KnE Life Sciences

ICMScE
International Conference On Mathematics And Science Education
Volume 2024

Knowledge E
enriching | engaging | empowering

Research Article

Implementing Time Series Cross Validation to Evaluate the Forecasting Model Performance

Winita Sulandari[1]*, Yudho Yudhanto[2], Sri Subanti[1], Etik Zukhronah[1], and Muhammad Zidni Subarkah[1]

[1]Study program of Statistics, Universitas Sebelas Maret, Indonesia
[2]Informatics Engineering, Vocational School, Universitas Sebelas Maret, Indonesia

ORCID
Winita Sulandari: https://orcid.org/0000-0002-8185-1274
Yudho Yudhanto: https://orcid.org/0000-0001-8998-8577
Sri Subanti: https://orcid.org/0000-0002-2493-4583
Etik Zukhronah: https://orcid.org/0000-0001-6387-4483

Abstract.
Theoretically, forecast error increases as the forecast horizon increases. This study aims to assess whether the statement is generally accepted or not. This study applies time series cross-validation to evaluate forecasting results up to seven steps ahead. As an illustration, we use Malaysia's hourly electricity load data. Each hour is considered a series of each, so there are 24 daily series. Time series cross-validation with a 334 window was applied to 24 data series, and then each daily series was modeled with the Autoregressive Integrated Moving Average (ARIMA), Neural Network Autoregressive (NNAR), ExponenTial Smoothing (ETS), Singular Spectrum Analysis (SSA), and General Regression Neural Network (GRNN) models. In terms of mean absolute percentage error (MAPE) from one to seven steps ahead, we then evaluate the performance of all models. The experimental results show that the MAPEs obtained from the GRNN model tend to increase along with the theory. However, MAPEs obtained from ETS increase by up to three steps ahead and decrease after that. Among the five models, ARIMA, NNAR, and SSA produce a reasonably stable MAPE value for one to seven steps ahead. However, SSA has the most stable error value compared to ARIMA and NNAR.

Keywords: time series, cross validation, evaluate, forecasting model performance

Corresponding Author: Winita Sulandari; email: winita@mipa.uns.ac.id

Published: 27 March 2024

Publishing services provided by Knowledge E

Selection and Peer-review under the responsibility of the ICMScE Conference Committee.

1. INTRODUCTION

In forecasting activities, accuracy is crucial and is the main aim that researchers want to achieve. Many techniques have been developed and implemented, from preprocessing, estimating parameters of the models, and combining two or more forecasting methods. Researchers usually decide the reliable model based on the accuracy performance obtained from the training and the testing data. Though, they prefer assessed for accuracy using testing data rather than goodness of fit to the training data [1]. Evaluation of the accuracy of forecasting results on the testing data needs to be done in an effort

to get a model that is not only good on training but also on new data or, in other words, suitable for forecasting up to several steps forward.

General way in determining the training and testing data are by divide the sample into two parts, named fixed origin cross validation or also called by holdout cross validation [2]. Fixed origin cross validation is considered the efficient and simple method of time series cross validation. However, for the small sample sizes of data, splitting them into two parts will limit both the training and the testing data. As a result, the parameter estimation model may not reflect the actual conditions. In addition, conclusions drawn from limited observations in the test data may not be reliable. To deal with this, we can use rolling window cross validation. Time series cross-validation with a rolling window is an advanced method of fixed origin cross validation [3, 4]. In this procedure, we pruning the oldest observation and add one new observation to training set in each update of the forecast origin. Thus, we have many pairs of training and test sets. Each test set consists of one observation and focuses on a single forecast horizon. Therefore, the accuracy measures is calculated from the average across all each single forecast horizon test sets rather than averaging across several forecast horizons. Note that in this approach, we do not use future observations to construct the forecast [5]. Therefore, we consider that accuracy measures obtained by the rolling window cross validation appropriate to select the best model that produce good multi-step ahead forecast value.

As an illustrative example, we considers Malaysia's hourly electricity load time series data. Some researchers showed that hourly load series has trend and multiple seasonal patterns [6, 7]. To lessen an hourly load series complexity, [8] and [9], partitioned the data into 24 daily time series, each showing every hour of the day. In this way, the daily seasonal pattern can be eliminated to simplify the pattern. However, the number of data samples is becoming more limited, which may raise a problem in defining a reliable model. Therefore, applying rolling window cross-validation will be the solution.

This study aims to evaluate the forecasting results by implementing rolling window cross validation. We compares autoregressive integrated moving average (ARIMA), neural network, exponential smoothing (ETS), and singular spectrum analysis (SSA) in terms of Mean Absolute Percentage Error (MAPE) evaluated from up to seven ahead of the forecast value. These methods or their combinations worked well in modeling and forecasting electricity load time series, see [10–12]. We can select the best time series model among all candidate models by implementing time-series cross-validation with a rolling window. Based on the experiment results, we show that a model that is good at predicting one step in the future is not necessarily good at predicting several steps in the future, and vice versa.

2. RESEARCH METHOD

This study observed the performance of five popular models, i.e., ARIMA, NNAR (Neural Network Autoregressive), ETS, SSA, and GRNN (General Regression Neural Network), in forecasting hourly electricity load data up to seven steps ahead. The accuracy forecasting values each of one to seven steps ahead obtained by implementing time series cross validation are evaluating by MAPE. Below is brief overview of ARIMA, NNAR, ETS, SSA, and GRNN methods.

1. ARIMA

Let $\{z_t,\ t = 1,\ 2,\ \ldots,\ n\}$ is a time series. ARIMA model for z_t, notated by ARIMA (p, d, q) is represented as (1)

$$\boxtimes(B)\,\nabla^d z_t = \theta_0 + \theta(B)\,a_t \qquad (1)$$

where $\boxtimes(B) = 1 - \boxtimes_1 B - \boxtimes_2 B^2 - \cdots - \boxtimes_p B^p$, $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$, and $\nabla = 1 - B$. $\nabla^d z_t$ is the differenced series of z_t with order *d* that follows stationary process $\boxtimes(B)$ is the autoregressive operator with the root of $\boxtimes(B) = 0$ lie outside the unit circle and $\theta(B)$ is moving average operator with the root of $\theta(B) = 0$ lie outside the unit circle [13].

1. NNAR

NNAR model is a feed forward neural network consists of input, hidden, and output layer. It represented as (2)

$$z_t = v_0 + \sum_{j=1}^{h} v_j g\left(w_{0j} + \sum_{i=1}^{m} w_{ij} y_{t-i}\right) + \boxtimes_t \qquad (2)$$

where w_{0j}, w_{ij}, v_0, and v_j $(i = 1,\ \ldots,\ m; j = 1,\ \ldots,\ h)$ are weights connected between two nodes. Symbol *m* represents the number of input, *h* denotes the number of node in hidden layer, and *g* is a sigmoid activation function that can be represented as $g(x) = 1/[1 + expexp(-x)]$ (3)

2. ETS

ETS model developed by [14] is an innovations state space models for exponential smoothing. It consists of a measurement equation and some state equations describe level, trend, and seasonal that change over time. These state models label by ETS (Error, Trend, Seasonal) where the possibilities for each state are Error = {A, M}, Trend = {N, A, Ad}, and Seasonal = {N, A, M}. Notations A, M, N, Ad mean Additive, Multiplicative, None, and Additive damped, respectively. Therefore, ETS (A, N, N) means simple exponential smoothing with additive errors (no trend and no seasonal).

3. SSA

   SSA is a non-parametric method for time series analysis and forecasting. This method consists of four main steps, i.e. embedding, decomposition, grouping, and diagonal averaging. The idea is to decompose time series into two subseries, signal and noise. Based on the signal, we can then calculate the forecast values by using linear recurrent formula. The important thing in modeling and forecasting time series using SSA is how we group the eigentriples by considering the w-correlation matrix. Some references discussed SSA can be found in [15, 16].

4. GRNN

   GRNN is an improved neural network with radial basis function neurons in its hidden layers [17]. It can learn quickly than NNAR since it has a single pass learning. Each training sample is associated with the center of radial basis neuron that commonly use Gaussian kernel function, $G(x,\ x_i) = expexp\,(-\ x\ --\ xi\ ^2/2^2)$ where x is the input, x_i is the center, and σ^2 is the smoothing parameter. The output of the network can be represented as $\hat{y}(x) = \sum_i^n y_i G(x,\ x_i)/\sum_i^n G\left(x,\ x_i\right)$. However, GRNN is sensitive to the smoothing parameter. Therefore, an optimization tool that minimizing forecast error measure can be used to select a suitable value for it [18]

In this work, we use R with the “forecast”, “tsfgrnn”, and “R_{SSA}” packages for all modeling and forecasting. Procedure of the study are explained in the following steps.

Step 1: Divide the hourly electricity time series data into 24 groups, each indicates hour in a day, 1, 2, ..., 24.

Step 2: Implement time series cross validation with a window.

1. State the window, that is the size of the training data

2. Obtain the training based on the chosen window and the testing data. As an illustrative example, see Fig. 1. Consider a time series with the sample size is 60. If we state the window is 48 and we will evaluate forecast accuracy up to five steps ahead, then the blue dots in each line in Fig. 1 are the training data to model and forecast the one (black), two (green), three (orange), four (red), and five (purple) steps ahead. In this case we have eights training data

Step 3: Modeling each subseries of the training data defined in Step 2a by ARIMA, NNAR, ETS, SSA, and GRNN.

Step 4: Calculate forecast values up to seven steps ahead using models obtained from Step 3.

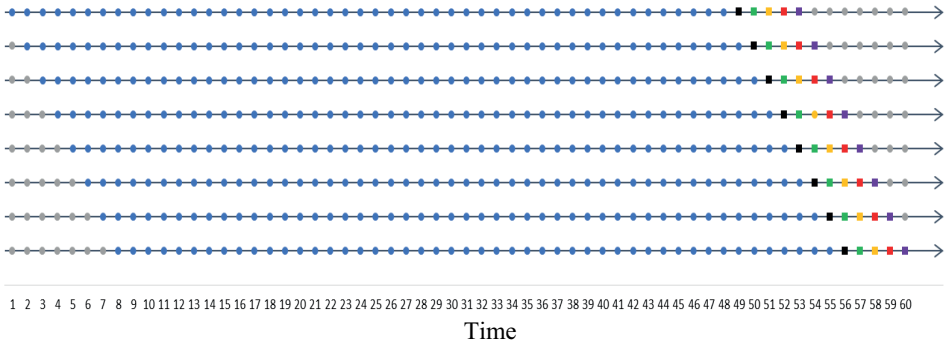

Figure 1: Illustration for time series cross validation with rolling window is 48 (blue points as the training data).

Step 5: Calculate MAPE for each step ahead by averaging across the corresponding the test sets.

3. RESULTS AND DISCUSSION

A Malaysia hourly electricity load time series data from January 2009 to December 2009 [19] is considered in this study. The data are depicted in Fig. 2 Further analysis showed that the data has double seasonal pattern with a daily and a weekly seasonal period (13). Inspired by [20, 21], we split the data into 24 groups, each group for each hour as presented in Fig. 3 to remove the daily seasonal. Figure 3 shows that the fluctuation of the data and their variances are different between one and another, especially during the work hours and sleep hours. For this reason, it possible that a model that is good for a certain series may not suitable for another.

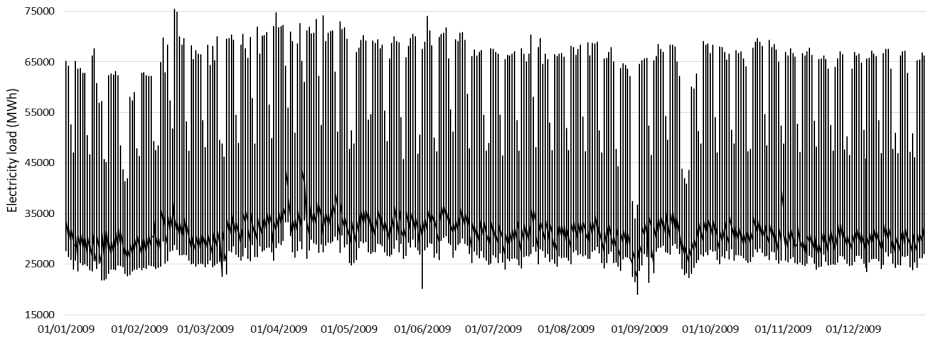

Figure 2: Malaysia's hourly electricity load from 1 January 2009 00:00 to 31 December 2009 23:00.

Later, we work with 24 daily time series data. Each series presented in Fig. 3 consists of 365 observations, except for the series of hour 24, which has 364 observations.

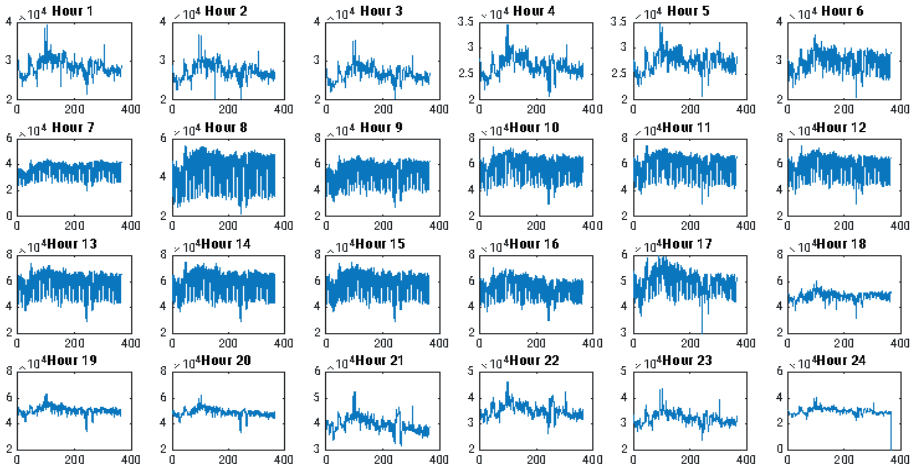

Figure 3: Malaysia electricity load for each hour from January 1 to December 31, 2009.

These 24 series are then divided into training and testing data using cross validation with a rolling window of 334. Thus, we have 25 series as the training data for the series of Hour 1 to Hour 23 and 24 series for Hour 24. As illustration, the distribution of training and testing data for Hour 1 is demonstrated in Fig. 4.

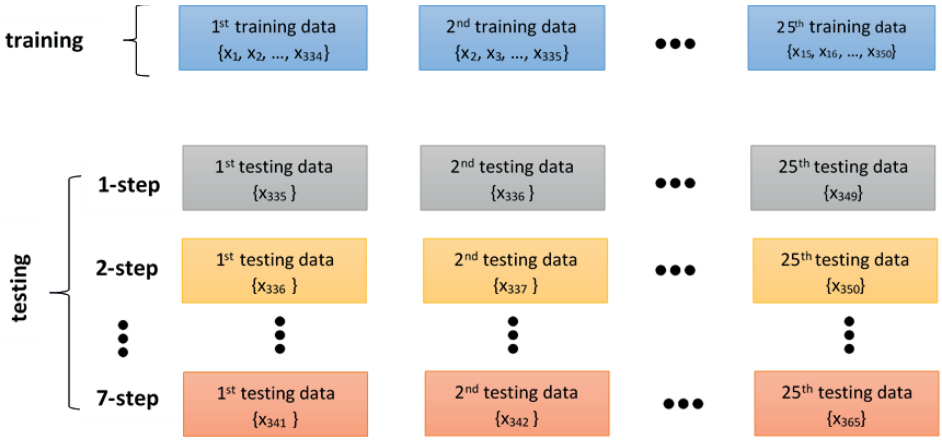

Figure 4: Example of time series cross validation for the series of Hour 1 with rolling window is 334.

Each training data is then modelled by ARIMA, NNAR, ETS, SSA, and GRNN. Therefore, we have 2995 models, obtained from (25 x 23) + 24 training data sets, each with five models. All calculations were done using R. In modelling ARIMA and ETS, we use “auto.arima” and “ets” function, respectively. These two function is available in the “forecast” package. The best model obtained from the training data is selected according to the AIC (Akaike’s Information Criterion) [22]. Meanwhile, for modelling NNAR and GRNN, we use “nnetar” and “grnn_forecasting” function, respectively. The function of “nnetar” is included in “caret” package while “grnn_forecasting” is included

in “tsfgrnn” package. Further, “ssa” and “rforecast” functions are needed to model and obtain the forecast values of SSA model. Readers may contact the corresponding author for the code we used to construct the models and obtain the forecast values.

The experiment results are summarized in Fig. 5. Based on Fig. 5, we can see that for Hour 1 to Hour 6 and Hour 18 to Hour 24, ETS produces a high MAPE at forecasting one step ahead and gets higher at two to three step ahead (even four step ahead, see Hour 6), decreases after that and reaches the lowest value at predicting seven step ahead. Meanwhile, GRNN provides a higher MAPE value than others. In this case, GRNN is not recommended for modelling and forecasting Malaysia daily electricity load time series data. Further studies may be done to develop this method by combining it with others.

Furthermore, the ARIMA as a conventional method appears to provide relatively smaller MAPE than the GRNN and ETS. However, it does not apply to forecasting the seven step ahead for Hour 1 to Hour 6 and Hour 19 to Hour 24 because ETS is the winner at this point. Meanwhile, SSA and NNAR consistently produce smaller MAPE values for forecasting up to seven steps ahead among the five models. We admit that the value is still above 2%. However, based on these results, we can focus on these two models’ development or consider both of them as components of a hybrid model to obtain a higher forecasting accuracy value. Finally, we cannot generally accept the statement that forecasting errors will increase with the forecast horizon. The MAPE values obtained from the ETS model for up to seven steps ahead, shown in Fig. 5, support this conclusion.

In addition, a model that yields good one step ahead forecasting accuracy is not necessarily suitable for predicting two or more steps ahead, and vice versa.

4. CONCLUSION

The main finding of this study is that the forecast error statement does not apply to increase in line with the forecast horizon. Time series cross validation with a rolling window can be considered when we have a limited number of data. Moreover, a model with good forecasting results for one step ahead may not necessarily provide good forecasts for multiple steps ahead. In the case of the Malaysia daily electricity load series discussed in this study, SSA and NNAR can be considered the methods that may develop to increase the forecast accuracy, not only for one step but also for multiple steps ahead. For future research, we may consider combining two or more methods so that we get a hybrid model with stable forecasting accuracy for one to several steps ahead.

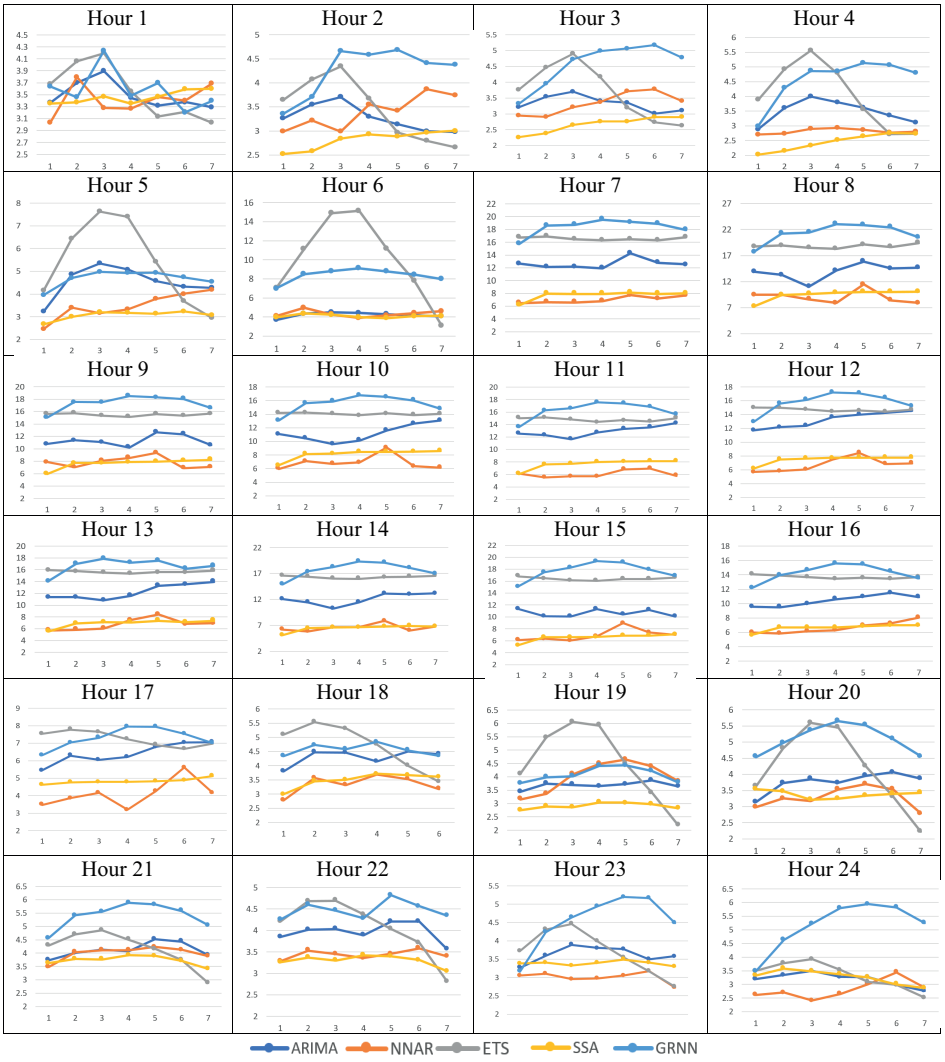

Figure 5: Comparisons of MAPEs obtained from forecast values for one to seven steps ahead using ARIMA, NNAR, ETS, SSA, and GRNN models.

Acknowledgments

We thank LPPM Universitas Sebelas Maret for supporting this research. This research was funded by the Ministry of Education, Culture, Research, and Technology Indonesia with source from the DIPA DIKTI RISTEK 2022, Number SP DIPA-023.17.1.690523/2022 (second revision on 22 April 2022), in the scheme of National Competitive Basic Research with Contract Number 096/E5/PG.02.00.PT/2022and673.1/UN27.22/PT.01.03/2022.

References

[1] Tashman LJ. Out-of-sample tests of forecasting accuracy: an analysis and review. Int J Forecast. 2000;16(4):437–50.

[2] Haris NA, Aziz AA, Nor NA, Sharif N. Improving Air Pollution Index (API) predictive accuracy using time series cross-validation technique. Journal of Fundamental and Applied Sciences. 2018;10(June):1257–67.

[3] R.J. Hyndman and G. Athanasopoulos, *Forecasting: principles and practice. OTexts*, 2018.

[4] Liu X, Yang MC. Simultaneous curve registration and clustering for functional data. Comput Stat Data Anal. 2009;53(4):1361–13776.

[5] R.J. Hyndman and G. Athanasopoulos, *Forecasting: principles and practice. OTexts*, 2018.

[6] De Livera AM, Hyndman RJ, Snyder RD. Forecasting time series with complex seasonal patterns using exponential smoothing. J Am Stat Assoc. 2011;106(496):1513–27.

[7] Lee JY, Kim S. Forecasting daily peak load by time series model with temperature and special days effect. The Korean Journal of Applied Statistics. 2019;32(1):161–71.

[8] Soares LJ, Medeiros MC. Modeling and forecasting short-term electricity load: A comparison of methods with an application to Brazilian data. Int J Forecast. 2008;24(4):630–44.

[9] Sulandari W, Subanar S, Suhartono S, Utami H, Lee MH, Rodrigues PC. SSA-based hybrid forecasting models and applications. Bulletin of Electrical Engineering and Informatics. 2020;9(5):2178–88.

[10] Arora S, Taylor JW. Short-term forecasting of anomalous load using rule-based triple seasonal methods. IEEE Trans Power Syst. 2013;28(3):3235–42.

[11] Sulandari W, Utami H. "Forecasting time series with trend and seasonal patterns based on SSA.," In: *2017 3rd International Conference on Science in Information Technology (ICSITech)*. pp. 648–653. *IEEE* (2017). https://doi.org/10.1109/ICSITech.2017.8257193.

[12] Sulandari W, Subanar S, Lee MH, Rodrigues PC. "Time series forecasting using singular spectrum analysis, fuzzy systems and neural networks.," *MethodsX*. vol. 7, p. 2020. https://doi.org/10.1016/j.mex.2020.101015.

[13] Box GE, Jenkins GM, Reinsel GC, Ljung GM. Time series analysis: forecasting and control. John Wiley & Sons; 2015.

[14] Hyndman R, Koehler AB, Ord JK, Snyder RD. Forecasting with exponential smoothing: the state space approach. Springer Science & Business Media; 2008. https://doi.org/10.1007/978-3-540-71918-2.

[15] Golyandina N, Korobeynikov A. Basic singular spectrum analysis and forecasting with R. Comput Stat Data Anal. 2014;71:934–54.

[16] Golyandina N, Korobeynikov A, Zhigljavsky A. Singular spectrum analysis with R. Springer Berlin Heidelberg; 2018. https://doi.org/10.1007/978-3-662-57380-8.

[17] Specht DF. *Brief Papers A General Regression Neural Network*.

[18] Martínez F, Charte F, Rivera AJ, Frías MP. Automatic time series forecasting with GRNN: A comparison with other models. In International Work-Conference on Artificial Neural Networks. Cham: Springer International Publishing; 2019. pp. 198–209.

[19] Sadei HJ, Silva PC, Guimaraes FG, Lee MH. Short-term load forecasting by using a combined method of convolutional neural networks and fuzzy time series. Energy. 2019;175:365–77.

[20] Soares LJ, Medeiros MC. Modeling and forecasting short-term electricity load: A comparison of methods with an application to Brazilian data. Int J Forecast. 2008;24(4):630–44.

[21] Sulandari W, Subanar S, Suhartono S, Utami H, Lee MH, Rodrigues PC. SSA-based hybrid forecasting models and applications. Bulletin of Electrical Engineering and Informatics. 2020;9(5):2178–88.

[22] Hyndman RJ, Khandakar Y. Automatic time series forecasting: The forecast Package for R. J Stat Softw. 2008;27(3):22.

Bottom of Form