

Maria Jadamus-Hacura

Akademia Ekonomiczna w Katowicach

Andrzej Hacura

Uniwersytet Śląski

WYKORZYSTANIE METOD WIELOWYMIAROWEJ ANALIZY DANYCH SPEKTRALNYCH DO IDENTYFIKACJI I KLASYFIKACJI POLSKICH MIODÓW

1. Wstęp

Wprowadzenie w roku 2006 rozporządzenia Unii Europejskiej [Rozporządzenie (WE) nr 1924/2006...] w sprawie oświadczeń żywieniowych i zdrowotnych dotyczących żywności spowodowało wzrost zainteresowania badaniami autentyczności produktów spożywczych. Do najczęściej fałszowanych produktów należą np. oleje roślinne, wina, miody [Kelly, Downey, Fouratie 2004], a także sery, herbaty i inne. Współczesne metody badawcze oparte na analitycznych badaniach spektroskopowych są stosowane coraz częściej [Bertelli, Plessi, Sabatini 2007; Ruoff, Luginbuhl 2006]. Celem artykułu jest przedstawienie metody klasyfikacji i badania autentyczności produktu na przykładzie miodu. Dla konsumenta miodu jego autentyczność oznacza oryginalność i zgodność z deklarowanym pochodzeniem zarówno botanicznym, jak i geograficznym.

W etapie pierwszym prezentowanej metody otrzymano dane spektralne absorpcji promieniowania podczerwonego dla szeregu dokładnie określonych pod względem botanicznego i regionalnego pochodzenia próbek miodu. Pomiary takie są metodą badań bardzo szybką, dokładną i powtarzalną, wymagającą małej ilości badanej próbki, która w trakcie badań nie ulega zniszczeniu.

W drugim etapie badania przeprowadzono analizę danych spektralnych metodą głównych składowych i metodą dyskryminacji, w wyniku której otrzymano funkcje klasyfikacyjne. Funkcje te mogą być wykorzystane do klasyfikacji nowych przypadków.

2. Materiał badawczy

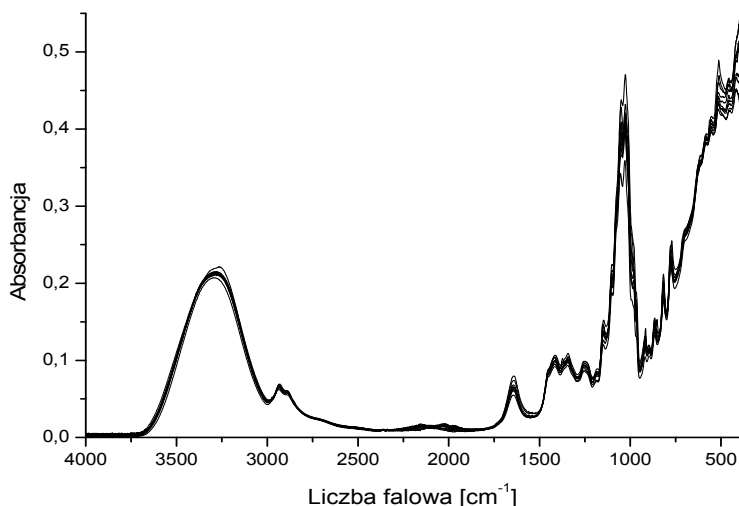
Badaniom absorpcji promieniowania podczerwonego poddano 50 próbek najczęściej spotykanych na rynku miódów polskich. Miody te wyprodukowano w różnych dużych pasiekach prywatnych i spółdzielczych zakładach pszczelarskich. Duża część z nich (23) została wytworzona przez pszczoły z nektaru różnych kwiatów i spadzi leśnej i została zaklasyfikowana jako miody wielokwiatowe. Pozostałe próbki miódów określono jako miody odmianowe¹ z nazwą rośliny, której udział nektaru był dominujący, jak np. miody akacjowe, lipowe, rzepakowe, gryczane, wrzosowe, malinowe, a cztery pozostałe próbki to miody spadziowe leśne. Miody te różnią się kolorem, smakiem, zapachem nie tylko z powodu botanicznego pochodzenia, ale także w zależności od pory roku zbiorów, regionu kraju, sposobów hodowli pszczół i pozyskiwania miodu. Ponadto różnią się one pewnymi mierzalnymi fizykochemicznymi własnościami, w tym również kształtem poszczególnych linii spektralnych, które otrzymujemy w wyniku analizy spektralnej – w naszym wypadku absorpcji promieniowania podczerwonego związanej ze strukturą molekularną próbek.

3. Badania doświadczalne

Pomiary osłabionego (przez absorpcję) całkowitego, odbitego promieniowania podczerwonego w zakresie liczb falowych od 4000 do 400 cm^{-1} zostały przeprowadzone za pomocą spektrometru fourierowskiego Bio-Rad FTS 6000. Spektrometr wyposażony był w przystawkę jednodbiciową ATR (*Attenuated Total Reflection*) z kryształem diamentu, co zapewniało dobrej jakości widma dla każdej z próbek w ilości mniejszej niż jedna kropla. Widma wszystkich próbek miódów, jak widać na rys. 1, są bardzo do siebie podobne. W praktyce do analiz wielowymiarowych danych została wykorzystana najbardziej charakterystyczna i zróżnicowana pod względem subtelnych zmian część widma z tzw. obszaru *finger print* od 1300 do 900 cm^{-1} . Widma były otrzymywane z rozdzielczością 1 cm^{-1} , co zapewniało bardzo dokładne zarejestrowanie tych małych różnic w kształcie linii widmowych badanych próbek. Widma próbek miodu przy tej wysokiej klasy aparaturze są powtarzalne, niemniej w celu wygładzenia szumów wywołanych dużym wzmocnieniem sygnału jednodbiciowego stosowano wielokrotne (32 razy) przemiatanie widma i uśrednianie (akumulację). W takich przypadkach czas pomiaru był nadal stosunkowo krótki i wynosił około 1 minuty.

Do zalet tej metody pomiarowej należą: krótki czas pomiaru, duża dokładność i powtarzalność wyników, mała ilość próbki potrzebnej do badań w sposób nieniszczący preparatu, choć w praktyce, ze względu na wygodę, za każdym razem w kolejnej serii pomiarowej badano pojedynczą, świeżą kroplę z większej ilości dostarczonej próbki. Wadą przedstawionej metody pomiarowej jest wysoka cena wykorzystywanej przez

¹ Zbadano 23 próbki miódów odmianowych.



Rys. 1. Typowe widma absorpcyjne podczerwieni 10 próbek miodów
 Źródło: pomiary własne.

nas aparatury, która jest bardzo dokładna i dość wszechstronna, a do takich rutynowych pomiarów można by użyć mniej rozbudowanych spektrometrów podczerwieni.

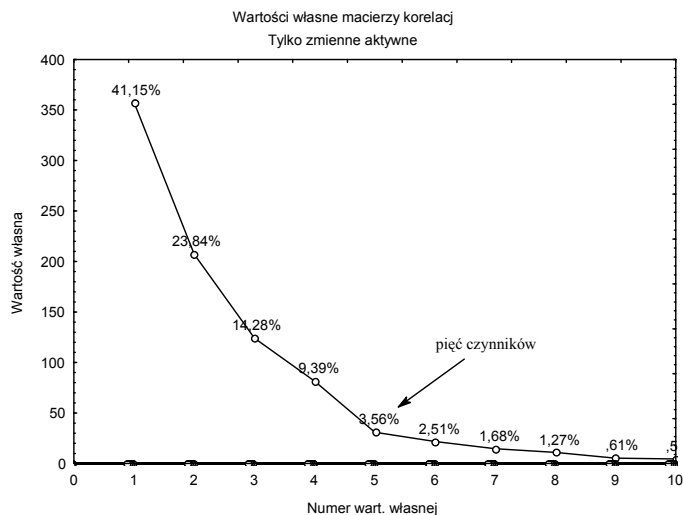
4. Wielowymiarowa analiza danych

W badaniu każdy z 50 miodów jest opisany przez 118 zmiennych, których wartości są rzędnymi punktów leżącymi na wykresie wybranej części widma danego miodu (wybrano co dziesiąty punkt). Zmienne te są silnie skorelowane. Analizę czynnikową wykorzystano w celu redukcji przestrzeni danych i przekształcenia skorelowanych zmiennych wyjściowych w nieskorelowane zmienne wyjaśniające przynajmniej w 90% zmienność danych pierwotnych. Do obliczeń wykorzystano komputerowy pakiet STATISTICA. Na podstawie otrzymanych wyników (zob. tab. 1 i rys. 2) do dalszej analizy wybrano 5 składowych głównych, które wyjaśniają 92,22% całkowitej zmienności oryginalnych danych.

Tabela 1. Wartości własne dla analizowanych zmiennych

Numer wartości	Wartość własna	% ogółu wariancji	Skumulowana wartość własna	Skumulowany %
1	357,5519	41,14521	357,5519	41,1452
2	207,2066	23,84426	564,7585	64,9895
3	124,1319	14,28445	688,8904	79,2739
4	81,6048	9,39065	770,4952	88,6646
5	30,9438	3,56085	801,4390	92,2254

Źródło: obliczenia własne.



Rys. 2. Wykres osypiska

Źródło: obliczenia własne.

Klasyfikację miodów przeprowadzono w nowej przestrzeni zdefiniowanej przez wybrane czynniki, wykorzystując w tym celu analizę dyskryminacyjną. Rozpoczęto ją od zbadania, na ile dobrze wybrane zmienne dyskryminujące (czyli pięć pierwszych czynników głównych) rozróżnią jedną grupę od drugiej. Jak wiadomo, standardową statystyką stosowaną do oceny statystycznej istotności mocy dyskryminacyjnej wybranego zbioru zmiennych jest statystyka lambda Wilksa². Jej wartość wynosi w przypadku naszych badań 0,14659, co świadczy o istotności statystycznej mocy dyskryminacyjnej wybranego modelu. Statystyki podsumowujące ten etap analizy przedstawiono w tab. 2.

Tabela 2. Statystyki zmiennych wprowadzonych do modelu

Wyszczególnienie	Lambda Wilksa	Cząstkowy Wilksa	F usunięcia	Poziom p	Tolerancja	Tolerancja
Czynnik 1	0,424264	0,345517	40,72555	0,000000	0,640419	0,359581
Czynnik 2	0,168991	0,867445	3,28544	0,047011	0,921853	0,078147
Czynnik 3	0,256052	0,572504	16,05430	0,000006	0,786749	0,213251
Czynnik 4	0,206957	0,708314	8,85377	0,000602	0,822620	0,177380
Czynnik 5	0,251055	0,583898	15,32147	0,000009	0,731409	0,268591

Źródło: obliczenia własne.

Analiza wartości cząstkowych statystyk lambda Wilksa pozwala na stwierdzenie, że wszystkie czynniki mają istotny wkład ($p < 0,05$) do dyskryminacji grup; największy jest wkład czynnika 1.

² Im wartość statystyki lambda Wilksa jest bliższa zeru, tym większa jest moc dyskryminacyjna zmiennych.

Po wykonaniu analizy kanonicznej otrzymano współczynniki funkcji dyskryminacyjnych przedstawione w tab. 3.

Tabela 3. Współczynniki funkcji dyskryminacyjnej

Wyszczególnienie	Pierwiastek 1	Pierwiastek 2
Czynnik 1	1,133154	-0,302285
Czynnik 2	0,339305	-0,390066
Czynnik 3	-0,687260	-0,704085
Czynnik 4	0,582241	-0,508009
Czynnik 5	-0,801559	-0,452253
Wart. wł.	3,349927	0,568237
Udział	0,854974	1,000000

Źródło: obliczenia własne.

Pierwsza funkcja dyskryminacyjna odpowiada za 85,5% wyjaśnionej wariancji, natomiast druga wyjaśnia 14,5% całej mocy dyskryminacyjnej. Do zbadania statystycznej istotności obu funkcji wykorzystujemy test χ^2 kolejnych pierwiastków³. Wyniki obliczeń związanych z tym testem zawiera tab. 4.

Tabela 4. Wyniki testu χ^2 kolejnych pierwiastków

Wyszczególnienie	Wartość własna	Kanoniczne R	Lambda Wilksa	χ^2	df	Poziom p
0	3,349927	0,877560	0,146591	86,40501	10	0,000000
1	0,568237	0,601948	0,637659	20,24785	4	0,000446

Źródło: obliczenia własne.

Jak widać, w naszym badaniu obie funkcje dyskryminacyjne są statystycznie istotne.

Po sprawdzeniu istotności funkcji dyskryminacyjnych chcemy poznać naturę dyskryminacji dla każdej funkcji. W tym celu wyznaczono średnie kanoniczne, a ich wartości zamieszczono w tab. 5.

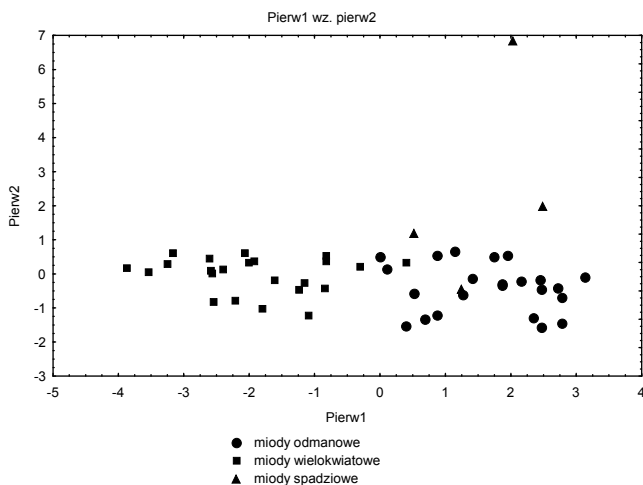
Tabela 5. Średnie zmiennych kanonicznych

Rodzaj miodu	Pierwiastek 1	Pierwiastek 2
Odmianowe	1,64938	-0,406907
Wielokwiatowe	-1,92252	-0,009187
Spadziowe	1,57052	2,392541

Źródło: obliczenia własne.

Wartości tych średnich pozwalają na stwierdzenie, że pierwsza funkcja dyskryminacyjna odróżnia głównie miody wielokwiatowe od pozostałych, natomiast druga wyróżnia miody spadziowe. Wykres rozrzutu wartości kanonicznych przedstawiony na rys. 3 potwierdza otrzymane wyniki.

³ W pierwszym wierszu tabeli zamieszczono wyniki związane z testem istotności dla obu pierwiastków, a w drugim – wartości związane z badaniem istotności tylko drugiego pierwiastka.



Rys. 3. Wykres rozrzutu wartości kanonicznych

Źródło: obliczenia własne.

Ostatnim etapem analizy dyskryminacyjnej jest wykorzystanie wyników analizy dyskryminacyjnej do klasyfikacji miodów. W tym celu dla każdej grupy wyznaczono funkcje klasyfikacyjne, które można również wykorzystać do celów predykcyjnych. Wiadomo, że dany przypadek (w naszym badaniu wybrany miód) klasyfikujemy do grupy, dla której wartość funkcji klasyfikacyjnej jest największa. Wartości obliczonych współczynników funkcji klasyfikujących dla trzech odmian miodów zawiera tab. 6.

Tabela 6. Współczynniki funkcji klasyfikujących

Wyszczególnienie	Wielokwiatowe	Odmianowe	Spadziowe
Czynnik 1	-2,90104	2,65607	1,40858
Czynnik 2	-0,65498	0,72528	-0,40422
Czynnik 3	1,52438	-0,97251	-3,17324
Czynnik 4	-1,17651	1,23176	-0,31770
Czynnik 5	1,69370	-1,24745	-2,56593
Stała	-2,62461	-2,21955	-6,62112

Źródło: obliczenia własne.

Po przeprowadzeniu klasyfikacji badanych 50 miodów otrzymano następujące wyniki (zob. tab. 7):

Tabela 7. Macierz klasyfikacji miodów

Wyszczególnienie	Procent	Odmianowe	Wielokwiatowe	Spadziowe
Odmianowe	100,00	23	0	0
Wielokwiatowe	95,65	1	22	0
Spadziowe	50,00	2	0	2
Razem	94,00	26	22	2

Źródło: obliczenia własne.

W pierwszej kolumnie tab. 7 widzimy procent przypadków, które zostały poprawnie zaklasyfikowane w każdej grupie przez aktualne funkcje klasyfikacyjne. Wyniki klasyfikacji nie są zadowalające tylko dla grupy miodów spadziowych, ale grupa ta stanowiła zaledwie 8% liczebności badanych miodów. Poza tym należy pamiętać, że klasyfikacja została przeprowadzona dla miodów, których badanie było podstawą wyznaczenia funkcji dyskryminacyjnych. Wykorzystanie funkcji klasyfikujących do nowych przypadków z pewnością zmniejszy liczbę trafnych klasyfikacji.

5. Podsumowanie

Przeprowadzona analiza ma charakter badania wstępnego. Przeprowadzono ją na stosunkowo małej próbie liczącej zaledwie 50 próbek miodów (w analizie składowych głównych jest to zdaniem wielu autorów wymagana minimalna liczebność próby), jednak wyniki są na tyle obiecujące, że zachęcają do dalszych analiz. W następnych etapach planuje się przeprowadzić klasyfikację miodów odmianowych i wykorzystać ją do celów predykcyjnych.

Przedstawiona metoda badania autentyczności miodu może być również wykorzystana do badania próbek innych produktów spożywczych pochodzenia biologicznego o dużym stopniu złożoności, takich jak: oleje roślinne, wina, herbaty i sery.

Literatura

- Bertelli D., Plessi M., Sabatini A. (2007), *Classification of Italian honeys by mid-infrared diffuse reflectance spectroscopy (DRIFTS)*, Food Chemistry, 101, 1565-1570.
- Kelly J.F.D., Downey G., Fouratier V. (2004), *Initial Study of honey adulteration by sugar solutions using midinfrared (MIR) spectroscopy and chemometrics*, "Journal of Agricultural and Food Chemistry", 52, 33-39.
- Metody statystycznej analizy wielowymiarowej w badaniach marketingowych* (2004), red. E. Gatnar, M. Walesiak, AE, Wrocław.
- Rozporządzenie (WE) nr 1924/2006 Parlamentu Europejskiego i Rady z dnia 20 grudnia 2006 r. w sprawie oświadczeń żywieniowych i zdrowotnych dotyczących żywności.
- Ruoff K., Luginbuhl W. (2006), *Authentication of the botanical and geographical origin of honey by mid-infrared spectroscopy*, "Journal of Agricultural and Food Chemistry", 54, 6873-6880.
- Stanisz A. (2007), *Przystępny kurs statystyki z zastosowaniem STATISTICA PL na przykładach medycyny*, t. 3, *Analizy wielowymiarowe*, Stat Soft Polska Sp. z o.o., Kraków.
- Walesiak M. (1996), *Metody analizy danych marketingowych*, PWN, Warszawa.

**THE APPLICATION OF MULTIVARIATE ANALYSIS
OF SPECTRA DATA IN THE IDENTIFICATION
AND CLASSIFICATION OF POLISH HONEYS**

Summary

Infrared Fourier transform spectroscopy with attenuated total reflection (ATR) sampling and multivariate analysis were used for the identification and classification of Polish honey. This method is rapid, non-destructive, with high precision and reliable detection. The 50 honey samples of multifloral and unifloral from different floral sources were precisely measured in mid-infrared region and analyzed with the use of principal component (PCA) and discriminant analysis.