УДК 621.3.08; 658.562; 620.179.1

SHORT TERM FORECASTING FOR AIR TEMPERATURE BASED ON PATTERN REPETITION

H.M. Hussein, A.G. Yakunin

In this work, a new prediction method based on adjusted pattern repetition (APR) has been proposed to assess the performance of temperature measurements. The formation of the APR model passes through several stages that include the measured data preparation, the template pattern extraction and the adjustment factor calculation. To evaluate the performance of the APR method, it has been applied to measured temperature sample, and compared to the alternative forecasting methods, which included, minimum mean square error (MMSE) forecasting with ARIMA model, Monte Carlo Forecasts and artificial neural networks model. These models have been built and simulated using MATLAB. The results indicated that the proposed APR method has the best prediction accuracy and is considerably faster than the other forecasting models, which makes it the most favorable for online validation of the weather sensors measurements. The proposed APR method has been used for online validation of the temperature sensors measurements. It succeeded in detecting the sensors failures, abnormal measurements and measurements deviations.

Key Words: ARIMA; Monte Carlo; Artificial intelligence; Temperature; Pattern.

Introduction

Weather forecasting is mainly concerned with the expectations of the weather conditions in the future. It provides valuable information about future weather.

There are several approaches involved in weather forecasting, from relatively simple observation of the sky to highly complex computerized mathematical models.

For short-term weather forecasting, several models have been developed. They can be divided into three main categories: parametric, nonparametric, and artificial intelligence methods. The parametric methods formulate a mathematical or statistical model by examining qualitative relationships between time series variables. These models include ARIMA models [1,2], Kalman filter [3–6], and multiple linear regression [7–9]. Nonparametric methods allow a forecast to be calculated directly from historical data [10–13]. Artificial intelligence based methods include artificial neural networks (ANN) and fuzzy logic [14–18].

In this work, a new effective prediction method will be proposed for validating the online temperature sensors measurements.

Notes

- The data for experimental test is extracted from the weather monitoring system database in Altai State Technical University "abc.altsu.ru";
- Matlab will be used for all simulation and experimental analysis.

1. Methodology.

The following steps summarize the formation of the proposed prediction model (APR) that will be used for evaluating the temperature sensors measurements.

1. Periodicity detection. This stage includes:

- Smoothing the measured data. To remove the noise and distortion, the measured temperature ***f(t)*** will be smoothed using Moving average, which is the most common smoothing method.
- Detecting the cyclic behavior of the temperature measurements. To detect the periodicity of the measured data, it will be transformed into frequency-domain representation using Discrete Fourier Transform "DFT", DFT can be calculated with the following equation:

$$Z_k = \sum_{n=0}^{N-1} f_n \cdot e^{-\frac{j2\pi kn}{N}} \qquad 1)$$

Where: *N is the measured data samples.*

The sinusoid's frequency is k cycles per N samples, where ***1 ≤ k ≤ N.***

Each Z_k is a complex number that includes both amplitude and phase of a sinusoidal component of function z_n.

The periodicity frequencies of the temperature series can be obtained from the observing of the significant peaks in the frequency-domain series Z.

2. Template pattern formation. The template pattern will be constructed from the temperature series according to the in order to procedure:

- The temperature series ***f(t)*** will be divided into time slot patterns "S", the period of each pattern "T" is the inverse of the periodicity frequency. Each point **P(t$_i$)** in the template pattern will be equal the average of the analogous points in every pattern in temperature series, and can be calculated according to the following equation:

$$P(t_i) = \frac{1}{S}\sum_{j=1}^{S} S_j\,(t_i) \qquad 2)$$

Where: i = 1, 2,…, m (Number of samples in every time slot) and S "the total number of time slots " = N/T.

- To get one pattern predicted horizon, P(t) will be normalized (by subtracting its average), then readjusted with adjustment factor C, which can be calculated using equation 3.

$$C = \begin{cases} \frac{\frac{m}{2}-i}{m}\bar{S} + \frac{\frac{m}{2}+i}{m}\bar{S}_{+1} & : i = 1,2,\dots,\frac{m}{2} \\ \frac{1.5*m-i}{m}\bar{S}_{+1} + \frac{i-\frac{m}{2}}{m}\bar{S}_{+2} & : i = \frac{m}{2}+1,\dots,m \end{cases} \qquad 3)$$

Where: $\bar{S}$ is the average of the last time slot.

The Calculation for the adjustment factor C requires exceptions for two leading time slots averages $\bar{S}_{+1}$ and $\bar{S}_{+2}$, that can be calculated from the measured data at the same time of the previous years.

- Then, the predicted pattern R(t) can be calculated with the following equation:

$$R(t) = P(t) + C\,(t) \qquad 4)$$

- The previous steps can be repeated to get a long prediction horizon.

To test the proposed prediction method, it has been applied to a measured temperature sample. The results will be summarized in the following section.

2. Experimental evaluations

To test the efficiency of the proposed method, it has been applied to a randomly selected sample of temperature measured every 30 second along one month (June 2014) (exactly 23 days, with total 66,240 sample) using the high resolution DS18S20 sensors, which is a part of a full academic weather monitoring project. "*More details about the project can be found on the website **abc.altstu.ru***".

The temperature sample has been passed through the following stages:

- Preparing the measured data. To remove the distortions, the measured data has been smoothed using the moving average method with half-hour (or 60 sample) moving window. Then the data has been divided into two series. The first series T_h (20 days) has been used to construct the template pattern. The other one T_a (3 days) has been used as actual data for prediction horizon to test the accuracy of the predicted data.
- Template pattern construction. The Template pattern P(t) has been constructed according to equation 2.
- The adjustment factor "C" calculation. The adjustment factor "C" has been calculated using equation 3. The averages $\bar{S}_{+1}$ and $\bar{S}_{+2}$ has been calculated from the same time of the last two years.
- One day ahead prediction. To predict on day ahead R_1, the predicted serried has been calculated using equation 4.

To predict more two days, the calculations for the factor C and R has been repeated.

- Predicted series validation. To validate the predicted series R, It has been compared to the actual series T_a. The average of the residuals between the two series was 0.33, which reflects the good accuracy of the predicted series

The result of previous procedure has been plotted in figure 1

3. Evaluation the proposed prediction method

To assess the performance of the proposed prediction method, it will be compared to alternative pioneer models for forecasting. These alternative models include:

- Minimum mean square error (MMSE) forecasting with ARIMA model. ARIMA *(Auto Regressive Integrated Moving Average)* models have been used widely for forecasting in different fields [19–24]. Seasonal ARIMA models can predict temperature series with good accuracy. Seasonal ARIMA model has been constructed using MATLAB. Based on that model, MMSE forecasts have been calculated recursively for the same temperature series T_h using the MATLAB function "forecast". The average of the residuals between MMSE forecasts and the actual series T_a was 1.34.

- Monte Carlo Forecasts (MC). This is an alternative to minimum mean square error (MMSE) forecasting, which provides an analytical forecast solution [25–28]. An advantage of Monte Carlo forecasting is that it can generate a complete distribution for future events, not just a point estimate and standard error.

The Monte Carlo forecast has been calculated for the same temperature series T_h using the MATLAB functions "infer" and "simulate". The average of the residuals between Monte Carlo forecast and the actual series T_a was 2.43.

Table 1 - Calculation results for the predicted series.

	Average	SD	r'	S_r	MSE	MAPE%	t_c (s)
Actual	25.84	3.59	-	-		-	-
APR	26.17	3.98	0.33	1.45	2.21	4.22	0.002
MMSE	27.18	4.68	1.34	2.40	7.54	8.50	8.025
MC	28.26	5.41	2.43	3.01	14.96	12.20	8.222
NN	26.33	4.23	0.90	2.26	6.06	6.61	25.43

- Artificial Neural Networks forecasting (NN). Artificial Neural Networks have been a pioneer achievement in the field of Intelligent Computation. Simulating the ability of the human brain in order to adapt to complex systems, it's one of the most fascinating fields of Computer Science.

The important characteristic of neural networks is their adaptive nature. This feature makes the ANN techniques common in application domains for solving highly nonlinear phenomena involving pattern classification and forecasting [29–39].

For forecasting the temperature series, ANN with 15 hidden neurons has been constructed using MATLAB (Fig. 2) and trained with the series T_h. The trained network has been used to predict 3 days temperature future. The residuals average for this prediction was 0.9.

Figure 3 shows that, the predicted temperature series using APR method is more closed to the actual series than the predicted series from the alternative models.

To assess the performance of the proposed method (APR) and the mentioned forecasting models, the calculations on the generated series have been carried out to get the average, Standard Deviation (SD), the average of the residuals (r'), the Standard Deviation of residuals (S_r), the mean squared error (MSE), the mean absolute percentage error (MAPE) and the calculation time in seconds (t_c).

The calculation results are summarized in the table 1. These results show that the proposed APR method has the best prediction accuracy and it is considerably faster than the other forecasting models. The best competitor is the NN forecasting, but it needs long time for training the network to get reasonable results.

From simulation results, it seems that the proposed APR method is the most favorable for online validation of the weather sensors measurements due to its accuracy and calculation speed.

The mean issue of the APR is its high dependency on the series seasonality. So, if the series changes randomly "i.e. it doesn't have a cyclic behavior", the performance of the APR may be unsatisfactory.

The proposed APR method has been used for online validation of the temperature sensors measurements. It succeeded to detect the sensors failures, abnormal measurements and measurements deviations.

Conclusion

The experimental results conclude that the proposed APR prediction method is much more effective than the alternative existing models not only in accuracy but also the execution speed. The performance of the APR method is highly dependent on the seasonality of the temperature series.

The proposed method is not applicable only for temperature series, but it can be generalized to involve, stock prices, power consumption in factories, and other similar series.

References

1. Palomares-Salas J.C. et al. ARIMA vs. neural networks for wind speed forecasting // 2009 IEEE International Conference on Computational Intelligence for Measurement Systems and Applications, CIMSA 2009. 2009. P. 129–133.

2. Lin H. et al. Time series analysis of Japanese encephalitis and weather in Linyi City, China. // Int. J. Public Health. 2011. Vol. 31. P. 98–105.

3. Liu H., Tian H. qi, Li Y. fei. Comparison of two new ARIMA-ANN and ARIMA-Kalman hybrid methods for wind speed prediction // Appl. Energy. 2012. Vol. 98. P. 415–424.

4. Welch G., Bishop G. An Introduction to the Kalman Filter // In Pract. 2006. Vol. 7. P. 1–16.

5. Cassola F., Burlando M. Wind speed and wind energy forecast through Kalman filtering of Numerical Weather Prediction model output // Applied Energy. 2012. Vol. 99. P. 154–166.

6. Anandalingam G., Chen L. Bayesian forecast combination and Kalman filtering // International Journal of Systems Science. 1989. Vol. 20. P. 1499–1507.

7. Marill K.A. Advanced Statistics: Linear Regression, Part II: Multiple Linear Regression // Acad. Emerg. Med. 2004. Vol. 11. P. 94–102.

8. Bates D.M., Watts D.G. Review of Linear Regression // Nonlinear Regression Analysis and Its Applications. 1988. P. 1–31.

9. Weisberg S. Applied Linear Regression // Statistics. 2005. Vol. 528. 336 p.

10. Meir R. Nonparametric Time Series Prediction Through Adaptive Model Selection // Mach. Learn. 2000. Vol. 39. P. 5–34.

11. Racine J., Li Q. Nonparametric estimation of regression functions with both categorical and continuous data // J. Econom. 2004. Vol. 119. P. 99–130.

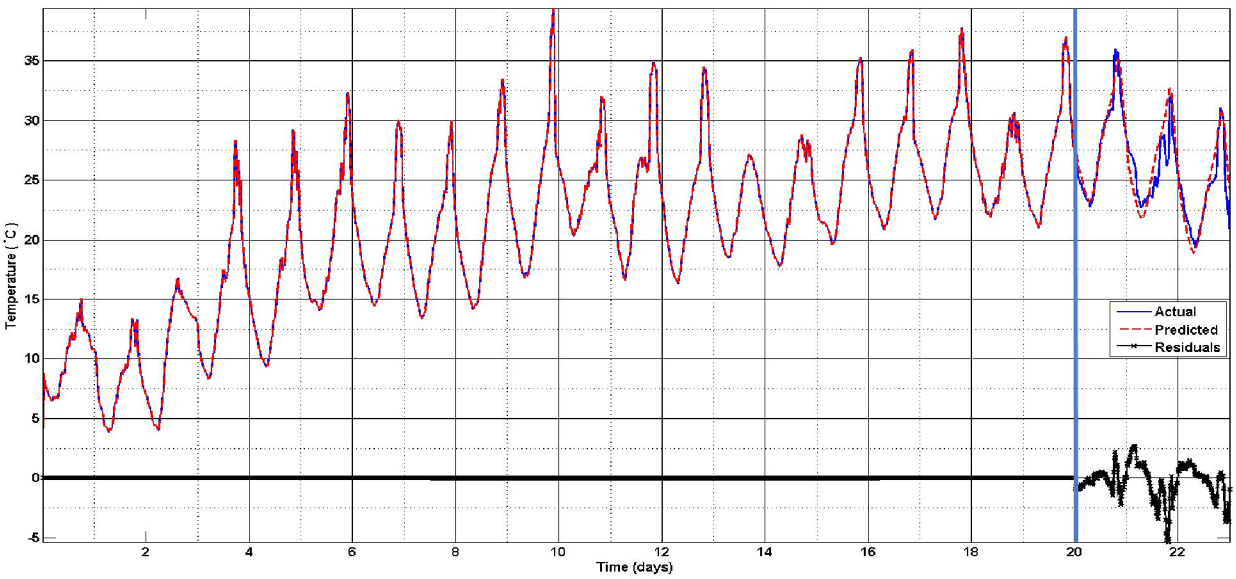

Fig. 1. The temperature actual and predicted series .

12.Ryabko B. Compression-based methods for nonparametric prediction and estimation of some characteristics of time series // IEEE Trans. Inf. Theory. 2009. Vol. 55. P. 4309–4315.

13.13. Aneiros-Pérez G., Vieu P. Nonparametric time series prediction: A semi-functional partial linear modeling // J. Multivar. Anal. 2008. Vol. 99. P. 834–857.

14.Zhang G.P. Neural Networks for Time-Series Forecasting // Handbook of Natural Computing. 2012. P. 461–477.

15.Bishop C.M. Neural Networks for Pattern Recognition // Journal of the American Statistical Association. 1995. Vol. 92. 482 p.

16.Zhang G.P., Patuwo E.B., Michael Y. H. Forecasting with artificial neural networks: The state of the art // Int. J. Forecast. 1998. Vol. 14. P. 35–62.

17.Hayward G., Davidson V. Fuzzy logic applications. // Analyst. 2003. Vol. 128. P. 1304–1306.

18.Saima H. et al. Intelligent methods for weather forecasting: A review // 2011 National Postgraduate Conference - Energy and Sustainability: Exploring the Innovative Minds, NPC 2011. 2011.

19.Zhang G.P. Time series forecasting using a hybrid ARIMA and neural network model // Neurocomputing. 2003. Vol. 50. P. 159–175.

20.Contreras J. et al. ARIMA models to predict next-day electricity prices // IEEE Trans. Power Syst. 2003. Vol. 18. P. 1014–1020.

21.Areekul P. et al. A hybrid ARIMA and neural network model for short-term price forecasting in deregulated market // IEEE Trans. Power Syst. 2010. Vol. 25. P. 524–530.

22.Ömer Faruk D. A hybrid neural network and ARIMA model for water quality tie series prediction // Eng. Appl. Artif. Intell. 2010. Vol. 23. P. 586–594.

23.Shu Y.S.Y. et al. Wireless traffic modeling and prediction using seasonal ARIMA models // IEEE Int. Conf. Commun. 2003. ICC ’03. 2003. Vol. 3.

24.Chen P., Yuan H., Shu X. Forecasting crime using the ARIMA model // Proceedings - 5th International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2008. 2008. Vol. 5. P. 627–630.

25.Cullen K.A., Virginia W. Forecasting Electricity Demand using Regression and Monte Carlo Simulation Under Conditions of Insufficient Data Division of Resource Management Forecasting Electricity Demand using Regression and Monte Carlo // Power. 1999.

26.Metropolis N., Ulam S. The Monte Carlo Method // J. Am. Stat. Assoc. 1949. Vol. 44. P. 335–341.

27.Zakhary A. et al. Forecasting hotel arrivals and occupancy using Monte Carlo simulation // Journal of Revenue and Pricing Management. 2011. Vol. 10. P. 344–366.

28.Berliner L.M. Monte Carlo based ensemble forecasting // Stat. Comput. 2001. Vol. 11. P. 269–275.

29.Nakamura E. Inflation forecasting using a neural network // Econ. Lett. 2005. Vol. 86. P. 373–378.

30.Maqsood I., Khan M., Abraham A. An ensemble of neural networks for weather forecasting // Neural Computing and Applications. 2004. Vol. 13.

31.Taylor J.W., Buizza R. Neural network load forecasting with weather ensemble predictions // IEEE Trans. Power Syst. 2002. Vol. 17. P. 626–632.

32.Teschl R., Randeu W.L., Teschl F. Improving weather radar estimates of rainfall using feed-forward neural networks // Neural Networks. 2007. Vol. 20. P. 519–527.

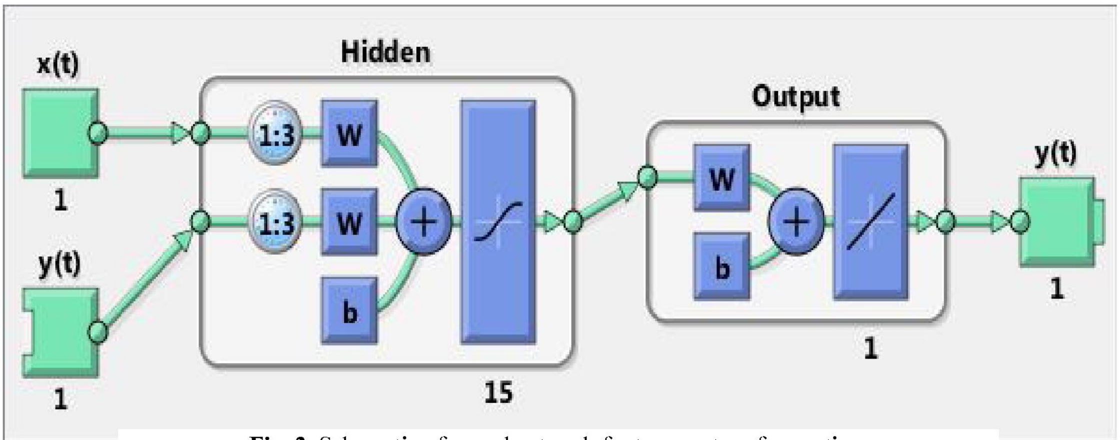

Fig. 2. Schematic of neural network for temperature forecasting.

33.Zhang G., Eddy Patuwo B., Y. Hu M. Forecasting with artificial neural networks: // International Journal of Forecasting. 1998. Vol. 14. P. 35–62.

34.Hippert H.S., Pedreira C.E., Souza R.C. Neural networks for short-term load forecasting: a review and evaluation // IEEE Transactions on Power Systems. 2001. Vol. 16. P. 44–55.

35.More A., Deo M.C. Forecasting wind with neural networks // Mar. Struct. 2003. Vol. 16. P. 35–49.

36.36. Qi M., Zhang G.P. Trend time-series modeling and forecasting with neural networks // IEEE Trans. Neural Networks. 2008. Vol. 19. P. 808–816.

37.Singhal D., Swarup K.S. Electricity price forecasting using artificial neural networks // International Journal of Electrical Power & Energy Systems. 2011. Vol. 33. P. 550–555.

38.Hinton G. et al. Deep Neural Networks for Acoustic Modeling in Speech Recognition // IEEE Signal Process. Mag. 2012. Vol. 29. P. 82–97.

39.Charytoniuk W., Mo-Shing C. Very short-term load forecasting using artificial neural networks // IEEE Trans. Power Syst. 2000. Vol. 15. P. 263–268.

Якунин А. Г. *- д.т.н., профессор, кафедра вычислительных систем и информационной безопасности ФГБОУ ВПО «Алтайский государственный технический университет им. И.И. Ползунова», г. Барнаул,*
E-mail: yakunin@agtu.secna.ru)

Хуссейн Х.М. *- аспирант, кафедра вычислительных систем и информационной безопасности ФГБОУ ВПО «Алтайский государственный технический университет им. И.И. Ползунова», г. Барнаул, (Египет,*
E-mail: helphs@yahoo.com)