

Wszczęcie przewodu doktorskiego

Proponowany tytuł pracy doktorskiej:

***Wykrywanie zafalszowań, potwierdzanie autentyczności oraz identyfikacja zagrożeń biologicznych z wykorzystaniem chromatografii i modelowania chemometrycznego***

mgr Barbara Krakowska

Opiekun pracy:  
dr hab. prof. UŚ Michał Daszykowski

Instytut Chemii  
Wydział Matematyki, Fizyki i Chemii  
Uniwersytet Śląski  
Katowice, 2015



## Spis treści

Wykaz skrótów stosowanych w pracy .....	4
1. Wstęp .....	6
2. Analiza chromatograficzna .....	8
3. Narzędzia chemometryczne do analizy chromatograficznych odcisków palca.....	9
3.1 Wstępne przygotowanie chromatograficznych odcisków palca .....	9
3.2 Metody modelowania stosowane do weryfikacji autentyczności próbek.....	12
3.3 Walidacja modelu dyskryminacyjnego.....	13
3.4 Metody wyboru zmiennych .....	14
3.5 Weryfikacja autentyczności wybranych produktów w oparciu o chromatograficzne odciski palca.....	15
4. Cele pracy .....	17
5. Przeprowadzone badania własne .....	18
5.1 Identyfikacja procederu fałszowania oleju napędowego .....	18
5.2 Identyfikacja zafalszowań leku Viagra® .....	25
6. Podsumowanie .....	31
7. Dalsze plany naukowe .....	32
8. Dorobek naukowy.....	33
9. Literatura.....	36

## Wykaz skrótów stosowanych w pracy

Skrót	Nazwa polska	Nazwa angielska
AUC	pole powierzchni pod krzywą	area under curve
CART	drzewa klasyfikacji i regresji	classification and regression trees
CCR	procent poprawnej klasyfikacji	correct classification rate
COW	metoda zoptymalizowanego nakładania sygnałów maksymalizująca ich wzajemną korelację	correlation optimized warping
DAD	detektor z matrycą diodową	diode array detector
ED-XRF	spektroskopia rentgenowska z dyspersją energii	energy-dispersive X-ray spectroscopy
GC-FID	chromatografia gazowa z detekcją płomieniowo-jonizacyjną	gas chromatography with flame ionization detector
HPLC-DAD	wysokosprawna chromatografia cieczowa z matrycą diodową	high-performance liquid chromatography with diode-array detector
LDA	liniowa analiza dyskryminacyjna	linear discriminant analysis
MS	spektrometria mas	mass spectrometry
NIR	bliska podczerwień	near infra red
NMR	magnetyczny rezonans jądrowy	nuclear magnetic resonance
PAsLS	metoda asymetrycznych najmniejszych kwadratów z funkcją kary	penalized asymmetric least squares
PCA	analiza czynników głównych	principal component analysis
PLS	metoda częściowych najmniejszych kwadratów	partial least squares
PLS-DA	dyskryminacyjny wariant metody częściowych najmniejszych kwadratów	partial least squares discriminant analysis
ROC	charakterystyka operacyjna odbiornika	receiver operating characteristic
SE	selektywność	selectivity
SIMCA	metoda modelowania indywidualnych grup	soft independent modelling of class analogies
SMC	metodę korelacji wieloczynnikowej	significance multivariate correlation
SP	specyficzność	specificity

SR	współczynnik selektywności	selectivity ratio
UVE	metoda eliminacji zmiennych nieistotnych	uninformative variable elimination
VIP	metoda zmiennych znaczących dla projekcji	variable importance in projection

## 1. Wstęp

Obecnie rynek produktów podrabianych i zafałszowanych rozwija się na szeroką skalę. Jest to spowodowane niższą ceną tego typu produktów. Falszowanie produktów jest celową ingerencją człowieka w ich skład, wygląd lub procedurę ich wytwarzania. W zależności czego dotyczy proceder fałszowania obserwuje się różne jego skutki. Gdy mamy do czynienia z fałszowaniem paliwa polegającym na usunięciu z niego dodatków akcyzowych, na szkodę narażony jest przede wszystkim budżet Państwa poprzez zaniżenie otrzymywanej kwoty podatku [1]. Natomiast, gdy problem dotyczy fałszowania leków, stawka jest dużo wyższa, gdyż zagrożone jest życie i zdrowie ludzkie. Wynika to z faktu, że nielegalne wytwarzanie leków najczęściej odbywa się w prymitywnych warunkach, w których nie są spełniane podstawowe normy czystości, a wytwarzane produkty nie podlegają żadnej kontroli. Zagrożenie w takim przypadku nie stanowi zaniżona zawartość substancji czynnej leku, a zanieczyszczenia pochodzące z produkcji [2]. Pod pojęciem zafałszowania kryje się też rozcieńczanie produktów tańszymi substancjami o podobnych właściwościach. Przykład stanowi dodawanie do miodu syropu kukurydzianego w celu zwiększenia jego objętości [3]. Tego typu działanie jest nielegalne i szkodzi bezpośrednio konsumentowi. Wszystkie te przykłady świadczą o dużej potrzebie kontrolowania parametrów produktów, których autentyczność może mieć wymiar finansowy, ale również może zagrażać zdrowiu i życiu ludzi. Pojęcie autentyczności produktu najczęściej związane jest z jego składem jakościowym bądź ilościowym. Może także być utożsamiane z pochodzeniem geograficznym danego produktu [4]. Ocena autentyczności różnego rodzaju produktów stanowi istotny problem badawczy, którego skuteczne rozwiązanie wymaga zaawansowanych podejść analitycznych i odpowiedniego modelowania danych wykorzystując narzędzia chemometryczne. Zazwyczaj analiza próbek obejmuje czynności mające na celu zbadanie wybranych parametrów produktu i porównanie uzyskanych wyników ze ściśle określonymi wymaganiami, w celu potwierdzenia ich zgodności. Analiza próbek o złożonym składzie jakimi są m.in. próbki żywności czy próbki środowiskowe to tylko jedno z wyzwań dotyczących zarówno analizy jakościowej jak i ilościowej.

Sygnaly instrumentalne uzyskane dla złożonych próbek i opisujące składniki chemiczne w nich zawarte posiadają duży zasób informacji na temat badanej próbki co pozwala rozpatrywać je jako tzw. chemiczne odciski palca charakteryzujące się unikalnością. Analiza tego typu danych opiera się na porównaniu uzyskanych sygnałów instrumentalnych pomiędzy sobą lub względem sygnałów uzyskanych dla próbek referencyjnych. Takie podejście jest stosowane do oceny autentyczności, ponieważ bardzo często zależy ona od całościowego składu chemicznego analizowanego produktu. Każda próbka opisywana jest przez sygnał analityczny odzwierciedlający informację chemiczną zawartą w próbce. Złożoność sygnałów analitycznych wynika z sumowania się informacji pochodzących od poszczególnych komponentów próbki. W przypadku sygnałów chromatograficznych, optymalny rozdział chromatograficzny stanowi potencjalną gwarancję możliwości wydobywania z niego maksimum informacji dotyczącej składu analizowanej próbki, jednak wymaga on m.in. odpowiedniego doboru kolumny chromatograficznej, składu fazy ruchomej oraz warunków rozdziału. Dla próbek naturalnych, ze względu na ich złożony skład, uzyskanie optymalnego rozdziału chromatograficznego jest trudne. Jedną z możliwości

poprawy jakości sygnału analitycznego wymaga zastosowania odpowiedniej procedury laboratoryjnej poprzedzającej analizę instrumentalną badanej próbki, takiej jak np. wstępne oczyszczanie próbki, jej zatężenie czy ekstrakcja. Poprawa jakości sygnału analitycznego wynikająca z zastosowanej techniki chromatograficznej może następować na skutek zwiększenia rozdzielczości (zastosowanie odpowiednich kolumn i rozpuszczalników) oraz poprzez wykorzystanie zaawansowanych detektorów takich jak np. spektrometr mas. Wykorzystanie chemometrycznych narzędzi do analizy złożonych danych eksperymentalnych ułatwia m.in. ich interpretację i eksplorację skutkując tym samym coraz szerszym zastosowaniem różnego rodzaju podejść chemometrycznych do analizy chromatograficznych odcisków palca w kontekście kontroli autentyczności wybranych produktów [5].

## 2. Analiza chromatograficzna

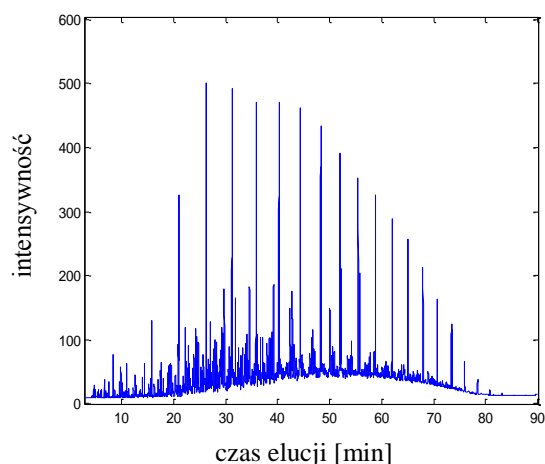
Analiza chromatograficzna jest techniką polegającą na rozdzieleniu składników próbki ze względu na ich powinowactwo do fazy stacjonarnej. Stanowi ona jedną z najczęściej stosowanych technik analitycznych wykorzystywanych do identyfikacji i analizy ilościowej składu chemicznego złożonych próbek. Wynika to z możliwości jednoczesnego oznaczania wielu składników próbki w trakcie jednego rozdzielania chromatograficznego.

Dzięki połączeniu standardowych technik chromatograficznych z różnego rodzaju zaawansowanymi detektorami wielokanałowymi (np. detektorem z matrycą diodową, z ang. *diode array detector* – DAD, lub spektrometrią mas, z ang. *mass spectrometry detector* – MS) uzyskano sprzężone techniki chromatograficzne, dzięki którym możliwe jest uzyskanie informacji o czystości pików chromatograficznych i polepszenie identyfikacji związków zawartych w badanej próbce. Z kolei dobry rozdział chromatograficzny komponentów próbki pozwala uzyskać pełniejszą informację o jej składzie chemicznym. Jednakże, analiza jakościowa poszczególnych substancji próbki oznaczonych podczas analizy chromatograficznej jest zazwyczaj skomplikowana, czasochłonna i wymaga zakupu kosztownych wzorców. W przypadku próbek o złożonym składzie chemicznym, całkowity rozdział komponentów oraz identyfikacja poszczególnych substancji chemicznych bywają wysoce skomplikowane, a nawet czasami nieosiągalne. Z tego powodu, często stosuje się sygnały instrumentalne stanowiące chemiczne odciski palca badanych próbek (z ang. *fingerprints*).

Chromatograficzny odcisk palca to charakterystyczny profil reprezentujący skład chemiczny próbki najlepiej jak to możliwe. Optymalny chemiczny odcisk palca to chromatogram, w którym obserwowana jest relatywnie duża rozdzielczość pików. Taka definicja implikuje konieczność odpowiedniego doboru warunków rozdzielania, które konsekwentnie są stosowane w analizie całego zbioru próbek [6].

Techniki chromatograficzne to doskonałe narzędzia do rejestracji chemicznych odcisków palca ze względu na stosunkowo nieskomplikowaną interpretację sygnałów. Jeden pik odpowiada jednej substancji, którą można scharakteryzować m.in. poprzez jej czas elucji, czy widmo MS porcji eluentu w danym czasie elucji. Cecha ta pozwala również na określenie sygnałów chromatograficznych jako unikalne źródło informacji o składzie próbki. Chromatograficzne odciski palca modelowane narzędziami chemometrycznymi jakimi są np. techniki dyskryminacyjne pozwalają konstruować modele diagnostyczne znajdujące swoje zastosowanie m.in. do badania autentyczności wybranych produktów. Dodatkowo, metody chemometryczne wspomagają analizę sygnałów, dla których nie został uzyskany optymalny rozdział chromatograficzny tzn. gdy niektóre piki nie są całkowicie od siebie oddzielone.





Rys. 1 Przykładowy chromatogram stanowiący chromatograficzny odcisk palca próbki oleju napędowego.

### 3. Narzędzia chemometryczne do analizy chromatograficznych odcisków palca

Analiza chemometryczna chromatograficznych odcisków palca może być trudna ze względu na obecność dodatkowych komponentów takich jak szum czy linia podstawowa. W zależności od wpływu tych komponentów, wyniki analizy surowych sygnałów (chromatograficznych odcisków palca) mogą być niewiarygodne. Przed przystąpieniem do modelowania danych chromatograficznych z użyciem różnego rodzaju metod chemometrycznych bardzo ważnym etapem jest wstępne przygotowanie danych polegające m.in. na eliminacji czynników wpływających na wartość informacyjną (szum, linia bazowa, przesunięcia pików).

#### 3.1 Wstępne przygotowanie chromatograficznych odcisków palca

Każdy sygnał instrumentalny w tym chromatograficzne odciski palca składają się z trzech komponentów: szumu, linii podstawowej oraz pików pochodzących od komponentów próbki [7]. Piki chromatograficzne stanowią główne źródło informacji mówiące o składzie próbki ilościowym i jakościowym. Natomiast linia podstawowa i szum są komponentami zakłócającymi sygnał. Zniekształcają piki pochodzące od poszczególnych składników i tym samym fałszują realny obraz składu próbki. Dzieje się tak, gdy linia podstawowa lub szum mają relatywnie dużą intensywność przez co analiza jakościowa i ilościowa jest trudna, a czasem nawet niemożliwa.

Szum definiuje się jako odchylenie standardowe od wartości średniej sygnału rejestrowanego przez dany przyrząd pomiarowy. Charakteryzuje go wielkość określająca stosunek sygnału do szumu (stosunek średniej z sygnału do jego odchylenia standardowego) [8]. Szum ma największą częstotliwość spośród wszystkich komponentów sygnału. Jego obecność jest warunkowana ograniczoną czułością stosowanego detektora powodującą fluktuacje sygnału

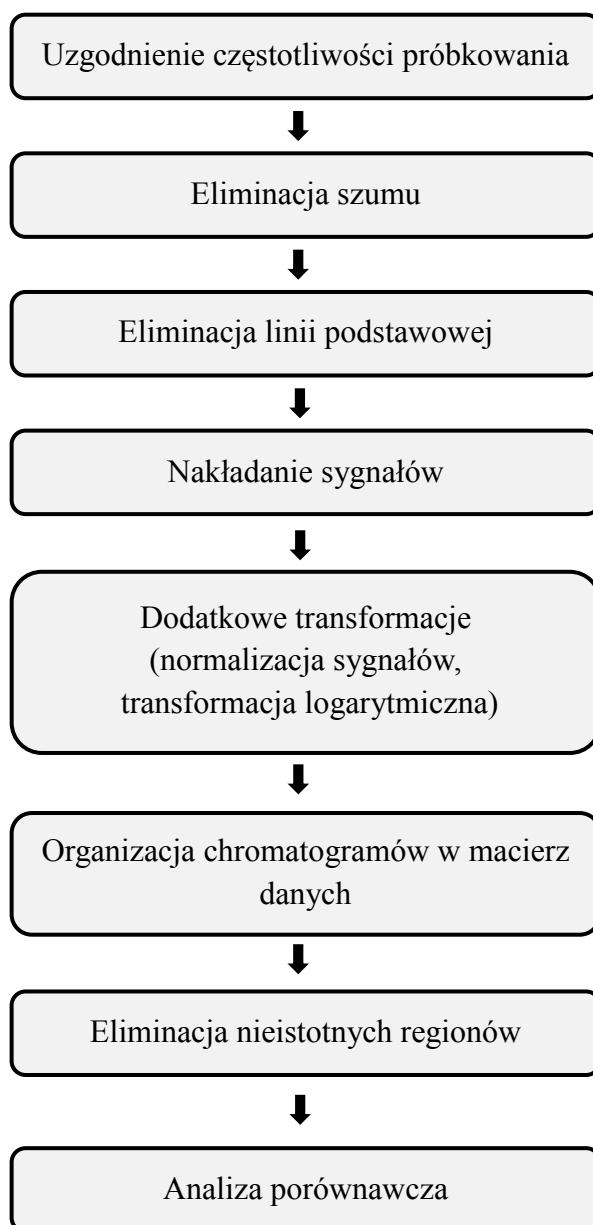
instrumentalnego oraz możliwością zachodzenia w trakcie analizy reakcji pomiędzy składnikami zawartymi w próbce. Wyróżnia się kilka rodzajów szumu, między innymi tzw. szum biały o rozkładzie gaussowskim, szum skorelowany oraz szum proporcjonalny do sygnału. Szum ze względu na to, że jest komponentem sygnału, nie wnoszącym istotnej analitycznej informacji, może negatywnie wpływać na dalszą analizę danych instrumentalnych. Można go zniwelować poprzez zastosowanie różnego rodzaju narzędzi chemometrycznych, np. używając filtra bazującego na medianie lub wartości średniej [9].

Linia podstawowa jest kolejnym składnikiem chromatograficznych odcisków palca. Charakteryzuje się ona najmniejszą częstotliwością i nie zawiera istotnych informacji o składzie chemicznym próbki. Z analitycznego punktu widzenia jest to sygnał instrumentalny rejestrowany dla próbki bez badanych analitów. Kształt linii podstawowej jest zmienny. Nawet dla zestawu próbek o tym samym pochodzeniu może się on znacznie różnić. Dlatego usunięcie linii podstawowej jest ważne przed dalszą analizą chemometryczną danych ze względu na możliwość zafałszowań uzyskanych wyników. W tym celu opracowano wiele metod umożliwiających eliminację linii podstawowej, m.in. metodę asymetrycznych najmniejszych kwadratów z funkcją kary (z ang. *penalized asymmetric least squares*, PAsLS) [10].

Kolejnym zjawiskiem utrudniającym analizę zbioru sygnałów instrumentalnych są przesunięcia pomiędzy pikami pochodzącymi od tych samych substancji. Ważnym jest, aby przed korekcją przesunąć usunąć linię podstawową, gdyż może ona mieć wpływ na nachylenie sygnału. Przesunięcia pomiędzy pikami są zazwyczaj spowodowane niestabilnością warunków rozdziału chromatograficznego, m.in. starzeniem się złoża kolumny chromatograficznej oraz wahaniami składu fazy ruchomej. Jest to szczególnie niekorzystne w przypadku wykorzystania metod bazujących na chemicznych odciskach palca, gdzie analiza próbek opiera się na porównywaniu ze sobą poszczególnych sygnałów. Wówczas, pik pochodzący od tej samej substancji różni się położeniem na poszczególnych chromatogramach i wyniki przeprowadzonej analizy będą błędne, gdyż pomimo podobieństwa składu próbek mogą one zostać zidentyfikowane jako różniące się ze względu na zawartość określonych komponentów. W celu usunięcia przesunąć pomiędzy pikami stosuje się techniki znane jako metody nakładania pików, m.in. metodę zoptymalizowanego nakładania widm, która maksymalizuje wzajemną korelację sygnałów (z ang. *correlation optimized warping*, COW) [11,12]. Przed przystąpieniem do analizy chemometrycznej należy jeśli jest to konieczne ujednolicić sygnały w taki sposób aby zarówno liczba punktów pomiarowych jak i gęstość ich próbkowania były takie same. Jest to konieczne w przypadku, gdy chromatogramy analizowanych próbek mają różną długość (różną liczbę punktów pomiarowych) co uniemożliwia zestawienie sygnałów w macierz danych.

Omówione zagadnienia dotyczące przygotowania sygnałów instrumentalnych w tym chromatograficznych odcisków palca są typowe dla wszystkich ich rodzajów i ich zastosowanie pozwala uzyskać lepsze rezultaty analizy porównawczej i modelowania. Dlatego też, przygotowanie danych do analizy obejmuje ocenę wpływu parametrów sygnału takich jak szum, czy występowanie przesunąć pików oraz obecność linii podstawowej oraz ich ewentualne usunięcie. Wstępnie przygotowane sygnały instrumentalne są następnie zestawiane w macierz danych w przypadku sygnałów mających postać wektora lub w tensor,

gdy pojedynczą próbkę opisują zbiór sygnałów – macierz danych. W przypadku sygnałów instrumentalnych, które dla każdej próbki mają postać wektora tak jak np. chromatogramy czy widma NIR, organizuje się je w macierz danych w taki sposób, że każdy z sygnałów reprezentujący badaną próbkę stanowi kolejny jej wiersz. Poszczególne etapy wstępnego przygotowania danych i zestawienia sygnałów w macierz przedstawia Rys. 2.



Rys. 2 Główne kroki postępowania z chromatograficznymi odciskami palca.

### 3.2 Metody modelowania stosowane do weryfikacji autentyczności próbek

Analizę chemometryczną chemicznych odcisków palca w kontekście badania autentyczności można prowadzić stosując różne metody. Jedną z nich są tzw. metody uczenia bez nadzoru. Mają one na celu ułatwienie identyfikacji grupowania się obiektów o podobnych właściwościach i/lub ujawnienie próbek znacznie różniących się od pozostałych w oparciu o zbiór danych. Do typowych technik uczenia bez nadzoru należy m.in. analiza czynników głównych (z ang. *principal component analysis*, PCA) [6,13]. Jej celem jest modelowanie, kompresja i wizualizacja wielowymiarowych danych. W analizie eksploracyjnej z wykorzystaniem metody PCA macierz danych  $\mathbf{X}$  o  $m$  obiektach i  $n$  parametrach jest przedstawiona jako iloczyn dwóch macierzy  $\mathbf{T}$  i  $\mathbf{P}$  o wymiarowości odpowiednio  $[m,f]$  i  $[f,n]$  (zob. Równanie 1). Macierz  $\mathbf{T}$  zawiera współrzędne obiektów, a macierz  $\mathbf{P}$  współrzędne parametrów na nowych ukrytych zmiennych tzw. czynnikach głównych.

$$\mathbf{X}_{[m,n]} = \mathbf{T}_{[m,f]} \mathbf{P}_{[f,n]}^T + \mathbf{E}_{[m,n]} \quad (1)$$

gdzie,  $\mathbf{X}$  to macierz wyjściowych danych,  $\mathbf{T}$  to macierz wyników,  $\mathbf{P}$  reprezentuje macierz wag, a  $\mathbf{E}$  macierz reszt od modelu,  $f$  to liczba czynników głównych, a  $m$  i  $n$  to odpowiednio macierz próbek i zmiennych.

Czynniki główne są konstruowane w sposób iteracyjny tak, aby maksymalizować opis wariancji danych. Każdy kolejny czynnik główny modeluje wariancję niewyjaśnioną przez poprzednie czynniki główne. Dlatego, wkład każdego kolejnego czynnika głównego w opisie całkowitej wariancji danych jest coraz mniejszy.

Wstępna ocena struktury danych pozwala określić zależności istniejące między próbkami i/lub parametrami, jak również ułatwia interpretacje wyników uzyskanych w kolejnych etapach analizy chemometrycznej. Główną zaletą analizy danych metodą PCA jest brak konieczności posiadania wiedzy na temat przynależności analizowanych danych do poszczególnych grup, co odróżnia ją od metod uczenia z nadzorem.

Metody uczenia z nadzorem w zależności od problemu badawczego, dzieli się na metody umożliwiające konstrukcję modeli kalibracyjnych, dyskryminacyjnych i klasyfikacyjnych. Do budowy tego typu modeli wykorzystuje się zbiór danych modelowych  $\mathbf{X}$  oraz zmienną zależną  $\mathbf{y}$ , która zawiera informację o przynależności próbki do danej grupy. Techniki dyskryminacyjne/klasyfikacyjne mają na celu przypisanie próbek do określonych grup na podstawie reguł logicznych opracowanych dla zbioru próbek uczących, należących do znanych grup. W metodach klasyfikacyjnych modele budowane są dla każdej grupy osobno. W konsekwencji, końcowy model uwzględnia fakt, że nowa próbka może należeć do jednej grupy, do kilku grup jednocześnie lub nie należała do żadnej z nich. Przykład mogą stanowić badania diagnostyczne klasyfikujące chorych ze względu na rodzaj nowotworu na jaki cierpią. Mogą oni należeć do dwóch grup jednocześnie, gdy zdiagnozowano u nich dwa rodzaje nowotworu, lub nie należeć do żadnej z grup jak jest w przypadku ludzi zdrowych. Do najczęściej stosowanych metod klasyfikacyjnych jest metoda modelowania indywidualnych grup, SIMCA (z ang. *soft independent modelling of class analogies*) [14].

W metodach dyskryminacyjnych, przestrzeń zmiennych eksperymentalnych zostaje podzielona na kilka wzajemnie wykluczających się podprzestrzeni, których liczba jest równa liczbie grup w analizowanym problemie dyskryminacyjnym. Ze względu na położenie próbek w przestrzeni zmiennych objaśniających jest ona zawsze przypisana tylko do jednej grupy. W przypadku problemu dwuklasowego przynależność próbek do określonych grup koduje zmienna zależna  $y$  mająca postać wektora. Poszczególne jego elementy definiują grupy za pomocą kodowania binarnego (0, 1) lub bipolarnego (-1, 1). Przyjęta etykieta dla danej grupy jest kwestią umowną. Ogólny model dyskryminacyjny, opisujący problem dwuklasowy, można przedstawić w postaci

$$\mathbf{y}_{[m,1]} = \mathbf{X}_{[m,n]} \mathbf{b}^T_{[n,1]} + \mathbf{e}_{[m,1]} \quad (2)$$

gdzie,  $y$  to zbiór zmiennych zależnych,  $\mathbf{X}$  to zbiór zmiennych objaśniających,  $\mathbf{b}$  to wektor współczynników korelacji,  $\mathbf{e}$  jest wektorem reszt od modelu, a  $m$  i  $n$  to odpowiednio macierz próbek i zmiennych.

Modele dyskryminacyjne konstruuje się wykorzystując takie techniki chemometryczne jak liniowa analiza dyskryminacyjna (z ang. *linear discriminant analysis*, LDA) [15], drzewa klasyfikacji i regresji (z ang. *classification and regression trees*, CART) [16] oraz dyskryminacyjna metoda częściowych najmniejszych kwadratów (z ang. *discriminant partial least squares*, PLS-DA) [17,18,19].

Do podziału zbioru danych na zbiór modelowy i testowy służą takie metody jak algorytm Duplex lub metoda Kennarda i Stonea [20]. Zasada ich działania skupia się na wyborze do zbioru modelowego próbek, które jak najlepiej opisują zmienność danych eksperymentalnych. Pozwala to zagwarantować, że zbudowany model dyskryminacyjny będzie optymalnie przewidywał nowe próbki.

Badanie produktów pod względem ich autentyczności to problem dwuklasowy - analizowana próbka może być autentyczna lub nie, co w pełni uzasadnia użycie metod dyskryminacyjnych. Gdy o autentyczności analizowanej próbki świadczy jej pochodzenie geograficzne, wówczas metody klasyfikacyjne są bardziej odpowiednie.

### 3.3 Walidacja modelu dyskryminacyjnego

Istnieje wiele parametrów walidacyjnych charakteryzujących efektywność działania modeli dyskryminacyjnych. Są one obliczane niezależnie dla zbioru modelowego i testowego. Najbardziej popularnym parametrem oceny modelu jest procent poprawnej klasyfikacji (z ang. *correct classification rate*, CCR), który mówi o liczbie próbek, które są właściwie rozpoznane przez model. Innymi parametrami oceny modelu są czułość i specyficzność. Do ich obliczenia wykorzystuje się informację o liczbie próbek poprawnie lub niepoprawnie rozpoznawanych dla każdej z analizowanych grup (oddzielnie lub w całym zbiorze próbek). Dla problemu dyskryminacyjnego, w którym dyskryminacji poddaje się tylko dwie grupy próbek, tak jak ma to miejsce w problemach identyfikacji autentyczności, zakłada się, że

próbki autentyczne stanowią grupę pozytywną, a zafalszowane grupę negatywną. Jako próbki prawdziwie pozytywne i prawdziwie negatywne uznaje się te, które są poprawnie przyporządkowane do danej grupy za pomocą danego modelu. Próbki fałszywie pozytywne są to próbki zafalszowane (negatywne) rozpoznawane przez model dyskryminacyjny jako autentyczne (pozytywne). Analogicznie, próbki fałszywie negatywne są to próbki autentyczne (pozytywne) błędnie przypisywane do grupy próbek zafalszowanych (negatywnych).

Czułość dla danej grupy próbek definiuje się jako iloraz liczby próbek prawdziwie pozytywnych i liczby wszystkich próbek pozytywnych. Specyficzność określa stosunek liczby próbek prawdziwie negatywnych do liczby wszystkich próbek negatywnych w analizowanym zbiorze danych.

Kolejnym sposobem oceny jakości modelu dyskryminacyjnego jest analiza krzywej ROC (z ang. *receiver operating characteristic*) [21]. Obrazuje ona zależność pomiędzy procentem próbek prawdziwie pozytywnych i fałszywie pozytywnych. Im bardziej krzywa ma wypukły kształt tym bardziej model dyskryminacyjny jest wiarygodny. Krzywa ROC może być również opisana za pomocą pola powierzchni pod jej wykresem, tzw. parametr AUC (z ang. *area under curve*). Najlepszą dyskryminację próbek zapewnia model, którego AUC wynosi 1.

W celu uzyskania bardziej realistycznych estymacji parametrów opisujących konstruowany model możliwe jest zastosowanie tzw. podejścia ‘bootstrap’ [22]. Polega ono na losowym wyborze próbek z każdej grupy do zbioru modelowego powtarzając dany wybór wielokrotnie. To pozwala symulować jak wpływa zmienność zbioru modelowego na konstrukcję modelu i jego działanie. Dla każdego wylosowanego zbioru modelowego konstruowane są modele, a zbiór testowy, który tworzą pozostałe próbki, służy do ich walidacji. Dzięki takiemu podejściu otrzymuje się rozkład określonych parametrów opisany przez wartość średnią i odchylenie standardowe w funkcji kompleksowości modelu, a tym samym zakresy niepewności dla uzyskanych wartości estymowanych parametrów walidacyjnych.

Alternatywnym podejściem jest metoda Monte Carlo. Polega ona na losowym podziale zbioru próbek na dwa podzbiory, który wykonuje się wielokrotnie. Za każdym razem do zbioru testowego włączana jest ta sama liczba próbek stanowiąca od 30% do 50% całkowitej liczby próbek [23]. Ponadto w celu określenia istotności konstruowanego modelu może być wykorzystywany test permutacyjny, który również polega na symulacji zmienności zbioru modelowego [24].

### **3.4 Metody wyboru zmiennych**

W przypadku, gdy konstruowany model ma bardzo dużo parametrów w stosunku do liczby próbek na podstawie której był konstruowany istnieje możliwość przeuczenia modelu. W takim przypadku model dopasowuje się do przypadkowych błędów obecnych w danych, a tym samym pogarszają się jego właściwości predykcyjne. Aby uniknąć przeuczenia modelu stosowane są metody wyboru zmiennych. Ich głównym celem jest identyfikacja zmiennych mających największy wkład do budowy modelu, a co za tym idzie rozróżnienia badanych

grup próbek. Model konstruowany dla wybranych zmiennych istotnych ma zbliżone lub lepsze parametry predykcyjne w stosunku do wyjściowego modelu. Wśród dużej liczby technik wyboru zmiennych wiele z nich jest dedykowane metodzie częściowych najmniejszych kwadratów, PLS. Są to m.in. metoda zmiennych znaczących dla projekcji (z ang. *variable importance in projection*, VIP) [25], współczynnik selektywności (z ang. *selectivity ratio*, SR) [26,27] oraz metoda eliminacji zmiennych nieistotnych (z ang. *uninformative variable elimination*, UVE) [28]. Idea wszystkich tych metod polega na obliczaniu różnego rodzaju parametrów charakteryzujących wkład danej zmiennej do budowy modelu. Następnie, wyznaczana jest wartość progowa parametru, a zmienne posiadające wartości większe niż wartość progowa są uznane za istotne i konieczne do konstrukcji modelu.

### **3.5 Weryfikacja autentyczności wybranych produktów w oparciu o chromatograficzne odciski palca**

Chromatograficzne odciski palca, zawierające duży zasób informacji na temat składu chemicznego badanych próbek, są często stosowane do badania autentyczności wybranych produktów determinowanej przez ich skład ilościowy/jakościowy lub ich pochodzenie geograficzne. Jednym z przykładów wykorzystania chromatograficznych odcisków palca jest monitorowanie jakości produktów ziołowych. Są one coraz bardziej popularne ze względu na ich naturalne pochodzenie, przez co uznaje się, że są z natury bezpieczne. Zioła i ich ekstrakty są bardzo złożone pod względem składu chemicznego, a dokładna identyfikacja ich komponentów jest niezbędna do oceny ryzyka związanego z ich zażywaniem. Wśród metod analitycznych, techniki bazujące na chromatograficznych odciskach palców są rekomendowane jako rzetelna metodologia do identyfikacji i kontroli jakości leków i preparatów ziołowych [29]. Badania tego typu produktów są prowadzone w celu uniknięcia zafałszowań produktów pod kątem ich obniżonej jakości. Obecnie analiza chromatograficzna produktów ziołowych bazująca na chromatograficznych odciskach palców w połączeniu z różnego rodzaju narzędziami chemometrycznymi stała się jedną z najbardziej popularnych metod w analizie materiałów roślinnych [30,31,32].

Podobnie jak produkty ziołowe mogą być także analizowane inne produkty farmaceutyczne, które są fałszowane stanowiąc tym samym realne niebezpieczeństwo dla zdrowia i życia ludzi. Fałszowane leki nie są poddawane kontroli jakości i tym samym ich bezpieczeństwo, jakość i skuteczność nie może być zagwarantowana. Obecnie, ocena bezpieczeństwa podrabianych leków opiera się głównie na identyfikacji i oznaczeniu zawartych w nich substancji czynnych. Jednak coraz częściej dokonuje się również analizy potencjalnych toksycznych składników pobocznych powstających w trakcie wytwarzania tego typu leków, takich jak np. pozostałości rozpuszczalników czy innych zanieczyszczeń [33,34].

Kolejnym obszarem wykorzystania chromatograficznych odcisków palca jest badanie produktów spożywczych. Kontrola autentyczności żywności w dużej mierze opiera się na identyfikacji pochodzenia żywności w związku z koniecznością weryfikacji znaków potwierdzających pochodzenie geograficzne. Pochodzenie geograficzne żywności, wraz z jej składem, powinno być znane konsumentowi i sprecyzowane na każdym produkcie

spożywczym. Działania te mają na celu zagwarantowanie bezpieczeństwa, autentyczności produktów oraz ochronę praw konsumentów. W związku szeroką gamą produktów wymagających kontroli oraz z ciągłą zmianą regulacji prawnych definiujących sposoby odnowy autentyczności produktów spożywczych istnieje duże zapotrzebowanie na opracowanie szybkich i tanich procedur mających na celu wykrywanie zafałszowań produktów i/lub sprawdzenie zgodności towaru ze specyfikacją podaną na etykiecie. W tym celu, opracowano szereg metod pozwalających na określenie pochodzenia geograficznego produktów spożywczych [35,36] oraz weryfikację ich autentyczności [37,38].

W taksonomii zwierząt i roślin pojęcie autentyczności wiąże się z poprawnym przyporządkowaniem zwierząt/roślin do odpowiednich podgrup ewolucyjnych (królestwo, typ, gromada, rząd, rodzina, gatunek) na podstawie wykonanych badań. Dziedziną taksonomii opierającą się na badaniu składu chemicznego jest chemotaksonomia, która wykorzystuje informacje o składzie chemicznym w celu ulepszenia systematyki organizmów żywych. Najczęściej wykonywanymi badaniami w taksonomii jest analiza metabolitów pierwotnych i wtórnych oraz nośników informacji genetycznej (kwasy nukleinowe i białka) za pomocą różnego rodzaju technik chromatograficznych [39,40].

W przemyśle petrochemicznym, ze względu na złożoność analizowanych produktów, analiza chromatograficznych odcisków palca jest bardzo często wykorzystywana m.in. do identyfikacji zafałszowania benzyny poprzez dodawanie rozpuszczalników organicznych [41], zafałszowania oleju napędowego poprzez dodawanie jadalnego oleju roślinnego [42], czy wykrycia prób nielegalnego usuwania dodatków akcyzowych [43].

Duża liczba publikacji poświęconych wykorzystaniu sygnałów chromatograficznych w połączeniu z narzędziami chemometrycznymi w kontekście badania autentyczności, świadczy o dużej potrzebie szukania tego typu rozwiązań do opracowywania nowych metodyk analitycznych.



## 4. Cele pracy

W ramach swojej pracy doktorskiej zamierzam zrealizować następujące cele badawcze:

- ustalenie optymalnego zestawu metod chemometrycznych w celu ekstrakcji użytecznej informacji ze złożonych sygnałów chromatograficznych, w kontekście weryfikacji specyfikacji wybranych produktów,
- opracowanie wieloparametrowych modeli diagnostycznych wspomagających wykrywanie procederu odbarwiania paliw na podstawie chromatograficznych odcisków palca (GC-FID, GC-MS),
- poszukiwanie obszarów sygnałów chromatograficznych, które różnicują grupy badanych próbek w kontekście badania autentyczności,
- potwierdzenie autentyczności preparatu Viagra<sup>®</sup> na podstawie chromatograficznych profili zanieczyszczeń,
- opracowanie i wykazanie użyteczności systemu eksperckiego skonstruowanego bazując na chromatograficznych odciskach palca i ich modelowanie z wykorzystaniem wybranych metod uczenia maszynowego w kontekście oceny ryzyka skażenia wody tributyllocyną i usprawnienia funkcjonowania laboratorium.

## 5. Przeprowadzone badania własne

W moich badaniach w głównej mierze skupiłam się na opracowaniu i wykorzystaniu chromatograficznych odcisków palca do weryfikacji autentyczności m.in. oleju napędowego i preparatu Viagra® [44]. W celu skutecznej ekstrakcji informacji ze zbioru chromatograficznych odcisków palca wykorzystałam odpowiednio dobrane narzędzia chemometryczne pozwalające wstępnie przygotować dane, skonstruować oraz rygorystycznie zwalidować opracowane modele dyskryminacyjne. W kolejnych podrozdziałach niniejszego opracowania przedstawię proponowane przeze mnie strategie wspomagające identyfikację zafałszowania paliw z użyciem chromatograficznych odcisków palca rejestrowanych techniką GC-FID oraz strategię wykorzystania chromatograficznych odcisków palca opisujących profil zanieczyszczeń preparatu Viagra®, rejestrowanych metodą HPLC-DAD do weryfikacji ich autentyczności.

### 5.1 Identyfikacja procederu fałszowania oleju napędowego

W krajach Unii Europejskiej cena paliwa zależy od czynników ekonomicznych i przepisów prawnych określających wysokość podatku akcyzowego. Na ogół, w wielu krajach przyjęte są różne stawki podatku akcyzowego ze względu na przeznaczenie paliwa. W Polsce na olej napędowy, wykorzystywany do celów grzewczych i napędzania maszyn rolniczych, nałożona jest niższa kwota podatku niż na olej napędowy przeznaczony do regularnego transportu. W celu łatwego wizualnego odróżnienia oleju napędowego ze względu na jego przeznaczenie dodawane są do niego komponenty fiskalne takie jak czerwony barwnik (Solvent Red 19 lub Solvent Red 164) oraz marker (Solvent Yellow 124) [45]. Stężenie tych komponentów w paliwie jest ściśle określone w Rozporządzeniu Ministra Finansów z dnia 20 sierpnia 2010 r. Obecność tych dodatków nie wpływa na właściwości fizykochemiczne paliwa, ani nie ogranicza jego dalszego przeznaczenia. Znacząca różnica w cenie oleju napędowego stanowi potencjalny impuls do nielegalnej praktyki usuwania dodatków akcyzowych z tańszego paliwa i sprzedawania go po wyższej cenie. Ta procedura jest znana jako tzw. odbarwianie paliwa, gdyż traci ono barwę. Odbarwianie paliwa może być wykonywane poprzez adsorpcję z wykorzystaniem powszechnie dostępnych materiałów. Proceder ten generuje znaczne straty w budżecie państwa, a zatem rozwój metod analitycznych służących do wykrywania zafałszowanego oleju napędowego jest bardzo potrzebny.

W mojej pracy podjęłam próbę opracowania nowego podejścia analitycznego w oparciu o standardową technikę badania składu paliwa jaką jest chromatografia gazowa. Do tej pory nasza grupa badawcza wykazała możliwość oznaczania zawartości barwnika i znacznika w oleju napędowym oraz wykrywania procederu odbarwiania tego paliwa za pomocą obrazów fluorescencyjnych [46]. Natomiast moim zamiarem było opracowanie metody pozwalającej na odróżnienie odbarwianego i nieodbarwianego oleju napędowego na podstawie sygnałów chromatograficznych wykorzystanych jako chemiczne odciski palca analizowanych próbek. W celu rozwiązania podjętego problemu badawczego w pierwszym etapie wykonano eksperyment polegający na symulacji procederu odbarwiania oleju napędowego za pomocą adsorpcji dodatków akcyzowych na złożu wybranego adsorbenta. Chromatograficzne odciski

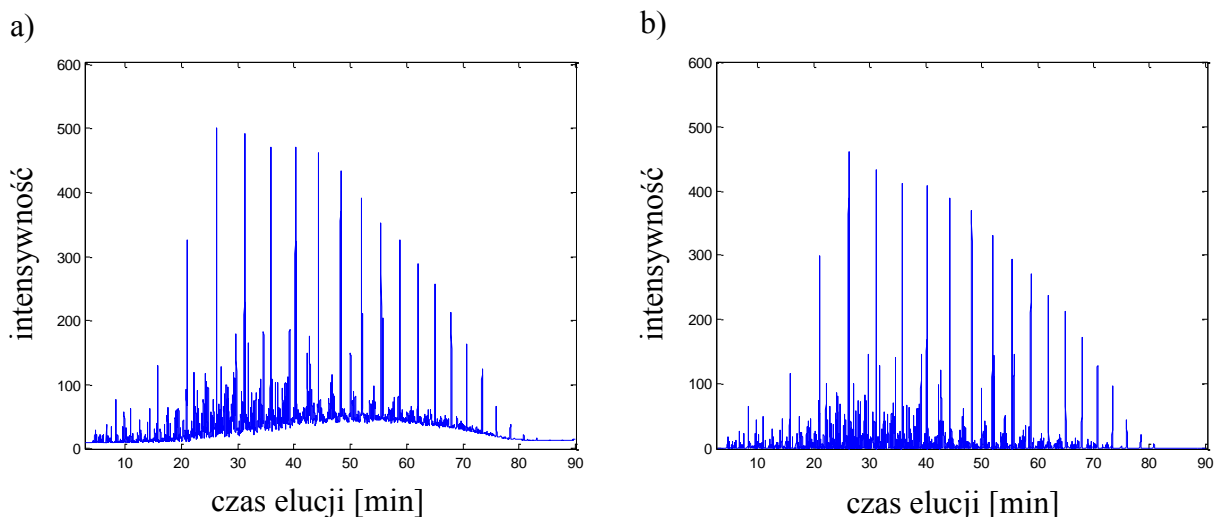
palca rejestrowane były dla próbek paliwa przed i po procesie odbarwiania stosując chromatografię gazowej z detekcją płomieniowo-jonizacyjną. Należy podkreślić, iż technika ta nie pozwala analizować zmian zawartości znacznika i barwnika, ponieważ w temperaturze w jakiej jest prowadzona analiza substancje te ulegają rozkładowi. Proponowana metoda z definicji nie może opierać się na oznaczaniu znacznika i barwnika, gdyż te w procesie wykonywania analizy zostają usunięte. Z tego powodu, proponuje by skupić się na profilach chromatograficznych, które pozwolą w sposób całościowy opisać różnice w ogólnym składzie próbek przed i po procesie odbarwiania. Takie podejście umożliwi skupienie się na fluktuacjach składu chemicznego paliwa wywołanych jego odbarwianiem. Eksperyment polegający na symulowanym odbarwianiu oleju napędowego wykonano w Laboratorium Izby Celnej w Białej Podlaskiej. Analizie poddano łącznie 31 próbek oleju napędowego pochodzących od różnych polskich dostawców paliwa pobranych zgodnie z wymogami określonymi w normie PN-EN ISO 3170: 2004. Chromatograficzne odciski palca zarejestrowano za pomocą chromatografu typu Agilent Technologies 6890N wyposażonego w detektor płomieniowo-jonizacyjny (GC-FID) i kolumnę Restek (60 m × 0,25 mm).

Dla zbioru chromatogramów zmniejszono częstość próbkowania poprzez redukcję liczby punktów pomiarowych z wykorzystaniem liniowej interpolacji. Początkowo chromatogramy zawierały 104 399 punktów pomiarowych. Ten zabieg zastosowano w celu przyspieszenia prowadzenia dalszych obliczeń oraz lepszego działania stosowanych algorytmów. Analizowano zmienną częstość próbkowania prowadzącą do uzyskania chromatogramów zawierających od 5 000 do 50 000 punktów pomiarowych. Jako optymalną rozdzielczość wybrano 25 000 punktów, co pozwala zachować ogólny kształt chromatogramów.

Wizualna ocena poszczególnych chromatogramów pozwoliła stwierdzić, że wymagają one usunięcia linii podstawowej oraz skorygowania przesunięć pików.

Do usunięcia linii podstawowej zastosowano algorytm asymetrycznych najmniejszych kwadratów z funkcją kary (z ang. *penalized asymmetric least squares*, PAsLS). Analizowano wyniki uzyskane dla różnych wartości parametrów wyjściowych. Najbardziej przybliżony kształt linii podstawowej uzyskano dla  $\lambda = 10\ 000$  i  $p = 0,001$ .

W kolejnym kroku, dokonano korekcji przesunięć pików techniką, maksymalizującą wzajemną korelację sygnałów (z ang. *correlation optimized warping*, COW). Wszystkie chromatogramy były nakładane względem sygnału wzorcowego. Wybrano go na podstawie najwyższej wartości średniej uzyskanych współczynników korelacji z pozostałymi sygnałami. Początkowa liczba sekcji jaką brano pod uwagę w trakcie nakładania wynosiła 500, a szerokość każdej z nich była równa średniej szerokości pików (około 50 punktów pomiarowych). Zadowolające rezultaty nakładania uzyskano wówczas, gdy sygnały GC-FID zostały podzielone na 250 sekcji, a parametr elastyczności był równy 3. Przykładowy chromatogram próbki oleju napędowego przed i po wstępnym przygotowaniu danych przedstawiono na Rys. 3.



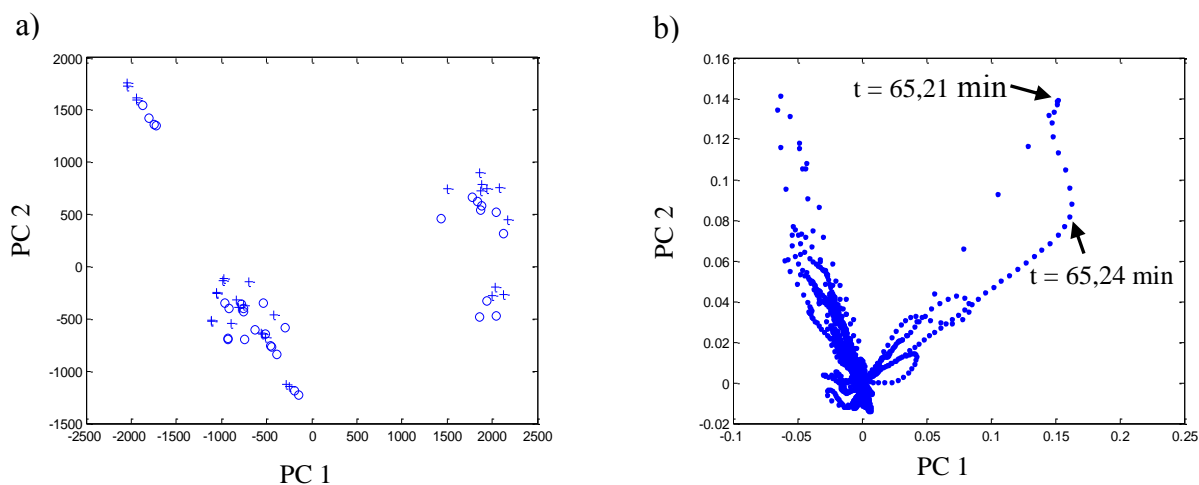
Rys. 3 Przykładowy chromatogram GC-FID próbki oleju napędowego (a) przed i (b) po wstępnym przygotowaniu danych.

Następnie, zbiór chromatograficznych odcisków palca po uprzedniej zmianie częstości próbkowania, usunięciu linii podstawowej i nałożeniu pików zestawiono w macierz danych. Eksploracja zbioru sygnałów GC-FID po ich wstępnym przygotowaniu została przeprowadzona z użyciem analizy czynników głównych. Efektywna kompresja danych wskazuje na dużą liczbę skorelowanych zmiennych (pierwsze dwa czynniki główne opisują 73,7% całkowitej wariancji danych). Projekcja próbek na przestrzeń zdefiniowaną przez pierwsze dwa czynniki główne umożliwia badanie ich chemicznego podobieństwa. Każdy punkt na projekcji PC1-PC 2 (zob. Rys. 4) to pojedynczy chromatogram GC-FID opisujący daną próbkę. Analizując położenie próbek można zaobserwować dwie dobrze oddzielone grupy próbek oleju napędowego wzdłuż osi PC 1. Jednakże, widać wyraźnie, iż na grupowanie się próbek nie ma wpływu przeprowadzony proces odbarwiania oleju napędowego. Taki wniosek można wyjaśnić tendencją grupowania, która nie odzwierciedla podziału próbek ze względu na przeprowadzony proces odbarwiania. Analiza projekcji parametrów opisanych przez pierwsze dwa czynniki główne pozwoliła zidentyfikować zakresy czasów elucji odpowiadające substancjom odpowiedzialnym za grupowanie się próbek. Są to dwa piki eluowane przy czasie około 65,21 i 65,24 min. Pochodzą one od estrów metylowych kwasów tłuszczowych (FAME).

Związki FAME nie mogą być uznane jako potencjalne markery do wykrywania procederu fałszowania oleju napędowego ponieważ są celowo dodawane do paliwa w trakcie jego produkcji. W Polsce dopuszcza się zawartość FAME do 7% (v/v), ale różnice w ich zawartości pomiędzy różnymi partiami oleju napędowego mogą być znaczne i najczęściej zależą od producenta. Grupa próbek oleju napędowego charakteryzowana przez pozytywne wartości PC 1 zawiera FAME, w ilości od 4,1% do 5,3 (v/v).

Grupowanie się próbek ze względu na ich odbarwienie nie zostało zaobserwowane dla projekcji opisanych przez pozostałe zestawy czynników głównych. Fakt, że analiza eksploracyjna z wykorzystaniem metody PCA nie pozwala rozróżnić grup próbek oleju

napędowego nie wyklucza potencjalnej możliwości ich dyskryminacji stosując metody uczenia z nadzorem – np. analizy dyskryminacyjnej.



Rys. 4 Projekcja próbek paliw na płaszczyznę zdefiniowaną przez pierwszy i drugi czynnik główny. Próbki oryginalne oznaczono jako (○), a próbki odbarwione jako (+).

Model PLS-DA skonstruowano dla zbioru złożonego z 21 chromatogramów oryginalnych próbek oleju napędowego wybranych za pomocą algorytmu Kennarda i Stonea oraz 21 chromatogramów tych samych próbek po odbarwieniu. Pozostałe chromatogramy 10 próbek paliwa przed i po procesie odbarwienia utworzyły zbiór testowy użyty w celu określenia właściwości predykcyjnych skonstruowanego modelu dyskryminacyjnego. Optymalny model PLS-DA posiadał kompleksowość 6, którą wybrano za pomocą walidacji krzyżowej typu 'leave-one-out'. Utworzona reguła logiczna pozwoliła poprawnie rozpoznać wszystkie próbki paliwa zarówno ze zbioru testowego jak i modelowego (zob. Tab. 1). W celu określenia zdolności predykcyjnych modeli liczone średnie wartości parametrów walidacyjnych z wykorzystaniem podejścia 'bootstrap'. Każdorazowo ze zbioru modelowego losowano ze zwracaniem 21 próbek nieodbarwionych, których chromatogramy przed i po odbarwieniu były wykorzystywane do konstrukcji modelu dyskryminacyjnego. Proces ten powtarzano 1000 razy otrzymując 1000 zestawów parametrów modeli PLS-DA.

Uzyskane rezultaty dyskryminacji wspierają hipotezę, że na podstawie sygnałów chromatograficznych GC-FID próbek oleju napędowego można rozróżnić próbki odbarwione od nieodbarwionych.

Następnie, zastosowano różne metody wyboru zmiennych istotnych w celu uniknięcia możliwości przeuczenia modelu dyskryminacyjnego i ustalenia istotnych regionów czasu elucji, w których uległy wymyciu ważne dla dyskryminacji próbki związku. W metodzie UVE-PLS-DA do każdego z 1000 wylosowanych zbiorów modelowych dodawano macierz danych o takiej samej liczbie wierszy co macierz stanowiąca zbiór modelowy i liczbie kolumn równej 10 000. Macierz ta zawierała zmienne nieistotne, którymi były liczby wylosowane z rozkładu normalnego, pomnożone przez współczynnik  $c=10^{-12}$ . Następnie skonstruowano model PLS-DA a jako istotne wybierano zmienne, których wartość bezwzględna stabilności

liczona na podstawie uzyskanych współczynników regresji była większa niż maksymalna wartość stabilności dla zmiennych nieistotnych.

Ten sam sposób wielokrotnego losowania ze zwracaniem zastosowano w pozostałych metodach wyboru zmiennych. W metodach VIP i SR w pierwszym kroku konstruowany jest model PLS-DA o założonej kompleksowości, a następnie dla każdej zmiennej obliczany jest odpowiedni parametr istotności odpowiednio zdefiniowany dla poszczególnej metody. Ze względu na zastosowanie podejścia ‘bootstrap’, dla każdej zmiennej uzyskano 1000 wartości zarówno parametru VIP jak i SR. Jako istotne wybierane były te zmienne, których wartości średnie danych parametrów przekraczały odpowiednie wartości progowe. W tym przypadku analizując różne wartości progowe jako optymalne dla metody VIP wybrano 0,8, natomiast dla SR 1,0. Współczynnik selektywności liczony był raz dla wszystkich zmiennych natomiast dla metody VIP zastosowano trzy iteracje, w których do kolejnej iteracji wykorzystywano zmienne uznane za istotne w poprzedniej. Wszystkie metody wyboru zmiennych realizowane w trakcie konstrukcji dyskryminacyjnego wariantu metody częściowych najmniejszych kwadratów, UVE-PLS-DA, SR-PLS-DA i VIP-PLS-DA wykazują ich potencjalną możliwość do wykrywania nielegalnego procederu odbarwiania paliw. Otrzymane wyniki dla każdej z zastosowanych metod dyskryminacji wraz z parametrami opisującymi modele dyskryminacyjne przedstawiono w Tabeli 1.

**Tabela 1** Wyniki modeli PLS-DA bez i z zastosowaniem metod wyboru zmiennych (UVE - metoda eliminacji zmiennych nieistotnych, VIP – metoda zmiennych znaczących dla projekcji, SR – współczynnik selektywności, n - liczba zmiennych, f - liczba czynników głównych, SE – selektywność, SP – specyficzność, AUC – parametr AUC).

Model	n	f	Zbiór modelowy			Zbiór testowy		
			SE [%]	SP [%]	AUC	SE [%]	SP [%]	AUC
PLS-DA	25 000	6	100,00	100,00	1,000	100,00	100,00	1,000
UVE	14	9	100,00	1,000	1,000	100,00	1,000	1,000
VIP	265	6	90,00	0,996	0,996	90,00	0,996	0,970
SR	16	3	100,00	1,000	1,000	100,00	1,000	1,000

Najgorsze parametry predykcyjne uzyskano dla modelu skonstruowanego za pomocą metody VIP-PLS-DA, wykorzystując do tego największą liczbę zmiennych istotnych. Pozostałe dwie metody wykazały czułość, procent poprawnej klasyfikacji równe 100% dla zbioru testowego dla niewielkiej liczby zmiennych (14 i 16 zmiennych). Dokładniejsza analiza wybranych zmiennych wykazała, że zmienne wybrane w metodzie SR-PLS-DA odpowiadają czasom elucji substancji polarnych, których stężenie obniża się w procesie odbarwiania paliwa co może być spowodowane przez ich adsorpcje na złożu, na którym są także sorbowane dodatki

akcyzowe paliwa. Dlatego, wydaje się, że metoda PLS-DA wraz z metodą wyboru zmiennych opartą na współczynniku selektywności jest najlepszą metodą do wykrywania nielegalnego odbarwienia oleju napędowego. W Tabelach 2 i 3 wyszczególniono związki odpowiadające czasom elucji wybranym przez metody UVE-PLS-DA i SR-PLS-DA. Przebadanie większej liczby próbek oleju napędowego pozwoliłoby dokładniej wyznaczyć parametry proponowanej metodologii. Jednakże uzyskane wyniki wyraźnie wykazują duży potencjał stosowanej metody, która może wspomóc wymiar sprawiedliwości.

**Tabela 2** Zidentyfikowane związki chemiczne wspierające diagnostykę procesu odbarwiania przy użyciu metody UVE-PLS-DA (NI - nie zidentyfikowane).

Numer piku	Czas elucji [min]	Zidentyfikowany związek
1	23,937	1-metylo-3-propylobenzen (C <sub>10</sub> H <sub>14</sub> )
	23,940	
	23,944	
2	24,786	1-metylo,4-propylobenzen (C <sub>10</sub> H <sub>14</sub> )
	24,789	
	24,793	
	24,796	
3	25,398	1-etylo-2,4-dimetylobenzen (C <sub>10</sub> H <sub>14</sub> )
	25,402	
4	27,556	1,2,3,5-tetrametylobenzen (C <sub>10</sub> H <sub>14</sub> )
	27,559	
	27,563	
5	40,881	n-paraffiny (C <sub>14</sub> )
	40,884	

**Tabela 3** Zidentyfikowane związki chemiczne wspierające diagnostykę procesu odbarwiania przy użyciu metody SR-PLS-DA (NI - związki nie zidentyfikowane).

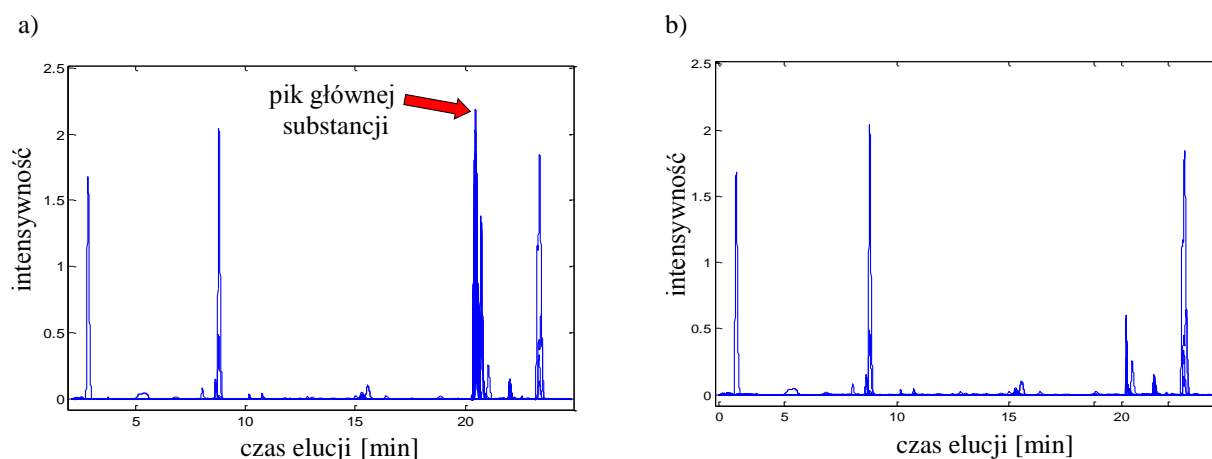
Numer piku	Czas elucji [min]	Zidentyfikowany związek
1	7,052	NI
	7,055	
2	23,982	4-etyloheptan (C <sub>9</sub> H <sub>20</sub> ) lub 1-octano, 2-butyl (C <sub>12</sub> H <sub>26</sub> )
	23,985	
	23,989	
	23,991	
	23,996	
	23,999	
3	32,292	fitol (C <sub>20</sub> H <sub>40</sub> O)
	32,296	
	32,299	
4	32,891	związki zawierające tlen np. 1-propeno,2-nitro-3-(1-cyklooktanyl) (C <sub>11</sub> H <sub>17</sub> NO <sub>2</sub> )
	32,894	
	32,898	
5	38,508	NI
6	47,058	3-metylopentadekan (C <sub>16</sub> H <sub>34</sub> )



## 5.2 Identyfikacja zafalszowań leku Viagra®

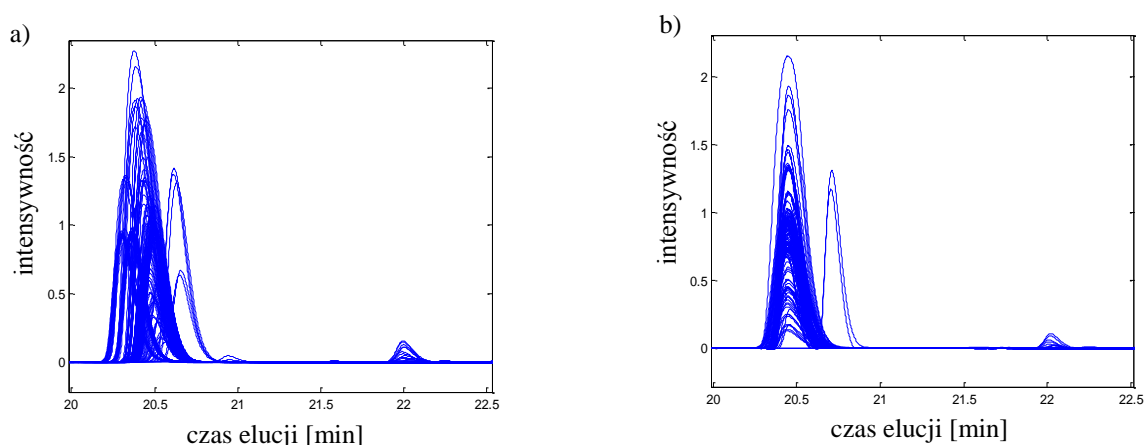
Produkcja zafalszowanych leków jest ogromnym problemem, który niesie poważne zagrożenie dla zdrowia publicznego. Znaczący wzrost dostępności na rynku leków, które nie spełniają wymogów jakości w ostatnich latach może być związany z łatwiejszym dostępem fałszerzy do technologii, które są wykorzystywane do ‘kopiowania’ oryginalnych leków [47]. W celu badania zgodności wytwarzanych leków z określonymi normami stosuje się zazwyczaj wiele technik instrumentalnych [48]. Produkty farmaceutyczne wytwarzane w postaci stałej (tabletki) są analizowane pod względem ich autentyczności poprzez zastosowanie różnego rodzaju technik instrumentalnych takich jak fluorescencja rentgenowska z dyspersją energii, ED-XRF (z ang. *energy-dispersive X-ray spectroscopy*) [49], spektroskopia magnetycznego rezonansu jądrowego, NMR (z ang. *nuclear magnetic resonance*) [50] czy spektroskopia Ramana [51]. Dla leków w postaci tabletek autentyczność może również być potwierdzana przez badanie ich parametrów fizycznych, takich jak grubość tabletki, jej długość oraz masa [34, 36] oraz poprzez analizę porównawczą zdjęć tabletek wykonanych w tych samych warunkach [52]. W moich badaniach do określenia składu zarówno ilościowego jak i jakościowego badanego leku wykorzystywano wysokosprawną chromatografię cieczową z matrycą diodową, HPLC-DAD (z ang. *high-performance liquid chromatography with diode-array detection*). Badanym materiałem było 46 oryginalnych i 97 zafalszowanych próbek leku Viagra®. Głównym celem badań było opracowanie nowych podejść chemometrycznych pozwalających zweryfikować autentyczność badanych próbek na podstawie sygnałów chromatograficznych traktowanych jako chromatograficzne odciski palca.

Analizie poddano chromatograficzne profile zanieczyszczeń próbek Viagry®, tzn. sygnałów otrzymanych poprzez usunięcie piku głównego składnika leku, Sildenafilu, który występuje przy czasie elucji ok. 22,5 min. Tym samym, prowadzona analiza ukierunkowana była na określenie substancji różnicujących próbki autentyczne i zafalszowane, które mogą stanowić markery tego nielegalnego procederu. Chromatogramy przed i po usunięciu piku przedstawia Rys. 5.



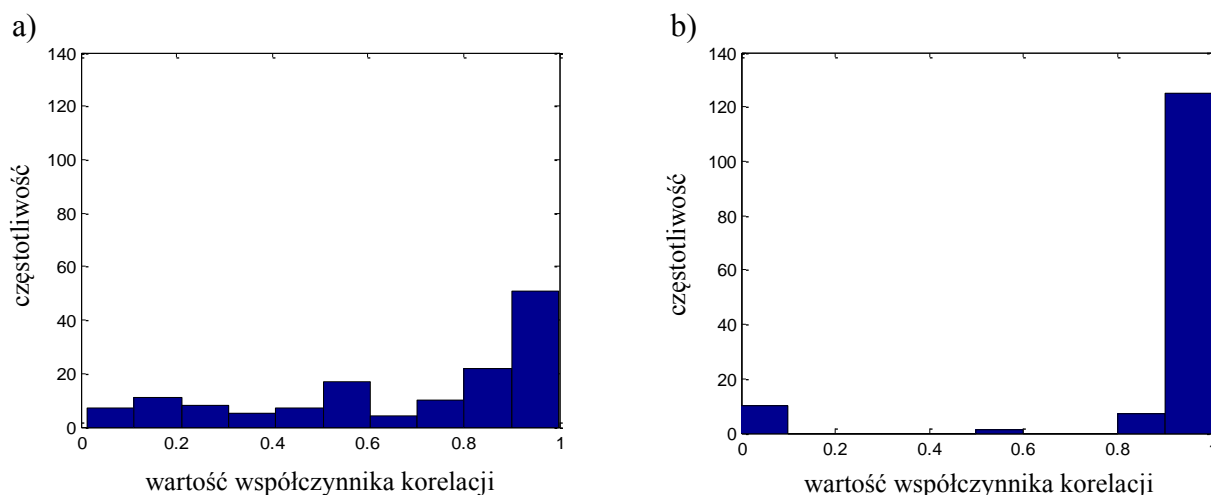
Rys. 5 Chromatogramy HPLC-DAD leku Viagra® (a) przed i (b) po usunięciu piku głównego składnika, Sildenafilu.

Analizę chromatograficzną próbek dokonano w Instytucie Zdrowia Publicznego w Brukseli. Uzyskane dane zostały wstępnie przygotowane i poddane analizie chemometrycznej. Ponieważ chromatograficzne odciski palca miały tę samą liczbę punktów pomiarowych (13 620), które odpowiadały tym samym czasom elucji, nie wymagały one ponownego próbkowania. Z powodu różnic występujących w kształcie linii podstawowej oraz jej relatywnie dużej intensywności do jej usunięcia zastosowano metodę częściowych najmniejszych kwadratów, PAsLS. Analizując różne parametry jak najbardziej odpowiedni zestaw uznano  $\lambda = 100\ 000$  i  $p = 0,001$ . Kolejnym krokiem wstępnego przygotowania danych była eliminacja przesunięć pików chromatograficznych, które głównie występowały dla pików głównej substancji aktywnej zawartej w badanym leku. W tym celu zastosowano metodę COW. Spośród wszystkich przeanalizowanych kombinacji wartości parametru elastyczności 's' i liczby sekcji najlepsze wyniki uzyskano gdy sygnały były podzielone na 28 części (większość sekcji zawierała 500 punktów pomiarowych), a parametr elastyczności wynosił 3. Dodatkowo, skorygowano przesunięcia pików w zakresach sygnałów od 4800 do 5000 oraz od 10 800 do 11 300 punktu pomiarowego z liczbami sekcji równymi odpowiednio 10 i 20 oraz parametrami elastyczności równymi 3 i 6. Na Rys. 6 przedstawiono fragment chromatograficznego odcisku palca zawierający pik pochodzący od głównej substancji przed i po wstępnym przygotowaniu danych.



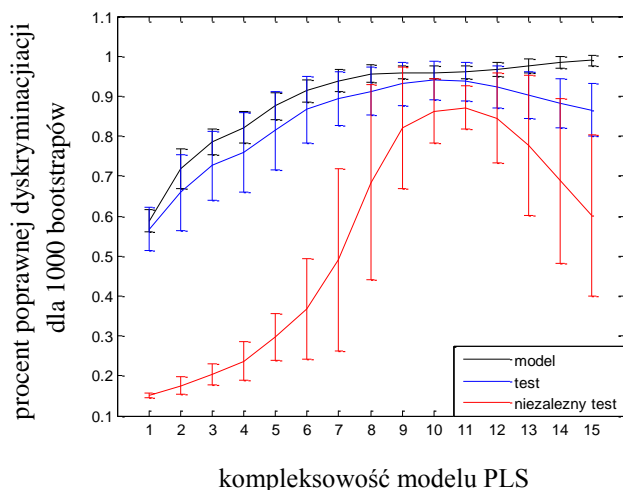
Rys. 6 Fragment chromatogramów zawierający pik pochodzący od głównej substancji aktywnej: (a) przed i (b) po wstępnym przygotowaniu danych.

Warto zauważyć, że piki występujące przy czasie elucji ok. 20,7 min, które nie zostały nałożone pochodzą od dodatkowych substancji mogących stanowić zanieczyszczenia. Świadczy o tym fakt, że dany pik nie jest obecny w chromatogramach oryginalnych próbek Viagry®. Początkowe wartości współczynników korelacji pomiędzy profilami zanieczyszczeń, a sygnałem względem, którego były one nakładane zawierały się w przedziale od 0,0134 do 0,9988. Natomiast po usunięciu linii podstawowej i korekcji przesunięć pików, 88% próbek miało współczynniki korelacji powyżej 0,9. W celu ułatwienia oceny poprawy jakości danych histogramy współczynników korelacji przed i po ich wstępnym przygotowaniu przedstawiono na Rys. 7.



Rys. 7 Histogramy współczynników korelacji obliczone między każdym chromatogramem, a sygnałem referencyjnym: (a) przed i (b) po zastosowaniu metody COW.

Po usunięciu linii podstawowej i skorygowaniu przesunięć pików w zbiorze sygnałów HPLC-DAD dla zafałszowanych i oryginalnych leków Viagra<sup>®</sup> usunięto fragment chromatogramu odpowiadający substancji czynnej badanego leku. Tak przygotowane profile zanieczyszczeń próbek użyto do budowy wieloparametrowego modelu z wykorzystaniem dyskryminacyjnego wariantu metody częściowych najmniejszych kwadratów. W celu zapewnienia reprezentatywności model dyskryminacyjny został skonstruowany na podstawie zbioru próbek wybranych przy użyciu algorytmu Kennarda i Stonea. Model miał na celu odróżnienie oryginalnych i zafałszowanych próbek Viagry<sup>®</sup>. Oryginalne próbki leku Viagra<sup>®</sup> zostały oznaczone jako '1', a zafałszowane jako '-1'. Do zbioru modelowego wybrano po 35 próbek ze zbioru próbek zafałszowanych i autentycznych, z których do konstrukcji pojedynczego modelu wybierano po 25 próbek z każdej z grup, a pozostałe 20 próbek stanowiło wewnątrz zbiór testowy, zależny od próbek wybranych do konstrukcji modelu. Niezależny zbiór testowy, który nie brał udziału w konstrukcji modelu zawierał 11 próbek autentycznych i 62 zafałszowanych, które nie zostały wybrane metodą Kennarda i Stonea. Zarówno wewnątrz jak i niezależny zbiór testowy służył do walidacji zdolności predykcyjnych skonstruowanego modelu dyskryminacyjnego. Optymalna liczba ukrytych zmiennych, potrzebnych do budowy modelu PLS-DA, została wybrana na podstawie procedury walidacji typu Monte Carlo. Podejście to polegało na wielokrotnym wyborze zbioru modelowego i testowego z 70 próbek wybranych metodą Kennarda i Stonea (1000 razy), a następnie konstrukcji modeli dla założonego przedziału czynników PLS. Taki sposób walidacji pozwolił wykreślić zależność pomiędzy liczbą ukrytych zmiennych, a procentem poprawnej dyskryminacji z uwzględnieniem przedziału niepewności wartości parametru CCR (zob. Rys. 8). Analiza rozkładu błędu dla modeli o danej kompleksowości pozwoliła ustalić optymalną liczbę ukrytych zmiennych równą 10.



Rys. 8 Wykres zależności procentu poprawnej dyskryminacji próbek od kompleksowości modelu PLS-DA niebieska linia - dla próbek wewnętrznego zbioru testowego, czarna linia - dla próbek zbioru modelowego, czerwona linia - dla próbek niezależnego zbioru testowego.

Pomimo, iż skonstruowany model PLS-DA ma bardzo dobre zdolności predykcyjne, z powodu dużej liczby zmiennych w stosunku do liczby analizowanych próbek może występować tzw. ryzyko przeuczenia modelu. Rozwiązaniem tego problemu może być ograniczenie liczby zmiennych poprzez wybór zmiennych istotnych dla budowy modelu dyskryminacyjnego. W pracy analizowano możliwości zastosowania czterech metod sposobu wyboru zmiennych w kontekście budowy modelu PLS-DA: metodę eliminacji zmiennych nieistotnych (UVE), metodę zmiennych znaczących dla projekcji (VIP), współczynnik selektywności (SR) oraz metodę korelacji wieloczynnikowej (SMC) [53]. W połączeniu z metodą PLS-DA pozwalały one uzyskać modele dyskryminacyjne o bardzo dobrych parametrach predykcyjnych, co obrazują wyniki przedstawione w Tabeli 5. W metodzie UVE-PLS-DA do każdego z 1000 wylosowanych zbiorów modelowych dodano macierz zmiennych nieistotnych (10000 zmiennych dla każdego sygnału), którymi były liczby losowe o rozkładzie normalnym, pomnożone przez współczynnik  $c=10^{-12}$ . Następnie konstruowano model PLS-DA a jako istotne wybierano zmienne, których wartość bezwzględna stabilności liczona na podstawie uzyskanych współczynników regresji była większa niż maksymalna wartość stabilności dla zmiennych nieistotnych. Ten sam sposób wielokrotnego losowania ze zwracaniem zastosowano w pozostałych metodach wyboru zmiennych. W metodach VIP, SMC i SR w pierwszym kroku konstruowany jest model PLS-DA o założonej kompleksowości, a następnie dla każdej zmiennej obliczany jest odpowiedni parametr istotności zależny od zastosowanej metody. Ze względu na zastosowanie walidacji krzyżowej typu Monte Carlo, dla każdej zmiennej uzyskano 1000 wartości każdego parametru. Jako istotne wybierane były te zmienne, których wartości średnie danych parametrów przekraczały odpowiednie wartości progowe. W tym przypadku jako optymalną wartość progową dla metody VIP wybrano 1 natomiast dla SR 0.9 natomiast w metodzie SMC wartość progowa została oszacowana na ok. 4.042. Dodatkowym kryterium wyboru zmiennych było zliczanie

częstotliwości ich wyboru w trakcie 1000 powtórzeń walidacji krzyżowej typu Monte Carlo. W tym przypadku jako istotne zmienne uznawane były te, które były wybierane co najmniej podczas 95% wszystkich losowań. W celu oceny zdolności predykcyjnych uzyskanych modeli obliczano średnie wartości parametru AUC dla modelu i testu uzyskane dla konstruowanych modeli. Wszystkie zastosowane metody wyboru zmiennych pozwoliły uzyskać modele o podobnych wartościach czułości i selektywności. Małe wartości parametru selektywności w stosunku do uzyskanych wartości specyficzności dla niezależnego zbioru testowego mogą być wynikiem nieproporcjonalnej liczby próbek z poszczególnych grup w nim zawartych. Jednak metody VIP-PLS-DA i SR-PLS-DA pozwoliły na wybór mniejszej liczby zmiennych niż pozostałe, dlatego wydaje się, że PLS-DA połączona z metodą wyboru zmiennych opierającą się na współczynniku selektywności lub zmiennych znaczących dla projekcji jest rekomendowana do wykrywania nielegalnego procederu fałszowania Viagry®. Wysokie wartości parametrów walidacyjnych dowodzą, że opracowane modele PLS-DA, skonstruowane na podstawie sygnałów HPLC-DAD z wykorzystaniem metod wyboru zmiennych, mają potencjał praktyczny.

**Tabela 5** Wyniki uzyskane dla modeli dyskryminacyjnych PLS-DA bez i z zastosowaniem metod wyboru zmiennych (UVE – metoda eliminacji zmiennych nieistotnych, VIP – metoda zmiennych znaczących dla projekcji, SR – współczynnik selektywności, SMC – metoda korelacji wieloczynnikowej, n - liczba zmiennych, f - liczba czynników głównych, CCR – procent poprawnej klasyfikacji, SE – selektywność, SP - specyficzność, AUC – parametr AUC).

Model	f	n	Zbiór modelowy (Monte Carlo)				Zbiór testowy (Monte Carlo)				Niezależny zbiór testowy			
			CCR [%]	SE [%]	SP [%]	AUC	CCR [%]	SE [%]	SP [%]	AUC	CCR [%]	SE [%]	SP [%]	AUC
PLS-DA	10	13 291	95,85	94,57	97,12	0,9983	94,04	94,23	93,84	0,983	86,02	47,85	93,00	0,967
			± 1,60	± 2,70	± 1,82	± 0,001	± 4,81	± 6,33	± 8,20	± 0,035	± 8,02	± 11,09	± 11,22	± 0,009
UVE	2	1 552	95,66	94,27	97,06	0,991	95,84	94,33	97,35	0,989	87,67	45,45	95,16	0,9520
			± 1,48	± 2,48	± 1,77	± 0,005	± 3,70	± 6,20	± 4,42	± 0,017	± 0,00	± 0,00	± 0,00	± 0,008
VIP	6	12	95,77	91,76	99,78	0,983	93,10	90,02	96,18	0,925	90,89	45,65	98,91	0,848
			± 1,74	± 3,34	± 0,95	± 0,013	± 4,68	± 7,75	± 6,74	± 0,060	± 1,09	± 1,61	± 1,29	± 0,070
SR	6	95	97,14	94,27	100,00	0,995	96,55	93,81	99,28	0,975	90,56	45,45	98,56	0,594
			± 1,24	± 2,48	± 0,00	± 0,005	± 3,71	± 7,02	± 2,62	± 0,037	± 0,92	± 0,00	± 0,01	± 0,048
SMC	1	1 156	95,70	91,41	100,00	0,991	95,74	91,48	100,00	0,9911	91,65	45,45	99,85	0,957
			± 1,45	± 2,89	± 0,00	± 0,005	± 3,61	± 7,23	± 0,00	± 0,013	± 0,40	± 0,00	± 0,47	± 0,006

## 6. Podsumowanie

Wyniki prowadzonych przeze mnie badań dowodzą, że metody chemometryczne stanowią znaczne wsparcie dla procesu identyfikacji i dyskryminacji próbek bazując na sygnałach instrumentalnych jakimi są chromatograficzne odciski palca. Odpowiednio dobrane metody chemometryczne dają możliwość eliminacji niepożądanych komponentów sygnałów instrumentalnych (szumu i linii podstawowej), które mogą negatywnie wpływać na późniejszą interpretację danych chromatograficznych.

Zastosowane techniki chromatograficzne doskonale nadają się do rejestracji sygnałów instrumentalnych stanowiących chemiczne odciski palca ze względu na proporcje pomiędzy liczbą pików a liczbą substancji (jeden pik odpowiada jednemu związkowi chemicznemu). Wszystkie te cechy świadczą o unikalnych cechach chromatograficznych odcisków palca w kontekście ogólnego opisu składu badanych próbek.

Proponowane w pracy rozwiązania oparte na zaawansowanym modelowaniu chromatograficznych odcisków palca narzędziami chemometrycznymi pozwalają na eliminację konieczności przeprowadzania analizy jakościowej.

Wybór istotnych zmiennych dla danego problemu dyskryminacyjnego może skrócić czas rozdziału chromatograficznego i wskazać obszary czasów elucji zawierających frakcje związków odpowiedzialnych za różnicowanie grup. Ich dokładna analiza i interpretacja ułatwia poznanie potencjalnych markerów procesu fałszowania.

Skonstruowane modele dyskryminacyjne, wykorzystane do dyskryminacji zafałszowanych próbek leków, pozwoliły z bardzo dobrą efektywnością określić autentyczność badanych próbek na podstawie profili zanieczyszczeń.

Proponowane strategie analizy zafałszowań oleju napędowego i Viagry<sup>®</sup> mogą być także implementowane w innych obszarach, w których problem badawczy dotyczy wykrywania zafałszowań określonych obiektów za pomocą innego typu odcisków palca.

## **7. Dalsze plany naukowe**

W moich dalszych badaniach zamierzam:

- zaproponować odpowiednią strategię analizy złożonych sygnałów chromatograficznych uzyskanych za pomocą detektora DAD, która umożliwi badanie/potwierdzenie specyfikacji wybranych produktów,
- wykazać użyteczność systemu eksperckiego skonstruowanego w oparciu o wybrane metody uczenia maszynowego w kontekście oceny ryzyka skażenia wody i usprawnienia funkcjonowania laboratorium.



## **8. Dorobek naukowy**

### **Publikacje zamieszczone w czasopismach z listy filadelfijskiej**

1. 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, *Analytical and Bioanalytical Chemistry*, 407 (2015) 1159-117, IF = 3,578
2. R. Sitko, B. Gliwińska (Krakowska), B. Zawisza, B. Feist, Ultrasound-assisted solid-phase extraction using multiwalled carbon nanotubes for determination of cadmium by flame atomic absorption spectrometry, *Journal of Analytical Atomic Spectrometry*, 28 (2013) 405-410, IF = 3,155

### **Publikacje skierowane do druku w czasopismach z listy filadelfijskiej**

1. M. Daszykowski, M. Korzeń, B. Krakowska, K. Fabiańczyk, Expert system for monitoring the tributyltin content in inland water samples, *Chemometrics and Intelligent Laboratory Systems*, IF = 2,381

### **Prezentacje ustne na konferencjach naukowych**

1. D. Custers, B. Krakowska, P. Courselle, M. Daszykowski, S. Apers, E. Deconinck, Testing of complementarity of PDA and MS detectors using chromatographic fingerprinting of genuine and counterfeit Viagra<sup>®</sup>, XXXVIII Sympozjum nt. Chromatograficzne metody badania związków organicznych, Szczyrk (26-29 maj 2015)
2. B. Krakowska, J. Orzel, M. Daszykowski, I. Grabowski, G. Zaleszczyk, M. Sznajder, Wykrywanie procederu fałszowania oleju napędowego przy użyciu metod chemometrycznych, Zjazd Wiosenny SSPTChem, 2014, Zawoja (9-13 kwiecień 2014)

### **Plakaty naukowe**

1. B. Krakowska, M. Daszykowski, K. Fabiańczyk, M. Korzeń, Identification of tributyltin in environmental water samples supported by means of discriminant models, XXXVIII Sympozjum nt. Chromatograficzne metody badania związków organicznych, Szczyrk (26-29 maj 2015)

2. B. Krakowska, I. Stanimirova, D. Custers, E. Deconinck, M. Daszykowski, Wykorzystanie profilowania zanieczyszczeń do potwierdzenia autentyczności Viagry, IX Seminarium Naukowe „Aktualne Problemy Chemii Analitycznej, Katowice (15 maj 2015)
3. B. Krakowska, M. Daszykowski, Wykorzystanie metod chemometrycznych w kontekście kontroli jakości produktów farmaceutycznych, III Ogólnopolska Konferencja „Pomiędzy naukami – Zjazd Fizyków i Chemików”, Chorzów (26 wrzesień 2014)
4. B. Krakowska, I. Stanimirova, D. Custers, E. Deconinck, M. Daszykowski, Detection of counterfeit medicines based on chromatographic impurity profiles, XXXVII Sympozjum nt. Chromatograficzne metody badania związków organicznych, Szczyrk (11-13 czerwiec 2014)
5. B. Krakowska, I. Stanimirova, M. Daszykowski, I. Grabowski, G. Zaleszczyk, M. Sznajder, Porównanie metod wyboru zmiennych na przykładzie ich zastosowania do budowy modelu dyskryminacyjnego, VIII Seminarium Naukowe „Aktualne Problemy Chemii Analitycznej, Katowice (16 maj 2014)
6. B. Krakowska, J. Orzeł, M. Daszykowski, I. Grabowski, G. Zaleszczyk, M. Sznajder, Partial Least Squares for detection of diesel oil adulteration, PLS 2014, Paryż (19-21 maj 2014), Francja
7. B. Krakowska, M. Daszykowski, I. Grabowski, G. Zaleszczyk, M. Sznajder, Zastosowanie metod chemometrycznych do wykrywania procederu fałszowania oleju napędowego, Zjazd Zimowy SSPTChem, Łódź (7 grudzień 2013)
8. B. Krakowska, M. Daszykowski, Wykorzystanie metod chemometrycznych w analizie jakości wybranych produktów, II Ogólnopolska Konferencja „Pomiędzy naukami – Zjazd Fizyków i Chemików”, Chorzów (27 wrzesień 2013)
9. B. Krakowska, M. Daszykowski, I. Grabowski, G. Zaleszczyk, M. Sznajder, Comparative analysis of diesel oil samples of different origin based on chromatographic fingerprints, XXXVI Sympozjum nt. Chromatograficzne metody badania związków organicznych, Szczyrk (5 czerwiec 2013)
10. B. Krakowska, M. Daszykowski, I. Grabowski, G. Zaleszczyk, M. Sznajder, Wykrywanie procederu fałszowania oleju napędowego przy użyciu metod chemometrycznych, VII Seminarium Naukowe „Aktualne Problemy Chemii Analitycznej, Katowice (17 maj 2013)

11. B. Krakowska, M. Daszykowski, I. Grabowski, G. Zaleszczyk, M. Sznajder, Eksploracja danych chromatograficznych opisujących próbki oleju napędowego, Zjazd Wiosenny SSPTChem, Augustów (11-18 kwiecień 2013)
12. B. Gliwińska (Krakowska), K. Kocot, B. Feist, R. Sitko, Zagęszczanie jonów miedzi i kadmu z wykorzystaniem wielościennych nanorurek węglowych techniką absorpcyjnej spektrometrii atomowej, VI Seminarium Naukowe „Aktualne Problemy Chemii Analitycznej”, Katowice (18 maj 2012)

## 9. Literatura

---

- [1] G. Zadora, Laundering of “Illegal” Fuels –a Forensic Chemistry Perspective, *Acta. Chim. Slov.*, 54 (2007) 110–113.
- [2] P.-Y. Sacré, E. Deconinck, M. Daszykowski, P. Courselle, R. Vancauwenberghe, P. Chiap, J. Crommen, J. O. De Beer, Impurity fingerprints for the identification of counterfeit medicines - A feasibility study, *Anal. Chim. Acta*, 7019 (2011) 224–231.
- [3] E.-S.M. Abdel-Aal, H.M. Ziena, M.M. Youssef, Adulteration of honey with high-fructose corn syrup: Detection by different methods, *Food Chemistry*, 2 (1993)209–212.
- [4] P.L. Pisano, M.F. Silva, A.C. Olivieri, Anthocyanins as markers for the classification of Argentinean wines according to botanical and geographical origin. Chemometric modeling of liquid chromatography–mass spectrometry data, *Food Chem.*, 175 (2015) 174–180.
- [5] M. Daszykowski, B. Walczak, Use and abuse of chemometrics in chromatography, *Trends Anal. Chem.*, 25 (2006) 1081–1096.
- [6] C. Tistaert, B. Dejaegher, Y. Vander Heyden, Chromatographic separation techniques and data handling methods for herbal fingerprints: A review, *Anal. Chim. Acta*, 690 (2011) 148–161.
- [7] Data preprocessing w: S.D. Brown, R. Tauler, B. Walczak, *Comprehensive Chemometrics*, t2, Elsevier, London 2009.
- [8] D. Skoog, D. West, J. Holler, S. Crouch, *Podstawy chemii analitycznej*, PWN, Warszawa, 2010.
- [9] 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, 1998.
- [10] P. Eilers, A perfect smoother, *Anal. Chem.*, 75 (2003) 3631-3636.
- [11] N. Nielsen, J. Carstensen, J. Smedsgaard, Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimised warping, *J. Chromatogr. A*, 805 (1998) 17-35.
- [12] M. Daszykowski, B. Walczak, Target selection for alignment of chromatographic signals obtained using monochannel detectors, *J. Chromatogr. A*, 1176 (2007) 1–11.
- [13] M. Daszykowski, B. Walczak, D.L. Massart, Projection methods in chemistry, *Chemom. Intell. Lab. Syst.*, 65 (2003) 97–112.
- [14] S. Wold, Pattern recognition by means of disjoint principal components models, *Pattern Recogn.*, 8 (1976) 127-139.
- [15] R. Fisher, The use of multiple measurements in taxonomic problems, *Annals of Eugenics*, 7 (1936) 179-188.
- [16] L. Breiman, J.H. Friedman, R.A. Olshen, C.G. Stone, *Classification and regression trees*, Wadsworth International Group, Belmont, 1984.

- 
- [17] M. Barker, W. Rayens, Partial least squares for discrimination, *J. Chemometr.*, 17 (2003) 166-73.
- [18] T. Næs, T. Isaksson, T. Fearn, T. Davies, *Multivariate calibration and classification*. NIR, Chichester, 2002.
- [19] H. Martens, T. Næs, *Multivariate calibration*, Wiley, Chichester, 1989.
- [20] M. Daszykowski, B. Walczak, D.L. Massart, Representative subset selection, *Anal. Chim. Acta*, 468 (2002) 91–103.
- [21] E.R. DeLong, D.M. DeLong, D.L. Clarke-Pearson, Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach, *Biometrics*, 44 (1988) 837-845.
- [22] R. Wehrens, H. Putter, L.M.C. Buydens, The bootstrap: a tutorial, *Chemom. Intell. Lab. Syst.*, 54 (2000) 35–52.
- [23] Q-S. Xu, Y.-Z. Liang, Monte Carlo cross validation, *Chemom. Intell. Lab. Syst.*, 56 (2001) 1-11.
- [24] J.A. Westerhuis, H.C.J. Hoefsloot, S. Smit, D.J. Vis, A.K. Smilde, E.J.J. van Velzen, J.P.M. van Duijnhoven, F.A. van Dorsten, Assessment of PLS-DA cross validation, *Metabolomics*, 4 (2008) 81–89.
- [25] C.M. Andersen, R. Bro, Variable selection in regression—a tutorial, *J. Chemometr.*, 24 (2010) 728–737.
- [26] O.M. Kvalheim, T.V. Karstang, Interpretation of latent-variable regression models, *Chemom. Intell. Lab. Syst.*, 7 (1989) 39–51.
- [27] T. Rajalahti, R. Arneberg, F. Berven, K.M. Myhr, R.J. Ulvik, O.M. Kvalheim, Biomarker discovery in mass spectral profiles by means of selectivity ratio plot, *Chemom. Intell. Lab. Syst.*, 95 (2009) 35–48.
- [28] V. Centner, D.L. Massart, O.E. Noord, S. de Jong, B.M. Vandeginste, C. Sterna, Elimination of uninformative variables for multivariate calibration, *Anal. Chem.*, 68 (1996) 3851–3858.
- [29] J. Mazina, M. Vaheer, M. Kuhtinskaja, L. Poryvkina, M. Kaljurand, Fluorescence, electrophoretic and chromatographic fingerprints of herbal medicines and their comparative chemometric analysis, *Talanta*, 139 (2015) 233-246.
- [30] Y. Vander Heyden, Extracting information from chromatographic herbal fingerprints, *LC-GC Europe*, 21 (2008) 438–443.
- [31] M. Goodarzi, P.J. Russell, Y. Vander Heyden, Similarity analyses of chromatographic herbal fingerprints: A review, *Anal. Chim. Acta*, 804 (2013) 16–28.
- [32] J. Orzel, M. Daszykowski, M. Kazura, D. de Beer, E. Joubert, A.E. Schulze, T. Beelders, A. de Villiers, C.J. Malherbe, B. Walczak, Modeling of the total antioxidant capacity of

---

rooibos (*Aspalathus linearis*) tea infusions from chromatographic fingerprints and identification of potential antioxidant markers, *J. Chromatogr. A*, 1366 (2014) 101-109.

[33] D. Custers, M. Canfyn, P. Courselle, J.O. De Beer, S. Apers, E. Deconinck, Headspace–gas chromatographic fingerprints to discriminate and classify counterfeit medicines, *Talanta*, 123 (2014) 78–88.

[34] P.-Y. Sacré, E. Deconinck, M. Daszykowski, P. Courselle, R. Vancauwenberghe, P. Chiap, J. Crommen, J.O. De Beer, Impurity fingerprints for the identification of counterfeit medicines-A feasibility study, *Anal. Chim. Acta*, 701, 2 (2011) 224–231.

[35] 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.

[36] 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, *Eur. Food Res. Technol.*, 231 (2010) 733-743.

[37] W. Yang, M. Hu, S. Chen, Q. Wang, S. Zhu, J. Dai, X. Li, Identification of Adulterated Cocoa Powder Using Chromatographic Fingerprints of Polysaccharides Coupled with Principal Component Analysis, *Food Anal. Methods*, DOI 10.1007/s12161-015-0126-6

[38] T.S. Villani, W. Reichert, M.G. Ferruzzi, G.M. Pasinetti, J.E. Simon, Q. Wu, Chemical investigation of commercial grape seed derived products to assess quality and detect adulteration, *Food Chemistry*, 170 (2015) 271–280.

[39] R.L. White, P.D. Wentzell, M.A. Beasy, Taxonomy of *Amanita* mushrooms by pattern recognition of amino acid chromatographic data, *Anal. Chim. Acta*, 277 (1993) 333-346.

[40] N.P. Mncwangi, A.M. Viljoen, J. Zhao, I. Vermaak, W. Chen, I. Khan, What the devil is in your phytomedicine? Exploring species substitution in *Harpagophytum* through chemometric modeling of <sup>1</sup>H-NMR and UHPLC-MS datasets, *Phytochemistry*, 106 (2014) 104–115.

[41] L.S.M. Wiedemann, L.A. d'Avila, D.A. Azevedo, Adulteration detection of Brazilian gasoline samples by statistical analysis, *Fuel*, 84 (2005) 467–473.

[42] L.F. Brandao, J.W.B. Braga, P.A. Suarez, Determination of vegetable oils and fats adulterants in diesel oil by high performance liquid chromatography and multivariate methods, *J. Chromatogr. A*, 1225 (2012) 150–157.

[43] J. Orzeł, M. Daszykowski, I. Grabowski, G. Zaleszczyk, M. Sznajder, B. Walczak, Simultaneous determination of Solvent Yellow 124 and Solvent Red 19 in diesel oil using fluorescence spectroscopy and chemometrics, *Talanta*, 101 (2012) 78-84.

[44] B. Krakowska, I. Stanimirova, J. Orzeł, 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.

- 
- [45] Rozporządzenie Ministra Finansów z dnia 20 sierpnia 2010 r. w sprawie znakowania i barwienia wyrobów energetycznych (Dz. U. 2010, Nr 157, poz. 1054).
- [46] J. Orzel, M. Daszykowski, I. Grabowski, G. Zaleszczyk, M. Sznajder, Identifying the illegal removal from diesel oil of certain chemical markers that designate excise duty, *Fuel*, 117 (2014) 224-229.
- [47] F.M. Fernandez, D. Hostetler, K. Powell, H. Kaur, M.D. Green, D.C. Mildenhall, P.N. Newton, Poor quality drugs: grand challenges in high throughput detection, countrywide sampling, and forensics in developing countries, *Analyst*, 136 (2011) 3073–3082.
- [48] R. Kumar, Recent Applications of Analytical techniques for counterfeit drug analysis: A Review, *Int. J. Pharm. Tech. Res.*, 6 (2014) 646–665.
- [49] M.J. Anzanello, R.S. Ortiz, R. Limberger, K. Mariotti, A framework for selecting analytical techniques in profiling authentic and counterfeit Viagra and Cialis, *Forensic Sci. Int.*, 235 (2014) 1–7.
- [50] U. Holzgrabe, M. Malet-Martino, Analytical challenges in drug counterfeiting and falsification–The NMR approach, *J. Pharm. Biomed. Anal.*, 55 (2011) 679687.
- [51] K. Dégardin, Y. Roggo, F. Been, P. Margot, Detection and chemical profiling of medicine counterfeits by Raman spectroscopy and chemometrics, *Anal. Chim. Acta*, 705 (2011) 334– 341.
- [52] C. R. Jung, R.S. Ortiz, R. Limberger, P. Mayorga, A new methodology for detection of counterfeit Viagra and Cialis tablets by image processing and statistical analysis, *Forensic Sci. Int.*, 216 (2012) 92–96.
- [53] T.N. Tran, N.L. Afanador, L.M.C. Buydens, L. Blanchet, Interpretation of variable importance in Partial Least Squares with Significance Multivariate Correlation (sMC), *Chemom. Intell. Lab. Syst.*, 138 (2014) 153-160.