

Aproksymacja profilu wysokościowego

1. Wstęp

Celem projektu było zaimplementowanie metod aproksymacji interpolacyjnych wykorzystujących wielomian interpolacyjny Lagrange'a, a także funkcje sklejane trzeciego stopnia. Zweryfikowano przydatność obu metod do tego zagadnienia.

Projekt został wykonany w języku Python. Do testów wykorzystano pięć wybranych tras o zróżnicowanym charakterze:

- SpacernikGdansk.csv - trasa płaska,
- Hel_yeah.csv - trasa zróżnicowana (wzniesienia i odcinki płaskie),
- WielkiKanionKolorado.csv - trasa o wielu stromych wzgórzach,
- MountEverest.csv - , trasa o jednym wyraźnym wzgórzem.

Dane pobrane z platformy Enauczanie stanowią 512 próbek składających się z odległości od początku trasy i wysokości w poszczególnych punktach.

2. Metodę wykorzystującą wielomian interpolacyjny Lagrange'a

Interpolacja Lagrange'a to globalna metoda przybliżania funkcji. Baza Lagrange do interpolacji jest określona wzorem:

$$\phi_i(x) = \prod_{j=1, j \neq i}^{n+1} \frac{(x - x_j)}{(x_i - x_j)}, \text{ gdzie } i = 1, 2, \dots, n+1$$

Do wyznaczenia funkcji interpolującej, wykorzystano wzór:

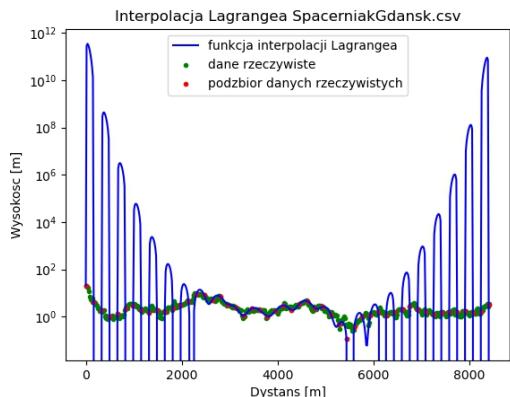
$$F(x) = \sum_{i=1}^n y_i \phi_i(x)$$

Dane wejściowe są traktowane jako dane rzeczywiste, natomiast do interpolacji konieczne jest utworzenie i wykorzystanie podzbiorów tych danych. W celu zbadania wpływu liczby punktów węzłowych na wyniki wykorzystano podzbiory danych o różnej liczbie węzłów i różnym rozmieszczeniu.

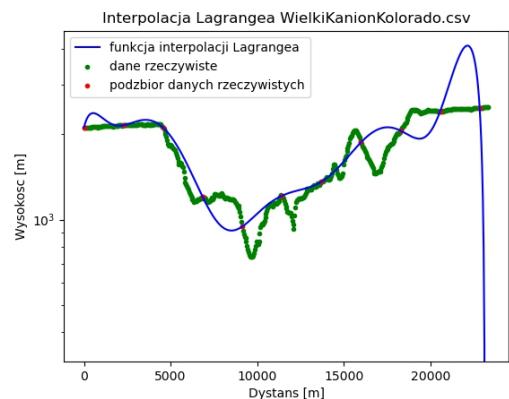
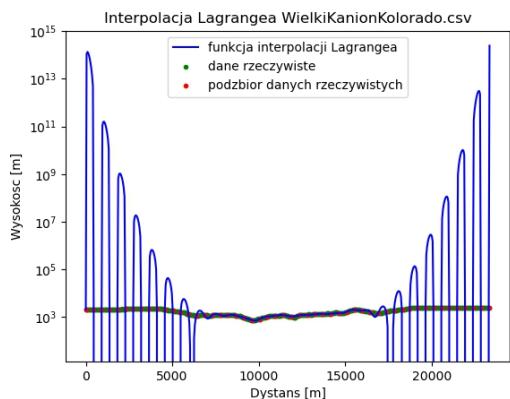
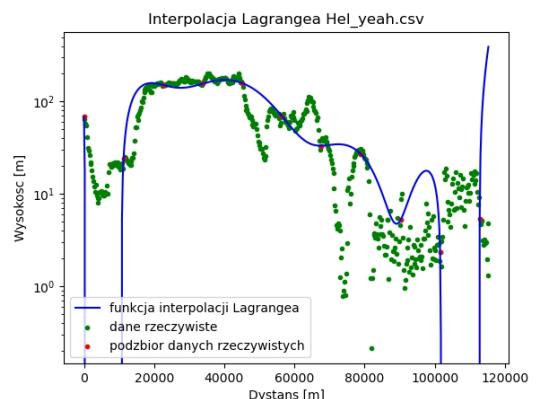
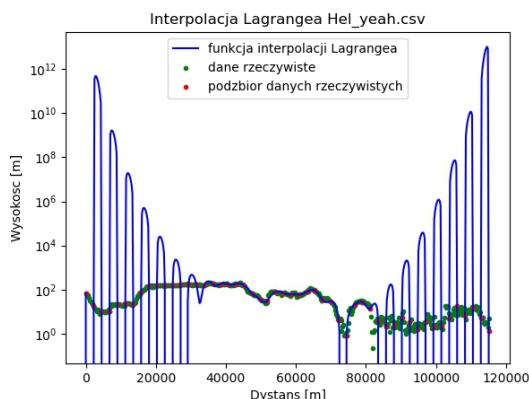
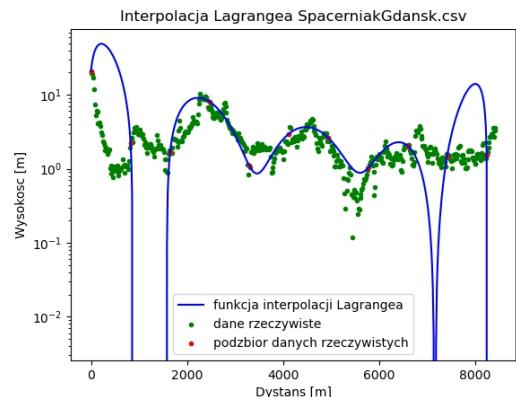
a. Wpływ liczby punktów węzłowych i rozmieszczenia na wyniki

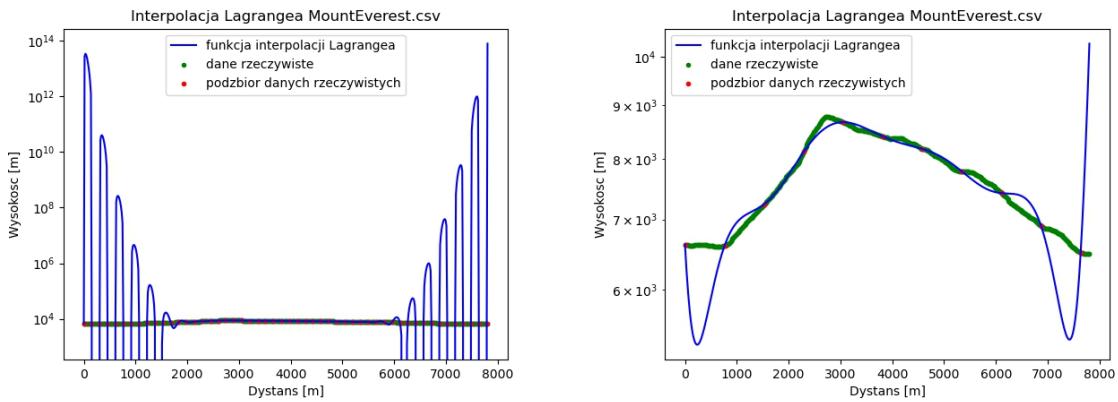
Poniższe wykresy przedstawiają wyniki przeprowadzonej interpolacji, do której wykorzystano podzbiór składający się z różnej liczby punktów o równomiernym rozmieszczeniu

Podzbiór składający się z 52. punktów, stanowiących co 10. punkt danych rzeczywistych.



Podzbiór składający się z 11. punktów, stanowiących co 50. punkt danych rzeczywistych.

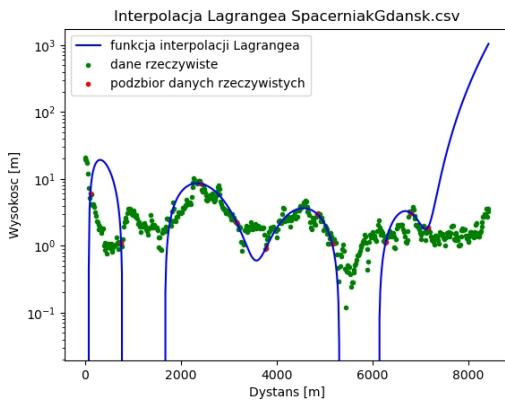




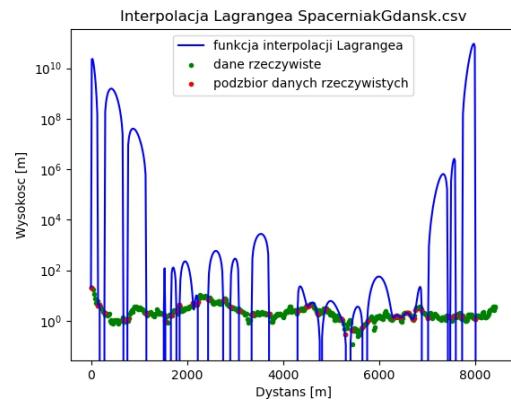
Porównując powyższe wykresy można stwierdzić, że liczba punktów węzłowych ma wpływ na uzyskane wyniki. Dla większej liczby punktów uzyskana funkcja jest bardziej dokładna, natomiast na krawędziach pojawiają się oscylacje. Jest to efekt Rungego. Problem pojawia się, gdy wykorzystuje się wielomiany wysokiego stopnia do interpolacji węzłów w równoodległych punktach. Dla mniejszej liczby punktów funkcja jest mniej dokładna. Interpolacja Lagrange'a jest ryzykowna. Za dużo węzłów powoduje efekt Rungego, za mało - niedokładną interpolację.

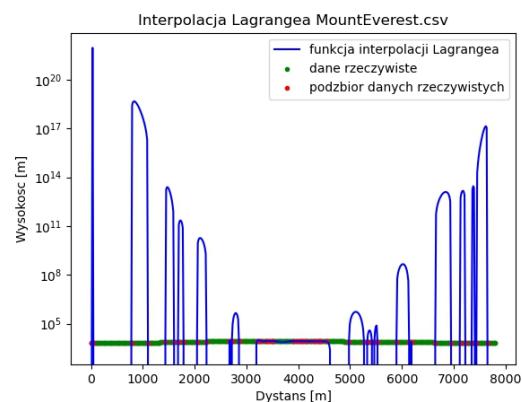
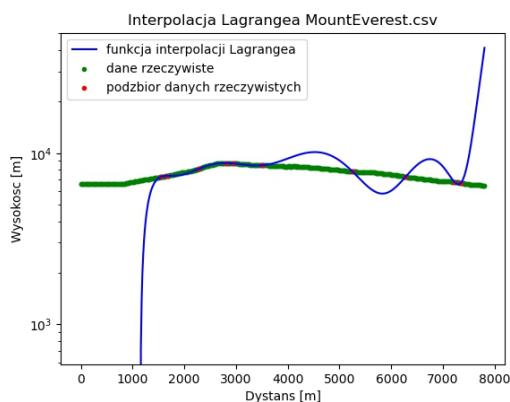
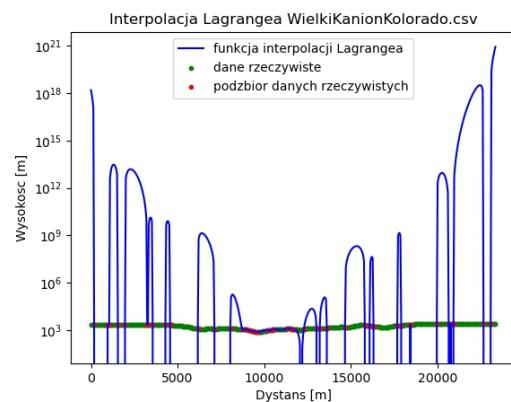
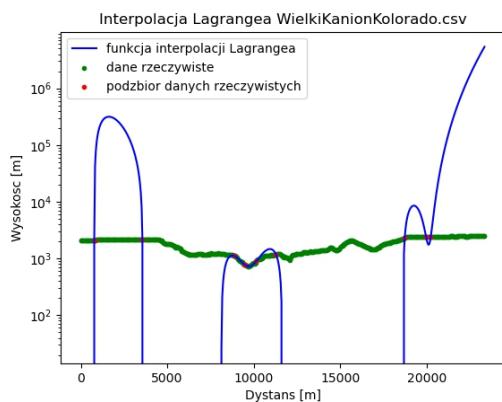
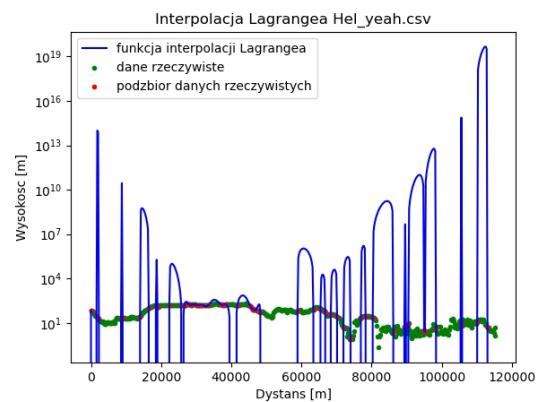
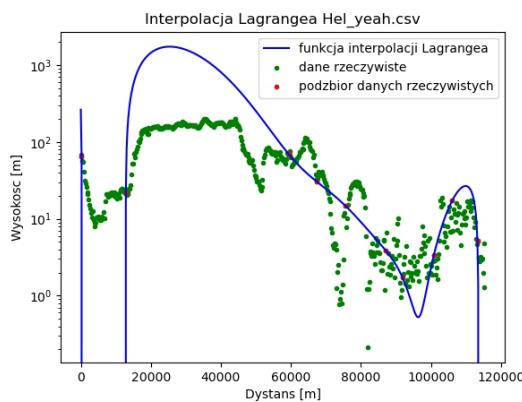
Poniższe wykresy przedstawiają wyniki przeprowadzonej interpolacji, do której wykorzystano podzbiór składający się z różnej liczby punktów o nierównomiernym rozmieszczeniu.

Podzbiór składający się z 10. losowo wybranych punktów.



Podzbiór składający się z 50. losowo wybranych punktów.



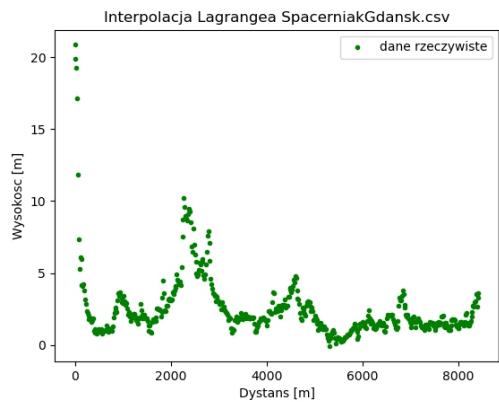


Na podstawie powyższych wykresów można stwierdzić, że losowe rozmieszczenie punktów, powoduje duże niedokładności. Zauważalne są również oscylacje na całym przedziale.

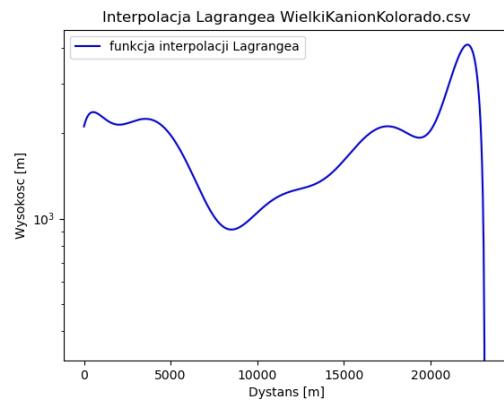
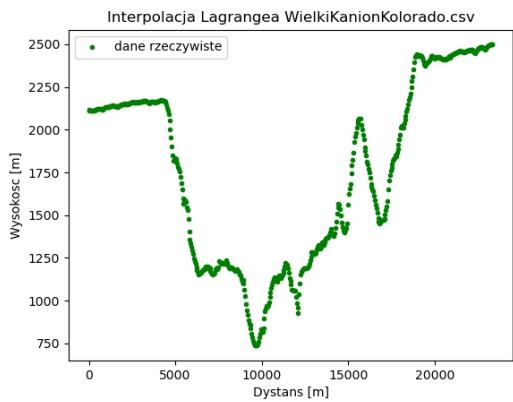
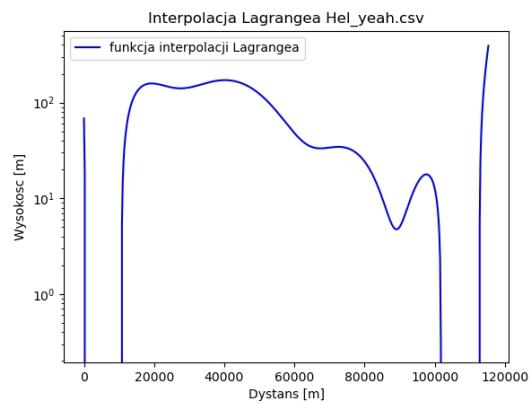
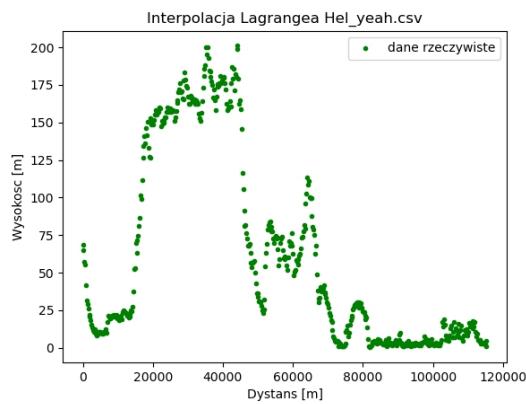
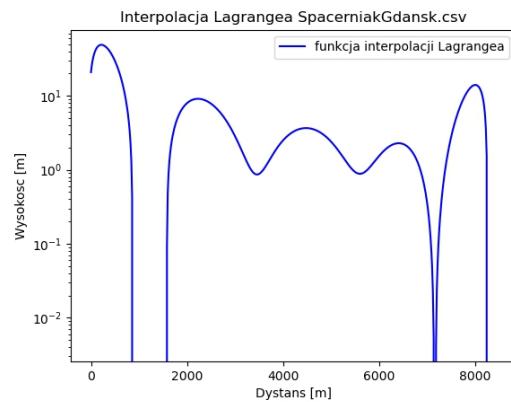
b. Wpływ charakteru trasy na wyniki

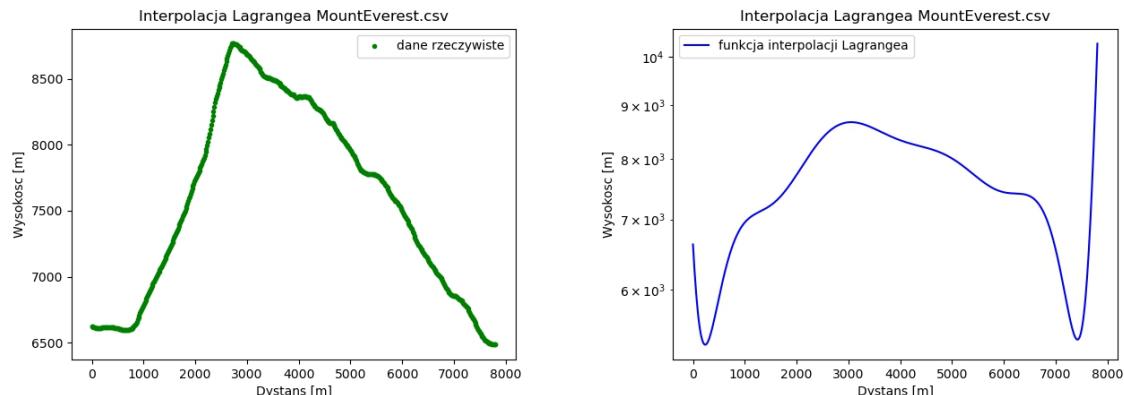
Poniższe wykresy przedstawiają osobno trasę i funkcję, będącą wynikiem interpolacji.

Trasa



Funkcja po interpolacji





Na podstawie powyższych wykresów, można zauważać, że charakter trasy ma wpływ na wyniki. W przypadku *Mount Everest* trasa utrzymuje stały charakter (najpierw ciągły wzrost wysokości w stosunku do odległości od punktu początkowego, następnie ciągły spadek). Takie właściwości trasy dały najbardziej przybliżone kształty funkcji interpolującej (pomijając oscylacje na krańcach). Trasa płaska *SpacerniakGdansk* również utrzymuje stały charakter i wyniki są zbliżone. Dla dynamicznie zmieniających się tras jak w przypadku *Hel yeah* i *WielkiKanionKolorado* utrudnione jest uzyskanie poprawnych wyników co skutkuje mniejszą dokładnością.

c. Błąd interpolacji

Do obliczenia błędu interpolacji wykorzystano algorytm RMS (Root Mean Square) o wzorze:

$$RMS = \sqrt{\frac{\sum_{i=1}^n (wartość\ dokładna - wartość\ obliczona)^2}{n}}$$

Dla zbioru składającego się z 52. punktów równomiernie rozłożonych błęd wynosi odpowiednio:

```
Blad interpolacji: Hel_yeah.csv: 2015467503706.952
Blad interpolacji: MountEverest.csv: 5602613733186.038
Blad interpolacji: SpacerniakGdansk.csv: 30521948151.664326
Blad interpolacji: WielkiKanionKolorado.csv: 18613533116128.78
```

Błąd interpolacji okazał się bardzo duży. Spowodowane jest to oscylacjami

Dla zbioru składającego się z 11. punktów równomiernie rozłożonych błęd wynosi odpowiednio:

```
Blad interpolacji: Hel_yeah.csv: 81.42729794764266
Blad interpolacji: MountEverest.csv: 500.9951775006126
Blad interpolacji: SpacerniakGdansk.csv: 10.967521369824993
Blad interpolacji: WielkiKanionKolorado.csv: 571.9955927832766
```

Mimo, że kształt funkcji interpolującej dla *Mount Everest* okazał się najbardziej zbliżony do oryginalnej, błęd jest największy. Spowodowane jest to dużymi różnicami pomiędzy wysokością uzyskaną a dokładną na krańcach. W przypadku *WielkiKanionKolorado* błęd

również jest duży ze względu na znaczne różnice wysokości - rzędu kilku tysięcy. Błędy są mniejsze dla *Hel Yeah* i *SpacerniakGdansk*, ponieważ różnica wysokości na tych odcinkach jest mała - rzędu 100 i 10.

Porównując uzyskane wartości można stwierdzić, że dla mniejszej liczby punktów błęd interpolacji jest mniejszy.

3. Metoda wykorzystująca funkcje sklejane trzeciego stopnia

Interpolacja funkcjami sklejonymi (splajanami) to interpolacja lokalna. Polega na wyznaczeniu wielomianów 3. stopnia między poszczególnymi węzłami podzbioru danych rzeczywistych.

Wielomiany te mają postać $S_i(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3$. Dla każdego przedziału należy wyznaczyć 4 współczynniki a_i, b_i, c_i, d_i , dlatego dla $n+1$ węzłów jest konieczność wyznaczenia układu o $4n$ równaniach:

- $S_j(x_j) = f(x_j)$, dla $j = 0, 1, \dots, n-1$
- $S_j(x_{j+1}) = f(x_{j+1})$, dla $j = 0, 1, \dots, n-1$
- dla węzłów wewnętrznych $S'_{j-1}(x_j) = S'_j(x_j)$, dla $j = 1, \dots, n-1$
- dla węzłów wewnętrznych $S''_{j-1}(x_j) = S''_j(x_j)$, dla $j = 1, \dots, n-1$
- na krawędziach $S''_0(x_0) = 0, S''_{n-1}(x_n) = 0$

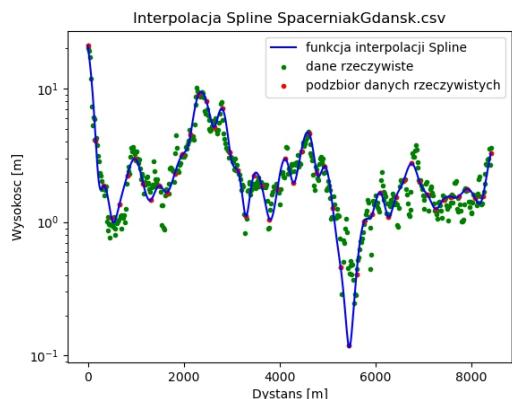
Powyższy układ równań można przedstawić w postaci $Ax = b$, gdzie A jest macierzą równań, x jest wektorem współczynników, b jest wektorem wartości

Tak otrzymany układ równań został rozwiązyany metodą faktoryzacji LU. Aby zapobiec dzieleniu przez zero na diagonali, zastosowano pivoting, czyli zamianę wierszami. Po rozwiązaniu układu równań wynikiem są współczynniki. Aby wyznaczyć wartość funkcji $S_i(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3$ sprawdzano w jakim przedziale znajduje się x i podstawiano odpowiednie dane do równania.

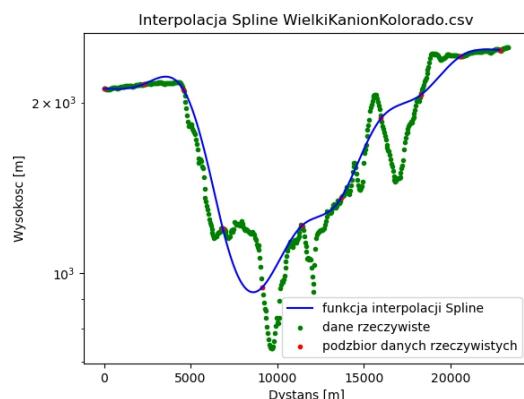
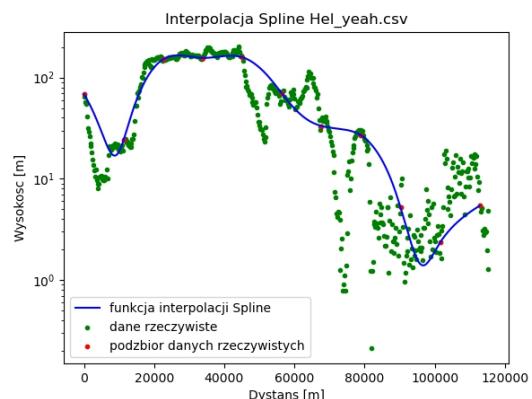
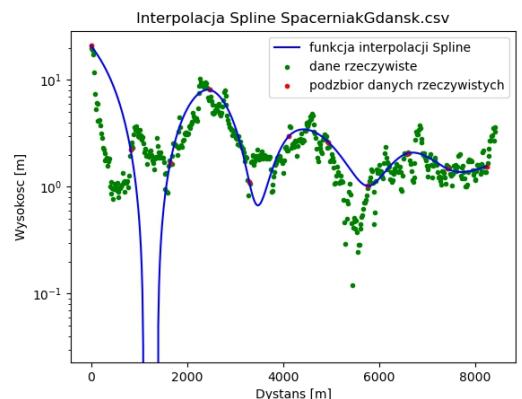
a. Wpływ liczby i rozmieszczenia punktów węzłowych na wyniki

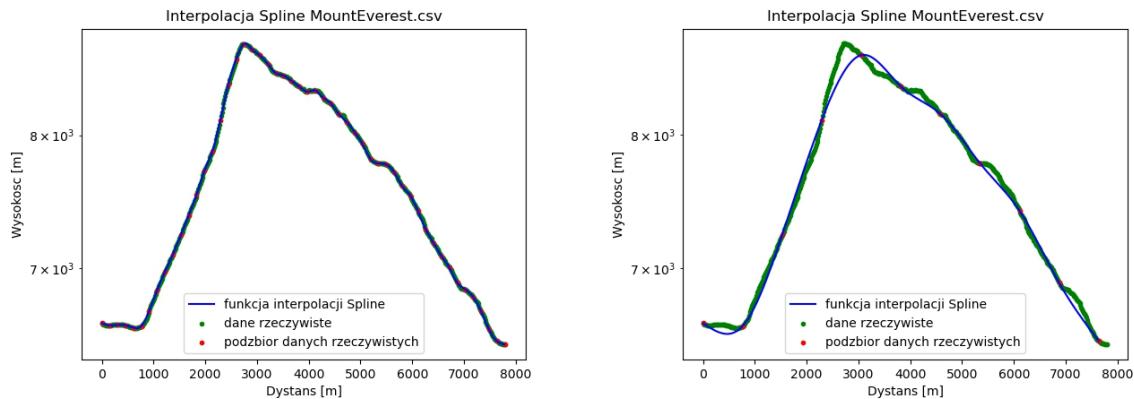
Poniższe wykresy przedstawiają wyniki przeprowadzonej interpolacji, do której wykorzystano podzbiór składający się z różnej liczby punktów o równomiernym rozmieszczeniu

Podzbiór składający się z 52. punktów, stanowiących co 10. punkt danych rzeczywistych.



Podzbiór składający się z 11. punktów, stanowiących co 50. punkt danych rzeczywistych.

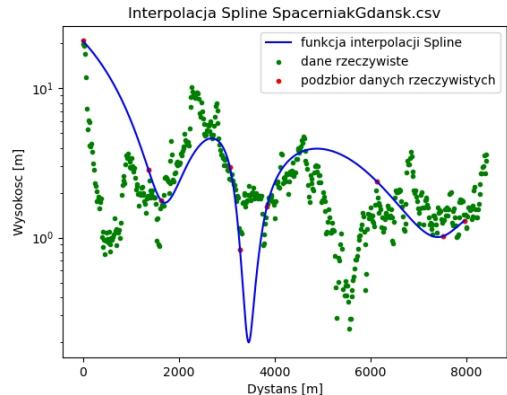




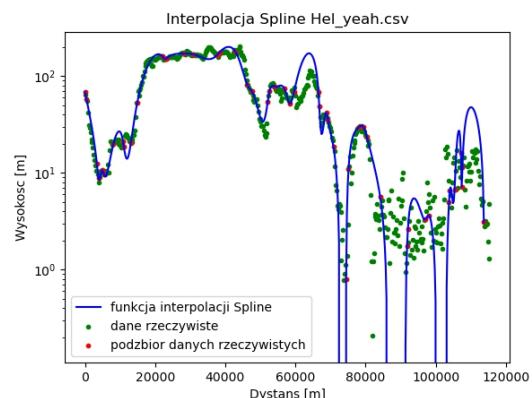
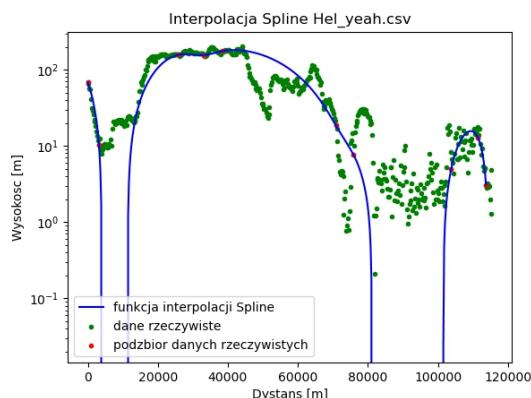
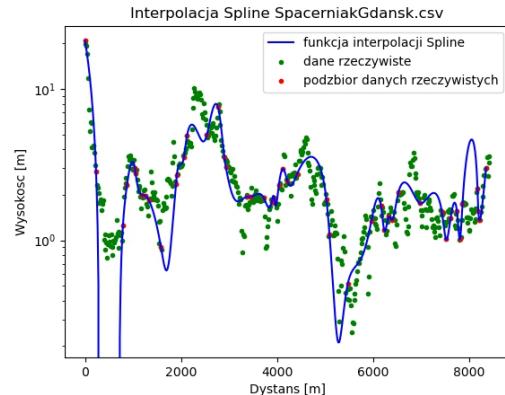
Porównując powyższe wykresy można stwierdzić, że dla tej metody liczba punktów węzłowych ma również wpływ na uzyskane wyniki. Dla większej liczby punktów, a za razem gęstsze rozmielenia uzyskana funkcja jest dokładniejsza.

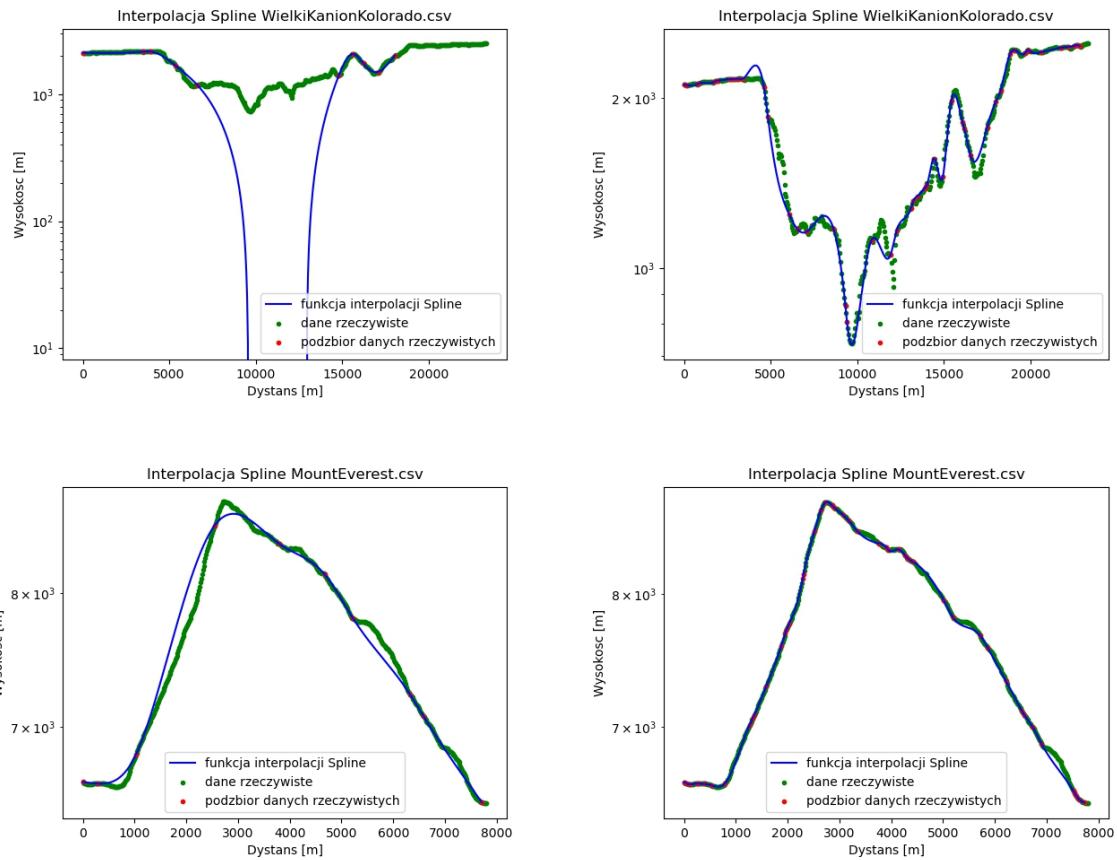
Poniższe wykresy przedstawiają wyniki przeprowadzonej interpolacji, do której wykorzystano podzbiór składający się z różnej liczby punktów o nierównomiernym rozmieleniu

Podzbiór składający się z 10. losowo wybranych punktów.



Podzbiór składający się z 50. losowo wybranych punktów.



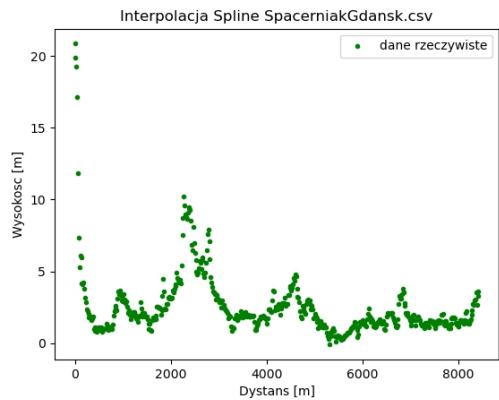


Na podstawie powyższych wykresów można stwierdzić, że dla tej metody, losowy wybór odpowiedniej liczby punktów węzłowych może doprowadzić do dość przybliżonego rozwiązania (jak w przypadku *Mount Everest*). Większa liczba punktów daje dokładniejsze przybliżenie.

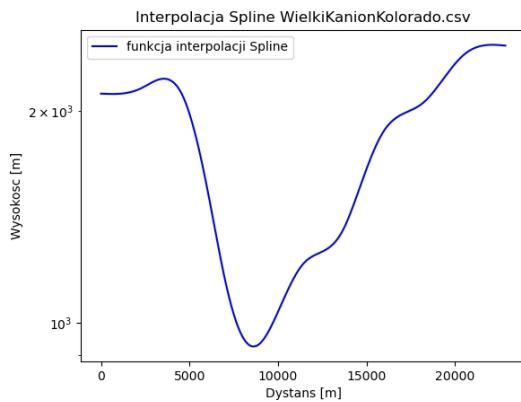
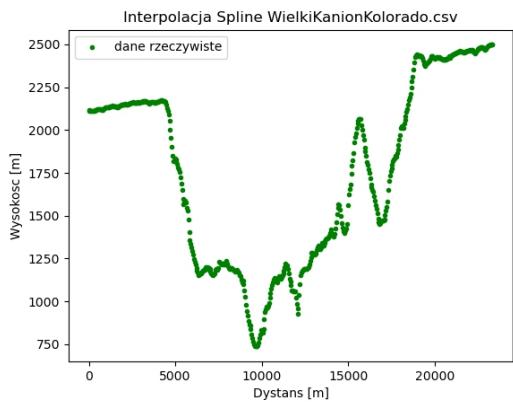
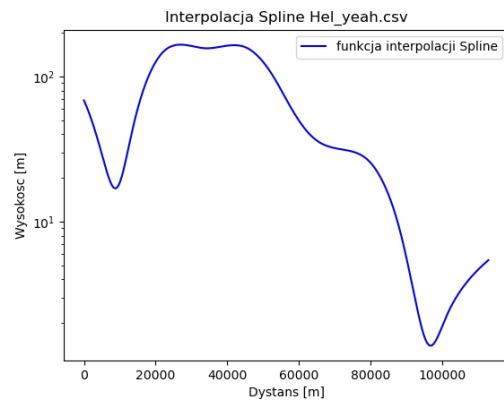
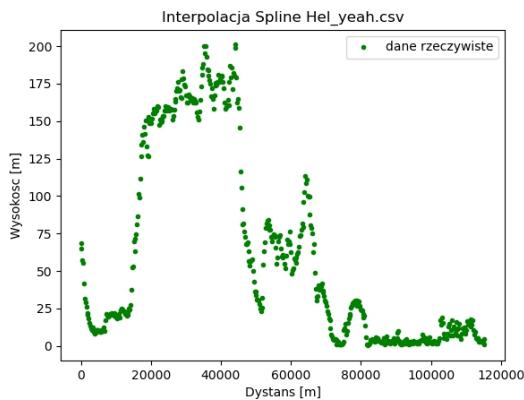
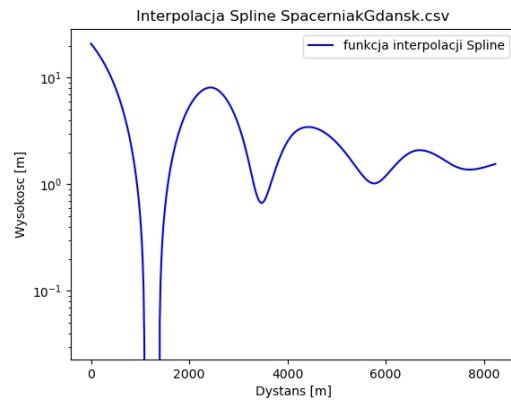
b. Wpływ charakteru trasy na wyniki

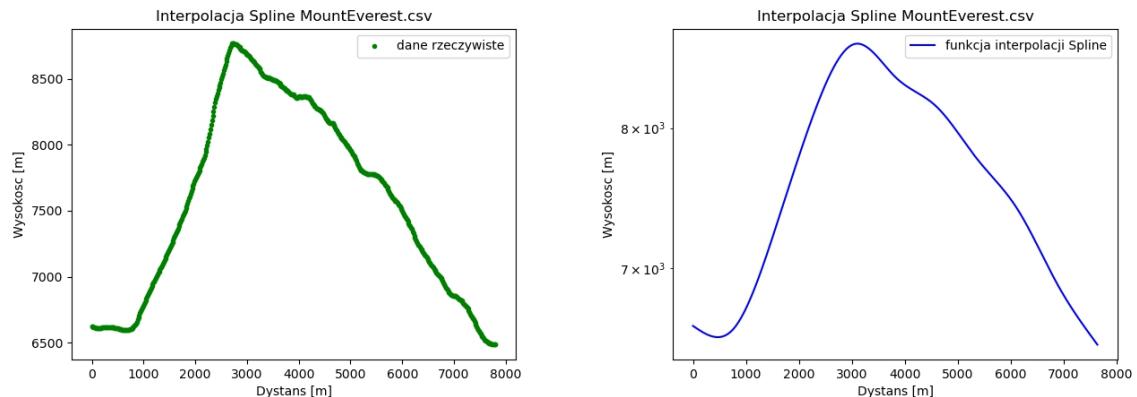
Poniższe wykresy przedstawiają osobno trasę i funkcję, będącą wynikiem interpolacji.

Trasa



Funkcja po interpolacji





Na podstawie powyższych wykresów, można zauważać, że charakter trasy ma wpływ na wyniki. Lepsze wyniki uzyskano np. dla *Mount Everest* i *SpacerniakGdansk*, gdzie trasa ma stały charakter (gwałtowny wzrost wysokości i gwałtowny spadek lub trasa płaska), niż dla danych zmiennych jak w przypadku *Hel Yeah* i *WielkiKanionKolorado*.

c. Błąd interpolacji

Do obliczenia błędu interpolacji wykorzystano ten sam wzór co dla metody Lagrange'a:

Dla zbioru składającego się z 52. punktów równomiernie rozłożonych błąd wynosi odpowiednio:

```
Blad interpolacji: Hel_yeah.csv: 7.583166992481072
Blad interpolacji: MountEverest.csv: 5.39232352218102
Blad interpolacji: SpacerniakGdansk.csv: 0.618214899512625
Blad interpolacji: WielkiKanionKolorado.csv: 28.833830254730927
```

Błąd interpolacji okazał się dość mały, szczególnie dla *SpacerniakGdansk*. Można stwierdzić, że w tym przypadku różnice między funkcją rzeczywistą a interpolującą są znikomo małe.

Dla zbioru składającego się z 11. punktów równomiernie rozłożonych błąd wynosi odpowiednio:

```
Blad interpolacji: Hel_yeah.csv: 25.88699115421372
Blad interpolacji: MountEverest.csv: 63.23375516235944
Blad interpolacji: SpacerniakGdansk.csv: 2.744212896511504
Blad interpolacji: WielkiKanionKolorado.csv: 155.33855775080423
```

Porównując uzyskane wartości można stwierdzić, że dla mniejszej liczby punktów błąd interpolacji jest większy.

4. Wnioski i porównanie obu metod

Na podstawie powyższej analizy można stwierdzić, że interpolacja metodą Lagrange'a jest ryzykowna, ponieważ może wystąpić efekt Rungego, spowodowany zbyt dużą liczbą węzłów. Zbyt mała liczba punktów węzłowych powoduje niedokładną interpolację.

Metoda wykorzystująca funkcje sklejane trzeciego stopnia wymaga tworzenia macierzy i rozwiązania układu równań. W projekcie nr 2 zaprezentowano, czas rozwiązania układu równań metodą faktoryzacji LU wzrasta wraz z liczbą niewiadomych. Aby wyznaczyć funkcję interpolującą dokładniej, należałyby zwiększyć liczbę węzłów, a tym samym zwiększyć liczbę przedziałów i równań.

Do interpolacji Splajnami można użyć losowo wybrane punkty węzlowe i przy ich odpowiedniej liczbie otrzymać dość dokładne wyniki, w przeciwieństwie do metody Lagrange'a. Błąd interpolacji okazał się znacznie mniejszy dla drugiej metody w każdym przypadku. Można stwierdzić, że metoda wykorzystująca funkcje sklejane trzeciego stopnia jest lepsza niż metoda Lagrange'a.

Metody mogą okazać się przydatne szczególnie dla tras o stałym charakterze, ponieważ można w ten sposób uzyskać dość dokładne wyniki. Metoda Lagrange'a jest przydatna w środkowej części trasy, natomiast na krawędziach należałyby zastosować inny sposób. W przypadku drugiej metody można uzyskać bardzo dokładne wyniki poprzez dobranie odpowiedniej liczby punktów (w tym wypadku ok. 50 punktów).