

Prof. dr hab. Małgorzata Jakubowska

Email: jakubows@agh.edu.pl

tel. 12 617 33 83 lub 507 825 755

Kraków 15.01.2024 r.

RECENZJA

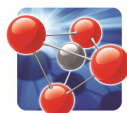
rozprawy doktorskiej mgr Andrzeja J. Kałki pt. „Development and Implementation of Computer-Aided Techniques of Data Modeling for Processing and Interpretation of Multicomponent Electronic Spectra Acquired for the Selected Organic Mixtures”,

zrealizowanej na Wydziale Chemii Uniwersytetu Jagiellońskiego

pod kierunkiem Promotora prof. dr hab. Andrzeja M. Turka

oraz Promotora pomocniczego dr hab. Mateusza Z. Breli

Problem analizy nakładających się składowych sygnałów jest od wielu dekad przedmiotem ciągłego zainteresowania w wielu dziedzinach nauki i także aktualnym zadaniem badawczym. Zagadnienie to formułowane jest na różne sposoby, a sprowadza się do uświadomienia, że wszystko, co obserwowalne lub mierzalne w rzeczywistym systemie, rzadko ma związek z jedną przyczyną lub źródłem, ale często jest fuzją prostszych składników. Różnice w zebranej odpowiedzi w funkcji czasu, przestrzeni, próbki lub różnych warunków np. eksperymentalnych, środowiskowych czy klinicznych stanowią globalny wyraz wkładu szeregu indywidualnych źródeł zmienności. Zrozumienie charakteru i udziału tych składowych na globalną odpowiedź jest niezbędne do zrozumienia badanego systemu. Separacja składowych sygnałów o nieznannej liczbie źródeł oraz bez dostępnej informacji o procesie sumowania, które nakładają się na siebie w dziedzinie czasu lub częstotliwości jest ważnym krokiem w badaniu przebiegów pomiarowych, których indywidualna rejestracja nie jest łatwa lub możliwa do przeprowadzenia. Celem działań podejmowanych w kontekście problemu separacji jest odtworzenie oryginalnych składowych z sygnału mieszaniny, a proponowane rozwiązania zazwyczaj zależą od informacji zawartych w źródłach i sposobu sumowania.



Istotnym problemem jest sposób sformułowania modelu mieszaniny. Gdy odpowiedź jest pomiarem instrumentalnym, wkład poszczególnych sygnałów można często opisać za pomocą modelu addytywnego. Jeżeli odpowiedź jest bardziej złożona, można mieć wątpliwości co do przydatności tak prostego podejścia. Jednakże w przypadku braku jasnej wiedzy na temat zależności pomiędzy udziałem poszczególnych składowych, strategia optymalizacji rozwiązania uzasadnia użycie możliwie najprostszego modelu do wyjaśnienia zjawiska. Dlatego też metody wielowymiarowe w separacji nakładających się składowych sygnałów rozwiązują problem analizy mieszanin najczęściej przy użyciu modelu dwuliniowego lub w przypadku danych wieloczynnikowych - jego wieloliniowego rozszerzenia. Zakres problemu analizy mieszaniny wiąże się także z różnorodnością pojęcia składnika. Ogólnie rzecz biorąc, komponentem jest dowolna jednostka, która może wykazywać odrębny profil jakościowy i ilościowy, który pozostaje niezmienny i ważny do opisu informacji. Ten odrębny profil można powiązać z prostymi sygnałami, np. czystym widmem związku chemicznego lub z bardziej złożonymi obiektami, np. widmowym odciskiem palca tkanki biologicznej lub profilem źródła zanieczyszczeń.

Brak jednoznacznych rozwiązań w stosowanych metodologiach obliczeniowych sprawia, że różne dwuliniowe lub wieloliniowe rozkłady tego samego zestawu danych z profilami zgodnymi z wcześniej wybranymi ograniczeniami odtwarzają oryginalny zestaw danych z optymalnym dopasowaniem. Te równoważne rozkłady zapewniają zakres możliwych rozwiązań, z których wszystkie są równie ważne w opisie informacji. W kontekście szerokiego spektrum problemów separacji składowych sygnałów akustycznych, biomedycznych, astronomicznych, chemicznych należy zauważyć, że jednym z aktualnych zagadnień jest szerokie spojrzenie na zagadnienie separacji składowych sygnałów na podstawie wielowymiarowych danych pozyskanych w eksperymentach spektroskopowych, podjęte przez Doktoranta. Rozwiązanie tego zadania jest interesujące ze względu na wiele praktycznych aplikacji w odniesieniu do badania mieszanin substancji organicznych z wykorzystaniem technik instrumentalnych. Z wymienionych powodów wybór tematyki badań, która łączy aspekt zaawansowanej analizy danych pozyskanych w pomiarach oraz potrzebę rozwoju wiarygodnych strategii interpretacyjnych dla spektroskopii należy uznać za w pełni uzasadniony. Działania prowadzone przez Doktoranta wskazują na szerokie możliwości wykorzystania zaproponowanej metodologii do rozwiązywania zadań z zakresu numerycznej separacji nakładających się składowych sygnałów, ze szczególnym naciskiem na fakt, że dane opisujące problem zostały pozyskane w różnych eksperymentach spektroskopowych.

Omówienie rozprawy doktorskiej

Obszerna rozprawa doktorska, zredagowana w języku angielskim, obejmująca 172 strony tekstu, w tym wiele rysunków, wykresów i tabel, została przygotowana w typowym układzie, z podziałem na część literaturową i badawczą. Opracowanie zawiera krótkie streszczenia w języku polskim i angielskim, a także opis dorobku naukowego Doktoranta, bibliografię, opis dostępnych dla czytelnika materiałów w wersji elektronicznej (kod zaimplementowanych przez Doktoranta algorytmów), załącznik zawierający dodatkowe wykresy, schematy i tabele oraz bibliografię, a także indeks rysunków, tabel i wzorów. Tekst rozprawy nie zawiera natomiast wydruku siedmiu artykułów Doktoranta, które ukazały się

w czasopiśmie o wysokim współczynniku oddziaływania (*Journal of Fluorescence*, *Spectrochimica Acta A*, *Physical Chemistry Chemical Physics*, *Journal of Chemical Physics*, *Applied Spectroscopy*). Tematyka tych prac, oznaczonych w tekście jako [A1]-[A7] jest bezpośrednio powiązana z zagadnieniami rozprawy, a w każdej z nich Doktorant jest pierwszym autorem. Układ rozprawy doktorskiej jest w pełni akceptowalny. Cele rozprawy zostały jasno sformułowane i zamieszczone w osobnym rozdziale.

Główną osią wstępu literaturowego, opartego na 166 źródłach są zagadnienia modelowania z wykorzystaniem procedur przetwarzania sygnałów (stosowane w pracy jako metody wstępnego przygotowania danych) oraz podejścia wielowymiarowego, dedykowanego do separacji nakładających się składowych poprzez implementację takich algorytmów jak: *Multivariate Curve Resolution*, *Self-Modeling Curve Resolution*, *Orthogonal Projection Approach*, *Rank Annihilation Method* czy *Parallel Factor Analysis*. Autor zaproponował ujednoczone oznaczenia struktur danych o różnym wymiarze (od wielkości skalarnych do 3D) oraz zamieścił przypomnienie podstaw algebry liniowej w zakresie pojęć i twierdzeń, wykorzystywanych w tej pracy. W osobnym rozdziale przedstawił także podstawowe koncepcje teorii funkcjonału gęstości DFT i opisał praktyczny aspekt wykorzystania symulacji kwantowo-chemicznych w interpretacji zebranych widm oraz ocenie wiarygodności uzyskanych wyników.

Prezentując zakres i efekty swoich badań Doktorant przedstawił spójną koncepcję rozwiązywania zadań, w których zgromadzone w macierzach widma spektroskopowe, pozyskane w ukierunkowanych eksperymentach stanowią dane wejściowe do algorytmów separacji składowych sygnałów i w kolejnym etapie pozwalają na rozwiązanie różnych sformułowanych zadań. Pierwszy etap tych działań to wytypowanie mieszaniny substancji organicznych, której składniki nie mogą być w łatwy sposób odseparowane, np. metodami fizycznymi lub chemicznymi. W celu uzyskania początkowej wiedzy na temat rozważanych systemów, badania kontynuował poprzez wszechstronne studium teoretyczne prowadzone zgodnie z obliczeniami DFT. Dalej kolejnym krokiem, realizowanym w celu zapewnienia wymaganej różnorodności danych i jednocześnie poprawnego działania algorytmów separacji było określenie czynnika zmienności, a tym czynnikiem może przykładowo być zmiana długości fali promieniowania, modyfikacja indywidualnych wkładów izomerów do sygnału wynikowego poprzez zmianę temperatury czy zastosowanie metody wygaszania fluorescencji. Wymienione przykładowe czynniki mają wpływ na badane obiekty i poprzez zaprogramowane oddziaływanie pozwalają na eksperymentalną obserwację odpowiedzi w formie widm spektroskopowych, które stanowią podstawę obliczeń chemometrycznych. Dalej następował zasadniczy etap, czyli separacja nakładających się składowych sygnałów z wykorzystaniem algorytmów chemometrycznych, poprzedzona niezbędnym przygotowaniem danych z zastosowaniem procedur przetwarzania sygnałów (przykładowo Autor stosował korektę linii bazowej, dekompozycje SVD oraz filtrację wektorów własnych po transformacji do dziedziny częstotliwościowej z wykorzystaniem FFT i potem transformacji odwrotnej). W kolejnym kroku analizował w szerokim aspekcie uzyskane efekty w kontekście rozwiązywanego problemu i stosował różne podejścia w zakresie weryfikacji wyników.

Zgodnie z określonym tytułem i zakresem pracy głównym zagadnieniem, które zainteresowało Doktoranta jest opracowanie, implementacja i wykorzystanie analizy wielowymiarowej do separacji nakładających się składowych sygnałów, z uwzględnieniem

etapu wstępnego przetwarzania danych. Należy zwrócić uwagę, że jest to zadanie o wysokim stopniu skomplikowania przede wszystkim dlatego, że widma poszczególnych składników mieszaniny mają złożony kształt. Autor w sposób wielokierunkowy i rzetelny poszukiwał wiarygodnych rozwiązań stosując metody eksperymentalne, symulację DFT, numeryczne przetwarzanie sygnałów oraz analizę wielowymiarową. W problemie dekompozycji widm dwuskładnikowego związku t-2StA na składowe pochodzące od dwóch izomerów konformacyjnych *s-cis* oraz *s-trans*, zebranych m.in. metodą wygaszania fluorescencji oraz rejestracji map wzbudzeniowo-emisyjnych EEM, wykazał zgodność efektów separacji profili emisyjnych, wykonanej z wykorzystaniem algorytmu RAFA i PARAFAC i dalej podobnie dla widm absorpcyjnych mieszaniny izomerów. Udowodnił, że mimo braku ograniczających założeń podczas analizy, uzyskane wyniki doskonale wpisują się w związany z nimi model fizykochemiczny, co dodatkowo podkreśla wysoką skuteczność dwóch zastosowanych niedeterministycznych technik analizy czynnikowej. Zagadnienie to kontynuował rozpatrując problem termicznej ewolucji widm absorpcyjnych dwuskładnikowej mieszaniny t-2StA. Podczas chemometrycznej separacji składowych, wykonanej na podstawie macierzy termospektralnej zastosował algorytm samomodelowania SMCR oraz ponownie RAFA, a także nowe podejście oparte na rachunku zakłóceń, przeprowadzonej z wykorzystaniem pochodnych spektralnych. Zaproponował ciekawe podejście, które bazuje na opracowaniu i optymalizacji algorytmu początkowo dla substancji modelowych antracenu i jego pochodnej difenylowej. Poprzez przesuwanie profili i poszerzanie pasm z wykorzystaniem funkcji Gaussa lub Lorentza uzyskał niemal identyczne wynikowe widma, co potwierdziło skuteczną redukcję niespecyficznych efektów termospektralnych wpływających na sygnał.

Kolejny problem to badanie nieoczekiwane przebiegu sygnału, ujawnionego dla mieszaniny fumaronitrylu i toluenu. Doktorant udowodnił, że przyczyną obserwowanego efektu jest powstanie optycznie aktywnego adduktu składającego się z molekuł wymienionych związków. Przeprowadził obliczenia z wykorzystaniem DFT, umożliwiające spojrzenie na badany system i procesy zachodzące w nim na poziomie molekularnym. Dalej w analizie danych uzyskanych w specjalnie zaprojektowanych eksperymentach spektroskopowych opracował oryginalny algorytm wykorzystujący RAFA, umożliwiający ekstrakcję sygnału pochodzącego z ukrytego kompleksu typu *charge-transfer* w badanej mieszaninie, po zastosowaniu wielu etapów wstępnego przetwarzania sygnałów. Dalsze badania ujawniły jednak, że obserwowane zjawisko zdecydowanie nie jest ograniczone do izolowanego przypadku, mając zaskakująco szeroki zasięg. Mianowicie, fumaronitryl okazał się tworzyć podobne addukty z dość szeroką gamą związków chemicznych. Udowodnił, że w przetwarzaniu sygnałów podejście do modelowania wektorów własnych obliczonych poprzez zastosowanie SVD jest szczególnie użyteczne, ponieważ pozwala wygodnie stosować różne efektywne modyfikacje zebranych zestawów widm eksperymentalnych.

W trzecim projekcie głównym celem było badanie układu regioizomerów *n*-[*trans*-(2-naftylo)winylo]okazolu (nNVOx). Zagadnienie to Autor rozwiązywał poprzez przeprowadzenie połączonego studium teoretycznego, eksperymentalnego i interpretacyjnego, pozyskując poszukiwaną informację na temat badanych nowo zsyntetyzowanych systemów poprzez numerycznie wspomaganą analizę i interpretację ich spektrów UV-Vis. W tym kontekście Doktorant wykorzystał algorytm projekcji ortogonalnej OPA, który umożliwił wygodne wydobycie istotnych informacji z pozyskanych zestawów danych spektralnych tj. map EEM oraz widm wygaszanej fluorescencji, a także algorytmy

RAFA i PARAFAC, których zastosowanie po raz kolejny umożliwiło skuteczne odzyskanie pierwotnych, indywidualnych profili spektralnych w mieszaninie 4NVOx oraz 5NVOx. Główne osiągnięcie Doktoranta związane jest z opracowaniem podejścia RAFA-SVD-SM do rozdzielania trójskładnikowych mieszanin spektralnych, zastosowanej do separacji sygnałów mieszaniny trzech dominujących form konformacyjnych związku 2NVOx. Taka hybrydowa kombinacja technik okazała się bardzo skuteczna w rozdzielaniu wielowymiarowych widm, oferując alternatywę np. dla procedury PARAFAC.

Analizując zaproponowaną strategię rozwiązywania problemów z punktu widzenia uczenia maszynowego należy zaznaczyć, że zastosowane algorytmy należą do grupy procedur uczenia bez nadzoru. Podejście to nie posiada wbudowanego mechanizmu walidacji i oceny uzyskanych efektów modelowania, w przeciwieństwie do metod z nadzorem. Dlatego też istotna jest ocena wiarygodności i użyteczności uzyskanych efektów z wykorzystaniem wiedzy dziedzinowej. Doktorant podjął różne działania prowadzące do oceny wyników modelowania, projektując oraz implementując na każdym etapie obliczeń adekwatne schematy interpretacji i oceny. Jako przykład można podać wykorzystanie nowych, zaproponowanych strategii obliczeniowych dla sygnałów rejestrowanych dla substancji modelowych czy ocena jakości wyników za pomocą zgodności z modelami fizykochemicznymi z zakresu podejmowanej tematyki. Jest to jeden z trudniejszych etapów realizowanych projektów, ponieważ w analizie wielowymiarowej praktycznie nie ma możliwości intuicyjnej oceny wyników modelowania, a ten krok jest niezbędny do udowodnienia wiarygodności efektów separacji.

Ocena pracy

Przechodząc do oceny pracy chcę podkreślić, że jest to znakomita rozprawa naukowa, która posiada wiele aspektów nowości. Autor efektywnie wykorzystał aktualny stan wiedzy na temat rozwiązywanych problemów, a także dostępny warsztat badawczy. Chociaż rozprawa jest wielowątkowa, jednak wspólna platforma prezentująca różne aplikacje separacji nakładających się składowych sygnałów mieszanin substancji organicznych w oparciu o dane pozyskane w eksperymentach spektroskopowych łączy poszczególne projekty.

Jeżeli chodzi o sposób redakcji pracy to nie budzi on wątpliwości. Szczególnie należy zwrócić uwagę na przyjęte jednolite oznaczenia i opracowanie uniwersalnego sposobu prezentacji wykresów w zestandaryzowanej formie, a także bardzo podobną strukturę trzech rozdziałów 4, 5, i 6, odnoszących się do dokonań Doktoranta, z których każdy stanowi pewną odrębną całość. Istotne, że na końcu każdego z tych rozdziałów Doktorant zamieszcza podziękowania dla współautorów badań, określając precyzyjnie ich wkład. Na tym etapie mam pewną sugestię dotyczącą redakcji rozdziału, prezentującego podstawy algebry liniowej. Proponuję, aby w swoich Autor w swoich przyszłych działaniach taki tekst formułował w sposób typowy dla prac matematycznych, np. używając stwierdzeń: zdefiniowano jako ..., warunek konieczny, warunek wystarczający czy wyraźnie podawał założenia twierdzeń. Chociaż trzeba przyznać, że obecny sposób redakcji jest spójny z formą całej rozprawy.

Przechodząc do oceny całości dokonania naukowego, a w szczególności podjętych badań uważam, że należy zauważyć interdyscyplinarną aktywność Doktoranta oraz Jego obszerny dorobek naukowy – publikacyjny, udział w konferencjach oraz kierowanie dwoma grantami. Do istotnych osiągnięć badawczych Autora, które posiadają aspekt nowości można zaliczyć:

- wprowadzenie do praktyki analizy mieszanin związków organicznych na podstawie badań spektroskopowych różnych niedeterministycznych algorytmów separacji nakładających się składowych sygnałów tj. typowych procedur po optymalizacji oraz nowych strategii obliczeniowych, zaproponowanych w rozprawie
- wykazanie, że w różnych nowych problemach związanych z analizą mieszanin substancji organicznych efektywne rozwiązanie można znaleźć poszukując wsparcia w obliczeniach chemometrycznych; dodatkowo takie rozwiązania mogą inspirować do szerszego spojrzenia na problematykę, która była przedmiotem zainteresowania oraz prowadzić do sformułowania i rozwiązania różnych problemów pokrewnych
- ścisły, niebudzący wątpliwości i adekwatny do potrzeb sposób opisu działania poszczególnych algorytmów separacji nakładających się składowych sygnałów i innych procedur obliczeniowych z wykorzystaniem formalizmu teorii macierzy (zestaw widm reprezentowany jest jako macierz, w związku z tym zaproponowana prezentacja algorytmów ma uniwersalny charakter i może zostać wprost wykorzystana także w innych zagadnieniach)
- opracowanie i optymalizację adekwatnych metod wstępnego przygotowania danych, tj. widm spektroskopowych (filtracja, korekta linii bazowej, SVD)
- zwrócenie uwagi na potrzebę walidacji i oceny poprawności uzyskanych efektów analizy wielowymiarowej, np. poprzez porównywanie kształtu krzywych uzyskanych różnymi metodami lub ocenę zgodności z modelami dziedzinowymi
- wykorzystanie obliczeń kwantowo-chemicznych DFT do opisu problemu na poziomie jego formułowania oraz przygotowania do fazy obliczeń chemometrycznych tj. wykazanie, że nowoczesna chemia obliczeniowa może stanowić doskonałe wsparcie dla eksperymentalnej części badań, nadając odpowiedni kierunek podczas projektowania eksperymentów, interpretacji widm, ale także ocenę wiarygodności ostatecznych wyników
- implementację, optymalizację oraz udostępnienie w wersji elektronicznej kodu procedur obliczeniowych.

Po szczegółowym zapoznaniu się z rozprawą proponuję cztery tematy (pytania) do dalszej dyskusji podczas obrony.

1. Czy Doktorant uznaje za użyteczne podjęcie działań związanych z oceną wiarygodności uzyskanych efektów separacji nakładających się widm, np. za pomocą wskaźników numerycznych zgodności kształtu krzywych (np. za pomocą pierwiastka z błędu średniokwadratowego RMSE) czy zgodności z modelem wykorzystującym zależności fizykochemiczne (współczynnik korelacji, współczynnik determinacji)? Proszę o ogólny komentarz, dotyczący oceny wyników modelowania wielowymiarowego bez nadzoru (*unsupervised learning*) w kontekście złożonych problemów z zakresu badań spektroskopowych.

2. Okazuję się, że przy stosowaniu algorytmów separacji istotne jest wstępne przetwarzanie sygnałów, m.in. korekta linii bazowej. Proszę o informację, jak taka operacja była wykonywana.
3. Doktorant podjął ogromny wysiłek i uzyskał znakomite efekty, biorąc pod uwagę widma spektroskopowe. Czy zaproponowane strategie obliczeniowe mogą zostać wprost wykorzystane w badaniach z wykorzystaniem innych metod analizy instrumentalnej, gdzie informacja przekazywana jest za pomocą sygnałów?
4. Jaki zakres badań należy wykonać, aby zestaw danych przydatny do analizy wielowymiarowej miał wystarczający rozmiar? Na wykresach Autor zaprezentował zazwyczaj kilka/kilkanaście widm. Czy jest to pełny zestaw do obliczeń, czy wymagana jest rejestracja szerszego zbioru sygnałów eksperymentalnych?

Podsumowanie recenzji

Podsumowując stwierdzam, że zgodnie z wymogami ustawowymi oceniana rozprawa doktorska Pana mgr Andrzeja J. Kałki prezentuje ogólną wiedzę teoretyczną osoby ubiegającej się o nadanie stopnia doktora, stanowi oryginalne rozwiązanie problemu naukowego, a Doktorant posiada umiejętność samodzielnego prowadzenia pracy badawczej. Praca przedstawia dużą wartość pod względem poznawczym, wnosi elementy nowości w zakresie chemometrycznej separacji składowych sygnałów rejestrowanych w eksperymentach chemicznych. Reasumując uważam, że przedłożona do oceny rozprawa pt. *„Development and Implementation of Computer-Aided Techniques of Data Modeling for Processing and Interpretation of Multicomponent Electronic Spectra Acquired for the Selected Organic Mixtures”* spełnia wszystkie wymagania merytoryczne i kryteria formalne nakładane na prace doktorskie w ustawie z dnia 20 lipca 2018 roku - Prawo o szkolnictwie wyższym i nauce (Dz. U. z 2023 r. poz. 742 z późn. zm.). Z pełnym przekonaniem wnoszę do Rady Dyscypliny Nauk Chemicznych Uniwersytetu Jagiellońskiego wniosek o dopuszczenie Doktoranta do dalszych etapów przewodu doktorskiego.

Ponieważ ta znakomita rozprawa, o szerokim aspekcie badawczym, której efekty zostały już opublikowane w czasopiśmie o wysokim współczynniku oddziaływania, swoim zakresem istotnie przekracza typowe ramy prac doktorskich, zwracam się do Rady Dyscypliny Nauk Chemicznych Uniwersytetu Jagiellońskiego o jej wyróżnienie.

