

Spektroskopia molekularna NMR

Jakub Giezgala

11 czerwca 2025

1 Wprowadzenie

1.1 NMR

NMR (Nuclear Magnetic Resonance) to technika analityczna wykorzystująca właściwości jąder atomowych (np. 1H , ^{13}C) umieszczonych w silnym polu magnetycznym. Jądra atomowe absorbują określone częstotliwości fal radiowych, co prowadzi do emisji charakterystycznych sygnałów związanych z ich środowiskiem chemicznym. Pozwala to na identyfikację struktury cząsteczki i ilości poprzez integrację sygnałów. Technika ta wykorzystuje fakt, że jądra atomów poddane działaniu fal radiowych, zachowują się jak magnesy. W polu magnetycznym ulegają specyficznej orientacji, a po wzbudzeniu emitują sygnał o częstotliwości zależnej od panujących warunków.

Częstotliwość absorpcji w obserwowanym widmie NMR zależy bezpośrednio od wartości indukcji pola magnetycznego w punkcie, w którym znajduje się jądro. Wartość ta jednak nie jest dokładnie równa wartości pola zewnętrznego, lecz jest zmodyfikowana przez orbitalny ruch elektronów (które wytwarzają pole magnetyczne), a także w przypadku układów otwartopowłokowych, przez moment magnetyczny związany ze spinem. W cieczech i gazach, gdzie występuje swobodna rotacja cząstek (skala czasowa eksperymentu NMR jest długa i eksperyment 'widzi' wszystkie wielkości, ale uśrednione po wszystkich 3 współrzędnych) współrzędnych szybszych ruchów) efektywne pole magnetyczne się uśrednia.

Na wartości pola magnetycznego nie wpływa tylko pole magnetyczne pochodzące od elektronów, ale także generowane przez inne jądra. Wynika z tego fakt, że niezależnie od orientacji spinów tych jąder, pole to możemy dodawać lub odejmować od pola zewnętrznego. Konsekwencją tego jest różna częstotliwość absorbowanego promieniowania. W obserwowanym widmie przejawia się to rozczepieniem sygnałów - tam gdzie spodziewamy się jednego, występuje kilka (*multiplet*). Najbardziej charakterystycznym przypadkiem jest dublet, kiedy to w widmie spin pojedynczego protonu z jednakowym prawdopodobieństwem wynosi $+1$ lub -1 .

To co jednak w praktyce mierzymy, to anizotropia, czyli zależność właściwości fizycznych lub mechanicznych materiału od kierunku, w którym są mierzone. Uśrednia się ona jednak do zera w cieczach i gazach pod wpływem swobodnej rotacji cząsteczki. Aby zidentyfikować cząstkę w praktyce mierzymy **przesunięcie chemiczne** - δ . Ze względu na bardzo małą miarę przyjęło się stosować pseudojednostki *ppm* (part per million). Nie mierzymy tego jednak bezpośrednio, a względnie, poprzez wykorzystanie sygnału wzorca. Do pomiarów przesunięcia chemicznego można stosować zarówno wzorce wewnętrzne, dodawane do roztworu mierzonej substancji, jak i zewnętrzne (kapilare z próbką umieszcza się w większej kapilarze, zawierającej wzorzec).

1.2 Metabolomika

Metabolomika natomiast jest dziedziną nauki zajmującą się całościowym pomiarem i analizą metabolitów – małych cząsteczek powstających w organizmach (zwyczajowo $\leq 1500\text{DA}$). Celem jest identyfikacja i pomiar stężeń w końcowych szlakach biochemicznych różnych próbek biologicznych. Taka analiza ilościowa, jak i jakościowa jest szeroko stosowana w odkrywaniu biomarkerów chorób, badaniu wpływu diety i czynników zewnętrznych na metabolizm, jak i w prawidłowym monitorowaniu odpowiedzi na leczenie.

Z tego powodu technika NMR jest szeroko stosowana w metabolomice, w szczególności ze względu na:

- Możliwość identyfikacji nowych metabolitów i śledzenia szlaków metabolicznych (np. przy użyciu izotopów). Technika jest nieinwazyjna, wykrywa nieznane związki, dzięki strukturalnej charakterystyce nowych metabolitów.
- Liniową odpowiedź sygnału względem stężenia. Dodatkowo wysoki poziom powtarzalności oraz dokładna ilościowość.
- Brak potrzeby oznaczania próbek i nie niszczący ich pomiar
- Wymagania minimalnej preparatyki, dzięki czemu możemy analizować mocz bez zbędnej ekstrakcji.
- Powtarzalność - wysoka precyzja między laboratoriami, która umożliwia wielośrodkowe badania kliniczne.

1.3 Cele

W tym projekcie zajmiemy się analizą danych pochodzących z badania, mającego na celu sprawdzenie jak zmienia się poziom metabolitów w moczu pochodzących od 6 wolontariuszy przebywających w ustandaryzowanym i stabilnym środowisku. Wszyscy wolontariusze dzielili ten sam rytm dobowy, dietę oraz byli wystawieni na działanie tych samych czynników zewnętrznych.

Hipoteza badawcza zakłada, że im dłużej wolontariusze będą przebywać w tym samym środowisku, tym bardziej podobny będzie skład metabolitów w ich moczu. Oznaczałoby to, że to środowisko zewnętrzne ma większy wpływ na nasz metabolizm niż różnice osobnicze. W tym celu dokonaliśmy analizy próbek moczu przy pomocy spektroskopii 1H NMR. Próbki były zbierane codziennie na przestrzeni 254 dni trwania eksperymentu oraz w dniach 713-716 od początku eksperymentu. Dodatkowo zmieszano ze sobą wszystkie próbki zebrane podczas eksperymentu oraz dokonano 32 pomiarów tak stworzonej próbki kontrolnej (**QC**). W naszych analizach skupimy się na analizie 29 widm NMR próbek pochodzących od jednego z wolontariuszy oraz jednego widma próbki QC z powodu ograniczeń w zasobach obliczeniowych.

Celami naszej analizy są ponadto próba oszacowania proporcji wybranych metabolitów w próbkach eksperymentalnych i kontroli. Monitorowanie zmian ich stężeń podczas przebiegu całego eksperymentu oraz ocena, czy analizowane warunki wpływają ostatecznie na metabolomikę naszych próbek. Dodatkowo znajdziemy odpowiednie hiperparametry oraz wybierzemy solver do uzyskania optymalnego rozwiązania. Na koniec wprowadzimy delikatne modyfikacje w naszej bazie referencyjnych metabolitów, aby zobaczyć, jaki wpływ ma to na naszą analizę. Cała praca została wykonana, dzięki narzędziu Magnestein [3].

2 Metody

2.1 Widma eksperymentalne

Widma eksperymentalne zostały uzyskane techniką NMR i stanowiły podstawę dokonanej analizy metabolicznej. W praktyce dane pochodziły wyłącznie od jednego wolontariusza, pobierane około co dziesięć dni, oraz w dniach 713-716 od początku trwania eksperymentu. W celu usprawnienia obliczeń wykorzystujemy ponadto jedno widmo próbki Quality Control, które powstało w wyniku uśrednienia wszystkich próbek. Widma znajdują się w plikach z rozszerzeniem *csv*, zawierających pozycje pików w *ppm* oraz ich intensywność. W ramach preprocessingu danych, usunięto z nich ujemne natężenia, które świadczą o błędzie pomiaru, oraz znormalizowano je w celu pozbycia się zmienności wewnątrz pomiarów. Wykonane kroki są standardową analizą, która powoduje, że widma pochodzące z różnych eksperymentów są gotowe do porównania między sobą.

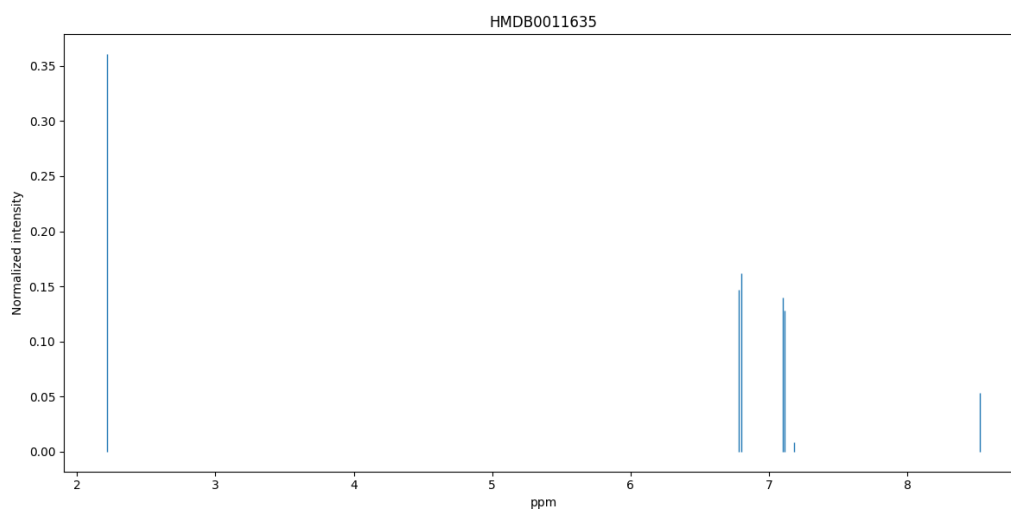
2.2 Biblioteka referencyjna

W celu przeprowadzenia analiz wykorzystano dwie biblioteki referencyjne. Jedną zaproponowaną *a priori*, obejmującą 35 metabolitów oraz drugą, która stanowi autorską modyfikację oryginalnej. Niewielka zmiana, jaką jest dodanie 5 nowych związków chemicznych, pozwala na zweryfikowanie wyników oraz pokaże, jaki jest wpływ minimalnych zmian na skład metabolitów w moczu. Dokładne listy metabolitów znajdują się w pliku `configuration.py`

Widma pobrano z bazy **HMDB** (Human Metabolome Database) oraz sparsowano przy pomocy odpowiednich funkcji zawartych w pliku `data_processing.py`. Jest to otwarta baza danych, zawierająca około 220 tysięcy ludzkich metabolitów oraz 9000 widm NMR. Ponadto zawiera bogate dane chemiczne, fizyczne, kliniczne i biochemiczne (np. stężenia w płynach, reakcje, powiązane białka) oraz zintegrowana jest z wieloma innymi bazami, takimi jak KEGG.

Prawidłowo sparsowano 216138 plików XML, zawierających widma, z czego 129540 dotyczy widm 1H oraz 128735 widm ^{13}C o ustalonych częstotliwościach. Wśród tych widm jest dokładnie 2906 uzyskanych eksperymentalnie, z czego 1680 dotyczy metabolitów domyślnie wykrytych w moczu. Wśród nich 924 jest wykrytych i nieskwantyfikowanych, a 756 skwantyfikowanych. To właśnie z tej puli wybrano 26 o największej możliwej częstotliwości (500 oraz 600 MHz), odpowiadających zaproponowanym metabolitom [Rysunek 1], dla pozostałych nie znaleziono widm znalezionych eksperymentalnie. Następnie spośród tych widm wybrano 5 nowych związków, które dodano do naszej oryginalnej biblioteki referencyjnej.

Analizując próbki moczu, zdecydowano się na wykorzystanie widm 1H1D , które są znacznie czulsze od widm ^{13}C . Dodatkowo metabolity w moczu to małe cząsteczki z wieloma atomami wodoru (np. mleczan, cytrynian, kreatynina, która znajduje się w obu bibliotekach referencyjnych). Warto dodać, że tego typu jest także więcej widm stwierdzonych eksperymentalnie, ze względu na szybkość oraz zdecydowanie mniejszy koszt.



Rysunek 1: Przykładowe referencyjne widmo - Siarczanu p-krezolu

2.3 Wybór solvera

W przeprowadzonych analizach na początku porównano wyniki dwóch różnych solverów, zalecanego Gurobi oraz podstawowego solvera do rozwiązywania problemów liniowych z biblioteki *pulp* - *LpSolverDefault*. Alternatywą byłby inny opensourcowy solver do rozwiązywania problemów liniowych, jak na przykład Mosek. Zwracane wyniki były jednak za każdym razem identyczne z dokładnością do precyzji obliczeń [Tabela 1]. Uwzględniając jednak ostatnio problemy, jakie miał solver Gurobi z normalizacją danych, zdecydowano się kosztem wydłużenia obliczeń na domyślny solver [Rysunek 2].

W celu znalezienia najlepszych parametrów zdecydowano się na grid searcha dla następnych parametrów $\kappa_{mixture} \in \{0.20; 0.25; 0.30\}$ oraz $\kappa_{compounds} \in \{0.20; 0.22; 0.25\}$. Pomimo uzyskiwania porównywalnych wyników, można było stwierdzić, że wraz ze wzrostem pierwszego parametru i spadkiem drugiego, rośnie proporcja sygnału, a tym samym spada odległość Wassersteina [1]. Pierwszy parametr odpowiada za maksymalny transport dla analizowanej mieszaniny. Podczas estymacji proporcji sygnał będzie przenoszony maksymalnie na taki właśnie dystans. W modelowaniu złożonych sygnałów (np. sprzężeń wielopikowych), może to być interpretowane jako kara za usunięcie szumu z mieszaniny, czyli sygnału niepasującego do naszych referencji. Drugi parametr odpowiada natomiast za transport każdej z referencji osobno. Jest to zatem kara za usuwanie sygnału poszczególnych metabolitów, będących częścią biblioteki referencyjnej. Wszystkie analizy dokonano dla $\kappa_{mixture} = 0.3$ oraz $\kappa_{compounds} = 0.2$ [Tabela 2].

3 Wyniki

3.1 Wybór solvera

W celu znalezienia optymalnego solvera do dokonywanych przez nas analiz, zdecydowano się porównać dla domyślnych parametrów wyniki dla próbki Quality Control (QC). Różnice w uzyskanych wynikach są minimalne i są prawdopodobnie kwestią błędów zaokrągleń.

Solver	Signal Proportion	Noise proportion	Total explanation
Gurobi	0.951 091 132	0.048 908 867 1	0.999 999
LpSolverDefault	0.951 091 136	0.048 908 867 0	1.00

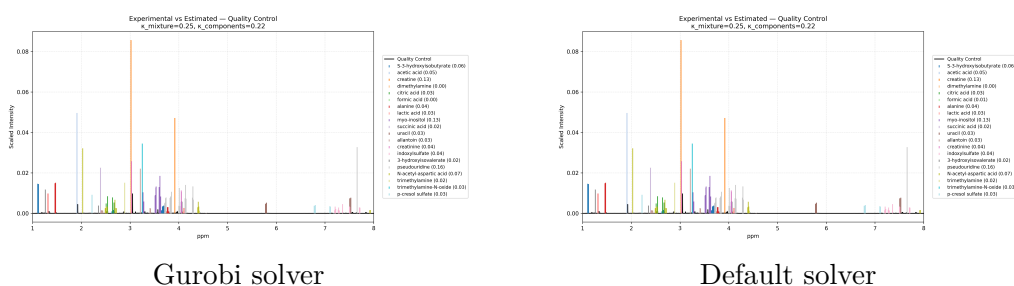
Tabela 1: Wyniki różnych solverów dla $\kappa_{mixture} = 0.25$ oraz $\kappa_{compounds} = 0.22$ dla próbki QC.

Na tej podstawie oraz doświadczeń z laboratorium, zdecydowano się na wybór domyślnego solvera biblioteki. *pulp*. Następnie wykonano poszukiwanie optymalnych hiperparametrów w celu ustalenia optymalnych $\kappa_{mixture}$ oraz $\kappa_{compounds}$.

$\kappa_{mixture}$	$\kappa_{compounds}$	Signal Proportion	Noise Proportion
0.30	0.25	0.953	0.047
0.30	0.22	0.955	0.045
0.30	0.20	0.958	0.042
0.25	0.25	0.949	0.042
0.25	0.22	0.951	0.051
0.25	0.20	0.952	0.049
0.20	0.25	0.945	0.055
0.20	0.22	0.946	0.054
0.20	0.20	0.948	0.052

Tabela 2: Porównywanie różnych parametrów dla próbki QC.

Dalsze analizy wykonano domyślnym solverem o zaznaczonych parametrach. Można zaobserwować, że wraz ze wzrostem $\kappa_{mixture}$ oraz ze spadkiem $\kappa_{compounds}$, zmniejsza się odległość Wassersteina, tym samym zwiększa się wyjaśnialność naszych widm. Szczegółowe wizualizacje znajdują się w folderach *solver_results* oraz *kappa_comparison*.

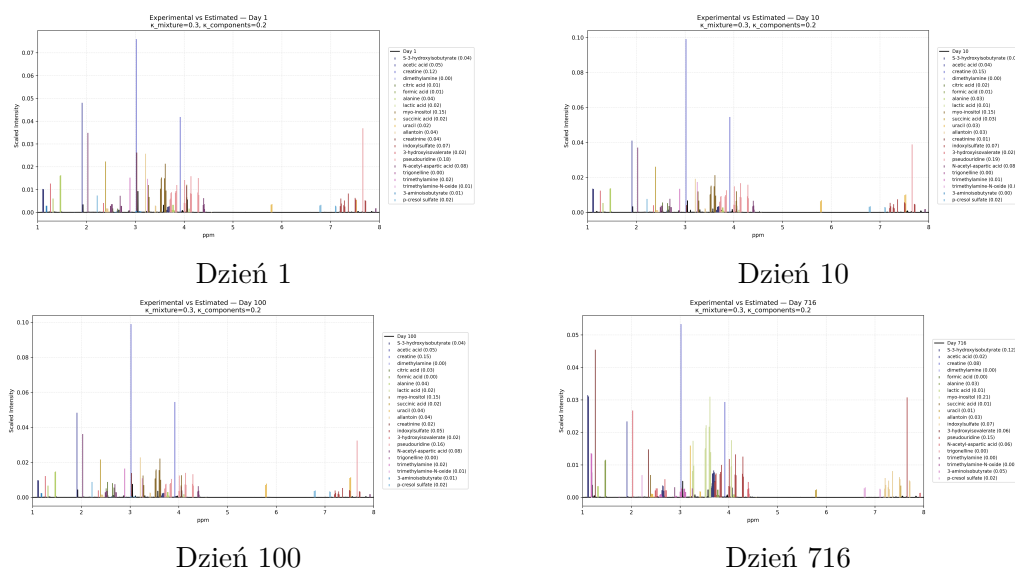


Rysunek 2: Porównanie stosunków metabolitów uzyskane przez różne solvery. Różnice występują ale są marginalne.

3.2 Proporcje metabolitów

We wszystkich eksperymentach użyto dokładnie tych samych (wcześniej wyznaczonych) parametrów oraz opcji porównywania **area**. Oznacza to, że uzyskane proporcje metabolitów są proporcjonalne do zajmowanego pola pod wykresem. Dodatkowo funkcja zwraca informacje o szumie w analizowanym eksperymencie oraz danych referencyjnych, oprócz proporcji poszczególnych metabolitów.

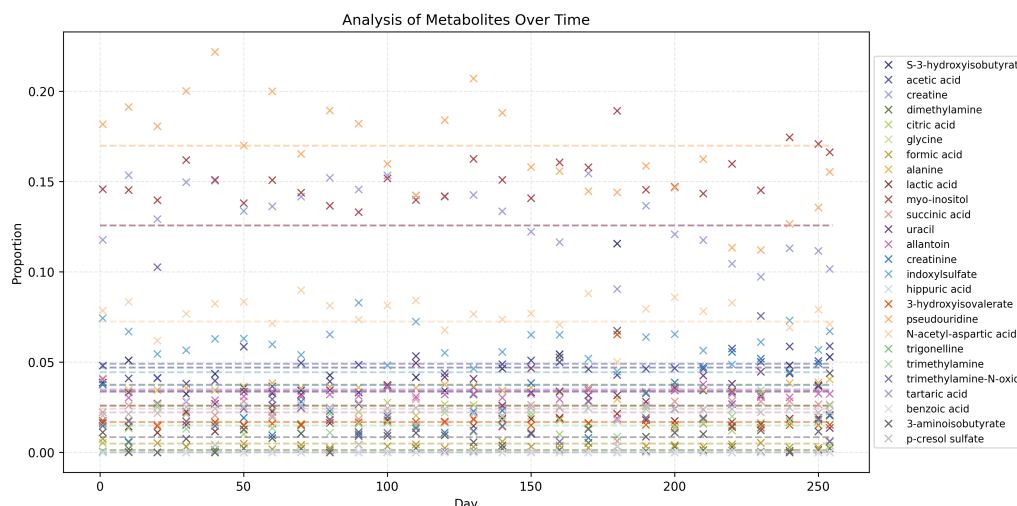
W naszej analizie wykorzystano pierwszą domyślną bibliotekę referencyjną, która zawierała 26 metabolitów, dla których były dostępne widma eksperymentalne o wysokiej częstotliwości. Do wizualizacji wybrano następujące dni: 1, 10, 100, 716 [Rysunek 3].



Rysunek 3: Porównanie ze sobą stosunków metabolitów z pierwszej biblioteki referencyjnej dla wybranych dni.

Dla większości dni możemy zaobserwować charakterystyczne piki w okolicy 2; 3; 4 i 7,8 ppm. Gwałtowna zmiana następuje podczas przerwy eksperymentu, szczególnie wyraźnie widać to dnia 716, kiedy pojawiają się piki substancji zupełnie wcześniej tam niewidzianych. Potwierdza to hipotezę, że skład metabolitów naszego moczu w czasie może się zmieniać. W celu dokładniejszej analizy, widma obcięto od 0-8 ppm.

Wykres przedstawiający jednak stosunek metabolitów w czasie od 1 do 255 dnia naszego eksperymentu, nie pozwala jednak wyciągnąć żadnych ścisłych wniosków [Rysunek 4]. Co prawda widać trend, że związki w czasie się coraz bardziej różnią i przestają się na siebie nakładać, ale nie sposób zaobserwować zbliżania się sygnałów do próbki QC w czasie. Dokładne dane można znaleźć w folderze *timecourse_results*.

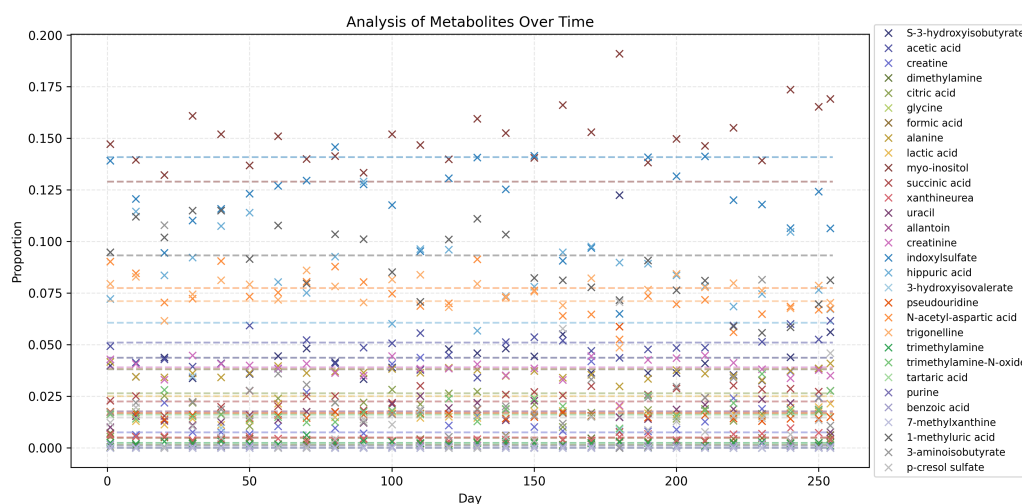


Rysunek 4: Stosunek wszystkich metabolitów dla wybranego przedziału czasowego. Linia przerywaną zaznaczono stosunek z próbki QC.

3.3 Nowa biblioteka referencyjna

Ze względu na to, że nasze dane pochodzą tylko od jednego pacjenta, zdecydowano się na niewielkie modyfikacje oryginalnej bazy referencyjnej. Rozszerzono ją o następujące metabolity:

- Urea (mocznik) - główny marker funkcji nerek. Jest końcowym produktem metabolizmu białek. Jego stężenie w moczu odzwierciedla wydolność nerek w filtrowaniu i usuwaniu zbędnych produktów przemiany materii.
- xanthinuurea - powstaje podczas degradacji ksantyny, związku pośredniego w metabolizmie puryn. Jego obecność w moczu, nawet śladowa, może wskazywać na ksanturię, czyli chorobę genetyczną prowadzącą do kamicy nerkowej.
- puryny - podstawowy prekursor kwasu moczowego. Puryny (np. adenina, guanina) pochodzą z rozpadu kwasów nukleinowych. Ich brak w moczu jest stanem fizjologicznym.



Rysunek 6: Stosunek wszystkich metabolitów dla wybranego przedziału czasowego dla zmodyfikowanej biblioteki referencyjnej.

Warto jednak odnotować, że dodanie w zasadzie tylko 3 nowych metabolitów, które zostały wykryte w naszych próbkach, znacznie poprawiło predykcję dla widma QC. Proporcja sygnału wzrosła z 0.951 do 0.966, co wskazuje, że kluczowy jest wybór odpowiedniej biblioteki referencyjnej. Dopiero na niej estymowanie najbardziej optymalnych hiperparametrów. Szczegółowe wyniki można odnaleźć w odpowiednich plikach `timecourse_data.csv`.

4 Dyskusja

Uzyskane przez nas wyniki nie są ani jednoznaczne, ani satysfakcjonujące. Z jednej strony uzyskaliśmy widma, które w dobry sposób odpowiadają naszym próbkom (żadna nie odbiega w negatywny sposób od normy), z drugiej skoro dodanie zaledwie dwóch metabolitów polepszyło wyniki, to kluczowy problem pojawia się przy wyborze biblioteki referencyjnej. Z jednej strony powinna zawierać związki chemiczne, których spodziewamy się w dużej ilości oraz takich, których brak świadczy o stanie fizjologicznym wolontariusza.

Dodatkowo analizowane przez nas próbki pochodzą tylko od jednej osoby, co może zaburzać sens naszej analizy. Z niej wynika, że czynniki egzogenne nie są jedynymi, które wpływają na skład naszego moczu. Wpływ czynników endogennych, takich jak funkcjonowanie nerek, produkcja hormonów, czy metabolizm, również może w znaczny sposób na niego oddziaływać.

Najważniejszym krokiem wydaje się zatem, przemyślany wybór biblioteki referencyjnej oraz widm o wysokiej częstotliwości, zapewniających poprawną analizę. W przyszłości można także zastanowić się nad ujednoczeniem wizualizacji w celu wykrywania różnic w stężeniach metabolitów. Na stan dzisiejszy, praktycznie wszystkie związki chemiczne zmieniają się skokowo w czasie. Jednak wykorzystując testy statystyczne można by zobaczyć, jakie zmiany są, a jakie nie statystycznie istotne. Aby jednak tego dokonać, musielibyśmy wielokrotnie powtórzyć cały proces badawczy, najlepiej na wielu próbkach, pochodzących od różnych pacjentów. Warte rozważenia wydaje się także uwspólnienie normalizacji, co może pomóc w porównywaniu natężeń sygnałów.

Literatura

- [1] S. Majewski, M. A. Ciach, M. Startek, W. Niemyska, B. Miasojedow, A. Gambin, "Wasserstein Distance as a Dissimilarity Measure for Mass Spectra with Application to Spectral Deconvolution", *WABI*, 2018.
- [2] M. A. Ciach, B. Miasojedow, G. Skoraczyński, S. Majewski, M. Startek, D. Valkenborg, A. Gambin, "Masserstein: Linear regression of mass spectra by optimal transport", *Rapid Communications in Mass Spectrometry* e8956, 2020.
- [3] B. Domżał, E. K. Nawrocka, D. Gołowicz; M. A. Ciach, B. Miasojedow, K. Kazimierzczuk, A. Gambin, "Magnetstein: An Open-Source Tool for Quantitative NMR Mixture Analysis Robust to Low Resolution, Distorted Lineshapes, and Peak Shifts", *Analytical Chemistry* 96, 188-196, 2023.