

## **Załącznik 2:** Autoreferat w wersji polskiej

dr Ivana Stanimirova-Daszykowska

*Instytut Chemii*

Uniwersytet Śląski w Katowicach

ul. Szkolna 9

40-006 Katowice

tel.: (32) 359-15-08

e-mail: [istanimi@us.edu.pl](mailto:istanimi@us.edu.pl)

www: <https://sites.google.com/site/chemomlab/members/ivana-stanimirova>

*Zaawansowane strategie chemometryczne poprawiające ekstrakcję chemicznej informacji z wielowymiarowych niekompletnych danych i danych o indukowanej strukturze*

dr IVANA STANIMIROVA-DASZYKOWSKA

Autoreferat przedstawiający opis dorobku i osiągnięć naukowych  
w związku z ubieganiem się o nadanie stopnia doktora habilitowanego

## Spis treści

1. Imię i Nazwisko .....	3
2. Posiadane dyplomy, stopnie naukowe.....	3
3. Informacje o dotychczasowym zatrudnieniu w jednostkach naukowych .....	3
4. Dotychczasowy przebieg pracy naukowej .....	3
4.1 Przed uzyskaniem stopnia doktora .....	3
4.2 Po uzyskaniu stopnia doktora.....	4
5. Wskazanie osiągnięcia wynikającego z art. 16 ust. 2 ustawy .....	4
a) Tytuł osiągnięcia naukowego.....	4
b) Wykaz publikacji stanowiących osiągnięcie naukowe zgodnie z art. 16 ust. 2 ustawy <sup>1</sup> .....	4
c) Omówienie celu naukowego ww. prac i osiągniętych wyników wraz z omówieniem ich ewentualnego wykorzystania .....	6
6. Omówienie pozostałych osiągnięć naukowo-badawczych.....	27
6.1 Wykaz pozostałych artykułów .....	27
6.2 Wykaz rozdziałów książkowych .....	30
6.3 Lista wykładów na zaproszenie imienne wygłoszonych na konferencjach krajowych oraz zagranicznych.....	30
6.4 Lista wykładów wygłoszonych na konferencjach krajowych i zagranicznych .....	31
6.5 Lista plakatów na konferencjach krajowych i zagranicznych .....	32
6.6 Lista przeprowadzonych kursów z zakresu szeroko pojętej chemometrii.....	35
6.7 Udział w projektach badawczych o zasięgu krajowym i międzynarodowym.....	36
6.8 Składane aplikacje grantowe .....	36
6.9 Odbyte staże w zagranicznych ośrodkach naukowych lub akademickich .....	37
6.10 Opieka i promotorstwo prac dyplomowych .....	37
6.11 Uzyskane nagrody i wyróżnienia za działalność naukową .....	37
6.12 Pełnienie roli edytora w czasopismach z listy filadelfijskiej.....	38
7. Syntetyczne zestawienie dorobku naukowego habilitanta .....	39
8. Aktywność publikacyjna w poszczególnych latach i liczba cytowań .....	40

## 1. Imię i Nazwisko

IVANA STANIMIROVA-DASZYKOWSKA

(posługująca się w publikacjach nazwiskiem Stanimirova)

## 2. Posiadane dyplomy, stopnie naukowe

- Dyplom doktora nauk farmaceutycznych, Wydział Medyczny i Farmaceutyczny, Wolny Uniwersytet Brukselski, Bruksela, Belgia, 27.09.2005 r.

Tytuł pracy doktorskiej: „*Exploratory analysis of environmental data*”

Promotorzy pracy doktorskiej: prof. D.L. Massart, prof. Y. Vander Heyden, prof. B. Walczak

- Dyplom magistra chemii (jednolite pięcioletnie studia magisterskie, specjalność chemia nieorganiczna i analityczna), Wydział Chemii, Uniwersytet Sofijski im. św. K. Okhridskiego, Sofia, Bułgaria, 1997 r.

Tytuł pracy magisterskiej: „*Receptor modeling of coastal sediment data*”

Promotor pracy magisterskiej: prof. V. Simeonov

## 3. Informacje o dotychczasowym zatrudnieniu w jednostkach naukowych

- 1997 - 2001            Instytut Chemii Fizycznej, Bułgarska Akademia Nauk, Sofia, Bułgaria:  
*naukowiec*
- 2001 - 2005            Wydział Medyczny i Farmaceutyczny, Wolny Uniwersytet Brukselski,  
Bruksela, Belgia: *studia doktoranckie*
- 2005 - 2007            Wydział Matematyki, Fizyki i Chemii, Instytut Chemii, Uniwersytet Śląski  
w Katowicach: *adiunkt naukowy* (finansowanie wynagrodzenia w ramach  
europejskiego projektu TRACE „*TRAcing food Commodities in Europe*”,  
FOOD-CT-2005-006942)
- 2008 - nadal            Wydział Matematyki, Fizyki i Chemii, Instytut Chemii, Uniwersytet Śląski  
w Katowicach: *adiunkt naukowo-dydaktyczny*

## 4. Dotychczasowy przebieg pracy naukowej

### 4.1 Przed uzyskaniem stopnia doktora

Po ukończeniu studiów magisterskich podjęłam pracę na stanowisku naukowca w Instytucie Chemii Fizycznej Bułgarskiej Akademii Nauk w Sofii, Bułgaria. Wówczas, moje główne zainteresowania badawcze obejmowały zagadnienia z zakresu elektrochemii półprzewodników. Prowadzone przeze mnie prace badawcze obejmowały zagadnienia z zakresu elektrochemii półprzewodników, w tym konstruowaniu cienkich warstw (tzw. filmów) miedzi na półprzewodnikach krzemowych typu ‘n-’

i 'p-' oraz na badaniach mechanizmów deponowania miedzi w kontekście elektrochemicznej teorii krystalizacji.

W 2001 r. skorzystałam z unikalnej możliwości rozwijania swoich zainteresowań badawczych w dziedzinie chemometrii. Rozpoczęłam studia doktoranckie w jednej z najbardziej rozpoznawalnych grup chemometrycznych prowadzonej przez prof. D.L. Massarta na Wydziale Medycyny i Farmacji, Wolny Uniwersytet Brukselski, Bruksela, Belgia.

#### **4.2 Po uzyskaniu stopnia doktora**

Po uzyskaniu stopnia doktora nauk farmaceutycznych rozpoczęłam podoktorskie badania w Instytucie Chemii Uniwersytetu Śląskiego w Katowicach w ramach projektu europejskiego „TRACE” (*Tracing food commodities in Europe*, [www.trace.eu.org](http://www.trace.eu.org)). Moje badania w głównej mierze koncentrowały się na rozwijaniu strategii chemometrycznych z nadzorem, które można zastosować do identyfikacji autentyczności różnych produktów spożywczych w oparciu o informację o ich składzie chemicznym. Niektóre z proponowanych metodologii zostały następnie zmodyfikowane w celu zaproponowania procedur umożliwiających dyskryminację próbek win ze względu na szereg winogron oraz rok produkcji. Ponadto, prowadziłam również badania w obszarze nauk o środowisku proponując kilka strategii chemometrycznych rozwijanych w kontekście analizy danych uzyskanych na drodze sekwencyjnej procedury ekstrakcji czy procesu oczyszczania wody.

Najważniejsze osiągnięcie naukowe związane z opracowaniem metod chemometrycznych, które umożliwiają przetwarzanie niekompletnych wielowymiarowych danych analitycznych zawierających obiekty odległe i danych o indukowanej strukturze, opisałam w dalszej części opracowania i zbiorze artykułów.

### **5. Wskazanie osiągnięcia wynikającego z art. 16 ust. 2 ustawy<sup>1</sup>**

#### **a) Tytuł osiągnięcia naukowego**

*Zaawansowane strategie chemometryczne poprawiające ekstrakcję chemicznej informacji z wielowymiarowych niekompletnych danych i danych o indukowanej strukturze*

#### **b) Wykaz publikacji stanowiących osiągnięcie naukowe zgodnie z art. 16 ust. 2 ustawy<sup>1</sup>**

Lp.	Publikacja	IF	L. pkt. <sup>2</sup>
1.	I. Stanimirova, M. Daszykowski, B. Walczak, Dealing with missing values and outliers in principal component analysis, <i>Talanta</i> , 72 (2007) 172-178	3,374	40

<sup>1</sup> Ustawa z dnia 14 marca 2003 r. o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (Dz. U. 2016 r. poz. 882 ze zm. w Dz. U. z 2016 r. poz. 1311)

<sup>2</sup> Liczba punktów obowiązująca w roku ukazania się publikacji w danym czasopiśmie zgodna z wykazem czasopism naukowych sporządzonym przez Ministra Nauki i Szkolnictwa Wyższego

2. I. Stanimirova, B. Walczak, Classification of data with missing elements and outliers, *Talanta*, 76 (2008) 602-609 3,206 40
3. I. Stanimirova, S. Serneels, P.J. Van Espen, B. Walczak, How to construct a multiple regression model for data with missing elements and outlying objects, *Analytica Chimica Acta*, 581 (2007) 324-332 3,186 40
4. I. Stanimirova, Practical approaches to principal component analysis for simultaneously dealing with missing and censored elements in chemical data, *Analytica Chimica Acta*, 796 (2013) 27-37 4,387 40
5. I. Stanimirova, R. Tauler, B. Walczak, A comparison of positive matrix factorization and the weighted multivariate curve resolution method. Application to environmental data, *Environmental Science & Technology* 45 (2011) 10102-10110 5,228 45
6. I. Stanimirova, A. Woźnica, T. Płociniczak, M. Kwaśniewski, J. Karczewski, A modified weighted mixture model for the interpretation of spatial and temporal changes in the microbial communities in drinking water reservoirs using compositional phospholipid fatty acid data, *Talanta*, 160 (2016) 146-156 4,035 40
7. I. Stanimirova, K. Michalik, Z. Drzazga, H. Trzeciak, P.D. Wentzell, B. Walczak, Interpretation of ANOVA models using PCA to assess the effect of a maternal anticancer treatment on the mineralization of rat bones, *Analytica Chimica Acta*, 689 (2011) 1-7 4,555 40  
(artykuł zamieszczony na okładce wydania gazety)
8. I. Stanimirova, C. Boucon, B. Walczak, Relating gas chromatographic profiles to sensory measurements describing the end products of the Maillard reaction, *Talanta*, 83 (2011) 1239-1246 3,794 40

Razem: **31,787 325**

### **c) Omówienie celu naukowego ww. prac i osiągniętych wyników wraz z omówieniem ich ewentualnego wykorzystania**

#### *Zaawansowane strategie chemometryczne poprawiające ekstrakcję chemicznej informacji z wielowymiarowych niekompletnych danych i danych o indukowanej strukturze*

Nowoczesna chemia analityczna jest postrzegana jako zaawansowane podejście [1] do rozwiązywania złożonych problemów chemicznych. A zatem, cele chemii analitycznej mają na względzie rozwijanie nowych lub rozbudowę istniejących eksperymentalnych protokołów jakościowych i ilościowych w szczególności dla nowych typów próbek, tworzenie tanich metodologii eksperymentalnych umożliwiających pomiary rozmaitych zjawisk, eksplorowanie zgromadzonych wielowymiarowych danych eksperymentalnych, a także modelowanie i interpretację zależności w danych istniejących pomiędzy mierzonymi własnościami (zmiennymi). Takie ujęcie roli chemii analitycznej wymaga od jej użytkowników szerokiej wiedzy w zakresie metodologii planowania eksperymentu, teorii pobierania próbek, technik analizy instrumentalnej, kalibracji metod, optymalizacji i walidacji, standaryzacji, statystyki, analizy wielowymiarowych danych i ich interpretacji. Przez to nie jest zaskoczeniem, że na co dzień chemicy analitycy są aktywnie zaangażowani w rozwiązywanie rozmaitych problemów wywodzących się z obszarów na styku wielu dziedzin nauk, jak np. nauki stosowane z zakresu ochrony środowiska, chemia żywności, chemia sądowa, toksykologia, chemia medyczna, geochemia i różne nauki typu –omika jak proteomika, genomika czy metabolomika.

Każde zadanie analityczne rozpoczyna się od zdefiniowania istoty problemu, który zamierzamy rozwiązać. Przykładowe problemy badawcze mogą być sformułowane w następujący sposób: czy zmiany sposobu wytwarzania i zarazem utrata jakości wybranych produktów spożywczych i napojów można wyjaśnić zmianami w składzie jakościowym i ilościowym niektórych specyficznych komponentów [2,3], jaki efekt w czasie wywołuje wpływ pH gleby na unieruchamianie niektórych metali ciężkich występujących w różnych formach w zanieczyszczonej glebie [4], czy jest możliwe właściwe diagnozowanie pacjentów cierpiących na obturacyjną chorobę płuc i bezdech senny na podstawie zbioru wybranych metabolitów obecnych w surowicy krwi, kondensacie wydychanego powietrza czy w moczu i ustalenie jakie jest prawdopodobieństwo podjęcia błędnej decyzji diagnostycznej [5]. Ogólnie, problemy badawcze mogą być sklasyfikowane jako jakościowe, ilościowe lub podstawowe, ale przeważnie ich identyfikacja z początku jest niejednoznaczna. Dalsze uszczegółowienie problemu badawczego ułatwia wybór rodzaju informacji, którą należy gromadzić, a także dobór właściwej techniki analizy instrumentalnej pod kątem wymaganej od niej czułości, specyficzności, dokładności, granicy oznaczalności, kosztu i czasu jednostkowej analizy. Po przyjęciu schematu próbkowania i zaplanowaniu procedury eksperymentu, które uwzględniają wszystkie wcześniej wspomniane kwestie włącznie z wyborem odpowiedniej procedury walidacyjnej metody, przeprowadza się kolejne eksperymenty i gromadzi dane analityczne. Kolejny istotny krok obejmuje eksplorację i modelowanie otrzymanych danych eksperymentalnych (w zgodzie celem badań), a w konsekwencji interpretację oraz wizualizację wzajemnych zależności pomiędzy badanymi własnościami by zaproponować rozwiązanie podjętego problemu badawczego, albo zaplanowanie nowego eksperymentu w celu dalszego uszczegółowienia badań i pozyskania komplementarnej informacji. Obecnie poznawanie określonego chemicznego zjawiska jest zazwyczaj realizowane

poprzez pobór relatywnie dużej liczby próbek i ich opis stosując dużą liczbę charakterystycznych parametrów w nadziei, że zgromadzone dane pomogą ujawnić pożądaną informację. Ta tendencja w postępowaniu analitycznym wraz z szybkim rozwojem nowych technologii analizy próbek i ich obrazowania jest coraz bardziej wyraźna. Uzyskane dane mogą być wielowymiarowe, wieloblokowe (wielotablicowe), a także wielomodalne (tensorowe) i dlatego ich dalsza analiza wymaga stosownego podejścia uwzględniającego wszystkie zmierzone parametry. Z tego właśnie powodu, nowoczesna chemia analityczna ma za zadanie, w głównej mierze, wspomagać pozyskiwanie kluczowej informacji o badanym procesie na podstawie zgromadzonych danych.

Chemometria oferuje swym użytkownikom szeroki wachlarz metod matematycznych i statystycznych usprawniających planowanie eksperymentów [6], eksplorację danych, a także modelowanie i interpretację zależności występujących pomiędzy badanymi zmiennymi. Choć niektóre z klasycznych metod chemometrycznych są znane od wielu lat, np. podstawowe techniki wieloparametrowej kalibracji, klasyfikacji, dyskryminacji oraz dekonwolucji sygnałów instrumentalnych, dopiero w ciągu ostatnich dwóch dekad obserwuje się znaczący wzrost zainteresowania środowiska analityków ich stosowaniem i dalszym rozwijaniem. Głównymi czynnikami, które stymulowały do tej pory rozwój metod chemometrycznych, były możliwość poprawy skuteczności ekstrakcji użytecznej chemicznej informacji poprzez nowe technologie pomiarowe przy jednoczesnym zmniejszeniu kosztu i czasu jednostkowej analizy, a także możliwość rozwijania szybkich lub bezpośrednich i nieinwazyjnych procedur wspierających podejmowanie decyzji [7, 8]. Wykorzystanie sprzężonych technik chromatograficznych w połączeniu z metodami dekonwolucji sygnałów instrumentalnych stanowi przykład bardzo udanego podejścia rozwiązującego zagadnienie oceny czystości pików chromatograficznych (zjawisko współwymywania związków) na potrzeby analizy różnych produktów farmaceutycznych i chemicznych [9, 10]. Innym interesującym przykładem jest wykorzystanie wieloparametrowych modeli klasyfikacyjnych lub dyskryminacyjnych budowanych w oparciu o kompletne sygnały instrumentalne stanowiące unikalne chemiczne odciski palca. Obecnie często stosuje się tego typu metodologię, aby skutecznie potwierdzić autentyczność czy pochodzenie geograficzne nieznanymi próbek [2, 11]. Dodatkowo, szeroka dostępność przyjaznego dla użytkownika oprogramowania wspomagającego analizy chemometryczne i implementacja niektórych klasycznych podejść chemometrycznych w oprogramowaniu sterującym urządzeniami pomiarowymi niewątpliwie przyczyniają się do popularyzacji i zwiększenia stosowalności technik chemometrycznych. Jednakże, mimo relatywnie dobrego dostępu do metod chemometrycznych bezwzględnie wymagana jest ich dobra znajomość i podstawowe zrozumienie stosowanych algorytmów, aby móc poprawnie wybrać właściwy sposób wstępnego przygotowania danych, a także przeprowadzić odpowiednią wizualizację i interpretację. Zazwyczaj podstawowym ograniczeniem jest umiejętność odpowiedniego interpretowania uzyskanych wyników obliczeń przez użytkownika, co bezpośredniego nie wynika z umiejętności wykonania obliczeń stosując dostępne oprogramowanie [8].

Konieczność analizy i interpretacji dużych i złożonych danych chemicznych zwiększyła zapotrzebowanie na rozwijanie nowych podejść. W głównej mierze zainteresowanie środowiska chemometryków zwróciło się w stronę rozwijania metod, które w trakcie analizy chemometrycznej uwzględniają informację o błędzie próbkowania, poziomie instrumentalnego szumu lub informację o innych źródłach zmienności (niepewności pomiarowej), a także ułatwiających analizę danych zawierających różne rodzaje mierzonych parametrów (pozyskane z odmiennych metod analitycznych), analizę chemicznych odcisków palca i danych zgromadzonych w oparciu o specyficzne statystyczne podejście planowania eksperymentów (dane o indukowanej strukturze). Tak obrane kierunki rozwoju



metod chemometrycznych pociągają za sobą konieczność prowadzenia dalszych badań w obszarze wstępnego przygotowania danych, wyboru istotnych zmiennych, rozwijania stabilnych wieloparametrowych metod analizy danych, analizy danych z brakującymi elementami, przetwarzania sygnałów instrumentalnych oraz zwiększania stosunku sygnału do szumu.

Istnieje wiele różnych podejść zwiększenia skuteczności ekstrakcji chemicznie użytecznej informacji ze zgromadzonych danych. Jednym ze sposobów jest bardziej efektywne planowanie eksperymentu i w konsekwencji pozyskanie informacji o zdecydowanie lepszej jakości dzięki zasadom statystycznego planowania eksperymentów. W praktyce analitycznej dobrze znane są plany kompletne, ułamkowe i centralne plany kompozycyjne, które są z powodzeniem stosowane w wariancie jednoczynnikowym w celu przesiewowej analizy lub optymalizacji warunków eksperymentalnych [12, 13]. Jednakże włączenie informacji o możliwej interakcji dwóch lub większej liczby czynników jest relatywnie nowym i wciąż problematycznym zagadnieniem. Ta kwestia zostanie zilustrowana przykładem zaczerpniętym z badań w obszarze chemii medycznej. Aby ocenić efekty związane z terapią antynowotworową z zastosowaniem cyklofosfamidu i cytarabiny podczas ciąży i ich wpływ na proces mineralizacji kości żuchwy u szczurów 7-, 14- i 28-dniowych [14], przyjęto następujące hipotezy zerowe: efekt czynnika 'wiek', 'leczenie' lub ich interakcji nie ma statystycznie istotnego wpływu na proces mineralizacji kości żuchwy. Wybrano dwuczynnikowy trójpoziomowy kompletny plan w celu zaplanowania eksperymentu oraz technikę dyspersyjnej rentgenowskiej fluorescencji jako niedestrukcyjną metodę charakteryzującą skład pierwiastkowy wybranych do analiz próbek. W tym przypadku głównym wyzwaniem było przeprowadzenie statystycznego testowania w oparciu o całościową informację zawartą w sygnałach XRF bez analizy ilościowej, co standardowo uzyskuje się poprzez dekonwolucję sygnałów instrumentalnych. Weryfikacja tak postulowanych hipotez zerowych na określonym poziomie istotności, na podstawie wielowymiarowych danych spektroskopowych o dużym stopniu korelacji wymaga użycia zaawansowanych metod chemometrycznych.

Inne praktyczne podejście stosowane by zwiększyć zawartość informacyjną danych i jednocześnie uzyskać całościowe spojrzenie na badane zjawiska polega na gromadzeniu danych z różnych platform analitycznych co w literaturze określa się mianem fuzji danych. Fuzja danych może być realizowana na kilka możliwych sposobów i na różnych poziomach w zależności od zamierzonego celu badań [15, 16]. Jej użyteczność uzasadnia fakt, że różne techniki analityczne pozwalają pozyskać komplementarną informację o próbkach, ponieważ niektóre grupy związków są widoczne lepiej stosując tylko określoną technikę. Jak wcześniej wspomniano, chromatograficzne techniki sprzężone takie jak chromatografia gazowa lub cieczowa sprzężona ze spektrometrią mas to obecnie standardowe techniki analityczne używane do badania złożonych mieszanin pochodzenia naturalnego, jednakże analiza połączonych danych pozyskanych techniką spektrometrii mas w trakcie rozdziału techniką chromatografii gazowej i cieczowej dla tego samego zestawu próbek to trudne zadanie [17]. Nadrzędnym celem tego typu badań jest ustalenie czy analiza poszczególnych zestawów parametrów czy też ich łączna analiza [18] pozwala w konsekwencji uzyskać pełniejszy opis danych, a w konsekwencji lepszy opis zjawiska. Z drugiej strony, w takiej sytuacji, liczba parametrów zazwyczaj znacząco przewyższa liczbę próbek co prowadzi do tzw. danych o pełnym rzędzie. To znaczy, że wiele zmiennych nie wnosi dodatkowej informacji do opisu zjawiska. Wówczas, należy rozważyć użycie któregoś wariantu metody wyboru istotnych zmiennych. Przygotowanie próbek oraz ich pobór można również realizować stosując zasady planowania eksperymentu. Ma to często miejsce w badaniach metabolomicznych [5], czy badaniach preferencji konsumentów [19] i prowadzi do

dotychczasowych komplikacji przy wyborze adekwatnej techniki eksploracji i modelowania wielowymiarowych danych.

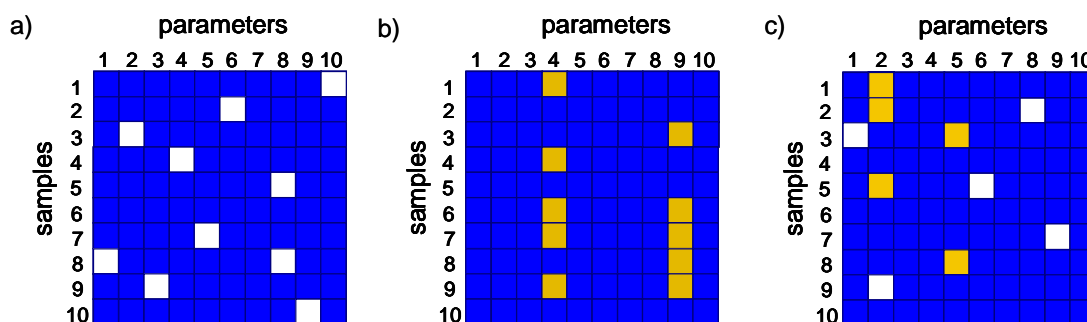
Ostatnim aspektem poprawy ekstrakcji chemicznie istotnej informacji z danych, lecz nie mniej ważnym, jest włączenie znanej z góry informacji na temat błędu próbkowania, poziomu instrumentalnego szumu, czy informacji o poziomie innych możliwych źródeł wariacji [20]. Fundamentalnym ograniczeniem wszystkich klasycznych metod analizy wielowymiarowych danych jest założenie, że błędy pomiarowe obserwowane dla uzyskanych wyników pomiarów mają niezależny i identyczny rozkład normalny (z ang. *i.i.d – independent and identically distributed*). W analizie mieszanin pochodzenia naturalnego takich jak np. próbki biologiczne czy środowiskowe to założenie jest najczęściej niespełnione. W przypadku tego typu pomiarów, niepewność pomiarowa uzyskanych danych zmienia się wraz ze wzrostem intensywności sygnału analitycznego. Wynikiem tego jest możliwość zignorowania w trakcie analizy chemometrycznej z użyciem klasycznych technik niektórych związków chemicznych obecnych w próbce w małych stężeniach, pomimo że mogą one mieć porównywalny stosunek sygnału do szumu jak w przypadku związków o znacznie wyższych stężeniach [21]. Z tego powodu rozwijanie technik chemometrycznych, które bezpośrednio włączają informację o wielkości niepewności pomiaru jest niezwykle istotna. Szczególną uwagę zwraca się na sytuację, gdy w danych ilościowych brakuje części wyników pomiarów przy jednoczesnej relatywnie dużej niepewności pomiarowej. Niektóre z pomiarów mogą być nierejestrowane z powodu nieprawidłowego funkcjonowania aparatury/czujnika czy wysokiego kosztu jednostkowej analizy lub z powodu wykonania oznaczenia poniżej lub powyżej granicy oznaczalności danej metody analitycznej. Niestety to zagadnienie jest często błędnie rozumiane, a nawet pomijane przez analityków na etapie wieloparametrowej analizy danych. Zwyczajowy sposób postępowania zakłada usunięcie kolumny lub wiersza macierzy, zawierających niekompletne dane. Jednakże takie postępowanie prowadzi do niepotrzebnej eliminacji znacznej części obserwowanych danych. Innym, relatywnie łatwym sposobem, jest zastąpienie brakujących elementów przez średnią kolumny lub średnią wiersza albo przez tzw. wartości oczekiwane. Także i ten sposób uchodzi za niewłaściwy ze względu na to, że średnia, odchylenie standardowe i kowariancja są wówczas źle estymowane. W przypadku analizy danych środowiskowych mogłoby to oznaczać rekomendację kosztownej procedury remediacji, podczas gdy nie jest to konieczne lub możliwe byłoby przeoczenie występowania skażenia. Kwestia analizy danych niekompletnych zazwyczaj dodatkowo komplikuje obecność w danych obiektów odległych. Są to próbki, dla których wartości niektórych parametrów znacznie różnią się w porównaniu do wartości tychże parametrów dla większości próbek w danych. Obiekty odległe powstają w wyniku popełnienia tzw. grubego błędu analitycznego lub mogą to być próbki o unikalnej charakterystyce. Wtedy estymacja parametrów modelu opartego na minimalizacji sumy kwadratów reszt jest znacznie obciążona obecnością w danych obiektów odległych, natomiast występowanie w danych brakujących elementów uniemożliwia poprawną identyfikację obiektów odległych [22]. Z tego powodu niezwykle pożądane są metody pozwalające analizować dane, w których współwystępują obiekty odległe i brakujące elementy.

Spójny cykl badań (opisany w publikacjach H1-H8 znajdujących się w załączniku opracowania) będący podstawą niniejszego osiągnięcia naukowego, miał na celu w głównej mierze opracowanie zaawansowane strategie analizy chemometrycznej wieloparametrowych i niekompletnych danych zawierających obiekty odległe, a także metody analizy wieloparametrowych danych o indukowanej strukturze. Przydatność zaproponowanych przeze mnie podejść, które zwiększają ekstrakcję użytecznej chemicznie informacji z danych analitycznych, wyczerpująco opisałam we wspomnianym cyklu artykułów. Jednakże dla lepszego zrozumienia przedstawianych w dokumentacji zagadnień

przybliżyć najważniejsze wyzwania badawcze podjęte przeze mnie w ramach pracy habilitacyjnej i moje główne osiągnięcia związane z realizacją obranych celów badawczych.

## 1. Strategie chemometryczne ulepszające ekstrakcję chemicznej informacji z wielowymiarowych i niekompletnych danych

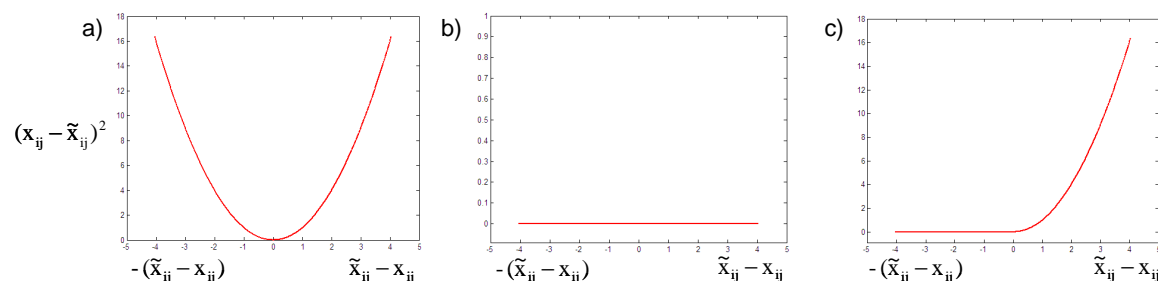
Wszystkie klasyczne metody wieloparametrowe wymagają, aby dane analityczne były kompletne. Tylko wówczas analiza danych jest możliwa. Zatem, jak wspomniano już wcześniej, zasadnym jest poszerzenie klasycznych algorytmów o możliwość przetwarzania niekompletnych danych. W praktyce, zważywszy mechanizm ich powstawania mamy do czynienia z trzema rodzajami brakujących elementów [23]. Gdy mechanizm powstawania w danych brakujących elementów nie zależy od parametrów, wówczas mamy do czynienia z danymi, w których brakujące elementy występują całkowicie losowo (MCAR). Nie tworzą one żadnego charakterystycznego wzoru w danych, co przykładowo ilustruje Rys. 1a. Elementy brakujące losowo powstają wówczas, gdy mechanizm ich powstawania zależy od zmiennych, które zawierają wszystkie elementy (MAR). Natomiast gdy brakujące elementy powstają w konsekwencji mechanizmu, który zależy od niekompletnych parametrów, wówczas takie braki w danych nazywamy brakującymi elementami w sposób nielosowy (NMAR). Tą sytuację przedstawiono na Rys. 1b. Wartości tychże elementów są poniżej lub powyżej granicy oznaczalności/wykrywalności danej metody analitycznej i określane są również mianem elementów cenzorowanych [24]. W danych analitycznych mogą występować MCAR lub elementy cenzorowane poniżej granicy oznaczalności/wykrywalności, ale również jedno i drugie (zob. Rys. 1c). Ponieważ elementy brakujące losowo, MAR, są w przypadku danych chemicznych rzadkością, z tego powodu nie były one przedmiotem dalszych rozważań.



Rys. 1 Rozkłady 10% brakujących elementów w danych powstających zgodnie z mechanizmem: a) MCAR, b) NMAR oraz c) MCAR i NMAR.

Istnieje kilka znanych podejść zastępowania brakujących elementów, które występują w danych całkowicie losowo [25]. Ponieważ podstawianie tychże elementów wartościami średnimi odpowiednich wierszy lub kolumn, wartościami oczekiwanymi czy innymi niszczy w wyraźny sposób zależności pomiędzy zmiennymi i pomiędzy próbkami, wspólną cechą wszystkich metod niedawno zaproponowanych jest takie zastąpienie brakujących elementów, aby struktura korelacyjna danych była zachowana. Nie dziwi zatem, że nowoczesne metody wykorzystują estymatory wartości średniej czy kowariancji lub opierają się na konstrukcji głównych składowych lub modelach regresji.

W szczególności algorytmy uwzględniają symetryczną kwadratową funkcję kosztów dla obserwowanych elementów (zob. Rys. 2a), zerową funkcję kosztów dla elementów typu MCAR (Rys. 2b) i/lub pół-kwadratową funkcję kosztów z górnym limitem dla elementów cenzorowanych poniżej granicy oznaczalności/wykrywalności (Rys. 2c).



Rys. 2 Funkcja kosztów dla: a) elementów obserwowanych, b) elementów MCAR oraz c) elementów cenzorowanych poniżej granicy oznaczalności/wykrywalności.

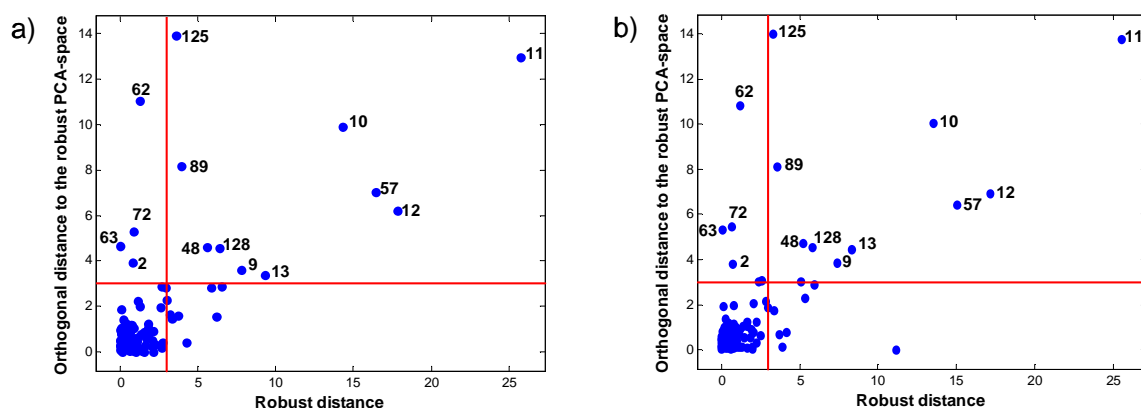
Iteracyjne podejście maksymalizacji wartości oczekiwanej (EM) to jeden z najpopularniejszych algorytmów [26] dedykowany przetwarzaniu danych z elementami typu MCAR. Procedurę rozpoczyna krok inicjalizacji brakujących elementów poprzez ich podstawienie wartościami średnimi odpowiednich wierszy lub kolumn macierzy danych. Po wstępnej estymacji parametrów modelu (może nim być jakikolwiek wieloparametrowy model), brakujące elementy są zastępowane ich przewidzianymi na podstawie modelu wartościami. Ten krok jest wielokrotnie powtarzany póki algorytm nie uzbieźni się, co oznacza, że w dwóch kolejnych iteracjach suma kwadratów reszt obliczona dla brakujących elementów jest poniżej ustalonej wartości progowej, np.  $10^{-8}$ . W przypadku przewidzianych wartości ( $\tilde{x}_{ij}$ ) dla elementów MCAR są one idealnie dopasowane do skonstruowanego modelu i mają zerowe reszty, tj.  $(\tilde{x}_{ij} - x_{ij}) = 0$ , co obrazuje Rys. 2b. Algorytm EM [26] jest niezwykle wszechstronny i stanowi podstawę wielu metod: probabilistycznej analizy głównych składowych [27], podejścia maksymalizacji wartości oczekiwanej w połączeniu z analizą głównych składowych z wykorzystaniem procedury dekompozycji macierzy do wartości własnych, podejścia maksymalizacji wartości oczekiwanej w połączeniu z metodą częściowych najmniejszych kwadratów, EM-PLS, podejścia maksymalizacji wartości oczekiwanej w połączeniu z metodą klasyfikacyjną SIMCA, a także algorytm N-modalnego Tuckera [28]. Jednakże algorytm EM jest nieodpowiedni dla danych z cenzorowanymi elementami i w sytuacji występowania w danych znacznej liczby brakujących elementów może prowadzić do bardzo czasochłonnych obliczeń. Oczywiście w literaturze zaproponowano nowe koncepcje i metody przetwarzania danych z elementami MCAR. Nadrzędny cel, jaki przyświecał ich twórcom, miał doprowadzić do opracowania algorytmu, który w relatywnie szybki sposób może dostarczyć stabilne rozwiązanie dla dużych danych, obciążone w niewielki sposób w sensie statystycznym w przypadku występowania brakujących elementów na poziomie powyżej 70%. Z drugiej strony, mając na względzie praktyczne aspekty wykorzystania tychże podejść, wydaje się zasadne rozważyć czy dla danych zawierających 80% brakujących elementów uzyska się rozwiązanie, któremu można nadać wiarygodną interpretację?

Zazwyczaj wielowymiarowe dane eksperymentalne zawierają obiekty odległe. Jak już wcześniej wspomniano, podstawowym ograniczeniem standardowych modeli, włącznie z tymi stosownymi do

analizy danych z brakującymi elementami, jest silny wpływ tychże obiektów na estymowane parametry modelu. To sprawia, że model niepoprawnie opisuje większości danych. Z góry oczywistym sposobem przewyciężenia tego problemu jest identyfikacja i wyeliminowanie obiektów odległych z danych, a następnie powtórzenie analizy dla 'czystego' zbioru danych. Inne, bardziej efektywne podejście zakłada użycie stabilnego wieloparametrowego modelu, co pozwala poprzez procedurę ważenia obiektów odległych zminimalizować ich wpływ na końcowe parametry modelu. Rozwijanie wieloparametrowych stabilnych modeli jest nadal dużym wyzwaniem. Do tej pory, w literaturze zaproponowano kilka podejść, aby móc estymować stabilne wektory wyników i wag w analizie głównych składowych [29-31]. Przykładami stabilnych modeli są: stabilny wariant metody regresji częściowych najmniejszych kwadratów [32-35], stabilny wariant metody klasyfikacyjnej SIMCA [36], stabilny wariant równoległej analizy czynnikowej PARAFAC [37], a także stabilny wariant regresji głównych składowych [38] są kolejnymi przykładami stabilnych modeli. To zagadnienie może być dodatkowo utrudnione, jeśli rozważamy analizę czy modelowanie niekompletnych wielowymiarowych danych zawierających także obiekty odległe. Trudność związana z obecnością obiektów odległych w niekompletnych danych bezpośrednio wynika z przeszkody poprawnego dopasowania brakujących elementów, podczas gdy obecność brakujących elementów nie pozwala na skuteczną identyfikację obiektów odległych. Rozwiązaniem tego bardzo istotnego problemu jest minimalizacja odpowiednio dobranej funkcji kosztów dla brakujących elementów (zob. Rys. 2b) podczas minimalizacji stabilnej funkcji kosztów dla obserwowanych elementów. Zaproponowany przeze mnie sposób rozwiązania tego zagadnienia opiera się na implementacji stabilnej wieloparametrowej techniki wewnątrz podstawowych kroków procedury EM.

W artykule H1, wchodzącym w skład cyklu habilitacyjnego, zaproponowałam efektywne i zarazem relatywnie proste podejście do eksploracji wieloparametrowych danych zawierających brakujące elementy i obiekty odległe jako alternatywę do klasycznej metody PCA. Poddałam analizie przydatność opracowanej metody, a także jej własności uzbieżnienia używając dane symulowane oraz dane eksperymentalne. Podejście to, o akronimie EM-SPCA [39], opiera się na jednym ze stabilnych wariantów metody PCA – tzw. sferycznej metodzie PCA [40], SPCA. W tymże stabilnym wariancie metody PCA, wieloparametrowe dane są rzutowane na hiper sferę o jednostkowym promieniu zakreślonym w punkcie będącym stabilnym środkiem danych, określonym za pomocą mediany typu 'L1'. Środek danych estymowany za pomocą mediany typu 'L1' to taki punkt w wielowymiarowej przestrzeni, od którego suma odległości Euklidesowych do wszystkich punktów jest najmniejsza. Jest to rozszerzenie klasycznego jednoparametrowego estymatora mediany do wielowymiarowego przypadku. Obiekty są następnie ważone biorąc pod uwagę ich odległość od stabilnego środka danych. Te, które są daleko od stabilnego środka danych, otrzymują dużą wagę i ich wpływ na konstrukcję modelu jest zanedbywalny. Stabilne wektory wyników i wag uzyskiwane są poprzez zbudowanie klasycznego modelu PCA dla danych, w których uprzednio obiekty poddano procedurze ważenia. Ostatecznie, identyfikacja obiektów odległych odbywa się za pomocą wykresu diagnostycznego, w którym dla każdego obiektu jego reszty od modelu PCA rysuje się względem stabilnych wartości odległości Mahalanobisa. Sferyczny wariant metody PCA wydaje się być lepszy od innych stabilnych wariantów PCA ze względu na swoją prostotę i relatywnie wysoki punkt załamania estymatora (z ang. *high breakdown point*). Zaproponowana procedura EM-SPCA zawiera te same kroki co algorytm EM wcześniej opisany w tekście. Kilka ważnych kwestii wymaga naszej uwagi i rozwiązania: i) w jaki sposób wstępnie przygotować dane, ii) jak oszacować liczbę czynników w modelu stosując algorytm, który prowadzi do uzyskania tzw. rozwiązań niezagnieżdżonych oraz iii) analiza własności uzbieżniania się algorytmu. Praktyczne własności zaproponowanego podejścia

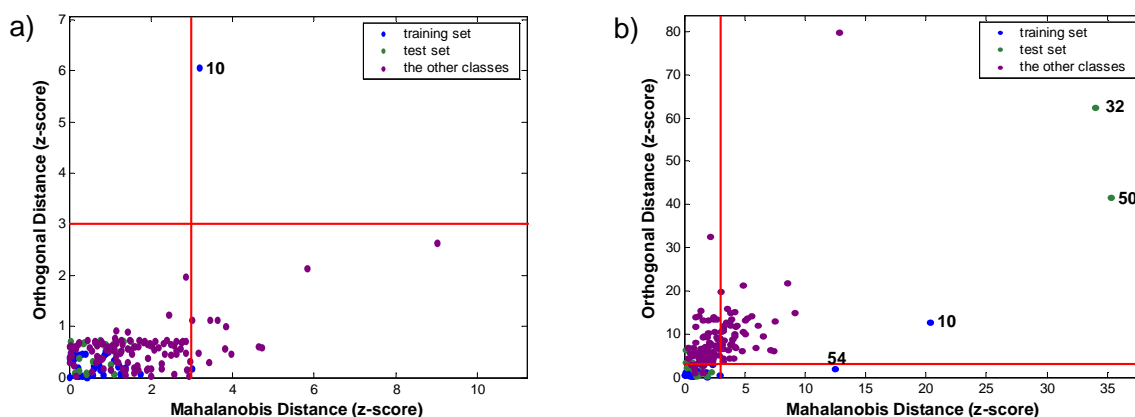
schematycznie przedstawia Rys. 3. Zidentyfikowane na podstawie wykresu diagnostycznego obiekty odległe dla danych zawierających 5% brakujących elementów (zob. Fig. 3b) są takie same jak w przypadku kompletnych danych (Fig. 3a).



Rys. 3 Wykres diagnostyczny obrazujący reszty od modelu EM-PCA względem stabilnych wartości odległości Mahalanobisa dla poszczególnych próbek sporządzony dla eksperymentalnych danych: a) bez brakujących elementów i b) z 5% pulą brakujących elementów.

Zazwyczaj metodę analizy głównych składowych wykorzystuje się aby eksplorować strukturę wielowymiarowych danych, ale często celem analizy chemometrycznej jest także budowa reguł klasyfikacyjnych czy dyskryminacyjnych, na podstawie których można przypisać nową próbkę do określonych grup. Jest to w szczególności użyteczne w przypadku badań autentyczności produktów czy badań metabolomicznych. Skonstruowane za pomocą metod dyskryminacyjnych reguły logiczne ułatwiają przypisanie nowej próbki do jednej konkretnej grupy, podczas gdy w przypadku metod klasyfikacyjnych skonstruowane reguły klasyfikacyjne dodatkowo pozwalają, by nowa próbka została przypisana do większej liczby grup albo żadnej. Do klasycznych metod dyskryminacyjnych [41] zaliczamy liniową i kwadratową analizę dyskryminacyjną (LDA, QDA), drzewa klasyfikacji i regresji (CART) oraz dyskryminacyjny wariant metody częściowych najmniejszych kwadratów (PLS-DA). Wśród metod klasyfikacyjnych do najpopularniejszych należą takie metody jak SIMCA [42] czy UNEQ [41]. Głównym celem metody SIMCA jest konstrukcja granic dla każdej z grup z osobna w oparciu o model PCA, zbudowany dla próbek modelowych z określonej gupy. Budowa indywidualnych modeli PCA wymaga w każdym przypadku dokonania wyboru liczby istotnych czynników głównych, szczególnej uwagi przy wyborze reprezentatywnych próbek do zbioru modelowego i do zbioru testowego oraz ustalenia efektywnego sposobu przypisywania nowych próbek do istniejących grup. Dodatkową trudność napotykamy w sytuacji, gdy dane są niekompletne i zawierają obiekty odległe. Mimo, że zaproponowanie metody EM-SPCA wydaje się być ułatwieniem, wciąż pożądane są dodatkowe strategie. W publikacji [36] przedstawiono stabilny wariant metody klasyfikacyjnej SIMCA wykorzystującej model SPCA i przebadano własności uzbieżniania nowej metody. Następnie rozszerzyłam to podejście o możliwość tworzenia modeli klasyfikacyjnych dla danych z brakującymi elementami, EM-S-SIMCA (zob. artykuł H2 w aneksie opracowania) [43]. Rys. 4 przedstawia różnice w zastosowaniu klasycznej metody EM-SIMCA oraz proponowanej metody EM-S-SIMCA. Gdy dane zawierają obiekty odległe i brakujące elementy, klasyczny wariant metody EM-SIMCA umożliwia jedynie przetwarzanie danych z brakującymi

elementami i prowadzi do niewłaściwej estymacji granic dla każdej z grup. To pociąga za sobą relatywnie niską specyficzność modelu, co oznacza, że próbki należące do innych grup (oznaczone na Rys. 4a kolorem magenta) są rozpoznawane za pomocą modelu jako należące do modelowanej grupy. Natomiast model EM-S-SIMCA, którego wyniki przedstawiono na Rys. 4b, zbudowany dla tej samej grupy próbek cechuje się dużą czułością i specyficznością.

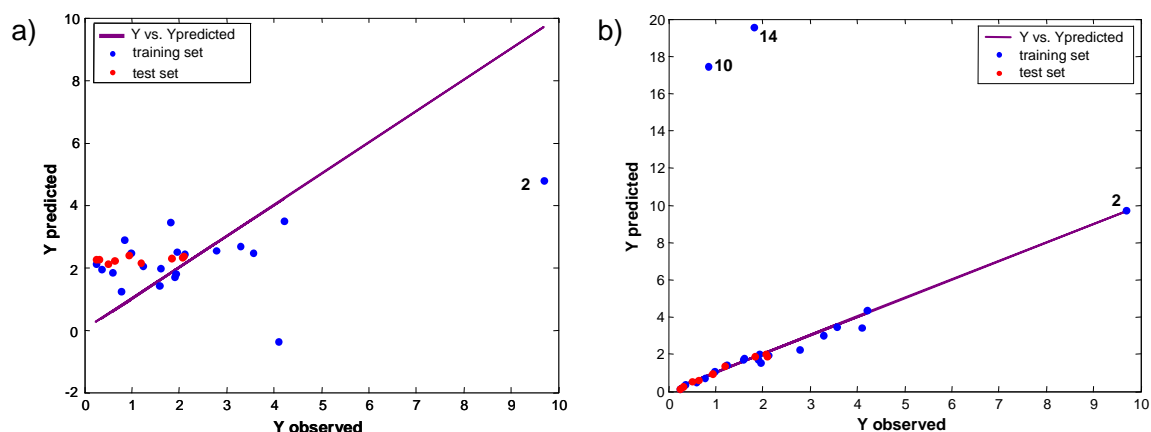


Rys. 4 Wykres diagnostyczny otrzymany stosując model: a) EM-SIMCA i b) EM-S-SIMCA dla danych eksperymentalnych zawierających 10% brakujących elementów i obiektów odległych.

By w sposób automatyczny ustalić kompleksowość modelu, zaproponowałam i wykorzystałam w algorytmie zmodyfikowany wariant metody ‘leverage correction’ [44]. W toku badań ustaliłam także, że wybór próbek modelowych i testowych w przypadku niekompletnych danych powinien być przeprowadzony z zastosowaniem macierzy czynników głównych o pełnym rzędzie, uzyskanej z procedury EM-PCA, która umożliwia optymalny wybór próbek do zbioru modelowego i testowego za pomocą metody Kennarda i Stonea [45] czy metody ‘duplex’ [46]. Wybór próbek może być także dokonany w sposób losowy, jeśli ich równomierny rozkład nie jest wymagany. W oparciu o metodę projekcji na płaszczyznę modelu (PMP) [47] uzyskałam satysfakcjonujące wyniki przewidywania dla próbek zawierających brakujące elementy. Z tego powodu wybrałam metodę PMP. Opracowany algorytm z powodzeniem został użyty do budowy modeli klasyfikacyjnych wspomagających wykrywanie zafałszowań produktów na podstawie ich profili chemicznych poprzez weryfikację specyfikacji produktu spożywczego. To rozwiązanie zostało włączone do oprogramowania stanowiącego jeden z wyników mojej działalności naukowej jako wykonawcy europejskiego projektu „TRACE” (*Tracing food commodities in Europe*; FOOD-CT-2005-006942), finansowanego w 6 programie ramowym UE.

Jak wspomniano już wcześniej, metoda częściowych najmniejszych kwadratów (PLS), może być użyta zarówno do budowy modeli dyskryminacyjnych i regresyjnych. To jedna z najbardziej popularnych technik chemometrycznych, gdyż pozwala na przetwarzanie danych zawierających wiele silnie ze sobą skorelowanych zmiennych. Jednakże, podobnie jak w przypadku wszystkich klasycznych metod, uzyskane wyniki z metody PLS mogą być znacznie obciążone obecnością w danych próbek odległych, a także nie może być stosowana w przypadku danych zawierających brakujące elementy. W ostatnich latach znajdujemy w literaturze kilka propozycji stabilnych wariantów metody PLS, jednakże wyczerpujące analizy symulacyjne wykazały, że ze wszystkich

stabilnych wariantów PLS metoda stabilnej częściowej M-regresji (PRM) ma najlepsze własności [32] estymacji współczynników regresji zważywszy na ewentualny stopień obciążenia estymatora, relatywnie dużą efektywność w przypadku rozkładu normalnego i jest mało czuła w przypadku rozkładów danych o tzw. ‘długich ogonach’. PRM to procedura iteracyjna, w której waży się próbki uwzględniając odległość Mahalanobisa w przestrzeni czynników modelu i reszty od modelu tak, by zminimalizować wpływ obiektów odległych na estymowane współczynniki regresji. Jednakże algorytm PRM nie może być bezpośrednio użyty do modelowania danych zawierających brakujące elementy. W związku z tym zaproponowałam nowe podejście, oparte o procedurę EM, polegające na włączeniu metody PRM do procedury iteracyjnej EM, aby dla brakujących elementów zminimalizować funkcję celu przedstawioną na Rys. 2b. Należy zauważyć, że w chwili obecnej zaproponowany algorytm może być użyty jedynie do rozwiązania problemów regresyjnych. Możliwość jego uzbieżniania były przedmiotem mojej wnikliwej analizy [48] podobnie jak skuteczne wykazanie użyteczności nowej metody w przypadku modelowania symulowanych zbiorów danych oraz eksperymentalnych danych. Ideę proponowanego algorytmu wraz z wynikami jego testowania przedstawiłam wyczerpująco w artykule H3, dołączonym do aneksu niniejszego opracowania.



Rys. 5 Diagramy przedstawiające wartości Y obserwowanego względem wartości Y przewidzianego uzyskane na podstawie metod: a) EM-PLS i b) EM-PRM dla zanieczyszczonych danych modelowych zawierających 30% brakujących elementów.

Analiza porównawcza metody EM-PLS i zaproponowanego podejścia EM-PRM potwierdza, że EM-PRM przewyższa EM-PLS pod względem wartości średniego błędu kwadratowego dla zbioru modelowego i testowego uwzględniającego określony, dopuszczalny poziom kontaminacji zbioru modelowego obiektami odległymi występującymi w zmiennej zależnej. Rys. 5 przedstawia różnice w działaniu dwóch algorytmów dla danych z obiektami odległymi i danych zawierających 30% brakujących elementów. Metoda EM-PLS jest w dużym stopniu czuła na obecność obiektów odległych, a ogólny trend modelu nie oddaje tendencji charakterystycznej dla większości danych ze zbioru modelowego (zob. Rys. 5a). W przypadku metody EM-PRM wpływ obiektów odległych jest zminimalizowany poprzez procedurę ich ważenia, co prowadzi do zdecydowanie lepszych wyników (zob. Rys. 5b). Metoda EM-PRM działa zgodnie z oczekiwaniem, to znaczy nieco gorzej niż EM-PLS w sytuacji, gdy w danych nie występują obiekty odległe i występuje do 50% brakujących elementów.



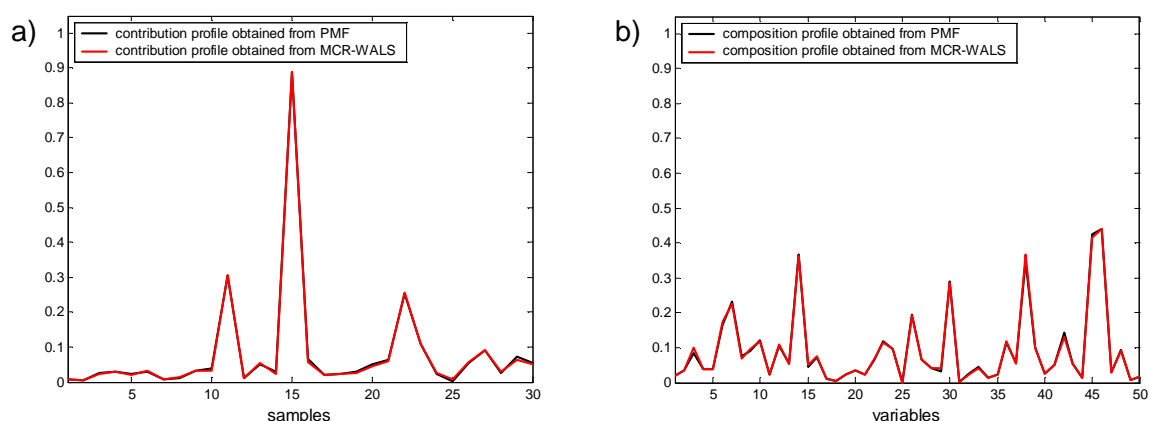
Podobne podejście, jak te omówione w trzech publikacjach, zostało wykorzystane przez innych autorów, aby skonstruować stabilny wariant regresji głównych składowych [38], dodatkowe warianty analizy głównych składowych [49] czy stabilny wariant współbieżnej analizy czynnikowej [37], z możliwością przetwarzania danych zawierających brakujące elementy typu MCAR. Głównym ograniczeniem tychże metod jest to, że nie mogą być stosowane, gdy dane zawierają brakujące elementy typu MCAR. Często w takich sytuacjach pada pytanie, w jaki sposób należy przetwarzać czy modelować dane zawierające elementy cenzorowane poniżej lub powyżej granicy oznaczalności/wykrywalności albo jedno i drugie wraz z brakującymi elementami typu MCAR? Ta kwestia była raczej skutecznie pomijana przez wielu analityków, gdyż zakłada się, że elementy cenzorowane poniżej granicy oznaczalności/wykrywalności mają zaniedbywalny wpływ na strukturę korelacyjną danych. Przyjmując właśnie takie założenie, w wielu przypadkach podstawia się je zerami, wartością równą połowie granicy detekcji metody czy wartościami oczekiwanymi. Stąd, przedmiotem moich zainteresowań było dokonanie analizy do jakiego stopnia obecność w wielowymiarowych danych elementów cenzorowanych poniżej granicy oznaczalności/wykrywalności wpływa na ich interpretację, która z dostępnych metod jest najbardziej odpowiednia by przetwarzać dane z elementami cenzorowanymi, jaki jest dopuszczalny procent elementów cenzorowanych obecnych w pojedynczej zmiennej, które można zastąpić stosując określoną metodę podstawiania i która metoda jest najbardziej skuteczna w przypadku danych zawierających elementy typu MCAR i elementy cenzorowane. Te rozważania wraz z uzyskanymi wnioskami badań przedstawiłam w publikacji H4, dołączonej w aneksie niniejszego opracowania. W toku prowadzonych przeze mnie badań [50], dokonałam w sposób wyczerpujący porównania kilku wariantów metody analizy głównych składowych i kilku metod podstawiania elementów dla symulowanych i eksperymentalnych danych. W szczególności zwróciłam uwagę na podejścia umożliwiające bezpośrednie włączenie na etapie budowy modeli niepewności pomiarowej lub na specyficznych metodach minimalizujących określoną funkcję kosztów. Obiektem moich zainteresowań było poznanie efektywności takich metod jak metoda największej wiarygodności PCA (MLPCA) [51, 52] i ogólna nieliniowa iteracyjna metoda częściowych najmniejszych kwadratów (GNIPALS) w stosunku do klasycznej PCA. Metoda GNIPALS [53] to sekwencyjna procedura iteracyjna, w której minimalizuje się wszystkie funkcje kosztów przedstawione na Rys. 2 dla każdego skonstruowanego czynnika. Forma funkcji kosztów dla elementów cenzorowanych poniżej granicy oznaczalności/wykrywalności (zob. Rys. 2c) sugeruje, że przewidzenie wartości powyżej górnej granicy oznaczalności prowadzi do błędu o określonej wielkości, gdy minimalizuje się sumę kwadratów reszt, natomiast przewidziana wartość poniżej górnej granicy oznaczalności przyjmuje wartość zero. Jest to tożsame z przewidzeniem przez model PCA o określonej kompleksowości dla elementów cenzorowanych poniżej granicy oznaczalności/wykrywalności wartości poniżej granicy oznaczalności. Problem minimalizacji funkcji kosztów rozwiązywany jest w oparciu o algorytm GNIPALS i procedurę wypukłej optymalizacji. Te same problemy minimalizacji można rozwiązywać również dla metod włączających informację o niepewności pomiarów [20]. Wszystkie one minimalizują sumę kwadratów ważonych reszt, przy czym wagę dla każdego obiektu definiuje się jako odwrotność wariancji jego błędu pomiarowego. Gdy dla wszystkich obiektów ich wagi przyjmują wartość jeden, wówczas uzyskamy rozwiązanie klasycznej metody PCA. Duże wartości odchylenia standardowego błędu na poziomie  $10^5$  (duża niepewność pomiarowa) dla elementów typu MCAR umożliwia wiarygodne przewidywanie brakującej informacji, a forma funkcji sumy kwadratów ważonych reszt jest taka sama jak na Rys. 2b. Przyjęcie wartości dla niepewności pomiarowych w przypadku elementów cenzorowanych poniżej granicy oznaczalności/wykrywalności nie jest oczywiste. W literaturze znajdujemy różne propozycje rozwiązania w tej kwestii. Ich wspólną cechą

jest założenie, że elementy cenzorowane poniżej granicy oznaczalności/wykrywalności powinny być przewidziane poniżej pewnej, z góry ustalonej wartości progowej. Wszystkie szczegóły dotyczące algorytmu i jego głównych kroków zostały opisane w publikacji H4, znajdującej się w załączniku niniejszego opracowania. Najważniejszy wniosek płynący z przeprowadzonej analizy porównawczej wskazuje, że elementy cenzorowane poniżej granicy oznaczalności/wykrywalności faktycznie można podstawić połową wartości granicy detekcji, jeśli ich liczba w danej zmiennej w przypadku wielowymiarowych danych chemicznych znajduje się w przedziale 30-40%. W sytuacji większej liczby zmiennych z elementami cenzorowanymi, metoda PCA prowadzi do obciążonego rozwiązania. Algorytm GNIPALS jest najlepszą metodą, by zwiększyć efektywność przetwarzania większej liczby elementów cenzorowanych poniżej granicy oznaczalności/wykrywalności w przeliczeniu na zmienną. Z kolei działanie algorytmu MLPCA dla elementów cenzorowanych poniżej granicy oznaczalności/wykrywalności w istotny sposób zależy od przyjętej definicji niepewności. Zgodność w działaniu EM-PCA, ogólnej metody NIPALS i MLPCA wykazano także dla danych zawierających brakujące elementy typu MCAR.

## **2. Strategie chemometryczne zwiększania ekstrakcji chemicznej informacji z wielowymiarowych danych poprzez włączenie informacji o niepewności pomiarowej**

Jak wspomniano już wcześniej w tekście, włączenie wiedzy o błędzie pomiarowym, znanym *a priori*, poziomu szumu instrumentalnego czy wielkości innych źródeł wariancji polepsza ekstrakcję chemicznej informacji z wielowymiarowych danych chemicznych. W tej sytuacji zaproponowano metodę PCA maksymalizującą wiarygodność (MLPCA) [51] jako zamiennik klasycznej PCA. Wielowymiarowa metoda dekonwolucji sygnałów w ważonym wariancie iteracyjnym najmniejszych kwadratów, MCR-WALS [54] oraz metoda pozytywnej faktoryzacji macierzy, PMF [55], są przykładami ważonej odmiany klasycznej iteracyjnej metody dekonwolucji sygnałów w wariancie iteracyjnym najmniejszych kwadratów, MCR-ALS [56, 57]. Do tej pory zwróciliśmy już uwagę na wykorzystanie metody MLPCA w celu przetwarzania danych z brakującymi elementami typu MCAR i elementami cenzorowanymi poniżej granicy oznaczalności/wykrywalności, ale uzasadnione jest dalsze zbadanie własności tychże metod w przypadku danych o wartościach błędów na różnych poziomach i z różną strukturą korelacyjną. W szczególności swoją uwagę poświęciłam porównaniu metod PMF i MCR-WALS dla danych o nieskorelowanej strukturze błędów pomiarowych [21] (artykuł H5 zawarty w aneksie opracowania). Metoda PMF jest rekomendowana przez Agencję Ochrony Środowiska Stanów Zjednoczonych jako narzędzie modelowania danych gromadzonych w trakcie kontroli jakości powietrza [58]. Metodę MCR-WALS skutecznie zastosowano do dekonwolucji źródeł zanieczyszczeń aerozoli [59] oraz analizy danych z mikromacierzy pozyskanych w trakcie eksperymentu realizowanego w czasie [54]. Metody PMF, MCR-WALS i MLPCA mają ten sam cel, ale algorytmy PMF i MCR-WALS zasadniczo różnią się od siebie. W przeciwieństwie do MLPCA, uwzględnienie warunku nieujemnego rozwiązania i brak warunku ortogonalności ukrytych zmiennych skonstruowanych za pomocą PMF i MCR-WALS daje możliwość uzyskania jednoznacznego lub prawie jednoznacznego rozwiązania, które można fizycznie interpretować. Uwzględnienie tychże warunków wewnątrz iteracyjnego algorytmu PMF wymaga ulepszenia funkcji kosztów stosując parametr regularyzacji i logarymiczny parametr kary. W rzeczywistości spełnienie na etapie konstrukcji ukrytych zmiennych w metodzie PMF innych warunków jak jednododalność, domknięcie, selektywność, lokalny rząd, równość lub nierówność jest zadaniem trudnym i wymaga większej złożoności procedury optymalizacyjnej dla wszystkich parametrów modelu. W trakcie badań

stwierdzono również, że szybkość uźbieźniania algorytmu i jakoć rozwiązania w dużej mierze zależą od przyjętego sposobu optymalizacji kroków algorytmu. W tym zakresie metoda MCR-WALS jest prostsza i bardziej wszechstronna. Wyniki uzyskane w ramach przeprowadzonych badań pozwalają wnioskować, że dwie metody działają praktycznie tak samo dla danych o różnych strukturach błędów. Ilustruje ten fakt Rys. 6, skonstruowany dla danych zawierających wyniki pomiarów obciążone nieskorelowanym i proporcjonalnym błędem pomiarowym.



Rys. 6 Wyniki uzyskane za pomocą metody PMF i MCR-WALS: a) profile wkładów i b) profile składu dla danych o nieskorelowanym i proporcjonalnym błędzie pomiarowym

Dodatkowo, w przypadku algorytmu MCR-WALS dokonano kolejnej jego modyfikacji [60], aby sprostać wymaganiom analizy tzw. danych zamkniętych. Dane zamknięte to szczególny typ danych chemicznych, w których ilościowa zawartość każdego z chemicznych składników jest wyrażona jako ułamek całkowitego składu, np. całkowita zawartość polifenoli, kwasów tłuszczowych czy składników mineralnych. Przestrzeń próbek tego typu danych opisana jest geometrią Aitchisona, a metody używane do ich analizy uwzględniają ich zamknięty charakter. Celem moich badań, szerzej dyskutowanym w artykule H6, dołączonym do tego opracowania, było zaproponowanie metodologii analizy danych zbudowanych z profili stężeniowych kwasów fosfolipidowych, PLFA, uzyskanych dla próbek wody w celu interpretacji przestrzennych i czasowych zmian obserwowanych w składzie biocenozy bakterii w zbiorniku wodnym. Do dnia dzisiejszego zaproponowano śledzenie źródeł niektórych form bakterii *coli* czy identyfikacji niektórych szczepów bakterii w oparciu o oznaczenie pojedynczych kwasów fosfolipidowych ignorując zarazem fakt, że te same kwasy tłuszczowe obecne są w różnej ilości w organizmach w zależności od warunków wzrostu. Ponadto, wykorzystanie nowoczesnej techniki analizy 16S rDNA aplikonów do identyfikacji patogenów wciąż pozostaje relatywnie drogim podejściem, aby można je stosować w rutynowej analizie. W celu interpretacji danych zamkniętych, zawierających całościowe profile PLFA, należy zdefiniować tzw. czyste wkłady ukrytych zmiennych opisujących profile PLFA, a następnie w celu dokonania identyfikacji należy porównać wkłady ukrytych zmiennych opisujących profile PLFA wybranych hodowlanych i niehodowlanych biocenozy bakterii. Dodatkowym utrudnieniem jest fakt, że zamknięte dane PLFA zawierają tzw. zaokrąglone zera, czyli elementy cenzorowane poniżej granicy oznaczalności/wykrywalności. Proponowany algorytm przewiduje ważenie każdej kolumny i każdego wiersza ukrytych wkładów i profili tak, aby uniknąć sytuacji, w której końcowy wynik uwidoczni

jedynie duże średnie zmiany pomiędzy zawartościami PLFA i ważenie każdego elementu stosując niepewność pomiarową by skutecznie przetwarzać elementy cenzorowane. Oczywiście warunki zamknięcia i nieujemności elementów danych są włączone w kroki iteracyjnego algorytmu najmniejszych kwadratów. Dodatkowo wymagana jest procedura oceny podobieństw pomiędzy każdym estymowanym składem PLFA i znanymi profilami PLFA dla każdej z rodzin bakterii, aby móc interpretować wyniki w oparciu o oznaczony skład biocenozy bakterii. Proponowane algorytmy zostały szczegółowo przedstawione w publikacji H6, dołączonej do opracowania. Zaproponowana metodologia została zwalidowana na podstawie wyników analizy 16S rDNA aplikonów kilku wybranych próbek wody. Główny wniosek wynikający z analizy uzyskanych wyników wskazuje, że połączenie relatywnie prostej i taniej metody analitycznej opierającej się na badaniach składu PLFA próbek wody wraz z zaawansowanymi algorytmami chemometrycznymi może polepszyć interpretację danych. Opracowaną metodykę z powodzeniem można stosować jako szybką, przesiewową procedurę w rutynowej kontroli jakości wody. W szczególności wykazano jej satysfakcjonującą czułość w stosunku do zmian w składzie biocenozy bakterii wpływających na jakość wody przed i po okresach powodzi.

### **3. Strategie chemometryczne zwiększania ekstrakcji chemicznej informacji z wielowymiarowych danych o indukowanej strukturze**

Stosowanie zasad planowania eksperymentu w protokołach analitycznych gwarantuje, że zgromadzone dane eksperymentalne będą reprezentowały wszystkie chemiczne czy biologiczne źródła wariacji związane z badanym efektem. Głównym celem moich badań było potwierdzenie bądź zaprzeczenie słuszności postulowanej hipotezy zerowej o wpływie rozważanych efektów. Zwyczajowo, sięga się po jedno-, dwu- albo N-czynnikową analizę wariacji, ANOVA [61], dla każdego oznaczonego ilościowo komponentu by oszacować istotność wpływu zmian stężeń/składu w przypadku różnych grup próbek. Na przykład, aby ocenić efekt leczenia przeciwnowotworowego cytarabiną w trakcie okresu ciąży na mineralizację kości żuchwy u nowonarodzonych szczurów, istotność zmian zawartości Zn w grupie leczonej i kontrolnej badano metodą ANOVA [62]. To podejście analizy danych nie uwzględnia możliwego istotnego wpływu pojedynczych komponentów na obserwowany efekt, lecz jedynie rozważa potencjalny istotny wpływ kombinacji chemicznych komponentów obecnych w próbkach na różnych poziomach. Z tego powodu należy sięgnąć po właściwą metodę analizy danych uwzględniającą indukowaną i silnie skorelowaną strukturę danych uzyskanych w trakcie planowanego eksperymentu. Takie możliwości zapewniają metody wieloparametrowe łączące w sobie statystyczne zalety metody ANOVA i analizy czynników głównych, PCA, czy jednoczesnej analizy czynnikowej, SCA [63], w celu przetwarzania danych skorelowanych. W ostatnich latach, w obszarze nauk o życiu szczególną popularność zyskały takie metody jak ANOVA-SCA (ASCA) [64] i ANOVA-PCA [65]. Metodologia analizy wielowymiarowych, wysoce ustrukturyzowanych danych obejmuje: i) wybór modelu zgodnie z określonym rodzajem zaplanowanego eksperymentu, ii) ocenę efektu czynników z użyciem jednoczynnikowego modelu ANOVA do każdej z kolumn (parametru) oryginalnych danych, gromadzenie tej informacji w odpowiednich macierzach czynników o tych samych wymiarach co oryginalne dane, iii) testowanie istotności efektu czynników i ostatecznie iv) dokonanie wizualizacji i interpretacji zgromadzonej informacji w indywidualnych macierzach czynników z zastosowaniem techniki redukcji wymiarowości danych. W zależności od specyfiki prowadzonych badań i stopnia złożoności planu eksperymentu, każdy z kroków ogólnej procedury powinien być odpowiednio

zmodyfikowany. Aby móc skutecznie przetwarzać niezbalansowane dane o strukturze indukowanej, konieczne jest dokonanie modyfikacji oryginalnego wariantu metody ASCA. Jej potrzeba wynika z braku spełnienia warunku ortogonalności przestrzeni kolumn macierzy czynników w przypadku niezbalansowanych danych, a w konsekwencji dekompozycja sumy kwadratów elementów jest nieadekwatna. Pociąga to także za sobą niepoprawne testowanie hipotez zerowych sformułowanych dla głównych czynników i ich interakcji. Dla danych niezbalansowanych i o indukowanej strukturze zaproponowano odpowiednią metodologię [14], by zbadać skutecznie efekt wywołany dawkowaniem cyklofosfamidu i cytarabiny, podawanych w trakcie okresu ciąży szczurom, na mineralizację kości żuchwy u 7-, 14- i 28-dniowych szczurów (zob. artykuł H7 w aneksie opracowania). Wybrano dwuczynnikowy ('wiek' i 'leczenie') kompletny trójpoziomowy plan eksperymentu dla każdego z czynników, a próbki opisano poprzez kompletne widma rentgenowskiej fluorescencji, XRF, bez zastosowania podejścia analizy ilościowej. Zgromadzone w trakcie eksperymentu dane były niezbalansowane, ponieważ każda grupa zróżnicowana ze względu na czynnik 'wiek' i czynnik 'leczenie' nie zawierała tej samej liczby próbek. Proponowanym rozwiązaniem tej kwestii jest uwzględnienie jedynie tej części wariancji, która jest charakterystyczna dla każdego czynnika przez co suma kwadratów każdego czynnika zostaje dostosowana do pozostałych. Innymi słowy, wykorzystuje się tzw. III rodzaj sumy kwadratów zamiast sekwencyjnego I rodzaju. Taka zmiana pociąga za sobą także różnice w procedurze testowania istotności czynników. Zamiast użycia I typu sumy kwadratów, dokonano porównania III typu sumy kwadratów dla danego czynnika stosując realne dane eksperymentalne z rozkładem III typu sumy kwadratów uzyskanego dla danych poprzez permutowanie porządku próbek. Stosując proponowane rozszerzenie podejścia ASCA wykazano, że dawkowanie cyklofosfamidu i cytarabiny w okresie ciąży wpływa na obniżenie zawartości potasu i cynku w kościach żuchwy u nowonarodzonych szczurów. Poziom potasu jest krytyczny na początkowych etapach rozwoju szkieletu (do 14 dni od urodzenia). Ponadto, ustalono kilka specyficznych różnic ze względu na proces mineralizacji kości żuchwy poszczególnych obszarów kości (różne zawartości potasu).

Statystyczne planowanie eksperymentu to także często wykorzystywana metodologia do optymalizacji produktów, które powinny sprostać oczekiwaniom konsumentów. Głównym celem prowadzonych w takim kontekście badań jest poznanie związku pomiędzy składem chemicznym złożonych mieszanin produktów a ich własnościami sensorycznymi (np. zapach i/lub smak). Często związki występujące w relatywnie małych stężeniach w mieszaninach próbek żywności mogą wywoływać bardzo silne wrażenie aromatu i smaku. Przeważnie wykorzystuje się zasady statystycznego planowania mieszanin by przygotować próbkę danego produktu, a następnie poddaje się ocenie smak i zapach przez panel ekspertów odpowiednio wyćwiczonych w ocenie starannie dobranego zestawu deskryptorów. Aby skutecznie osiągnąć cel prowadzonych badań, analiza danych powinna pozwolić uzyskać możliwie pełny obraz produktu wykorzystując do jego opisu informację o składzie chemicznym poszerzoną o dokonaną ocenę parametrów sensorycznych [19]. Co do zasady, poszukuje się dobrze określonej wspólnej reprezentacji próbek, która podkreśla indywidualne i wspólne własności zmiennych opisujących skład chemiczny i parametry sensoryczne. W literaturze opisano kilka podejść nadających się do zdefiniowania wspólnej (łączej) przestrzeni tych dwóch typów zmiennych. Celem prowadzonych przeze mnie badań, szerzej przedstawionych w artykule H8 znajdującej się w aneksie opracowania, było dokonanie analizy danych opisujących produkty reakcji Maillarda. W szczególności porównano działanie takich metod jak: zgodna analiza głównych składowych, CPCA [66], SUM-PCA [66], wieloraka analiza czynnikowa, MFA [67, 68] i metoda STATIS [69]. Wybrane aminokwasy i cukry redukujące zmieszano w odpowiednich proporcjach

stosując statystyczny plan mieszanin. Produkty ogrzewania mieszanin analizowano za pomocą chromatografii gazowej z techniką dozowania znad powierzchni próbki. Dodatkowo przeprowadzono analizę sensoryczną za pomocą panelu ekspertów pod kątem kilku wybranych deskryptorów zapachu. Wspomniany zestaw metod chemometrycznych został dobrany ze względu na różny sposób przetwarzania wariacji zmiennych i z uwzględnieniem efektu wielkości bloków zmiennych. Poszczególne algorytmy metod i ich własności szerzej opisano w artykule H8. Uzyskane wyniki analiz wskazują, że metody SUM-PCA i MFA są przydatne w sytuacji, gdy celem jest uzyskanie wspólnej struktury zestawu parametrów, które zawierają podobną informację, ale mają także unikane części. Metoda MFA w przypadku danych z reakcji Maillarda była preferowaną techniką analizy, ze względu na możliwość redukcji dominującego wpływu zmiennych danych chromatograficznych w tworzeniu kompromisowej reprezentacji mieszanin i możliwości otrzymania pełniejszej informacji z zastosowaniem hierarchicznej metody grupowania danych Warda, by wyjaśnić podobieństwa pomiędzy próbkami mieszanin w oparciu o parametry sensoryczne.

Celem badań realizowanych w ramach mojej pracy habilitacyjnej było opracowanie efektywnych chemometrycznych strategii analizy wielowymiarowych danych chemicznych polepszających ekstrakcję chemicznej informacji. W szczególności opracowałam nowe stabilne rozwinięcia wariantów takich metod jak PCA, SIMCA i PRM, które pozwalają na przetwarzanie danych zawierających brakujące elementy typu MCAR i wykazałam ich przydatność [39, 43, 48], a w oparciu o symulowane i realne zbiory danych dokonałam analizy ich własności uzbieźniania. Zaproponowałam również skuteczną strategię przetwarzania wielowymiarowych danych zawierających zarówno brakujące elementy typu MCAR jak i elementy cenzorowane poniżej granicy oznaczalności/wykrywalności [50]. Udowodniłam, że metody MLPCA, EM-PCA i ogólny algorytm NIPALS są tożsame w działaniu w przypadku analizy danych zawierających brakujące elementy typu MCAR. W oparciu o wyczerpującą analizę porównawczą wykazałam, że wyniki uzyskane metodami PMF i MCR-WALS są tożsame dla danych z różną strukturą błędu pomiarowego [21]. Ponadto, zwróciłam uwagę na zdecydowanie większą wszechstronność algorytmu MCR-WALS w porównaniu do algorytmu PMF. Dokonałam dodatkowej modyfikacji algorytmu MCR-WALS [60], aby rozszerzyć zakres jego stosowalności o analizę danych zamkniętych. Przydatność i czułość nowej metody zilustrowałam na przykładzie danych zamkniętych, opisujących zmiany jakościowe w biocenozach bakterii przed i po okresie powodzi. W końcowym etapie pracy, zaproponowałam ogólne podejście przetwarzania informacji zawartej w danych niezbalansowanych o indukowanej strukturze [14]. Gdy nadrzędnym celem jest badanie wspólnej struktury różnych zbiorów parametrów (rejestrowanych za pomocą różnych platform analitycznych), które mają pewien zasób wspólnej informacji, a także posiadają pewne części unikalne, skutecznym podejściem będzie użycie metody MFA wzbogacającej możliwości interpretacyjne wieloblokowych danych [19]. Opracowane strategie zostały użyte w różnych badaniach o charakterze aplikacyjnym [5, 70].

Pomimo relatywnie dużych osiągnięć w zakresie rozwijania stabilnych metod analizy i modelowania niekompletnych danych, nadal istnieje konieczność opracowywania stabilnych metod dedykowanych analizie wielowymiarowych danych, zawierających elementy cenzorowane poniżej granicy oznaczalności/wykrywalności cenzorowane. Ponadto niezwykle pożądane są również algorytmy, które efektywnie i szybko dają możliwość przetwarzania dużych wielowymiarowych i zarazem niekompletnych danych. Nowe obszary zwiększania funkcjonalności metod obejmują próby usprawniania narzędzi dyskryminacyjnych, które powinny uwzględniać niepewność pomiarową jak np. probabilistyczny wariant metody SIMCA, probabilistyczny wariant metody LDA. Tego typu

rozwiązania w znaczącym stopniu wsparłyby interpretację niezwykle złożonych danych jak na przykład dane pozyskiwane za pomocą platform mikromacierzy DNA. Szerokim i zarazem interesującym polem do dalszych badań są metody stabilne, których kierunki modyfikacji powinny wspierać analizę wielowymiarowych niekompletnych danych o charakterze zamkniętym. W szczególności, niezwykle istotnym problemem w analizie danych o charakterze zamkniętym wymagającym rozwiązania jest skuteczne rozstrzygnięcie kwestii tych elementów, które są poniżej granicy wykrywalności/oznaczalności.

#### 4. Literatura

- [1] D. Harvey, *Modern Analytical Chemistry*, The McGraw-Hill Companies Inc., 2000, The USA.
- [2] I. Stanimirova, B. Üstün, T. Cajka, K. Ridelova, J. Hajslova, L.M.C. Buydens, B. Walczak, Tracing the geographical origin of honeys based on volatile compounds profiles assessment using pattern recognition techniques, *Food Chem.* 118 (2010) 171-176.
- [3] I. Stanimirova, M. Kazura, D. de Beer, E. Joubert, A.E. Schulze, T. Beelders, A. de Villiers, B. Walczak, High-dimensional nested analysis of variance to assess the effect of production season, quality grade and steam pasteurization on the phenolic composition of fermented rooibos herbal tea, *Talanta* 115 (2013) 590-599.
- [4] I. Stanimirova, A. Kita, E. Malkowski, E. John, B. Walczak, N-way exploration of environmental data obtained from sequential extraction procedure, *Chemometr. Intell. Lab.* 96 (2009) 203-209.
- [5] A. Zabek, I. Stanimirova, S. Deja, W. Barg, A. Kowal, A. Korzeniewska, M. Orczyk-Pawłowicz, D. Baranowski, Z. Gdaniec, R. Jankowska, P. Młynarz, Fusion of the <sup>1</sup>H NMR data of serum, urine and exhaled breath condensate in order to discriminate chronic obstructive pulmonary disease and obstructive sleep apnea syndrome, *Metabolomics* 11 (2015) 1563-1574.
- [6] D.L. Massart, B.G.M. Vandeginste, S.N. Deming, Y. Michotte, L. Kaufman, *Chemometrics: a textbook*, Elsevier Science Publishers B.V., Amsterdam, The Netherlands, 1988.
- [7] B.K. Lavine, J. Workman Jr., *Chemometrics*, *Anal. Chem.* 82 (2010) 4699-4711.
- [8] B.K. Lavine, J. Workman Jr., *Chemometrics*, *Anal. Chem.* 85 (2013) 705-714.
- [9] S.C. Rutan, A. de Juan, R. Tauler, *Introduction to Multivariate Curve Resolution in Comprehensive Chemometrics*, S.D. Brown, R. Tauler, B. Walczak (Eds.), Elsevier, Amsterdam, The Netherlands, Vol. 2, 2009, pp 249.
- [10] M. Daszykowski, B. Walczak, Use and abuse of chemometrics in chromatography, *TrAC* 25 (2006) 1081-1096.
- [11] B. Krakowska, I. Stanimirova, J. Orzel, M. Daszykowski, I. Grabowski, G. Zaleszczyk, M. Sznajder, Detection of discoloration in diesel fuel based on gas chromatographic fingerprints, *Anal. Bioanal. Chem.* 407 (2015) 1159-1170.
- [12] C.D. Stalikas, G.A. Pilidis, Development of a method for the simultaneous determination of phosphoric and amino acid group containing pesticides by gas chromatography with mass-selective detection. Optimization of the derivatization procedure using an experimental design approach, *J. Chromatogr. A* 872 (2000) 215-225.

- [13] J. Orzel, I. Stanimirova, B. Czarnik-Matusiewicz, M. Daszykowski, Prediction of hydrophilic antioxidant capacity of tomato pastes from the IR and fluorescence excitation-emission spectra of extracts and intact samples, *Talanta* 138 (2015) 64-70.
- [14] I. Stanimirova, K. Michalik, Z. Drzazga, H. Trzeciak, P.D. Wentzell, B. Walczak, Interpretation of ANOVA models using PCA to assess the effect of a maternal anticancer treatment on the mineralization of rat bones, *Anal. Chim. Acta* 689 (2011) 1-7.
- [15] S. Rouse, V. Bellon-Maurel, J.M. Roger, P. Grenier, Fusion of aroma, FT-IR and UV sensor data based on the Bayesian inference. Application to the discrimination of white grape varieties, *Chemometr. Intell. Lab.* 65 (2003) 209-219.
- [16] R. Bro, H.J. Nielsen, F. Savorani, K. Kjeldahl, I.J. Christensen, N. Brunner, A.J. Lawaetz, Data fusion in metabolomic cancer diagnostics, *Metabolomics* 9 (2013) 3-8.
- [17] A.K. Smilde, M.J. van der Werf, S. Bijlsma, B.J.C. van der Werff-van der Vat, R.H. Jellema, Fusion of mass spectrometry-based metabolomics data, *Anal. Chem.* 77 (2005) 6729-6736.
- [18] K. Van Deun, A.K. Smilde, M.J. van der Werf, H.A.L. Kiers, I. Van Mechelen, A structured overview of simultaneous component based data integration, *BMC Bioinformatics* 10 (2009) 246-260.
- [19] I. Stanimirova, C. Boucon, B. Walczak, Relating gas chromatographic profiles to sensory measurements describing the end products of the Maillard reaction, *Talanta* 83 (2011) 1239-1246.
- [20] P.D. Wentzell, Other topics in soft-modeling: maximum likelihood-based soft-modeling methods, in *Comprehensive Chemometrics*, S.D. Brown, R. Tauler, B. Walczak (Eds.) Elsevier, Amsterdam, The Netherlands, Vol. 2, 2009, pp. 507.
- [21] I. Stanimirova, R. Tauler, B. Walczak, A comparison of positive matrix factorization and the weighted multivariate curve resolution method. Application to environmental data, *Environ. Sci. Technol.* 45 (2011) 10102-10110.
- [22] P. Rousseeuw, M. Leroy, *Robust Regression and Outlier Detection*, John Wiley and Sons, Inc., The USA, 1987.
- [23] D.B. Rubin, Inference and missing data, *Biometrika* 63 (1976) 581-592.
- [24] D. Helsel, *Nondetects and data analysis. Statistics for censored environmental data*, John Wiley and Sons Inc., The USA, 2005.
- [25] R.J.A. Little, D.B. Rubin, *Statistical Analysis with Missing Data*, John Wiley and Sons Inc., The USA, 1987.
- [26] A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum likelihood for incomplete data via the EM algorithm (with discussions). *J. R. Statist. Soc. B* 39 (1977) 1-38.
- [27] M.E. Tipping, C.M. Bishop, Probabilistic principal component analysis, *J. R. Statist. Soc. B* 61 (1999) 611-622.
- [28] B. Walczak, D.L. Massart, Dealing with missing data. Part I, *Chemometr. Intell. Lab.* 58 (2001) 15-27.
- [29] C. Croux, A. Ruiz-Gazen, A fast algorithm for robust principal components based on Projection Pursuit, *COMPSTAT: Proceedings in Computational Statistics 1996*, Physica-Verlag, Heidelberg, 1996, pp. 211-217.



- [30] M. Hubert, P. Rousseeuw, S. Verboven, A fast method for robust principal components with application to chemometrics, *Chemometr. Intell. Lab.* 60 (2002) 101-111.
- [31] R. Maronna, Principal components and orthogonal regression based on robust scales, *Technometrics* 47 (2005) 264-273.
- [32] S. Serneels, C. Croux, P. Filzmoser, P.J. Van Espen, Partial robust M regression *Chemometr. Intell. Lab.* 79 (2005) 55-64.
- [33] S. Serneels, E. De Nolf, P.J. Van Espen, Spatial sign preprocessing: a simple way to import moderate robustness to multivariate estimators, *J. Chem. Inf. Model* 46 (2006) 1402-1409.
- [34] A. Smolinski, B. Walczak, J.W. Einax, Exploratory analysis of data sets with missing elements and outliers, *Chemosphere* 49 (2002) 233-245.
- [35] M. Hubert, K. Vanden Branden, Robust methods for partial least squares regression, *J. Chemometr.* 17 (2003) 537-549.
- [36] M. Daszykowski, K. Kaczmarek, I. Stanimirova, Y. Vander Heyden, B. Walczak, Robust SIMCA-bounding influence of outliers, *Chemometr. Intell. Lab.* 87 (2007) 121-129.
- [37] M. Hubert, J. Van Kerckhoven, T. Verdonck, Robust PARAFAC for incomplete data, *J. Chemometr.* 26 (2012) 290-298.
- [38] S. Serneels, T. Verdonck, Principal component regression for data containing outliers and missing elements, *Comput. Stat. Data An.* 53 (2009) 3855-3863.
- [39] I. Stanimirova, M. Daszykowski, B. Walczak, Dealing with missing values and outliers in principal component analysis, *Talanta* 72 (2007) 172-178.
- [40] N. Locantore, J.S. Marron, D.G. Simpson, N. Tripoli, J.T. Zhang, K.L. Cohen, *Sociedad de Estadística e Investigación Operativa, Test* 8 (1999) 1-74.
- [41] B.G.M. Vandeginste, D.L. Massart, L.M.C. Buydens, S. De Jong, P.J. Lewi, J. Smeyers-Verbeke, *Handbook of Chemometrics and Qualimetrics: Part B*, Elsevier, Amsterdam, The Netherlands, 1998.
- [42] S. Wold, Pattern recognition by means of disjoint principal component models, *Pattern Recogn.* 8 (1976) 127-139.
- [43] I. Stanimirova, B. Walczak, Classification of data with missing elements and outliers, *Talanta* 76 (2008) 602-609.
- [44] H. Martens, T. Naes, *Multivariate Calibration*, John Wiley & Sons, Gulford, The Great Britain, 1989.
- [45] R.W. Kennard, L.A. Stone, Computer Aided Design of Experiments, *Technometrics* 11 (1969) 137-148.
- [46] R.D. Snee, Validation of regression models: methods and examples, *Technometrics* 19 (1977) 415-428.
- [47] F. Arteaga, A. Ferrer, Dealing with missing data in MSPC: several methods, different interpretations, some examples, *J. Chemometr.* 16 (2002) 408-418.
- [48] I. Stanimirova, S. Serneels, P.J. Van Espen, B. Walczak, How to construct a multiple regression model for data with missing elements and outlying objects, *Anal. Chim. Acta* 581 (2007) 324-332.

- [49] S. Serneels, T. Verdonck, Principal component analysis for data containing outliers and missing elements, *Comput. Stat. Data An.* 52 (2008) 1712-1727.
- [50] I. Stanimirova, Practical approaches to principal component analysis for simultaneously dealing with missing and censored elements in chemical data, *Anal. Chim. Acta* 796 (2013) 27-37.
- [51] P.D. Wentzell, D.T. Andrews, D.C. Hamilton, K. Faber, B.R. Kowalski, Maximum likelihood principal component analysis, *J. Chemometr.* 11 (1997) 339-366.
- [52] D.T. Andrews, P.D. Wentzell, Applications of maximum likelihood principal component analysis: incomplete data sets and calibration transfer, *Anal. Chim. Acta* 350 (1997) 341-352.
- [53] J. Ramon, F. Costa, Handling missing values and censored data in PCA of pharmacological matrices, *StReBio'09*, ACM 978-1-60558-667-0, Paris, France, 2009.
- [54] P.D. Wentzell, T.K. Karakach, S. Roy, M.J. Martinez, C.P. Allen, M. Werner-Washburne, Multivariate curve resolution of time course microarray data, *BMC Bioinformatics* 7 (2006) 343-361.
- [55] P. Paatero, U. Tapper, Positive Matrix Factorization: A non-negative factor model with optimal utilization of error estimates of data values, *Environmetrics* 5 (1994) 111-126.
- [56] S.C. Rutan, A. de Juan, R. Tauler, Introduction to multivariate curve resolution in *Comprehensive Chemometrics*: S.D. Brown, R. Tauler, B. Walczak (Eds.) Elsevier, Amsterdam, The Netherlands, 2009, Vol. 2, pp. 249.
- [57] A de Juan, R. Tauler, Chemometrics applied to unravel multicomponent processes and mixtures. Revisiting latest trends in multivariate resolution, *Anal. Chim. Acta* 500 (2003) 195-210.
- [58] EPA Positive Matrix Factorization (PMF), Fundamentals and User Guide v. 3.0.
- [59] R. Tauler, M. Viana, X. Querol, A. Alastuey, R.M. Flight, P.D. Wentzell, P.K. Hopke, Comparison of the results obtained by four receptor modelling methods in aerosol source apportionment studies, *Atmos. Environ.* 43 (2009) 3989-3997.
- [60] I. Stanimirova, A. Woznica, T. Plociniczak, M. Kwasniewski, J. Karczewski, A modified weighted mixture model for the interpretation of spatial and temporal changes in the microbial communities in drinking water reservoirs using compositional phospholipid fatty acid data, *Talanta* 160 (2016) 148-156.
- [61] R.A. Johnson, D.W. Wichern, *Applied multivariate statistical analysis*, Sixth Edition, Pearson Education Inc., Upper Saddle River, New Jersey, The USA, 2007.
- [62] Z. Drzazga, K. Michalik, K. Maciejewska, H. Trzeciak, M. Kaszuba, BioFactors Role of endogenous zinc in bones of newborn rats, *BioFactors* 30 (2007) 243-248.
- [63] J.J. Jansen, H.C.J. Hoefsloot, J. van der Greef, M.E. Timmerman, A.K. Smilde, Multilevel Component Analysis of time-resolved metabolomics data, *Anal. Chim. Acta* 530 (2005) 173-183.
- [64] J.J. Jansen, H.C.J. Hoefsloot, J. Van der Geert, M.E. Timmerman, J.A. Westerhuis, A.K. Smilde, ASCA: analysis of multivariate data obtained from an experimental design, *J. Chemometr.* 19 (2005) 469-481.
- [65] P.B. Harrington, N.E. Vieira, J. Espinoza, J. Kae Nien, R. Romero, A. Yergey, Analysis of variance-principal component analysis: A soft tool for proteomic discovery, *Anal. Chim. Acta* 544 (2005) 118-127.

- [66] A.K. Smilde, J. Westerhuis, S. de Jong, A framework for sequential multiblock component methods, *J. Chemometr.* 17 (2003) 323-337.
- [67] B. Escofier, J. Pages, *Analyses Factorielles Simples et Multiples: Objectifs, Methodes, Interpretations*, Dunod, Paris, 1990.
- [68] B. Escofier, J. Pagès, Multiple factor analysis (AFMULT package), *Comput. Stat. Data An.* 18 (1994) 121-140.
- [69] Ch. Lavit, Y. Escoufier, R. Sabatier, P. Traissac, The ACT (STATIS method), *Comput. Stat. Data An.* 18 (1994) 97-119.
- [70] M. Horochowska, I. Stanimirova, B. Czarnik-Matusiewicz, Studying the influence of enflurane, isoflurane, and sevoflurane on the DPPC lipid bilayer using the analysis of variance and parallel factor analysis, *Chemometr. Intell. Lab.* 153 (2016) 146–152.

## 6. Omówienie pozostałych osiągnięć naukowo-badawczych

### 6.1 Wykaz pozostałych artykułów

Lp.	Publikacja	IF	L. pkt. <sup>3</sup>
1.	I. Stanimirova, S. Tsakovski, V. Simeonov, Multivariate statistical analysis of coastal sediment data, <i>Fresenius Journal of Analytical Chemistry</i> , 365 (1999) 489-493	1,428	
2.	V. Simeonov, I. Stanimirova, S. Tsakovski, Multivariate statistical interpretation of coastal sediment monitoring data, <i>Fresenius Journal of Analytical Chemistry</i> , 370 (2001) 719-722	1,649	
3.	V. Simeonov, J.W. Einax, I. Stanimirova, J. Kraft, Environmetric modeling and interpretation of river water monitoring data, <i>Analytical and Bioanalytical Chemistry</i> , 374 (2002) 898-905	1,948	
4.	I. Stanimirova, P. Simeonova, V. Simeonov, Assessment of the Atmospheric Pollution in the Industrial Region of Devnia, Bulgaria, <i>Chemical and Engineering Ecology</i> , 10 (2003) 381-388	0	
5.	I. Stanimirova, B. Walczak, D.L. Massart, V. Simeonov, A comparison between two robust PCA algorithms, <i>Chemometrics and Intelligent Laboratory Systems</i> , 71 (2004) 83-95	1,889	
6.	I. Stanimirova, B. Walczak, D.L. Massart, V. Simeonov, C.A. Saby, E. Di Crescenzo, STATIS, a 3-way method for data analysis. Application to environmental data, <i>Chemometrics and Intelligent Laboratory Systems</i> , 73 (2004) 219-233	1,899	
7.	I. Stanimirova, M. Daszykowski, D.L. Massart, F. Questier, V. Simeonov, H. Puxbaum, Chemometrical Exploration of the Wet Precipitation Chemistry from the Austrian Monitoring Network (1988-1999), <i>Journal of Environmental Management</i> , 74 (2005) 349-363	1,163	
8.	I. Stanimirova, B. Walczak, D.L. Massart, Multiple Factor Analysis in environmental chemistry, <i>Analytica Chimica Acta</i> , 545 (2005) 1-12	2,760	
9.	I. Stanimirova, V. Simeonov, Modeling of environmental four-way data from air quality control, <i>Chemometrics and Intelligent Laboratory Systems</i> , 77 (2005) 115-121	1,770	
10.	I. Stanimirova, M. Daszykowski, E. Van Gyseghen, F.F. Bensaid, M. Lees, J. Smeyers-Verbeke, D.L. Massart, Y. Vander Heyden, Chemometrical exploration of an isotopic ratio pharmaceutical data set of acetylsalicylic acid, <i>Analytica Chimica Acta</i> , 552 (2005) 1-12	2,760	

---

<sup>3</sup> Liczba punktów obowiązująca w roku ukazania się publikacji w danym czasopiśmie zgodna z wykazem czasopism naukowych sporządzonym przez Ministra Nauki i Szkolnictwa Wyższego

11. M. Daszykowski, I. Stanimirova, B. Walczak, D. Coomans, Explaining a presence of groups in analytical data in terms of original variables, *Chemometrics and Intelligent Laboratory Systems*, 78 (2005) 19-29 1,770
12. M. Daszykowski, I. Stanimirova, B. Walczak, Q.-S. Xu, F. Daeyaert, M.R. de Jonge, J. Heeres, L.M.H. Koymans, P.J. Lewi, H.M. Vinkers, P.A. Janssen, D.L. Massart, Improving models for biological activity of HIV Reverse Transcriptase inhibitors: aspects of outlier detection and variable selection, *Talanta*, 68 (2005) 54-60 2,391
13. S. Gourvenec, I. Stanimirova, C.A. Sabi, C.Y. Airiau, D.L. Massart, Monitoring batch processes with the STATIS approach, *Journal of Chemometrics*, 19 (2005) 288-300 1,875 35
14. I. Stanimirova, K. Zehl, D.L. Massart, Y. Vander Heyden, J.W. Einax, Chemometric analysis of soil pollution data applying the Tucker N-way method, *Analytical and Bioanalytical Chemistry*, 385 (2006) 771-773 2,591 40
15. M. Daszykowski, K. Kaczmarek, I. Stanimirova, Y. Vander Heyden, B. Walczak, Robust SIMCA - bounding influence of outliers, *Chemometrics and Intelligent Laboratory Systems*, 87 (2007) 121-129 2,053 40
16. M. Daszykowski, I. Stanimirova, A. Bodzon-Kulakowska, J. Silberring, G. Lubec, B. Walczak, The start-to-end processing of two-dimensional gel electrophoretic images, *Journal of Chromatography A*, 1158 (2007) 306-317 3,641 40
17. I. Stanimirova, M. Polowniak, R. Skorek, A. Kita, E. John, F. Buhl, B. Walczak, Chemometric analysis of the water purification process data, *Talanta*, 74 (2007) 153-162 3,374 40
18. I. Stanimirova, A. Kubik, B. Walczak, J.W. Einax, Discrimination of biofilm samples using pattern recognition techniques, *Analytical and Bioanalytical Chemistry*, 390 (2008) 1273-1282 3,328 35
19. E. Deconinck, A.M. van Nederkassel, I. Stanimirova, M. Daszykowski, F. Bensaid, M. Lees, G.J. Martin, J.R. Desmurs, J. Smeyers-Verbeke, Y. Vander Heyden, Isotopic ratios to detect infringements of patents or proprietary processes of pharmaceuticals: Two case studies, *Journal of Pharmaceutical and Biomedical Analysis*, 48 (2008) 27-41 2,629 30
20. I. Stanimirova, A. Kita, E. Malkowski, E. John, B. Walczak, N-way exploration of environmental data obtained from sequential extraction procedure, *Chemometrics and Intelligent Laboratory Systems*, 96 (2009) 203-209 2,111 40
21. M. Dumarey, A.M. van Nederkassel, I. Stanimirova, M. Daszykowski, F. Bensaid, M. Lees, G.J. Martin, J.R. Desmurs, J. Smeyers-Verbeke, Y. Vander Heyden, Recognizing paracetamol formulations with the same synthesis pathway based on their trace-enriched chromatographic impurity profiles, *Analytica Chimica Acta*, 655 (2009) 43-51 3,757 40

22. I. Stanimirova, B. Üstün, T. Cajka, K. Riddellova, J. Hajslova, L.M.C. Buydens, B. Walczak, Tracing the geographical origin of honeys using the GCxGC-MS and pattern recognition techniques, *Food Chemistry*, 118 (2010) 171-176 3,458 45
23. A.J. Charlton, M.S. Wrobel, I. Stanimirova, M. Daszykowski, H. Grundy, B. Walczak, Multivariate discrimination of wines with respect to their grape varieties and vintages, *European Food Research and Technology*, 231 (2010) 733-743 1,5857 30
24. J. Polak, M. Bartoszek, I. Stanimirova, A study of the antioxidant properties of beers using electron paramagnetic resonance, *Food Chemistry*, 141 (2013) 3042-3049 3,655 45
25. I. Stanimirova, M. Kazura, D. de Beer, E. Joubert, A.E. Schulze, T. Beelders, A. de Villiers, B. Walczak, High-dimensional nested analysis of variance to assess the effect of production season, quality grade and steam pasteurization on the phenolic composition of fermented rooibos herbal tea, *Talanta*, 115 (2013) 590-599 3,498 40
26. A. Martyna, G. Zadora, I. Stanimirova, D. Ramos, Wine authenticity verification as a forensic problem. An application of likelihood ratio test to label verification, *Food Chemistry*, 150 (2014) 287-295 3,259 40
27. S. Deja, I. Porębska, A. Kowal, A. Ząbek, W. Barg, K. Pawełczyk, I. Stanimirova, M. Daszykowski, A. Korzeniewska, R. Jankowska, P. Młynarz, Metabolomics provides new insights on lung cancer staging and discrimination from chronic obstructive pulmonary disease, *Journal of Pharmaceutical and Biomedical Analysis*, 100 (2014) 369-380 2,829 35
28. A. Białek, A. Stawarska, A. Tokarz, K. Czuba, A. Konarska, M. Mazurkiewicz, I. Stanimirova, Influence of maternal diet enrichment with Conjugated Linoleic Acids on fatty acids profile and delta-6 and delta-5 desaturase activity in livers and hepatic microsomes of their offspring with 7,12-dimethylbenz[a]anthracene induced mammary tumours, *Acta Poloniae Pharmaceutica*, 71 (2014) 747-761 0,693 15
29. M. Orłowska, I. Stanimirova, D. Staszek, M. Sajewicz, T. Waksmundzka-Hajnos, T. Kowalska, Optimization of extraction based on the thin-layer chromatographic fingerprints of common thyme, *Journal of AOAC International*, 97 (2014) 1274-1281 1,340 25
30. A. Stawarska, A. Białek, I. Stanimirova, T. Stawarski, A. Tokarz, The effect of conjugated linoleic acids (CLA) supplementation on the activity of enzymes participating in the formation of arachidonic acid in liver microsomes of rats – probable mechanism of CLA anticancer activity, *Nutrition and Cancer: An International Journal*, 67 (2015) 145-155 2,322 25

- |     |   |       |    |
|-----|---|-------|----|
| 31. | B. Krakowska, I. Stanimirova, J. Orzel, M. Daszykowski, I. Grabowski, G. Zaleszczyk, M. Sznajder, Detection of discoloration in diesel fuel based on gas chromatographic fingerprints, <i>Analytical and Bioanalytical Chemistry</i> , 407 (2015) 1159-1170   | 3,436 | 35 |
| 32. | J. Orzel, I. Stanimirova, B. Czarnik-Matusewicz, M. Daszykowski, Prediction of hydrophilic antioxidant capacity of tomato pastes from the IR and fluorescence excitation-emission spectra of extracts and intact samples, <i>Talanta</i> , 138 (2015) 64–70   | 3,545 | 40 |
| 33. | A. Ząbek, I. Stanimirova, S. Deja, W. Barg, A. Kowal, A. Korzeniewska, M. Orczyk-Pawiłowicz, D. Baranowski, Z. Gdaniec, R. Jankowska, P. Młynarz, Fusion of the <sup>1</sup> HNMR data of serum, urine and exhaled breath condensate in order to discriminate chronic obstructive pulmonary disease and obstructive sleep apnea syndrome, <i>Metabolomics</i> , 11 (2015) 1563-1574 | 3,855 | 30 |
| 34. | M. Horochowska, I. Stanimirova, B. Czarnik-Matusewicz, Studying the influence of Enflurane, Isoflurane and Sevoflurane on the DPPC lipid bilayer using the analysis of variance and parallel factor analysis, <i>Chemometrics and Intelligent Laboratory Systems</i> , 153 (2016) 146-152   | 2,217 | 40 |

Razem: **82,4787**

## 6.2 Wykaz rozdziałów książkowych

1. I. Stanimirova, M. Daszykowski, B. Walczak, Robust methods in analysis of multivariate food chemistry data, w: F. Marini, Ed., *Chemometrics in Food Chemistry*, vol. 28, Chennai: Elsevier, 2013, pp. 315-340 (ISBN 978-0-444-59528-7)
2. M. Daszykowski, I. Stanimirova, B. Walczak, Robust methods in QSAR, w: *Recent Advances in QSAR Studies: Methods and Applications*, M.T.D. Cronin, J. Leszczyński i T. Puzyn (Eds.), Springer, 2010 (ISBN 978-1-4020-9782-9)
3. I. Stanimirova, M. Daszykowski, B. Walczak, *Metody uczenia z nadzorem - kalibracja, dyskryminacja i klasyfikacja*, w: A. Parczewski i D. Zuba, Eds., *Chemometria w analityce*, IES, Kraków, 2008 (ISBN 83-87425-13-3)

## 6.3 Lista wykładów na zaproszenie imienne wygłoszonych na konferencjach krajowych oraz zagranicznych

1. I. Stanimirova, Electrochemical deposition of copper onto n- and p- type silicon substrates Max Planck Institute, Berlin (październik 2000 r.), Niemcy
2. I. Stanimirova, D.L. Massart, V. Simeonov, New chemometric approaches in environmental data mining, Technical University, Wiedeń (maj 2002 r.), Austria

3. I. Stanimirova, D.L. Massart, Exploratory Analysis of Environmental Data, Friedrich Schiller University, Jena (maj 2003 r.) Niemcy
4. I. Stanimirova, B. Walczak, Relating chromatographic data to sensory measurements, The 31st Conference on "Chromatographic methods of investigating the organic compounds", Szczyrk (6.06.2007 r.)
5. I. Stanimirova, Data fusion - enhancement of systems interpretability, National Research Council, Halifax (21.08.2009 r.), Kanada
6. I. Stanimirova, B. Walczak, Simultaneous processing of data from multiple sources, Seventh Winter Symposium on Chemometrics, Petersburg (17.02.2010 r.), Rosja
7. I. Stanimirova, Interpretation of experimental results with the use of chemometric methods, Seminar of Department of Medical Physics and Polish Society of Medical Physics, Instytut Fizyki, Uniwersytet Śląski w Katowicach, Katowice (15.03.2010 r.)
8. I. Stanimirova, Dealing with missing elements, outliers and measurement uncertainty of environmental data, Euroanalysis 2011, Belgrad (12.09.2011 r.), Serbia (keynote speaker)

#### **6.4 Lista wykładów wygłoszonych na konferencjach krajowych i zagranicznych**

1. I. Stanimirova, B. Walczak, D.L. Massart, Robust principal component analysis, ChemoAC meeting, Wolny Uniwersytet Brukselski, Bruksela (21-22.11.2002 r.), Belgia
2. I. Stanimirova, B. Walczak, D.L. Massart, STATIS, an alternative for 3-way data analysis, ChemoAC meeting, Wolny Uniwersytet Brukselski, Bruksela (15-16.01.2004 r.), Belgia
3. M. Daszykowski, I. Stanimirova, B. Walczak, F. Dayaert, P. Lewi, D.L. Massart, Robust modeling of Reverse Transcriptase HIV inhibitors, ChemoAC meeting, Wolny Uniwersytet Brukselski, Bruksela (10-11.06.2004 r.), Belgia
4. S. Gourvenec, I. Stanimirova, D.L. Massart, Monitoring outlying batch process trajectories with "STATIS approach", ChemoAC meeting, Wolny Uniwersytet Brukselski, Bruksela (10-11.06.2004 r.), Belgia
5. S. Gourvenec, I. Stanimirova, C-A. Saby, C.Y. Airiau, D.L. Massart, Monitoring batch processes with the STATIS approach, "Chimiométrie 2004" kongres, Paryż (30.11-1.12.2004 r.), Francja
6. I. Stanimirova, B. Walczak, D.L. Massart, Multiple Factor Analysis, ChemoAC meeting, Wolny Uniwersytet Brukselski, Bruksela (9-10.12.2004 r.), Belgia
7. I. Stanimirova, B. Walczak, D.L. Massart, Y. Vander Heyden, Exploratory analysis of environmental data, ChemoAC meeting, Wolny Uniwersytet Brukselski, Bruksela (16-17.06.2005 r.), Belgia
8. B. Walczak, I. Stanimirova, K. Kaczmarek, SIMCA and robust SIMCA for TRACE Toolbox, TRACE meeting, Tarragona (16-17.03.2006 r.), Hiszpania
9. B. Walczak, I. Stanimirova, K. Kaczmarek, Exploration of data with missing elements and outliers, TRACE meeting, Tarragona (16-17.03.2006 r.), Hiszpania



10. I. Stanimirova, K. Kaczmarek, M. Daszykowski, B. Walczak, Robust exploratory data analysis dealing with outliers, missing values and censored data, The 2nd annual TRACE meeting, Praga (24-26.04.2006 r.), Republika Czeska
11. M. Daszykowski, I. Stanimirova, B. Walczak, Classification of samples using chromatographic data, The 30th Symposium "Chromatographic methods of investigating the organic compounds", Szczyrk (13.06.2006 r.)
12. M. Daszykowski, I. Stanimirova, B. Walczak, From Projection Pursuit to PCA, robust PCA, ICA and other approaches, The 10th International Conference on Chemometrics in Analytical Chemistry, Aguas de Lindoia (14.09.2006 r.), San Paulo, Brazylia
13. I. Stanimirova, B. Walczak, N-way analysis of environmental data, The 3rd Conference: "Chemometrics - Methods and Applications", Zakopane (20.10.2006 r.)
14. I. Stanimirova, B. Walczak, Dealing with missing data, The 3rd Conference: "Chemometrics - Methods and Applications", Zakopane (20.10.2006 r.)
15. B. Walczak, I. Stanimirova, Analysis of mineral water data, Trace WP6 meeting, Genua (28.09.2006 r.), Włochy
16. M. Michalska, I. Stanimirova, M. Daszykowski, B. Walczak, Zastosowanie metod chemometrycznych do eksploracji danych z kontroli jakości wody pitnej, Wiosenne konwersatorium, Instytut Chemii, Uniwersytet Śląski w Katowicach, Katowice (4.06.2007 r.)
17. B. Walczak, I. Stanimirova, Tracing origin of honeys and olive oils, Trace WP6 meeting, Ljubljana (7.03.2007 r.), Słowenia
18. I. Stanimirova, M. Daszykowski, B. Walczak, Metody uczenia z nadzorem - kalibracja, dyskryminacja i klasyfikacja, The 4th Conference: "Chemometrics – Methods and Applications", Zakopane (24.10.2008 r.)
19. I. Stanimirova, B. Walczak, Practical approaches for dealing with missing and censored data, The 13th International Conference on Chemometrics in Analytical Chemistry, Budapeszt (27.06.2012 r.), Węgry
20. I. Stanimirova, High dimensional nested analysis of variance of HPLC-DAD fingerprints to assess the effect of production season, quality grade and steam pasteurization on the phenolic composition of fermented rooibos herbal tea, The 36th Symposium on "Chromatographic methods of investigating the organic compounds", Szczyrk (5.06.2013 r.)
21. A. Martyna, G. Zadora, I. Stanimirova, D. Ramos, A. Parczewski, Verification of food products authenticity supported by chemometric and statistical tools - a forensic problem, Euroanalysis, Warszawa (29.08.2013 r.)

## **6.5 Lista plakatów na konferencjach krajowych i zagranicznych**

1. I. Stanimirova, Probabilistic SIMCA – a classification tool that incorporates measurement uncertainty information, The 16th Chemometrics in Analytical Chemistry, Barcelona (6-10.06.2016 r.), Hiszpania

2. M. Horochowska, I. Stanimirova B. Czarnik-Matuszewicz, Studying the effect of enflurane, isoflurane and sevoflurane on DPPC lipid bilayers using ANOVA and PARAFAC (PARAFASCA), CAC 2015 – The 15th Chemometrics in Analytical Chemistry, Changsha (22-26.06.2015 r.), Chiny
3. M. Horochowska, I. Stanimirova, B. Czarnik-Matuszewicz, Badanie działania enfluranu, izofluranu i sewofluranu na strukturę błony lipidowej DPPC z wykorzystaniem metod ANOVA i PARAFAC, 9 Seminarium „Aktualne problemy chemii analitycznej”, Instytut Chemii, Uniwersytet Śląski w Katowicach, Katowice (15.05.2015 r.)
4. J. Orzel, I. Stanimirova, M. Daszykowski, A novel method for rapid evaluation of total antioxidant capacity of food commodities based on nondestructive fluorescence measurements supported by chemometric modeling, Food Analysis Congress, Barcelona (29-30.10.2014 r.), Hiszpania
5. J. Orzel, I. Stanimirova, M. Daszykowski, Modelowanie sygnałów spektroskopowych w kontekście oceny zdolności antyoksydacyjnej past pomidorowych, The 57th PTChem and SITPChem Meeting, Częstochowa (14-18.09.2014 r.)
6. J. Orzel, I. Stanimirova, M. Daszykowski, Ocena zdolności antyoksydacyjnej przecierów pomidorowych, 8 Seminarium „Aktualne problemy chemii analitycznej”, Instytut Chemii, Uniwersytet Śląski w Katowicach, Katowice (16.05.2014 r.)
7. M. Orłowska, I. Stanimirova, D. Staszek, M. Sajewicz, M. Waksmundzka-Hajnos, T. Kowalska, Ekstrakcja kwasów fenolowych dla gatunku tymianek pospolity za pomocą aparatu Soxhletha i przyspieszonej ekstrakcji rozpuszczalnikiem, V Podlaskie Spotkanie Chromatograficzne, Chlewiska (15-18.09.2013 r.)
8. M. Orłowska, I. Stanimirova, D. Staszek, M. Sajewicz, M. Waksmundzka-Hajnos, T. Kowalska, Extraction of Phenolic Acids from Common Thyme by Means of Soxhlet Apparatus and Accelerated Solvent Extraction (ASE), The 9th Balaton Symposium on High-Performance Separation Methods, Siófok (4-6.09.2013 r.), Węgry
9. M. Orłowska, I. Stanimirova, D. Staszek, K. Rykulska, S. Słomczyńska, M. Sajewicz, M. Waksmundzka-Hajnos, T. Kowalska, A comparison of the plant extraction methods upon an example of common thyme (*Thymus vulgaris*), The 36th conference on “Chromatographic methods of investigating the organic compounds”, Szczyrk (5-7.06.2013 r.)
10. L. Adam, K. Drab, I. Stanimirova, A. Smolinska, L. Blanchet, L. Buydens, B. Walczak, Zastosowanie fuzji danych do analizy widm  $^1\text{H-NMR}$ , XV Polish Symposium on applications of spectroscopic methods for investigation of materials and chemical compounds, the Poznan University of Economics, Poznań (25.05.2011 r.)
11. I. Stanimirova, K. Michalik, Z. Drzazga, H. Trzeciak, B. Walczak, Ocena wpływu terapii nowotworowej na proces mineralizacji kości u szczurów z wykorzystaniem modeli ANOVA I analizy czynników głównych, The 15th Polish Symposium on applications of spectroscopic methods for investigation of materials and chemical compounds, the Poznan University of Economics, Poznań (25.05.2011 r.)
12. K. Drab, L. Adam, I. Stanimirova, A. Smolinska, L. Blanchet, L. Buydens, B. Walczak, Analiza widm  $^1\text{H-NMR}$  przy użyciu metody ANOVA-PCA, 5 Seminarium “Aktualne problem chemii analitycznej”, Instytut Chemii, Uniwersytet Śląski w Katowicach, Katowice (13.05.2011 r.)

13. I. Stanimirova, R. Tauler, P.D. Wentzell, Beata Walczak, A comparison of positive matrix factorisation and the weighted multivariate curve resolution method, The 12th International Conference on Chemometrics in Analytical Chemistry, Antwerpia (21.10.2010 r.), Belgia
14. I. Stanimirova, B. Walczak, Fusion of multisource chemical data, The 8th Polish Conference on Analytical Chemistry, Kraków (6.06.2010 r.)
15. I. Stanimirova, B. Walczak, Combining data from different sources, RENACQ-4, Beni Mellal (12-13.03.2010 r.), Maroko
16. I. Stanimirova, K. Michalik, Z. Drzazga, M. Kaszuba, H. Trzeciak, B. Walczak, Application of ASCA to assess the effect of a maternal anticancer drug treatment on the mineralization of bones in rats, The 6th Symposium on Medical Physics, The 4th International Symposium on Medical Physics, Szczyrk (15-18.06.2009 r.)
17. M. Lelatko, I. Stanimirova, B. Walczak, Dealing with censored data in the chemometric analysis of chromatographic data, Conference on "Chromatographic methods of investigating the organic compounds", Szczyrk (3.06.2009 r.)
18. I. Stanimirova, B. Walczak, Dealing with missing elements and outliers in PCA, SIMCA and PLS, The 11th Conference on Chemometrics in Analytical Chemistry, Montpellier (31.06.2008 r.), Francja
19. I. Stanimirova, B. Walczak, Chemometric analysis of environmental data - a case study, 2 Seminarium „Aktualne problem chemii analitycznej”, Instytut Chemii, Uniwersytet Śląski w Katowicach, Katowice (30.05.2008 r.)
20. B. Vandeginste, N.F. Pérez, R. Boqué, J. Ferré, Y. Vander Heyden, S. Caetano, A. Durand, M. Novic, N. Grosej, B. Walczak, I. Stanimirova, L. Buydens, B. Üstün, ChemTRACE Chemometrics Toolbox for Food Traceability, TRACE 4th Annual Meeting, Torremolinos (23-25.04.2008 r.), Hiszpania
21. A. Smolinska, I. Stanimirova, M. Daszykowski, B. Walczak, How to select representative samples in tracing origin of food products, The 31st Conference on "Chromatographic methods of investigating the organic compounds", Szczyrk (5.06.2007 r.)
22. M. Michalska, I. Stanimirova, M. Daszykowski, B. Walczak, Application of chemometric methods to exploration of water quality data, Spring Conversatorium 2007, Instytut Chemii, Uniwersytet Śląski w Katowicach, Katowice (24.05.2007 r.)
23. I. Stanimirova, B. Walczak, Multiple factor analysis as a tool for exploring multiblock data, The 10th International Conference on Chemometrics in Analytical Chemistry CAC-2006, Aguas de Lindoia (14.09.2006 r.), San Paulo, Brazylia
24. I. Stanimirova, M. Daszykowski, K. Kaczmarek, B. Walczak, Robust exploration of data with missing elements and outliers, TRACE, The 2nd annual TRACE meeting, Praga (24-26.04.2006 r.), Republika Czeska
25. I. Stanimirova, B. Walczak, D.L. Massart, Y. Vander Heyden, A strategy for exploratory analysis of environmental data, ChemoAC meeting, Wolny Uniwersytet Brukselski, Bruksela (16-17.06.2005 r.), Belgia

26. K. Zehl, I. Stanimirova, D.L. Massart, J. Einax, Projection Pursuit und PARAFAC zur Auswertung eines Mehrwege Datensatzes, ANAKON International Analytical Conference, Regensburg, (15-18.03.2005 r.), Niemcy
27. I. Stanimirova, D.L. Massart, Variable selection in Multiple Factor Analysis (MFA), ChemoAC meeting, Wolny Uniwersytet Brukselski, Bruksela (9-10.10.2004 r.), Belgia
28. I. Stanimirova, M. Daszykowski, B. Walczak, F. Daeyaert, P.J. Lewi, D. L. Massart, Studying differences between crystal structures of HIV Reverse Transcriptase complexed with HIV inhibitors with STATIS, 9-th International Conference Chemometrics in Analytical Chemistry (New Challenges for a Maturing Science), Lizbona (20-23.09.2004 r.), Portugalia
29. M. Daszykowski, I. Stanimirova, B. Walczak, D.L. Massart, Projection Pursuit combined with Multivariate Regression Trees for the exploratory analysis of analytical data, The 9th International Conference Chemometrics in Analytical Chemistry (New Challenges for a Maturing Science), Lizbona (20-23.09.2004 r.), Portugalia
30. I. Stanimirova, M. Daszykowski, B. Walczak, F. Dayaert, P. Lewi, D.L. Massart, STATIS for analyzing differences between crystal structures of HIV Reverse Transcriptase, ChemoAC meeting, Wolny Uniwersytet Brukselski, Bruksela (10-11.04.2004 r.), Belgia
31. M. Daszykowski, I. Stanimirova, B. Walczak, F. Dayaert, P. Lewi, D.L. Massart, Interpretation of the data structure in terms of exploratory variables by Multivariate Regression Trees and Projection Pursuit, ChemoAC meeting, Wolny Uniwersytet Brukselski, Bruksela (10-11.06.2004 r.), Belgia
32. I. Stanimirova, B. Walczak, V. Simeonov, D.L. Massart, Insight into the 3-way data structure with STATIS, ChemoAC meeting, Wolny Uniwersytet Brukselski, Bruksela (15-16.01.2004 r.), Belgia
33. I. Stanimirova, B. Walczak, D.L. Massart, Processing sets with different types of variables (quantitative, qualitative, ...) with Multiple Factor Analysis, ChemoAC meeting, Wolny Uniwersytet Brukselski, Bruksela (10-11.06.2004 r.), Belgia
34. I. Stanimirova, B. Walczak, D.L. Massart, Robust Principal Component Analysis, The 2nd Conference, Chemometrics - Methods and Applications, Zakopane (16-19.10.2003 r.)
35. I. Stanimirova, B. Walczak, D.L. Massart, The ACT (STATIS) Method, ChemoAC meeting, Wolny Uniwersytet Brukselski, Bruksela (26.03.2003 r.), Belgia
36. I. Stanimirova, V. Simeonov, D.L. Massart, Preferred methods for exploratory analysis of environmental data, ChemoAC meeting, Wolny Uniwersytet Brukselski, Bruksela (21-22.11.2002 r.), Belgia
37. I. Stanimirova, V. Simeonov, D.L. Massart, Exploratory data analysis of environmental data, ChemoAC meeting, Wolny Uniwersytet Brukselski, Bruksela (24-25.04.2002 r.), Belgia
38. I. Stanimirova, V. Simeonov, D.L. Massart, Time trend analysis of environmental data, ChemoAC meeting, Wolny Uniwersytet Brukselski, Bruksela (24-25.04.2002 r.), Belgia

## **6.6 Lista przeprowadzonych kursów z zakresu szeroko pojętej chemometrii**

1. *“Classification with the SIMCA and robust SIMCA methods, Chemometrics in food traceability studies”*; w ramach projektu europejskiego TRACE, Radboud University, Nijmegen (28.06.2007 r.), Holandia
2. *„Planowanie eksperymentu”* dla pracowników i doktorantów Konsorcjum Wrocławskiego Centrum Biotechnologii posiadającego status Krajowego Naukowego Ośrodka Wiodącego (KNOW) w ramach działań projakościowych, Wydział Chemii, Politechnika Wroclawska, Wrocław (21-22.11.2015 r.)
3. *„Metody klasyfikacyjne i dyskryminacyjne”* dla pracowników i doktorantów Konsorcjum Wrocławskiego Centrum Biotechnologii posiadającego status Krajowego Naukowego Ośrodka Wiodącego (KNOW) w ramach działań projakościowych, Wydział Chemii, Politechnika Wroclawska, Wrocław (12-13.12.2015 r.)

### **6.7 Udział w projektach badawczych o zasięgu krajowym i międzynarodowym**

1. Udział w projektach badawczych realizowanych dla Konsorcjum ChemoAC (*Chemometrics in Analytical Chemistry*), działającym na Wolnym Uniwersytecie Brukselskim, Bruksela, Belgia, 2001-2005 r.
2. *„TRacing food Commodities in Europe”*, FOOD-CT-2005-006942, Komisja Europejska; szósty program ramowy, Charakter udziału: wykonawca, 2005-2007 r.
3. *„Analytical strategy to detect infringement of patents or proprietary processes of pharmaceuticals and fine chemicals”*, G6RD-CT-2002-00849, Komisja Europejska; piąty program ramowy, Charakter udziału: wykonawca, 2004-2005 r.
4. Projekt pt. *„Zintegrowany system wspomagający zarządzaniem i ochroną zbiornika zaporowego (ZiZOzap)”* realizowany w ramach Programu Operacyjnego Innowacyjna Gospodarka 2007-2013, Priorytet 1. Badania i rozwój nowoczesnych technologii. Charakter udziału: wykonawca, 2012 r.
5. Projekt OPUS 7, finansowany przez Narodowe Centrum Nauki, pt. *„Badanie stabilności wybranych związków azowych i diazowych jako dodatków fiskalnych paliw pod wpływem kontrolowanych czynników”*, (2014/13/B/ST4/05007); Charakter udziału: główny wykonawca, 2015-2017 r.

### **6.8 Składane aplikacje grantowe**

Złożyłam w ramach konkursu Fundacji na Rzecz nauki Polskiej projekt w ramach programu „Pomost” pod tytułem *„An ecotoxic risk assessment of residue products from medical waste incineration”* (2012).

Byłam współtwórcą projektu pt. *„Inteligentny system zintegrowanego zarządzania stanem wód”* złożonego w kolejnych dwóch edycjach programu BIOSTRATEG w ramach środków Narodowego Centrum Badań i Rozwoju (2014 i 2015).

Ponadto, uczestniczyłam w pracach związanych z opracowaniem wniosków aplikacyjnych do dwóch edycji programu TEMPUS (2009 i 2011).

## 6.9 Odbyte staże w zagranicznych ośrodkach naukowych lub akademickich

1. *Environmental Assessment and Water Research* (IDAEA), Spanish Council of Scientific Research, Barcelona, Spain, wrzesień 2009 r., staż naukowy pod opieką prof. R. Taulera
2. *Laboratory for Chemometrics*, Department of Chemistry at Dalhousie University, Halifax, Nova Scotia, Kanada, maj – sierpień 2009 r., staż naukowy (postdoctoral fellow) pod opieką prof. P.D. Wentzella
3. *Department of Pharmaceutical and Biomedical Analysis*, Wydział Medycyny i Farmacji, Wolny Uniwersytet Brukselski, Bruksela, Belgia, 2001-2005 r., studia doktoranckie i staż naukowy realizowany w ramach działalności konsorcjum *ChemoAC* pod opieką prof. D.L. Massarta

## 6.10 Opieka i promotorstwo prac dyplomowych

1. Jagoda Kurowska – praca magisterska pt. *„Wykorzystanie liniowej analizy dyskryminacyjnej w klasyfikacji okruchów szklanych dla potrzeb wymiaru sprawiedliwości”*, Instytut Chemii, czerwiec 2016 r. – promotor pracy
2. Magdalena Pawlaczyk – praca inżynierska pt. *„Optymalizacja procesu pomiaru pienistości piwa”*, Instytut Chemii, czerwiec 2015 r. – promotor pracy (praca uzyskała wyróżnienie dla najlepszej pracy inżynierskiej nadane przez Polskie Towarzystwo Chemiczne Oddział w Katowicach, 2015 r.)
3. Klaudia Drab – praca magisterska pt. *„Analiza wariancji danych metabolicznych”*, Instytut Chemii, czerwiec 2011 r. – promotor pracy
4. Katarzyna Kozicka, praca magisterska pt. *„Chemometryczna analiza sygnałów instrumentalnych zarejestrowanych różnymi technikami analitycznymi dla próbek win o różnym pochodzeniu geograficznym”*, Instytut Chemii, czerwiec 2010 r. – opiekun pracy
5. Małgorzata Kóleczo, praca magisterska pt. *„Badanie zależności pomiędzy profilami chromatograficznymi próbek żywności, a ich parametrami sensorycznymi”*, Instytut Chemii, czerwiec 2010 r. – opiekun pracy
6. Marcin Lełątko, praca magisterska pt. *„Chemometryczne strategie analizy danych z brakującymi i/lub cenzorowanymi elementami”*, Instytut Chemii, czerwiec 2009 r. – opiekun pracy
7. Agnieszka Hebda, praca magisterska pt. *„Zastosowanie modeli kalibracyjnych do kontroli jakości tabletek escitalopramu w oparciu o dane z bliskiej podczerwieni i Ramana”*, Instytut Chemii, czerwiec 2008 r. – opiekun pracy
8. Marta Michalska, praca magisterska pt. *„Zastosowanie metod chemometrycznych do eksploracji danych z kontroli jakości wody pitnej”*, Instytut Chemii, czerwiec 2007 r. – opiekun pracy

## 6.11 Uzyskane nagrody i wyróżnienia za działalność naukową

1. Nagroda za najlepszy poster przyznana na międzynarodowej konferencji: M. Horochowska, I. Stanimirova, B. Czarnik-Matusiewicz, Studying the effect of enflurane, isoflurane and

sevoflurane on DPPC lipid bilayers using ANOVA and PARAFAC (PARAFASCA), The 15th Chemometrics in Analytical Chemistry, Changsha (22-26.06.2015), Chiny

2. Dyplom za promotorstwo najlepszej pracy inżynierskiej pt. „Optymalizacja procesu pomiaru pienistości piwa” wykonanej przez inż. Magdalenę Pawlaczyk, obronionej w Instytucie Chemii na Wydziale Matematyki, Fizyki i Chemii w roku akademickim 2014/2015 przyznana przez Polskie Towarzystwo Chemiczne Oddział w Katowicach (14.04.2015 r.)
3. Nagroda Rektora Uniwersytetu Śląskiego III stopnia za działalność naukowo-badawczą, 2014 r.
4. Dodatek specjalny Rektora Uniwersytetu Śląskiego przyznany w drodze konkursowej za ponadprzeciętne osiągnięcia naukowe, 2012 r.
5. Nagroda dla najbardziej aktywnych recenzentów za działalność recenzencką dla czasopisma *Analytica Chimica Acta*, 2010 r.

#### **6.12 Pełnienie roli edytora w czasopismach z listy filadelfijskiej**

1. Redaktor z uprawnieniami do decydowania o przyjmowaniu prac do druku: *Acta Chromatographica* (ISSN: 1233-2356, IF = 0,523), od marca 2011 r. do nadal (wydawca czasopisma: Akademia Kiado i Uniwersytet Śląski w Katowicach)
2. Ko-edytor specjalnego wydania *Journal AOAC International* (IF = 1,385), 97 (2014), pt. “*Chemometrics in Analytical Chemistry Workflow*”

## 7. Syntetyczne zestawienie dorobku naukowego habilitanta<sup>4</sup>

### Do doktoratu:

Publikacje:	13 artykułów ogłoszonych drukiem w czasopismach z listy filadelfijskiej (sumaryczny współczynnik wpływu <sup>5</sup> = 23,302)
Rozdziały książkowe:	-
Wykłady:	10 wykładów na konferencjach zagranicznych w tym 3 na zaproszenie imienne
Postery:	14 posterów na konferencjach zagranicznych

### Po doktoracie:

Publikacje:	29 artykułów ogłoszonych drukiem w czasopismach z listy filadelfijskiej (sumaryczny współczynnik wpływu = 90,942)
Rozdziały książkowe:	3
Wykłady:	19 wykładów na konferencjach krajowych i zagranicznych w tym 5 na zaproszenie imienne
Postery:	24 posterów na konferencjach krajowych i zagranicznych

### Sumaryczny dorobek naukowy do habilitacji:

Cykl habilitacyjny:	8 publikacji (sumaryczny współczynnik wpływu = 31,787, średnia na publikację = 3,973)
Łączna liczba publikacji:	42 (sumaryczny współczynnik wpływu = 114,244, średnia na publikację = 2,720)
Rozdziały w książkach:	3
Wykłady:	29 wykładów na konferencjach krajowych i zagranicznych w tym 8 na zaproszenie imienne
Postery:	38 posterów na konferencjach krajowych i zagranicznych
Liczba cytowań artykułów:	655 (611 bez autocytowań)
Indeks Hirscha:	14

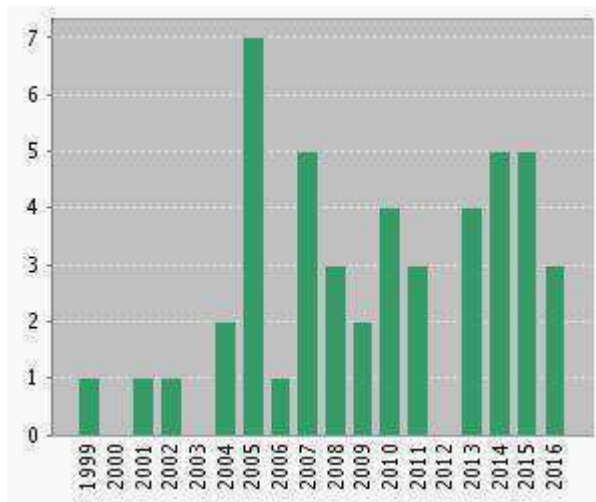
*Anna Stanisława - Paszykowska*

<sup>4</sup> dane wedle bazy Web of Science na dzień 19.12.2016 r.

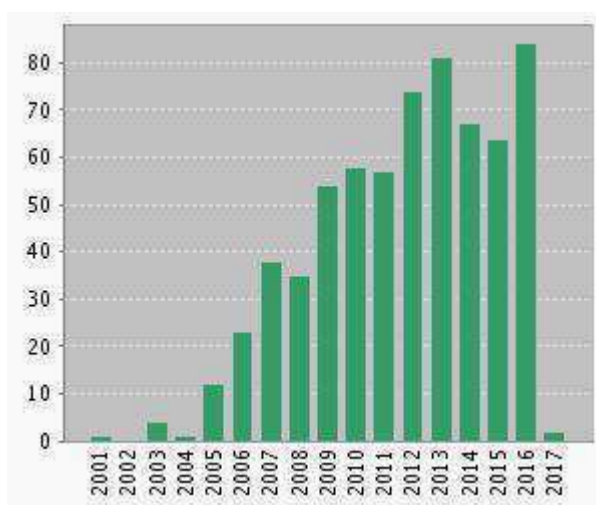
<sup>5</sup> Współczynnik wpływu (IF) tzw. 'impact factor' jest zgodny obowiązującym w roku wydania danej publikacji wedle *Journal Citation Reports* (JCR)



## 8. Aktywność publikacyjna w poszczególnych latach i liczba cytowań



Rys. 1 Liczba opublikowanych artykułów w poszczególnych latach (wedle bazy Web of Science na dzień 19.12.2016 r.)



Rys. 2 Liczba cytowań artykułów w poszczególnych latach (wedle bazy Web of Science na dzień 19.12.2016 r.)