

Czasowo–częstotliwościowe metody analizy i syntezy sygnałów

Projekt 2019 L

Predykcja AR

Igor Markiewicz

Prowadzący – dr inż. Rafał Piotr Rytel–Andrianik

Spis treści

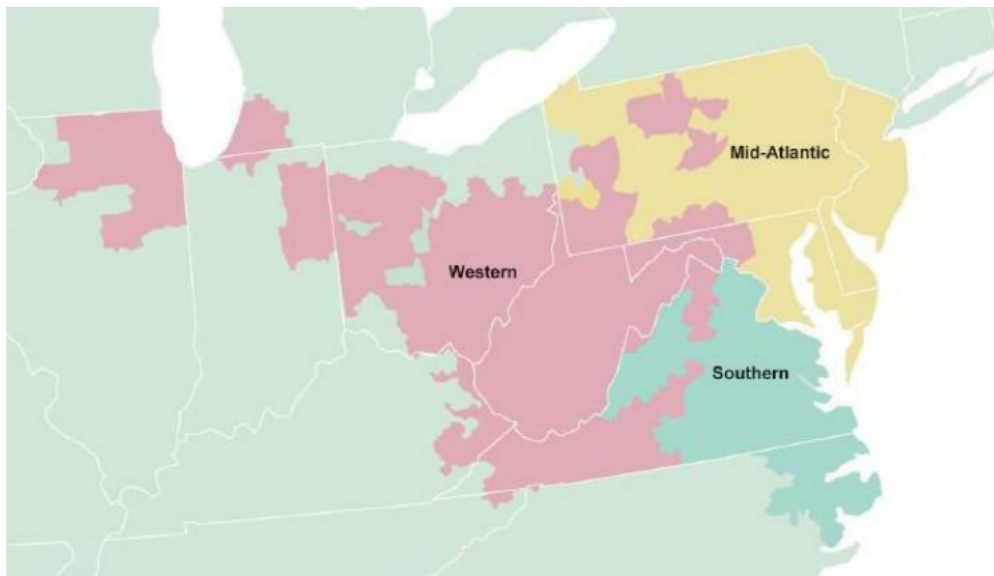
1	Opis zadania	2
2	Opis danych	2
3	Analiza danych	2
3.1	Wstępna analiza	3
3.2	Model AR	5
3.2.1	Opis modelu AR	5
3.2.2	Opis procedury testowej	6
3.2.3	Wyniki i ich analiza	7
4	Bibliografia	11

1 Opis zadania

Głównym celem projektu jest analiza szeregu czasowego (sygnału) przy użyciu modelu AR wraz z doбором odpowiedniego rzędu modelu wg. kryterium AIC. Został również przeprowadzony szereg innych analiz mający za zadanie uzupełnić model AR.

2 Opis danych

Jako przykładowe dane został wybrany zbiór zawierający godzinowe zapotrzebowanie na energię elektryczną w megawatach przez zachodni region *PJM Interconnection LLC* [1, 2] (z lat 2001 – 2018) – jednego z dostawców energii elektrycznej na wschodnim wybrzeżu Stanów Zjednoczonych.



Rys. 1: Podział regionów *PJM Interconnection LLC* [3]

Zdecydowano się na podjęcie dwóch działań względem oryginalnego zbioru danych zanim zostanie użyty on do testów:

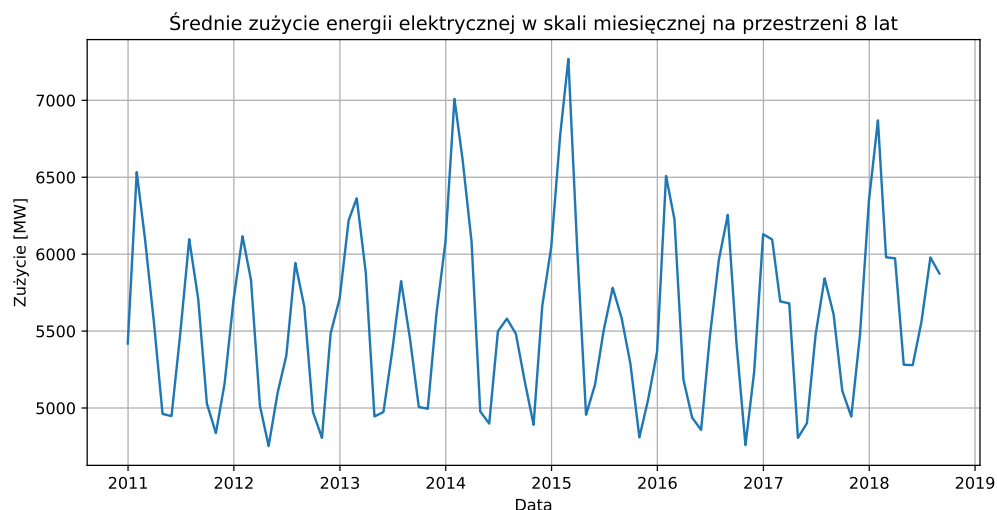
- obcięcie danych do ostatnich 8 lat
- uśrednienie danych z częstotliwością miesięczną zamiast godzinową (dane zawierają znaczniki datowo–czasowe)

W efekcie otrzymano 93 próbki.

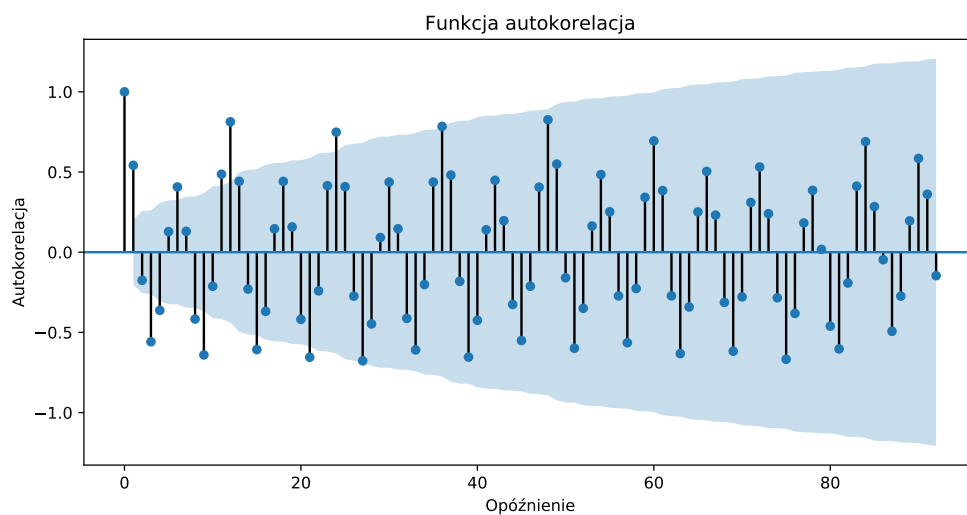
3 Analiza danych

W niniejszym punkcie została przeprowadzona analiza danych obejmująca zarówno model AR jak i inne techniki mające na celu jego uzupełnienie. Jako język został wybrany *Python 3* wraz z różnymi pakietami w tym *statsmodels* [4] zapewniającym m.in predykcję AR.

3.1 Wstępna analiza

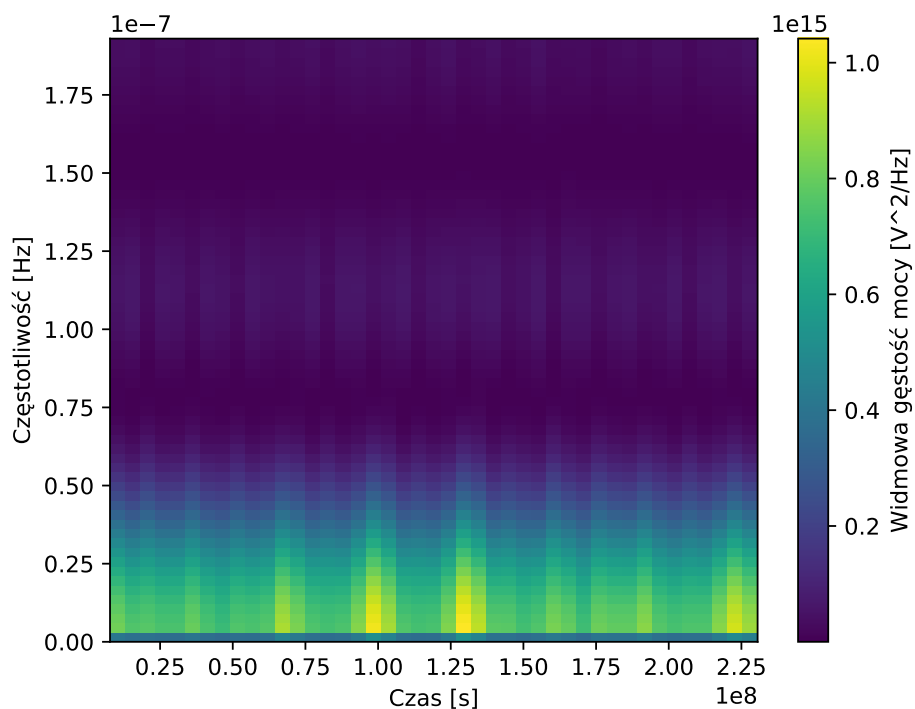


Rys. 2: Badany szereg czasowy



Rys. 3: Wykres autokorelacji danych (z poźnieniem miesięcznym). Kolorem jasnoniebieskim został oznaczony 95 % przedział ufności.

Możemy zauważyć że autokorelacja jest okresowo, wolno zanikająca co odpowiada quasi okresowej naturze danych – w miesiącach zimowych oraz letnich istnieje największe zapotrzebowanie na energię elektryczną. Odpowiada za to najprawdopodobniej zwiększony pobór mocy wynikający odpowiednio z ogrzewania i chłodzenia.

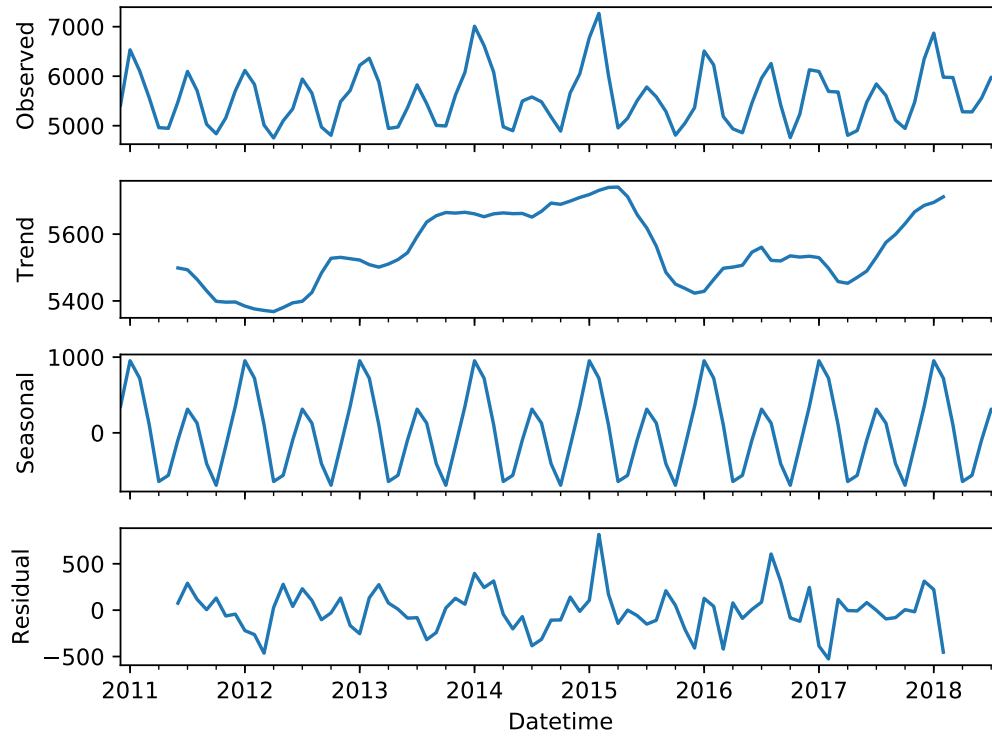


Rys. 4: Spektrogram

Na Rys. 4 przedstawiony został spektrogram z następującymi parametrami:

- $f_s = 3.86 \cdot 10^{-7} \text{ s}$ (uśrednienie miesięczne)
- $nfft = 128$
- okno – *tukey*(0.25)
- długość okna = 6
- ilość zachodzących punktów między kolejnymi przesunięciami okna = 4

Możemy na nim zaobserwować dominację głównie niskich częstotliwości oraz segmenty czasu z wyraźnie większą widmową gęstością mocy. Odpowiadają one przedziałom czasu z największymi zmianami zużyciami energii.



Rys. 5: Dekompozycja sygnału na trend, zmiany okresowe oraz residuum (w modelu addytywnym).

Dekompozycja sygnału na składowe jest przeprowadzana w trzech krokach:

1. Zostaje przeprowadzane wygładzanie sygnału z użyciem filtru i konwolucji w efekcie czego otrzymujemy trend
2. Od sygnału zostają odjęte trend w efekcie czego otrzymujemy składową sezonową. Składowa ta jest następnie dzielona na okresy które są ze sobą uśredniane w efekcie czego otrzymujemy jeden uśredniony okres (powtarzany potem wielokrotnie).
3. residuum powstaje jako różnica między obserwowanym przebiegiem oraz trendem i składową sezonową

3.2 Model AR

3.2.1 Opis modelu AR

Modelem AR rzędu P nazwiemy [5, 6]

$$x(n) = const - \sum_{k=1}^P a_k x(n-k) + u(n) \quad (1)$$

gdzie $u(n)$ jest szumem białym najczęściej zakładanym jako normalny $\mathcal{N}(0, \sigma^2)$. Składową stałą $const$ można włączyć do szumu $u(n)$ otrzymując wzór w bardziej zwartej postaci. Zadaniem jest

estymacja parametrów $const, a_1, a_2, \dots, a_P, \sigma$, w przypadku projektu zrealizowana przez metodę *Maximum Likelihood Estimation* z solverem *Broyden-Fletcher-Goldfarb-Shanno*. Korzystając z transformaty \mathcal{Z} po włączeniu składowej stałej do czynnika szumowego oraz traktując $u(n)$ jako wejście filtru liniowego można otrzymać jego transmitancję:

$$H(z) = \frac{1}{1 + \sum_{k=1}^P a_k z^{-k}} \quad (2)$$

Warunkiem stabilności powyższego filtru jest aby wszystkie pierwiastki wielomianu:

$$z^P + a_1 z^{P-1} + a_2 z^{P-2} + \dots + a_P = 0 \quad (3)$$

leżały wewnątrz okręgu jednostkowego na płaszczyźnie zespolonej.

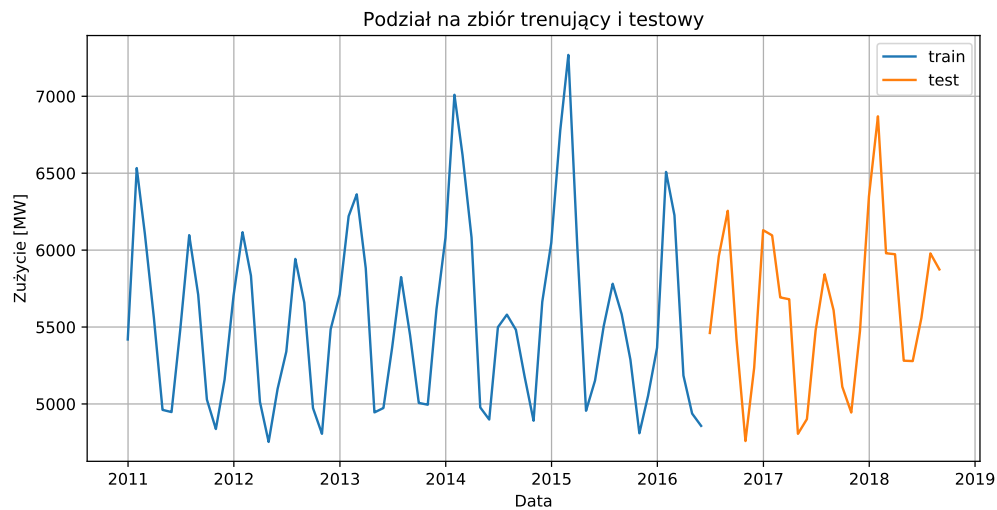
3.2.2 Opis procedury testowej

Procedura testowa miała następujący przebieg:

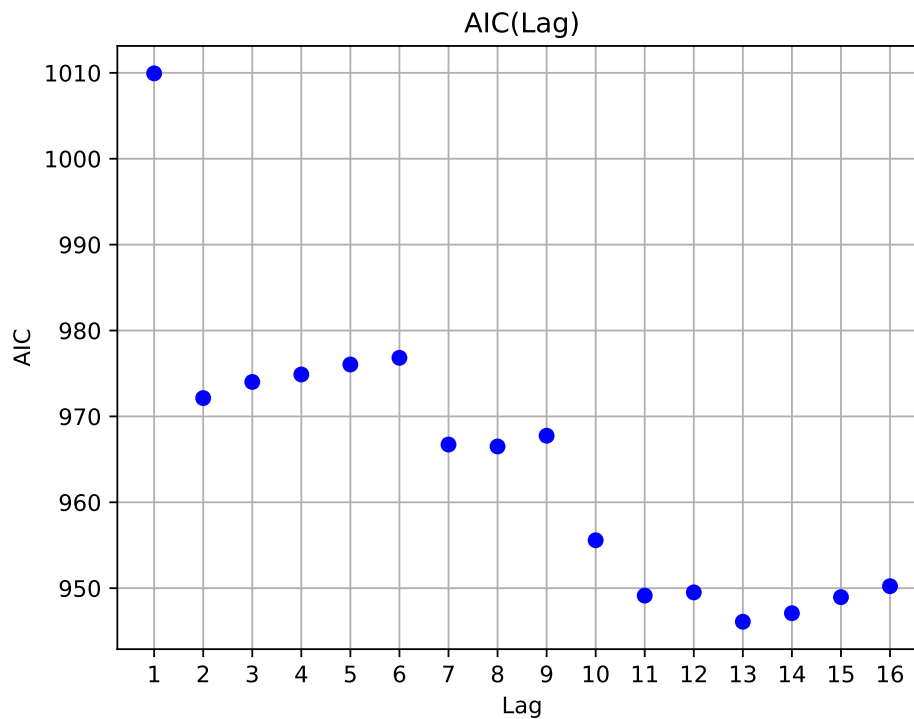
1. Podział na zbiór uczący i testowy - odpowiednio $\approx 70\%$ i $\approx 30\%$
2. Na zbiorze uczącym została przeprowadzona optymalizacja rzędu modelu wg. kryterium AIC
3. Na zbiorze testowym zastosowano rząd modelu wybrany w poprzednim punkcie a następnie przeprowadzono procedurę dynamicznego uczenia-testowania: cyklicznie następowało uczenie modelu na zbiorze o długości zbioru uczącego, predykcja jednego przykładu a następnie przesunięcie okna o jeden element w przód po czym cała procedura się powtarzała

W trakcie procesu została zastosowana metoda wymuszenia stacjonarności modelu (wbudowana w wykorzystywaną bibliotekę) przez transformację parametrów oparta o [7].

3.2.3 Wyniki i ich analiza

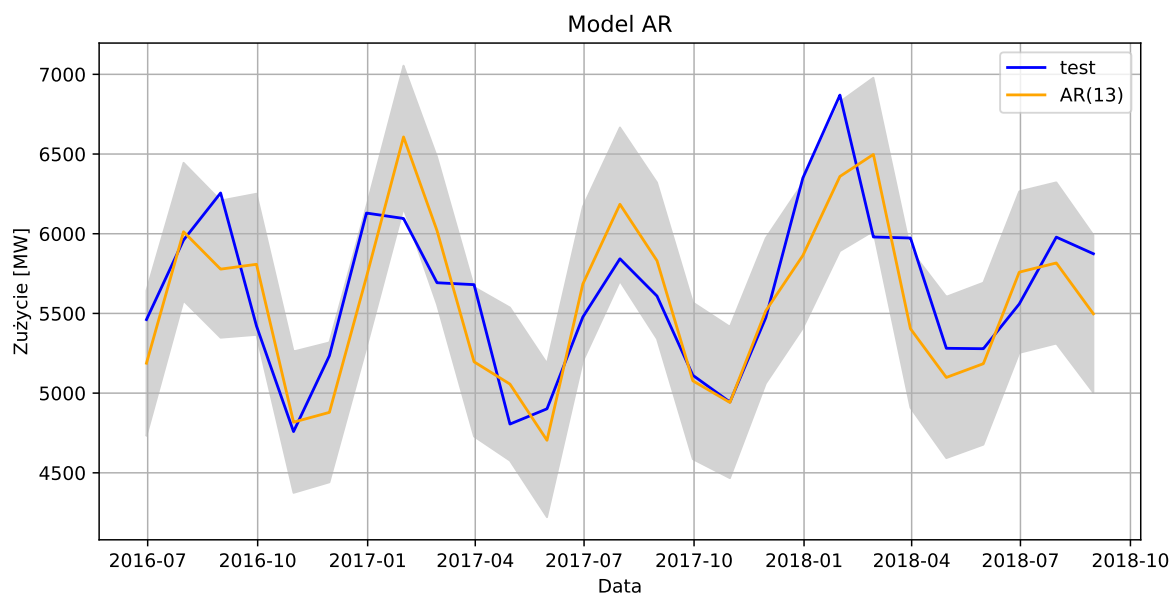


Rys. 6: Podział na zbiór uczący i testowy

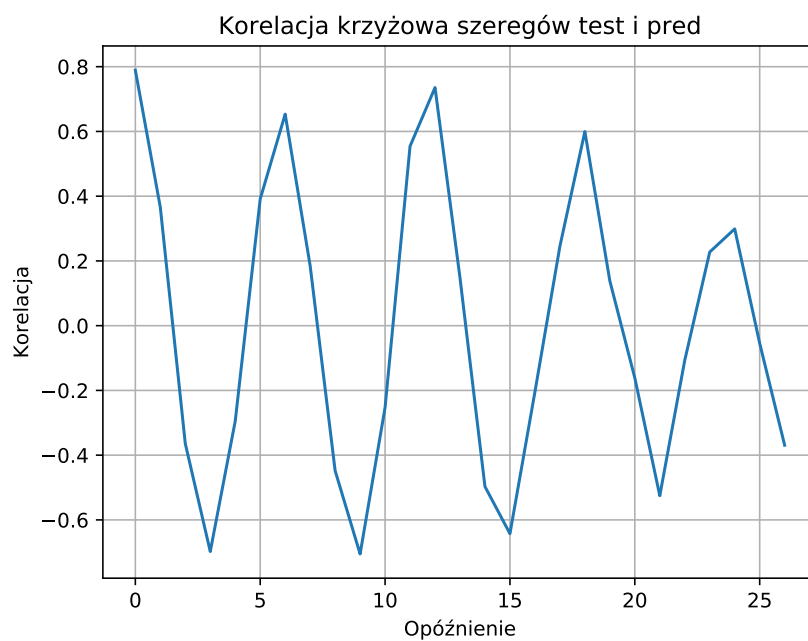


Rys. 7: Wartość AIC w zależności od rzędu modelu

Możemy zauważyć że w rozważanym zakresie rzędów modelu najlepsze rezultaty zostały osiągnięte dla liczby 13.

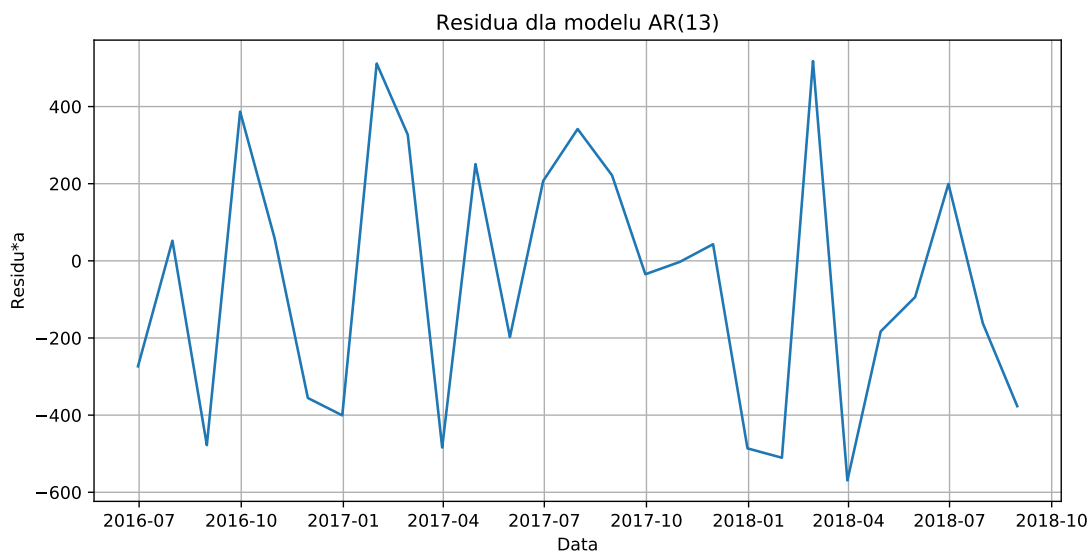


Rys. 8: Porównanie prawdziwych i przewidywanych wartości na zbiorze testowym. Szarym kolorem zostały zaznaczone 95 % przedziały ufności modelu.

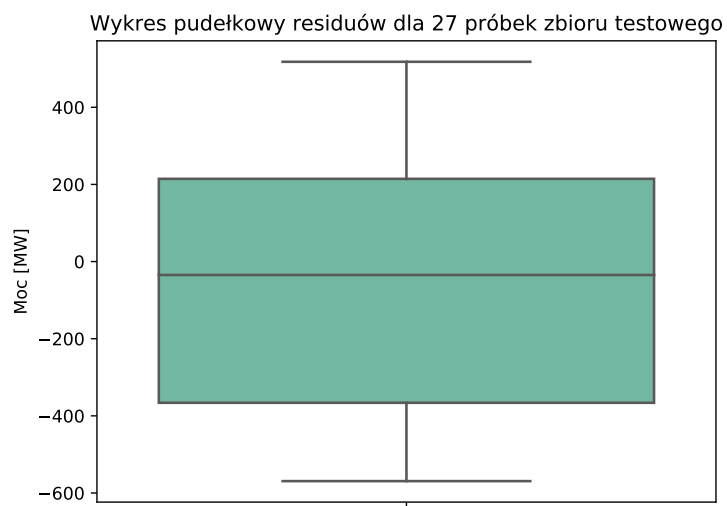


Rys. 9: Autokorelacja predykcji i wartości prawdziwych.

Na Rys. 8 możemy zauważyć że system faktycznie wyuczał się wartości pseudo okresowych.

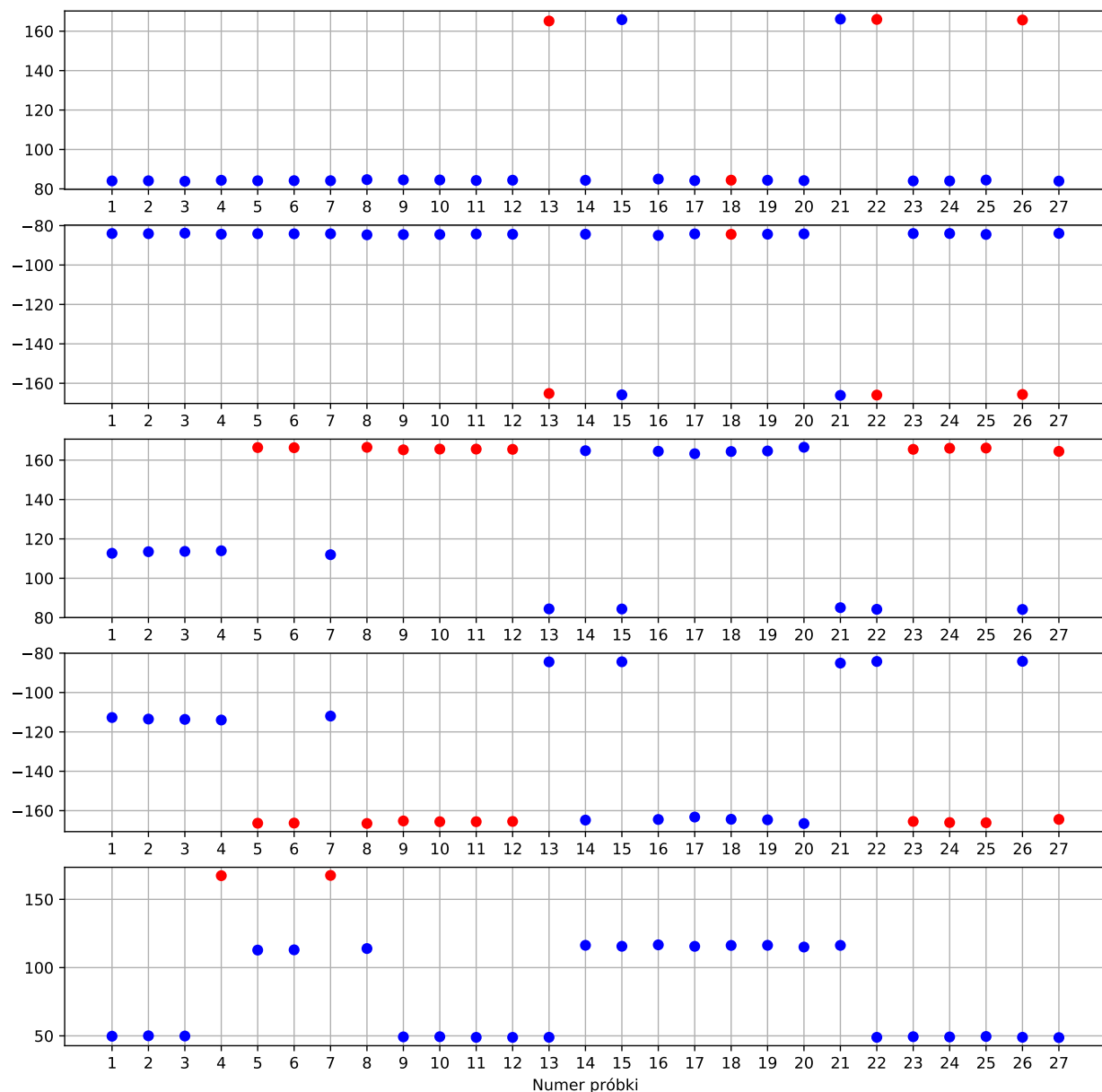


Rys. 10: Residua między wartościami przewidywanymi a prawdziwymi.



Rys. 11: Wykres pudełkowy residuów.

Możemy zauważyć że model ma niewielką tendencję do wartości zaniżania predykcji.



Rys. 12: Wykresy kątów pierwiastków mianownika transmitancji o pięciu największych promieniach (od góry w dół nierosnąco) w czasie. Kąty są wyskalowane w przedziale $(-180^\circ, 180^\circ]$. Niebieskim kolorem zaznaczono wartości dla pierwiastków znajdujących się wewnątrz okręgu jednostkowego, a czerwonym wewnątrz ale po zastosowaniu inwersji.

Na Rys. 12 możemy zauważyć iż nie wszystkie pierwiastki mianownika transmitancji znajdowały się wewnątrz okręgu jednostkowego, co pozwala przypuszczać że filtr odpowiadający temu procesowi jest niestabilny. W takim przypadku zastosowano inwersję liczby zespolonej:

$$inv(z) = \overline{\left(\frac{1}{z}\right)} = \frac{1}{\bar{z}} \quad (4)$$

pozwalające na sprowadzenie liczby zespolonej z poza okręgu jednostkowego do jego wnętrza. Kolejną rzeczą wartą odnotowania jest występowanie par sprzężonych pierwiastków – pierwsze dwa wykresy stanowią pierwiastki o taki samych promieniach i przeciwnych fazach (podobnie jak kolejne dwa). Taka sytuacja wynika wprost z twierdzenia o parach sprzężonych pierwiastków zespolonych [8] – dla wielomianu zmiennej zespolonej o współczynnikach rzeczywistych jeśli z_i jest pierwiastkiem, to również jest nim \bar{z}_i . Możemy również zauważyć że pierwsze dwa pierwiastki wykazują się najmniejszymi wahaniami, co sugeruje że mają największy wpływ w predykcji.

4 Bibliografia

- [1] https://en.wikipedia.org/wiki/PJM_Interconnection
- [2] https://www.kaggle.com/robikscube/hourly-energy-consumption#PJM_hourly.csv
- [3] https://www.researchgate.net/figure/PJM-Interconnection-Region_fig1_228652530
- [4] <https://www.statsmodels.org/stable/index.html>
- [5] Wykłady przedmiotowe
- [6] https://en.wikipedia.org/wiki/Autoregressive_model
- [7] <https://www.tandfonline.com/doi/abs/10.1080/00401706.1980.10486171>
- [8] https://en.wikipedia.org/wiki/Complex_conjugate_root_theorem