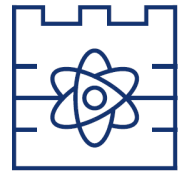




**Politechnika Krakowska  
im. Tadeusza Kościuszki**

Wydział Inżynierii Materiałowej i Fizyki



**Karolina Krupińska**

Nr albumu: 113892

**Długofalowa analiza zmian klimatu na podstawie  
danych z rdzeni lodowych. Zastosowanie metod  
uczenia maszynowego.**

**Long-term analysis of climate change based on ice  
core data. Application of machine learning  
methods.**

Praca magisterska  
na kierunku FIZYKA TECHNICZNA

Praca wykonana pod kierunkiem  
**dra Radosława Kyci**  
Instytut Fizyki

Uzgodniona ocena: .....

.....  
podpisy promotora i recenzenta

Kraków 2021

*Niniejszą pracę dedykuję moim kochanym rodzicom Ewie i Arkadiuszowi, dzięki którym miałam możliwość kształcić się i zdobywać cenną wiedzę na Politechnice Krakowskiej, którzy stale mnie mobilizowali i wspierali podczas trwania studiów.*

*Za nieocenioną pomoc oraz inspirację, pragnę złożyć serdecznie podziękowania dla mojego promotora i wykładowcy dr Radosława Kyci.*

# Spis treści

<b>Wstęp</b>	<b>3</b>
0.0.1 Cel . . . . .	3
0.0.2 Zakres . . . . .	3
0.0.3 Metodyka . . . . .	3
<b>1 Część teoretyczna</b>	<b>4</b>
1.1 Klimat . . . . .	5
1.1.1 Temperatura powietrza atmosferycznego . . . . .	5
1.1.2 Efekt cieplarniany . . . . .	6
1.2 Cykle klimatyczne . . . . .	7
1.2.1 Cykle lodowcowe i międzylodowcowe . . . . .	7
1.2.2 Cykl słoneczny . . . . .	8
1.2.3 Cykle klimatyczne w skali tysiąclecia . . . . .	9
1.2.4 Cykle klimatyczne w skali stulecia . . . . .	10
1.2.5 Cykle klimatyczne na przestrzeni dekad . . . . .	10
1.3 Odwierty w lodowcach . . . . .	12
1.3.1 Rdzenie lodowe . . . . .	12
1.3.2 Wiercenie rdzeni lodowych . . . . .	13
1.4 Zastosowanie języka Python w analizie danych . . . . .	15
1.4.1 Analityka danych . . . . .	15
1.4.2 Data Science - Języki Programowania . . . . .	15
1.4.3 Python . . . . .	16
1.4.4 Uczenie maszynowe . . . . .	17
1.5 Dyskretna Transformata Fouriera . . . . .	19
<b>2 Część praktyczna</b>	<b>20</b>
2.1 Analiza temperatury w latach 1850-2015 . . . . .	21
2.1.1 Opis zbioru danych . . . . .	21
2.1.2 Wizualizacja danych . . . . .	23
2.1.3 Dopasowanie metodą regresji liniowej . . . . .	32
2.2 Analiza stężenia $CO_2$ w latach 1850-2015 . . . . .	34
2.3 Analiza Fouriera . . . . .	41
2.4 Wnioski . . . . .	44
2.5 Dodatek . . . . .	45
<b>Bibliografia</b>	<b>63</b>

# Wstęp

Zmiany klimatyczne na Ziemi oraz przyczyny ich występowania są jednymi z najbardziej popularnych zagadnień współczesnej nauki. Zmiany te ustalane są na podstawie badania zmiennych takich jak wielkość opadów atmosferycznych, temperatura czy w siła wiatru [1]. Gromadzenie danych z wieloletnich obserwacji pogody pozwala na określenie klimatu dla danego regionu.

## 0.0.1 Cel

Głównym celem pracy jest analiza jednego ze składników klimatu - temperatury oraz jej związkami z  $CO_2$ . Dzięki rozpoczętym w przeszłości systematycznym pomiarom tej wielkości fizycznej możemy obecnie przeprowadzić analizę zgromadzonych danych, która pozwoli na prześledzenie zmian temperatury jakie występowały w przeszłości oraz prognozowanie jej wartości w przyszłości.

## 0.0.2 Zakres

W niniejszej pracy przeanalizowane zostaną dane dotyczące średniej temperatury lądu a także zostanie wykonana analiza danych dotyczących średniej temperatury lądu i oceanów oraz jej porównanie ze średnią temperaturą lądu w latach od 1850 do 2015 roku, ze szczególnym uwzględnieniem okresu po 1960 roku. Przy pomocy narzędzi nadzorowanego uczenia maszynowego (regresji) uzyskane zostaną tendencje zmian średniej temperatury lądu. Przeanalizowane zostaną niepewności pomiarowe: średniej, minimalnej oraz maksymalnej temperatury lądu jak również niepewności średniej temperatury lądu i oceanów. Wykonana zostanie również animacja mapy świata z zaznaczoną temperaturą w danym regionie dla określonego roku. Na podstawie przeprowadzonego badania, wyznaczona zostanie także predykcja wartości średniej temperatury lądu na najbliższe trzydzieści lat. Panuje powszechne przekonanie, że istnieje zależność między stężeniem  $CO_2$  w atmosferze a temperaturą na Ziemi [2]. W analizie zbadana została korelacja średniej temperatury lądu wraz z stężeniem  $CO_2$  w atmosferze w danym roku. Na podstawie danych o stężeniu dwutlenku węgla w określonym roku przeprowadzono w pracy również predykcję wartości temperatury w tymże roku. Finalnie zrealizowana została także dyskretna transformacja Fouriera dla szeregu czasowego średniej temperatury lądu oraz stężenia  $CO_2$  w atmosferze w celu sprawdzenia periodyczności tych wielkości.

## 0.0.3 Metodyka

W pracy wykorzystane zostały biblioteki języka Python takie jak: Scikit-learn, Pandas, SciPy, NumPy, Matplotlib, które pozwalają na użycie nowoczesnych metod analizy oraz wizualizacji danych będących obecnie standardem w Data Science. Implementacja kodu wykonana została w środowisku Jupiter Notebook. Analizowane dane pochodzą z platformy Kaggle [3].

# Rozdział 1

## Część teoretyczna

## 1.1 Klimat

Zjawiska pogodowe oraz towarzyszące im procesy atmosferyczne, które stale utrzymują się na określonym obszarze nazywamy klimatem. Kształtowany jest on przez trzy kluczowe czynniki: obieg wody, obieg ciepła oraz cyrkulacje powietrza. Do jego określenia potrzebne są pomiary i dotyczą one: opadów atmosferycznych, temperatury oraz wiatru. Klimat wyznaczyć można tylko na podstawie długotrwałych obserwacji nie krótszych niż 30 lat [4].

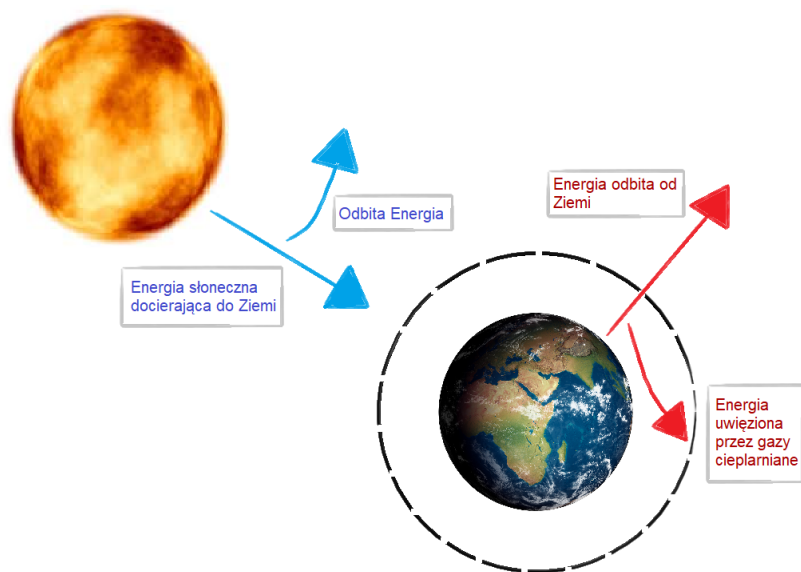
### 1.1.1 Temperatura powietrza atmosferycznego

Czynnikami wpływającymi na zróżnicowanie temperatury powietrza atmosferycznego są:

- szerokość geograficzna - wraz ze wzrostem szerokości geograficznej zmniejsza się wysokość Słońca, spada ilość dostarczanej energii słonecznej, a zatem spada temperatura powietrza.
- wysokość nad poziomem morza - wraz ze wzrostem wysokości nad poziomem morza temperatura zmniejsza się ok.  $0.6^{\circ}\text{C}$  na 100 m w powietrzu wilgotnym, natomiast przy powietrzu suchym spadek temperatury odnotowuje się na ok.  $1^{\circ}\text{C}$  na 100 m.
- rodzaj podłoża - woda powoli się nagrzewa, jak i powoli oddaje ciepło, natomiast ląd szybko się nagrzewa i szybko oddaje ciepło. Ciemne podłoże nagrzewa się szybciej pochłaniając większą ilość promieni słonecznych w porównaniu do jasnego podłoża, które to odbija znaczną część promieniowania.
- prądy morskie - prowadzą do powstawania anomalii termicznych. Ciepłe prądy morskie podnoszą temperaturę powietrza, a ich obecność kieruje izotermę na mapach w stronę wysokich szerokości geograficznych. Z kolei zimne prądy morskie obniżają temperaturę powietrza, co powoduje, że izotermę na mapach skierowane są w stronę niskich szerokości geograficznych.
- odległość od morza - wraz ze wzrostem odległości od mórz i oceanów roczne temperatury powietrza rosną, lata stają się cieplejsze, natomiast zimy chłodniejsze.
- rzeźba powierzchni Ziemi - stoki które mają ekspozycję południową przyjmują znacznie większą ilość promieniowania słonecznego i są znacznie cieplejsze w porównaniu do stoków o ekspozycji północnej. W dolinach oraz kotlinach występują zastoiska chłodnego powietrza. Powietrze chłodne jako cięższe gromadzi się u podnóża, w związku z tym tworzy inwersję temperatury. Przy powierzchni Ziemi zalega powietrze chłodne i ciężkie, a powyżej powietrze ciepłe, stan taki jest możliwy gdy nie ma prądów pionowych i wiatru. W dolinach powstają lokalne wiatry górskie, w Europie przykładem tego zjawiska może być pasmo górskie Alp; które zatrzymuje ciepłe masy powietrza z nad Afryki i ogranicza ich przepływ do Europy Środkowej [5].

### 1.1.2 Efekt cieplarniany

Efekt cieplarniany nazywamy wzrost temperatury powietrza przy powierzchni Ziemi, który spowodowany jest wzrostem stężenia gazów cieplarnianych. Efekt ten związany jest z promieniowaniem zwrotnym atmosfery. Promieniowanie słoneczne przechodzi przez atmosferę bez poważnych strat, natomiast promieniowanie ziemskie zostaje w przeważającej części pochłonięte w atmosferze i następnie wypromieniowane ku powierzchni Ziemi jako dodatkowy strumień energii, wpływając na wzrost temperatury powietrza. Promieniowanie zwrotne atmosfery rośnie wraz ze zwiększaniem się w nim zawartości wody, pary wodnej oraz gazów cieplarnianych np.  $CO_2$ , tlenków azotu, metanu i freonów [2]. Naturalny efekt cieplarniany jest bardzo istotnym zjawiskiem i przynosi korzystne skutki dla formowania warunków życia na Ziemi. Szacuje się, że zwiększa on temperaturę powierzchni naszej planety od  $20^{\circ}C$  do  $34^{\circ}C$ . Obecnie średnia temperatura Ziemi wynosi około  $15^{\circ}C$  [6], w przypadku braku efektu cieplarnianego temperatura na planecie wyniosłaby około  $-19^{\circ}C$ . Ciekawe jest to, że efekt ten nie występuje tylko na Ziemi. Znajdująca się w atmosferze Wenus para wodna absorbowała promieniowanie, które było emitowane przez planetę. Spowodowało to zwiększenie temperatury na planecie. Wysokie stężenie pary wodnej zgromadzone w górnych warstwach atmosfery pozwoliło na zajście fotodysocjacji. W związku z tym wodór, który jest lżejszy ulotnił się z atmosfery natomiast tlen reagował ze skałami powierzchniowymi. Woda znajdująca się na planecie nie uległa tak szybko fotodysocjacji, dzięki czemu deuter nie uciekł z atmosfery. Stosunek ilości deuteru do wodoru na Wenus jest aż 150 razy większy niż na Ziemi [7].



Rysunek 1.1: Uproszczony schemat efektu cieplarnianego.

## 1.2 Cykle klimatyczne

### 1.2.1 Cykle lodowcowe i międzylodowcowe

W przeszłości wielkie kontynentalne pokrywy lodowe na półkuli północnej wielokrotnie zmieniały swoją objętość zwiększając się i kurcząc. Czasy gdy wielkie pokrywy lodowe zajmują znaczne obszary Ziemi nazywamy okresami lodowcowymi (lub epokami lodowcowymi), a czasy bez dużych łądolodów okresami międzylodowcowymi. Okres ostatniego zlodowacenia szacuje się między około 120 000 a 11 500 lat temu. Od tego czasu Ziemia znajduje się w okresie międzylodowcowym, który zwany jest holocenem. Cykle lodowcowe są suche i znacznie zimniejsze, niż okresy międzylodowcowe. Zdarzenia te spowodowane są zmianą orbity Ziemi w czasie co skutkuje zmianą ilość promieniowania słonecznego, jakie Ziemia otrzymuje w każdym sezonie [8].

Jednym z najbardziej znanych przykładów naglej zmiany klimatu jest Młodszy dryas. Około 14500 lat temu zimny klimat Ziemi zaczął gwałtownie zmieniać się na cieplejszy wchodząc w okres międzylodowcowy. W połowie tego przejścia temperatury na półkuli północnej nagle powróciły do podobnych jak za czasów ubiegłej epoki lodowcowej. Koniec młodszego dryasu nastąpił około 11500 lat temu i jego skutki były odczuwalne chociażby na Grenlandii, gdzie w ciągu dekady temperatury wzrosły o  $10^{\circ}\text{C}$  [9].

Od około 11500 lat temu, klimat naszej Ziemi zaczął zachowywać się bardzo podobnie do znanego nam dziś. Łądolody skurczyły się, temperatura wzrosła, podniósł się poziom mórz, a monsuny przybrały na sile. Około 8200 lat temu miało miejsce zdumiewające wydarzenie, które to zostało nazwane "8.2 ka Event" i zostało po raz pierwszy odkryte w grenlandzkim rdzeniu lodowym GISP2, gdzie analizy o wysokiej rozdzielczości wskazały, że w przeciągu dwóch dekad temperatura spadła o około  $3.3^{\circ}\text{C}$  na Grenlandii. Wydarzenie to mogło przyczynić się do powstania cywilizacji Sumerijskiej. Przez susze oraz ochładzanie klimatu nastąpił nagły rozwój rolnictwa między innymi powstały systemy nawadniające uprawy [10]. Wydarzenie to trwało około 150 lat a następnie temperatury wzrosły, powracając do poprzedniego poziomu [11]. Podnosząca się temperatura, na Ziemi, doprowadziła do rozszerzenia się osad ludzkich. W Afryce deszcze monsunowe widocznie nasiliły się i rozprzestrzeniły na północ nad Saharę. Piaszczysta pustynia, którą teraz znamy, była stepem pokrytym trawami i krzewami. W sprzyjających rozwojowi kultury warunkach ludzie rozpoczęli udomawianie bawołów i kóz oraz opracowali wczesny system sztuki symbolicznej [12].

Jednym z najważniejszych wydarzeń w historii ludzkości, był tak zwany 4,2 - kiloyear event (wydarzenie 4.2 kiloroku), kiedy wielkie cywilizacje na całym świecie popadły w anarchię i społeczny chaos. Od tego momentu rozpoczęło się ochładzanie klimatu i powszechne osuszanie, obniżając produkcję żywności rolnej i komfort życia ludzi. Przypuszcza się, że susze oraz głód doprowadziły do upadku Starego Królestwa Egipskiego, Imperium Akadyjskiego w Mezopotamii i kultury Liangzhu [13].

W latach 2600 p.n.e. a 1200 n.e. pochodząca z Półwyspu Jukatan starożytna cywilizacja Majów zajmowała szeroki obszar Mezoameryki. Osiągnęła kulturowy rozkwit między 600 a 800 rokiem n.e., rozwijając wyrafinowane koncepcje astronomiczne i matematyczne oraz konstruując tysiące budowli architektonicznych. W latach 800 a 950 rokiem n.e. wiele południowych miast zostało opuszczonych, a okres ten przeszedł do historii jako upadek klasycznej cywilizacji Majów. Przyczyna do tej pory jest jedną z największych archeologicznych tajemnic, a uczeni debatują nad nią od prawie wieku. Według badań najprawdopodobniej to okres intensywnej suszy doprowadził do nieszczęścia i upadku wielkiej cywilizacji Majów [14].

Warto więc skupić się na analizie zmian klimatu na przestrzeni lat, gdyż to one w głównej mierze kształtowały historię naszej cywilizacji.

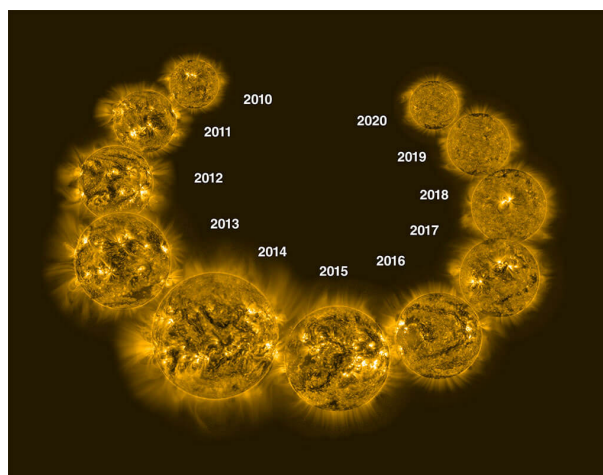


### 1.2.2 Cykl słoneczny

Nasze Słońce jest ogromną kulą składającą się z gorącej plazmy. Gaz ten porusza się wytwarzając przy tym ogromne pole magnetyczne, które przechodzi cykl zwany cyklem słonecznym. Co około 11 lat pole magnetyczne Słońca zmienia się całkowicie, północny i południowy biegun Słońca zamieniają się miejscami. Następnie musi minąć kolejne 11 lat, zanim północne i południowe bieguny znów się odwrócą. Cykl ten znacznie wpływa na aktywność na powierzchni Słońca przykładem mogą być plamy słoneczne, które są spowodowane przez pola magnetyczne Słońca. Wraz ze zmianą pól magnetycznych zmienia się także aktywność na powierzchni.

Obliczenie ilości plam słonecznych jest jednym ze sposobów śledzenia tegoż cyklu. Jego początek to minimum słoneczne lub gdy Słońce ma najmniej plam. Wraz ze zwiększającą się liczbą plam słonecznych rośnie jego aktywność. Maksimum słoneczne osiąga w środku cyklu lub kiedy ilość plam słonecznych jest największa. Gdy cykl się kończy, zanika z powrotem do minimum słonecznego, a następnie rozpoczyna się nowy cykl.

W czasie trwania tego okresu nasilają się również ogromne erupcje na Słońcu, takie jak rozbłyski słoneczne i koronalne wyrzuty masy. Erupcje te wysyłają w przestrzeń kosmiczną potężne wybuchy energii i materiału. Aktywność ta może mieć duży wpływ na naszą planetę. Przykładem mogą być erupcje, przez które można zobaczyć światła na niebie zwane zorzą polarną bądź takie, które wpływają na komunikację radiową. Ekstremalne koronalne wyrzuty masy mogą nawet wpływać na sieci elektroenergetyczne na Ziemi [15]. Istnieją też cykle z dłuższymi okresami: 88, 210, 2300 lat, które tłumaczą powstawanie epok lodowcowych i innych zmiany klimatu obserwowanych na planecie [16]. Stałą słoneczną nazywamy ilość energii promieniowania słonecznego (poza atmosferą Ziemi) w jednostkowym czasie, które dociera do powierzchni ustawionej prostopadle do promieniowania w odległości 1 AU. Wartość stałej słonecznej w maksimum aktywności jest równa  $1367 \text{ W/m}^2$ , zależna od ilości ciemnych plam na Słońcu im jest ich najwięcej tym stała ta osiąga najwyższą wartość [17].



Rysunek 1.2: Ewolucja Słońca w ekstremalnym świetle ultrafioletowym w latach 2010 do 2020, wykonana teleskopem znajdującym się na pokładzie europejskiej sondy PROBA2. Źródło: Dan Seaton / Europejska Agencja Kosmiczna (Kolaż: NOAA / JPL-Caltech).

### 1.2.3 Cykle klimatyczne w skali tysiąclecia

Okresy lodowcowe główne (zimne) i międzylodowcowe, którym przypisuje się charakter ciepłych, inicjowane są przez zmiany orbity Ziemi wokół Słońca. Zostały one nazwane cyklami Milankovitcha - od nazwiska serbskiego naukowca, który jako pierwszy opisał te cykle jako naczelną mechanizm występowania epok lodowcowych [18]. To właśnie im przypisuje się wpływ na przebieg minionych epok lodowcowych wywołanym zmianami orbity Ziemi. Wyróżniamy trzy główne zmiany orbity naszej planety:

- Ekscentryczność czyli kształt orbity Ziemi wokół Słońca waha się między prawie okrągłym a eliptycznym kształtem.
- Nachylenie Ziemi względem Słońca oscyluje między  $22.5^{\circ}\text{C}$  a  $24.5^{\circ}\text{C}$ .
- Precesję, czyli obrót Ziemi wokół własnej osi, oś ziemską zakreśla na niebie okrąg i trwa to około 26 000 lat.

Skonsolidowany efekt cykli Milankovitcha powoduje długo trwające zmiany w ilości padającego światła słonecznego na Ziemię, w szczególności na dużych szerokościach geograficznych o różnych porach roku. Przykładem ocieplenia wywołanego przez cykle orbitalne może być stopienie ogromnej ilości lodu ok. 19 000 lat temu na wysokich szerokościach geograficznych w skutek czego słodka woda zalała oceany. Zakłóciło to cyrkulację prądów powierzchniowych i głębokich w Oceanie Atlantyckim (AMOC<sup>1</sup>). W rezultacie wywołując silne zmiany temperatury między półkulami Ziemi [20]. Ląd i oceany ociepliły się jako pierwsze na półkuli południowej, począwszy od około 18 000 lat temu. Wraz ze wzrostem temperatury w oceanach rozpuszczalność  $\text{CO}_2$  spadła w wyniku czego oceany oddały więcej dwutlenku węgla do atmosfery [21]. Istnieje przekonanie że mechanizm emitowania  $\text{CO}_2$  do atmosfery przez oceany wynika z pionowego mieszania się oceanów [22]. Zwiększony poziom  $\text{CO}_2$  w atmosferze przez emisje z oceanów wzmacnia pierwotne ocieplenie. Dwutlenek węgla (ditlenek węgla) z Oceanu Południowego miesza się z atmosferą rozprzestrzeniając ocieplenie na północną półkulę Ziemi [23][24]. Nawet niewielki wzrost promieniowania słonecznego na północnych szerokościach geograficznych naszego globu może doprowadzić do zwiększenia topnienia się lodu, a w konsekwencji zmniejszenia ilości światła słonecznego odbijającego się od powierzchni lodu. Jednocześnie już niewielkie ocieplenie zapoczątkowane przez zmiany orbity Ziemi doprowadza do podwyższenia temperatur w wodach i oceanach, co z kolei implikuje globalny wzrost dwutlenku węgla. Wyższe stężenie dwutlenku węgla prowadzi do zwiększonego ocieplenia Ziemi tworząc tym samym wzmocniony efekt cieplarniany. Takie powtarzające i narastające cykle klimatyczne, będące swoistymi niekontrolowanymi ociepleniami klimatu, towarzyszą Ziemi od tysięcy lat i stanowią kolebkę epok lodowcowych i międzylodowcowych [25].

---

<sup>1</sup>Atlantic Meridional Overturning Circulation co oznacza Atlantycka Południkowa Cyrkulacja Wymienna [19]

#### 1.2.4 Cykle klimatyczne w skali stulecia

Nie wszystkie cykle klimatyczne do swojego zapoczątkowania i rozwoju potrzebują jednak tysięcy lat. Oprócz wielotysięcznych cykli lodowcowych i międzylodowcowych istnieją także krótsze, które rozwijają się na przestrzeni od 200 do 1500 lat. Mechanizmy, które stanowią ich źródło do dziś nie są nam dobrze znane, jednak możliwe jest że napędzane są one przez serie zmian zachodzących na Słońcu wraz z współistniejącymi cyrkulacjami oceanicznymi. Te ostatnie z kolei są efektem nierównomiernego docierania promieni słonecznych do powierzchni Ziemi. Większość energii słonecznej trafia bowiem do obszarów okołorównikowych powodując tym samym dużą dysproporcję w temperaturze pomiędzy poszczególnymi regionami globu. Oceany pełnią w tym procesie znaczącą rolę poprzez transportowanie ciepła z równika ku biegunom. Na cyrkulacje oceaniczną składają się zasadniczo dwie sprzężone ze sobą przyczyny: różnica w gęstości wody oceanów w poszczególnych miejscach oraz prądy morskie. Nie bez znaczenia dla powstania takiego zjawiska jest także temperatura wody oraz stopień jej zasolenia. Takie rodzaje cyrkulacji noszą w literaturze nazwę „termohalinowych” i mają miejsce w rejonach północnej części kuli ziemskiej oraz na wodach wokół Antarktyki [26]. Średniowieczny okres ciepły oraz mała epoka lodowcowa, które przypadały na okres 1450-1900 n.e. są przykładami fazy cieplej oraz zimnej w każdym z tych cykli. Chociaż nie miały one globalnego charakteru, a ich konsekwencje rozpościerały się zaledwie na poszczególne regiony, zrozumienie ich wzorców i ich odtworzenie jest kluczowe dla analizy przyszłych skutków zmian klimatycznych w poszczególnych częściach globu [25].

#### 1.2.5 Cykle klimatyczne na przestrzeni dekad

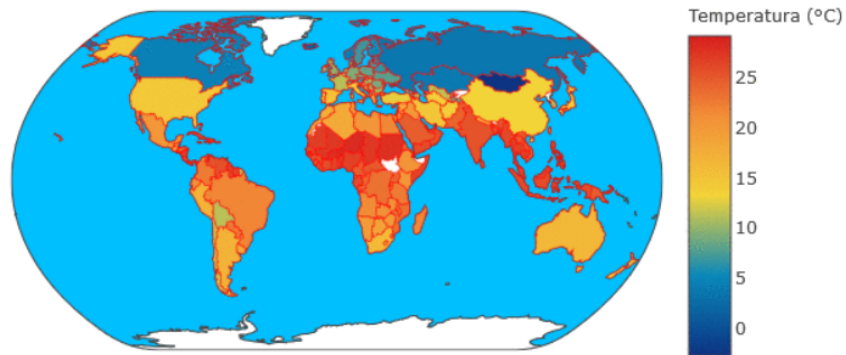
Najkrótszymi i najczęściej spotykanymi są cykle klimatyczne na przestrzeni dekad. Te dynamiczne i regularne zmiany klimatu są efektem interakcji oceanu z atmosferą, a ich efekty zauważalne są już na przestrzeni kilku lat. Jednym z najbardziej znanych tego typu zjawisk klimatycznych jest tzw. Enso<sup>2</sup>, który dzieli się na trzy fazy. Pierwsze dwie z nich są sobie przeciwstawne i noszą nazwy El Nino i La Nina<sup>3</sup>. El Nino polega na ociepleniu oceanu ponad jego przeciętną temperaturę. Zjawisko to dotyczy najczęściej środkowej i wschodniej części Oceanu Spokojnego. Wówczas opady atmosferyczne w rejonie Indonezji zmniejszają się, podczas gdy nad tropikalnym Pacyfikiem ich ilość gwałtownie rośnie. Cykl La Nina jako przeciwstawny poprzedniemu, polega na ochłodzeniu się powierzchni oceanu poniżej jego przeciętnej temperatury. Wówczas opady w rejonie Indonezji rosną na sile, a środkowa część Oceanu Spokojnego znacznie się osusza. Zjawisko to ma miejsce średnio co 3 do 7 lat i spowodowane jest głównie przez zmieniające się warunki pogodowe w różnych częściach świata. Mimo, iż zmiany te nie należą do gwałtownych i długo się utrzymujących poznanie ich może dostarczyć informacji na temat regionalnych skutków zmiany klimatu, co z kolei może pomóc w opracowaniu skutecznych strategii adaptacyjnych zarówno dla ludzi jak i zwierząt zamieszkałych na Ziemi. Cykle, które towarzyszą Nam aktualnie są o wiele bardziej dynamiczne niż poprzedzające je w przeszłości zmiany klimatu. Tempo globalnych zmian klimatycznych zdaje się stale rosnać, co w efekcie powoduje, iż coraz ciężiej jest nam przewidzieć ich efekty [27][28].

---

<sup>2</sup>(El Nino-Southern Oscillation) Oscylacja południowa El Niño) są to wahania temperatury i wiatrów występujące okresowo i nieregularnie na powierzchni morza wschodniego Pacyfiku. Wpływają one znacząco na klimat tropików i strefy podzwrotnikowej [27].

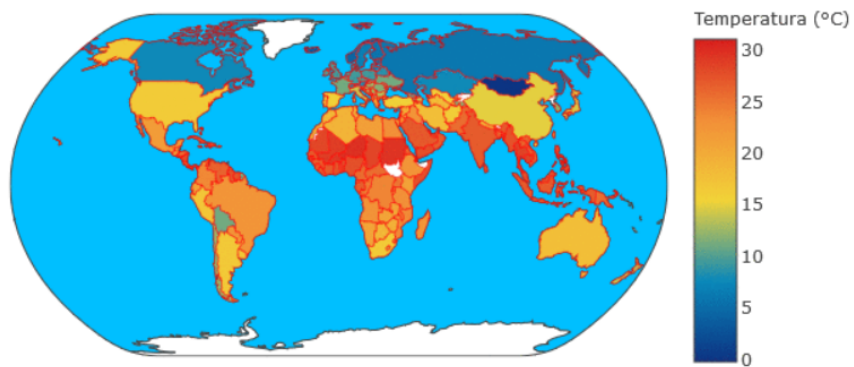
<sup>3</sup>El Nino, La Nina - z języka hiszpańskiego tłumaczenie odpowiednio chłopiec i dziewczyna

Średnia temperatura lądu w 1943 roku



Rysunek 1.3: Średnia temperatura lądu w 1943 roku.

Średnia temperatura lądu w 2013 roku



Rysunek 1.4: Średnia temperatura lądu w 2013 roku.

Rys. 1.3 oraz 1.4 wygenerowane zostały w środowisku Jupyter Notebook i przedstawiają wizualizację mapy Ziemi wraz z temperaturą na danym obszarze w roku 1943 i 2013<sup>4</sup>.

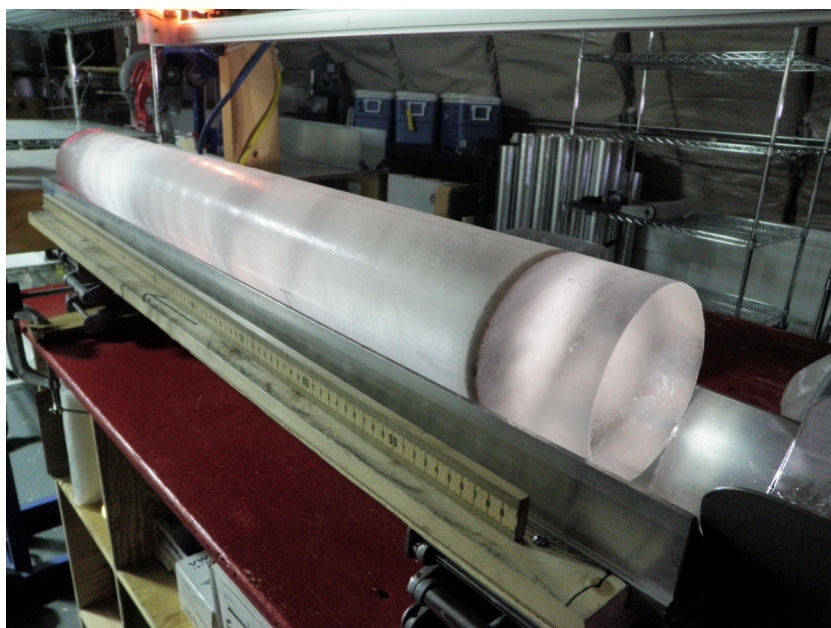
---

<sup>4</sup>W Dodatku do pracy można znaleźć kod źródłowy do wygenerowania powyższej mapy.

## 1.3 Odwierty w lodowcach

### 1.3.1 Rdzenie lodowe

Gromadzące się warstwy śniegu różniące się pod względem chemicznym jak i faktur tworząc lodowce. Warstwa śniegu zgromadzona w okresie letnim różni się od tego w okresie zimowym. Z biegiem czasu pod wpływem ciężaru górnych warstw, śnieg kompresuje się tworząc lód. Częścią lodu stają się pęcherzyki uwiecznionego powietrza, cząstki stałe oraz chemikalia, które zgromadziły się wraz ze spadającym śniegiem. Gromadzenie się warstw lodu latami niezależnie od pory roku, tworzyły swego rodzaju zapis warunków klimatycznych: lokalnej temperatury, stężenia gazów cieplarnianych, aktywności wulkanicznej i słonecznej oraz składu chemicznego atmosfery [29].



Rysunek 1.5: Rdzeń lodowy z Zachodniej Antarktyki. Ciemny pas to warstwa popiołu wulkanicznego, który osiadł na pokrywie lodowej około 21 000 lat temu. fot. Heidi Roop, NSF [29].

W lodowcach oraz pokrywach lodowych wierce się otwory o kształcie cylindrycznym, są to tak zwane rdzenie lodowe. Można nadać im miano zamrożonych "kapsuł czasu", dzięki którym możemy zrekonstruować klimat jaki panował na Ziemi. Rdzenie lodowe dzielą się na warstwy, które to odpowiadają za lata i pory roku, gdzie najstarszy lód znajduje się na dole rdzenia a najmłodszy na górze. Odzyskany lód z głębi lodowców pomaga określić skład i zachowanie atmosfery w przeszłości, klimat, ilość opadów śniegu a także rozmiar pokryw lodowych i lodowców. Informacje te pozwalają określić w jaki sposób i dlaczego klimat na Ziemi zmieniał się w przeszłości, oraz jesteśmy w stanie udoskonalić prognozy zmian klimatu w przyszłości. National Science Foundation Ice Core Facility (NSF-ICF) to ośrodek, w którym przechowywane i badane są rdzenie lodowe z całego świata. Zapewnia on naukowcom możliwość przeprowadzania badań i pomiarów rdzeni lodowych, które przechowywane są tak aby w przyszłości mogły posłużyć też do badań. Ośrodek ten przechowuje obecnie ponad 17 000 metrów rdzeni lodowych zebranych z różnych lokalizacji na między innymi na Antarktydzie, Grenlandii czy w Ameryce Północnej [29].



Rysunek 1.6: Ośrodek NSF-ICF. Każda srebrna rura na tych półkach zawiera 1-metrowy odcinek rdzenia lodowego. fot. Peter Rejcek, NSFT [29]

### 1.3.2 Wiercenie rdzeni lodowych

Na wszystkich kontynentach Ziemi są wykonywane odwierty rdzeni lodowych. Jednak największa ilość pobranego lodu pochodzi z Antarktydy i Grenlandii. Rdzenie te sięgają nawet 3 kilometrów. Najlepiej zachowane i szczegółowe zapisy klimatyczne pochodzą z zimnych wewnętrznych rejonów lądolodów polarnych. Jest tak, gdyż w tym regionie materiał nie uległ częściowemu stopieniu dzięki czemu nie zakłóca to zapisów zanieczyszczeń i uwieczonych gazów. Najstarsze nieprzerwane zapisy pobrane z rdzeni lodowych sięgają 800 000 lat na Antarktydzie oraz 130 000 lat na Grenlandii [29]. Rdzenie lodowe wiercone są za pomocą wiertel mechanicznych lub termicznych, wycinają one pierścień wokół centralnego, pionowego rdzenia. Wiertło mechaniczne to obracająca się rura lub bęben wiertniczy z nożami umiejscowionymi na głowicy. Przy obrocie noże wycinają okrąg wokół lodu, który ma zostać wydrążony a pozostałości są transportowane do komory na wióry. Wiertło obracane jest siłą fizyczną lub elektromechanicznym napędem silnikowym. Stosowane są też wiertła termiczne, wykorzystują element grzejny do topienia pierścienia otaczającego rdzeń lodowy a woda jest odprowadzana do zbiornika w wiertle [29].



Rysunek 1.7: Mechaniczna głowica wiertła z frezami używanymi do wycinania pierścienia wokół rdzenia lodu. Fot. Steven Profaizer, NSF [29].



Rysunek 1.8: Głowica wiertła termicznego bez frezów. Wiertła termiczne wykorzystują element grzejny do topienia pierścienia wokół lodu. Fot. Tony Wendricks, Univ. Wisconsin [29].

Na tym kończymy przegląd zagadnień z geofizyki. W następnej sekcji zajmiemy się zagadnieniami związanymi z analizą danych.

## 1.4 Zastosowanie języka Python w analizie danych

### 1.4.1 Analityka danych

„Data science”, a więc tzw. analityka danych (Danologia) to interdyscyplinarna dziedzina badań, która znajduje zastosowanie zarówno w świecie nauki jak i biznesu. Łączy ona metody matematyczne, statystykę oraz informatykę w celu wydobywania konkretnych wartości z zespołu danych. Wykorzystuje popularne narzędzia takie jak sieci neuronowe, Big data, Machine Learning oraz wiele różnych języków programowania. Dyscyplina ta jest stosunkowo młoda. Obejmuje zagadnienia takie jak generowanie i przetwarzanie danych, analizę matematyczną i biznesową oraz wizualizację [30]. Jej początki datuje się na rok 2002, kiedy to Magazyn Data Science Journal zadebiutował jako publikacja Międzynarodowej Rady Nauki (ICSU): Komitetu Danych dla Nauki i Techniki (CODATA). Przed 2008 roku zaczęto z kolei posługiwać się określeniem „analityka danych”, a dyscyplina zaczęła dynamicznie się rozwijać. Najczęściej spotykać się można z Data Science w sektorze medycznym, badaniach akademickich, bankach oraz przy tworzeniu najnowszych technologii [31]. Jednym z przykładów wykorzystania analityki danych w medycynie jest stworzenie mechanizmów do diagnozowania i rozpoznawania złośliwych zmian skórnych u człowieka.

### 1.4.2 Data Science - Języki Programowania

Na świecie istnieje wiele języków programowania, co roku powstają nowe a najpopularniejsze są na bieżąco aktualizowane. Istnieją języki, które sprawdzają się najbardziej w pisaniu oprogramowania inne w grach komputerowych, są też takie które dysponują narzędziami do analizy dużych zespołów danych. W Data Science dominującymi językami programowania są Python i R, jednak nie są to jedyne możliwości. Najpopularniejszymi językami w analityce danych są:

- Python to język o otwartym kodzie (Open Source), nazywany jest jednym z najłatwiejszych języków programowania, bardzo popularny wśród programistów Data Science. Największą jego zaletą jest to że z łatwością można połączyć go z algorytmami napisanymi w innych językach, które to są bardziej skomplikowane takie jak C. Najczęściej stosowany w takich dziedzinach, jak finanse, media, badania naukowe, elektronika, procesy produkcji czy technika [32].
- R jest językiem o otwartym kodzie (Open Source) i jest to środowisko dedykowane do obliczeń i analiz statystycznych a także do tworzenia wizualizacji danych. Jego zaletą jest możliwość korzystania z modeli statystycznych, wielu analityków tworzy w nim swoje aplikacje. Największą popularność zdobył w sektorze medycznym. W porównaniu do Pythona nie jest to język ogólnego przeznaczenia [33].
- SQL to kolejny również popularny język wykorzystywany w Data Science. Używany głównie do edytowania i tworzenia zapytań dla relacyjnej bazy danych i zawartych w niej danych informacji. Sprawdza się przy zarządzaniu dużymi bazami danych, dzięki szybszemu procesowaniu znacząco skraca czas wysyłanych zapytań online. Wadą tego języka jest ograniczenie do wykonywania podstawowych działań matematycznych [34].
- Julia to dynamiczny język programowania wysokiego poziomu. Bardzo dobrze sprawdza się przy analizie numerycznej oraz naukach obliczeniowych oraz radzi sobie z projektami zawierającymi dużą ilość zestawów danych. Jego wadą jest obecnie zbyt mała liczba bibliotek [35].



- Scala to język ogólnego przeznaczenia, dość dobrze znany wśród programistów, oparty jest na otwartym kodzie źródłowym. Wykorzystuje wirtualną maszynę Javy dzięki czemu może również współdziałać z tym językiem [36].
- Matlab jest to system komercyjny, stabilny i szybki, bardzo popularny wśród naukowców, który umożliwia pracę nad algorytmami do analizy numerycznej, cyfrowej oraz modelowania matematycznego. Nie jest językiem ogólnego przeznaczenia [37].

Wraz z rozwijaniem się usług internetowych oraz cyfryzacją wytwarzane są ogromne pokłady danych, rośnie zapotrzebowanie na narzędzia i metody do ich przetwarzania i rozwiązywania problemów wcześniej nie osiągalnych. Dzięki temu interdyscyplinarna dziedzina jaką jest Data Science zyskuje coraz większą popularność [38].

### 1.4.3 Python

Jednym z najbardziej popularnych języków analizy danych jest Python. Stanowi on swiste narzędzie ułatwiające poruszanie się po bazach danych w celu sprawnego wyszukiwania określonych wartości. Używanie go pozwala na przeprowadzanie różnych eksperymentów i obliczeń. Umożliwia on także korzystanie z kodów w innych językach takich jak Java, Fortran, Julia, R czy C oraz pozwala na podnoszenie wydajności tychże skryptów. Język ten działa sprawnie na wielu systemach operacyjnych. Bez zarzutów można z niego korzystać więc na systemie Windows, IOS, czy Linux. Również mniej znane systemy operacyjne bez problemu poradzą sobie z przetworzeniem kodu tego programu, gdyż mimo jego stosunkowo dużej objętości działa on dynamicznie przy zastosowaniu odpowiedniego kompilatora. Charakteryzuje się wysoką wydajnością oraz sprawnym zarządzaniem pamięcią, przy obciążaniu jej w minimalnym stopniu. Duży zbiór danych oraz duża ilość przekształceń na nim dokonywanych nie stanowi żadnego problemu, gdyż stworzony on jest z myślą o błyskawicznym odzyskiwaniu danych z pamięci również w odniesieniu do poprzednich kroków przetwarzania. Główną zaletą tego języka jest jednak ogromna liczba bibliotek, w które jest on wyposażony oraz prostota w jego obsłudze. Do analizy danych poprzez Python potrzebne jest pobranie pakietów, które umożliwiają rozszerzenie podstawowych możliwości tegoż języka popularnymi i często wykorzystywanymi są:

- SciPy, który posiada ogromną bazę precyzyjnych algorytmów i matematycznych rozwiązań. Służy on do optymalizacji i segregacji danych, a także do algebry liniowej i szybkiej transformaty Fouriera wykorzystywanej w uczeniu maszynowym. Aby z niego skorzystać, należy zainstalować pakiet za pomocą komendy *pip install scipy* [39].
- Matplotlib to biblioteką tablic dzięki której można tworzyć wykresy edytować je ale także przedstawiać w interaktywny sposób wartości. Do jego instalacji potrzebne jest zastosowanie polecenia *pip install matplotlib* [40].
- NumPy umożliwiający edycję i tworzenie wielowymiarowych tablic. Pakiet ten oferuje szeroki zakres matematycznych operacji i rozwiązań na tablicach, z których można skorzystać poprzez jego instalację za pomocą komendy: *pip install numpy* [41].
- Pandas umożliwia on zarządzanie bardzo dużymi tablicami złożonymi z danych różnych typów a także pozwala na tworzenie szeregów czasowych czyli ciągu uporządkowanych informacji w czasie. Instalacja pakietu odbywa się po wpisaniu komendy *pip install pandas* [42].

Powłoka Jupyter umożliwia korzystanie z wielu języków programowania, co czyni ją niezwykle przydatną i wielofunkcyjną. Jej funkcje sprawdzają się przy obliczeniach równoległych [43].

#### 1.4.4 Uczucie maszynowe

Uczenie maszynowe (ang. Machine learning) jest obecnie jednym z najważniejszych działów w dziedzinie informatyki zajmującymi się poszukiwaniem innowacyjnych rozwiązań wykorzystania danych. Polega ono na technologicznym zapamiętywaniu pewnych zachowań, których maszyna wcześniej „doświadczyła”. To właśnie dzięki temu możemy analizować dane a także znaleźć możliwe rozwiązania, które sprawdziły się w innych sytuacjach, znanych już w przeszłości maszynie [44]. Odpowiednie modele uczenia maszynowego, mają zdolność do ciągłego przewidywania zmian. Ponieważ dane są ciągle dodawane, modele te zapewniają, że rozwiązanie jest stale aktualizowane. Jeśli użyje się odpowiednich i stale zmieniających się danych, ma się możliwość przewidzenia następujących zmian w przyszłości. Uczenie maszynowe jest formą SI (Sztuczna Intelgencja, ang. Artificial Intelligence – AI), która umożliwia systemowi uczenie się z danych, a nie poprzez wyraźne programowanie. Jednakże, nauka maszynowa nie jest prostym procesem ponieważ maszynę należy trenować, aby unikać błędów i sytuacji, które dałoby się przewidzieć wcześniej. Uczenie to wykorzystuje różne algorytmy, które iteracyjnie uczą się na podstawie danych, aby poprawić, opisać i przewidzieć wyniki. Rozróżniamy trzy typy uczenia maszynowego:

- uczenie nadzorowane (ang. supervised learning) polega na tym że program otrzymuje oznaczone dane wejściowe i oczekiwane wyniki. Systemowi podaje się konkretnie, czego ma szukać, dlatego model jest szkolony, dopóki nie wykryje podstawowych wzorców i relacji, umożliwiając mu uzyskanie dobrych wyników, gdy zostaną przedstawione nigdy wcześniej nie widziane dane. Uczenie nadzorowane jest dobre w rozwiązywaniu problemów związanych z klasyfikacją i regresją, takich jak określanie kategorii, do której należy artykuł lub przewidywanie wielkości sprzedaży w określonym przyszłym terminie. W uczeniu nadzorowanym celem jest nadanie sensu danym konkretnym pomiarom. W przeciwieństwie do uczenia nadzorowanego jest to metoda uczenia się bez nadzoru, która stara się nadać sens danym samym w sobie. Nie ma żadnych zewnętrznych pomiarów ani wytycznych w uczeniu bez nadzoru: algorytm musi po prostu zrozumieć dane i wykryć wzorce lub podobieństwa [45].
- uczenie nienadzorowane (ang. unsupervised learning) w tym przypadku znana jest poprawna odpowiedź jeszcze przed uruchomieniem trenowania danego modelu, natomiast w uczeniu przez wzmacnianie korzystamy z regulatora do definiowania wartości nagród dla poszczególnych działań. Z kolei korzystając z technik uczenia nienadzorowanego, działamy na nieoznakowanych danych lub takich o nieznannej strukturze. Dzięki modelom stworzonym dla uczenia nienadzorowanego potrafimy poznać strukturę przetwarzanych danych oraz uzyskać użyteczne informacje bez stosowania funkcji nagrody lub znanej zmiennej wyjściowej [45].

- uczenie przez wzmacnianie (ang. reinforcement learning) jego celem jest utworzenie systemu agenta lub regulatora, który na podstawie interakcji ze środowiskiem potrafi poprawiać własną skuteczność. Model ten można uznać za powiązany z uczeniem nadzorowanym ponieważ pobierane informacje o bieżącym stanie środowiska często zawierają także tzw. sygnał nagrody. Istotną jest tu wartość skuteczności naszego pomiaru działania poprzez funkcję nagrody. Regulator oddziałuje ze środowiskiem dzięki temu może wykorzystywać uczenie przez wzmacnianie do trenowania wielu działań które dążą do maksymalizowania nagrody metodą planowania lub po prostu prób i błędów [45].

Jednym z najprostszych algorytmów uczenia nadzorowanego jest regresja liniowa. Prosta konstrukcja jest jej ogromną zaletą która zapewnia pełną interpretowalność modelu. Jej celem jest modelowanie związku między dwiema zmiennymi dzięki dopasowaniu równania liniowego do obserwowanych danych. Pierwsza zmienna jest zmienną objaśniającą, a druga zmienną objaśnianą. Współczynnik korelacji jest istotną liczbową miarą liniowego związku między dwiema zmiennymi oraz wartością między  $-1$  a  $1$  i ukazują siłę powiązania liniowego obserwowanych danych dla tych zmiennych. Model regresji liniowej ma równanie w postaci:

$$Y = a + bX, \tag{1.1}$$

gdzie  $X$  to zmienna objaśniająca,  $Y$  jest zmienną objaśnianą. Współczynnik kierunkowy to  $b$ , punktem przecięcia z osią jest  $a$ , zaś  $b$  to nachylenie prostej (wartość  $y$ , gdy  $X = 0$ ).

Najpopularniejszą metodą dopasowania linii regresji jest metoda najmniejszych kwadratów. Oblicza ona najlepiej dopasowaną linię dla interesujących nas danych przy czym minimalizuje sumę kwadratów odchyleń pionowych od każdego punktu danych do linii. Punkt leżący dokładnie na dopasowanej linii ma odchylenie pionowe wynoszące  $0$ . Na początku odchylenia są podnoszone do kwadratu, następnie są sumowane dzięki czemu wartości dodatnie i ujemne nie zostają zniwelowane. Po wyznaczeniu linii regresji dla obserwowanych danych punkt, który leży daleko od linii a więc jego wartość rezydualna jest duża nazywany jest wartością odstającą. Punkty te mogą wskazywać na błędne dane lub ukazują źle dopasowaną linię regresji. Punkty leżące z dala od innych w kierunku poziomym, mogą mieć istotny wpływ na nachylenie linii regresji i jest to nazywane obserwacją wpływową [46].

## 1.5 Dyskretna Transformata Fouriera

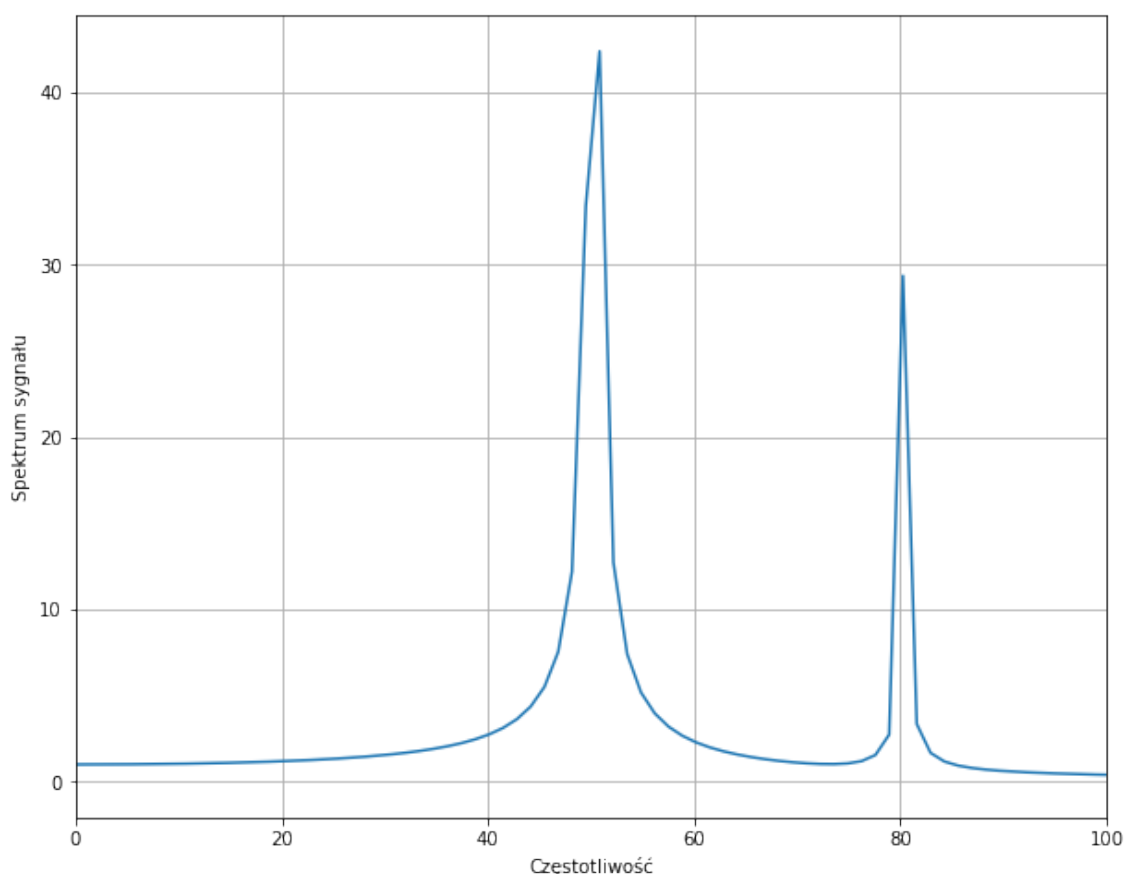
Dyskretne przekształcenie Fouriera (transformata, ang. Discrete Fourier Transform : DFT) jest bardzo wydajną i jedną z najpopularniejszych procedur spotykanych w obszarze cyfrowego przetwarzania sygnałów. DFT pozwala analizować, syntetyzować i badać sygnały w metodą niemożliwą do wykorzystania przy przetwarzaniu sygnałów analogowych. Dyskretne przekształcenie Fouriera jest procedurą matematyczną wykorzystywaną do wyznaczenia zawartości częstotliwościowej lub harmonicznej sygnału dyskretnego. DFT jest użyteczne w analizie dowolnych ciągów dyskretnych [47]. Ciągłe przekształcenie Fouriera  $X(f)$  ma równanie:

$$X(f) = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft} dt, \quad (1.2)$$

gdzie  $j = \sqrt{-1}$  natomiast  $x(t)$  to sygnał ciągły w dziedzinie czasu. Dyskretny ciąg  $X(m)$  w dziedzinie częstotliwości:

$$X(m) = \sum_{n=0}^{N-1} x(n)e^{-\frac{j2\pi nm}{N}}, \quad (1.3)$$

gdzie  $x(n)$  to dyskretny ciąg próbkowanych wartości w dziedzinie czasu ciągłej zmiennej  $x(t)$ ,  $e$  to podstawa logarytmu naturalnego. Przykładowy wykres transformaty Fouriera przedstawiono na Rys. 1.9.



Rysunek 1.9: Wykres zależności transformaty Fouriera sygnału  $\sin(100\pi x) + 0.5\sin(160\pi x)$  od częstotliwości.

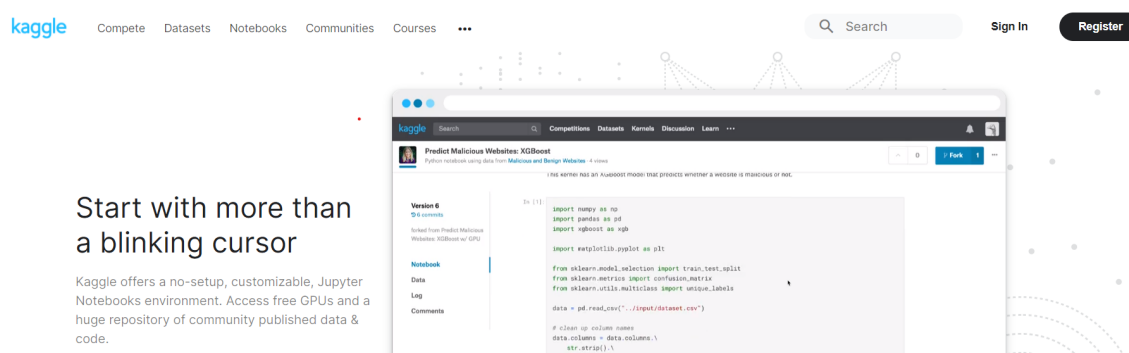
## Rozdział 2

### Cześć praktyczna

## 2.1 Analiza temperatury w latach 1850-2015

### 2.1.1 Opis zbioru danych

Pierwsze z analizowanych danych dotyczące temperatury na Ziemi pobrane zostały z platformy Kaggle [48]. Jest to strona internetowa stworzona z myślą o naukowcach zajmujących się pracą z danymi oraz dla osób praktykujących uczenie maszynowe. Kaggle pozwala wyszukiwać oraz publikować ogromne zestawy danych, a także pracować nad nimi i budować modele przy współpracy z innymi naukowcami. Baza strony posiada aż 536 000 aktywnych członków ze 194 krajów oraz otrzymuje około 150 000 zgłoszeń miesięcznie [48].



Rysunek 2.1: Strona internetowa Kaggle.

W skład repozytorium wchodziło pięć plików:

- *GlobalLandTemperaturesByCity.csv*
- *GlobalLandTemperaturesByCountry.csv*
- *GlobalLandTemperaturesByMajorCity.csv*
- *GlobalLandTemperaturesByState.csv*
- *GlobalTemperatures.csv*

z których wykorzystane w badaniu zostały pierwszy i ostatni z powyższej listy, przy czym w niniejszym rozdziale przedstawione zostały wyniki analizy danych z pliku *GlobalTemperatures.csv*. Dane z repozytorium zostały zebrane przez Berkeley Earth. Jest to amerykańska organizacja typu non-profit zajmująca się gromadzeniem i analizą danych związanych ze środowiskiem. Organizacja ta związana jest z Lawrence Berkeley National Laboratory<sup>1</sup>. Dane które analizowaliśmy dotyczą zmian temperatury (mierzonej w stopniach Celsjusza) na Ziemi w okresie od roku 1750 do roku 2015. Pomiar temperatury w okresie przed 1980 rokiem wykonywany był za pomocą termometrów rtęciowych, natomiast w latach późniejszych wykorzystywano termometry elektroniczne.

<sup>1</sup>Jest to multidyscyplinarny ośrodek badawczy założony przez fizyka Ernesta Lawrence w 1931 roku na Uniwersytecie Kalifornijskim. Prowadzone są tam badania z zakresu nauk podstawowych jak również inżynierskich czy biologicznych.

Oryginalne dane pochodzące z pliku *GlobalTemperatures.csv* zawierały 3192 rekordy i 9 kolumn. W każdym rekordzie znajdowały się wyniki pomiarów cech zawartych w kolumnach. Cechy te to:

- data wykonania pomiaru
- średnia temperatura lądu
- niepewność pomiaru średniej temperatury lądu
- maksymalna temperatura lądu
- niepewność pomiaru maksymalnej temperatury lądu
- minimalna temperatura lądu
- niepewność pomiaru minimalnej temperatury lądu
- średnia temperatura lądu i oceanów
- niepewność pomiaru średniej temperatury lądu i oceanów

Przed przystąpieniem do właściwej analizy wykonana została procedura wykrywania braków w danych, w wyniku której otrzymane zostały informacje iż w kolumnie opisującej datę wykonania pomiaru nie występują luki. W kolumnach opisujących średnią temperaturę lądu oraz niepewność pomiaru średniej temperatury lądu występuje 12 luk, a w pozostałych kolumnach 1200 pozycji (Rys. 2.3). Ze względu na charakter przetwarzanych przez nas danych do dalszej analizy użyto jedynie tych rekordów, w których nie zaobserwowano braków. W wyniku czego w dalszym badaniu wykorzystane zostały 1992 rekordy, co miało przełożenie na ograniczenie zakresu czasowego danych do lat 1850-2015. Po wykonaniu powyżej opisanej preparacji danych dokonano ich agregacji w ten sposób, że każdemu z lat w przedziale od 1850 do 2015 roku przyporządkowano średnią temperaturę z danego roku. Procedura ta została zastosowana dla każdej z cech występujących w badanych danych. W tej części pracy wykorzystane zostały narzędzia zawarte w pakietach dostępnych dla języka Python takich jak: *Numpy*, *Pandas*, *Seaborn*, *Matplotlib* oraz *Scikit-learn*<sup>2</sup>. Poniżej przedstawione zostały wyniki przeprowadzonej analizy, głównie w postaci wykresów szeregów czasowych danych cech.

---

<sup>2</sup>Opis zastosowanych tutaj pakietów można znaleźć w rozdziale trzecim niniejszej pracy.

## 2.1.2 Wizualizacja danych

	Średnia_temperatura_lądu	Niepewność_pomiaru_średniej_temperatury_lądu	Max_temperatura_lądu	Niepewność_pomiaru_max_temperatury_lądu
count	1992.00	1992.00	1992.00	1992.00
mean	8.57	0.28	14.35	0.48
std	4.26	0.22	4.31	0.58
min	0.40	0.03	5.90	0.04
25%	4.43	0.10	10.21	0.14
50%	8.85	0.23	14.76	0.25
75%	12.86	0.35	18.45	0.54
max	15.48	1.49	21.32	4.37

Rysunek 2.2: Zrzut ekranu z środowiska Jupyter Notebook. Dane dla temperatury.

	Min_temperatura_lądu	Niepewność_pomiaru_min_temperatury_lądu	Średnia_temperatura_lądu_i_oceanów	Niepewność_pomiaru_średniej_temperatury_lądu_i_oceanów
count	1992.00	1992.00	1992.00	1992.00
mean	2.74	0.43	15.21	0.13
std	4.16	0.45	1.27	0.07
min	-5.41	0.04	12.48	0.04
25%	-1.33	0.16	14.05	0.06
50%	2.95	0.28	15.25	0.12
75%	6.78	0.46	16.40	0.15
max	9.72	3.50	17.61	0.46

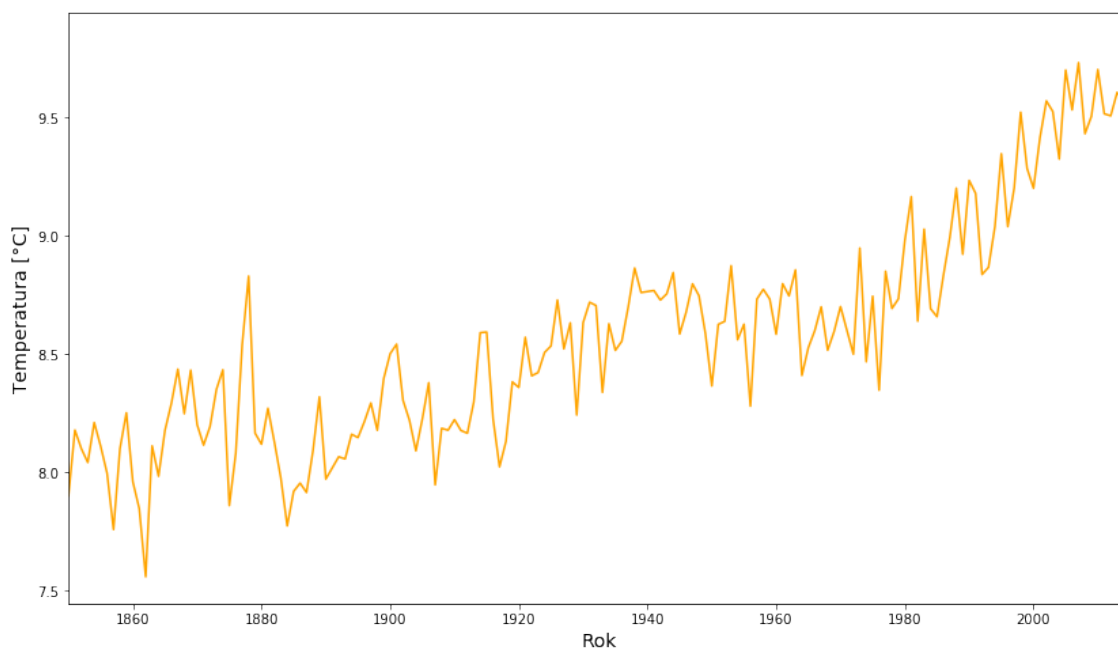
Rysunek 2.3: Zrzut ekranu z środowiska Jupyter Notebook. Dane dla temperatury.

Suma braków w kolumnach:

```
Data          0
Średnia_temperatura_lądu      12
Niepewność_pomiaru_średniej_temperatury_lądu      12
Max_temperatura_lądu          1200
Niepewność_pomiaru_max_temperatury_lądu          1200
Min_temperatura_lądu          1200
Niepewność_pomiaru_min_temperatury_lądu          1200
Średnia_temperatura_lądu_i_oceanów          1200
Niepewność_pomiaru_średniej_temperatury_lądu_i_oceanów          1200
dtype: int64
```

Rysunek 2.4: Zrzut ekranu z środowiska Jupyter Notebook. Suma braków w poszczególnych kolumnach.



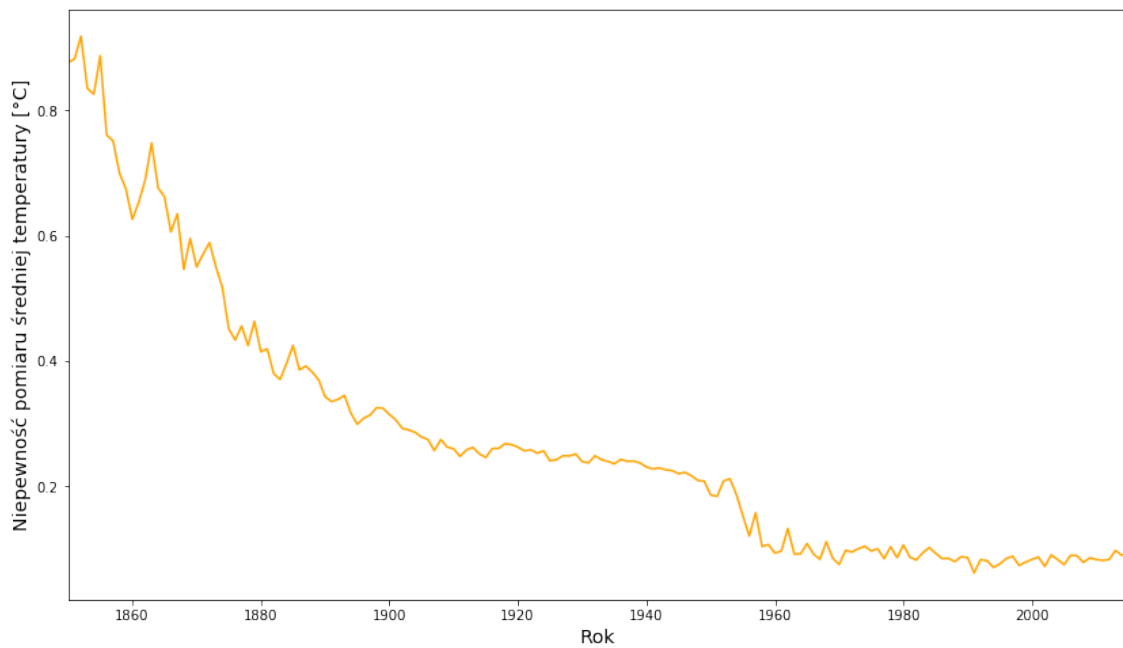


Rysunek 2.5: Średnia temperatura lądu w latach od 1850 do 2015.

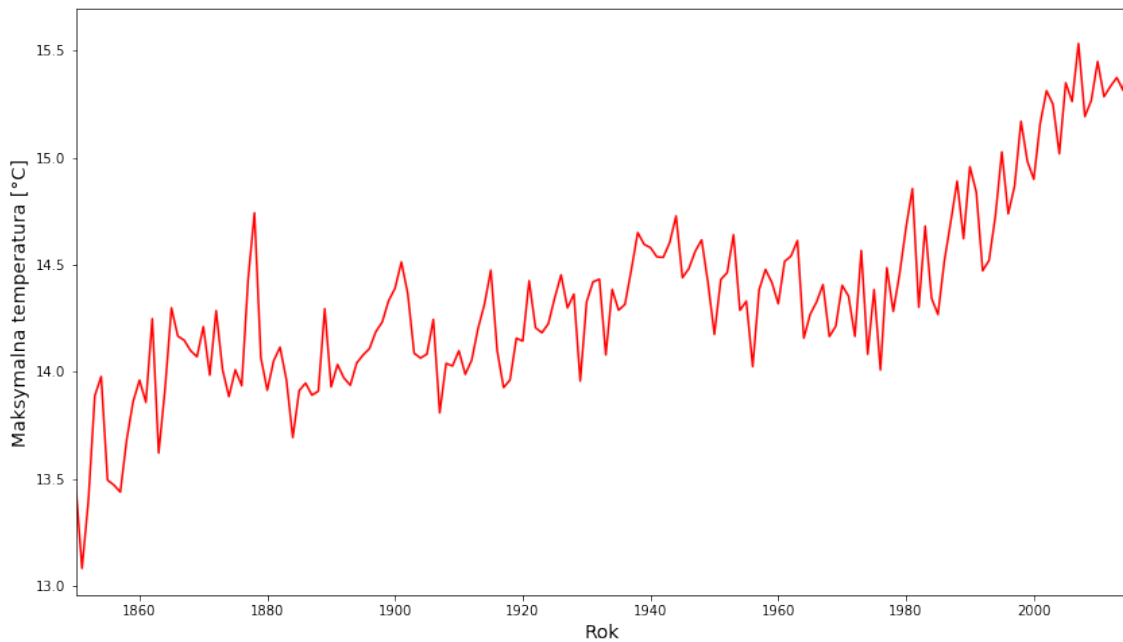
Powyższy wykres 2.5 przedstawia szereg czasowy opisujący średnią temperaturę lądu w danym roku w zakresie lat od 1850 do 2015. Z wykresu wywnioskować można istnienie tendencji rozwojowej która w przeważającym okresie obserwacji jest funkcją rosnącą. Obserwuje się gwałtowny wzrost temperatury po 1960 roku w porównaniu z latami wcześniejszymi. Efekt ten może być związany z industrializacją Azji. Wzbogacanie się krajów azjatyckich miało początek w latach sześćdziesiątych XX wieku i w fazie początkowej dotyczyło państw takich jak: Hongkong, Singapur, Tajwan oraz Korea Południowa. W późniejszym okresie do krajów tych dołączyły Malezja i Tajlandia a pod koniec lat siedemdziesiątych Indonezja<sup>3</sup> oraz Chiny. Rozwój przemysłu ciężkiego jak również lekkiego<sup>4</sup> doprowadził do ciągłego wzrostu zapotrzebowania na surowce takie jak węgiel czy ropa naftowa. Nadmierna eksploatacja zasobów naturalnych przełożyła się na wzrost emisji dwutlenku węgla jak również ogólną degenerację środowiska naturalnego, co może być jedną z przyczyn zwiększonej szybkości wzrostu temperatury po 1960 roku [49]. Poniżej, na Rys. 2.6 przedstawiona została zależność między niepewnością pomiaru średniej temperatury lądu a rokiem wykonania tegoż pomiaru. Wyraźnie widoczne zmniejszenie wartości niepewności pomiarowej w czasie związane jest ze zwiększeniem dokładności używanych przyrządów pomiarowych. Kolejny Rys. 2.7 przedstawia zależność maksymalnej temperatury lądu w funkcji czasu. Z wykresu tego możemy odczytać, że w analizowanym okresie czasu maksymalna temperatura zmieniała się w zakresie od  $5.9^{\circ}C$  do  $21.32^{\circ}C$ . Przez maksymalną temperaturę rozumiana jest tutaj globalna średnia z maksymalnych temperatur.

<sup>3</sup>Indonezja z punktu widzenia geograficznego jest częścią Oceanii natomiast często w opracowaniach ekonomicznych zaliczana jest do regionu przemysłowego Azji Wschodniej.

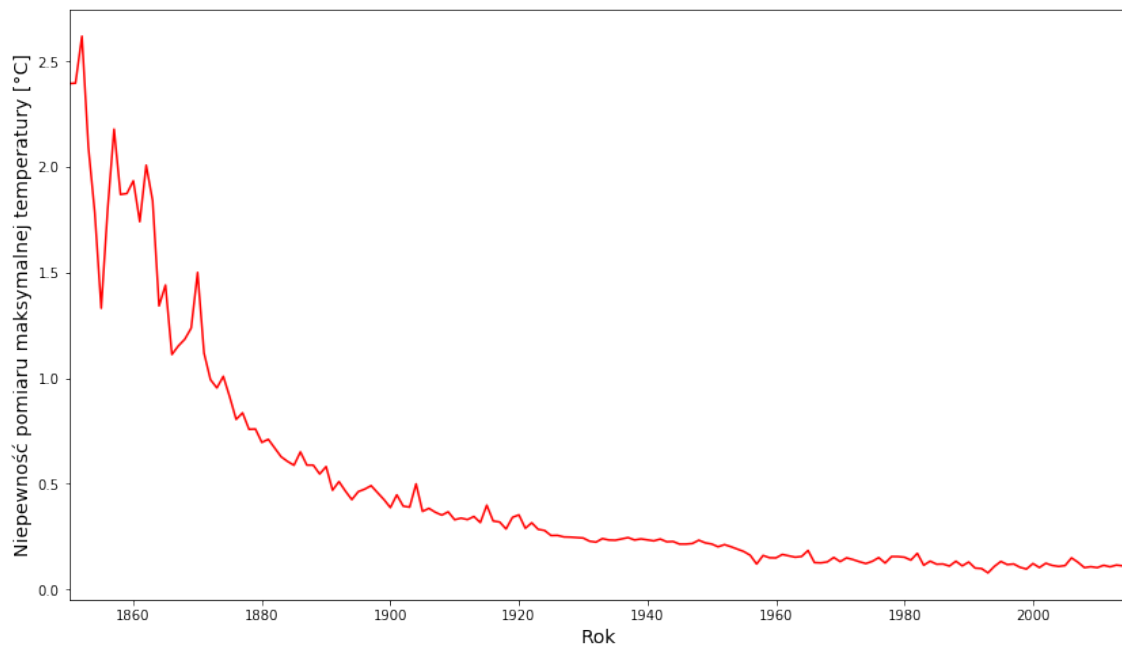
<sup>4</sup>Do rozwoju przemysłu w krajach azjatyckich przyczyniła się w dużej mierze relokacja zakładów przemysłowych z krajów wysoko rozwiniętych.



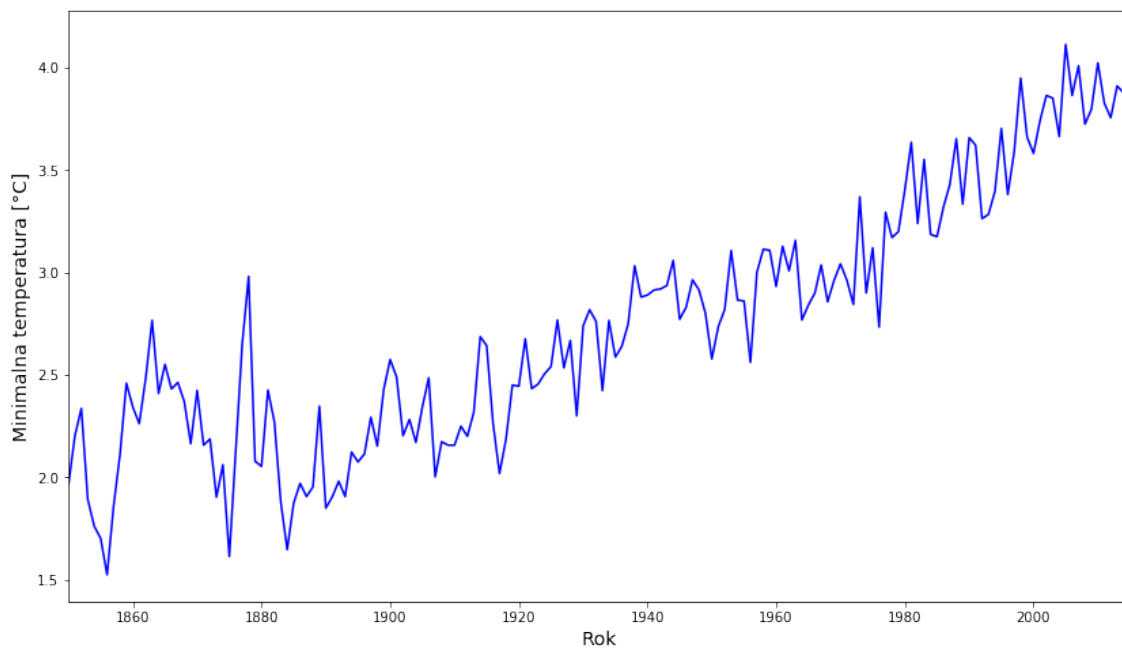
Rysunek 2.6: Niepewność pomiaru średniej temperatury lądu w latach 1850 do 2015.



Rysunek 2.7: Maksymalna temperatura lądu w latach 1850 do 2015.

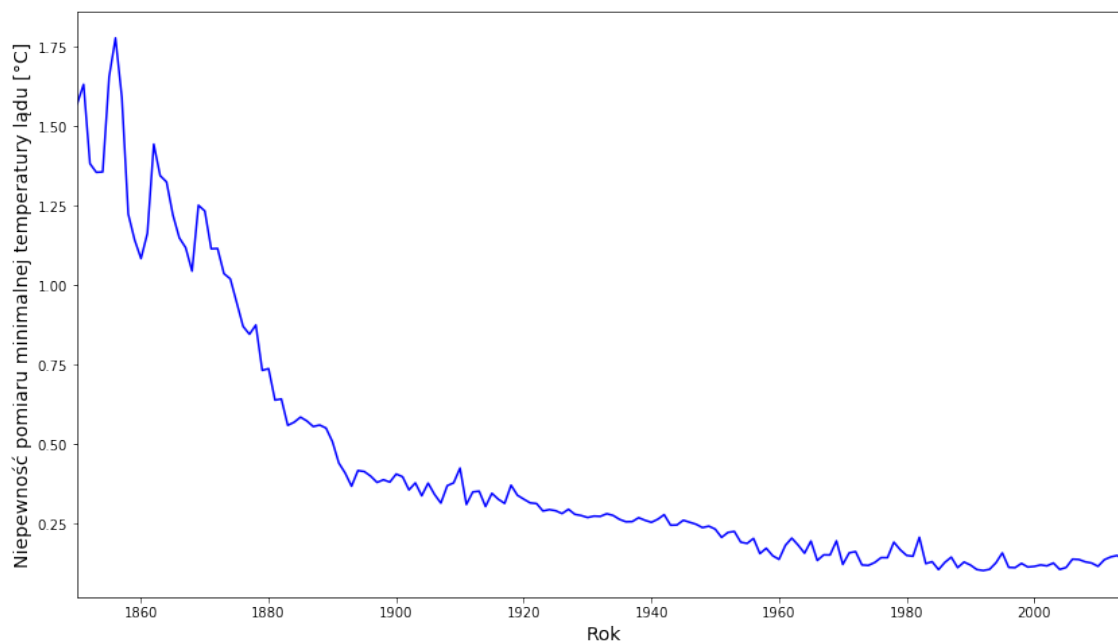


Rysunek 2.8: Niepewność pomiaru maksymalnej temperatury lądu w latach 1850 do 2015.



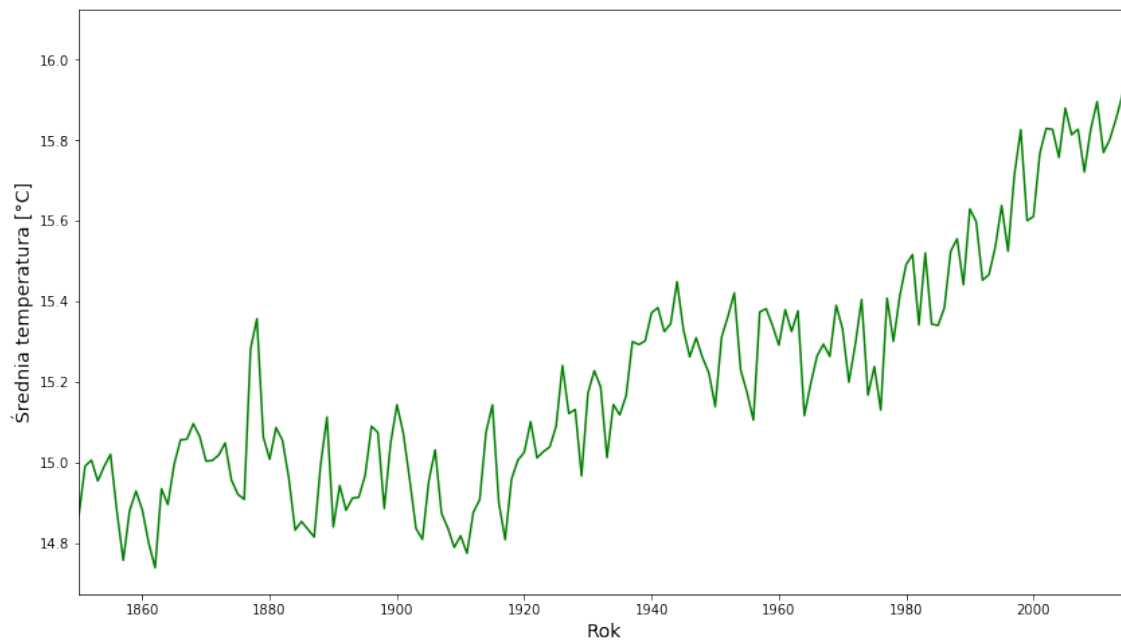
Rysunek 2.9: Minimalna temperatura lądu latach w 1850 do 2015.

Na Rys. 2.8 przedstawiona została zależność między niepewnością pomiaru maksymalnej temperatury lądu a rokiem wykonania tegoż pomiaru. Wyraźnie widoczne zmniejszenie wartości niepewności pomiarowej w czasie związane jest ze zwiększeniem dokładności używanych przyrządów pomiarowych. Kolejny Rys. 2.9 przedstawia zależność minimalnej temperatury lądu w funkcji czasu. Z danych wynika, że w analizowanym okresie czasu minimalna temperatura zmieniała się w zakresie od  $1.52^{\circ}\text{C}$  do  $-9.71^{\circ}\text{C}$ . Przez minimalną temperaturę rozumiana jest tutaj globalna średnia z minimalnych temperatur.

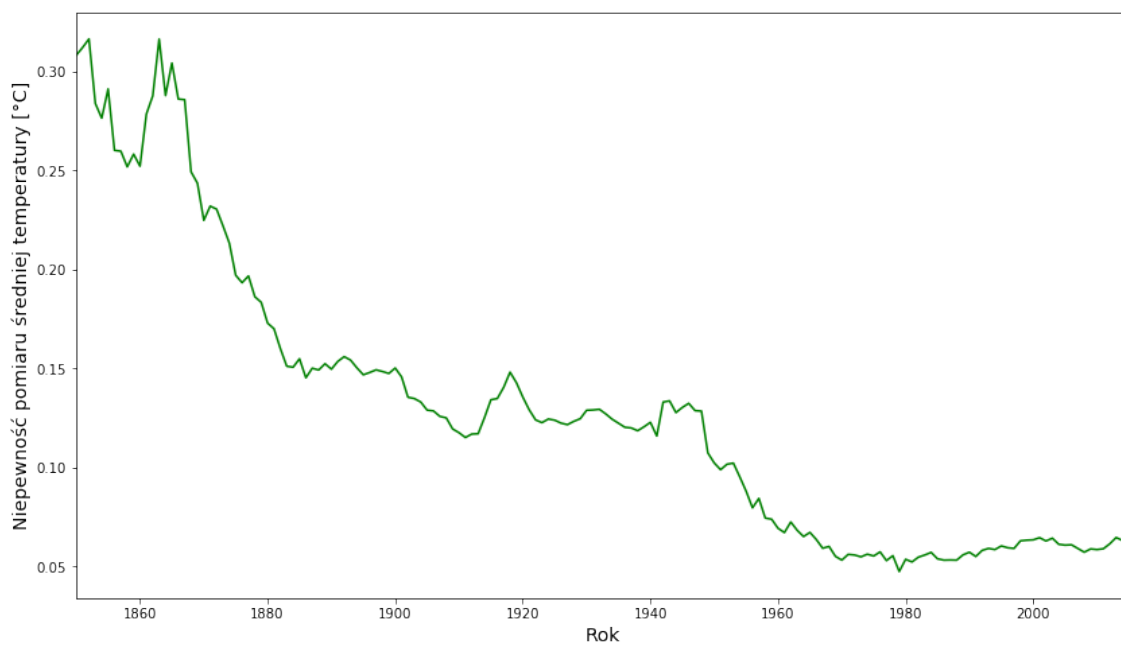


Rysunek 2.10: Niepewność pomiaru minimalnej temperatury lądu w latach 1850 do 2015.

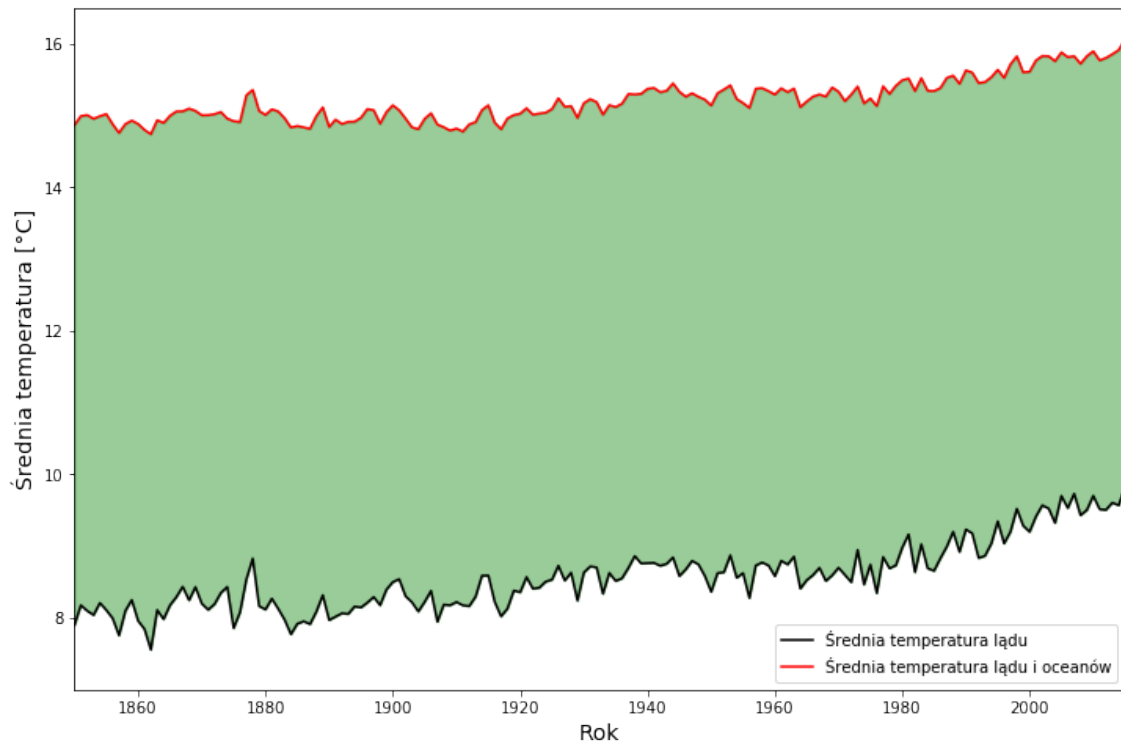
Na Rys. 2.10 przedstawiona została zależność między niepewnością pomiaru minimalnej temperatury lądu a rokiem wykonania tegoż pomiaru. Wyraźnie widoczne zmniejszenie wartości niepewności pomiarowej w czasie związane jest ze zwiększeniem dokładności używanych przyrządów pomiarowych. Na Rys. 2.11 oraz Rys. 2.11 zamieszczonych na kolejnej stronie przedstawione są odpowiednio wizualizację średniej temperatury lądu i oceanów oraz niepewności pomiaru tej temperatury w latach od 1850 do 2015. Z pierwszego wykresu wnioskować można, że podobnie jak w przypadku średniej temperatury lądu występuje tutaj tendencja rozwojowa, która w przeważającym okresie obserwacji jest funkcją rosnącą. Podobnie jak we wcześniejszych przypadkach wykres niepewności średniej temperatury lądu i oceanów ukazuje spadek wartości niepewności mierzonej wielkości, co może być wyjaśnione poprawą dokładności pomiarów wynikającą z zastosowania nowszych i dokładniejszych mierników.



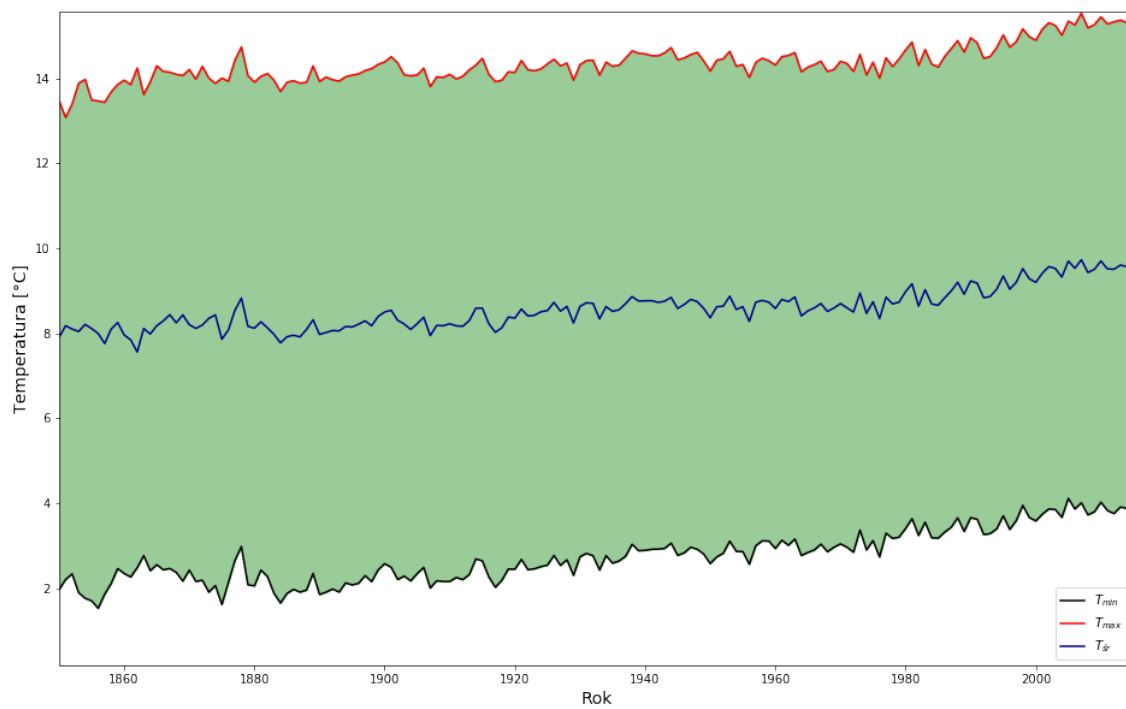
Rysunek 2.11: Średnia temperatura lądu i oceanów w latach 1850 do 2015.



Rysunek 2.12: Niepewność pomiaru średniej temperatury lądu i oceanów w latach 1850 do 2015.



Rysunek 2.13: Porównanie średniej temperatury lądu i oceanów.



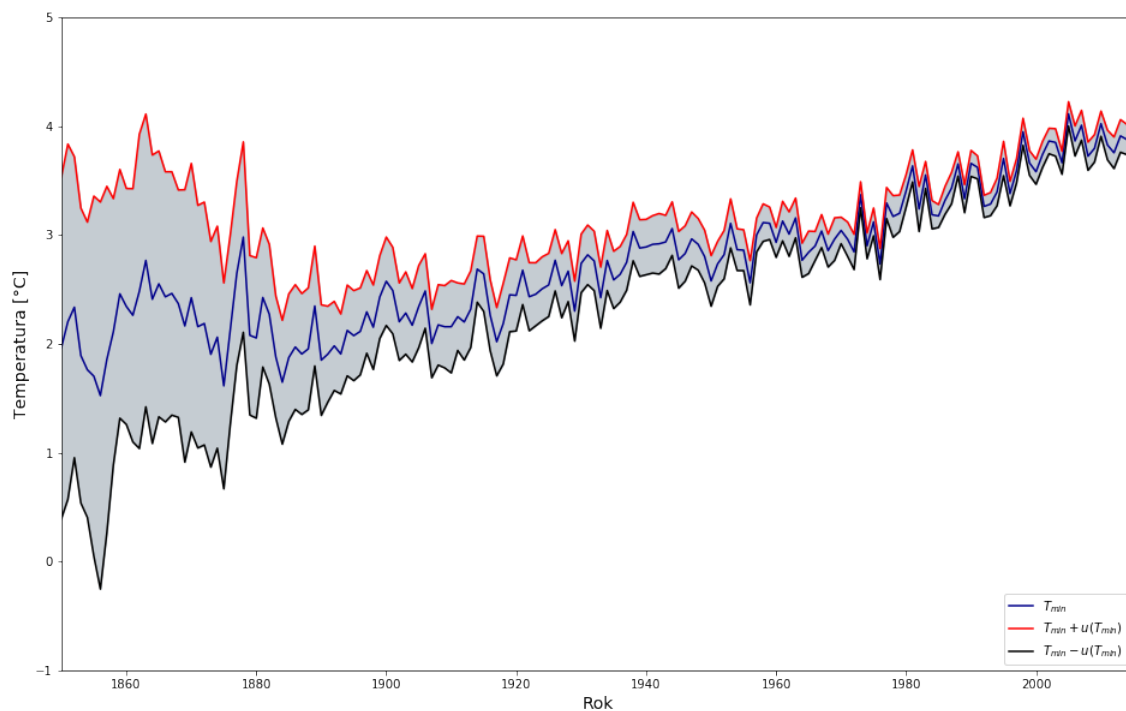
Rysunek 2.14: Wahania temperatury lądu.

Na Rys. 2.13 zamieszczonym na poprzedniej stronie obserwujemy, że w całym analizowanym okresie czasu średnia temperatura łądu i oceanów jest większa od średniej temperatury łądu o około  $6.6^{\circ}\text{C}$ . Wypełnieniem (zielony kolor) zaznaczona została różnica pomiędzy średnimi temperaturami. Na kolejnym Rys. 2.14 przedstawione zostało porównanie maksymalnej, minimalnej i średniej temperatury łądu w danym okresie. Wypełnieniem (zielony kolor) zaznaczono zakres zmienności. Krzywa na wykresie Rys. 2.15 opisuje różnicę temperatury maksymalnej i minimalnej i na jej podstawie można zaobserwować, że wahania się nie zmieniają, tylko średnia temperatura rośnie. Implikuje to zatem, że zmiany klimatu nie sprawiają, iż temperatura w zimie jest mniejsza, a w lecie większa przy odniesieniu względem średniej.

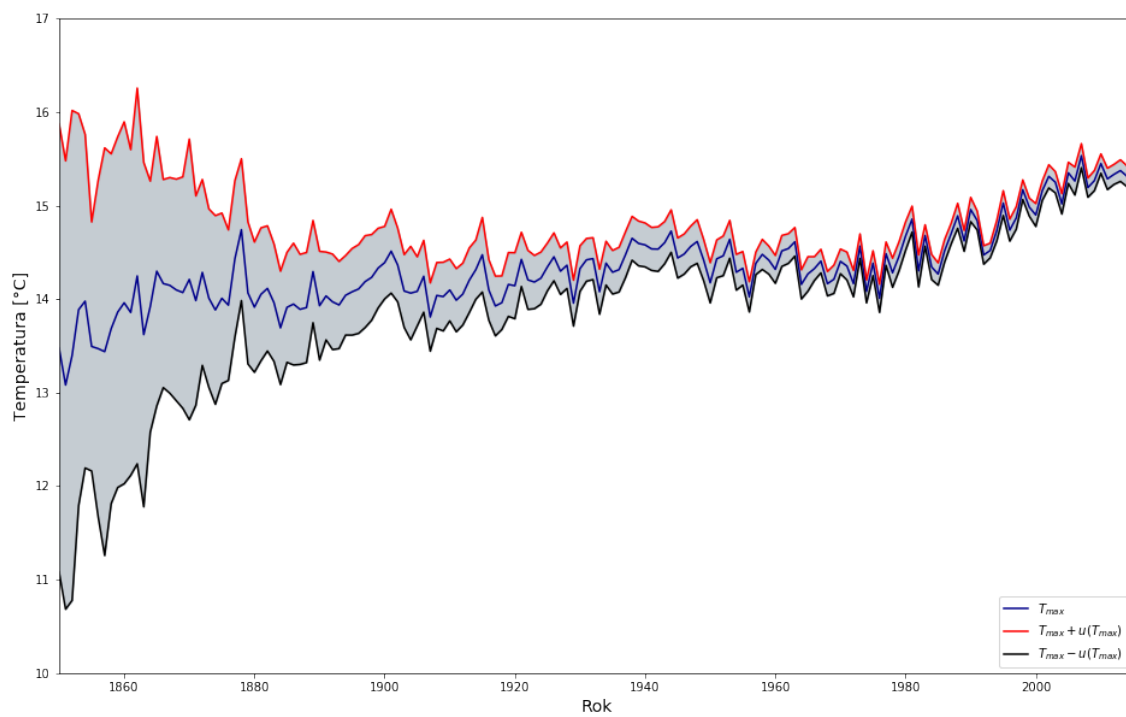


Rysunek 2.15: Różnica temperatury maksymalnej i minimalnej w latach 1850 do 2015.

Rys. 2.16, Rys. 2.17 i Rys. 2.18 zamieszczone na kolejnych stronach przedstawiają zależności odpowiednio temperatury: minimalnej łądu, maksymalnej łądu oraz średniej łądu wraz z błędami w funkcji czasu. Błędy te dla minimalnej i maksymalnej temperatury zostały określone jako wartości bezwzględne z niepewności pomiaru odpowiednio dla minimalnej i maksymalnej temperatury łądu. Błąd dla średniej temperatury to średnia arytmetyczna błędów dla temperatury minimalnej i maksymalnej. W każdym przypadku można stwierdzić, że wraz z upływem czasu błędy stają się coraz mniejsze.

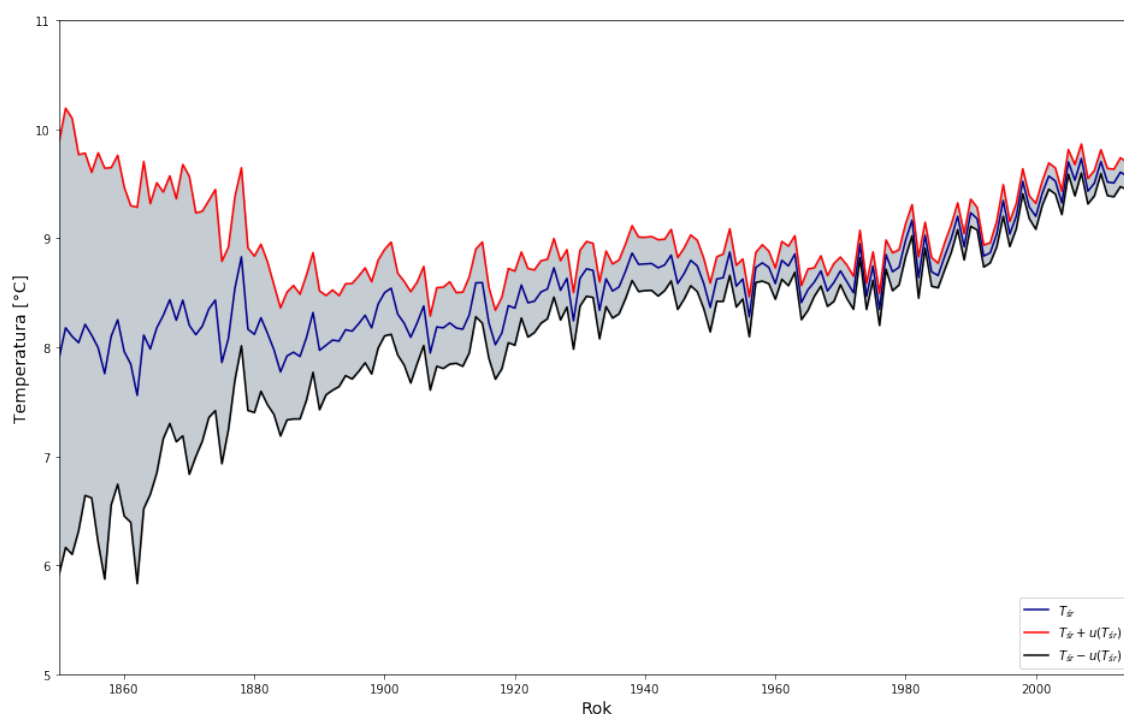


Rysunek 2.16: Temperatura minimalna wraz z błędem.



Rysunek 2.17: Temperatura maksymalna wraz z błędem.

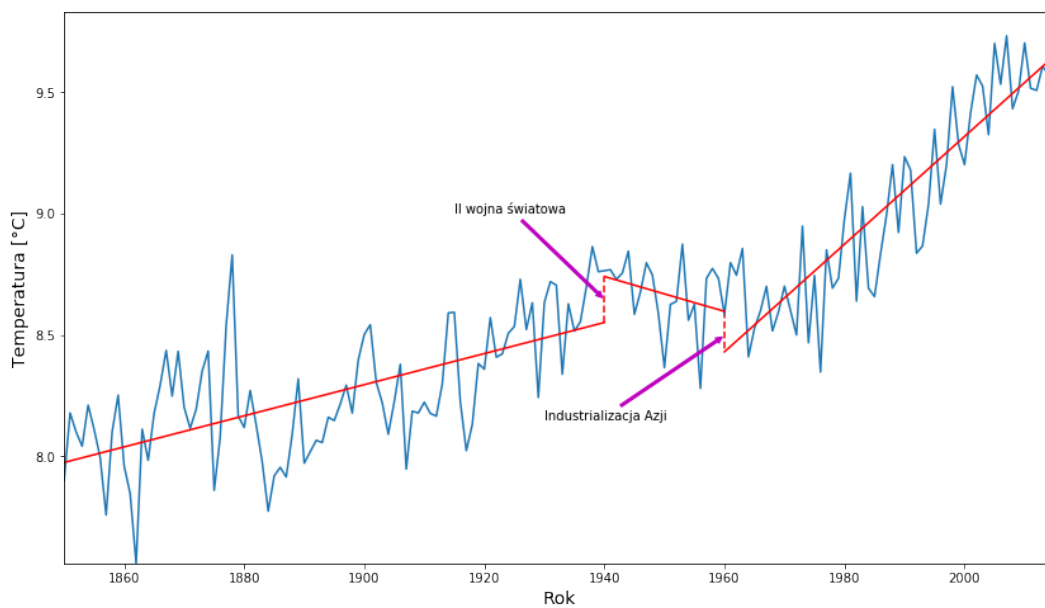




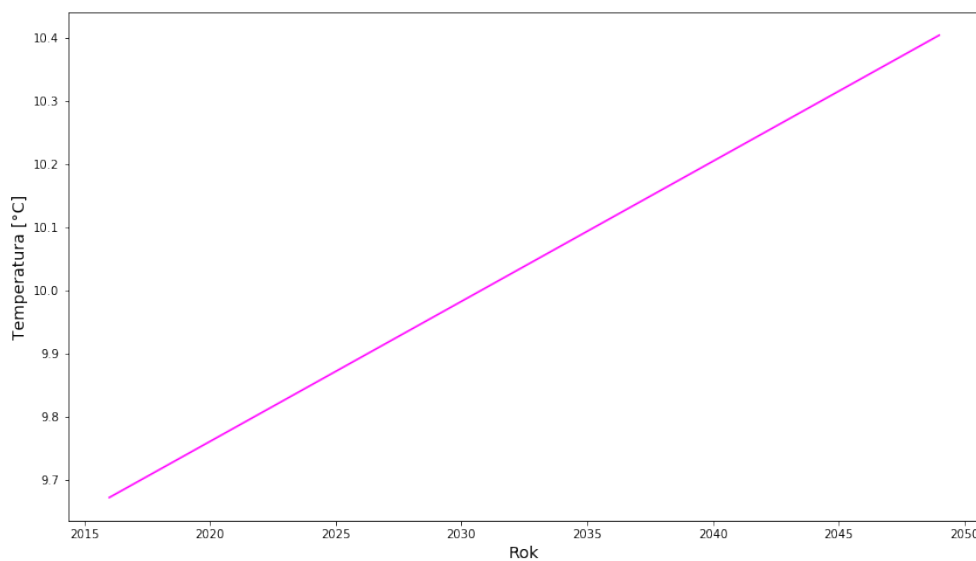
Rysunek 2.18: Średnia temperatura wraz z błędem.

### 2.1.3 Dopasowanie metodą regresji liniowej

Na Rys. 2.19 przedstawione zostały dopasowania trendu dla szeregu czasowego średniej temperatury lądu. Badany okres został podzielony na trzy części. W pierwszym okresie od 1850 roku do 1940 roku dopasowana linia trendu wskazuje na wzrost temperatury. W literaturze [50] rok 1850 kończy małą epokę lodowcową i rozpoczyna okres współczesnego ocieplenia. Drugi okres to lata od 1940 do 1960. W tym przedziale czasowym linia trendu wskazuje na spadek średniej temperatury lądu czego przyczyny można upatrywać w zniszczeniach infrastruktury przemysłowej, które nastąpiły w wyniku II wojny światowej. Ostatni wydzielony przez nas okres obejmuje przedział czasowy od roku 1960 do 2015. W okresie tym obserwujemy rosnącą funkcję trendu, która może być wyjaśniona odbudową i rozwojem światowego przemysłu oraz wzrostem populacji. Można zauważyć, że funkcja trendu otrzymana za pomocą regresji liniowej w okresie trzecim jest nachylona pod większym kątem niż ta która została otrzymana dla danych z pierwszego okresu. Różnica ta wynikać może z mniejszego uprzemysłowienia świata w pierwszym okresie w porównaniu do okresu trzeciego, co z kolei implikować może wpływ rozwoju cywilizacji na zmiany klimatu.



Rysunek 2.19: Średnia temperatura lądu wraz z dopasowaną funkcją trendu.



Rysunek 2.20: Średnia temperatura lądu - predykcja.

Rys. 2.20 ilustruje predykcję średniej temperatury lądu wykonaną na podstawie wyznaczonej wcześniej funkcji trendu z danych dla przedziału czasowego od 1960 do 2015 roku. Przewidywanie zostało wykonane do roku 2050 i jeśli wyznaczona tendencja utrzyma się to średnia temperatura lądu za trzydzieści lat osiągnie wartość około  $10.5^{\circ}\text{C}$ .

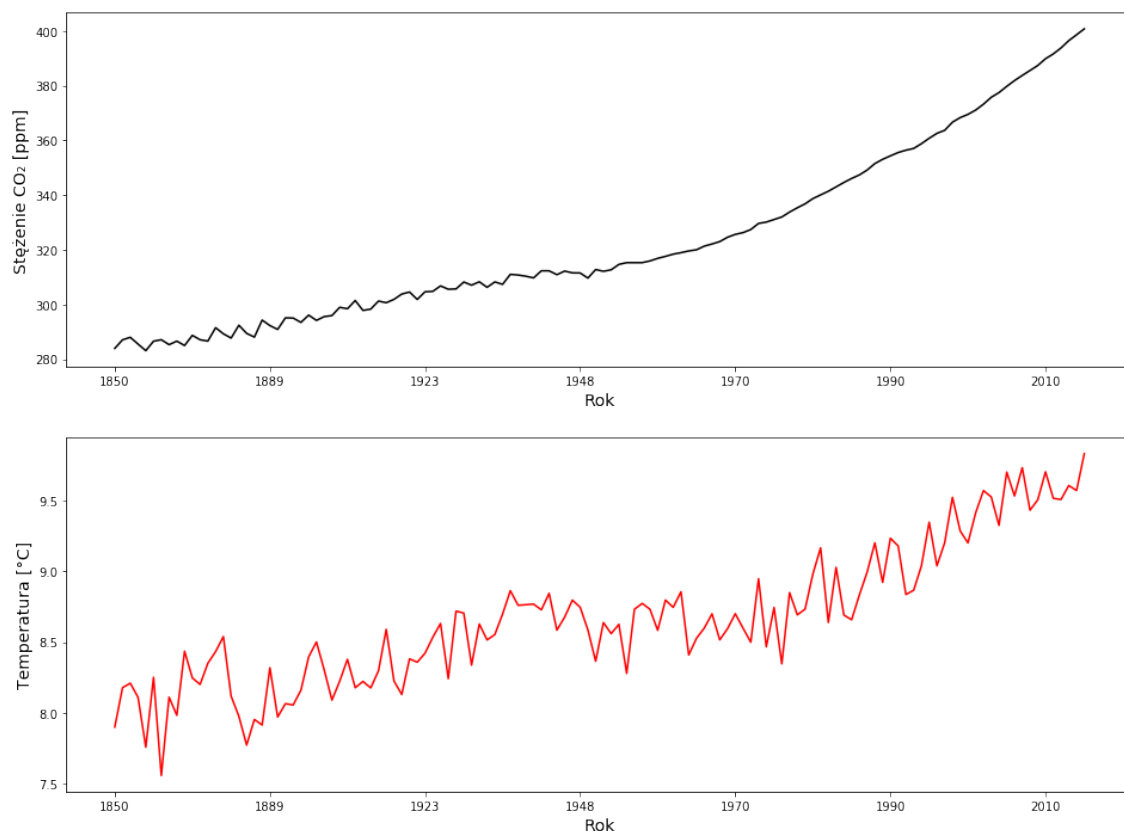
## 2.2 Analiza stężenia $CO_2$ w latach 1850-2015

Dane dotyczące stężenia dwutlenku węgla w atmosferze pobrane zostały ze strony Our World in Data [51]. OurWorldinData.org to strona powstała dzięki pracy naukowców z Uniwersytetu Oksfordzkiego, którzy są naukowymi redaktorami zawartości witryny a także organizacji non-profit Global Change Data Lab, która to publikuje, utrzymuje stronę internetową i narzędzia do obsługi danych, które umożliwiają nam pracę. Na Uniwersytecie w Oksfordzie prowadzony jest również program Oxford Martin dotyczący rozwoju globalnego. Witryna ta skupia się na gromadzeniu i udostępnieniu za darmo wszystkim danych dotyczących najważniejszych problemów ludzi i naszej planety takich jak: bieda, głód, wojny, choroby, zmiany klimatyczne.



Rysunek 2.21: Strona internetowa OurWordInData.

Oryginalne dane zawierały 1871 rekordów oraz cztery kolumny z cechami, które określały region pomiaru stężenia  $CO_2$  (w niniejszej pracy analizowanym regionem był cały świat), kod opisujący dany region, rok w którym stwierdzono występowanie danego stężenia  $CO_2$  oraz samo stężenie. Przed przystąpieniem do właściwej analizy wykonana została procedura wykrywania braków w danych, w wyniku której otrzymaliśmy informację że w analizowanych danych nie występują luki. Aby możliwe było zbadanie wpływu ilości  $CO_2$  w atmosferze na temperaturę dane dotyczące stężenia dwutlenku węgla zostały ograniczone do lat od 1850 do 2015. Poniżej przedstawione zostaną wyniki przeprowadzonej analizy. Na Rys. 2.22 zamieszczonym na następnej stronie zilustrowano zależność stężenia  $CO_2$  oraz średniej temperatury lądu w zależności od czasu dla okresu od 1850 do 2015 roku. Na obydwu wykresach zauważyć można podobny wzrost obydwu wielkości co może sugerować związek pomiędzy nimi.



Rysunek 2.22: Porównanie zależności stężenia  $CO_2$  i średniej temperatury lądu w latach od 1850 do 2015.

Na Rys. 2.23 widać zależność temperatury od stężenia  $CO_2$  w atmosferze w latach od 1850 do 2015. Na wykresie tym można zauważyć tendencję wzrostową temperatury oraz widoczna jest nieliniowa zależność w okolicy 280-320 ppm<sup>5</sup>, co wynika z faktu że dla takiego samego stężenia w różnych latach występowały różne temperatury. Kolejny Rys. 2.24 ilustruje zależność temperatury od stężenia  $CO_2$  w atmosferze w latach od 1945 do 2015. Na wykresie tym można zauważyć tendencję wzrostową temperatury.

Współczynnik korelacji Pearsona wykorzystuje się do określenia poziomu zależności liniowej między zmiennymi losowymi. Wartość tego współczynnika mieści się w przedziale  $[-1, 1]$ . Im jego wartość bezwzględna jest większa, tym silniejsza jest zależność liniowa między zmiennymi, gdzie wartość 0 mówi o braku liniowej zależności, 1 zależność dodatnią, natomiast  $-1$  zależność ujemną między cechami [52]. Współczynnik ten obliczamy na podstawie wzoru:

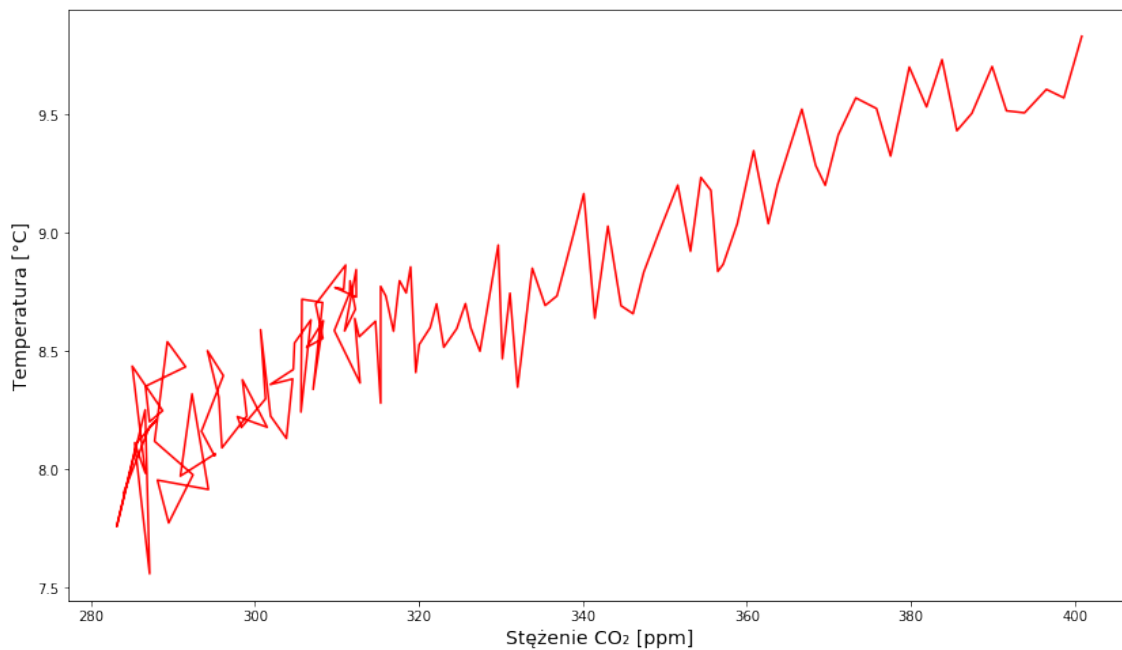
$$r_{xy} = \frac{cov(x, y)}{Sd_x \cdot Sd_y} \quad (2.1)$$

gdzie:

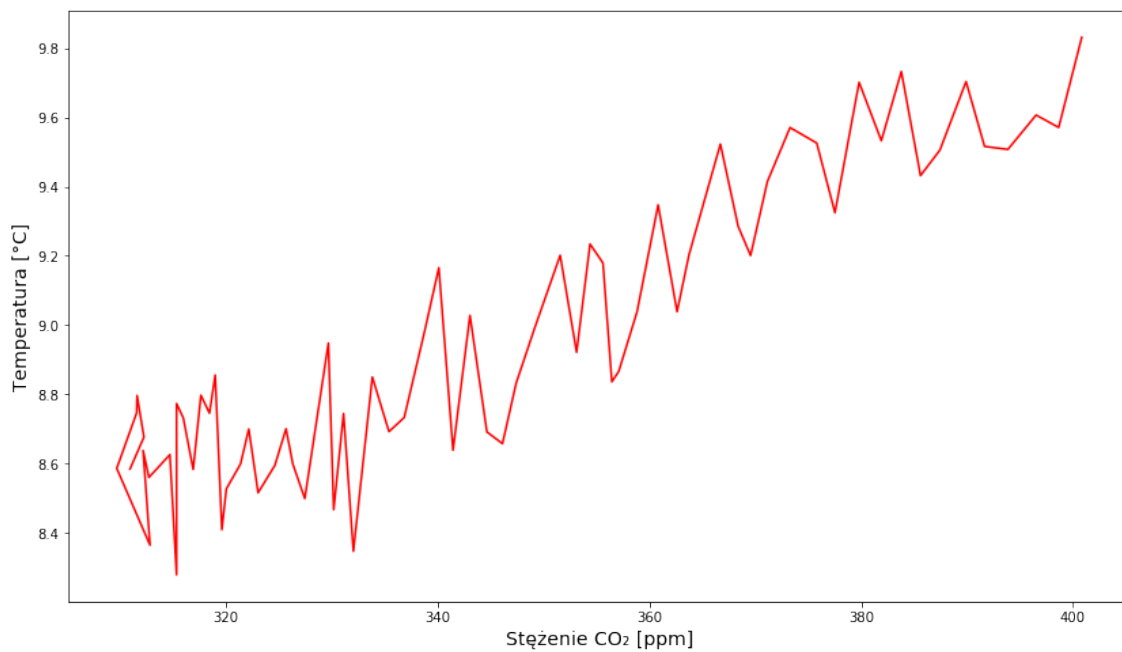
$$cov(x, y) = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n} \quad (2.2)$$

jest kowariancją między zmiennymi  $x$  i  $y$ , natomiast  $Sd_x$  i  $Sd_y$  to odchylenia standardowe tych zmiennych. Współczynnik korelacji Pearsona pomiędzy temperaturą a stężeniem  $CO_2$  wynosi: 0.9.

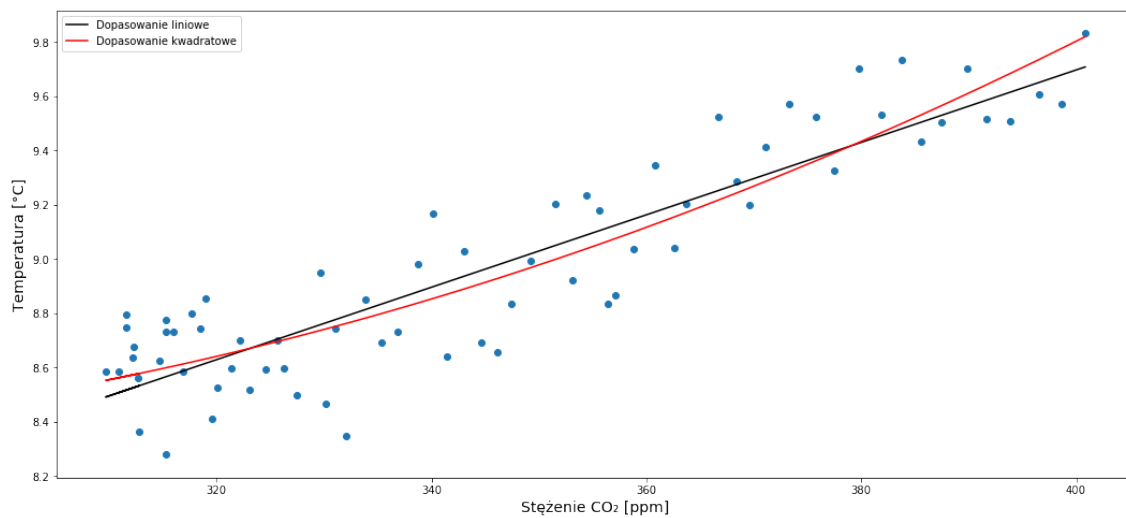
<sup>5</sup>ppm - parts per milion (liczba części na milion)



Rysunek 2.23: Koncentracja  $CO_2$  w funkcji temperatury w latach of 1850 do 2015.

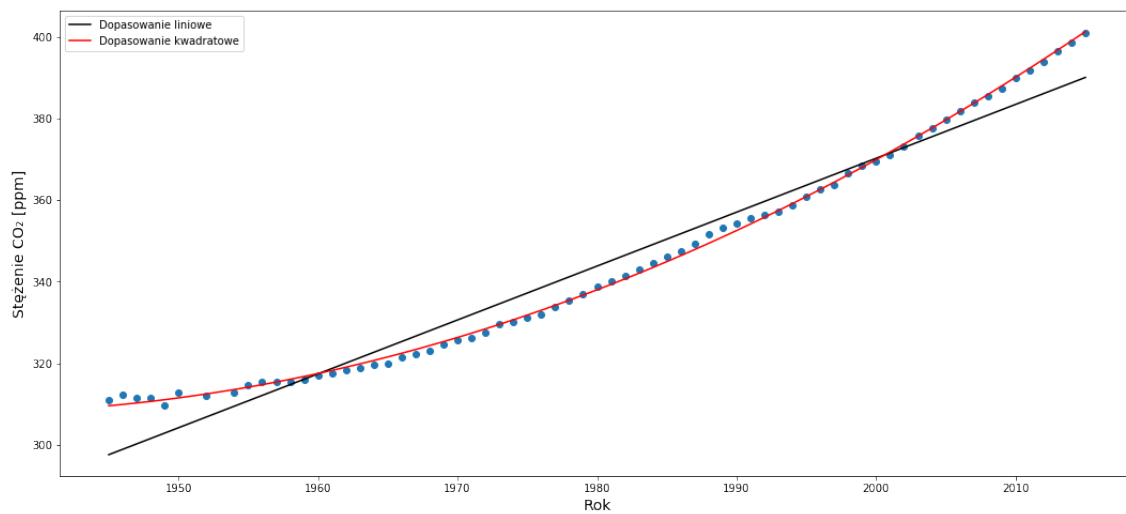


Rysunek 2.24: Koncentracja  $CO_2$  w funkcji temperatury w latach od 1945 do 2015.



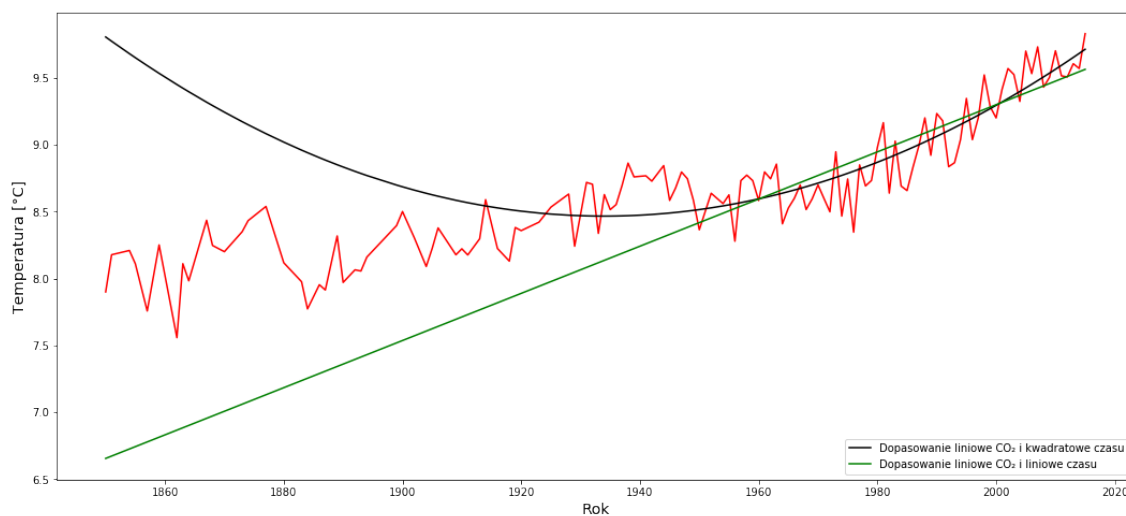
Rysunek 2.25: Zależność temperatury od stężenia  $CO_2$  w latach 1945 do 2015.

Dane zilustrowane na Rys. 2.25 przedstawiają zależność temperatury od stężenia  $CO_2$  w latach od 1945 do 2015 oraz dopasowane przy użyciu narzędzi z pakietu Scikit-learn linie trendu jako wielomiany stopnia pierwszego i drugiego.

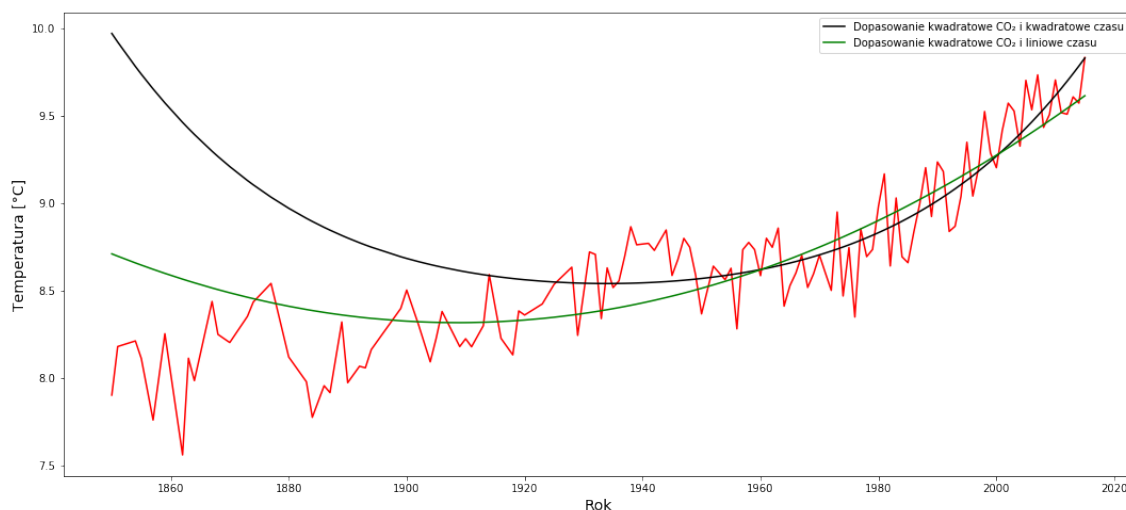


Rysunek 2.26: Wartości stężenia  $CO_2$  w danym roku w latach od 1945 do 2015.

Dane zilustrowane na Rys. 2.26 przedstawiają wartości stężenia  $CO_2$  w zależności od czasu w latach od 1945 do 2015 oraz dopasowane przy użyciu narzędzi z pakietu Scikit-learn linie trendu jako wielomiany stopnia pierwszego i drugiego.

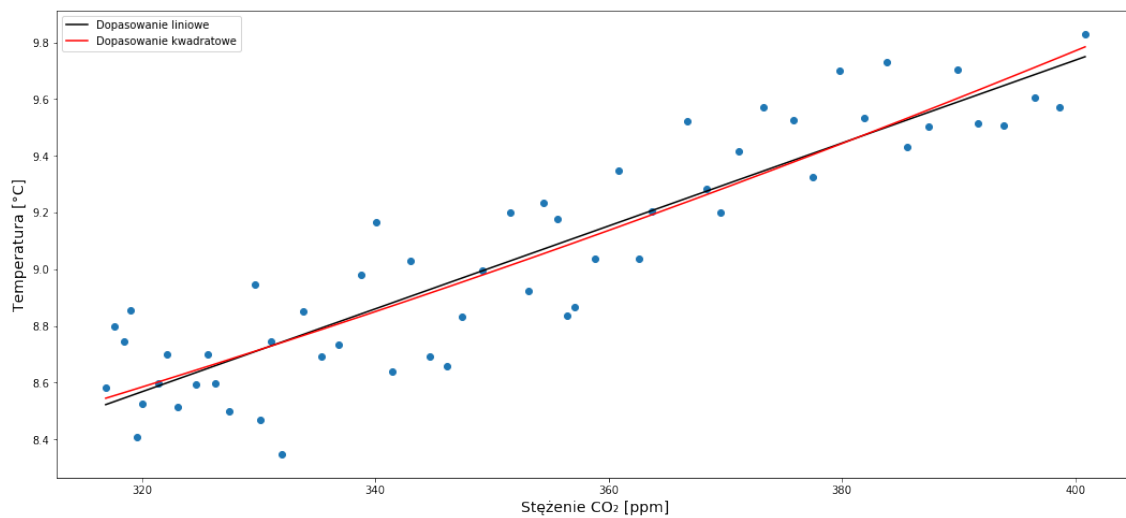


Rysunek 2.27: Zależność średniej temperatury lądu w funkcji czasu wraz z wyznaczonymi trendami, liniowym i nieliniowym w latach od 1850 do 2015.

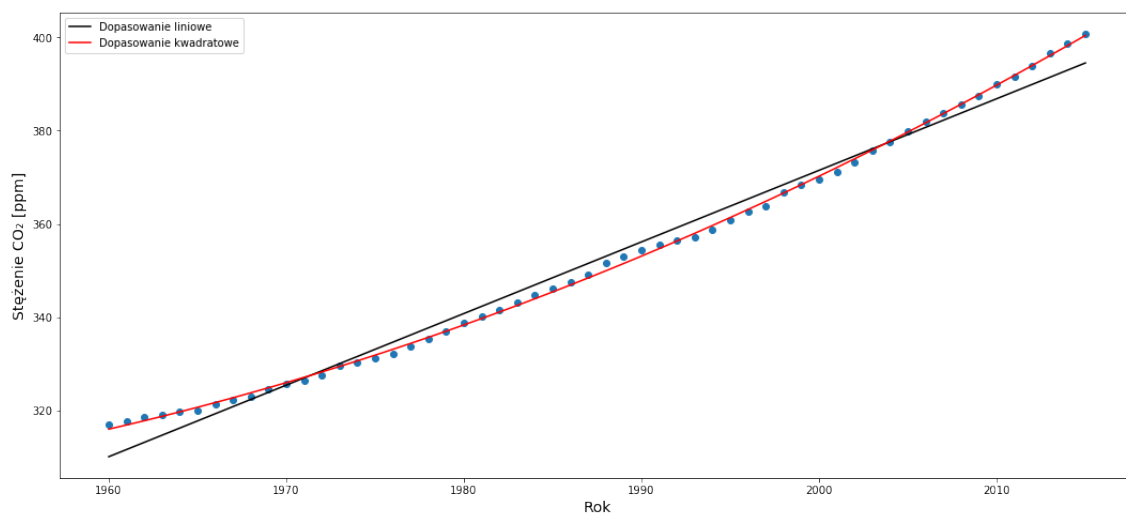


Rysunek 2.28: Zależność średniej temperatury lądu w funkcji czasu wraz z wyznaczonymi trendami nieliniowymi w latach od 1850 do 2015.

Na Rys. 2.27 oraz 2.28 widzimy zilustrowaną zależność średniej temperatury lądu od czasu z wyznaczonymi trendami. Można zaobserwować, że najlepsze dopasowanie występuje po 1960 roku.



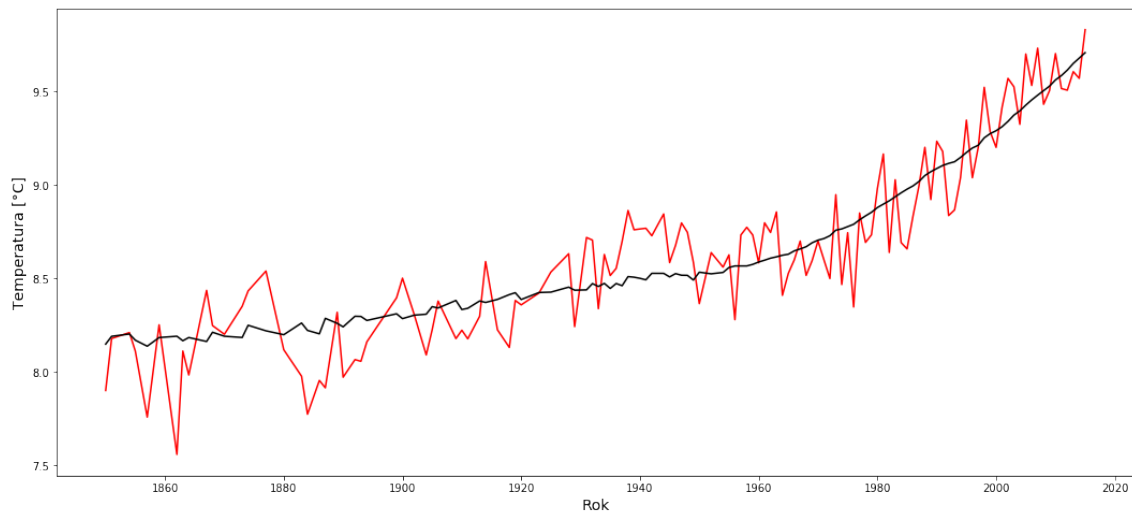
Rysunek 2.29: Zależność temperatury od  $CO_2$  w latach 1960 do 2015.



Rysunek 2.30: Stężenie  $CO_2$  w funkcji czasu w latach od 1960 do 2015.

Na Rys. 2.29 i 2.30 wyznaczono odpowiednio za pomocą regresji liniowej i kwadratowej dopasowanie trendu do zależności średniej temperatury lądu od stężenia  $CO_2$  oraz stężenia  $CO_2$  od czasu dla lat od 1960 do 2015. Na Rys. 2.28 można zaobserwować bardzo dobrą zgodność modelu kwadratowego i danych pomiarowych.

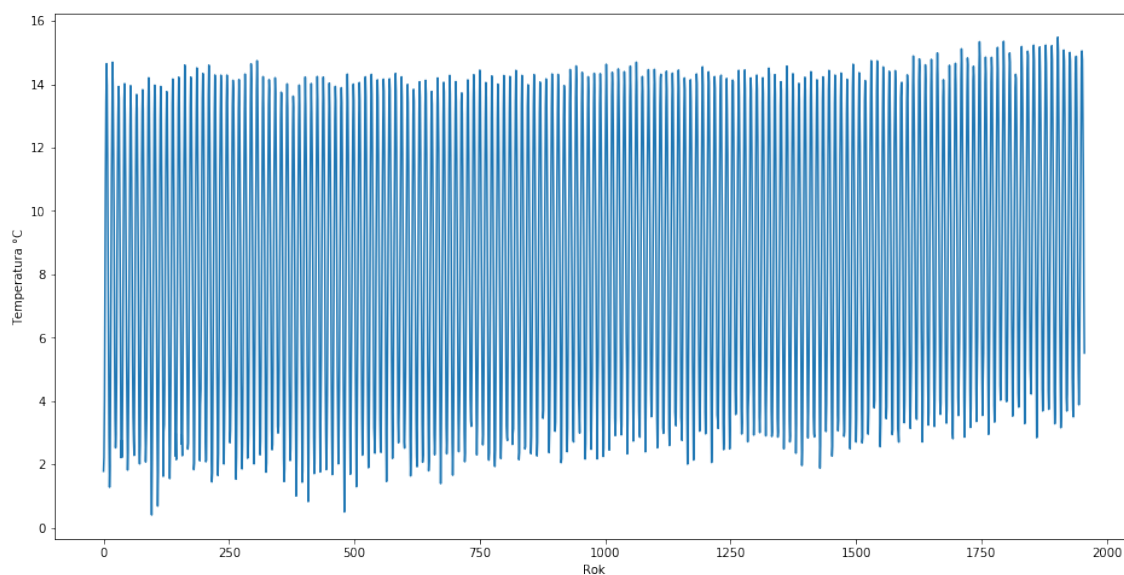




Rysunek 2.31: Zależność temperatury w funkcji czasu z wyznaczonym trendem w latach od 1850 do 2015.

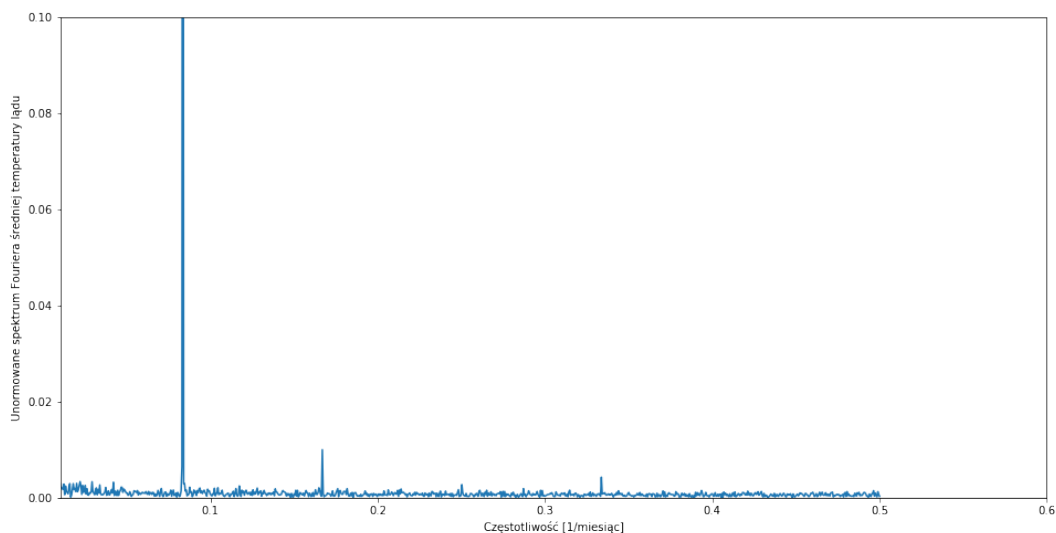
Na wykresie 2.31 zależność temperatury od czasu wyznaczona została linia trendu przy wykorzystaniu współczynników dopasowania liniowego temperatury i stężenia  $CO_2$ .

## 2.3 Analiza Fouriera

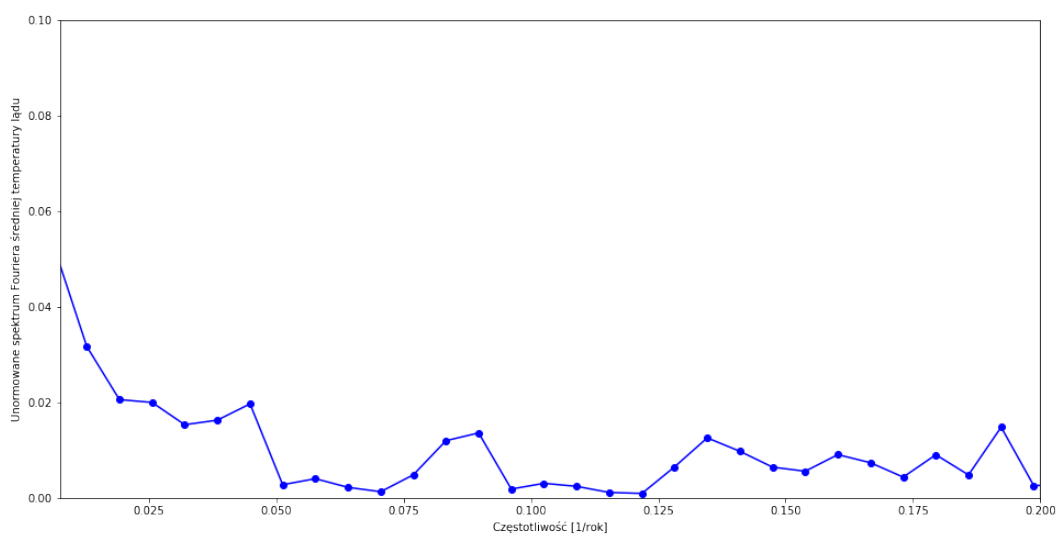


Rysunek 2.32: Wykres zależności średniej temperatury łądu od czasu.

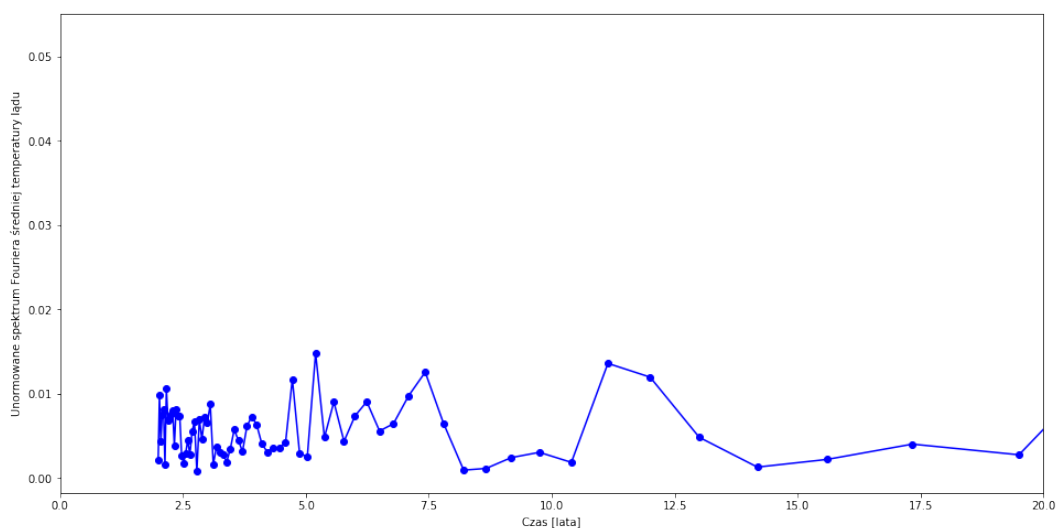
Na Rys. 2.33 zaobserwować można piki dla częstotliwości:  $0.083$  [1/miesiąc],  $0.167$  [1/miesiąc] oraz  $0.334$  [1/miesiąc]. Pierwszy z nich znacząco większy od pozostałych ukazuje roczną zmienność danych, natomiast kolejne odpowiednio zmienność sześciomiesięczną i trzymiesięczną. Wykres ten również uwidacznia czteromiesięczną zmienność danych.



Rysunek 2.33: Wykres zależności dyskretnej transformaty Fouriera temperatury od częstotliwości.

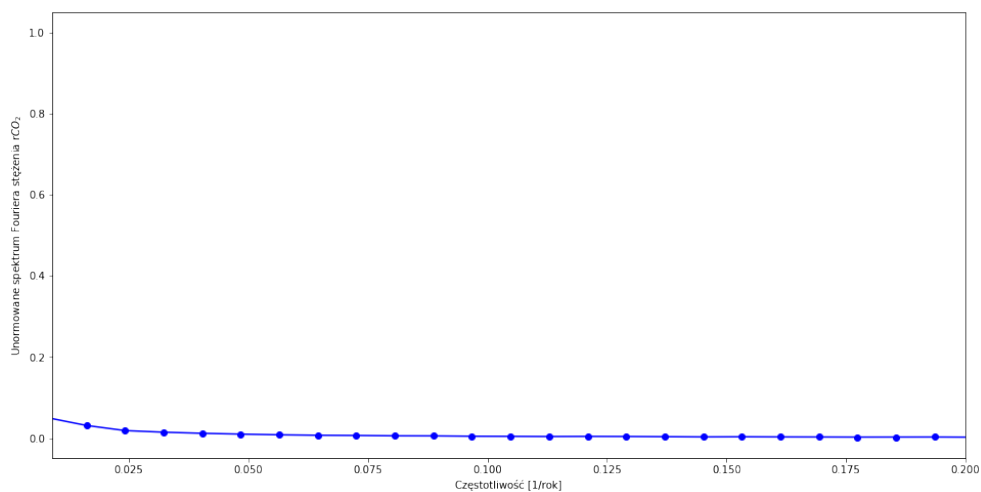


Rysunek 2.34: Wykres zależności dyskretnej transformaty Fouriera temperatury od częstotliwości.

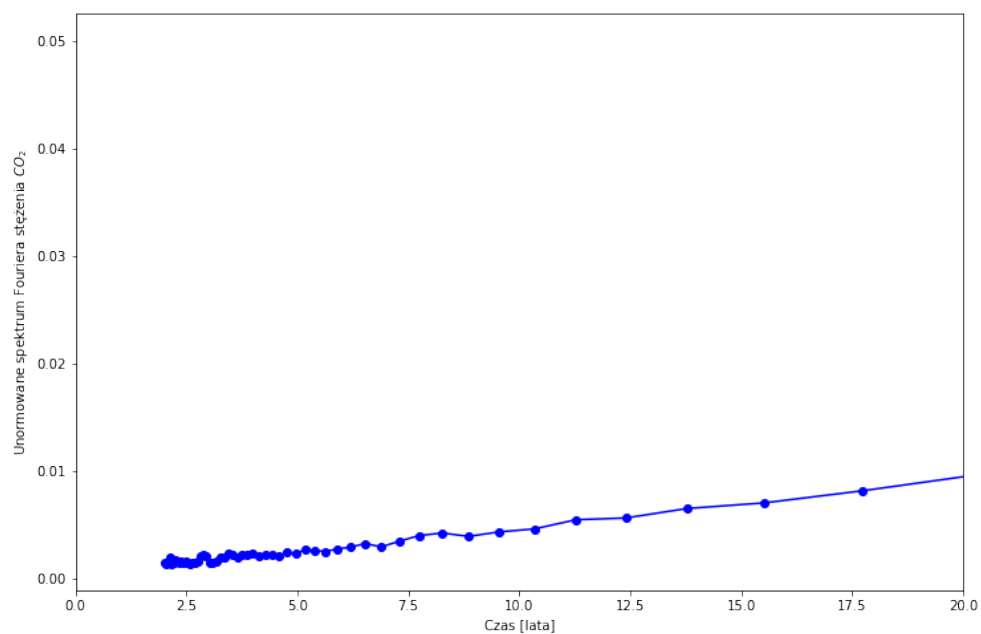


Rysunek 2.35: Wykres zależności dyskretnej transformaty Fouriera temperatury od czasu.

Na Rys. 2.34 można zauważyć maksimum dla częstotliwości 0.092 [1/rok] co może zostać zinterpretowane jako 11-letnią zmienność danych powiązana z cyklem słonecznym. Na Rys. 2.35 można wyraźniej zaobserwować wspomniany wyżej cykl słoneczny.



Rysunek 2.36: Wykres zależności dyskretnej transformaty Fouriera stężenia  $CO_2$  od częstotliwości.



Rysunek 2.37: Wykres zależności dyskretnej transformaty Fouriera stężenia  $CO_2$  od czasu.

Na Rysunkach 2.36 oraz 2.37 przedstawiona została wizualizacja wyników analizy Fouriera dla emisji  $CO_2$ . Z otrzymanych wyników widać brak oscylacji w emisji dwutlenku węgla do atmosfery.

## 2.4 Wnioski

Głównym celem niniejszej pracy, który został osiągnięty, była analiza danych dotyczących zmian średniej temperatury lądu na podstawie danych historycznych oraz związku tych zmian ze stężeniem dwutlenku węgla w atmosferze w danym okresie czasu. Stosując metody nadzorowanego uczenia maszynowego wyznaczone zostały trendy występujące w szeregu czasowym temperatury w badanym okresie. Na podstawie otrzymanych wyników można zaobserwować, że do roku około 1940 otrzymany trend był rosnący. W kolejnych latach do 1960 roku wartość średniej temperatury wykazywała trend spadkowy, natomiast od roku 1960 do 2015 nastąpiła zmiana tendencji na wzrostową przy czym, wzrost ten był znacząco szybszy niż w pierwszym z analizowanych okresów. Wykonana na podstawie uzyskanych wyników predykcja temperatury do roku 2050 przewiduje wzrost średniej temperatury lądu do wartości  $10.4^{\circ}\text{C}$ , co w sposób znaczący wpłynęłoby na zmianę klimatu na Ziemi. Obecnie istotnym problemem z punktu widzenia analizy zmian klimatycznych jest odpowiedź na pytanie jaki wpływ na temperaturę ma stężenie  $\text{CO}_2$  w atmosferze. Wykonana w pracy analiza zależności temperatury od wielkości stężenia dwutlenku węgla wskazuje na dodatnią korelację pomiędzy tymi wielkościami z czego można wnioskować, że wzrost emisji dwutlenku węgla przyczynia się do ocieplenia klimatu [53]. W ostatniej części pracy przeprowadzona została analiza Fourierska danych dotyczących temperatury i stężenia dwutlenku węgla, z której wynika istnienie 11-letniej, rocznej, półrocznej, czteromiesięcznej oraz trzymiesięcznej zmienności temperatury i braku zmienności dla stężenia  $\text{CO}_2$ . Uzyskane w pracy wyniki pokazują możliwości zastosowania nowoczesnej analizy danych z wykorzystaniem metod uczenia maszynowego do rozwiązywania problemów związanych z analizą danych klimatycznych.

## 2.5 Dodatek

```
#Poniżej wczytane zostały moduły Pythona potrzebne
#do analizy

import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression

d1 = pd.read_csv('1.csv')

#Wczytujemy plik z danymi pobrany ze strony kaggle:
#https://www.kaggle.com/timokerremans/
#how-does-climat-change-affect-different-places
#Przeanalizujemy plik z danymi dla temperatury globalnej
#zmieniającej się na przestrzeni lat 1850 do 2015

d1.head()

#Ustawione zostały nazwy kolumn
d1.columns = ['Data', 'Średnia_temperatura_lądu',
              'Niepewność_pomiaru_średniej_temperatury_lądu',
              'Max_temperatura_lądu',
              'Niepewność_pomiaru_max_temperatury_lądu',
              'Min_temperatura_lądu',
              'Niepewność_pomiaru_min_temperatury_lądu',
              'Średnia_temperatura_lądu_i_oceanów',
              'Niepewność_pomiaru_średniej_temperatury_lądu_i_oceanów']

# Wyznaczona została suma obserwacji brakujących danych w kolumnach
print('Suma braków w kolumnach:\n')
d1.isnull().sum()

d1 = d1.dropna(axis = 0)
# Usunięte zostały wszystkie wiersze w których występują brakujące dane

z = len(d1.index)
d1 = d1.set_index(np.arange(1,z+1))
# Wszystkie wiersze wyświetlamy tak aby numery kolejnych
# wierszy były po kolei

d1.head()

d1['Data'] = pd.DatetimeIndex(d1['Data']).year
d1.head()

# Poniżej za pomocą pakietu Matplotlib wygenerowane zostały wykresy
```

```

plt.figure(figsize = (14,8))
d1['Średnia_temperatura_lądu'].groupby(d1['Data'])
    .agg('mean').plot(color = 'orange')
plt.ylabel('Temperatura [C]', fontsize = 14)
plt.xlabel('Rok', fontsize = 14)
plt.show()

plt.figure(figsize = (14,8))
d1['Niepewność_pomiaru_średniej_temperatury_lądu'].groupby(d1['Data'])
    .agg('mean').plot(color = 'orange')
plt.ylabel('Niepewność pomiaru średniej temperatury [C]',
    , fontsize = 14)
plt.xlabel('Rok', fontsize = 14)
plt.show()

plt.figure(figsize = (14,8))
d1['Max_temperatura_lądu'].groupby(d1['Data']).agg('mean')
    .plot(color = 'red')
plt.ylabel('Maksymalna temperatura [C]', fontsize = 14)
plt.xlabel('Rok', fontsize = 14)
plt.show()

plt.figure(figsize = (14,8))
d1['Niepewność_pomiaru_max_temperatury_lądu'].groupby(d1['Data'])
    .agg('mean').plot(color = 'red')
plt.ylabel('Niepewność pomiaru maksymalnej temperatury [C]',
    fontsize = 14)
plt.xlabel('Rok', fontsize = 14)
plt.show()

plt.figure(figsize = (14,8))
d1['Min_temperatura_lądu'].groupby(d1['Data']).agg('mean')
    .plot(color = 'BLUE')
plt.ylabel('Minimalna temperatura [C]', fontsize = 14)
plt.xlabel('Rok', fontsize = 14)
plt.show()

plt.figure(figsize = (14,8))
d1['Min_temperatura_lądu'].groupby(d1['Data'])
    .agg('mean').plot(color = 'BLUE')
plt.ylabel('Niepewność pomiaru minimalnej temperatury lądu [C]',
    , fontsize = 14)
plt.xlabel('Rok', fontsize = 14)
plt.show()

plt.figure(figsize = (14,8))
d1['Średnia_temperatura_lądu_i_oceanów'].groupby(d1['Data'])
    .agg('mean').plot(color = 'Green')

```

```

plt.ylabel('Średnia temperatura [C]', fontsize = 14)
plt.xlabel('Rok', fontsize = 14)
plt.show()

plt.figure(figsize = (14,8))
d1['Niepewność_pomiaru_średniej_temperatury_lądu_i_oceanów']
.groupby(d1['Data']).agg('mean').plot(color = 'Green')
plt.ylabel('Niepewność pomiaru średniej temperatury [C]',
, fontsize = 14)
plt.xlabel('Rok', fontsize = 14)
plt.show()

t = d1['Średnia_temperatura_lądu']
.groupby(d1['Data']).agg('mean').index
SL = d1['Średnia_temperatura_lądu']
.groupby(d1['Data']).agg('mean').values
S0 = d1['Średnia_temperatura_lądu_i_oceanów']
.groupby(d1['Data']).agg('mean').values
plt.figure(figsize = (12,8))
plt.plot(t,SL, color = 'black',
label = 'Średnia temperatura lądu')
plt.plot(t,S0, color = 'red',
label = 'Średnia temperatura lądu i oceanów')
plt.fill_between(t, SL, S0,
color = 'green', alpha = 0.4)
plt.ylabel('Średnia temperatura [C]',
fontsize = 14)
plt.xlabel('Rok', fontsize = 14)
plt.axis([min(t),max(t),7,16.5])
plt.legend(loc = 4)
plt.show()

x = d1['Min_temperatura_lądu'].groupby(d1['Data'])
.agg('mean').index
y1 = d1['Min_temperatura_lądu'].groupby(d1['Data'])
.agg('mean').values
y2 = d1['Max_temperatura_lądu'].groupby(d1['Data'])
.agg('mean').values
y3 = d1['Średnia_temperatura_lądu'].groupby(d1['Data'])
.agg('mean').values
y4 = y2 - y1
y5 = np.absolute(d1['Niepewność_pomiaru_max_temperatury_lądu']
.groupby(d1['Data']).agg('mean').values)
y6 = np.absolute(d1['Niepewność_pomiaru_min_temperatury_lądu']
.groupby(d1['Data']).agg('mean').values)
y7 = y5 + y6
plt.figure(figsize = (16,10))
#plt.title('Wahania temperatury lądu', fontsize = 18)
plt.plot(x,y1, color = 'black', label = r'$T_{\min}$')

```



```

plt.plot(x,y2, color = 'red', label = r'$T_{\max}$')
plt.plot(x,y3, color = 'darkblue', label = r'$T_{\text{sr}}$')
plt.fill_between(x, y1, y2, color = 'green', alpha = 0.4)
plt.ylabel('Temperatura [C]', fontsize = 14)
plt.xlabel('Rok', fontsize = 14)
plt.axis([min(x),max(x),min(y7),max(y2)])
plt.legend(loc = 'best')
plt.show()

```

```

plt.figure(figsize = (16,10))
plt.plot(x, y4, color = 'orange',
label = r'$T_{\max}-T_{\min}$')
plt.ylabel('Temperatura [C]', fontsize = 14)
plt.xlabel('Rok', fontsize = 14)
plt.axis([min(x),max(x), 8, 14])
plt.legend(loc = 4)
plt.show()

```

```

plt.figure(figsize = (16,10))
plt.plot(x, y1, color = 'darkblue',
label = r'$T_{\min}$')
plt.plot(x, y1 + y6, color = 'red',
label = r'$T_{\min} + u(T_{\min})$')
plt.plot(x, y1 - y6, color = 'black',
label = r'$T_{\min} - u(T_{\min})$')
plt.fill_between(x, y1 - y6, y1 + y6,
color = 'slategray', alpha = 0.4)
plt.ylabel('Temperatura [C]', fontsize = 14)
plt.xlabel('Rok', fontsize = 14)
plt.axis([min(x),max(x),-1, 5])
plt.legend(loc = 4)
plt.show()

```

```

plt.figure(figsize = (16,10))
plt.plot(x, y2, color = 'darkblue',
label = r'$T_{\max}$')
plt.plot(x, y2 + y5, color = 'red',
label = r'$T_{\max} + u(T_{\max})$')
plt.plot(x, y2 - y5, color = 'black',
label = r'$T_{\max} - u(T_{\max})$')
plt.fill_between(x, y2 - y5, y2 + y5,
color = 'slategray', alpha = 0.4)
plt.ylabel('Temperatura [C]', fontsize = 14)
plt.xlabel('Rok', fontsize = 14)
plt.axis([min(x),max(x),10, 17])
plt.legend(loc = 4)
plt.show()

```

```

plt.figure(figsize = (16,10))

```

```

plt.plot(x, y3, color = 'darkblue',
label = r'$T_{\acute{s}r}$')
plt.plot(x, y3 + 0.5*y7, color = 'red',
label = r'$T_{\acute{s}r} + u(T_{\acute{s}r})$')
plt.plot(x, y3 - 0.5*y7, color = 'black',
label = r'$T_{\acute{s}r} - u(T_{\acute{s}r})$')
plt.fill_between(x, y3 - 0.5*y7, y3 + 0.5*y7,
color = 'slategray', alpha = 0.4)
plt.ylabel('Temperatura [C]', fontsize = 14)
plt.xlabel('Rok', fontsize = 14)
plt.axis([min(x),max(x), 5, 11])
plt.legend(loc = 4)
plt.show()

x = d1['Średnia_temperatura_lądu'].groupby(d1['Data'])
.agg('mean').index
y = d1['Średnia_temperatura_lądu'].groupby(d1['Data'])
.agg('mean').values

#Wyszukane zostały indeksy dla lat 1940 i 1960.
a1 = int(*np.nonzero(x == 1940))
a2 = int(*np.nonzero(x == 1960))

# Stworzone zostały instancje klasy LinearRegression
# i przeprowadzona została regresja liniowa

reg1 = LinearRegression()
reg2 = LinearRegression()
reg3 = LinearRegression()

X = [[i] for i in x]
X1 = [[i] for i in x[:a1+1]]
X2 = [[i] for i in x[a1:a2+1]]
X3 = [[i] for i in x[a2:]]

reg1.fit(X1, y[:a1+1])
reg2.fit(X2, y[a1:a2+1])
reg3.fit(X3,y[a2:])
print('Współczynnik nachylenia dla okresu 1850-1940:
%s\n' % round(reg1.coef_[0],3))
print('Wyraz wolny dla okresu 1850-1940:
%s\n' % round(reg1.intercept_,3))
print('Współczynnik nachylenia dla okresu 1940-1960:
%s\n' % round(reg2.coef_[0],3))
print('Wyraz wolny dla okresu 1940-1960:
%s\n' % round(reg2.intercept_,3))
print('Współczynnik nachylenia dla okresu 1960-2015:
%s\n' % round(reg3.coef_[0],3))
print('Wyraz wolny dla okresu 1960-2015:

```

```

%s' % round(reg3.intercept_,3))
plt.figure(figsize = (14,8))
plt.plot(X, y)
plt.plot(X1,reg1.predict(X1), color = 'red')
plt.vlines(1940, 8.55128838, 8.74064935, colors='r', linestyle='dashed')
plt.plot(X2,reg2.predict(X2), color = 'red')
plt.vlines(1960, 8.59708874, 8.43025924, colors='r', linestyle='dashed')
plt.plot(X3,reg3.predict(X3), color = 'red')
plt.xlabel('Rok', fontsize = 14)
plt.ylabel('Temperatura [C]', fontsize = 14)
plt.annotate("Pierwsze wydarzenie", xy=(1940, 8.645968865),
xytext=(1915, 9),
arrowprops = dict(arrowstyle = 'simple', color = 'm'))
plt.annotate("Drugie wydarzenie", xy=(1960, 8.49708874),
xytext=(1930, 8.15),
arrowprops = dict(arrowstyle = 'simple', color = 'm'))
plt.axis([min(x),max(x),min(y),max(y)])
plt.show()

```

```
X1 = [[i] for i in range(2016,2050)]
```

```

# Liniowy model predykcji na przyszłe 30 lat
# Przy utrzymaniu obecnego trendu, przewidywana temperatura
# w latach2016-2050

```

```

plt.figure(figsize = (14,8))
plt.plot(X1, reg3.predict(X1), color = 'magenta')
plt.xlabel('Rok', fontsize = 14)
plt.ylabel('Temperatura [C]', fontsize = 14)
plt.show()

```

```

# Instalacja pakietu plotly służącego do generowania
# między innymi wykresów interaktywnych

```

```
pip install plotly
```

```

import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import plotly.offline as py

```

```
d2 = pd.read_csv('2.csv')
```

```

d2.columns = ['Data', 'Średnia_temperatura',
              'Niepewność_pomiaru_średniej_temperatury_lądu',
              'Miasto', 'Państwo', 'Szerokość_geograficzna',
              'Długość_geograficzna']

```

```

print('Suma braków w kolumnach:\n')
d2.isnull().sum()

d2 = d2.dropna(axis = 0)

z = len(d2.index)
d2 = d2.set_index(np.arange(1,z+1))

# Sformatowanie daty do roku

d2['Data'] = pd.DatetimeIndex(d2['Data']).year
d2.head()

# Wybranie unikatowych wartości z kolumny Państwo
# i przyporządkowanie im średnich temperatur

srednia_temperatura = []
p = d2['Państwo'].unique()
for x in p:
    srednia_temperatura.append(d2[d2['Państwo'] ==x]
    ['Średnia_temperatura'].mean())

# Stworzenie interaktywnej mapy
dane = {'type': 'choropleth', 'locations': p, 'z': srednia_temperatura,
        'colorscale': 'Portland', 'locationmode': 'country names',
        'text': p, 'marker': {'line': {'color': 'red', 'width': 0.5}},
        'colorbar': {'title': 'Temperatura (C)'}}

układ = {'title': 'Średnia temperatura lądu',
        'geo': {'showocean': True, 'oceancolor': 'deepskyblue',
        'projection': {'type': 'natural earth'}}}

rysunek = {'data': [dane], 'layout': układ}

py.iplot(rysunek, validate = False)

def ftemp(r):

    temperatura = []
    p = d2['Państwo'].unique()
    for x in p:
        temperatura.append(d2[(d2['Państwo'] == x)
        & (d2['Data'] == r)]
        ['Średnia_temperatura'].mean())
    return temperatura

# Ograniczenie ilustrowanych danych do lat od 1850 do 2013
for r in range(1850,2014,1):

```

```

zt = ftemp(r)

dane = {'type': 'choropleth',
        'locations': p, 'z': zt, 'colorscale': 'Portland',
        'locationmode': 'country names', 'text': p, 'marker':
        {'line': {'color': 'red', 'width': 0.5}}, 'colorbar'
        : {'title': 'Temperatura (C)'}}

układ = {'title': 'Średnia temperatura lądu w %s roku'%r, 'geo
        {'showocean': True, 'oceancolor': 'deepskyblue',
        'projection': {'type': 'natural earth'}}}

rysunek = {'data': [dane], 'layout': układ}

py.iplot(rysunek, validate = False)

import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import matplotlib.ticker as ticker
from sklearn.linear_model import LinearRegression
from sklearn.preprocessing import PolynomialFeatures
from sklearn.metrics import r2_score
from sklearn.metrics import mean_squared_error
import scipy as sp
import statsmodels.api as sm

d1 = pd.read_csv('1.csv')

d1.columns = ['Data', 'Średnia_temperatura_lądu',
              'Niepewność_pomiaru_średniej_temperatury_lądu',
              'Max_temperatura_lądu',
              'Niepewność_pomiaru_max_temperatury_lądu',
              'Min_temperatura_lądu',
              'Niepewność_pomiaru_min_temperatury_lądu',
              'Średnia_temperatura_lądu_i_oceanów',
              'Niepewność_pomiaru_średniej_temperatury_lądu_i_oceanów']

d1 = d1.dropna(axis = 0)

z = len(d1.index)
d1 = d1.set_index(np.arange(1,z+1))

d1['Data'] = pd.DatetimeIndex(d1['Data']).year

y = d1['Średnia_temperatura_lądu']
.groupby(d1['Data']).agg('mean').values

```

```

d3 = pd.read_csv('3.csv')

d3.columns = ['Region', 'Kod', 'Rok', 'Ilość']

d3.head()

x1 = d3['Rok'].values
y1 = d3['Ilość'].values

a1 = int(*np.nonzero(x1 == '1850'))
a2 = int(*np.nonzero(x1 == '2015'))

X = [x for x in x1[a1:a2+1]]
Y = [y for y in y1[a1:a2+1]]

G = np.arange(1850,2016)
G1 = set([str(g) for g in G])
G2 = set(X)
G3 = list(G1-G2)
G4 = [int(h) for h in G3]

for i in range(len(G)):
    if G[i] in G4:
        G[i] = 0
    else:
        G[i] = 1

log = [bool(g) for g in G]

plt.figure(figsize = (16,12))
ax1 = plt.subplot(2,1,1)
plt.plot(X,Y, color = "k")
ax1.xaxis.set_major_locator(ticker.MultipleLocator(20))
plt.ylabel('Stężenie CO [ppm]', fontsize = 14)
plt.xlabel('Rok', fontsize = 14)
ax2 = plt.subplot(2,1,2)
plt.ylabel('Temperatura [C]', fontsize = 14)
plt.xlabel('Rok', fontsize = 14)
plt.plot(X, y[log], color = "r")
ax2.xaxis.set_major_locator(ticker.MultipleLocator(20))
plt.show()

plt.figure(figsize = (14,8))
plt.ylabel('Temperatura [C]', fontsize = 14)
plt.xlabel('Stężenie CO [ppm]', fontsize = 14)
plt.plot(Y, y[log], color = "r")
plt.show()

```

```

a3 = int(*np.nonzero(np.array(X) == '1945'))

plt.figure(figsize = (14,8))
plt.ylabel('Temperatura [C]', fontsize = 14)
plt.xlabel('Stężenie CO [ppm]', fontsize = 14)
plt.plot(Y[a3:], y[log][a3:], color = "r")
plt.show()

# Za pomocą modułu SciPy wyznaczony został
# współczynnik korelacji Pearsona

pear = sp.stats.pearsonr(y[log][a3:],Y[a3:])
print('Współczynnik korelacji Pearsona pomiędzy
      temperaturą a stężeniem CO wynosi: %s.\n' %round(pear[0],2))

# Wykonana została regresja wielomianowa
y_r = y[log][a3:]
x_r = [[x] for x in Y[a3:]]

reg1 = LinearRegression()
reg2 = LinearRegression()

q = PolynomialFeatures(degree = 2)

x_q = q.fit_transform(x_r)

reg1.fit(x_r, y_r)
y_1 = reg1.predict(x_r)

reg2.fit(x_q, y_r)
y_2 = reg2.predict(q.fit_transform(x_r))

plt.figure(figsize = (18,8))
plt.scatter(x_r, y_r)
plt.plot(x_r, y_1, color = 'black', label = 'Dopasowanie liniowe')
plt.plot(x_r, y_2, color = 'red', label = 'Dopasowanie kwadratowe')
plt.ylabel('Temperatura [C]', fontsize = 14)
plt.xlabel('Stężenie CO [ppm]', fontsize = 14)
plt.legend(loc = 2)
plt.show()

print('Błąd średniokwadratowy dla dopasowania liniowego wynosi:
      %s natomiast współczynnik determinacji: %s \n'
      % (round(mean_squared_error(y_r, y_1),2),
        round(r2_score(y_r, y_1),3)))
print('Błąd średniokwadratowy dla dopasowania kwadratowego wynosi:
      %s natomiast współczynnik determinacji: %s \n'
      % (round(mean_squared_error(y_r, y_2),2),
        round(r2_score(y_r, y_2),3)))

```

```

y_r_n = Y[a3:]
w = np.array([float(x) for x in X])
x_r_n = [[x] for x in w[a3:]]

reg3 = LinearRegression()
reg4 = LinearRegression()

reg3.fit(x_r_n, y_r_n)
y_3 = reg3.predict(x_r_n)

q1 = PolynomialFeatures(degree = 2)

x_q1 = q1.fit_transform(x_r_n)

reg4.fit(x_q1, y_r_n)
y_4 = reg4.predict(q1.fit_transform(x_r_n))

plt.figure(figsize = (18,8))
#plt.title('Stężenie CO w funkcji czasu', fontsize = 18)
plt.scatter(x_r_n, y_r_n)
plt.plot(x_r_n, y_3, color = 'black', label = 'Dopasowanie liniowe')
plt.plot(x_r_n, y_4, color = 'red', label = 'Dopasowanie kwadratowe')
plt.ylabel('Stężenie CO [ppm]', fontsize = 14)
plt.xlabel('Rok', fontsize = 14)
plt.legend(loc = 2)
plt.show()

A = reg1.coef_[0]
B = reg1.intercept_

y_n = A*(reg3.coef_[0]*w + reg3.intercept_) + B
y_n_q = A*(reg4.coef_[2]*w**2 + reg4.coef_[1]*w + reg4.intercept_) + B

plt.figure(figsize = (18,8))
ax = plt.axes()
plt.ylabel('Temperatura [C]', fontsize = 14)
plt.xlabel('Rok', fontsize = 14)
plt.plot(w, y[log], color = "r")
plt.plot(w, y_n_q, color = 'black',
label = 'Dopasowanie liniowe CO i kwadratowe czasu')
plt.plot(w, y_n, color = 'green',
label = 'Dopasowanie liniowe CO i liniowe czasu')
plt.legend(loc = 'best')
ax.xaxis.set_major_locator(ticker.MultipleLocator(20))
plt.show()

A1 = reg2.coef_[2]

```



```

B1 = reg2.coef_[1]
C1 = reg2.intercept_

y_n1 = A1*(reg3.coef_[0]*w + reg3.intercept_)**2 +
      B1*(reg3.coef_[0]*w + reg3.intercept_) + C1
y_n_q1 = A1*(reg4.coef_[2]*w**2 + reg4.coef_[1]*w +
           reg4.intercept_)**2 + B1*(reg4.coef_[2]*w**2 +
           reg4.coef_[1]*w + reg4.intercept_) + C1

plt.figure(figsize = (18,8))
ax = plt.axes()
plt.ylabel('Temperatura [C]', fontsize = 14)
plt.xlabel('Rok', fontsize = 14)
plt.plot(w, y[log], color = "r")
plt.plot(w, y_n_q1, color = 'black',
         label = 'Dopasowanie kwadratowe CO i kwadratowe czasu')
plt.plot(w, y_n1, color = 'green',
         label = 'Dopasowanie kwadratowe CO i liniowe czasu')
plt.legend(loc = 'best')
ax.xaxis.set_major_locator(ticker.MultipleLocator(20))
plt.show()

a4 = int(*np.nonzero(np.array(X) == '1960'))

y_rr = y[log][a4:]
x_rr = [[x] for x in Y[a4:]]

reg1r = LinearRegression()
reg2r = LinearRegression()

qr = PolynomialFeatures(degree = 2)
x_qr = qr.fit_transform(x_rr)

reg1r.fit(x_rr, y_rr)
y_1r = reg1r.predict(x_rr)

reg2r.fit(x_qr, y_rr)
y_2r = reg2r.predict(qr.fit_transform(x_rr))

plt.figure(figsize = (18,8))
plt.scatter(x_rr, y_rr)
plt.plot(x_rr, y_1r, color = 'black', label = 'Dopasowanie liniowe')
plt.plot(x_rr, y_2r, color = 'red', label = 'Dopasowanie kwadratowe')
plt.ylabel('Temperatura [C]', fontsize = 14)
plt.xlabel('Stężenie CO [ppm]', fontsize = 14)
plt.legend(loc = 2)
plt.show()

y_r_nr = Y[a4:]

```

```

x_r_nr = [[x] for x in w[a4:]]

reg3r = LinearRegression()
reg4r = LinearRegression()

reg3r.fit(x_r_nr, y_r_nr)
y_3r = reg3r.predict(x_r_nr)

q1r = PolynomialFeatures(degree = 2)

x_q1r = q1r.fit_transform(x_r_nr)

reg4r.fit(x_q1r, y_r_nr)
y_4r = reg4r.predict(q1r.fit_transform(x_r_nr))

plt.figure(figsize = (18,8))
plt.scatter(x_r_nr, y_r_nr)
plt.plot(x_r_nr, y_3r, color = 'black', label = 'Dopasowanie liniowe')
plt.plot(x_r_nr, y_4r, color = 'red', label = 'Dopasowanie kwadratowe')
plt.ylabel('Stężenie CO [ppm]', fontsize = 14)
plt.xlabel('Rok', fontsize = 14)
plt.legend(loc = 2)
plt.show()

Ar = reg1r.coef_[0]
Br = reg1r.intercept_

y_nr = Ar*(reg3r.coef_[0]*w + reg3r.intercept_) + Br
y_n_qr = Ar*(reg4r.coef_[2]*w**2 + reg4r.coef_[1]*
w + reg4r.intercept_) + Br

plt.figure(figsize = (18,8))
ax = plt.axes()
plt.ylabel('Temperatura [C]', fontsize = 14)
plt.xlabel('Rok', fontsize = 14)
plt.plot(w, y[log], color = "r")
plt.plot(w, y_n_qr, color = 'black',
         label = 'Dopasowanie liniowe CO i kwadratowe czasu')
plt.plot(w, y_nr, color = 'green',
         label = 'Dopasowanie liniowe CO i liniowe czasu')
plt.legend(loc = 'best')
ax.xaxis.set_major_locator(ticker.MultipleLocator(20))
plt.show()

A1r = reg2r.coef_[2]
B1r = reg2r.coef_[1]
C1r = reg2r.intercept_

```

```

y_n1r = A1r*(reg3r.coef_[0]*w +
           reg3r.intercept_)**2 + B1r*(reg3r.coef_[0]*w
           + reg3r.intercept_) + C1r
y_n_q1r = A1r*(reg4r.coef_[2]*w**2 +
             reg4r.coef_[1]*w + reg4r.intercept_)**2 +
           B1r*(reg4r.coef_[2]*w**2 + reg4r.coef_[1]*
           w + reg4r.intercept_) + C1r

plt.figure(figsize = (18,8))
ax = plt.axes()
plt.ylabel('Temperatura [C]', fontsize = 14)
plt.xlabel('Rok', fontsize = 14)
plt.plot(w, y[log], color = "r")
plt.plot(w, y_n_q1r, color = 'black', label =
'Dopasowanie kwadratowe CO i kwadratowe czasu')
plt.plot(w, y_n1r, color = 'green', label =
'Dopasowanie kwadratowe CO i liniowe czasu')
plt.legend(loc = 'best')
ax.xaxis.set_major_locator(ticker.MultipleLocator(20))
plt.show()

plt.figure(figsize = (18,8))
ax = plt.axes()
plt.ylabel('Temperatura [C]', fontsize = 14)
plt.xlabel('Rok', fontsize = 14)
plt.plot(w, y[log], color = "r")
plt.plot(w, reg1.coef_[0] * np.array(Y) + reg1.intercept_, color = "k")
ax.xaxis.set_major_locator(ticker.MultipleLocator(20))
plt.show()

import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import matplotlib.ticker as ticker
from sklearn.linear_model import LinearRegression
from sklearn.preprocessing import PolynomialFeatures
from sklearn.metrics import r2_score
from sklearn.metrics import mean_squared_error
import scipy as sp

d1 = pd.read_csv('1.csv')
d1.columns = ['Data', 'Średnia_temperatura_lądu',
             'Niepewność_pomiaru_średniej_temperatury_lądu',
             'Max_temperatura_lądu',
             'Niepewność_pomiaru_max_temperatury_lądu',
             'Min_temperatura_lądu',
             'Niepewność_pomiaru_min_temperatury_lądu',
             'Średnia_temperatura_lądu_i_oceanów',

```

```

        'Niepewność_pomiaru_średniej_temperatury_lądu_i_oceanów']
d1 = d1.dropna(axis = 0)
z = len(d1.index)
d1 = d1.set_index(np.arange(1,z+1))
d1['Data'] = pd.DatetimeIndex(d1['Data']).month
y = d1['Średnia_temperatura_lądu'].groupby(d1['Data']).agg('mean').values

ST2= d1['Średnia_temperatura_lądu'][36:]
plt.figure(figsize=(16,8))
plt.plot(range(z-36),ST2)
plt.ylabel("Temperatura C")
plt.xlabel("Rok")
plt.show()

# Wyznaczenie dyskretnej transformaty Fouriera,
# charakterystyki amplitudowej i jej normalizacja
fft = np.fft.fft(ST2)
fftabs = np.abs(fft)
spectrumNormalised = fftabs/fftabs[0]

N=ST2.size
Nyrs = 1
# próbkowanie co 1 miesiąc, dokładniej z częstotliwością 1/(1 miesiąc)
f = np.linspace(0.0, 1.0/(2.0*Nyrs), N//2)

plt.figure(figsize=(16,8))
plt.ylabel("Unormowane spektrum Fouriera średniej temperatury lądu")
plt.xlabel("Częstotliwość [1/miesiąc]")
plt.xlim(0.01,0.6)
plt.ylim(0.0,0.4)
plt.plot(f,spectrumNormalised[0:N//2])
plt.show()

plt.figure(figsize=(16,8))
plt.ylabel("Unormowane spektrum Fouriera średniej temperatury lądu")
plt.xlabel("Częstotliwość [1/miesiąc]")
plt.xlim(0.01,0.6)
plt.ylim(0.0,0.1)
plt.plot(f,spectrumNormalised[0:N//2])
plt.show()

# Wyszukiwanie indeksów danych dla których
# wartość amplitudy jest większa niż 0,05
SP = spectrumNormalised[0:N//2]
SP[SP>0.05]
np.where(SP>0.05)

```

```

# Jak działa FFT [54]
import numpy as np
from scipy import fft
# Liczba punktów próbkowania
N = 600
# Częstotliwość próbkowania
T = 1.0 / 800.0 # 1 próbowkoanie na 800 s
#punkty próbkowania:
x = np.linspace(0.0, N*T, N)
#sygnał testowy 50Hz + 80 Hz
y = np.sin(50.0 * 2.0*np.pi*x) + 0.5*np.sin(80.0 * 2.0*np.pi*x)
yf = fft(y)
#oś częstotliwości - połowa punktów
xf = np.linspace(0.0, 1.0/(2.0*T), N//2)

import matplotlib.pyplot as plt
plt.figure(figsize = (10,8))
plt.xlim(0.0,100.0)

plt.plot(xf, np.abs(yf[0:N//2])/np.abs(yf[0]))

plt.grid()
plt.ylabel("Spektrum sygnału")
plt.xlabel("Częstotliwość")
plt.show()

for i,x in zip(f,spectrumNormalised[0:N//2]):
    if x >= 0.00270 and i > 0.084:
        print(round(i,4), round(x,5))
    else:
        continue

d1 = pd.read_csv('1.csv')
d1.columns = ['Data', 'Średnia_temperatura_lądu',
              'Niepewność_pomiaru_średniej_temperatury_lądu',
              'Max_temperatura_lądu',
              'Niepewność_pomiaru_max_temperatury_lądu',
              'Min_temperatura_lądu',
              'Niepewność_pomiaru_min_temperatury_lądu',
              'Średnia_temperatura_lądu_i_oceanów',
              'Niepewność_pomiaru_średniej_temperatury_lądu_i_oceanów']
d1 = d1.dropna(axis = 0)
z = len(d1.index)
d1 = d1.set_index(np.arange(1,z+1))
d1['Data'] = pd.DatetimeIndex(d1['Data']).year
y = d1['Średnia_temperatura_lądu'].groupby(d1['Data']).agg('mean').values

ST2= d1['Średnia_temperatura_lądu']

```

```

plt.figure(figsize=(16,8))
#plt.plot(range(z),ST2)
plt.plot(range(y.shape[0]),y)
plt.xlabel("Lata od 1850 do 2015")
plt.ylabel("Temperatura")
plt.show()

fft = np.fft.fft(y)
fftabs = np.abs(fft)
spectrumNormalised = fftabs/fftabs[0]

N=y.size
Nyrs = 1
f = np.linspace(0.0, 1.0/(2.0*Nyrs), N//2)

plt.figure(figsize=(16,8))
plt.ylabel("Unormowane spektrum Fouriera średniej temperatury lądu")
plt.xlabel("Częstotliwość [1/rok]")
plt.xlim(0.0075,0.2)
plt.ylim(0.0,0.025)
plt.plot(f,spectrumNormalised[0:N//2], 'bo-')
plt.show()

plt.figure(figsize=(16,8))
plt.ylabel("Unormowane spektrum Fouriera średniej temperatury lądu")
plt.xlabel("Czas [lata]")
plt.xlim(0.0,20.)
plt.plot(1.0/f,spectrumNormalised[0:N//2], 'bo-')
plt.show()

y = d1['Średnia_temperatura_lądu']

Z = y.iloc[98::12].values

fft = np.fft.fft(Z)

fftabs = np.abs(fft)
spectrumNormalised = fftabs/fftabs[0]

N=Z.size
Nyrs = 1
f = np.linspace(0.0, 1.0/(2.0*Nyrs), N//2)

plt.figure(figsize=(16,8))
plt.ylabel("Unormowane spektrum Fouriera średniej temperatury lądu")

```

```

plt.xlabel("Częstotliwość [1/rok]")
plt.xlim(0.0075,0.2)
plt.ylim(0.0,0.1)
plt.plot(f,spectrumNormalised[0:N//2], 'bo-')
plt.show()

plt.figure(figsize=(16,8))
plt.ylabel("Unormowane spektrum Fouriera średniej temperatury łądu")
plt.xlabel("Czas [lata]")
plt.xlim(0.0,20.)
plt.plot(1.0/f,spectrumNormalised[0:N//2], 'bo-')
plt.show()

d3 = pd.read_csv('3.csv')

d3.columns = ['Region','Kod','Rok','Ilość']

d3.head()

x1 = d3['Rok'].values
y1 = d3['Ilość'].values
a1 = int(*np.nonzero(x1 == '1850'))
a2 = int(*np.nonzero(x1 == '2015'))
X = [x for x in x1[a1:a2+1]]
Y = np.array([y for y in y1[a1:a2+1]])

fft = np.fft.fft(Y)
fftabs = np.abs(fft)
spectrumNormalised = fftabs/fftabs[0]

N=Y.size
Nyrs = 1
f = np.linspace(0.0, 1.0/(2.0*Nyrs), N//2)

#Plot FFT
plt.figure(figsize=(16,8))
plt.ylabel("Unormowane spektrum Fouriera stężenia $CO_2$")
plt.xlabel("Częstotliwość [1/rok]")
plt.xlim(0.009,0.2)
plt.plot(f,spectrumNormalised[0:N//2], 'bo-')
plt.show()

#Plot FFT CO2
plt.figure(figsize = (12,8))
plt.ylabel("Unormowane spektrum Fouriera stężenia $CO_2$")
plt.xlabel("Czas [lata]")
plt.xlim(0.0,20.)
plt.plot(1.0/f,spectrumNormalised[0:N//2], 'bo-')
plt.show()

```

# Bibliografia

- [1] P. E. Lydolph, *Weather and Climate*, University of Wisconsin–Madison, Madison, 1985.
- [2] M. Popkiewicz, A. Kardaś, S. Malinowski, *Nauka o klimacie*, Warszawa 2019.
- [3] <https://www.kaggle.com/berkeleyearth/climate-change-earth-surface-temperature-data>, (dostęp: 01.06.2020).
- [4] <https://glossary.ametsoc.org/wiki/Climate>, (dostęp: 08.08.2020).
- [5] <https://www.ekologia.pl/wiedza/slowniki/leksykon-ekologii-i-ochrony-srodowiska/klimat> (dostęp: 10.11.2020).
- [6] <https://www.space.com/17816-earth-temperature.html> (dostęp: 12.11.2020)
- [7] A. P. Ingersoll, *The Runaway Greenhouse: A History of Water on Venus*, 1969
- [8] <https://www.ncdc.noaa.gov/abrupt-climate-change/Glacial-Interglacial%20Cycles> (dostęp: 12.12.2020)
- [9] R. B. Alley, *The Younger Dryas cold interval as viewed from central Greenland*, 2020
- [10] Van der Plicht, J. Akkermans, P. G. Nieuwenhuys, O. Kaneda, A. Russell, A. Tell Sabi Abyad, Syria: Radiocarbon Chronology, Cultural Change, and the 8.2 ka Event, 2011
- [11] Praca zbiorowa, *Precise timing and characterization of abrupt climate change 8200 years ago from air trapped in polar ice*, 2007
- [12] <https://www.ncdc.noaa.gov/abrupt-climate-change/End%20of%20the%20African%20Humid%20Period> (dostęp: 12.12.2020)
- [13] <https://science.sciencemag.org/content/261/5124/985> (dostęp: 12.12.2020)
- [14] David A. Hodell, Jason H. Curtis Mark Brenner, *Possible role of climate in the collapse of Classic Maya civilization*, 1995
- [15] <https://spaceplace.nasa.gov/solar-cycles/en/> (dostęp: 12.12.2020)
- [16] L. E. A. Vieira, L. Norton, A. Kretzschmar, M. Schmidt, G. A. Cheung, *Jak nachylenie orbity Ziemi wpływa na przychodzące promieniowanie słoneczne*, 2012
- [17] <https://encyklopedia.pwn.pl/haslo/sloneczna-stala;3976492.html> (dostęp: 11.01.2021)
- [18] M. Popkiewicz, *Świat na rozdrożu*, Sonia Draga, Warszawa, 2012
- [19] <https://klimatziemi.pl/amoc-atlantycka-poludnikowa-cyrkulacja-wymienna> (dostęp: 24.06.2020)



- [20] Praca Zbiorowa, Global Warming Preceded by Increasing Carbon Dioxide Concentrations during the Last Deglaciation, 2012.
- [21] P. Martin, D. Archer, D. W. Lea, Paleooceanography and Paleoclimatology, Role of deep sea temperature in the carbon cycle during the last glacial, 2005.
- [22] J. R. Toggweiler, Paleooceanography and Paleoclimatology, Variation of atmospheric CO<sub>2</sub> by ventilation of the ocean's deepest water, 1999
- [23] Kurt M. Cuffey, Covariation of carbon dioxide and temperature from the Vostok ice core after deuterium-excess correction, 2001
- [24] <https://www.skepticalscience.com/co2-lags-temperature-intermediate.htm>, (dostęp: 02.08.2020)
- [25] <https://www.fs.usda.gov/ccrc/climate-basics/climate-primer/natural-climate-cycles> (dostęp: 02.06.2020)
- [26] H. M. van Aken, The Oceanic Thermohaline Circulation: An Introduction, 2007.
- [27] E. S. Sarachik, M. A. Cane, The El Niño-Southern Oscillation Phenomenon, 2010.
- [28] <https://www.climate.gov/news-features/blogs/enso/what-el-ni%C3%B1o%E2%80%93southern-oscillation-enso-nutshell> (dostęp: 02.06.2020)
- [29] <https://icecores.org/about-ice-cores> (dostęp: 03.06.2020)
- [30] A. Boschetti, L. Massaron, Python. Podstawy nauki o danych, 2017
- [31] William Vance, Data Science: A Comprehensive Beginner's Guide to Learn the Realms Of Data Science, 2020
- [32] <https://pl.python.org/> (dostęp: 11.01.2021)
- [33] <https://www.r-project.org/> (dostęp: 11.01.2021)
- [34] <https://www.mysql.com/> (dostęp: 11.01.2021)
- [35] <https://julialang.org/> (dostęp: 11.01.2021)
- [36] <https://www.scala-lang.org/> (dostęp: 11.01.2021)
- [37] <https://www.mathworks.com/products/matlab.html> (dostęp: 11.01.2021)
- [38] <http://projektfreelancer.pl/data-science-6-jezykow-programowania-ktorych-znajomosc-procentuje/> (dostęp: 11.12.2020)

- [39] <https://www.scipy.org/> (data dostępu: 20.08.2020)
- [40] <https://matplotlib.org/> (data dostępu: 20.08.2020)
- [41] <http://www.numpy.org/> (data dostępu: 20.08.2020)
- [42] <https://pandas.pydata.org/> (data dostępu: 20.08.2020)
- [43] <http://jupyter.org/> (data dostępu: 20.08.2020)
- [44] <https://aibusiness.pl/czym-jest-uczenie-maszynowe-machine-learning/>  
(dostęp: 02.01.2021)
- [45] S. Raschka, Python Uczenie maszynowe, 2016
- [46] <http://www.stat.yale.edu/Courses/1997-98/101/linreg.htm> (dostęp: 11.12.2020)
- [47] [http://zet10.ipee.pwr.wroc.pl/cyf/cps\\_w07\\_v15.pdf](http://zet10.ipee.pwr.wroc.pl/cyf/cps_w07_v15.pdf) (dostp : 11.12.2020)
- [48] <https://www.kaggle.com/> (dostp : 11.12.2020)
- [49] <https://www.wri.org/blog/2014/05/history-carbon-dioxide-emissions>  
(dostp : 11.12.2020)
- [50] Loyd D. Keigwin, *The Little Ice Age and Medieval Warm Period in the Sargasso Sea*, 1996
- [51] <https://ourworldindata.org/co2-and-other-greenhouse-gas-emissions>  
(dostp : 11.12.2020)
- [52] <https://stat.gov.pl/metainformacje/slownik-pojec/pojecia-stosowane-w-statystyce-publicznej/3033,pojecie.html> (dostp : 11.12.2020)
- [53] N. L. Bindoff : *Detection of Climate Change and Attribution of Causes : The Science of Climate Change*, 1995.
- [54] <https://www.ritchievink.com/blog/2017/04/23/understanding-the-fourier-transform-by-example/> (dostp : 11.01.2020)