SILESIAN UNIVERSITY OF TECHNOLOGY PUBLISHING HOUSE

SCIENTIFIC PAPERS OF SILESIAN UNIVERSITY OF TECHNOLOGY 2023
ORGANIZATION AND MANAGEMENT SERIES NO. 171

A DECOMPOSITION STUDY OF THE TIME SERIES OF ELECTRICITY CONSUMPTION FORECASTING ERRORS

Maciej WOLNY

Silesian University of Technology, Zabrze; maciej.wolny@polsl.pl, ORCID: 0000-0002-8872-7794

Purpose: This paper attempts to present a method for studying hourly time series of forecasting errors in electricity consumption in the context of daily and weekly seasonality.
Design/methodology/approach: The proposed approach is based on MSTL (Bandara, Hyndman, Bergmeir, 2021) decomposition of hourly forecast error series. The method is presented using the example of household electricity consumption based on data (Makonin, 2019). The time series was divided into a training set and a test set. The forecast was made based on the training set for the test period. Next, the time series of differences between the actual (test set) and forecast values was examined. Calculations were performed in the R environment.
Findings: Decomposition of the forecasting error time series makes it possible to isolate the seasonal (systematic) components of forecasting errors. The values of daily and weekly errors show how forecast values deviate from actual values in a systematic way. These values can be used to adjust forecasts in subsequent periods.
Research limitations/implications: Identification and inclusion of seasonal components in forecasting errors may not improve forecast quality. The assumption made about daily and weekly seasonality of the error may not be met.
Originality/value: This paper presents a decompositional approach to examining the time series of forecasting errors in order to identify deterministic error components for possible adjustment of future forecasts.

Keywords: forecasting error decomposition, error time series, forecasting error analysis, multi-seasonal decomposition.

Category of the paper: Research paper, Conceptual paper.

1. Introduction

Forecasting electricity consumption is of significant economic importance. It is primarily related to electricity demand planning.

Various approaches and models are used to forecast electricity consumption (Islam et al., 2022; Ghalehkhondabi et al., 2017). In addition to classical methods (Hyndman, Athanasopoulos, 2018), artificial intelligence models are also used (Ahn, Kim, 2022).

Electricity consumption series are characterized by seasonality (daily, weekly and yearly) which forecasting models should take into account (Shao et al., 2017).

In business practice, profiles (mainly annual) of electricity consumption are often used (Duarte et al., 2022; Anvari et al., 2022; Bargiel et al., 2019). The main reasons for using consumption profiles are: there are historical data gaps (Kowalska-Styczeń et al., 2022); consumption readings are taken at a different frequency than required.

As a result, forecasts have large errors, which can be subjected to various analyses. In general, averaged forecasting errors are determined, such as Mean Absolute Error, Mean Squared Error. A more detailed study of forecasting errors can involve Fourier analysis (Yang et al., 2022). In such a situation, the time series of forecasting errors can be analyzed like any time series.

In this paper, it is assumed ad hoc that the forecasting error series is characterized by the same seasonality as the electricity consumption series. For the identification of individual components, the considerations presented here refer to a forecasting perspective in which the error time series is sufficiently long.

The main objective of this paper is to present the author's method for investigating forecasting errors in the context of daily and weekly seasonality based on the MSTL decomposition (Bandara et al., 2021), for hourly data. The identification of the daily and weekly components of the error series shows what proportion of the overall error is accounted for by daily and weekly systematic variation. The identification of systematic components can be used to adjust the forecasting model in subsequent forecasting periods.

2. Materials and methods

Secondary data was analyzed (Makonin, 2018). The time series "Residential_1.tab", which shows electricity consumption [kWh] in bungalows (belongs to single-level houses built in the 1940s and 1950s). Hourly data from 2012-06-01 to 2015-05-31 were used for the analysis.

The time series was divided into two subsets (two time windows). Each subset covered one year.

The first subset covered the period from 2012-06-01 to 2013-05-31 and constituted the learning part (training set). Based on this set, a forecast was made for the following year.

The second subset of data covered the period from 2013-06-01 to 2014-05-31 and constituted the test part (test set) for forecasts built on the basis of the first subset. Based on the second subset and the forecasts covering this period, a series of forecasting errors was determined and analyzed.

The data and the division of the data into subsets are shown in Figure 1.

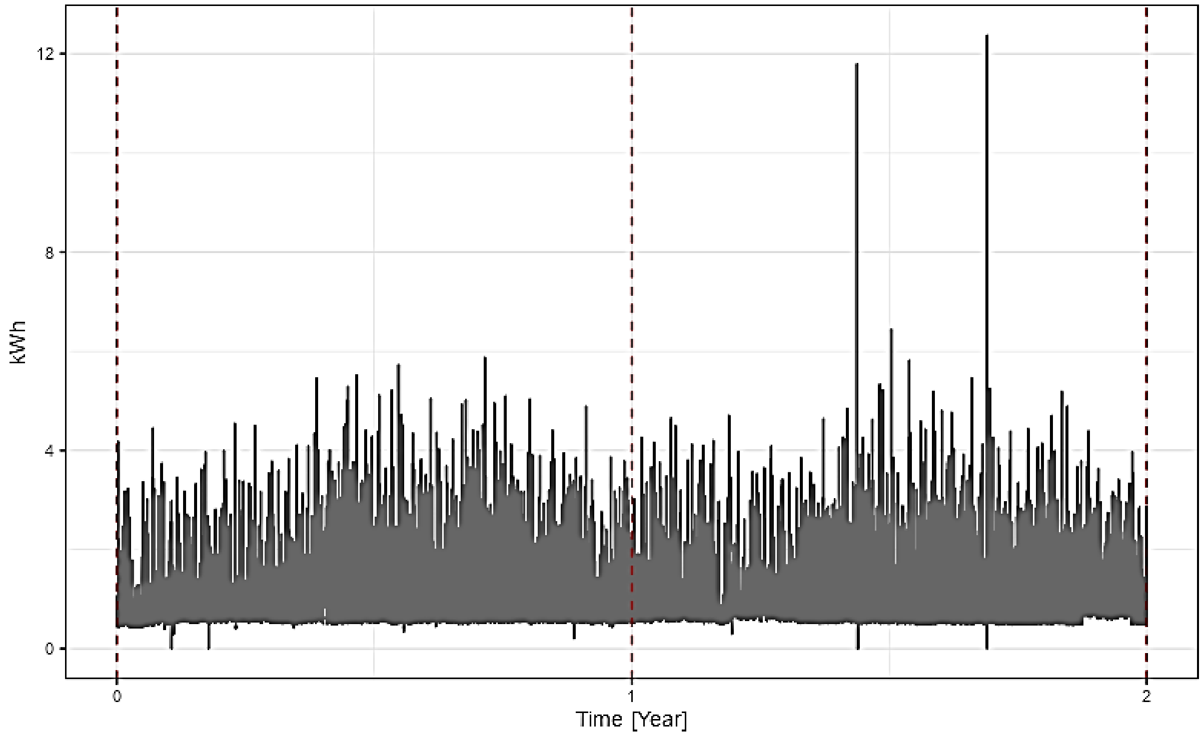

Figure 1. Hourly electricity consumption by training and test sets.

The research was carried out in stages related to the division of data into subsets. In the first stage, a forecast for the following year was built. A seasonal naive model was used, taking into account annual seasonality.

In the second stage, forecasting errors were determined for the test set and an MSTL (Bandara et al., 2021) decomposition of the error series was performed. The average daily and weekly profile of forecasting errors was determined. The strength of seasonality and trend in the errors was examined using measures (Wang et al., 2006):

$$F_T = \max\left(0, 1 - \frac{Var(R_t)}{Var(T_t + R_t)}\right), \tag{1}$$

$$F_S = \max\left(0, 1 - \frac{Var(R_t)}{Var(S_t + R_t)}\right), \tag{2}$$

where:

T_t is the smoothed trend component,

S_t is the seasonal component,

R_t is a remainder component.

Equation (1) determines the strength of the trend component and equation (2) the seasonal component. The MSTL decomposition is additive, so the strength of the seasonal component (2) was counted separately for the analyzed seasonality and together. Measures of component strength (1) and (2) take values from 0 to 1. The closer the value to 1, the stronger the component.

All calculations were performed in the R environment (R Core Team, 2022), using the "forecast" package (Hyndman et al., 2022).

3. Results

The data subsets studied have a similar distribution of values (Figure 2).

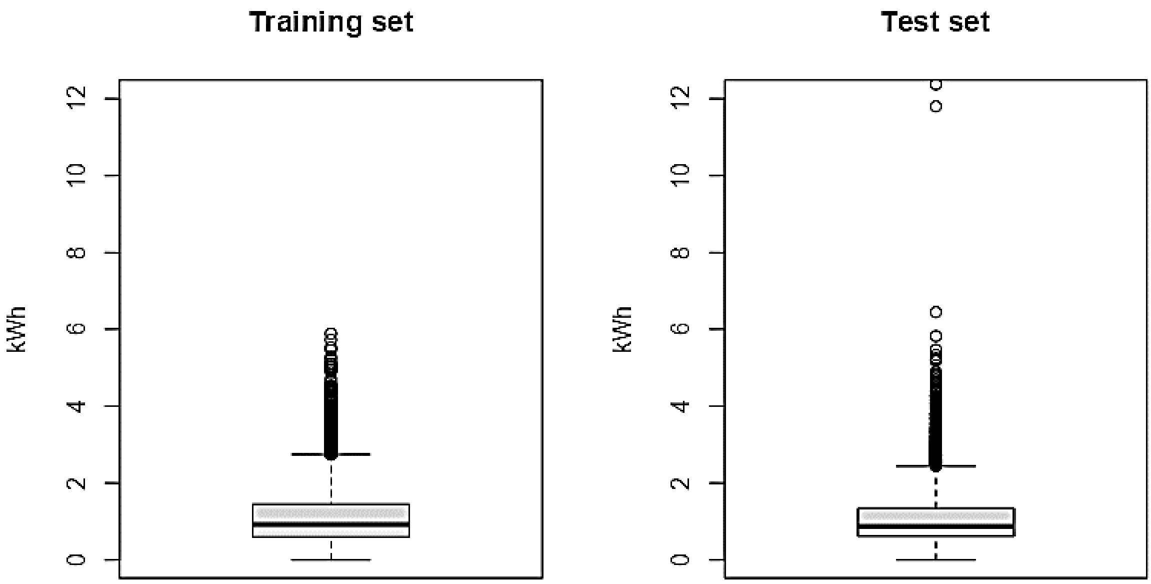

Figure 2. Distribution of values of the considered data sets.

The basic numerical descriptive characteristics are shown in Table 1.

Table 1.
Basic descriptive characteristics of the data sets under consideration

Measure	Training set	Test set
Minimum	0.000	0.000
1st quartile	0.596	0.614
Median	0.918	0.873
Mean	1.141	1.098
3rd quartile	1.454	1.343
Maximum	5.880	12.372
Standard deviation	0.699	0.697

The main difference in the distributions relates to the maximum values. There are two definite outlier observations in the test set. Since the problem relates to electricity consumption for two hours during the year, this has no significant impact. Moreover, there is no indication that the recorded consumption is a consequence of the error. Thus, these observations were left in the set.

Based on the training set, a forecast was made for the following year, that is, for the time period of the test set. The forecasting results are shown in Figure 3.

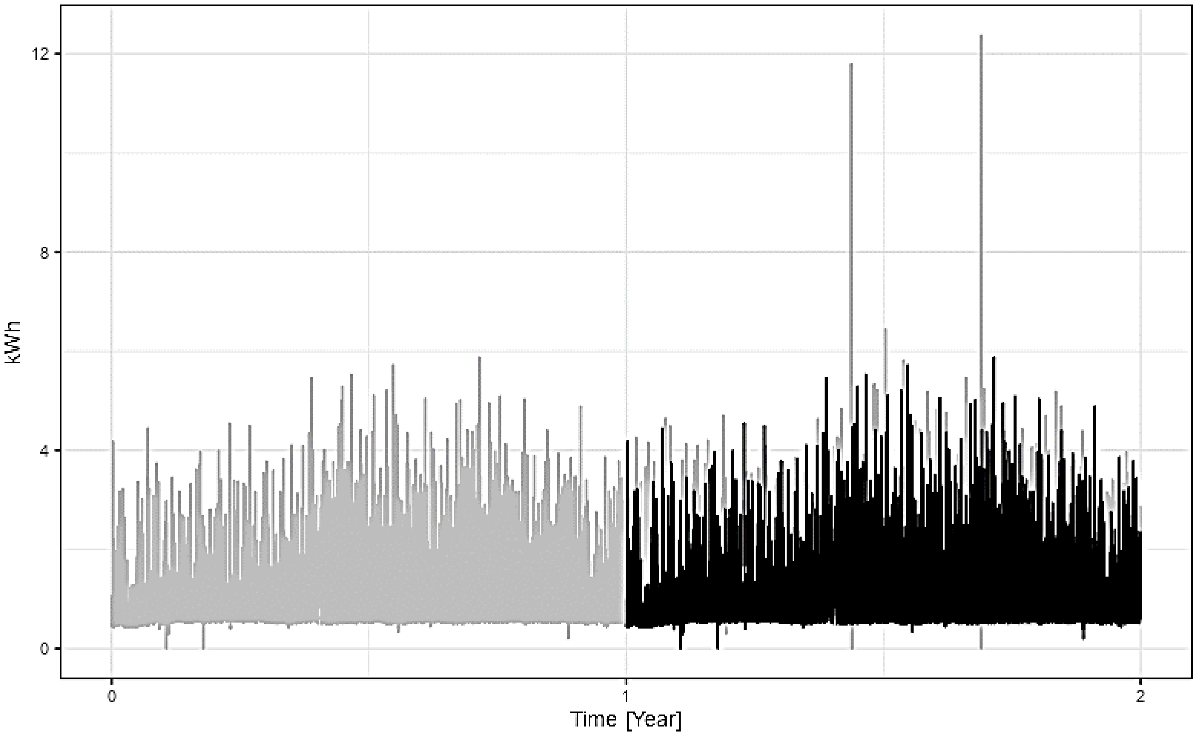

Figure 3. Actual (gray) and forecast (black) values of electricity consumption.

For the test set, forecasting errors were calculated as the difference between actual and forecast values. Figure 4 shows the time plot of forecasting errors, the autocorrelation function (ACF) and the partial autocorrelation function (PACF).

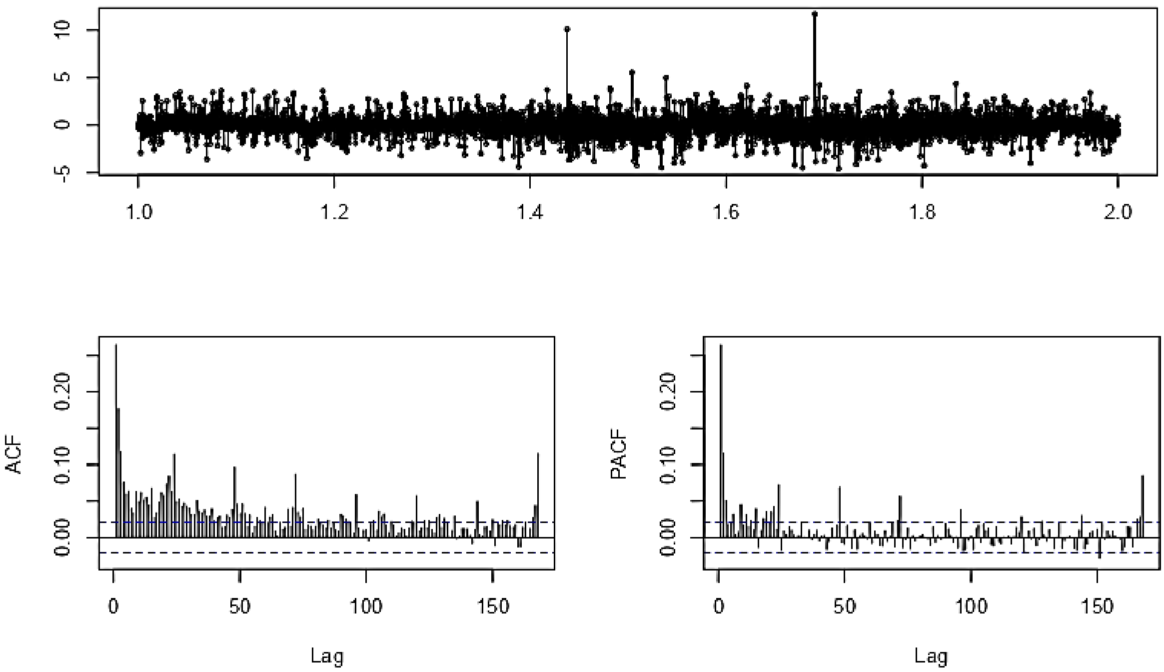

Figure 4. Forecasting error series and ACF and PACF charts.

In the ACF and PACF charts, one can see significant peaks for 24-hour delays (and multiples). One can also see a clearly greater autocorrelation for 168-hour delays. This confirms the presence of daily and weekly seasonality in the series.

Decomposition of the error series was performed using the MSTL method (Bandara et al., 2021). The individual components of the decomposed forecasting error series are shown in Figure 5.

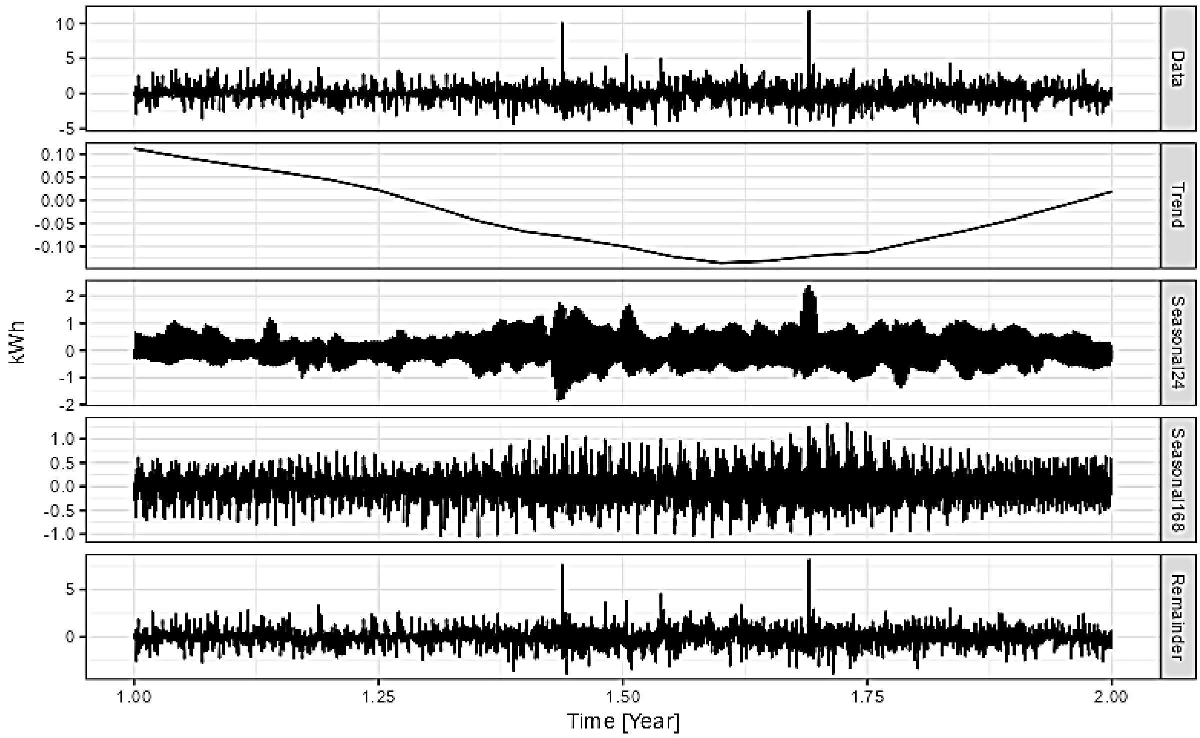

Figure 5. Error time series decomposition.

In Figure 5, the “Data” time series shows the development of forecasting errors. The following time series: “Trend,” “Seasonal24” and “Seasonal168” present the deterministic components of the forecasting error series. The “Remainder” component, on the other hand, should present the error series related to randomness, although technically it is the difference between the “Data” error series and its deterministic component.

By comparing the average forecasting errors counted for the “Data” series and those counted for the “Remainder” series, it is possible to assess how the elimination of the deterministic components of forecasting errors affects the quality of the forecast. The following errors were determined for the considered series and are shown in Table 2: Mean Absolute Error (MAE), Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE), Median Absolute Error (MdAE), Median Squared Error (MdSE), Median Absolute Percented Error (MdAPE).

Table 2.
Total forecasting errors and errors after elimination of deterministic components

Error	Time series	
	"Data"	"Remainder"
MAE	0.535	0.427
MSE	0.704	0.414
MAPE	50.4%	41.1%
MdAE	0.310	0.271
MdSE	0.096	0.074
MdAPE	30.4%	28.4%

Given the large size of the analyzed set (8,760 observations), it can be assumed that for all calculated measures, significantly lower values were obtained for the "Remainder" series.

The strength of the trend component determined according to formula (1) is 0.085. This means that the trend component in the forecasting error is very weak. For the seasonality components, considered together, the value of measure (2) is 0.377. For the component of daily seasonality (Seasonal24) is 0.274. While for weekly seasonality (Seasonal168) it is 0.182. This means that the seasonality in the analyzed series is weak. At the same time, daily seasonality is relatively stronger than weekly seasonality.

The components of the series are determined using the STL method (Cleveland et al., 1990). Consequently, the seasonal components are not regular. By averaging (using medians) the daily deviations (Seasonal24) and the weekly deviations (Seasonal168), the daily error profile and the weekly error profile can be obtained, as shown in Figure 6.

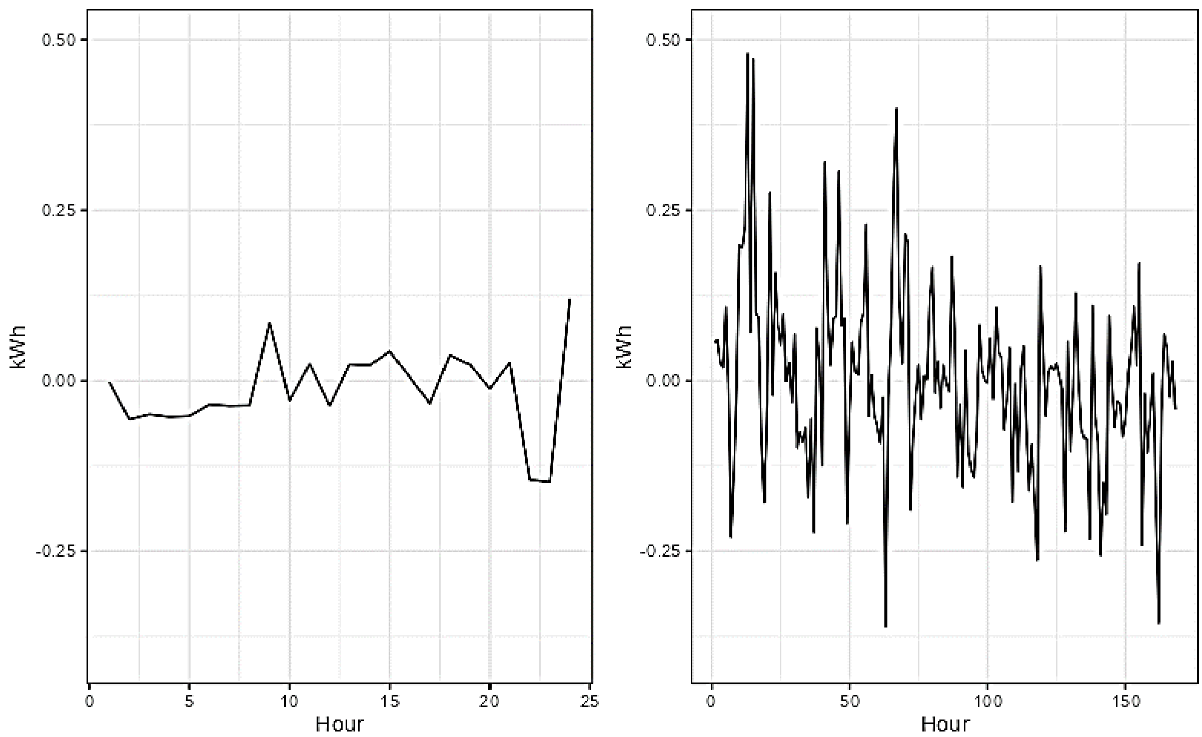

Figure 6. Median-averaged daily and weekly forecasting error profile.

The averaged profiles shown in Figure 6 present the average forecasting error for the corresponding hour during the day (for the daily profile) and during the week (for the weekly profile). Of interest from the analyst's point of view is whether there is regularity in the errors made. Analysis of the graphs and the course of errors over time indicates that there is no obvious regularity. The daily profile's randomness test (Wald, Wolfowitz, 1940) indicates that it can be considered random (statistic = -0.835, runs = 11, n1 = 12, n2 = 12, n = 24, p-value = 0.404). On the other hand, the results of the randomness test for the weekly profile (statistic = -3.250, runs = 64, n1 = 84, n2 = 84, n = 168, p-value = 0.001) indicate a lack of randomness. At the same time, it is difficult to pinpoint a clear reason for the lack of randomness in this series. However, in technical terms, it is a consequence of too many runs.

4. Discussion and conclusions

The paper proposes to study a range of forecasting errors using MSTL decomposition. The decomposition makes it possible to extract the deterministic components of the error, which can be used to correct the forecasting model.

Deterministic components show systematic forecasting error, which, when eliminated, can improve the quality of forecasts. This is shown by the values in Table 2. At the same time, it should be noted that forcing seasonality does not necessarily make it obvious. Thus, it is reasonable to study the strength of the relevant components of the time series in advance (Wang et al., 2006).

Accounting for multi-seasonality in the decomposition of a time series can also be done using harmonic analysis (De Livera et al., 2011). Preliminary consideration of the corresponding harmonics (Fourier analysis) can be an alternative to the proposed approach. Harmonics corresponding to daily and weekly seasonality will be regular. On the one hand, they will not take into account the change in seasonality over time. On the other, the determination of daily and weekly profiles will be relatively simple, unencumbered by variability.

Notably, the analyzed time series of forecasting errors was characterized by relatively weak seasonality. This fact may undermine the assumption of the presence of multi-seasonality in the errors or indicate a good forecasting model. Furthermore, one can consider whether decomposition with a single dominant seasonality could be sufficient. Then, the classical decomposition or STL decomposition could be used.

In summary, the proposed method can be particularly useful when:

- forecasts are subject to large errors,
- the predictive model is based on incomplete or aggregated data.

In general, decomposition study of a range of forecasting errors:

- involves breaking down the total forecast error into seasonal components and determining the contribution of individual systematic factors to the error,
- aims to identify the causes of forecasting errors and enable a better understanding of why the forecast was inaccurate,
- can be used to assess the quality of different forecasting methods and compare their effectiveness.

References

1. Ahn, Y., Kim, B.S. (2022). Prediction of building power consumption using transfer learning-based reference building and simulation dataset. *Energy and Buildings, 258,* 111717.
2. Anvari, M., Proedrou, E., Schäfer, B., Beck, C., Kantz, H., Timme, M. (2022). Data-driven load profiles and the dynamics of residential electricity consumption. *Nature Communications, 13(1)*, 4593.
3. Bandara, K., Hyndman, R.J., Bergmeir, C. (2021). MSTL: *A Seasonal-Trend Decomposition Algorithm for Time Series with Multiple Seasonal Patterns*. arXiv preprint arXiv:2107.13462.
4. Bargiel, J., Halinka, A., Niedopytalski, M. (2019). Analiza profili zużycia energii elektrycznej odbiorców w sieci niskiego napięcia. *Zeszyty Naukowe Wydziału Elektrotechniki i Automatyki Politechniki Gdańskiej, 63,* 51-54.
5. Cleveland, R.B., Cleveland, W.S., McRae, J.E., Terpenning, I.J. (1990). STL: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics, 6(1),* 3-33.
6. De Livera, A.M., Hyndman, R.J., Snyder, R.D. (2011). Forecasting time series with complex seasonal patterns using exponential smoothing. *J. American Statistical Association, 106(496),* 1513-1527.
7. Duarte, O.G., Rosero, J.A., Pegalajar, M.D. (2022). Data Preparation and Visualization of Electricity Consumption for Load Profiling. *Energies, 15(20),* 7557.
8. Ghalehkhondabi, I., Ardjmand, E., Weckman, G.R., Young, W.A. (2017). An overview of energy demand forecasting methods published in 2005-2015. *Energy Systems, 8*, 411-447.
9. Hyndman, R., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., O'Hara-Wild, M., Petropoulos, F., Razbash, S., Wang, E., Yasmeen, F. (2022). *_forecast: Forecasting functions for time series and linear models_. R package version 8.18*, https://pkg.robjhyndman.com/forecast/.
10. Hyndman, R.J., Athanasopoulos, G. (2018). *Forecasting: principles and practice.* Melbourne, Australia: OTexts. OTexts.com/fpp2, 2023-02-13.

11. Islam, M.A., Che, H.S., Hasanuzzaman, M., Rahim, N.A. (2020). *Energy demand forecasting. In Energy for sustainable development.* Academic Press, pp. 105-123.
12. Kowalska-Styczeń, A., Owczarek, T., Siwy, J., Sojda, A., Wolny, M. (2022). Analysis of Business Customers' Energy Consumption Data Registered by Trading Companies in Poland. *Energies, 15(14),* 5129.
13. Makonin, S. (2018). *HUE: The Hourly Usage of Energy Dataset for Buildings in British Columbia,* https://doi.org/10.7910/DVN/N3HGRN, Harvard Dataverse, V5, UNF:6:F2ursIn9woKDFzaliyA5EA== [fileUNF].
14. Makonin, S. (2019). HUE: The Hourly Usage of Energy Dataset for Buildings in British Columbia. *Data in Brief, vol. 23, no. 103744*, pp. 1-4.
15. Nizami, S.J., Al-Garni, A. (1995). Forecasting electric energy consumption using neural networks. *Energy policy, 23(12),* 1097-1104.
16. R Core Team (2022). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.
17. Shao, Z., Chao, F., Yang, S.L., Zhou, K.L. (2017). A review of the decomposition methodology for extracting and identifying the fluctuation characteristics in electricity demand forecasting. *Renewable and Sustainable Energy Reviews, 75*, 123-136.
18. Wald, A., Wolfowitz, J. (1940). On a test whether two samples are from the same population. *The Annals of Mathematical Statistics 11,* 147-162. doi:10.1214/aoms/1177731909.
19. Wang, X., Smith, K.A., Hyndman, R.J. (2006). Characteristic-based clustering for time series data. *Data Mining and Knowledge Discovery, 13(3),* 335-364.
20. Yang, Y., Bremner, S., Menictas, C., Kay, M. (2022). Forecasting error processing techniques and frequency domain decomposition for forecasting error compensation and renewable energy firming in hybrid systems. *Applied Energy, 313*, 118748.