

Analiza i prognozowanie szeregów czasowych o złożonej sezonowości

Case study: Prognozowanie zapotrzebowania mocy
Krajowego Systemu Energetycznego (KSE)

Adam Zagdański

QUANTUP

<http://www.quantup.pl>

12 marca 2014

Spis treści

1 Szeregi czasowe o złożonej sezonowości	2
2 Modelowanie złożonej sezonowości – klasyczne i nowe metody	2
3 Prognozowanie zapotrzebowania mocy KSE – porównanie efektywności metod	2
3.1 Dane	2
3.2 Cel analizy	5
3.3 Metody	5
3.4 Wyniki	6
STL – metoda nie uwzględniająca złożonych sezonowości	6
TBATS – metoda uwzględniająca sezonowość dobową i tygodniową	7
Porównanie metod	8
3.5 Podsumowanie	11
4 Bibliografia	11

1 Szeregi czasowe o złożonej sezonowości

Wiele spotykanych w praktyce szeregów czasowych cechuje występowanie niestandardowych i często skomplikowanych wzorców sezonowych (sezonowości). Typowe sytuacje z jakimi możemy się spotkać to:

- **Szeregi o ułamkowym (niecałkowitym) okresie**
Przykłady: szeregi zawierające tygodniowe wielkości produkcji lub wielkości sprzedaży wykazują często sezonowość roczną o okresie $365.25/7 \approx 52.18$.
- **Szeregi o dużej częstotliwości i wielookresowej sezonowości**
Przykłady: dane godzinowe dotyczące zapotrzebowania na energię elektryczną, gaz, wodę, itp. (*utility demand*) charakteryzuje zwykle sezonowość dobową i tygodniową. Obserwowane w dłuższym okresie szeregi te wykazują często także sezonowość roczną.

2 Modelowanie złożonej sezonowości – klasyczne i nowe metody

Większość standardowych metod, stosowanych w analizie i prognozowaniu szeregów czasowych, opracowano z myślą o prostych wzorcach wahań sezonowych występujących w danych. Metody te są zwykle skuteczne w przypadku sezonowości o małym (całkowitym) okresie, takim jak $s = 12$ dla danych miesięcznych, lub $s = 4$ dla szeregów kwartalnych. Ponadto, standardowe metody nie uwzględniają poprawnie takich specyficznych własności danych, jak występowanie sezonowości o nakładających się okresach (np. sezonowości dziennej i tygodniowej).

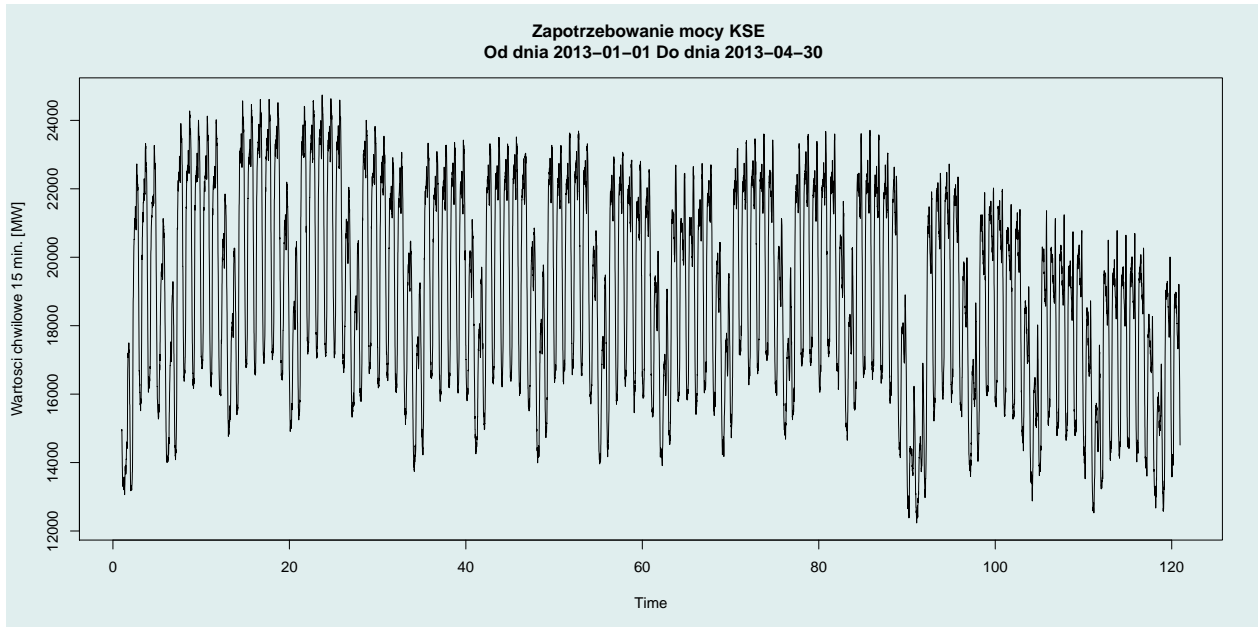
Istnieją również specjalistyczne modele i metody dedykowane szeregom o skomplikowanych wzorcach sezonowości. Narzędzia te są istotną alternatywą dla klasycznych metod, a prace nad ich udoskonalaniem ciągle trwają (np. kilka najnowszych propozycji można znaleźć w artykule De Livera i inni (2011)). Z uwagi na złożoność zagadnienia, istotny jest jednak wybór odpowiedniego wariantu metody i odpowiednie przygotowanie danych przez analityka.

3 Prognozowanie zapotrzebowania mocy KSE – porównanie efektywności metod

3.1 Dane

W analizie wykorzystamy szereg czasowy zawierający wielkości zapotrzebowania mocy Krajowego Systemu Energetycznego (KSE) w okresie: od 2013-01-01 do 2013-04-30. Dane są dostępne na stronie PSE (Polskie Sieci Elektroenergetyczne S.A.) pod adresem <http://www.pse-operator.pl/index.php?dzid=77>. Szereg ten zawiera wartości chwilowe, rejestrowane w mega watach [MW], w odstępach co 15 min. Łącznie mamy więc wartości zapotrzebowania mocy dla kolejnych 120 dni (4 miesięcy) bieżącego roku.

Wykres danych surowych przedstawia Rys. 1. Przyjęta jednostka czasu (oś X) to 1 dzień. Oprócz charakterystycznych efektów sezonowych, na wykresie łatwo zauważamy także wyraźny spadek zapotrzebowania na początku roku (1 stycznia 2013) oraz w okresie Świąt Wielkanocnych (1 kwietnia 2013).



Rysunek 1: Zapotrzebowanie mocy KSE od dnia 2013-01-01 do dnia 2013-04-30. Dane chwilowe 15 min. (w MW).

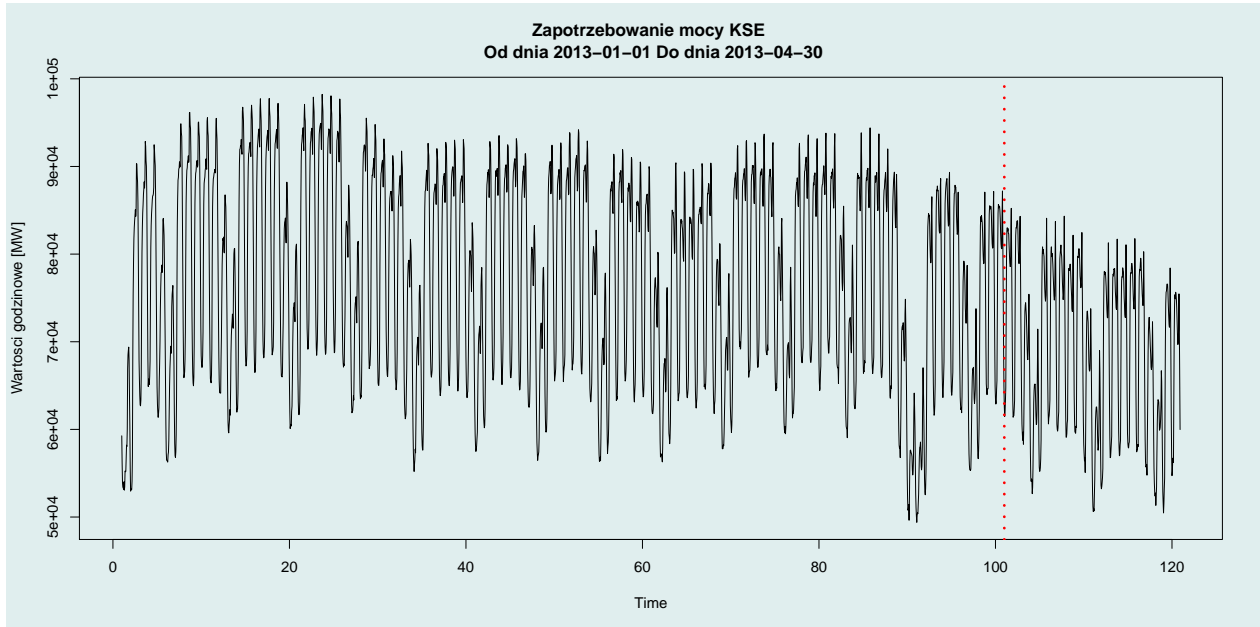
Przed przystąpieniem do dalszej analizy przeprowadzimy agregację do danych godzinowych. Otrzymujemy w ten sposób szereg zawierający $n = 2879$ wartości ($120 \text{ dni} \times 24 \text{h}$). Aby porównać skuteczność różnych metod prognozowania, podzielimy dane na dwie części (Rys. 2):

- **część konstrukcyjną** – wartości dla pierwszych 100 dni (począwszy od 1 stycznia 2013), które wykorzystamy do kalibracji modeli; szereg zawierający 2400 wartości,
- **część testową** – pozostałe wartości dla 20 dni; szereg zawierający 479 wartości.

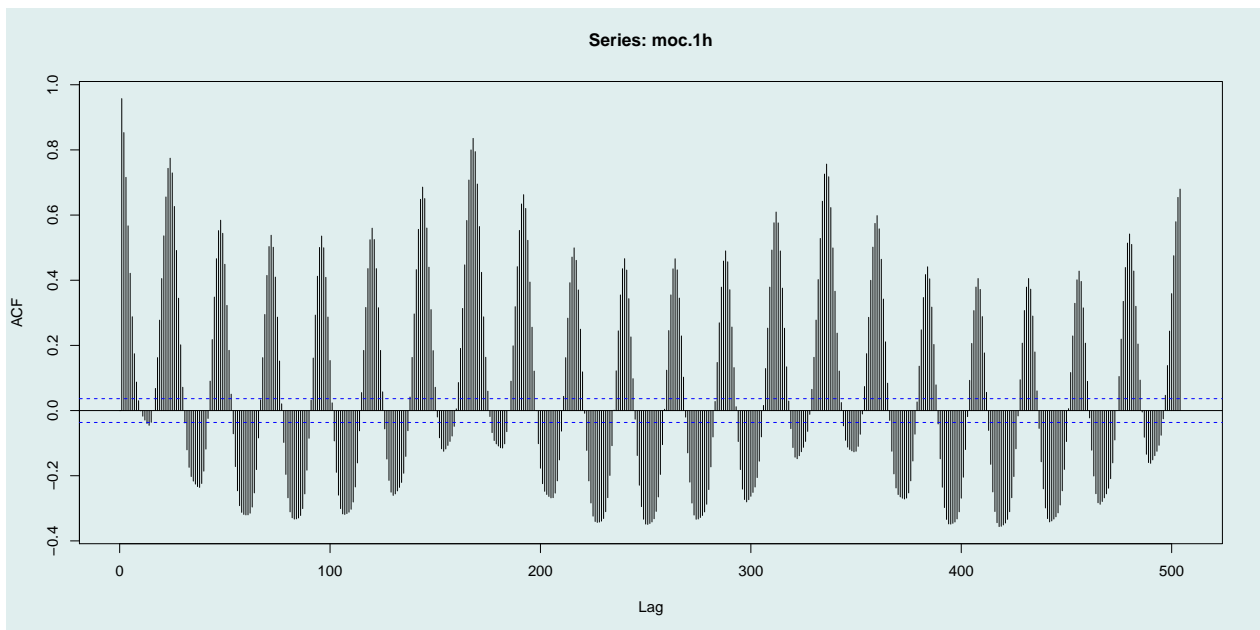
Aby przyjrzeć się zależnościom występującym w naszych danych, wyznaczmy wykres funkcji autokorelacji (ACF) (Rys. 3). Analizując wykres wyraźnie dostrzegamy sezonowości o dwóch okresach:

- $s_1 = 24$ – okres odpowiadający sezonowości dobowej,
- $s_2 = 7 * 24 = 168$ – okres odpowiadający sezonowości tygodniowej.

Analizując szereg zapotrzebowania na energię w dłuższym okresie (np. kilku lat) prawdopodobnie mielibyśmy do czynienia również z sezonowością roczną.



Rysunek 2: Dane po agregacji godzinowej. Zaznaczony podział na część konstrukcyjną i część testową.



Rysunek 3: Wykres autokorelacji (ACF) dla danych godzinowych. Widoczna sezonowość dobową (okres: $s_1 = 24$) oraz tygodniową (okres: $s_2 = 168$).

3.2 Cel analizy

W analizie danych związanych z zapotrzebowaniem na energię elektryczną celem może być konstrukcja prognoz krótko-, średnio- lub długoterminowych. Często ważna jest także identyfikacja składowych związanych z poszczególnymi rodzajami sezonowości (dobowej, tygodniowej czy rocznej) oraz identyfikacja pozostałych charakterystycznych własności danych (np. obecność trendów długoterminowych, wpływ świąt, "blackoutów", itp.).

W naszej analizie skoncentrujemy się na:

- ilustracji działania wybranych metod analitycznych, dostępnych w środowisku R,
- przykładach zastosowania metod do dekompozycji oraz prognozowania,
- porównaniu skuteczności wykorzystywanych metod.

3.3 Metody

W analizie wykorzystamy następujące metody:

- **metoda STL** – Seasonal Decomposition of Time Series by Loess (Cleveland i inni (1990)).

Zaawansowana metoda dekompozycji i prognozowania szeregów, stosowana w przypadku danych o *standardowej sezonowości* (np. miesięcznej lub kwartalnej).

- **metoda TBATS** – Trigonometric BATS (Box–Cox transform, ARMA errors, Trend, Seasonal components) (De Livera i inni (2011)).

Zaawansowana metoda dekompozycji i prognozowania szeregów, opracowana dla szeregów o *złożonej (wielookresowej) sezonowości*. Algorytm pozwala na uwzględnienie dodatkowych własności danych, takich jak obecność trendów długoterminowych czy niejednorodna wariancja w czasie.

Do porównania dokładności prognoz wykorzystamy następujące kryteria:

- **MAPE** – Mean Absolute Percentage Error (średni bezwzględny błąd procentowy, wyrażony w %):

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{X_t - F_t}{X_t} \right| \cdot 100\%,$$

gdzie: X_t - rzeczywista wartość zapotrzebowania, F_t - prognoza.

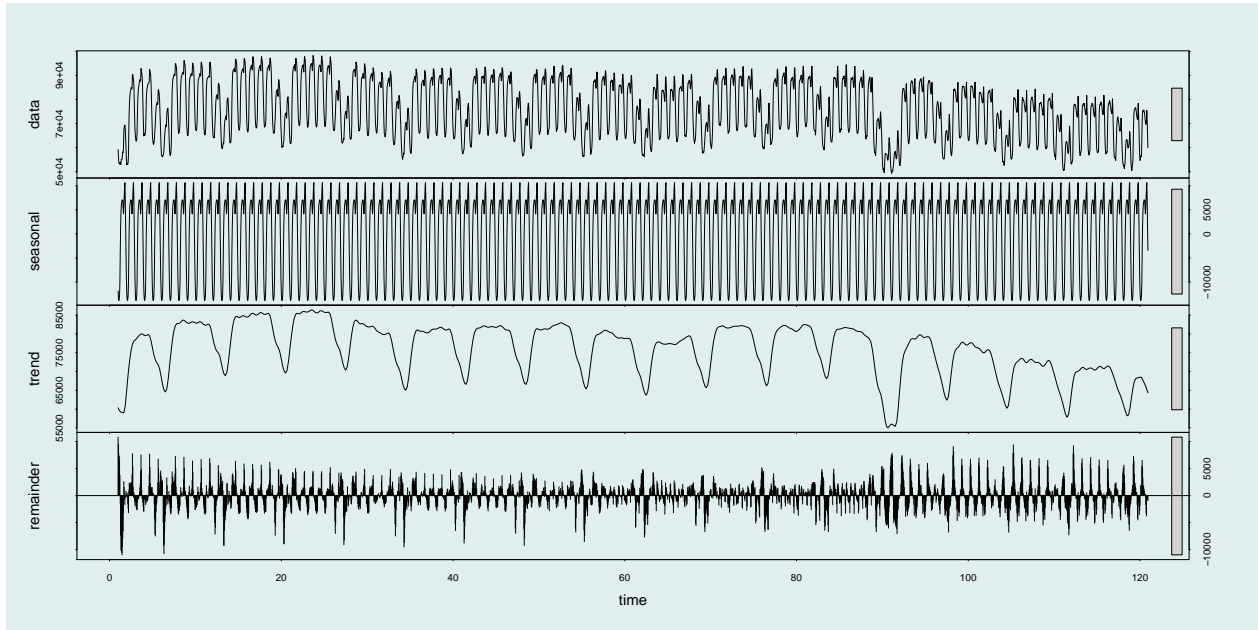
- **RMSE** – Root Mean Squared Error (pierwiastek błędu średniokwadratowego):

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (X_t - F_t)^2}{n}}. \quad (1)$$

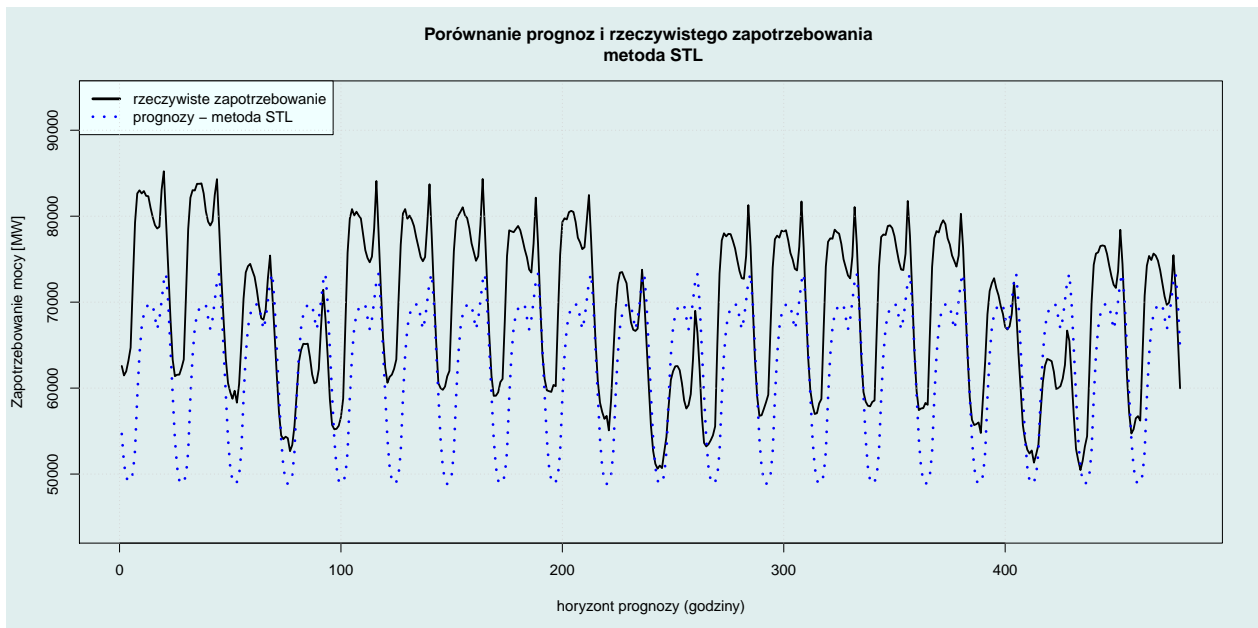
W analizie wykorzystano metody i narzędzia dostępne dla środowiska R (www.r-project.org).

3.4 Wyniki

STL – metoda nie uwzględniająca złożonych sezonowości



Rysunek 4: Dekompozycja na bazie metody STL.

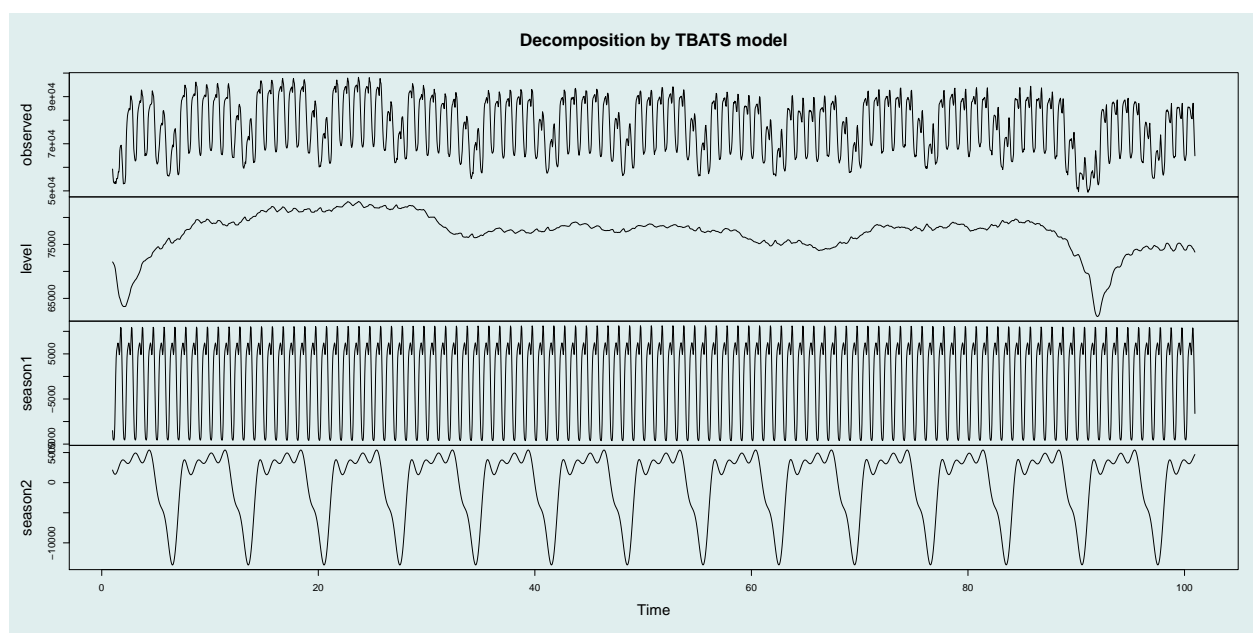


Rysunek 5: Prognozy na bazie metody STL.

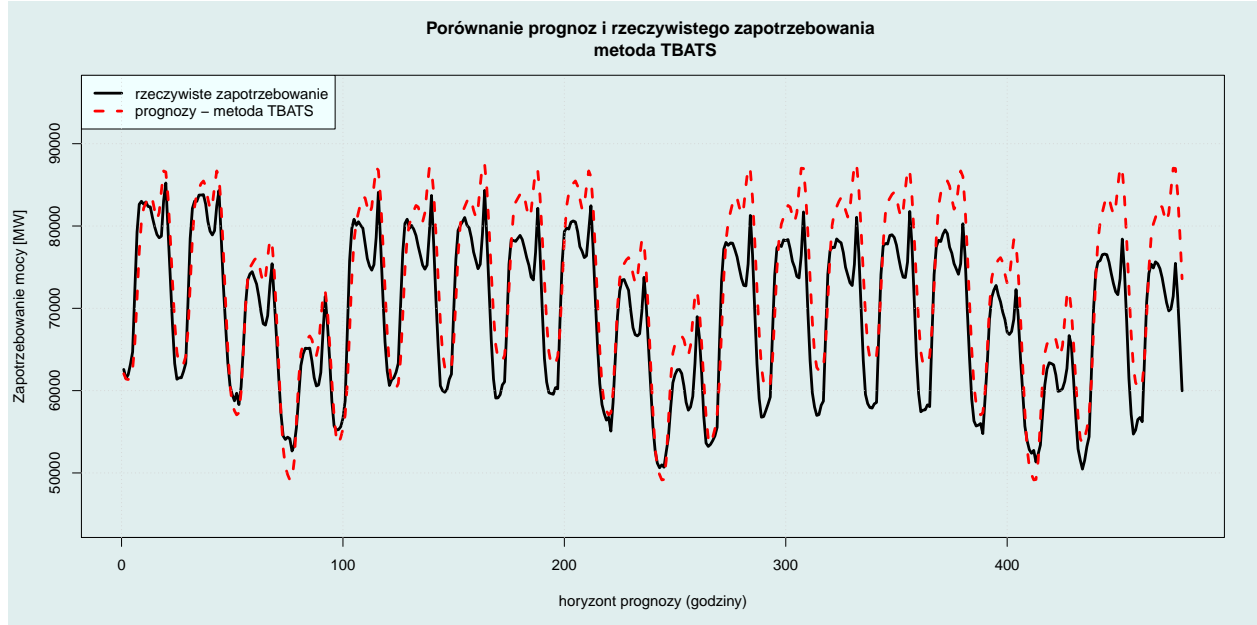
STL – Wnioski

- Analizując wyniki dekompozycji STL (Rys. 4) widzimy, że uwzględnienie tylko pojedynczej sezonowości (sezonowości dobowej) powoduje, że w składowej trendu (tendencji długoterminowej) widać niepożądane zachowania sezonowe, związane z sezonowością tygodniową.
- Uzyskane na bazie STL prognozy (Rys. 5) są również bardzo złej jakości, nawet dla krótkiego horyzontu. W większości okresów prognozowane wartości są dużo niższe od rzeczywistego zapotrzebowania.
- STL – metoda, która nie uwzględnia sezonowości o dwóch nakładających się okresach – nie daje zadowalających wyników.

TBATS – metoda uwzględniająca sezonowość dobową i tygodniową



Rysunek 6: Dekompozycja na bazie metody TBATS.



Rysunek 7: Prognozy na bazie metody TBATS.

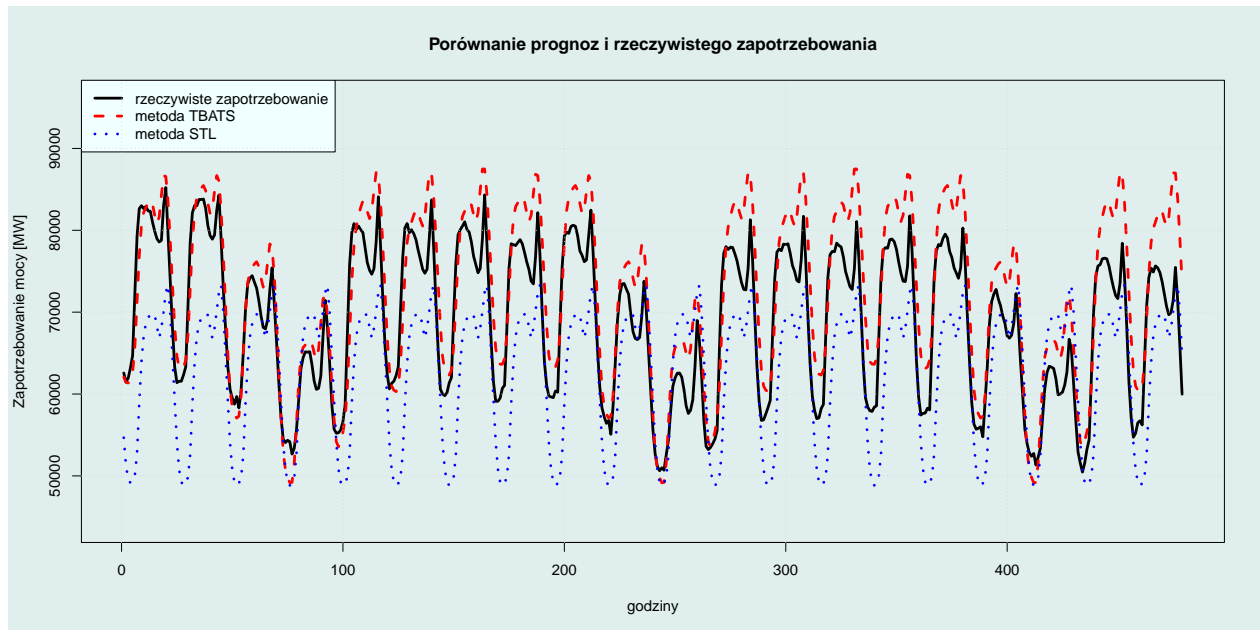
TBATS – wnioski

- Dekompozycja uzyskana za pomocą metody TBATS (Rys. 6) pozwala łatwo zidentyfikować regularności występujące w analizowanych danych.
- Oprócz sezonowości dobowej (season 1) i tygodniowej (season 2) na wykresie składowej *level* łatwo widać wyraźny spadek zapotrzebowania w okresie świąt (1 stycznia oraz 1 kwietnia). Sugeruje to konieczność uzupełnienia modelu o dodatkowe informacje, np. w postaci zmiennych 0-1 kodujących występowanie nietypowych zdarzeń, które mogą wpływać na wielkość zapotrzebowania na energię.
- Uzyskane na bazie metody TBATS prognozy (Rys. 7) są zadowalające, szczególnie w krótkim horyzoncie (prognozy dla około 100 kolejnych godzin).
- Możemy spodziewać się, że dokładność prognoz można by jeszcze poprawić uwzględniając w analizie dodatkowe własności danych, takie jak występowanie świąt.

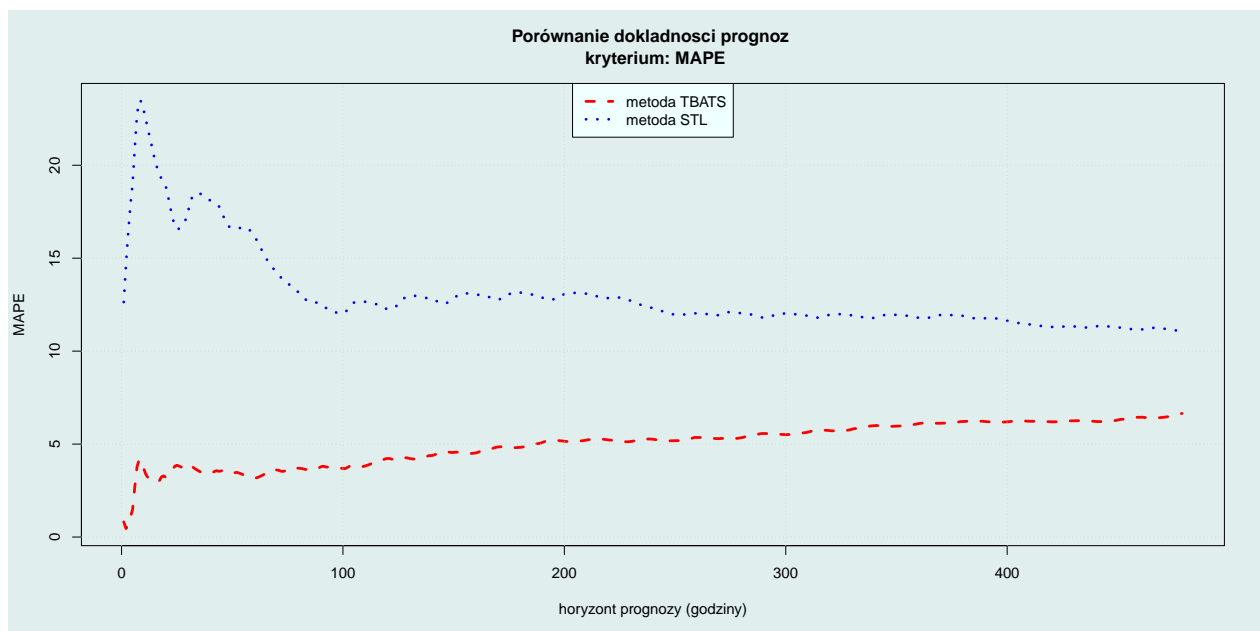
Porównanie metod

Porównajmy jeszcze obie metody na jednym wykresie (Rys. 8). Wyraźnie widzimy przewagę metody TBATS w stosunku do metody STL.

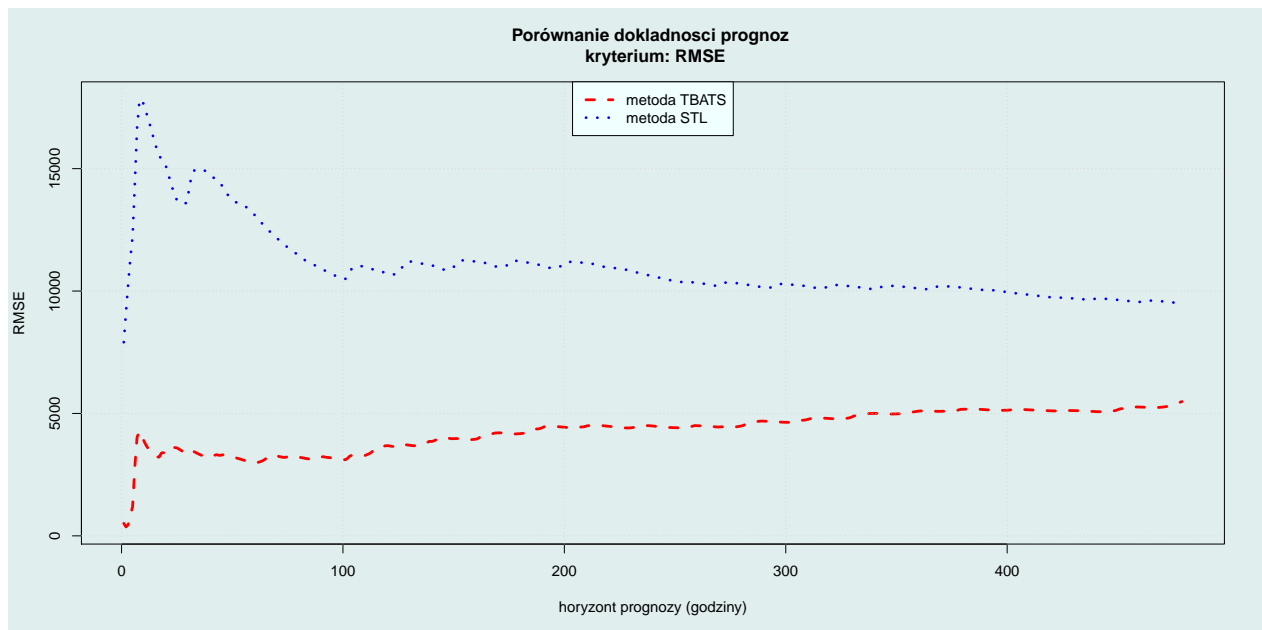
Rysunki 9 i 10 przedstawiają porównanie błędów prognoz (MAPE i RMSE) dla obu metod i różnych horyzontów prognoz. Ponownie widzimy wyraźną przewagę metody TBATS.



Rysunek 8: Porównanie prognoz i rzeczywistego zapotrzebowania (wszystkie metody).



Rysunek 9: Porównanie dokładności prognoz – kryterium MAPE



Rysunek 10: Porównanie dokładności prognoz – kryterium RMSE

3.5 Podsumowanie

- Metoda STL – zaawansowana metody dekompozycji, uwzględniająca tylko sezonowość dobową prowadzi do złych prognoz. Prognozowane wartości są dla większości okresów zdecydowanie zaniżone, w porównaniu do rzeczywistego zapotrzebowania.
- Metoda TBATS, która bezpośrednio uwzględnia nakładające się sezonowości (dobową i tygodniową) daje zadowalające prognozy.
- Oba kryteria (MAPE i RMSE) wskazują na dobrą jakość prognoz dla metody TBATS w krótkim horyzoncie (prognozy zapotrzebowania na około 100 kolejnych godzin). Dokładność prognoz maleje wraz ze wzrostem horyzontu prognozy.
- Metoda TBATS pozwala na łatwą identyfikację regularnych tendencji występujących w analizowanych danych, w tym wyodrębnienie obu składowych sezonowych (dobowej i tygodniowej).
- Dokładność prognoz można by poprawić uwzględniając dodatkowe własności danych, takie jak występowanie spadku zapotrzebowania w okresie świąt (np.: w okresie Nowego Rok i Świąt Wielkanocnych).

4 Bibliografia

- [1] Strona domowa pakietu R, *The R Project for Statistical Computing*, www.r-project.org.
- [2] De Livera, A.M., Hyndman, R.J., & Snyder, R. D. (2011), *Forecasting time series with complex seasonal patterns using exponential smoothing*, *Journal of the American Statistical Association*, 106(496), 1513-1527.
- [3] R. B. Cleveland, W. S. Cleveland, J.E. McRae, and I. Terpenning (1990), *STL: A Seasonal-Trend Decomposition Procedure Based on Loess*. *Journal of Official Statistics*, 6, 3–7.