

# Analiza danych rzeczywistych natężenia prądu plazmy przy pomocy modelu ARMA

Dominika Lewandowska, Nikodem Drelak

7 lutego 2026

## Spis treści

<b>1</b>	<b>Wstęp</b>	<b>2</b>
1.1	Cel pracy . . . . .	2
1.2	Opis i źródło danych . . . . .	2
1.3	Charakterystyka próby i wizualizacja . . . . .	2
<b>2</b>	<b>Przygotowanie danych do analizy</b>	<b>4</b>
2.1	Analiza jakości danych . . . . .	4
2.1.1	Analiza autokorelacji surowych danych . . . . .	4
2.2	Dekompozycja szeregu czasowego . . . . .	6
2.2.1	Różnicowanie danych . . . . .	6
2.2.2	Ocena autokorelacji danych po dekompozycji . . . . .	6
<b>3</b>	<b>Modelowanie przy pomocy ARMA</b>	<b>6</b>
3.1	Dobranie rzędu modelu i parametrów . . . . .	9
3.2	Ocena dopasowania modelu . . . . .	10
3.2.1	Porównanie empirycznych i teoretycznych ACV i PACV . . . . .	10
3.2.2	Analiza trajektorii ARMA w porównaniu do badanego szeregu . . . . .	10
<b>4</b>	<b>Weryfikacja założeń dotyczących szumu</b>	<b>12</b>
4.1	Założenie dotyczące średniej . . . . .	12
4.2	Założenie dotyczące wariancji . . . . .	12
4.3	Założenie dotyczące niezależności . . . . .	14
4.3.1	Autokorelacja reszt . . . . .	14
4.4	Założenie dotyczące normalności rozkładu . . . . .	14
<b>5</b>	<b>Wnioski</b>	<b>17</b>
	<b>Literatura</b>	<b>18</b>

# 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 [1, 2].

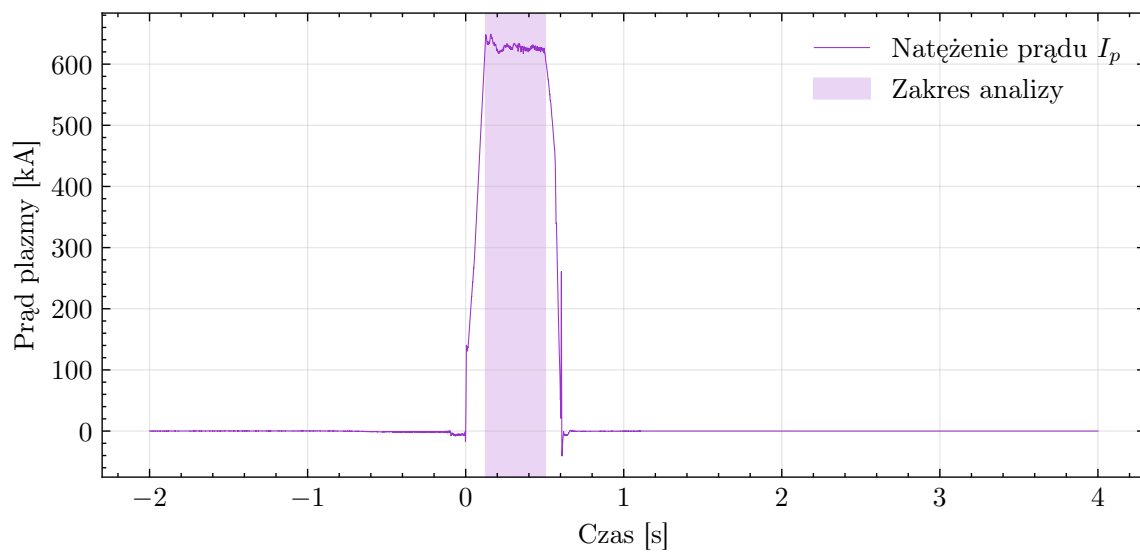
Interpretacja fizyczna zmiennej: Analizowana zmienna to natężenie prądu plazmy ( $I_p$ ), wyrażone w kiloamperach (kA). W uproszczeniu, parametr ten jest wskaźnikiem „życia” eksperymentu:

- 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).

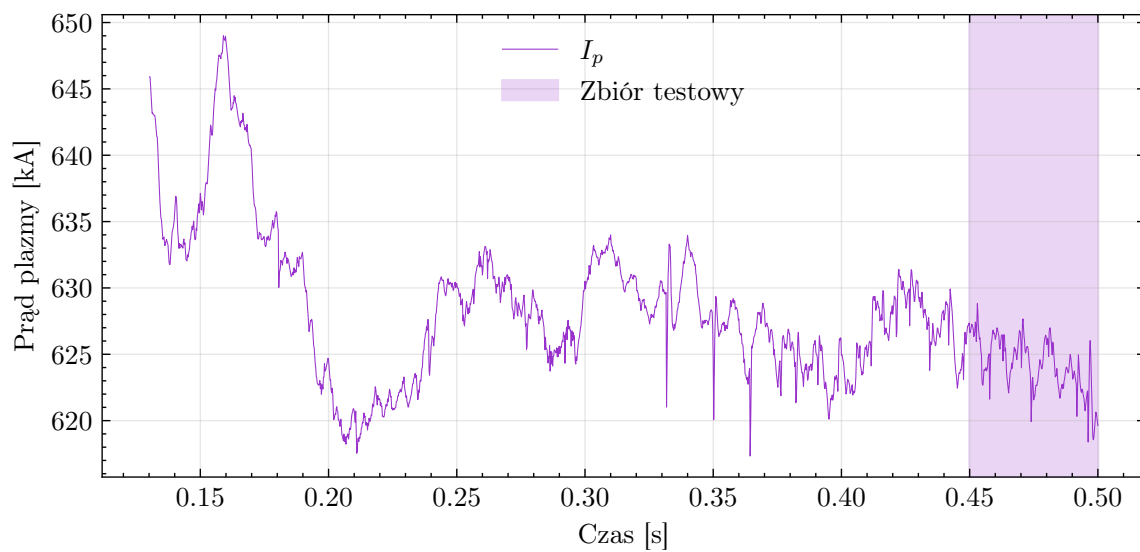
Dla potrzeb analizy szeregów czasowych, fluktuacje tego prądu w fazie stabilnej traktujemy jako proces stochastyczny, wynikający z turbulencji wewnątrz gorącego gazu oraz działania systemów sterowania reaktora.

## 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

Na wstępnym etapie analizy dokonano oceny jakości danych pomiarowych. Sprawdzono obecność braków danych, wartości odstających oraz nieciągłości czasowych. Analizowany szereg czasowy nie zawiera brakujących obserwacji ani duplikatów, a odstępy czasowe pomiędzy kolejnymi próbkami są jednorodne.

Wartości natężenia prądu mieszczą się w zakresie fizycznie uzasadnionym dla pracy tokamaka i nie zaobserwowano anomalii mogących wskazywać na błędy pomiarowe. Na tej podstawie dane uznano za poprawne i odpowiednie do dalszej analizy statystycznej.

#### 2.1.1 Analiza autokorelacji surowych danych

W celu wstępnej identyfikacji struktury zależności czasowych w szeregu obliczono funkcję autokorelacji (ACF) oraz funkcję częściowej autokorelacji (PACF).

Funkcja autokorelacji ACF opisuje stopień liniowej zależności pomiędzy obserwacjami oddzielonymi o  $h$  kroków czasowych i dana jest wzorem:

$$\rho(h) = \frac{\text{Cov}(X_t, X_{t-h})}{\text{Var}(X_t)}. \quad (1)$$

W praktyce, dysponując skończoną próbą  $n$  obserwacji  $x_1, x_2, \dots, x_n$ , wyznacza się empiryczną funkcję autokorelacji (estymator funkcji ACF) według wzoru:

$$\hat{\rho}(h) = \frac{\sum_{t=1}^{n-h} (x_t - \bar{x})(x_{t+h} - \bar{x})}{\sum_{t=1}^n (x_t - \bar{x})^2}, \quad (2)$$

gdzie  $\bar{x}$  oznacza średnią arytmetyczną z próby.

Z kolei funkcja częściowej autokorelacji (PACF) mierzy korelację pomiędzy  $X_t$  i  $X_{t-h}$  po wyeliminowaniu wpływu opóźnień pośrednich  $1, 2, \dots, h-1$ . Teoretyczna wartość PACF dla rzędu  $h$ , oznaczana jako  $\phi_{hh}$ , jest zdefiniowana jako ostatni współczynnik w modelu autoregresji rzędu  $h$ :

$$X_t = \phi_{h1}X_{t-1} + \phi_{h2}X_{t-2} + \dots + \phi_{hh}X_{t-h} + e_t. \quad (3)$$

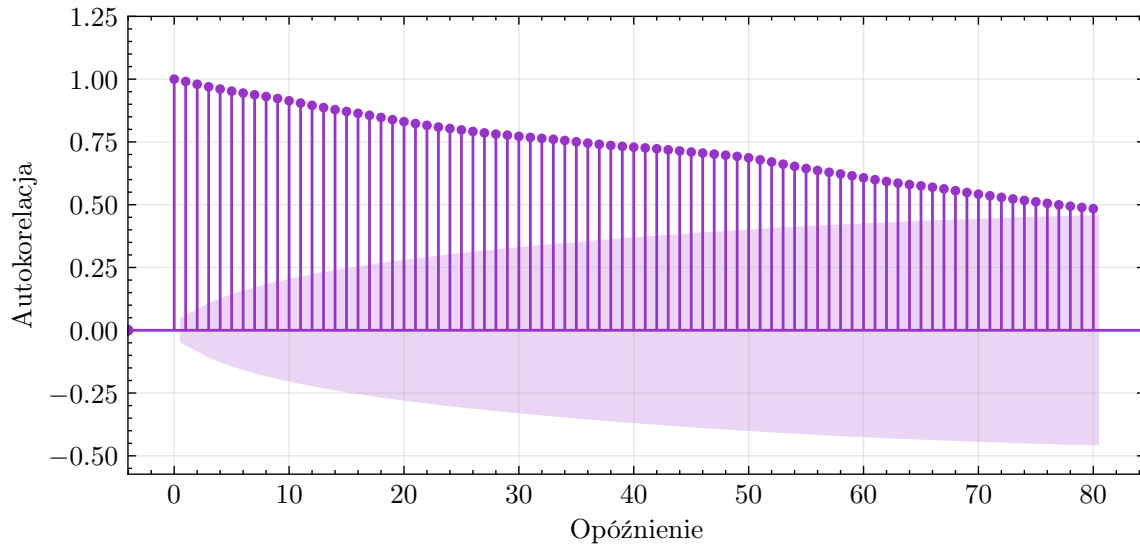
Empiryczna funkcja PACF, oznaczana jako  $\hat{\phi}_{hh}$ , jest zazwyczaj obliczana rekurencyjnie przy użyciu algorytmu Durбина-Levinsona:

$$\hat{\phi}_{hh} = \frac{\hat{\rho}(h) - \sum_{j=1}^{h-1} \hat{\phi}_{h-1,j} \hat{\rho}(h-j)}{1 - \sum_{j=1}^{h-1} \hat{\phi}_{h-1,j} \hat{\rho}(j)}, \quad (4)$$

gdzie  $\hat{\phi}_{11} = \hat{\rho}(1)$ , co pozwala na lepszą identyfikację rzędu części autoregresyjnej modelu.

Rysunek 3 przedstawia wykres empirycznej funkcji autokorelacji ACF dla surowych danych ze zbioru treningowego. Zaznaczony na wykresie półprzezroczysty obszar stanowi

przedział ufności skonstruowany dla hipotezy zerowej, zakładającej, że badany proces jest czystym szumem białym. Widoczna jest bardzo wolno malejąca autokorelacja oraz wysokie wartości współczynnika dla dużych opóźnień, znacząco przekraczające granice przedziału ufności. Taki kształt ACF jest charakterystyczny dla procesów niestacjonarnych.



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

Rysunek 4 prezentuje empiryczną funkcję częściowej autokorelacji (PACF). Na wykresie dominują wysokie wartości dla dwóch pierwszych opóźnień ( $h = 0$  i  $h = 1$ ). Pomijając autokorelację własną w punkcie zero, obserwujemy gwałtowne załamanie funkcji po pierwszym opóźnieniu. Taki obraz PACF, w zestawieniu z wolno malejącym ACF, jest typowy dla procesów niestacjonarnych i wskazuje na konieczność różnicowania szeregu.

Dodatkowo przeprowadzono test Augmented Dickey–Fullera (ADF), weryfikujący hipotezę zerową o istnieniu pierwiastka jednostkowego. Statystyka ta, będąca odpowiednikiem statystyki  $t$  dla rozkładu niesymetrycznego, pozwala ocenić siłę dowodów przeciwko niestacjonarności. Im niższa (bardziej ujemna) jest jej wartość, tym bardziej prawdopodobna jest stacjonarność procesu.

W analizowanym przypadku otrzymano statystykę testową równą  $-0.665$  oraz wartość  $p$  równą  $0.427$ , co ze względu na wynik znacznie wyższy od wartości krytycznych oznacza brak podstaw do odrzucenia hipotezy o niestacjonarności szeregu.

## 2.2 Dekompozycja szeregu czasowego

### 2.2.1 Różnicowanie danych

W celu uzyskania stacjonarności szeregu zastosowano różnicowanie pierwszego rzędu. Operacja ta polega na zastąpieniu oryginalnego szeregu  $X_t$  nowym szeregiem:

$$Y_t = X_t - X_{t-1}, \quad (5)$$

który opisuje zmiany wartości natężenia prądu pomiędzy kolejnymi chwilami czasu. Różnicowanie jest standardową metodą eliminacji trendu oraz wolnozmiennych składowych deterministycznych w analizie szeregów czasowych.

Rysunek 5 przedstawia trajektorię szeregu po różnicowaniu. Widoczne jest usunięcie trendu oraz oscylowanie wartości wokół zera, co wskazuje na poprawę własności stacjonarnych.

### 2.2.2 Ocena autokorelacji danych po dekompozycji

Na rysunkach 6 i 7 przedstawiono odpowiednio wykresy funkcji autokorelacji (ACF) oraz częściowej autokorelacji (PACF) dla szeregu po zróżnicowaniu.

W przeciwieństwie do danych surowych, funkcja ACF charakteryzuje się szybkim, gwałtownym zanikiem wartości współczynników, co potwierdza brak długookresowej pamięci (trendu) w przekształconym procesie.

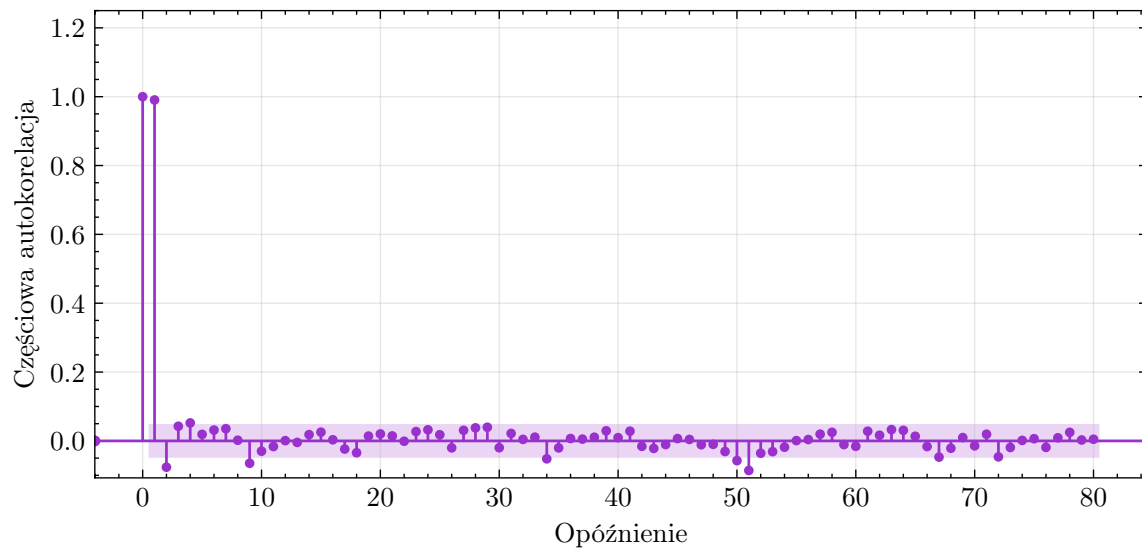
Analiza wykresu PACF wskazuje na ustąpienie dominującej korelacji dla pierwszego opóźnienia ( $h = 1$ ), która we wcześniejszym etapie sugerowała niestacjonarność. Zdecydowana większość empirycznych wartości autokorelacji mieści się wewnątrz przedziału ufności, co świadczy o statystycznej nieistotności tych zależności i skutecznym sprowadzeniu szeregu do postaci stacjonarnej.

Test ADF przeprowadzony dla szeregu po dekompozycji dał statystykę 14.148 oraz wartość  $p$  równą  $1.7039 \times 10^{-25} \approx 0$ , co jednoznacznie prowadzi do odrzucenia hipotezy o niestacjonarności. Oznacza to, że uzyskany szereg jest stacjonarny i może być modelowany przy użyciu modeli klasy ARMA.

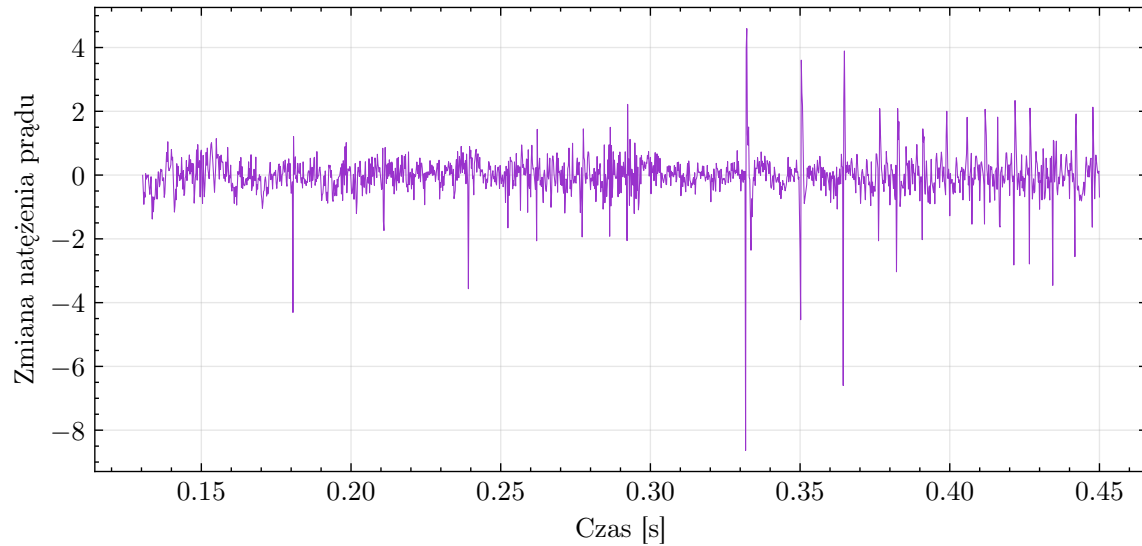
## 3 Modelowanie przy pomocy ARMA

Po sprowadzeniu szeregu do postaci stacjonarnej, przystąpiono do modelowania jego dynamiki przy wykorzystaniu modeli klasy ARMA (ang. *AutoRegressive Moving Average*). Struktura ta pozwala na efektywny opis zależności czasowych poprzez liniową kombinację przeszłych obserwacji (część autoregresyjna) oraz przeszłych błędów losowych (część średniej ruchomej).

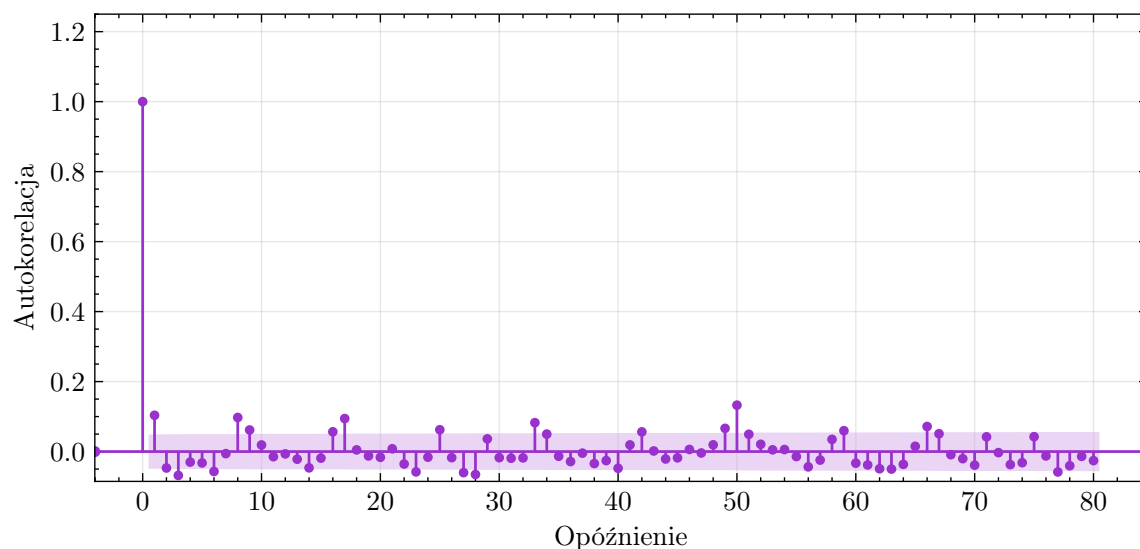
W niniejszej sekcji przeprowadzono proces identyfikacji rzędów modelu  $p$  i  $q$ , estymację parametrów oraz diagnostykę dopasowania. Cały proces estymacji parametrów modelu oraz wyznaczenia wartości kryteriów informacyjnych zrealizowano przy użyciu biblioteki `statsmodels` [3].



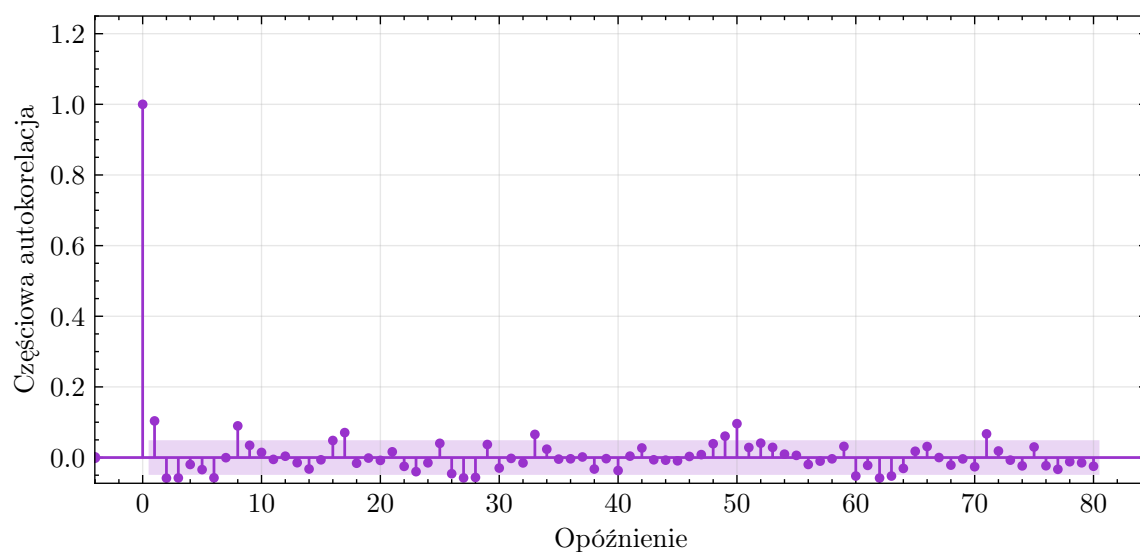
Rysunek 4: Wykres współczynnika częściowej autokorelacji surowych danych ze zbioru treningowego w zależności od opóźnienia (parametru  $h$ ). Półprzezroczystym obszarem zaznaczono przedział ufności o poziomie 95% dla hipotezy o braku korelacji (szum biały).



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% dla hipotezy o braku korelacji (szum biały).



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% dla hipotezy o braku korelacji (szum biały).



### 3.1 Dobranie rzędu modelu i parametrów

Model  $\text{ARMA}(p, q)$  łączy w sobie składnik autoregresyjny rzędu  $p$  oraz składnik średniej ruchomej rzędu  $q$ . Jego postać ogólna dana jest wzorem:

$$X_t = \sum_{i=1}^p \Phi_i X_{t-1} + \sum_{j=1}^q \theta_j \epsilon_{t-1} + \epsilon_t, \quad (6)$$

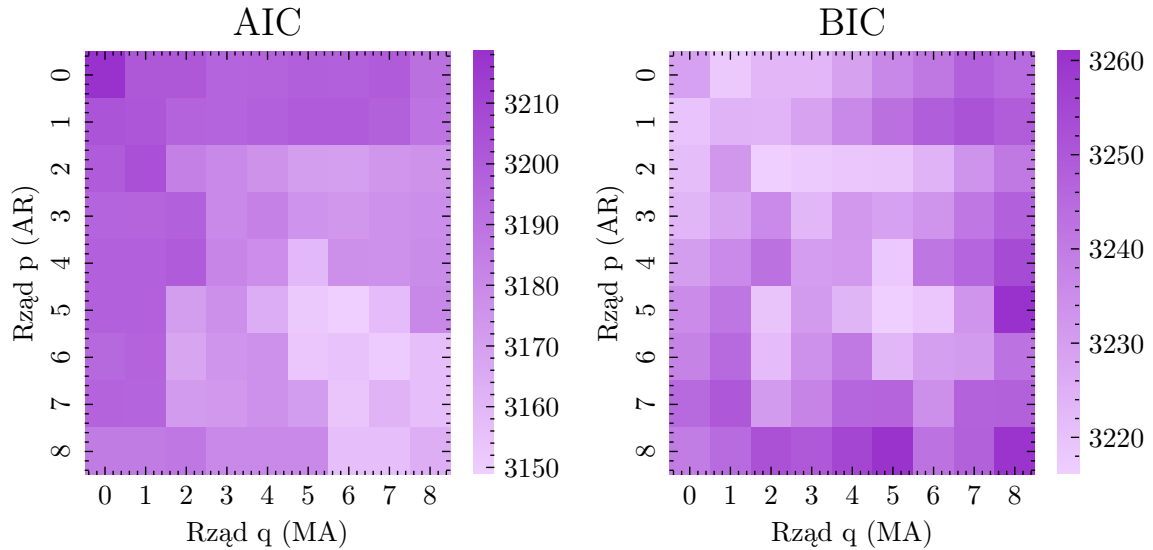
gdzie  $\epsilon_t$  jest białym szumem. Dobór rzędów  $p$  i  $q$  przeprowadzono na podstawie kryteriów informacyjnych Akaikego ( $AIC$ ) oraz Bayesowskiego ( $BIC$ ), zdefiniowanych jako:

$$AIC = -2 \ln(L) + 2k, \quad BIC = -2 \ln(L) + k \ln(n), \quad (7)$$

gdzie  $L$  jest wartością funkcji wiarygodności,  $k$  liczbą parametrów, a  $n$  liczbą obserwacji.

Rysunek 8 przedstawia wartości kryteriów informacyjnych AIC oraz BIC dla różnych kombinacji rzędów  $p$  i  $q$ . Analiza map ciepła wskazuje, że minimum kryterium AIC osiągnięto dla modelu  $\text{ARMA}(5,6)$ , natomiast minimum kryterium BIC przypada dla modelu  $\text{ARMA}(5,5)$ . Biorąc pod uwagę zbliżone wartości kryteriów, do dalszej analizy wybrano model  $\text{ARMA}(5,5)$ .

Parametry modelu zostały wyznaczone metodą największej wiarygodności (MLE), a ich wartości zestawiono w tabeli 1.

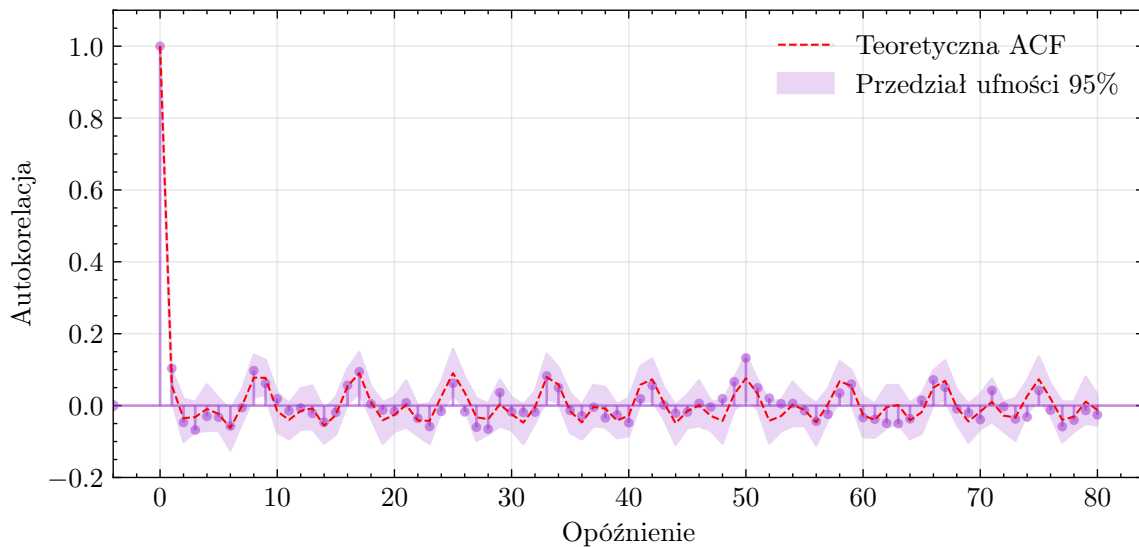


Rysunek 8: Heatmapa wartości kryteriów AIC oraz BIC w zależności od rzędów  $p$  i  $q$  modelu  $\text{ARMA}(p, q)$  badanego szeregu. Jaśniejszy kolor oznacza niższą wartość kryterium.

## 3.2 Ocena dopasowania modelu

### 3.2.1 Porównanie empirycznych i teoretycznych ACV i PACV

Rysunki 9 i 10 przedstawiają porównanie empirycznych funkcji ACF i PACF z ich teoretycznymi odpowiednikami dla dopasowanego modelu ARMA(5,5). Widoczne na wykresach granice przedziałów ufności wyznaczono metodą Monte Carlo, generując  $N = 1000$  trajektorii procesu o parametrach zgodnych z oszacowanym modelem. Obserwowana wysoka zgodność przebiegów oraz fakt, że wartości empiryczne w większości mieszczą się w wyznaczonych symulacyjnie granicach, świadczą o poprawnym odwzorowaniu struktury zależności czasowych przez model.



Rysunek 9: Wykres empirycznego i teoretycznego współczynnika autokorelacji modelu ARMA(5, 5) dopasowanego do zbioru treningowego, w zależności od opóźnienia (parametru  $h$ ). Półprzezroczystym obszarem zaznaczono przedział ufności o poziomie 95% wyznaczony metodą Monte Carlo ( $N = 1000$ ) na podstawie wygenerowanych trajektorii.

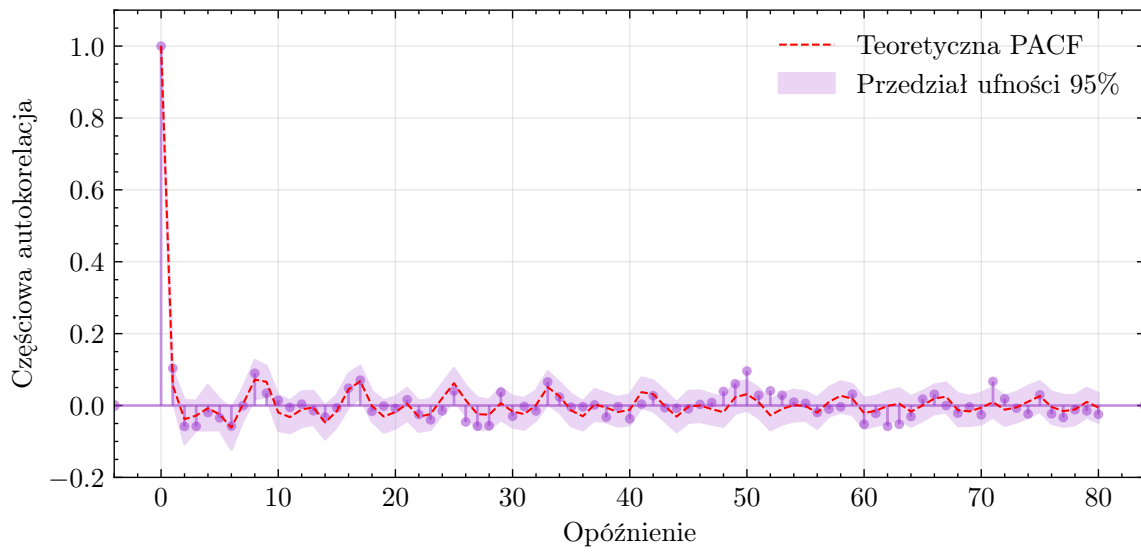
### 3.2.2 Analiza trajektorii ARMA w porównaniu do badanego szeregu

Rysunek 11 przedstawia porównanie trajektorii rzeczywistego szeregu po różnicowaniu z dopasowaniem modelu ARMA(5,5) oraz wyznaczonymi przedziałami ufności na poziomie 95%. 96.56% rzeczywistych obserwacji mieści się wewnątrz wygenerowanych granic przedziału ufności, co oznacza, że model poprawnie odwzorowuje dynamikę trajektorii.

Rysunek 12 przedstawia porównanie prognozy modelu dla zbioru testowego z rzeczywistym przebiegiem szeregu oraz wyznaczonymi przedziałami ufności na poziomie 95%. 90.80% rzeczywistych obserwacji mieści się wewnątrz wygenerowanych granic przedziału ufności, co

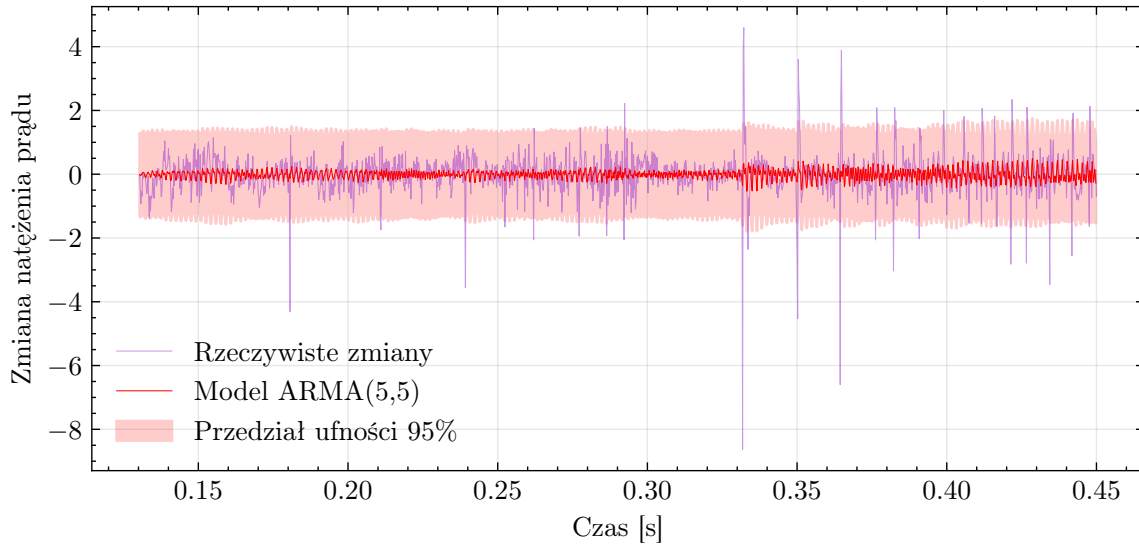
Tabela 1: Parametry modelu ARMA(5, 5) dopasowanego do szeregu ze zbioru treningowego, wyznaczone metodą największej wiarygodności (MLE).

AR(5)	$\phi_1$	$\phi_2$	$\phi_3$	$\phi_4$	$\phi_5$
	-0.57475292	0.61155579	0.5790856	-0.55870469	0.97329654
MA(5)	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_4$	$\theta_5$
	-0.53449985	0.5807027	0.601362	-0.55516097	0.9431987



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% wyznaczony metodą Monte Carlo ( $N = 1000$ ) na podstawie wygenerowanych trajektorii.

pomimo niewielkiego odchylenia od założonego poziomu, oznacza, że model zachowuje wysoką zdolność predykcyjną na nowych danych.



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

## 4 Weryfikacja założeń dotyczących szumu

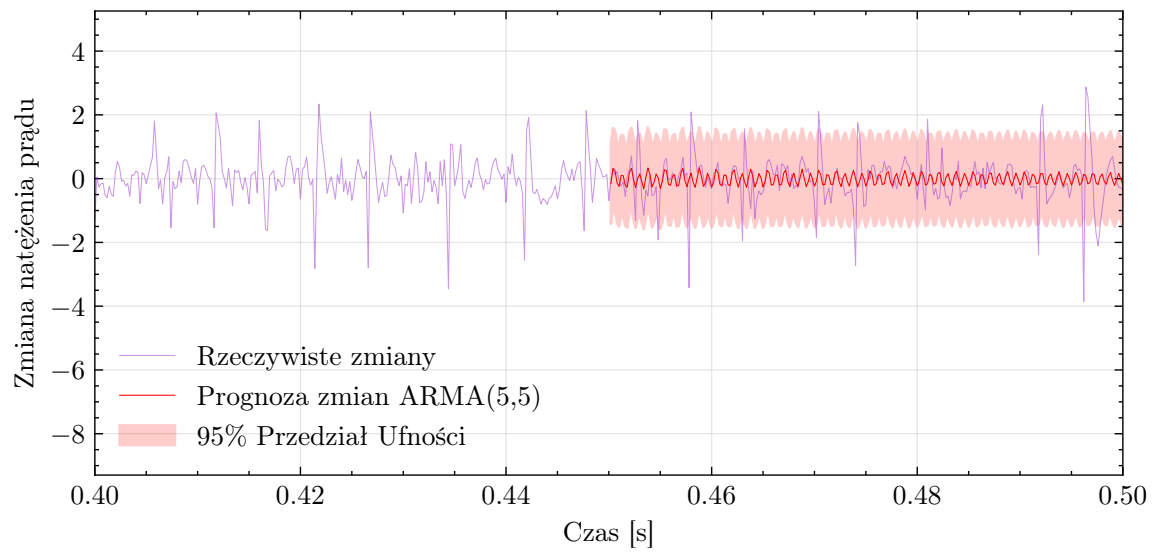
Reszty (błędy) w modelu ARMA/ARIMA reprezentują różnicę między obserwowanymi wartościami szeregu czasowego a wartościami przewidywanymi (dopasowanymi) przez model. Poprawne reszty powinny zachowywać się jak „biały szum”, czyli mieć stałą wariancję, zerową średnią, brak autokorelacji i wykazywać normalność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

### 4.1 Założenie dotyczące średniej

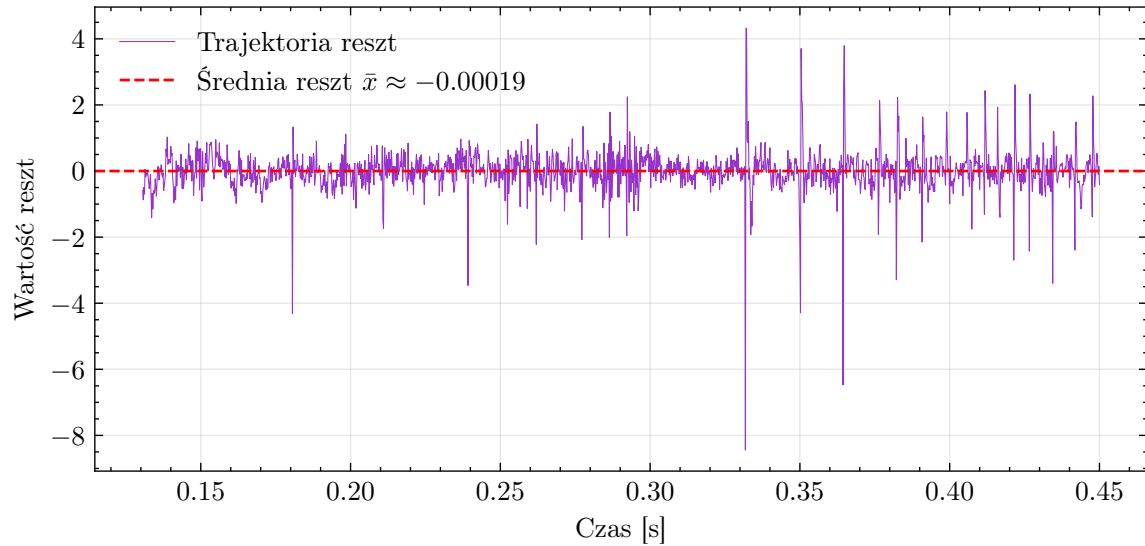
W celu weryfikacji nieobciążoności modelu przeprowadzono test t-Studenta dla średniej wartości reszt. Test ten sprawdza hipotezę zerową  $H_0 : \mu = 0$ , zakładającą, że wartość oczekiwana błędu jest równa zero, co oznacza brak systematycznego trendu w błędach prognozy. Otrzymana wartość  $p = 0.9389$  nie daje podstaw do odrzucenia hipotezy zerowej, co potwierdza, że średnia reszt jest statystycznie równa zero (rysunek 13).

### 4.2 Założenie dotyczące wariancji

W celu weryfikacji założenia o stałej wariancji reszt zastosowano test ARCH (Autoregressive Conditional Heteroskedasticity). Test ten sprawdza, czy wariancja reszt w danym momencie



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  $-0.00123$ .

czasu zależy od kwadratów reszt z poprzednich chwil, co jest charakterystyczne dla procesów wykazujących heteroscedastyczność (niejednorodność wariancji).

Hipoteza zerowa testu ARCH zakłada brak efektu ARCH, czyli stałą wariancję reszt w czasie. Odrzucenie hipotezy zerowej oznacza występowanie heteroskedastyczności warunkowej, czyli zmiennej wariancji zależnej od przeszłych zaburzeń losowych.

Otrzymana wartość  $p = 2.0279 \times 10^{-32} \approx 0$  prowadzi do jednoznacznego odrzucenia hipotezy zerowej, co wskazuje na obecność efektu ARCH w resztach modelu ARMA. Oznacza to, że wariancja procesu nie jest stała w czasie, a model ARMA nie opisuje w pełni struktury zmienności danych.

W celu lokalizacji zmiany reżimu wariancji wykorzystano sumy skumulowanych kwadratów reszt:

$$C_j = \sum_{i=1}^j \epsilon_i^2, \quad (8)$$

oraz funkcję błędu:

$$V(j) = \frac{j}{n} C_n - C_j. \quad (9)$$

Rysunek 14 przedstawia trajektorię szumu (reszt), przebieg funkcji skumulowanych kwadratów  $C_j$  oraz wykres funkcji  $V(j)$ . Wyznaczony na rysunku punkt  $l$  odpowiada momentowi, w którym funkcja  $V(j)$  osiąga ekstremum, co wskazuje na punkt największej zmiany wariancji w analizowanym procesie [4].

### 4.3 Założenie dotyczące niezależności

Do weryfikacji założenia o niezależności reszt zastosowano test Ljunga–Boxa (przyjmując rząd opóźnień  $h = 10$ ), który bada, czy grupa autokorelacji do zadanego opóźnienia różni się istotnie od zera. Test ten sprawdza hipotezę zerową, że reszty są nieskorelowane, czyli mają charakter białego szumu.

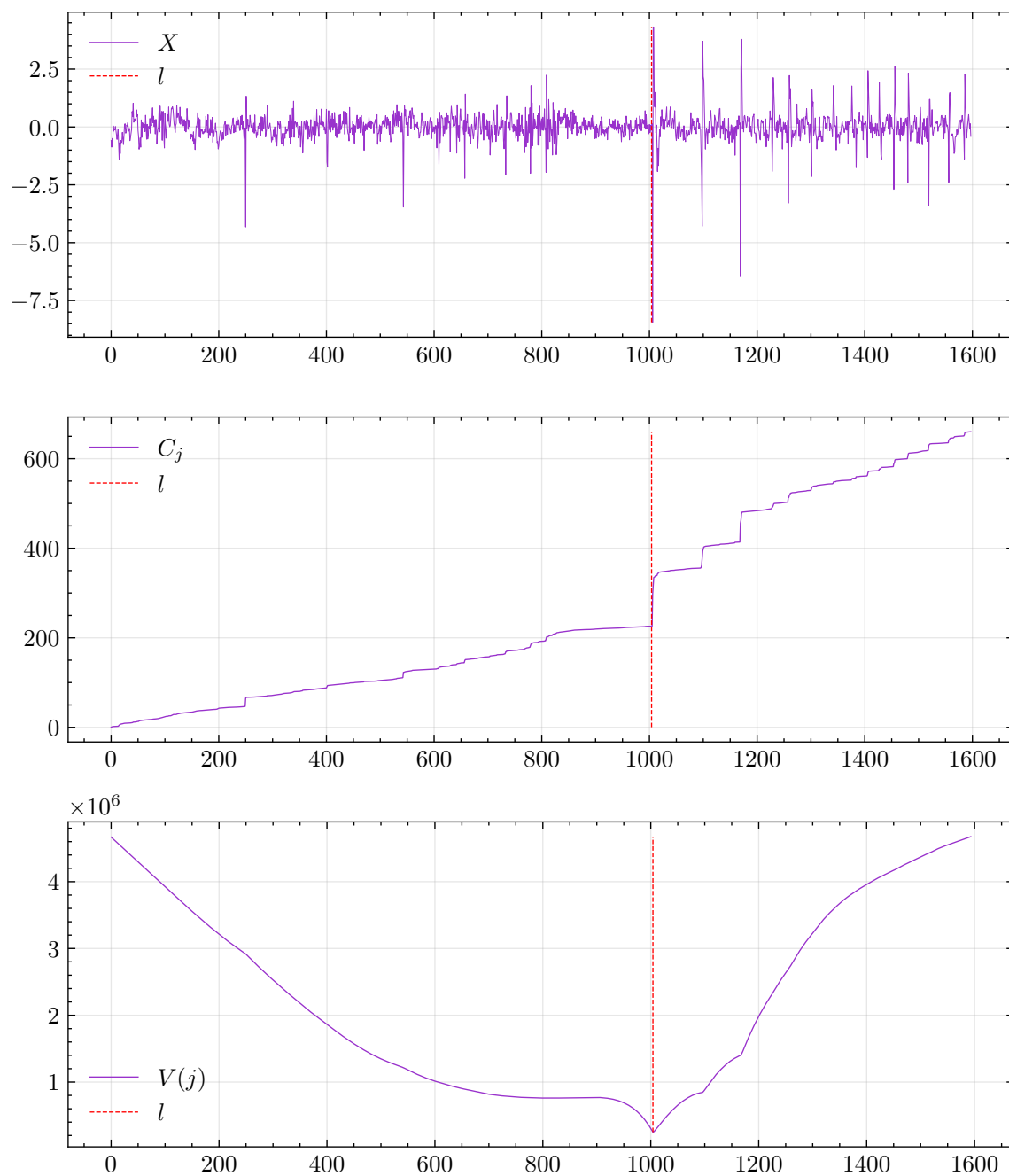
Otrzymana wartość  $p = 0.2137$  nie daje podstaw do odrzucenia hipotezy zerowej, co oznacza, że w resztach nie występuje istotna autokorelacja, a model poprawnie opisuje zależności czasowe w średniej procesie.

#### 4.3.1 Autokorelacja reszt

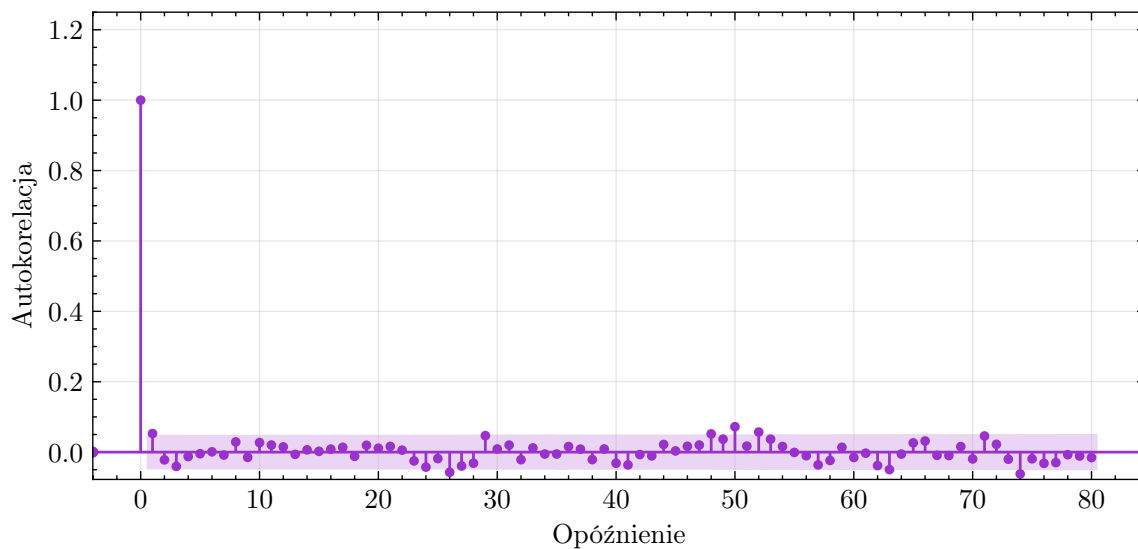
Rysunki 15 i 16 przedstawiają wykresy ACF i PACF reszt. Wszystkie współczynniki mieszczą się w granicach przedziałów ufności (dla hipotezy o braku korelacji), co potwierdza brak autokorelacji i spełnienie założenia białego szumu.

### 4.4 Założenie dotyczące normalności rozkładu

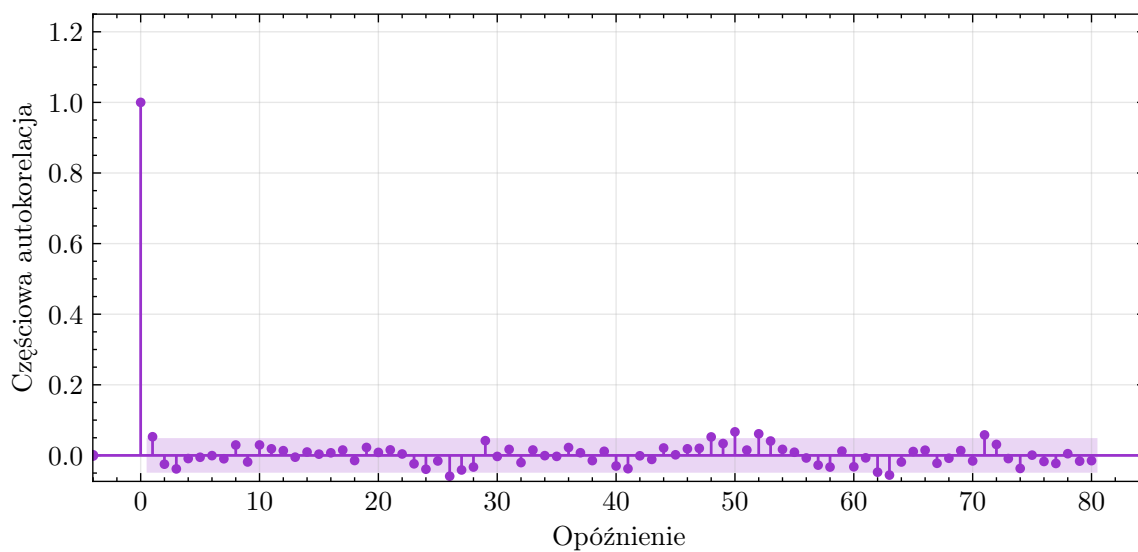
Normalność rozkładu reszt zweryfikowano przy użyciu testu Jarque–Bera, który opiera się na analizie skośności i kurtozy rozkładu. Hipoteza zerowa testu zakłada zgodność rozkładu reszt z rozkładem normalnym.



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



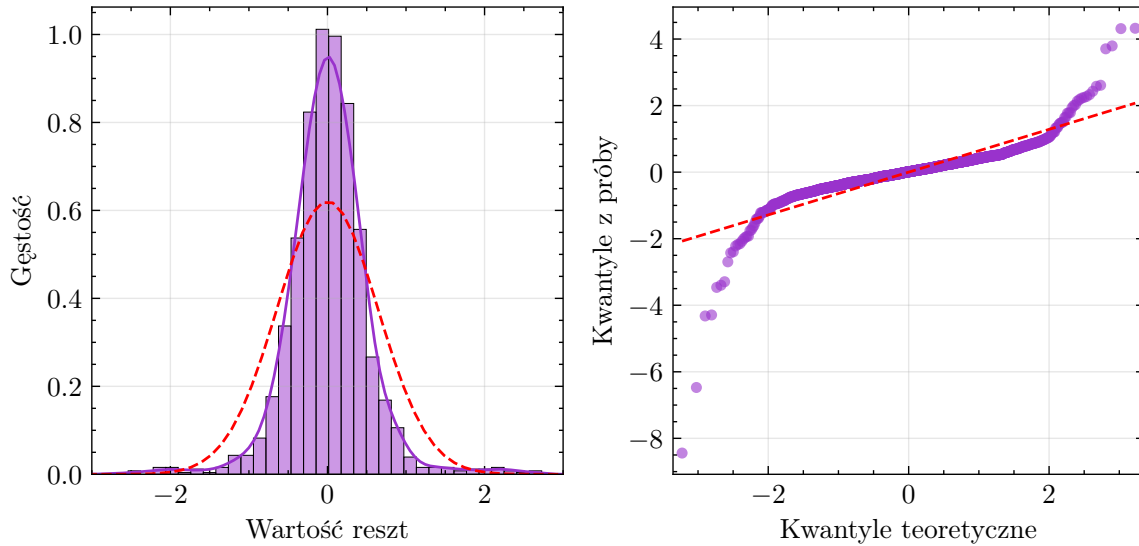
Rysunek 15: Wykres współczynnika autokorelacji reszt modelu ARMA(5, 5) w zależności od opóźnienia (parametru  $h$ ). Półprzezroczystym obszarem zaznaczono przedział ufności o poziomie 95% dla hipotezy o braku korelacji (szum biały).



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% dla hipotezy o braku korelacji (szum biały).



Otrzymana wartość  $p = 0$  prowadzi do odrzucenia hipotezy zerowej, co oznacza, że rozkład reszt istotnie odbiega od normalnego. Tego typu zachowanie jest często obserwowane w danych pochodzących z procesów fizycznych o charakterze turbulentnym. Histogram oraz wykres kwantylowy (QQ-plot) (Rysunek 17) wskazują na cięższe ogony rozkładu, co jest typowe dla omawianego zjawiska.



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).

## 5 Wnioski

Przeprowadzona analiza natężenia prądu plazmy w fazie stabilnej eksperymentu tokamakowego pozwoliła na ocenę możliwości opisu badanego szeregu czasowego przy użyciu modelu ARMA. Wykazano, że surowe dane są niestacjonarne, jednak zastosowanie różnicowania pierwszego rzędu skutecznie doprowadziło do uzyskania szeregu stacjonarnego, co potwierdziły zarówno analiza ACF i PACF, jak i test Augmented Dickey–Fullera.

W pracy zastosowano klasyczne metody analizy szeregów czasowych do modelowania rzeczywistych danych pochodzących z eksperymentu fizycznego. Uzyskane wyniki potwierdzają użyteczność modeli ARMA w opisie krótkoterminowej dynamiki badanego zjawiska, wskazując jednocześnie na granice ich stosowalności wynikające z silnej heteroskedastyczności oraz odstępstw od rozkładu normalnego.

Jednocześnie weryfikacja założeń modelowych ujawniła istotne ograniczenia. Pomimo spełnienia założeń o zerowej średniej i braku autokorelacji reszt, stwierdzono występowanie zmiennej wariancji oraz odstępstwa od normalności rozkładu. Wskazuje to, że model ARMA

nie opisuje w pełni statystycznych własności fluktuacji prądu plazmy.

W związku z tym dalsze badania powinny obejmować wykorzystanie modeli uwzględniających zmienność wariancji, takich jak rodzina modeli GARCH (w tym warianty asymetryczne np. EGARCH, TGARCH) czy modele stochastycznej zmienności (SV), które mogą lepiej odpowiadać naturze analizowanych danych.

## Literatura

- [1] 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.
- [2] 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.
- [3] Skipper Seabold and Josef Perktold. statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*, 2010.
- [4] J. Gajda, G. Sikora, and A. Wylomańska. Regime variance testing — a quantile approach. *Acta Physica Polonica B*, 44(5):1015, 2013.