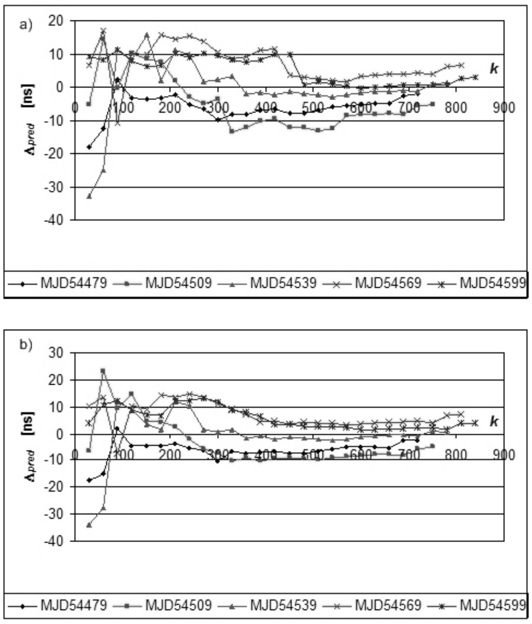

Fig. 5. Influence of learning data vector length on the result of prediction of corrections for the UTC(PL) using time series analysis method: a) ts1, b) ts2.

In the case of predicting the corrections for the UTC(PL) using the regression method, the number of data given to the input of the GMDH neural network in the learning process must be much larger, to ensure that the prediction errors will not exceed the value of ±10 ns (Fig. 6). This condition is fulfilled only for the data prepared on the basis of time series ts1. In this case the number of elements (k) of time series ts1 may not be less than 685 elements (Fig. 6a). Having (k) number of elements for time series ts2 (Fig. 6b and 6c) did not allow the determination of the minimum value of k, for which the prediction errors would not exceed the value of ±10 ns.

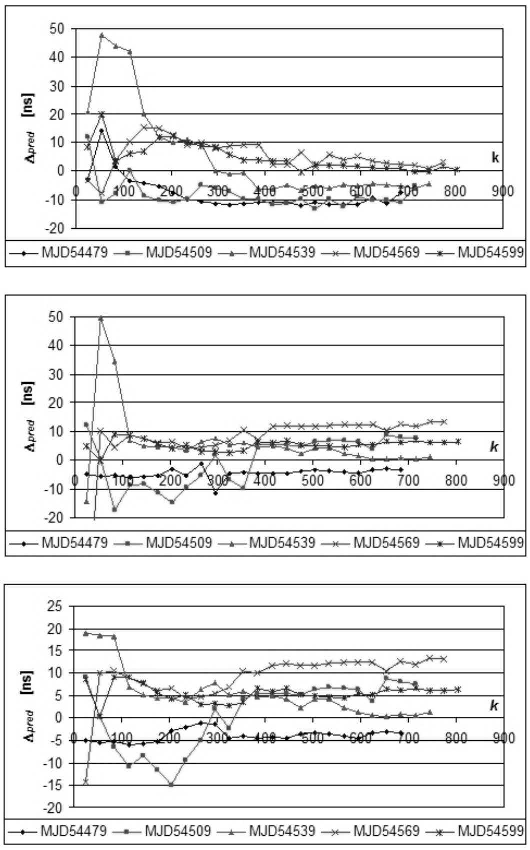

Fig. 6. Influence of learning data vector length on the result of prediction of corrections for the UTC(PL) using the regression method for time series: a) ts1, b) ts2 with regression coefficients a_1 i a_0, c) ts2 with regression coefficient a_1.

The research shows that the atomic clock can be used to implement the UTC(PL) after a certain period of time from its launch. When using time series analysis to predict the corrections for the UTC(PL) this time may not be less than 19 months for time series ts1 and 14 months for time series ts2. When using the regression method to predict the corrections for the UTC(PL) this time cannot be shorter than 23 months (for time series ts1). In the case of GUM, which has currently three atomic clocks, the indicated lengths of time do not restrict the use of the GMDH neural networks to predict the corrections. An atomic clock can be used for the implementation of UTC(PL) after about two years from the commencement of its work.

4.4. Predicting the corrections for UTC(PL) *for another day of the month*

Due to delays in the publication of corrections for the UTC(PL) for the end-of-the-month period (about 8 to 12 days) appearing in the bulletin „*Circular T*", the first prediction can be determined only on the 15th day of the following month.

Subsequent studies using the GMDH neural network were carried out setting prediction of corrections for the UTC(PL*)* on selected days (m) occurring between successive dates of publication of the „*Circular T*" bulletin. Prediction of the corrections was made using only the time series analysis, which at the current stage of research can achieve smaller prediction errors than prediction using the regression method. Table 3 presents the prediction errors for the selected day of prediction $(m = 15,\ 20,\ 25,\ 30,\ 35\ \text{and}\ 40)$ for five consecutive months.

Table 3. Influence of m day of prediction on prediction error correction to the UTC(PL).

Day of prediction $m =$		15	20	25	30	35	40
		Δ_{pred} [ns]					
First prognosis on 15th day, that is MJD=54479	ts1	-1.6	-12.3	-21.6	-28.6	-26.2	-32.1
	ts2	-1.9	-13.0	-22.7	-30.6	-28.6	-34.9
	GUM	-2.0	-13.2	-21.6	-28.4	-27.0	-32.7
First prognosis on 15th day, that is MJD=54509	ts1	-4.9	-6.8	-7.9	4.7	7.1	7.1
	ts2	-4.8	-5.9	-7.0	5.4	7.2	7.6
	GUM	5.1	9.2	11.8	27.8	33.0	37.1
First prognosis on 15th day, that is MJD=54539	ts1	1.5	0.3	-7.1	-16.1	-17.6	-18.5
	ts2	1.6	33.2	20.7	25.0	26.1	-6.4
	GUM	1.2	1.4	-5.1	-13.6	-15.2	-15.0
First prognosis on 15th day, that is MJD=54569	ts1	7.0	-2.4	-4.8	-9.2	-7.5	-6.7
	ts2	7.5	-4.6	-7.1	-12.4	-11.4	-10.8
	GUM	12.3	6.5	6.6	4.3	7.6	10.1
First prognosis on 15th day, that is MJD=54599	ts1	3.5	-7.0	-6.4	-3.8	-9.0	-17.3
	ts2	4.4	-8.8	-8.7	-5.8	-11.1	-18.8
	GUM	7.0	-1.2	1.3	5.2	2.1	-4.2

This research shows that the GMDH neural networks for the input data based on time series ts1 behave similarly to prediction using the analytical linear regression method used so far by the GUM. Significantly larger prediction errors were obtained in the case of predicting for the input data based on time series ts2.

5. Conclusions

The study has confirmed that GMDH neural networks can be used to predict the corrections for the UTC(PL) in order to ensure compliance with UTC at a level not exceeding ±10 ns. This state is feasible if:

- time series analysis is used for prediction,
- input data for the GMDH neural network is prepared on the basis of relation (3), creating time series ts1, and with the number of elements greater than 570.

An important advantage of the GMDH neural networks is the fact that these networks are self-organizing neural networks. Their automatic adjustment of the structure and number of neurons to the changing nature of the time series supplied to its input in the learning process enables the implementation of the results of the predicted corrections in the UTC(PL) in a short time (a few seconds). In the case of predicting corrections using MLP and RBF neural networks, the person who performs the learning process of the network selects the structure and quantity of the neurons in the hidden layer, which greatly extends the time to complete the final result of prediction of corrections (up to several hours). GMDH neural networks are also much faster and produce much clearer results of predicted corrections for the UTC(PL) than the prediction used so far by GUM using the analytical linear regression method.

However, at this stage the research carried out has not confirmed unequivocally the possibility of applying the regression method to predict the correction for the UTC(PL) using GMDH neural networks. Prediction of corrections for the 15th day with input data for the GMDH neural networks based on time series ts1 allows the calculation of results of prediction of correction that ensure the compliance of the UTC(PL) with UTC at a level not exceeding ±10 ns. However, the number of elements of the time series that were available while doing the research was too small to fully confirm the usefulness of the regression method in predicting the corrections for the UTC(PL). This requires further research with a larger number of elements in the time series and the development of another form of transition function for the neuron.

The research on how to prepare two time series (ts1 and ts2), presented in subsection 2.2, has shown that in using the polynomial transition function for the neuron better results were obtained for time series ts1 (without trend elimination). For the same set of input data for MLP and RBF neural networks, the time series ts2 was favorable (with trend elimination) [4].

References

[1] Czubla, A., Konopka, J., Nawrocki, J. (2006). Realization of atomic SI second definition in context UTC(PL) and TA(PL). *Metrology and Measurement Systems*, 2, 149-159.

[2] Panfilo, G., Tavella, P. (2008). Atomic clock prediction based on stochastic differential equations. *Metrology*, 45, 108-116.

[3] Bernier, L.G. (2003). Use of the Allan Deviation and Linear Prediction for the Determination of the Uncertainty on Time Calibrations Against Predicted Timescales. *IEEE Transactions on Instrumentation and Measurement*, 52(2), 483-486.

[4] Miczulski, W., Cepowski, M. (2010). Influence of type of neural network and selection of data pre-processing method on UTC-UTC(PL) prediction result. *The Measurements, Automation and Monitoring*, 11, 1330-1332. (in Polish)

[5] Farlow, S.J. (1984). *Self-organizing Methods in Modelling: GMDH-type Algorithms*, 54, New York: Marcel Dekker Inc.

[6] Masters, T. (1993). *Practical neural networks recipes in C++*. Academic Press, Inc.