

Analiza danych rzeczywistych natężenia prądu plazmy przy pomocy modelu ARMA

Dominika Lewandowska, Nikodem Drelak

31 stycznia 2026

Spis treści

1	Wstęp	2
1.1	Cel pracy	2
1.2	Opis i źródło danych	2
1.3	Charakterystyka próby i wizualizacja	2
2	Przygotowanie danych do analizy	4
2.1	Analiza jakości danych	4
2.1.1	Analiza autokorelacji surowych danych	4
2.2	Dekompozycja szeregu czasowego	4
2.2.1	Różnicowanie danych	4
2.2.2	Ocena autokorelacji danych po dekompozycji	5
3	Modelowanie przy pomocy ARMA	5
3.1	Dobranie rzędu modelu i parametrów	5
3.2	Ocena dopasowania modelu	5
3.2.1	Porównanie empirycznych i teoretycznych ACV i PACV	5
3.2.2	Analiza trajektorii ARMA w porównaniu do badanego szeregu	5
4	Weryfikacja założeń dotyczących szumu	10
4.1	Założenie dotyczące średniej	10
4.2	Założenie dotyczące wariancji	10
4.3	Założenie dotyczące niezależności	10
4.3.1	Autokorelacja reszt	10
4.4	Założenie dotyczące normalności rozkładu	10
5	Wnioski	10
6	Podsumowanie	14
	Literatura	16

1 Wstęp

1.1 Cel pracy

Celem niniejszej pracy jest analiza statystyczna oraz modelowanie dynamiki szeregu czasowego pochodzącego z rzeczywistego eksperymentu fizycznego. Przedmiotem badań jest natężenie prądu plazmy (ang. *Plasma Current*) zarejestrowane w urządzeniu typu tokamak.

Głównym zadaniem jest weryfikacja hipotezy o możliwości opisu fluktuacji prądu w fazie stabilnej (tzw. *flat-top*) za pomocą liniowego modelu stochastycznego klasy ARMA. Analiza obejmuje zbadanie stacjonarności procesu, identyfikację rzędu modelu, estymację parametrów oraz weryfikację założeń dotyczących reszt.

1.2 Opis i źródło danych

Dane wykorzystane w projekcie pochodzą z reaktora fuzyjnego MAST (Mega Ampere Spherical Tokamak), znajdującego się w Culham Centre for Fusion Energy w Wielkiej Brytanii. Zostały pobrane za pośrednictwem otwartego interfejsu API udostępnionego w ramach projektu [FAIR-MAST](#).

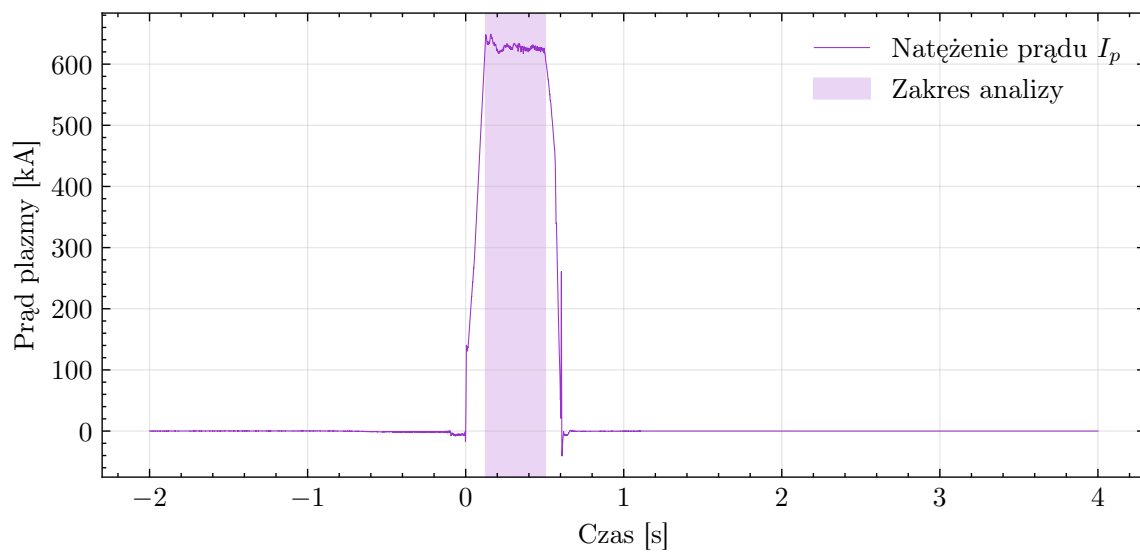
Interpretacja fizyczna zmiennej: Analizowana zmienna to natężenie prądu plazmy (I_p), wyrażone w kiloamperach (kA). W uproszczeniu, parametr ten jest kluczowym wskaźnikiem „życia” eksperymentu:

- Wzrost prądu oznacza formowanie się plazmy.
- Utrzymywanie stałej wartości (plateau) oznacza fazę stabilną, w której przeprowadza się właściwe eksperymenty.
- Nagły spadek wartości do zera może sygnalizować niekontrolowaną utratę stabilności (tzw. disruption).

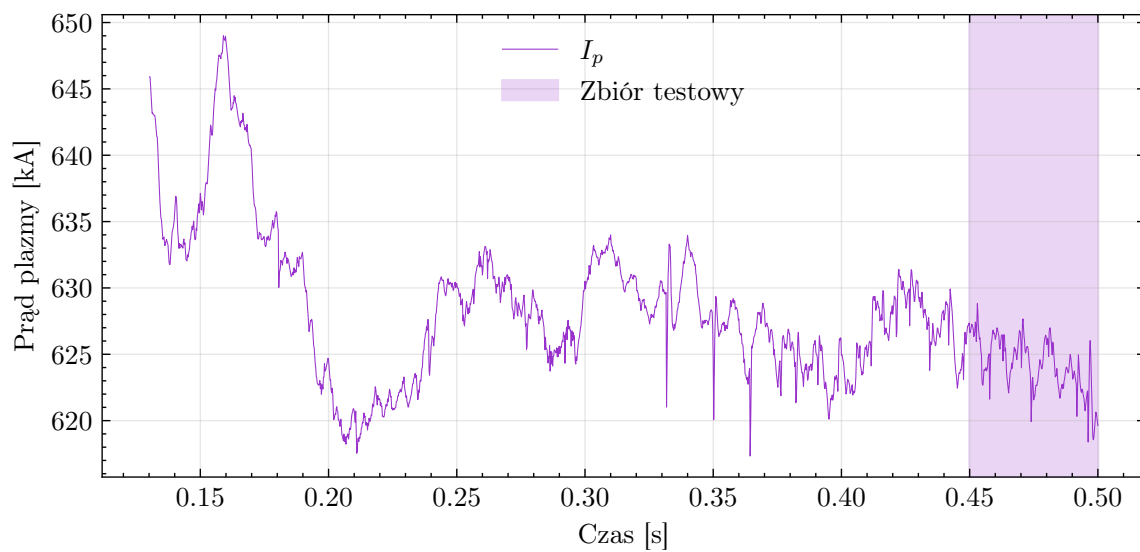
Dla potrzeb analizy szeregów czasowych, fluktuacje tego prądu w fazie stabilnej traktujemy jako proces stochastyczny, wynikający z turbulencji wewnątrz gorącego gazu oraz działania systemów sterowania reaktora.

1.3 Charakterystyka próby i wizualizacja

Do analizy wybrano eksperyment (tzw. *shot*) o numerze ID: 30421. Pełny przebieg eksperymentu (przedstawiony na rysunku 1) trwa około 6 sekund. Ze względu na niestacjonarny charakter całego procesu (faza rozruchu i wygaszania), do modelowania ARMA wyodrębniono wycinek czasowy odpowiadający fazie stabilnej (rysunek 2). Zbiór analityczny odpowiada 0,37 sekundy obserwacji i zawiera 1850 próbek. W celu późniejszej weryfikacji poprawności modelowania ARMA dla przyszłych wartości, został on podzielony na zbiór treningowy - pierwsze 0,32 sekundy (1500 próbek) oraz zbiór testowy - ostatnie 0.05 sekund (250 próbek).



Rysunek 1: Pełny przebieg natężenia prądu plazmy w czasie eksperymentu. Półprzezroczystym obszarem zaznaczono obszar fazy stabilnej wybrany do analizy.



Rysunek 2: Wyodrębniony fragment szeregu czasowego (faza *flat-top*) poddany modelowaniu ARMA. Półprzezroczystym obszarem zaznaczono obszar zbioru testowego.

2 Przygotowanie danych do analizy

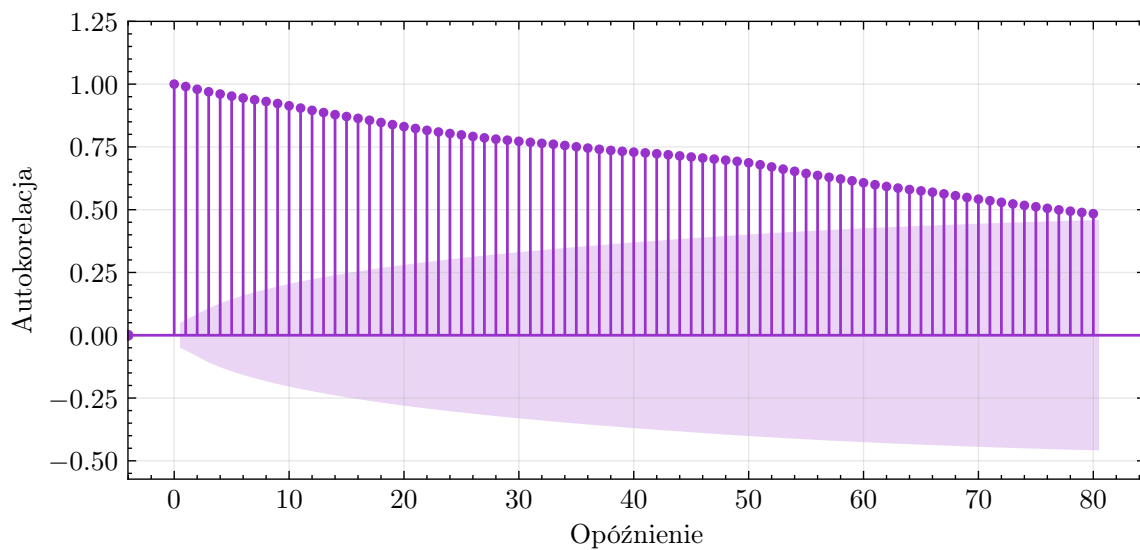
2.1 Analiza jakości danych

Coś tam że dane nie mają braków itd.

2.1.1 Analiza autokorelacji surowych danych

wzory i opisanie ACF i PACF co to i wgl

Rysunek 3 przedstawia to i to. Możemy wywnioskować z tego, że coś tam. Interpretujemy to tak i tak



Rysunek 3: Wykres współczynnika autokorelacji surowych danych ze zbioru treningowego w zależności od opóźnienia (parametru h). Półprzezroczystym obszarem zaznaczono przedział ufności o poziomie 95%.

Rysunek 4 przedstawia to i to. Możemy wywnioskować z tego, że coś tam. Interpretujemy to tak i tak

test ADF weryfikujący hipotezę o niestacjonarności dla surowych danych (Augmented Dickey-Fuller Test) dał takie wyniki: Statystyka testowa = -0.665, p-value = 0.427 =, Non-stationary.

2.2 Dekompozycja szeregu czasowego

2.2.1 Różnicowanie danych

co to różnicowanie danych itd

Rysunek 5 przedstawia to i to. Możemy wywnioskować z tego, że coś tam. Interpretujemy to tak i tak

2.2.2 Ocena autokorelacji danych po dekompozycji

Rysunek 6 przedstawia to i to. Możemy wywnioskować z tego, że coś tam. Interpretujemy to tak i tak

Rysunek 7 przedstawia to i to. Możemy wywnioskować z tego, że coś tam. Interpretujemy to tak i tak

test ADF weryfikujący hipotezę o niestacjonarności dla danych po dekompozycji (Augmented Dickey-Fuller Test) dał takie wyniki: Statystyka testowa = -14.148, p-Value = 0.0 =, Stationary.

3 Modelowanie przy pomocy ARMA

3.1 Dobranie rzędu modelu i parametrów

co to rząd modelu, BIC, AIC, jak się dobiera itd BIC (Bayesowskie Kryterium Informacyjne) i AIC (Kryterium Informacyjne Akaikego) dobraliśmy rząd modelu korzystając z kryteriów informacyjnych w pakiecie pdarima??? wyszło: Best model: ARMA(7,6)??? (my użyliśmy 5,5 z rysunku 8)

parametry: TODO (tabelka najlepiej)

3.2 Ocena dopasowania modelu

3.2.1 Porównanie empirycznych i teoretycznych ACV i PACV

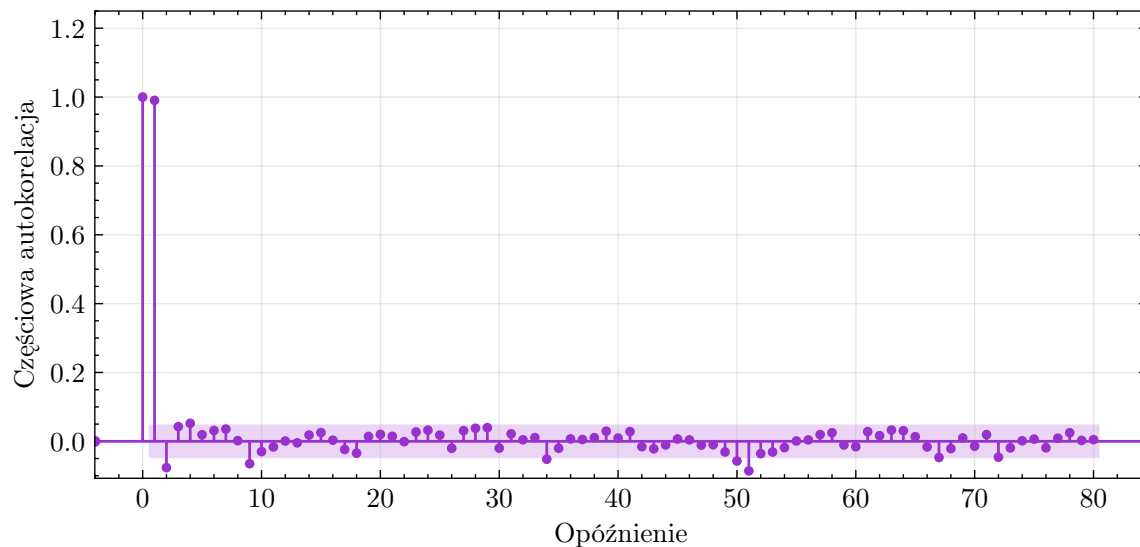
Rysunek 9 przedstawia to i to. Możemy wywnioskować z tego, że coś tam. Interpretujemy to tak i tak

Rysunek 10 przedstawia to i to. Możemy wywnioskować z tego, że coś tam. Interpretujemy to tak i tak

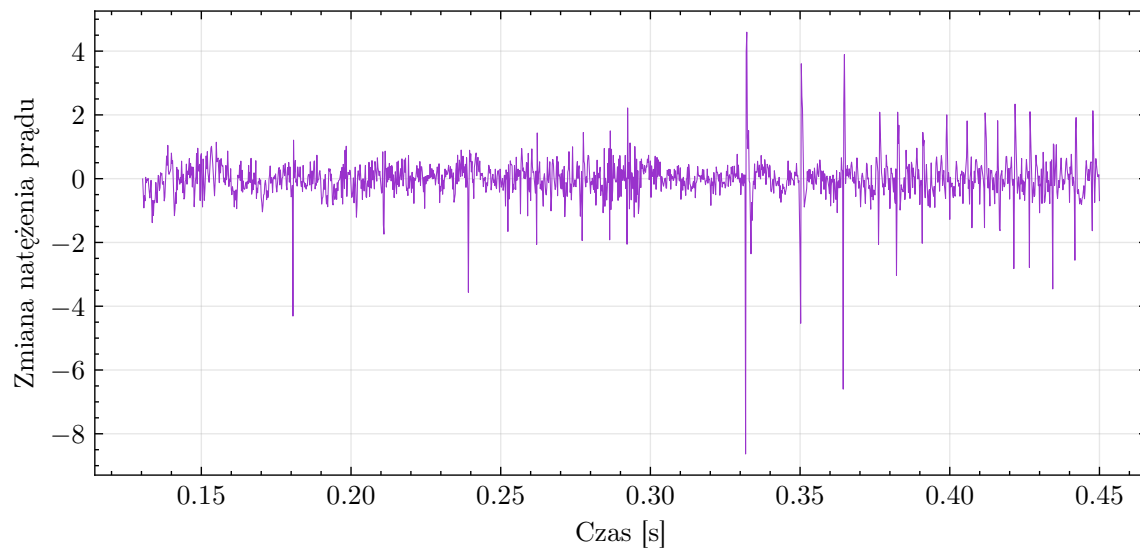
3.2.2 Analiza trajektorii ARMA w porównaniu do badanego szeregu

Rysunek 11 przedstawia to i to. Możemy wywnioskować z tego, że coś tam. Interpretujemy to tak i tak

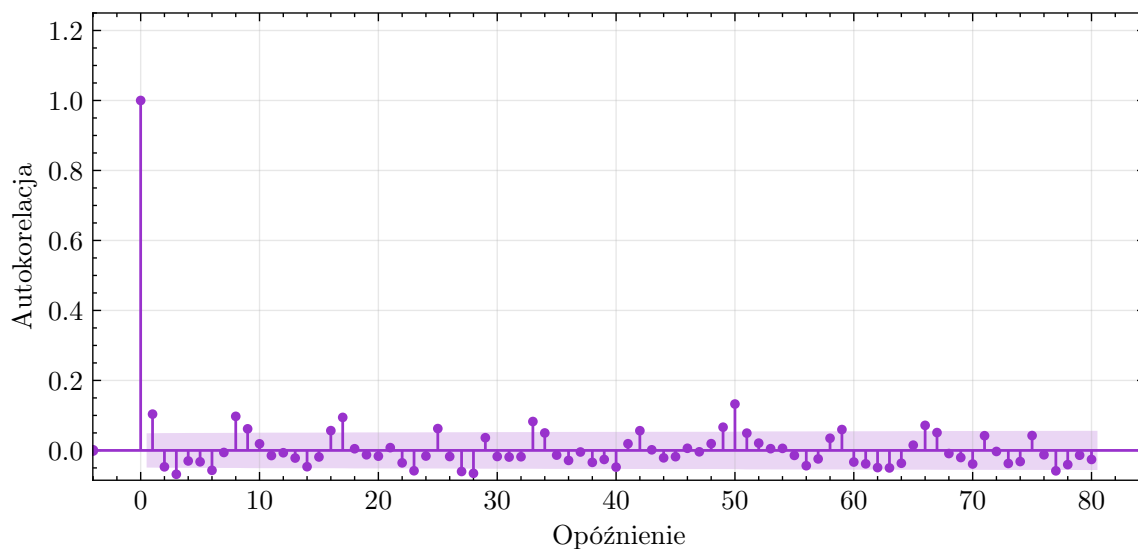
Rysunek 12 przedstawia to i to. Możemy wywnioskować z tego, że coś tam. Interpretujemy to tak i tak



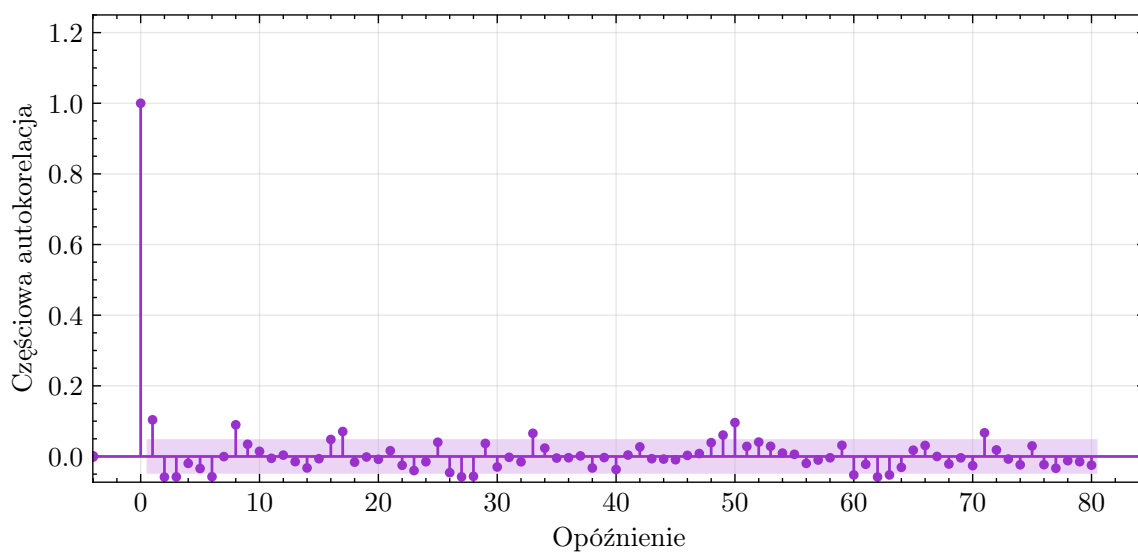
Rysunek 4: Wykres współczynnika częściowej autokorelacji surowych danych ze zbioru treningowego w zależności od opóźnienia (parametru h). Półprzezroczystym obszarem zaznaczono przedział ufności o poziomie 95%.



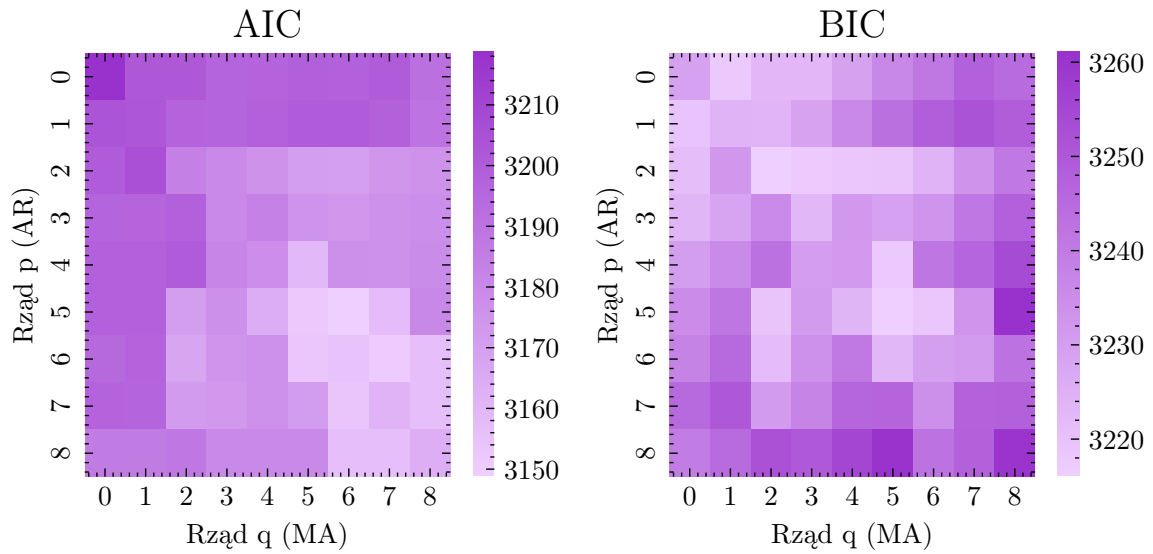
Rysunek 5: Szereg czasowy ze zbioru treningowego poddany różnicowaniu. Przedstawia zmianę natężenia prądu w zależności od czasu.



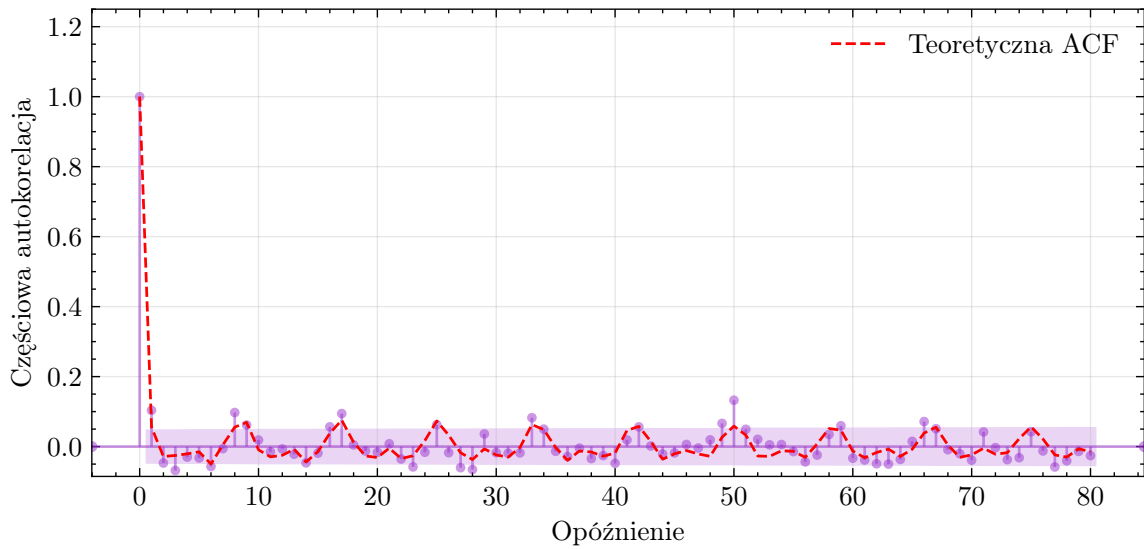
Rysunek 6: Wykres współczynnika autokorelacji danych ze zbioru treningowego poddanych różnicowaniu w zależności od opóźnienia (parametru h). Półprzezroczystym obszarem zaznaczono przedział ufności o poziomie 95%.



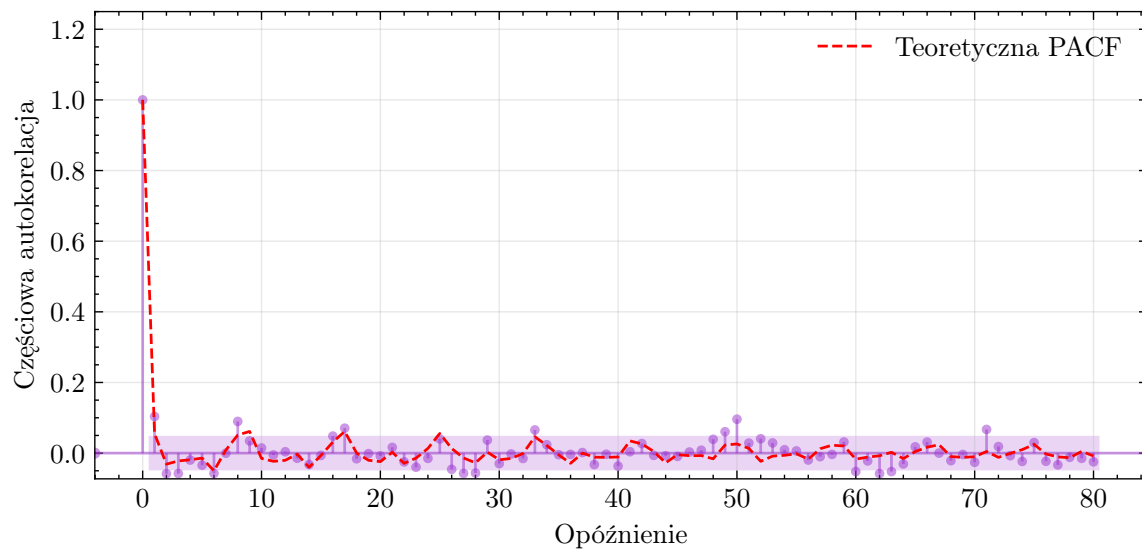
Rysunek 7: Wykres współczynnika częściowej autokorelacji danych ze zbioru treningowego poddanych różnicowaniu w zależności od opóźnienia (parametru h). Półprzezroczystym obszarem zaznaczono przedział ufności o poziomie 95%.



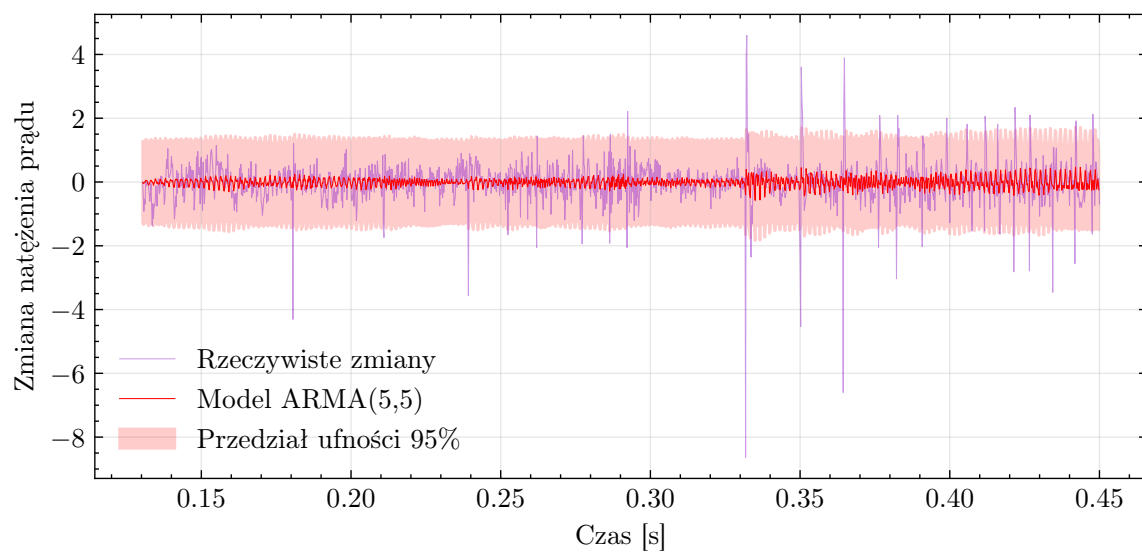
Rysunek 8: Heatmapa wartości kryteriów AIC oraz BIC w zależności od rzędów p i q modelu $\text{ARMA}(p, q)$ badanego szeregu. Jaśniejszy kolor oznacza niższą wartość kryterium.



Rysunek 9: Wykres empirycznego i teoretycznego współczynnika autokorelacji modelu $\text{ARMA}(5, 5)$ dopasowanego do zbioru treningowego, w zależności od opóźnienia (parametru h). Półprzezroczystym obszarem zaznaczono przedział ufności o poziomie 95%.



Rysunek 10: Wykres empirycznego i teoretycznego współczynnika częściowej autokorelacji modelu ARMA(5, 5) dopasowanego do zbioru treningowego, w zależności od opóźnienia (parametru h). Półprzezroczystym obszarem zaznaczono przedział ufności o poziomie 95%.



Rysunek 11: Trajektorja szeregu czasowego ze zbioru treningowego poddanego różnicowaniu oraz trajektorja dopasowanego modelu ARMA(5, 5) wraz z przedziałami ufności o poziomie 95%.

4 Weryfikacja założeń dotyczących szumu

4.1 Założenie dotyczące średniej

Rysunek 13 co to za test i wgl: Test Średniej (T-test): Średnia reszt: -1.2344e-03 P-value: 0.9389 WNIOSEK: Brak podstaw do odrzucenia H_0 . Średnia jest równa 0.

4.2 Założenie dotyczące wariancji

opisać co to za test itd Test Wariancji (ARCH Test): P-value: 0.0000 WNIOSEK: Wykryto zmienną wariancję (efekt ARCH).

sprawdzimy jak zmienia się wariancja: opisać jak się liczyło tą zmianę rezimu wariancji, co to C_j , $V_j \dots$

Rysunek 14

4.3 Założenie dotyczące niezależności

co to za test i wgl: Test Niezależności (Ljung-Box, lag=10): P-value: 0.2137 WNIOSEK: Reszty są niezależne (brak autokorelacji, biały szum).

4.3.1 Autokorelacja reszt

Reszty (błędy) w modelu ARMA/ARIMA reprezentują różnicę między obserwowanymi wartościami szeregu czasowego a wartościami przewidywanymi (dopasowanymi) przez model. Poprawne reszty powinny zachowywać się jak „biały szum”, czyli mieć stałą wariancję, zerową średnią i brak autokorelacji.

Rysunek 15 przedstawia to i to. Możemy wywnioskować z tego, że coś tam. Interpretujemy to tak i tak

Rysunek 16 przedstawia to i to. Możemy wywnioskować z tego, że coś tam. Interpretujemy to tak i tak

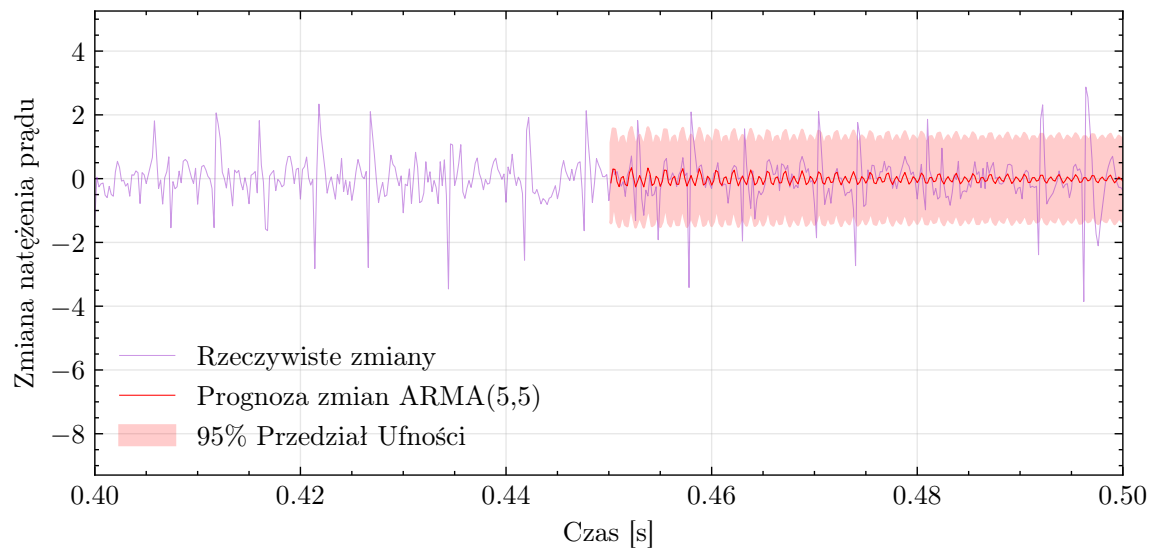
4.4 Założenie dotyczące normalności rozkładu

Rysunek 17

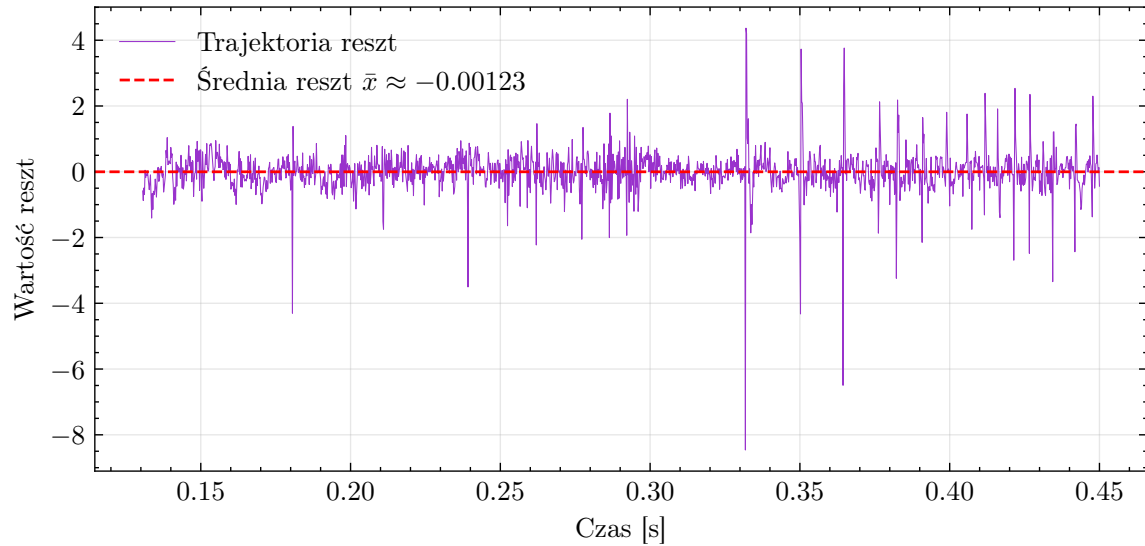
co to za test i wgl: Test Normalności (Jarque-Bera): P-value: 0.0000 WNIOSEK: Rozkład nie jest idealnie normalny (częste w danych fizycznych).

5 Wnioski

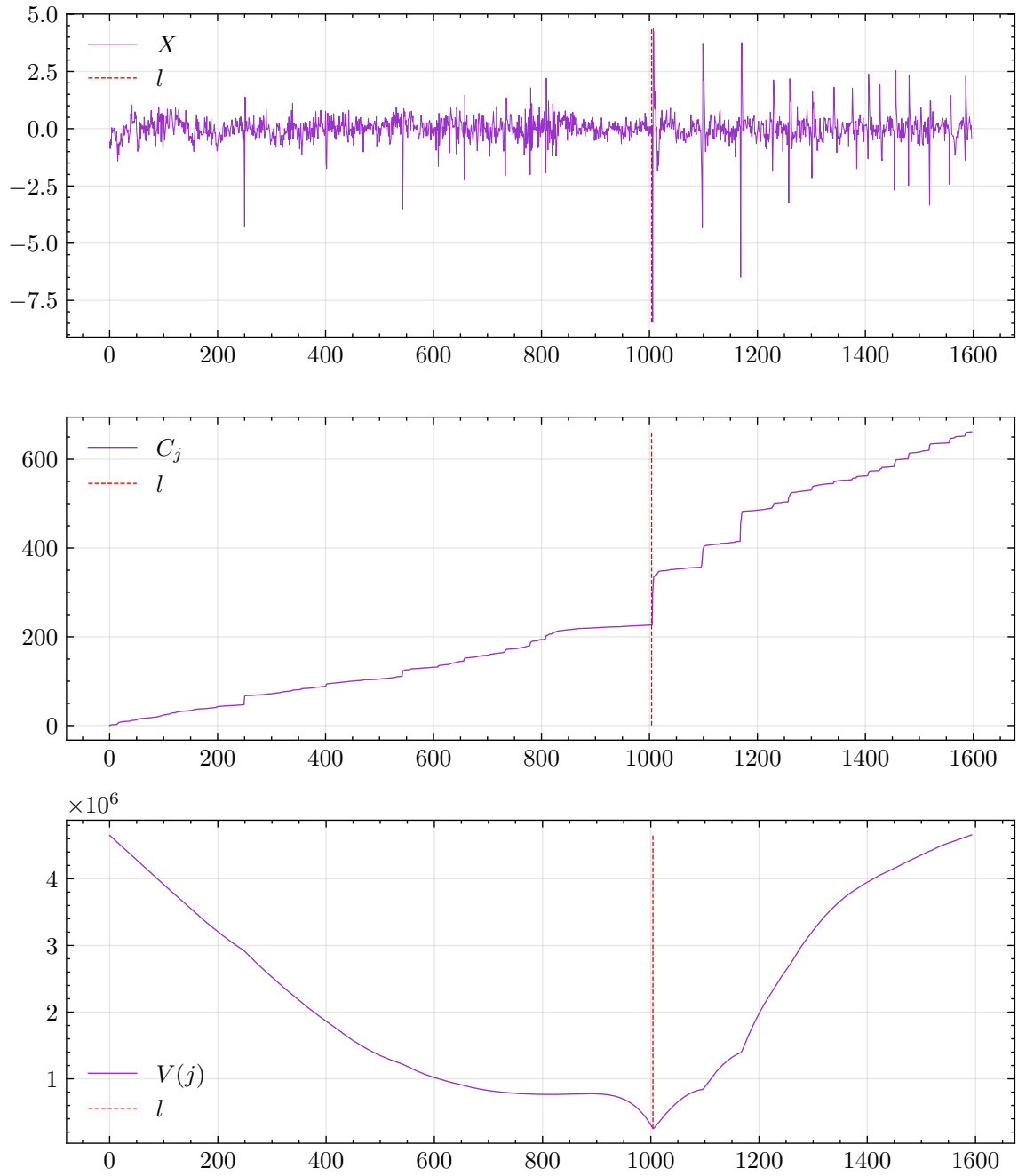
nie zgadza się wariancja ani normalność więc nie można użyć ARMA do modelowania tego typu danych



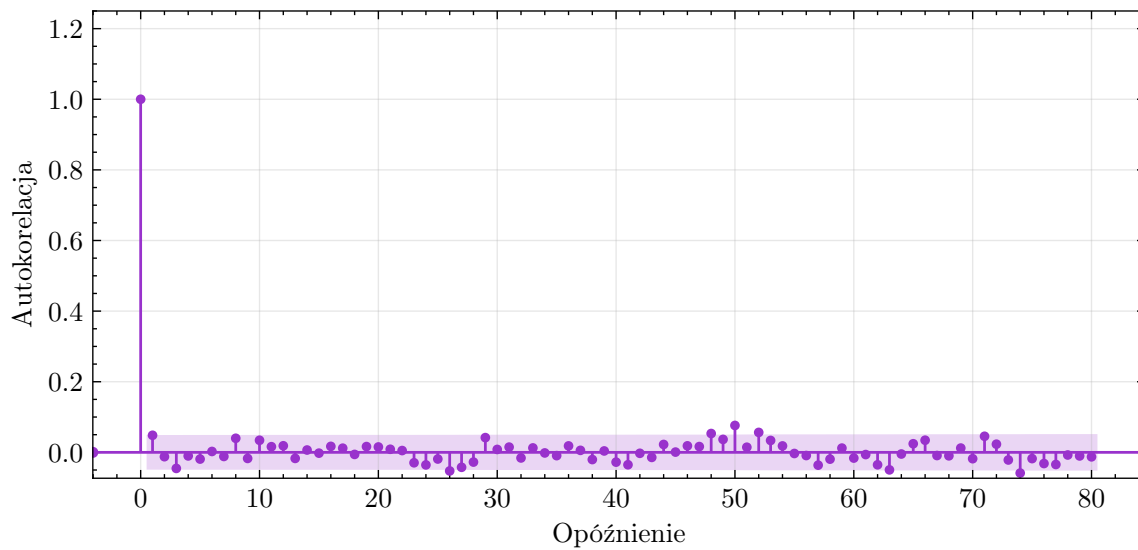
Rysunek 12: Trajektoria szeregu czasowego ze zbioru treningowego i ze zbioru testowego poddanych różnicowaniu, oraz prognoza trajektorii dopasowanego modelu ARMA(5, 5) dla zbioru testowego wraz z przedziałami ufności o poziomie 95%.



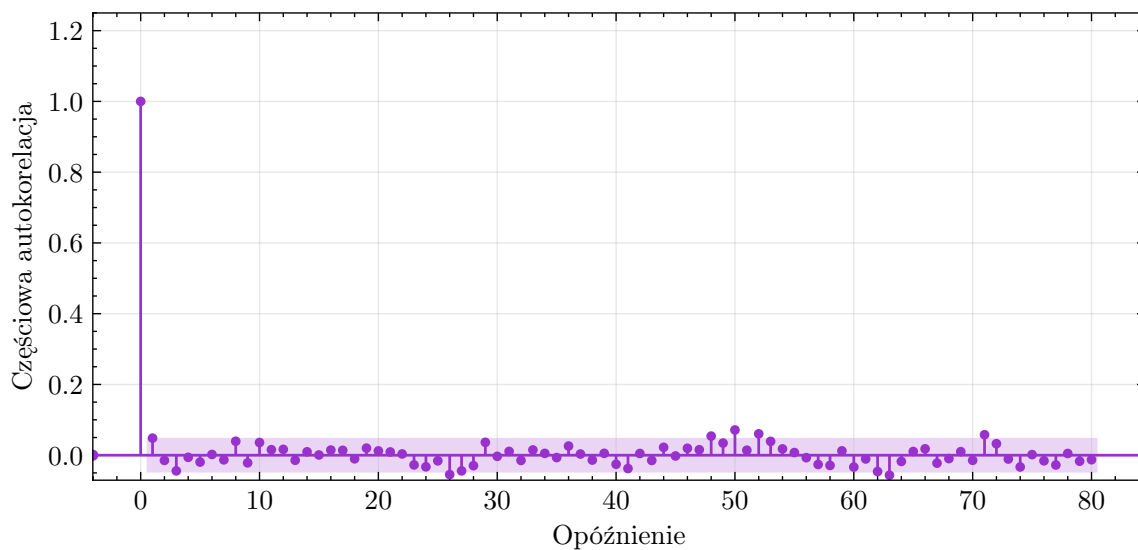
Rysunek 13: Wykres wartości reszt modelu ARMA(5, 5) dopasowanego do zbioru treningowego. Czerwoną przerywaną linią zaznaczono średnią reszt o wartości około -0.00123 .



Rysunek 14: Wykresy przedstawiające kolejno: trajektorię reszt modelu ARMA(5, 5) dopasowanego do zbioru treningowego, sumę skumulowanych kwadratów reszt (j) oraz sumę błędów kwadratowych $V(j)$, w zależności od czasu. Czerwoną linią przerywaną zaznaczono punkt l , czyli punkt największej zmiany reżimu wariancji.



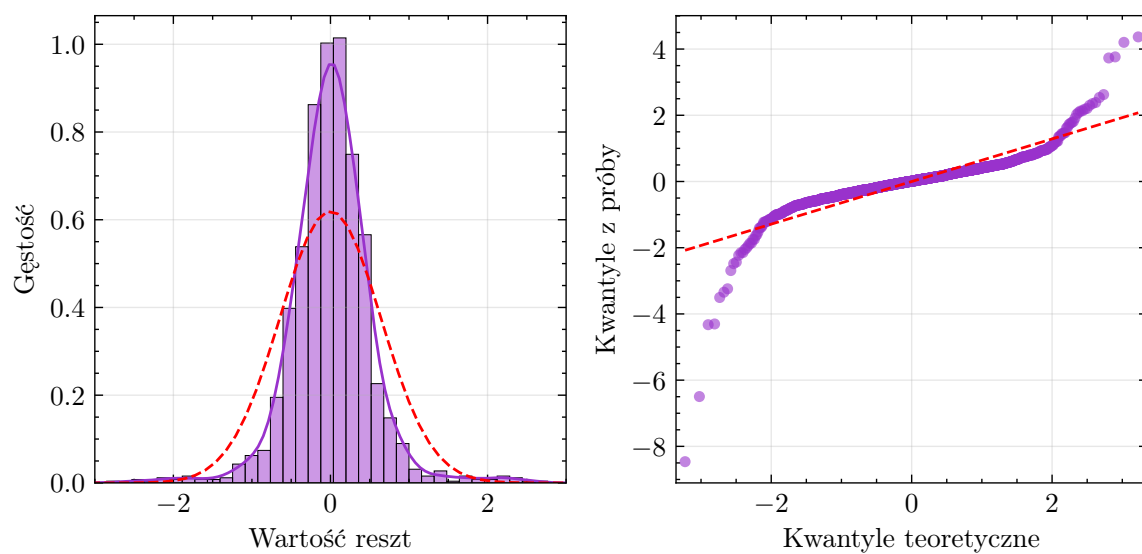
Rysunek 15: Wykres współczynnika autokorelacji reszt modelu ARMA(5, 5) w zależności od opóźnienia (parametru h). Półprzezroczystym obszarem zaznaczono przedział ufności o poziomie 95%.



Rysunek 16: Wykres współczynnika częściowej autokorelacji reszt modelu ARMA(5, 5) w zależności od opóźnienia (parametru h). Półprzezroczystym obszarem zaznaczono przedział ufności o poziomie 95%.

6 Podsumowanie

nie można używać arma ale i tak daje wyniki zawierające się w przedziale ufności. i do tego typu danych lepiej używać innych modeli które uwzględniają zmiany wariancji i brak normalności np ARCH, GARCH...



Rysunek 17: Histogram i empiryczna gęstość oraz wykres kwantylowy (QQ-plot) wartości reszt w porównaniu do rozkładu normalnego (czerwona, przerywana linia) o parametrach wyznaczonych na podstawie badanej próby reszt (średniej i wariancji z próby).

Literatura

- [1] J. Gajda, G. Sikora, and A. Wyłomańska. Regime variance testing — a quantile approach. *Acta Physica Polonica B*, 44(5):1015, 2013.
- [2] Samuel Jackson, Saiful Khan, Nathan Cummings, James Hodson, Shaun de Witt, Stanislas Pamela, Rob Akers, and Jeyan Thiyagalingam. An Open Data Service for Supporting Research in Machine Learning on Tokamak Data. *IEEE Transactions on Plasma Science*, 2025.
- [3] Samuel Jackson, Saiful Khan, Nathan Cummings, James Hodson, Shaun de Witt, Stanislas Pamela, Rob Akers, Jeyan Thiyagalingam, and The MAST Team. Fair-mast: A fusion device data management system. *SoftwareX*, 27:101869, 2024.
- [4] Skipper Seabold and Josef Perktold. statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*, 2010.