Methods of time series preparation based on UTC and UTCr scales for predicting the [UTC-UTC(PL)]

L Sobolewski¹, W Miczulski¹

¹ Institute of Metrology, Electronics and Computer Science, University of Zielona Gora, prof. Z. Szafrana 2, 65-246 Zielona Gora, Poland

E-mail: l.sobolewski@imei.uz.zgora.pl

Abstract. The article presents the results of methods of preparation of two time series on the quality of predicting the [UTC-UTC(PL)] for the Polish Timescale UTC(PL) using GMDH neural networks. The first time series (TS1) was based on the [UTC-UTC(PL)] deviations designated by the BIPM. In the second time series (TS2) the deviations designated by the BIPM on the basis of the UTC and UTC Rapid scales were applied. The obtained results indicate that the use of time series for predicting the [UTC-UTC(PL)], based on deviations determined by the UTC and UTC Rapid scales, allowed to obtain more accurate predictions.

1. Introduction

The organization responsible for determining the quality of national timescales UTC(k) is the International Bureau of Weights and Measures (BIPM, Bureau International des Poids et Mesures). Every month for each UTC(k) the \( xb(t) \) deviations are determined by the BIPM according to relation

\[
xb(t) = UTC(t) - UTC_{k}(t),
\]

(1)

defining the differences of the national timescale relative to UTC. These deviations are calculated as averages per day on days MJD (Modified Julian Date) ends with the digits 4 and 9. The process of calculating the UTC scale, based on specific algorithms [1], is very complex and time-consuming. The measurement data for the previous month are sent to the BIPM from the National Metrology Institutes NMIs. The assembly, proper preparation of measurement data and calculation of the UTC scale by the BIPM, cause a delay in publishing by the BIPM "Circular T" bulletin containing calculated deviations of UTC(k) in relation to UTC for a given month. This bulletin is published between the 8th and 12th day of the following month. Considering the delay in publication of the "Circular T" bulletin, ensuring the highest possible compatibility of UTC(k) in relation to the UTC is only possible by predicting the [UTC-UTC(k)]. The designated value of the prediction would be the basis for correcting the UTC(k) scale. Part of the NMIs laboratories are predicting the [UTC-UTC(k)]. In the literature for predicting the [UTC-UTC(k)] only methods that are based on statistical models are presented, for example [2, 3, 4, 5].

Polish Timescale UTC(PL) is the local implementation of Coordinated Universal Time (UTC). Central Office of Measures (GUM, Główny Urząd Miar) is responsible for maintaining the UTC(PL). The physical realization of the UTC(PL) is based on one of the selected caesium atomic clocks available in the GUM. The system implementing the UTC(PL) is equipped with a control device (Phase-Microstepper Austron 2055), enabling the introduction of appropriate adjustments to the time
scale. For predicting the [UTC-UTC(PL)], to 13.12.2012, a laborious procedure based on analytical linear regression was used in the GUM.

The Institute of Metrology, Electronics and Computer Science of the University of Zielona Gora, in cooperation with the GUM, began a research on the application of neural networks for predicting the [UTC-UTC(PL)] in 2006. This choice was due to the properties of neural networks [6]. For predicting the GMDH (Group Method of Data Handling), GRNN (Generalized Regression Neural Network), MLP (Multilayer Perceptron) and RBF (Radial Basis Function) neural networks were used. The input data for these neural networks were prepared in the form of time series as described in section 2.1. The results of the studies, presented in [7, 8 and 9], shows that the best results were achieved for GMDH neural network using time series analysis method.

Delay in publication of the \( x_b(t) \) deviations by the BIPM adversely affect on the result of the prediction, and consequently on maintaining the best convergence of the UTC(k) with the UTC. Therefore, in order to expedite the transfer of information about the differences of the UTC(k) in relation to the UTC, BIPM developed A Rapid UTC project. On the basis of UTC Rapid (UTC\( r \)) scale, every Wednesday on the BIPMs ftp server the \( x_{br}(t) \) deviations are published, determined according to the relationship

\[
x_{br}(t) = UTC_r(t) - UTC_k(t)
\]

for the previous week, for each UTC(k). On 1 July 2013 this project became official.

In this paper, on the example of UTC(PL) scale and using GMDH neural networks, the results of research on the methods of preparation of two time series on the quality of predicting the [UTC-UTC(PL)] are presented. For the preparation of these time series the \( x_b(t) \) and \( x_{br}(t) \) deviations were used.

2. Preparation of the input data for neural network

Predicting the [UTC-UTC(PL)] based on neural networks requires a process of training, the quality of which depends on the number of training data and the method of its preparation [10, 11]. Aiming to achieve the most accurate predictions forces also in the prediction process to maintain the smallest possible time horizon between the date of the determination of the prediction, and the date of the last known value of the time series. Thus, there is also a necessity to supplement the time series with new data published by BIPM, which means that the repetition of the training process of the neural network is needed.

2.1. Input data prepared on the basis of deviations determined according to the UTC scale

The basis of preparing the input data for neural network based on the UTC scale was the relation

\[
x_1(t) = x_a(t) + x_b(t) = UTC(t) - clock_{PL}(t),
\]

where:

\( x_a(t) \) – measurements of the phase time between 1 pps signal from the UTC(PL) and a single caesium clock realizing the UTC(PL) (\( clock_{PL} \)), determined on each day according to relation

\[
x_a(t) = UTC_{PL}(t) - clock_{PL}(t),
\]

\( x_b(t) \) – values of deviations determined by the BIPM for the UTC(PL) in relation to the UTC according to relation (1). For the set of \( x_b(t) \) deviations, an interpolation function PCHIP (Piecewise Cubic Hermite Interpolating Polynomial), available in MATLAB was used. A mathematical model was determined that allows to extend a set of [UTC-UTC(PL)] deviations by calculating its values for each day of analysed data set.

The data set \( x_1(t) \), called time series TS1, characterize on each day the time instability of \( clock_{PL} \) in relation to the UTC. Time series TS1 was created based on a historical data of \( x_a(t) \) and \( x_b(t) \) from MJD 55899 to MJD 56694. Determining the value of [UTC-UTC(PL)] prediction (marked further as \( x_{bp}(t) \)) was carried out from MJD 56514 to MJD 56694. For the first \( x_{bp}(t) \) prediction designated by
the GMDH neural network on day \((t_{\text{pred}})\) MJD 56514 (August 2013) the training data set for time series TS1 was available from day \(t_0\) (figure 1) MJD 55899 (December 2011) to day \(t_{(i)}\) (figure 1) MJD 56504 (July 2013). This means that time series TS1 in the training process of the GMDH neural network contains 605 elements. The required minimum number of elements of times series TS1 ensuring the achievement of predictions with best possible accuracies, determined on the basis of studies [11], is 570. For the predictions designated in 6 consecutive months the number of training data were increased each time by approximately 30 elements, which required re-run of training of the neural network.

2.2. Input data prepared on the basis of deviations determined according to the UTCr scale

The basis for preparation of second time series (TS2) were the \(x_b(t)\) and \(x_{br}(t)\) deviations designated based on the UTC and UTCr scales. This time series consisted of two subsets of elements prepared according to the rules defined on figure 1. The first subset contains a group of data (from 1 to \(i\)) having a values of time series TS1, designated based on relation (3), from day \(t_0\) to day \(t_{(i)}\), for which the last value of \(x_1(t_{(i)})\) before the publication day \(t_{\text{pub}(i)}\) was known. The second subset was a complement of time series TS2 by a group of data having values determined based on relation

\[
x_i(t) = xa(t) + xbr(t) = UTCr(t) - clock_{PL}(t),
\]

between days \(t_{(i)}\) and \(t_{nr}\) published by the BIPM on day \(t_{\text{pubr}}\). The publication day of \(x_{br}(t)\) deviations could also be the day \(t_{\text{predr}}\) of designating the \(xb_{pl}(t)\) prediction. The values of \(xa(t)\) subset, as in the case of data prepared on the basis of the UTC scale, are the measurements of the phase time between the UTC(PL) and clock_{PL}.

![Figure 1. Creation of time series TS2.](image)

3. Research results

The studies were carried out on the basis of prepared time series TS1 and TS2. GMDH neural network used for research belongs to a group of self-organizing networks. In the training process of this network the selection of a degree of the polynomial transfer function of the neuron and the automatic selection of its structure (the number of hidden layers and the number of neurons in these layers) was
chosen. In most cases of training the neural network the selected activation function of the neuron was a linear function. Then the obtained structures of the GMDH neural network consisted of two or one hidden layer having two neurons.

For each month (for TS1) and each week (for TS2) the predictions of $x_{bp}(t)$ on the MJD days ending with digits 4 and 9 were designated. For time series TS1 in a given month a several predictions (from 4 to 7) with prediction horizon of 10, 15, 20, 25, 30, 35 and 40 days could be designated. However for time series TS2 in a given week one or two predictions could be designated. Figure 2 presents a sample set of designated $x_{bp}(t)$ predictions for time series TS2. For comparison also the values of $xb(t)$ and $xbr(t)$ deviations at the same days were presented. A similar course of $x_{bp}(t)$ prediction values and $xb(t)$ deviations were obtained for time series TS1.

Figure 2. Values of $x_{bp}(t)$ predictions and the values of $xb(t)$ and $xbr(t)$ deviations designated by the BIPM for time series TS2.

Figure 3 presents the calculated values of $r$ residuals, determined according to relation:

$$ r = xb(t) - x_{bp}(t), $$

for time series TS1 and TS2. The residuals ($r$) determine the differences between the predicted values of $x_{bp}(t)$ and the $xb(t)$ published by the BIPM for the same day of prediction. In the vast number of cases the designated first predictions in a given month (TS1) or week (TS2) have smaller values of residuals than the consecutive predictions in the same period of time. Practically this means, that only first designated predictions could be used for correcting the UTC(PL) scale. This fact decided that the further analysis of research results was related only to these predictions.

Figure 3. Values of $r(t)$ residuals for time series TS1 a) and for time series TS2 b).
The obtained values of the $r$ residuals were used to evaluate the quality of predictions, which was performed on the basis of selected measures of quality of predictions [12]: Mean Error ($ME$), Mean Absolute Error ($MAE$), Mean Squared Error ($MSE$), Root Mean Squared Error ($RMSE$) and linear correlation coefficient between the $xb(t)$ and $xb_p(t)$. Also a three components of a $MSE$ error ($MSE_1$, $MSE_2$ and $MSE_3$) were designated, which determine the quality of designated predictions. The $MSE_1$ component represents the prediction load, $MSE_2$ is associated with insufficient flexibility of the prediction (so the inaccuracy of estimation of the variation of the predicted variable). The $MSE_3$ component reports of an error related to the insufficient compliance of a change in prediction direction with a change in direction of predicted value. The results of calculations are presented in Table 1.

Table 1. Values of measures of quality of prediction for examined time series.

| Measures of quality of prediction | Time series |
|----------------------------------|-------------|
|                                  | TS1 | TS2 |
| $r_{\text{max}}$ [ns]            | 15.2 | 9.9 |
| $r_{\text{min}}$ [ns]            | -8.0 | -6.4 |
| $ME$ [ns]                        | 3.0  | -2.9 |
| $MAE$ [ns]                       | 6.6  | 4.0 |
| $MSE$ [ns$^2$]                   | 64   | 22 |
| $MSE_1$ [ns$^2$]                 | 8.9  | 8.4 |
| $MSE_2$ [ns$^2$]                 | 37   | 0.2 |
| $MSE_3$ [ns$^2$]                 | 19   | 14 |
| $RMSE$ [ns]                      | 8.0  | 4.7 |

The presented values of $xb_p(t)$ predictions (figure 2) and values of residuals (figure 3) and calculated values of measures of quality of prediction (Table 1) lead to the following conclusions:

- The comparison of the values of residuals and all of the measures of quality of prediction for analysed time series show that the best results of predicting the $xb_p(t)$ were obtained for time series TS2. This is due to the short prediction horizon, which was achieved by introducing into this time series the values of $xb_r(t)$ deviations determined by BIPM.

- The calculated values of correlation coefficient for both time series close to unity, and the values of $xb(t)$ and $xb_p(t)$ (figure 2) show a very good convergence between these values. It also means that absence of correction of the UTC(PL) scale or its use does not affect the predicting process of the [UTC-UTC(PL)] by GMDH neural network.

- Very good convergence also occurs between the values of $xb(t)$ and $xb_p(t)$ (figure 2) determined by the BIPM. For analysed period of time the differences between these values were in the range of -3.2 ns to +2.7 ns. This fact decided on trying to use the values of $xb_r(t)$ deviations for constructing time series TS2.

- The obtained values of residuals for time series TS2 are much smaller than the residuals obtained for time series TS1. The highest residual value for time series TS2 is equal to 9.9 ns, and for time series TS1 15.2 ns. Larger values of residuals for time series TS1 would be expected due to the longer prediction horizon for first prediction in a given month (from 10 to 20 days). It was dictated by using for construction of this time series only the values of $xb(t)$ deviations. Hence, for this time series the larger values of individual measures of quality of prediction, mainly $MSE$ error and its component $MSE_2$ were obtained.

- Predictions of the $xb_p(t)$ designated based on time series TS2 allow to obtain the values of residuals not exceeding the limit of ±10 ns. This means that predicting the $xb_p(t)$ based on data
prepared using this method can contribute to qualify the UTC(PL) to the first group of timescales.

- Comparison of values of $ME$ and $MAE$ errors and value of $MSE_1$ component between time series TS1 and TS2 indicates that designated predictions are underestimated (high number of positive values of $r$ residuals).
- Time series TS2 allows the highest flexibility in determining the prediction (the lowest values of $MSE_2$) in relation to time series TS1. This means that variation of predicted values is good compared to the variation of observed values.
- Time series TS2 also lead to smaller values of $MSE_3$ error, which indicates the higher degree of compliance of a change in prediction direction with a change in direction of predicted value.

4. Conclusions

Our research on methods of preparation of time series for predicting the $xb_p(t)$ has utilized the fact that the introduction of the UTCr scale by the BIPM allows the determination of prediction for the UTC(PL) with a several-day prediction horizon, compared to the horizons of predictions from 10 to 20 days in the case of UTC scale. This enables also more frequent prediction of $xb_p(t)$ (once a week) compared to one prediction a month in the case of UTC scale. This decreased to a large extent a divergence between the $xb_p(t)$ predictions designated using GMDH neural network and $xb(t)$ deviations published by BIPM. The results at a level not exceeding the limit of $\pm 10\text{ ns}$ and the lack of impact of the correction procedure of the UTC(PL) scale on the results of predicting [UTC-UTC(PL)] by GMDH neural network encourages the authors to develop a procedure for predicting the $xb_p(t)$ for the UTC(PL) based on this neural network.

The observed very good convergence between $xbr(t)$ and $xb(t)$ values for UTC(PL) has also been confirmed for other national time scales realized based on single caesium atomic clocks, e.g. UTC(AUS) in Australia, UTC(BEV) in Austria, UTC(CNMP) in Panama, UTC(LT) in Lithuania and UTC(NRC) in Canada. This allows using GMDH neural network for predicting the [UTC-UTC(k)] based on time series TS2 for other timescales than UTC(PL).

References

[1] Arias E F, Panfilo G and Petit G 2011 Timescales at the BIPM Metrologia 48 pp 145–153
[2] 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 pp 483–486
[3] Czubla A, Konopka J and Nawrocki J 2006 Realization of atomic SI second definition in context UTC(PL) and TA(PL) Metrology and Measurement Systems 2 pp 149–159
[4] Davis J A, Shemar S L and Whibberley P B 2011 A Kalman filter UTC(k) prediction and steering algorithm Proc. Joint IEEE FCS EFTF (San Francisco, USA, 2-5 May 2011) pp 779–784
[5] Panfilo G and Tavella P 2008 Atomic clock prediction based on stochastic differential equations Metrologia 45 pp 108–116
[6] Norgard M, Ravn O, Poulsen N and Hansen L 2000 Networks for Modelling and Control of Dynamic Systems (London: Springer)
[7] Luzar M, Sobolewski Ł, Miczulski W and Korbicz J 2013 Prediction of corrections for the Polish Timescale UTC(PL) using artificial neural networks Bulletin of the Polish Academy of Sciences: Technical Sciences 61 pp 589–594
[8] Sobolewski Ł 2013 Predicting the Polish timescale UTC PL based on the corrections designated by the UTC and UTCr scale Proc. Joint UFFC EFTF and PFM (Prague, Czech Republic, 21-25 July 2013) pp 658–661
[9] Sobolewski Ł 2014 Predicting the corrections for the Polish Timescale UTC(PL) using GMDH and GRNN neural networks Proc. European Frequency & Time Forum (Neuchatel, Switzerland, 23-26 June 2014) pp 475–478
[10] Masters T 1993 *Practical neural networks recipes in C++* (San Diego, CA, USA: Academic Press)

[11] Miczulski W and Sobolewski Ł 2012 Influence of the GMDH neural network data preparation method on UTC(PL) correction prediction results *Metrology and Measurement Systems* 19 pp 123–132

[12] Pindyck R S and Rubinfeld D C 1998 *Econometric Models and Economic Forecasts* (Boston: McGraw-Hill)