

METODY CHEMOMETRYCZNE W IDENTYFIKACJI ŹRÓDEŁ POCHODZENIA AMFETAMINY

Waldemar S. Krawczyk

*Centralne Laboratorium Kryminalistyczne Komendy Głównej Policji, Warszawa
(praca obroniona na Wydziale Chemii Uniwersytetu Jagiellońskiego)*

W trakcie syntezy amfetaminy powstają produkty różnych reakcji ubocznych. Związki te, stanowiące zanieczyszczenia amfetaminy, są charakterystyczne zarówno dla metody syntezy jak i dla warunków prowadzenia reakcji czy sposobu wydzielania końcowego produktu. Przeprowadzone badania wskazują na duże różnice pomiędzy składem zanieczyszczeń amfetaminy otrzymywanej w różnych warunkach. Próbkami amfetaminy otrzymywane według tej samej receptury mają podobny skład zanieczyszczeń. Obraz chromatograficzny tych zanieczyszczeń, czyli *profil zanieczyszczeń*, pozwala określić metodę syntezy amfetaminy oraz, w oparciu o metody statystyczne, umożliwia klasyfikację próbki do grupy związanej z jej źródłem (nielegalnym laboratorium) czyli tak zwane *profilowanie* [1].

Profilowanie amfetaminy składa się z następujących etapów:

- 1 analiza chromatograficzna syntezowych zanieczyszczeń amfetaminy – gromadzenie danych (profili zanieczyszczeń),
- 2 statystyczna analiza wyników: obliczanie współczynników korelacji, analiza składowych głównych, analiza skupień,
- 3 przypisanie profili do poszczególnych grup związanych z nielegalnymi laboratoriami.

Zanieczyszczenia amfetaminy po ich odpowiednim zateżeniu, poddawano analizie metodą chromatografii gazowej, otrzymując w ten sposób tzw. *profil zanieczyszczeń*. Spośród kilkudziesięciu pików stanowiących wspomniany profil, wybrano piętnaście najbardziej charakterystycznych dla syntezy amfetaminy. Związki chemiczne odpowiadające tym pikom są związane zarówno z metodą jak i warunkami otrzy-

mywania oraz wydzielania końcowego produktu.

Pola powierzchni 15 pików wprowadzono do bazy danych, a następnie poddawano kilkustopniowej obróbce statystycznej [2]. W pierwszym kroku obliczono współczynniki korelacji liniowej pomiędzy poszczególnymi profilami. Wysokie wartości tych współczynników (>0.95) wskazują, że próbki amfetaminy najprawdopodobniej pochodzą z tego samego źródła i są przypisywane do tej samej grupy podobieństwa. Do grupy tej przypisywane są również profile o współczynnikach korelacji z innymi profilami zawartymi w przedziale 0.8 do 0.95, jeśli wizualne porównanie chromatogramów wskazuje na ich duże podobieństwo. W oparciu o powyższe kryteria, 1000 zbadanych próbek amfetaminy podzielono na 37 grup. W celu zweryfikowania poprawności klasyfikacji amfetaminy do każdej z tych grup (różnych źródeł próbek amfetaminy) zastosowano analizę składowych głównych (Principal Component Analysis – PCA) oraz analizę skupień (Cluster Analysis – CA).

Klasyfikacja profili zanieczyszczeń i weryfikacja wyników przebiega według następującego schematu:

- klasyfikacja do poszczególnych grup na podstawie współczynnika korelacji pomiędzy chromatogramami (profilami) i ich wizualnego porównania,
- analiza składowych głównych i analiza skupień w oparciu o składowe objaśniające 90% wariancji,
- weryfikacja przypisania chromatogramu do danej grupy na podstawie *analizy skupień*,

–graficzna analiza wyników analizy składowych głównych w trójwymiarowym układzie współrzędnych.

1 ANALIZA SKŁADOWYCH GŁÓWNYCH

Celem przeprowadzonej analizy głównych składowych było zbadanie struktury bazy danych, oraz określenie liczby składowych niezbędnych do jej opisu. Analizie składowych głównych poddano zarówno bazę 1000 profili, potraktowaną jako jedna macierz danych, jak i mniejsze macierze obejmujące profile związane z poszczególnymi grupami podobieństwa, a wytypowane w oparciu o współczynnik korelacji. W badaniach przyjęto, że liczba obiektów (chromatogramów) musi przewyższać liczbę zmiennych (powierzchnia pików) dlatego obliczenia przeprowadzono tylko dla tych grup, które zawierały powyżej 15 chromatogramów (11 grup).

Wyniki analizy składowych głównych (PCA) znajdują się w tabeli 1. W tabeli tej podano liczbę wartości własnych większych od 1, policzonych dla poszczególnych grup podobieństwa wstępnie wytypowanych w oparciu o współczynniki korelacji, oraz procent wariacji opisywanej przez odpowiadającą im liczbę składowych. W tabeli umieszczono również liczbę składowych głównych wyjaśniających określony procent wariacji. Tabela zawiera także wyniki obliczeń przeprowadzonych dla całej bazy 1000 profili.

Tabela 1. Wyniki analizy składowych głównych.

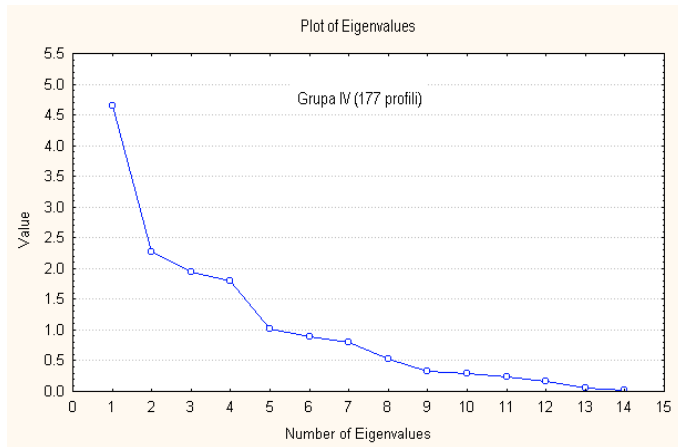
| Grupa | L.P. | W.w.>1 | % war. | Licz. głównych składowych wyjaśniających wariację: | | |
|-----------|------|--------|--------|--|------|------|
| | | | | >90% | >95% | >98% |
| I | 141 | 5 | 84.8 | 7 | 8 | 10 |
| IV | 177 | 5 | 77.8 | 8 | 10 | 11 |
| VII | 28 | 3 | 81.0 | 3 | 5 | 7 |
| VIII | 24 | 4 | 78.1 | 7 | 9 | 11 |
| X | 37 | 5 | 81.1 | 7 | 9 | 10 |
| XVI | 69 | 5 | 74.2 | 8 | 10 | 12 |
| XIX | 25 | 3 | 82.7 | 5 | 7 | 9 |
| XXII | 46 | 4 | 81.2 | 6 | 8 | 9 |
| XXIV | 42 | 5 | 88.9 | 6 | 7 | 8 |
| XXVIII | 16 | 4 | 90.0 | 4 | 5 | 7 |
| XXXIII | 30 | 2 | 86.2 | 3 | 5 | 6 |
| Cała baza | 1000 | 6 | 65.9 | 11 | 12 | 13 |

Poszczególne kolumny mają następujące znaczenie: *Grupa* - numer grupy podobieństwa wytypowanej na podstawie współczynników korelacji, *L.P.* - liczba chromatogramów (profili)

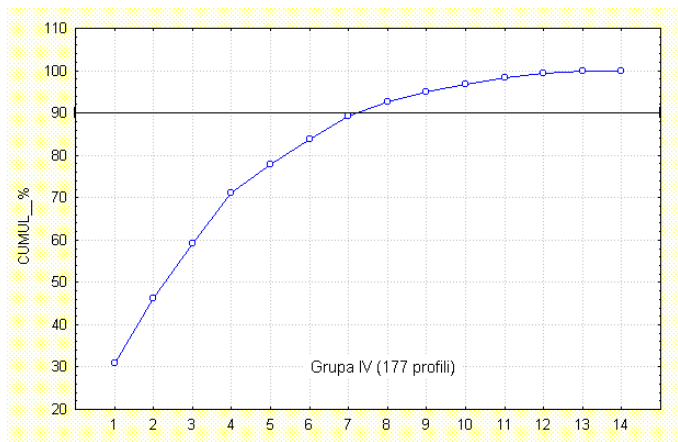
w poszczególnych grupach, *W.w.>1* - liczba wartości własnych większych od 1, *% war.* - procent wariacji opisywanej przez liczbę składowych odpowiadającą wartościom własnym większym od 1, *Liczba głównych składowych wyjaśniających wariację:* >90%, >95%, >98% - liczba składowych opisujących założony procent wariacji.

W przypadku PCA przeprowadzonej wewnątrz wcześniej wytypowanych grup podobieństwa, mamy od 2 do 5 wartości własnych większych od 1. Odpowiadająca im liczba składowych opisuje od 74.2% do 90.0% całkowitej wariacji. Dla całej bazy 1000 profili, 6 składowych opisuje 65.9% wariacji. Wyniki te wskazują, że w poszczególnych grupach podobieństwa, jest znacznie bardziej uporządkowana struktura danych niż w całej bazie (mniej wartości własnych większych od 1, większy % wariacji wyjaśnianej przez składowe główne), ale jednocześnie struktura ta jest różna dla różnych grup.

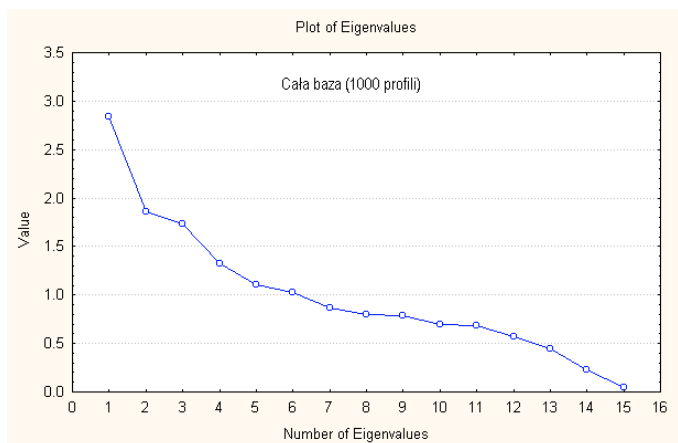
Istotnym problemem jest stwierdzenie, jaki procent wariacji powinien być wyjaśniany przez składowe główne, a w konsekwencji odpowiedź na pytanie, jakie kryterium przyjąć dla określenia wystarczającej liczby tych składowych. Przyjęcie liczby składowych wynikających z wartości własnych większych od jedynki wydaje się nie wystarczające, gdyż składowe te opisują bardzo różny procent wariacji (od 65.9% do 90.0%) w zależności od analizowanej bazy danych. Wydaje się słuszniejsze, oparcie się na określonym poziomie wyjaśnienia wariacji tzn. 90%, 95%, lub 98%. Pomocą w rozstrzygnięciu, którą z tych wartości przyjąć jako wystarczającą, jest analiza wykresów wartości własnych w zależności od ich numeru. Rysunki 1 i 2 pokazują zależność wartości własnych od ich numeru, a rysunki 1a - 2a ilustrują zależność skumulowanej wariacji od liczby składowych głównych, dla grupy IV i całej bazy danych.



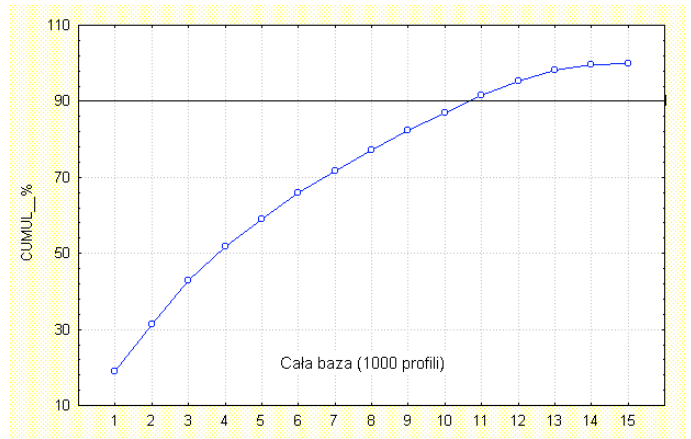
Rys. 1. Wykres wartości własnych dla grupy IV zawierającej 177 profili amfetaminy



Rys. 1a. Zależność skumulowanej wariancji od liczby składowych głównych (Grupa IV). Ciągła linia oznacza granicę 90% objaśnianej wariancji

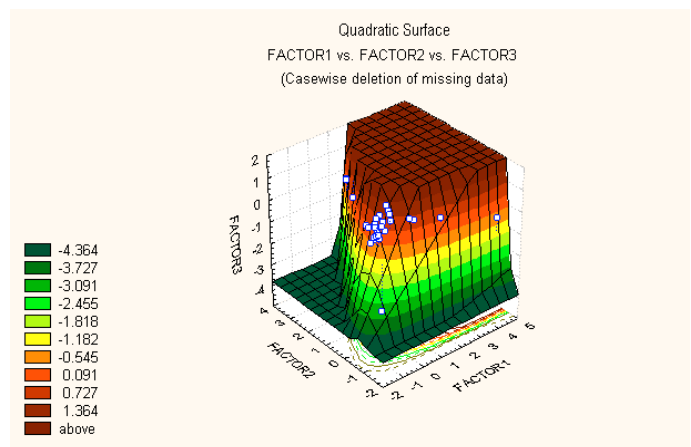


Rys. 2. Wykres wartości własnych dla całej bazy zawierającej 1000 profili amfetaminy



Rys. 2a. Zależność skumulowanej wariancji od liczby składowych głównych (Cała baza). Ciągła linia oznacza granicę 90% objaśnianej wariancji

Jedną z istotnych korzyści zastosowania PCA jest możliwość przynajmniej częściowej wizualizacji danych (obiektów) na drodze redukcji wymiaru przestrzeni czynnikowej do 2 lub 3 wymiarów. Obraz danych rzutowanych na pierwsze dwie lub trzy osie główne zachowuje szereg cech prawdziwej struktury danych i pozwala na wyciągnięcie wniosków wynikających z tej struktury, m. in. dotyczących zgrupowań obiektów oraz obiektów odbiegających. Rysunek 3 przedstawia w przestrzeni 3-wymiarowej, profile amfetaminy przypisanej do jednej z grup. Na osiach wykresu znajdują się współrzędne trzech składowych. Składowe te opisują ponad 90% wariancji.



Rys. 3. Profile amfetaminy przypisane do grupy XXXIII opisane trzema składowymi obliczonymi metodą PCA

Na rysunku tym widoczne są wyraźnie punkty odpowiadające próbkom amfetaminy, odbiegającym od jądra grupy. Przypisanie tych próbek do grupy XXXIII należy zweryfikować.

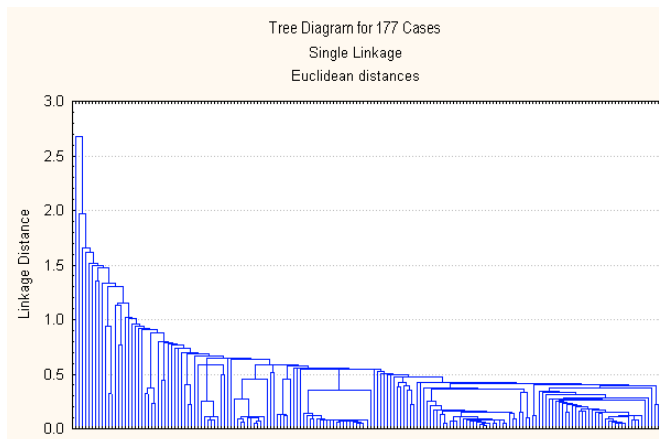
W przypadku bazy zawierającej wszystkie 1000 profili analiza składowych głównych

praktycznie nie pozwala zredukować liczby zmiennych i nie upraszcza opisu danych.

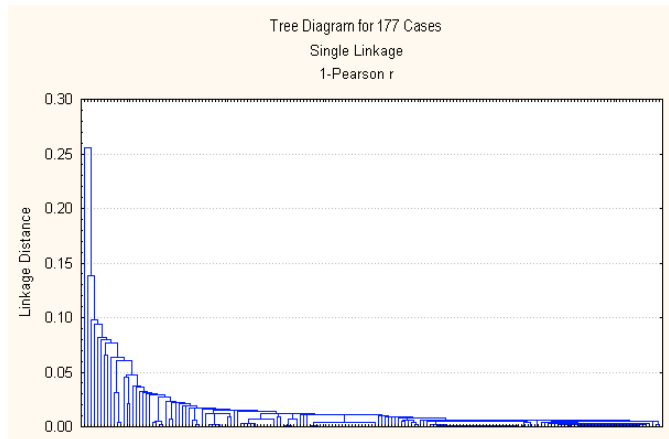
2 ANALIZA SKUPIEŃ (KLASTRÓW)

Skutecznej weryfikacji podziału profili na poszczególne grupy podobieństwa można dokonać za pomocą analizy skupień. Podstawą skutecznego wykorzystania tej metody jest jednak określenie najlepszego sposobu definiowania odległości pomiędzy profilami oraz zdefiniowanie wielkości tego dystansu, dla obiektów, które można uznać za podobne, należące do jednej grupy (jednego klastra) oraz odległości, wskazującej, że mamy do czynienia z różnymi grupami (klastrami).

Rys. 4 i 5 ilustrują wynik analizy skupień w oparciu o różne miary odległości, dla 177 profili (15 współrzędnych dla każdego profilu) wstępnie zakwalifikowanych do grupy IV. Na rysunkach widoczne są podgrupy, w ramach których obserwuje się bardzo niewielkie odległości pomiędzy profilami (odległość Euklidesowa poniżej 0.3, kwadrat odległości Euklidesowej poniżej 0.1, odległość Manhattan poniżej 0.5, odległość Czebyszewa poniżej 0.1, odległość oparta na współczynniku korelacji poniżej 0.01). Podgrupy te prawdopodobnie odpowiadają próbkom amfetaminy pochodzącym z tego samego rzutu syntezy (współczynnik korelacji pomiędzy nimi jest większy od 0.99).



Rys. 4. Analiza skupień 177 profili amfetaminy przypisanych wstępnie do grupy IV. Odległość Euklidesowa

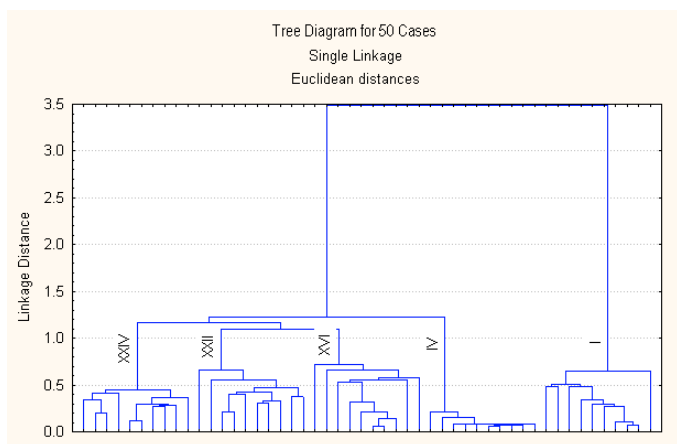


Rys. 5. Analiza skupień 177 profili amfetaminy przypisanych wstępnie do grupy IV. Odległość oparta na współczynniku korelacji

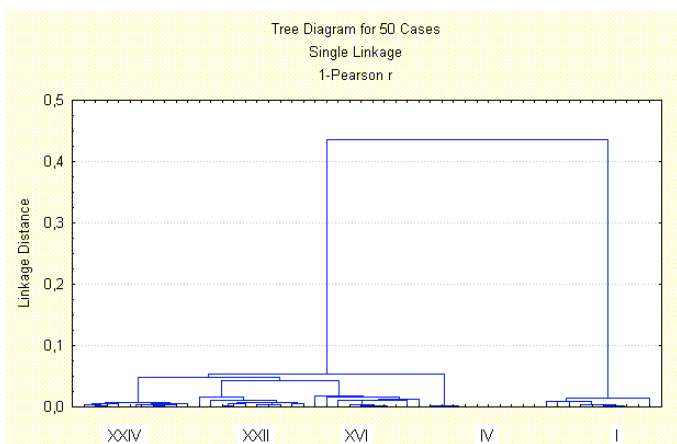
Głównym celem analizy skupień jest potwierdzenie, podziału na grupy podobieństwa dokonanego w oparciu o współczynnik korelacji. Jest to szczególnie istotne dla nowych profili, dla których współczynnik korelacji z innymi profilami, wcześniej przypisanymi do danej grupy, jest zawarty pomiędzy 0.8 a 0.95. Szczegółowa analiza poszczególnych diagramów wskazuje, że jako profile podobne można uznać, te dla których odległość euklidesowa jest mniejsza od 1, o odległość mierzona w oparciu o współczynnik korelacji jest mniejsza od 0.04.

Obserwacje te potwierdzają się w modelowym przykładzie opartym na 50 profilach amfetaminy pochodzących z pięciu najliczniejszych grup: I, IV, XVI, XXII i XXIV (tabela 1). Wzorcową bazę danych zbudowano w ten sposób, że z każdej grupy wzięto 10 profili amfetaminy. Warunkiem, dla wybranych profili był współczynnik korelacji pomiędzy wybranymi chromatogramami, większy od 0.95. Powstała w ten sposób macierz zawierająca 50 obiektów, z których każdy był opisany 15 współrzędnymi (powierzchniami 15 pików po standaryzacji). Rysunki 6 i 7 przedstawiają wyniki analizy skupień.

Oba rysunki jednoznacznie wskazują na pięć grup podobieństwa (pięć skupień). Odległość Euklidesowa pomiędzy skupieniami jest większa od 1, a odległość oparta o współczynnik korelacji jest większa od 0.04.



Rys. 6. Analiza skupień 50 profili amfetaminy pochodzących z 5 grup. Odległość Euklidesowa



Rys. 7. Analiza skupień 50 profili amfetaminy pochodzących z 5 grup. Odległość oparta na współczynniku korelacji

W oparciu o przeprowadzone obliczenia "podobieństwa" profili zanieczyszczeń wewnątrz grup oraz obliczenia modelowe stwierdzono, że profile mogą być przypisane do jednej grupy podobieństwa, jeżeli odległość Euklidesowa pomiędzy nimi jest mniejsza niż 1, a odległość oparta na współczynniku korelacji jest mniejsza od 0.04. Profile zakwalifikowane do jednej grupy podobieństwa, odpowiadają próbkom amfetaminy pochodzącym z jednego źródła (nielegalnego laboratorium).

BIBLIOGRAFIA

- 1) Krawczyk W., Profilowanie narkotyków, Wydawnictwo Centralnego Laboratorium Kryminalistycznego KGP, Warszawa 1998.
- 2) Pakiet *STATISTICA*, Wersja 5.0, StatSoft.