

Politechnika Gdańska
Wydział Chemiczny
Katedra Technologii Leków i Biochemii

Rozprawa doktorska

**Zastosowanie spektroskopii bliskiej podczerwieni i klasycznej
podczerwieni w analizie jakościowej i ilościowej surowców
w przemyśle farmaceutycznym i kosmetycznym**

mgr inż. Patrycja Wrosz

Promotor: prof. dr hab. inż. Jan Mazerski

Gdańsk 2012

WYKAZ SKRÓTÓW I AKRONIMÓW	5
WYKAZ SYMBOLI	6
I WPROWADZENIE	7
II CZĘŚĆ TEORETYCZNA	9
1. PODSTAWY SPEKTROSKOPII OSCYLACYJNEJ	9
1.1. Oscylator harmoniczny i anharmoniczny	9
1.2. Zalety i wady spektroskopii NIR	12
1.3. Porównanie spektroskopii NIR i IR	13
1.4. Metody pomiaru	15
2. METODY CHEMOMETRYCZNE STOSOWANE W SPEKTROSKOPII	16
2.1. Transformacja widm	17
2.2. Techniki chemometryczne stosowane w analizie jakościowej	20
2.2.1. Analiza podobieństwa	20
2.2.2. Klasyfikacja – rozpoznawanie obrazów z nauczycielem	22
2.3. Techniki chemometryczne stosowane w analizie ilościowej	24
2.3.1. Kalibracja jednowymiarowa	25
2.3.2. Kalibracja wielowymiarowa	26
2.3.2.1. Wielokrotna regresja liniowa (MLR-Multiple Linear Regression)	27
2.3.2.2. Regresja głównych składowych	28
2.3.2.3. Częstkowa metoda najmniejszych kwadratów	29
3. PROCEDURA POSTĘPOWANIA PODCZAS OPRACOWYWANIA METOD JAKOŚCIOWYCH I ILOŚCIOWYCH	31
3.1. Metoda jakościowa	31
3.2. Metoda ilościowa	33
4. PUNKTY ODBIEGAJĄCE	37
4.1. Rodzaje punktów odbiegających	37
4.2. Wykrywanie punktów odbiegających typu x	38
5. ZASTOSOWANIE SPEKTROSKOPII NIR W PRZEMYSŁE	39
5.1. Przemysł farmaceutyczny	39
5.1.1. Identyfikacja substancji	40
5.1.2. Analiza ilościowa	41
5.1.3. Polimorfizm i izomery	41
5.1.4. Opakowania	42
5.1.5. Wilgotność	42
5.1.6. Twardość	42
5.1.7. Wielkość cząstek	43
5.2. Przemysł spożywczy	43
5.2.1. Mięso i ryby	43
5.2.2. Produkty mleczne	44
5.2.3. Owoce i warzywa	45
III CEL PRACY	47
IV METODY I MATERIAŁY	50
4.1. APARATURA	50
4.1.1. Szkło laboratoryjne	51
4.1.2. Oprogramowanie	51

4.1.3. Ustawienia sprzętowe.....	50
4.2. MATERIAŁY, ODCZYNNIKI.....	51
4.3. PROCEDURA POMIAROWA.....	52
4.3.1. Tworzenie metody identyfikacyjnej.....	52
4.3.1.1. Transformacja widm.....	52
4.3.1.2. Analiza głównych składowych.....	53
4.3.1.3. Klasyfikacja.....	53
4.3.1.4. Walidacja.....	53
4.4. TWORZENIE METOD ILOŚCIOWYCH.....	54
4.4.1. Wyznaczenie wartości referencyjnej badanego parametru.....	55
4.4.2. Skład badawczy serii kalibracyjnych i walidacyjnych.....	56
4.4.3. Wybór metody regresji.....	57
4.4.4. Transformacja widm.....	57
4.4.5. Optymalizacja i kompleksowość modeli.....	57
4.4.6. Ocena jakości skonstruowanych modeli kalibracyjnych.....	58
4.4.7. Walidacja modeli ilościowych.....	58
4.5. ANALIZA CHEMOMETRYCZNA DANYCH.....	61
V WYNIKI I DYSKUSJA.....	63
5.1. ANALIZA CHEMOMETRYCZNA.....	63
5.1.1. Ocena zmienności warunków pomiarowych.....	63
5.1.1.1. Ocena wpływu ilości skanów na jakość opracowanego modelu jakościowego.....	63
5.1.1.2. Ocena wpływu ilości głównych składowych na jakość końcowych wyników modeli tożsamościowych.....	74
5.1.1.3. Powtarzalność pomiarów.....	76
5.1.1.4. Ocena wpływu położenia próbki na okienku pomiarowym.....	78
5.1.1.5. Ocena wpływu środowiska na jakość opracowywanych modeli.....	80
5.1.1.6. Wpływ wstępnego przygotowania danych na konstrukcję czynników głównych.....	83
5.1.1.7. Wpływ punktów odbiegających na konstrukcję czynników głównych.....	88
5.2. ANALIZA JAKOŚCIOWA.....	93
5.2.1. Procedura opracowywania metod jakościowych.....	93
5.2.1.1. Wybór zakresu analizy.....	93
5.2.1.2. Podział widm.....	95
5.2.1.3. Wstępna obróbka widm.....	95
5.2.1.4. Analiza głównych składowych.....	96
5.2.1.5. Klasyfikacja SIMCA.....	98
5.2.1.6. Walidacja wewnętrzna oraz zewnętrzna.....	98
5.2.2. Wpływ transformacji danych oraz liczby istotnych czynników na zdolności predykcyjne opracowanych modeli.....	99
5.2.3. Określenie przydatności poszczególnych technik transformacji do minimalizacji zmienności pochodzącej z różnych źródeł.....	112
5.2.3.1. Dobór metod technik chemometrycznych stosowanych podczas transformacji surowych widm.....	112
5.2.3.2. Ocena wpływu normalizacji na zdolności predykcyjne modeli opracowanych dla surowców stałych oraz surowców płynnych.....	115
5.2.4. Wpływ doboru widm do zbioru uczącego na zdolność predykcyjną modelu jakościowego.....	124
5.2.5. Porównanie zastosowania spektroskopii NIR oraz IR w analizie tożsamościowej surowców kosmetycznych i farmaceutycznych.....	127
5.2.5.1. Podział widm na trzy zbiory.....	128
5.2.5.2. Wstępna obróbka danych.....	128
5.2.5.3. Klasyfikacja widm zbioru uczącego.....	129
5.2.5.4. Walidacja zewnętrzna oraz wewnętrzna.....	130
5.3. ANALIZA ILOŚCIOWA.....	132
5.3.1. Oznaczanie w mg/g zawartości substancji aktywnej – ketoprofen w żelu Ketoprofen.....	132
5.3.1.1. Analiza widm, wybór zakresu analizy.....	132

5.3.1.2. Zastosowanie algorytmu PCR do oznaczania zawartości ketoprofenu.....	133
5.3.1.3. Zastosowanie algorytmu PLS1 do oznaczania zawartości ketoprofenu.....	137
5.3.1.4. Zastosowanie algorytmu PLS2 do oznaczania zawartości w mg/g ketoprofenu	141
5.3.2. Oznaczanie zawartości filtrów UV w emulsjach do opalania	143
5.3.2.1. Dobór zakresu analizy, analiza widmowa	143
5.3.2.2. Transformacja surowych widm.....	144
5.3.2.3. Optymalizacja i kompleksowość modeli.....	144
5.3.3. Oznaczanie zawartości anionowych związków powierzchniowo czynnych w składnikach środków myjących	148
VI PODSUMOWANIE	151
VII LITERATURA	156
VIII STRESZCZENIE.....	165
IX ABSTRACT	167
X DOROBEK NAUKOWY.....	168

WYKAZ SKRÓTÓW I AKRONIMÓW

Skrót	Termin polskojęzyczny	Termin anglojęzyczny
AZPC	anionowe związki powierzchniowo czynne	-
ATR	całkowite wewnętrzne odbicie	Attenuated Total Reflection
BIAS	średni wartość resztkowych oznaczanego składnika	Mean residual value
IR	spektroskopia podczerwieni	Infrared Spectroscopy
KNN	klasyfikator k najbliższych sąsiadów	K-Nearest Discriminant Analysis
LDA	liniowa analiza dyskryminacyjna	Linear Discriminant Analysis
LV	wskaźnik dźwigni	Leverage Ratio
MD	odległość Mahalanobiasa	Model Distance
MLR	liniowa regresja wieloraka	Multiple Linear Regression
MSC	normalizacja	Multiply Scatter Correction
NIR	spektroskopia bliskiej podczerwieni	Near Infrared Spectroscopy
PC	główna składowa	Principal Component
PCA	analiza głównych składowych	Principal Component Analysis
PCR	regresja czynników głównych	Principal Component Regression
PLS	metoda częściowych najmniejszych kwadratów	Partial Least Squares
PLS-1	metoda częściowych najmniejszych kwadratów dla jednej zmiennej	-
PLS-2	metoda częściowych najmniejszych kwadratów dla wielu zmiennych	-
RR	wskaźnik reszt	Residual Ratio
RSD	względne odchylenie standardowe	Residual Standard Deviation
RMSECV	średni błąd kwadratowy krosvalidacji	Root Mean Square Error of Cross Validation
RMSEP	średni błąd kwadratowy predykcji	Root Mean Square Error of Prediction
SEC	standardowy błąd kalibracji	Standard Error of Calibration
SEP	standardowy błąd przewidywania	Standard Error of Prediction
SIMCA	klasyfikator SIMCA	Soft Independent Modeling of Class Analogies
SNV	transformacja SNV	Standard Normal Variate
ZPC	związki powierzchniowo czynne	-

WYKAZ SYMBOLI

Symbol

a	współczynnik kierunkowy prostej
b	nachylenie prostej
\mathbf{b}	wektor współczynników regresji
c	prędkość światła
C	stężenie molowe roztworu
\mathbf{e}	wektor reszt modelu
E	macierz błędów
F	wartość F-Snedecora
\mathbf{G}	macierz reszt
k	współczynnik siły
m	liczba zmiennych
n	liczba próbek
P	macierz wag
R	reflektancja
R	współczynnik korelacji
S	odchylenie standardowe
T	macierz wyników
T_r	transmitancja
T^*	transfleksja
U	macierz wyników
W	macierz wag PLS
X	macierz danych
y	zmienna zależna
Y	macierz zmiennych zależnych
Q	macierz wag
Q_w	współczynnik walidacji

I WPROWADZENIE

Dla wszystkich laboratoriów kluczowe zagadnienia to czas analizy i koszt analizy. Jest to szczególnie widoczne w przemysłowych laboratoriach kontroli jakości. Ich pracownicy stoją pod presją z jednej strony jak najszybszego uzyskania wyniku a z drugiej strony dokładności i wiarygodności wykonywanych analiz. Długotrwałe oczekiwania na wynik, od którego często zależy kolejny etap produkcji przynosi firmom straty finansowe. Jednak z drugiej strony błędny wynik może oznaczać jeszcze większe straty finansowe, a także zagrożenie dla zdrowia odbiorcy oraz złą reputację. Dlatego też poszukuje się coraz to nowszych metod analitycznych w celu skrócenia czasu analizy a także ograniczenia kosztów rutynowych analiz przy zachowaniu poprawności wyniku. Odpowiedzią na takie potrzeby jest spektroskopia w zakresie klasycznej podczerwieni (IR -*infrared spectroscopy*) i bliskiej podczerwieni (NIR- *near infrared spectroscopy*) [1].

Historia spektroskopii w podczerwieni sięga początku XIX wieku. W 1800 roku angielski muzyk i astronom Frederick William Herschel pracował nad pozornie prostym astronomicznym zagadnieniem. Starał się dowiedzieć, który kolor pochodzący od widzialnego białego światła słonecznego dostarcza ciepło. W swoim eksperymencie zastosował szklany pryzmat oraz termometr owinięty w czarny papier. Taki prymitywny sprzęt posłużył mu do pomiaru zmian temperatury wywoływanej przez światło o danym kolorze. Herschel w kwietniu 1800r zaobserwował, że termometr wskazuje najwyższą temperaturę, gdy jest on ustawiony tuż za czerwonym krańcem światła słonecznego. To odkrycie Herschel opisał w pracy „Experiments on the refrangibility of the invisible rays of the Sun”[2-4]. Zakres ten później nazwano promieniowaniem podczerwonym.

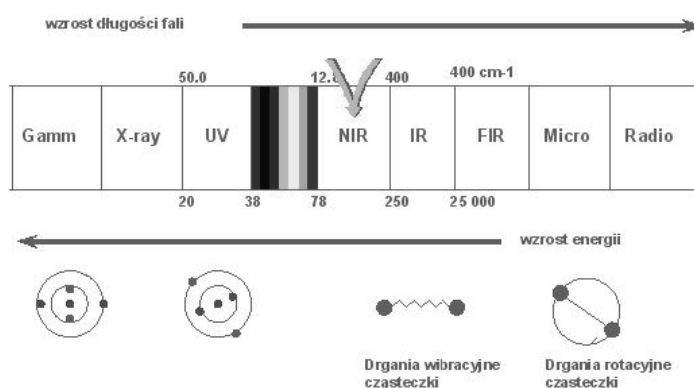
Podczas gdy spektroskopia IR uzyskała ogólną akceptację i zastosowanie głównie w analizie tożsamościowej, spektroskopia NIR była zaniedbywana przez spektroskopów. Zaniebanie to wynikało z faktu, że naukowcy nie mogli znaleźć w tym zakresie użytecznych informacji, wynikających z szerokich, nałożonych i słabych pasm absorpcji. Przełomem okazał się rok 1881 kiedy to Abney i Festing zarejestrowali widma organicznych związków w zakresie 1200-700 nm. W swojej pracy wykazali obecność organicznych grup funkcyjnych, szczególnie wodoru, w widmie NIR. Pod wpływem tych wyników 1900 r. Coblentz przeprowadził eksperyment i opisał właściwości wiązania C-H. Prawdziwy rozwój spektroskopii bliskiej podczerwieni nastąpił po drugiej wojnie światowej. W 1960 r Karl Noris z Amerykańskiego Departamentu Rolnictwa, wykazał

możliwość stosowania spektroskopii NIR do oznaczania zawartości wody w produktach rolnych[5].



Rys.1 Aparatura, za której pomocą, Herschel badał promieniowanie podczerwone Słońca[6]

Analiza NIR wykorzystuje promieniowanie w zakresie $10000\text{-}4000\text{ cm}^{-1}$ dostarczając informacji na temat właściwości fizykochemicznych i stanu fizycznego badanej próbki jak i jej składu chemicznego. Pasma absorpcji w spektroskopii podczerwieni można podzielić na: pasma podstawowe, pasma nadtonów oraz pasma kombinacyjne. Pasma podstawowe powstają, gdy układ oscylujący przechodzi ze stanu podstawowego na pierwszy poziom wzbudzony, takie pasma obserwuje się w klasycznej spektroskopii IR. Pasma nadtonów oraz pasma kombinacyjne widoczne są w spektroskopii NIR. Nadtony powstają na wskutek przejścia układu drgającego ze stanu podstawowego na 2, 3 lub wyższy wzbudzony poziom energetyczny. Z drganiami kombinacyjnymi natomiast mamy do czynienia wówczas, gdy dwa lub kilka układów drgających oddziałuje ze sobą dając równocześnie wspólny pik na widmie absorpcyjnym. Zatem częstotliwość tych oddziaływań jest sumą częstotliwości poszczególnych oddziaływań.[7]



Rys.2 Zakres widma w podczerwieni [8]

II CZĘŚĆ TEORETYCZNA

1. Podstawy spektroskopii oscylacyjnej

Atomy lub grupy atomów oddziałujące ze sobą drgają względem siebie z częstotliwością, która zależy od siły ich wzajemnych oddziaływań. Źródłem tej energii są fotony, których energia (E_p) dana jest wzorem (1)

$$E_p = h\nu = \frac{hc}{\lambda} \quad (1)$$

gdzie:

- λ → długość fali
- h → stała Plancka
- c → prędkość światła

Promieniowanie elektromagnetyczne ma nieciągłą kwantową naturę. Składa się ono z określonych porcji energii zwanych kwantami promieniowania lub fotonami. Wielkość energii niesionej przez pojedynczy kwant związana jest z długością fali padającego promieniowania. Sposób oddziaływania promieniowania elektromagnetycznego z materią zależy od energii jego kwantów. Energia niesiona przez promieniowanie z przedziału NIR mieści się w zakresie mogącym spowodować, iż grupa atomów znajdzie się na wyższym poziomie oscylacyjnym niż stan podstawowy.

1.1. Oscylator harmoniczny i anharmoniczny

Promieniowanie elektromagnetyczne absorbowane przez cząsteczki powoduje nie tylko wzbudzenie elektronów ale także zmiany energii oscylacyjnej i rotacyjnej cząsteczki. Oscylacja jest ruchem periodycznym, w czasie którego atomy na przemian oddalają i przybliżają się do siebie. Modelem makroskopowym drgającej cząsteczki dwuatomowej mogą być dwie kulki o masach m_1 oraz m_2 , połączone sprężyną. W stanie równowagi siła F jaką sprężyna działa na masy jest równa 0. Zgodnie z prawem Hooke'a siła F wychylająca atomy ze stanu równowagi jest proporcjonalna do wychyleń atomów. Opisany układ mechaniczny nosi nazwę oscylatora harmonicznego [9-10]

Absorpcja promieniowania elektromagnetycznego z zakresu podczerwieni jest możliwa tylko wtedy, gdy drgający układ jest dipolem. Częstotliwość drgających dipoli zależy od siły wiązania i masy związanych ze sobą atomów. Kiedy harmoniczny oscylator drga, energia oscylacyjna stale ulega przemianie z energii kinetycznej na potencjalną.

Całkowita energia wiązań jest proporcjonalna do częstotliwości tych drgań. Częstotliwość drgań oscylatora harmonicznego jest przedstawiona następującym wzorem:

$$\nu = \frac{h}{2\pi} \sqrt{\frac{k}{\mu}} \quad (2)$$

gdzie :

h → stała Plancka

k → stała siłowa

μ → zredukowana masa związanych atomów

$$\mu = \frac{m_1 m_2}{m_1 + m_2} \quad (3)$$

m_1 → masa jednego atomu

m_2 → masa drugiego atomu

Częstotliwość drgań układu jest tym większa im większa jest stała siłowa (gdy wiązanie pomiędzy dwoma atomami jest silniejsze). Z drugiej strony częstotliwość drgań maleje wraz ze wzrostem oscylujących mas.

Przy pomocy modelu oscylatora harmonicznego można względnie poprawnie przewidzieć energię drgań czy też częstotliwości absorpcji odpowiadające oscylacjom zginającym, rozciągającym, nożycowym, wachlarzowym i to zarówno symetrycznym jak i asymetrycznym [2,11-12]. Jednakże w przypadku nadtonów obserwowanych w regionie NIR model ten nie jest w stanie przewidzieć poprawnie pozycji pasm absorpcyjnych drgających cząsteczek.

Teoria mechaniki kwantowej dla oscylatora harmonicznego zakłada, że energia oscylatora jest kwantowana i energia poszczególnych poziomów energetycznych wyraża się następującym wzorem:

$$E_v = h\nu \left(n + \frac{1}{2} \right) \quad (4)$$

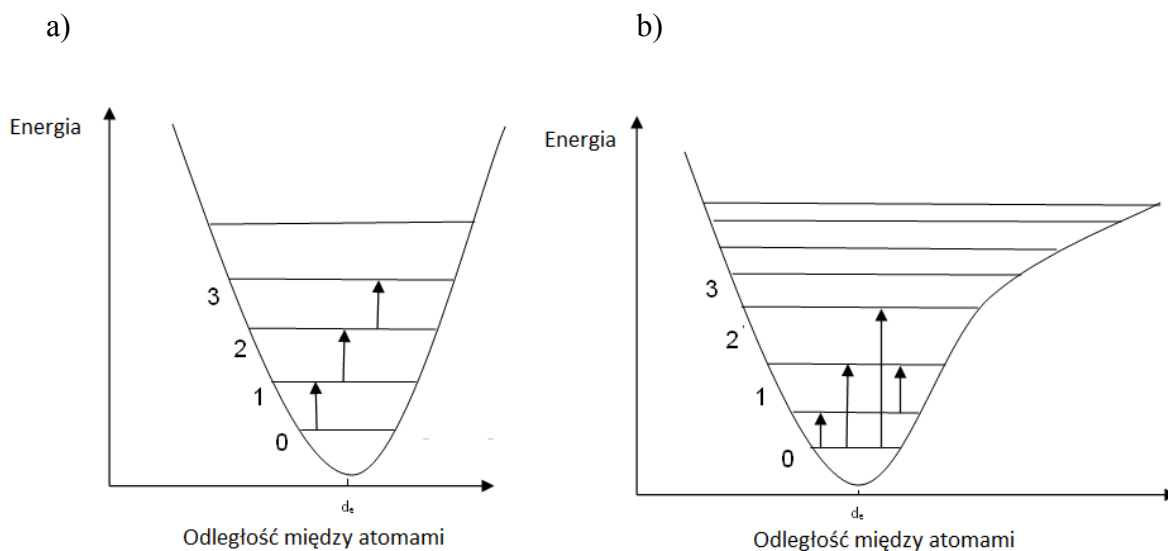
gdzie:

n → oscylacyjna liczba kwantowa

ν → częstotliwość drgań na poziomie zerowym

Zgodnie z tym opisem różnice pomiędzy sąsiednimi poziomami energetycznymi są stałe, niezależne od wyjściowego poziomu energetycznego.

Poziomy oscylacyjne są kwantowane i oscylator może zmieniać swoje stany tylko skokowo. Zgodnie z „regułą wyboru” możliwe są tylko przejścia między sąsiednimi poziomami. Zatem możliwe są tylko przejścia, dla których $\Delta n = \pm 1$, **Rys.3**. Z tej reguły wynika, że w widmie cząsteczki nie powinny występować nadtony lub mieć bardzo małą intensywność jako pasma „wzbronione”.



Rys.3 Studnia potencjału dla oscylatora harmonicznego (a) i anharmonicznego (b)

Oscylator harmoniczny nie opisuje dokładnie oscylacji cząsteczek. W rzeczywistości siła przeciwdziałająca odkształceniu wiązania nie jest proporcjonalna do wartości tego odkształcenia. Zależność energii potencjalnej wiązania od odchylenia od stanu podstawowego nie jest parabolą, lecz krzywą bardziej skomplikowaną. Oscylację cząsteczek dużo lepiej opisuje oscylator anharmoniczny. Krzywa energii potencjalnej takiego oscylatora nie ma kształtu paraboli, a aproksymuje się ją tzw. krzywą Morse'a.

Rys.3b. Energia potencjalna modelu anharmonicznego wyrażona jest poniższym wzorem:

$$E_{osc.anhar.} = hv \left(n + \frac{1}{2} \right) - hvx \left(n + \frac{1}{2} \right)^2 \quad (5)$$

gdzie :

$x \rightarrow$ współczynnik anharmoniczności opisujący stopień odstępstwa oscylatora od modelu harmonicznego

Posługując się równaniem (5) można łatwo wykazać, że odległości dowolnych sąsiadujących ze sobą poziomów energetycznych są zmienne, maleją one wraz ze wzrostem liczby kwantowej tym szybciej, im większą wartość ma współczynnik anharmoniczności. Ponadto w modelu anharmonicznym dozwolone są przejścia $\Delta n = \pm 2, \pm 3$ [12]

Aby promieniowanie elektromagnetyczne mogło oddziaływać z oscylującymi cząsteczkami musi być spełniony jeszcze jeden warunek. Oscylacjom musi towarzyszyć zmiana momentu dipolowego cząsteczki drgającej. Zatem aby wystąpiła znaczna absorpcja promieniowania wiązanie musi wykazywać częściowy charakter jonowy (wiązanie kowalencyjne spolaryzowane). Częstość absorbowanej fali zależy od częstości drgań molekuly, natomiast intensywność absorpcji zależy od efektywności przekazu energii fotonu molekule. Wodór jako najlżejszym z atomów oscyluje z największą amplitudą oraz wykazuje największe odchylenia od modelu harmonicznego. Dlatego silnie absorbującymi grupami w spektroskopii NIR są grupy: C-H, N-H, O-H, S-H i R-H [13].

1.2. Zalety i wady spektroskopii NIR

Spektroskopia bliskiej podczerwieni ze względu na szybkie i nieniszczące pomiary ilościowe oraz jakościowe stanowi interesującą alternatywę dla tradycyjnych metod analitycznych. Zaletami wykorzystywania spektroskopii NIR są: 1) nieniszczący, nieinwazyjny charakter analizy, 2) brak konieczności przygotowania próbki przed analizą, 3) prostota i szybkość wykonania rutynowej analizy, 4) wysoka czułość, 5) niski koszt, 6) uzyskanie kompleksowego obrazu badanego materiału, 7) zwiększenie bezpieczeństwa pracy 8) wpływ na ochronę środowiska ze względu na wyeliminowanie zużycia rozpuszczalników do przygotowywania próbek przed analizą. Oprócz ww. zalet technika ta posiada również pewną wadę: surowe widma NIR trudne są w interpretacji i wymagają znajomości zaawansowanych technik chemometrycznych. Trudność ta wynika przede wszystkim z faktu że [14-15]:

- a) użyteczna informacja niesiona przez widma NIR nie jest bezpośrednio dostępna. Użyteczna informacja rozproszona jest w praktycznie całym zakresie częstości.
- b) próbki opisane są przez dziesiątki a nawet tysiące zmiennych (ogromna wielowymiarowość danych)
- c) obserwuje się niską powtarzalność widm

Surowe widma otrzymane bezpośrednio ze spektrofotometru obarczone są szumem fotometrycznym. Szumy charakterystyczne dla pomiarów NIR mogą pochodzić od wielu czynników takich jak: interferencja różnych czynników fizycznych i/lub chemicznych, niedoskonałości aparatury wynikającej ze starzeniem się sprzętu oraz wielu losowych czynników. Zbyt duży szum powoduje zaburzenia stosunku szumu do sygnału, rozdzielczości widm oraz dokładności i precyzji końcowych modeli. Inne niepożądane czynniki niekorzystnie wpływające na widma NIR często dotyczą złożonego tła pomiarowego oraz linii bazowych, które pomimo, że nie są związane ze składem, stężeniem badanej substancji posiadają duży wpływ na dane widmowe. Dotyczy to wielkości ziaren substancji badanych w wyniku czego następuje przesunięcie oraz zakrzywienie linii bazowej. Na intensywność oraz kształt widma ma również wpływ wiele czynników nieswoistych takich jak temperatura, wilgotność, grubość i własności optyczne próbki np. zdolność do rozpraszania światła, wiek próbki, a nawet zawartość CO₂ w powietrzu otaczającym próbkę. Tego typu widmowe wariacje nie tylko mogłyby doprowadzić do niepotrzebnie skomplikowanych modeli kalibracyjnych, ale także do zafałszowania wyników końcowych. Dlatego przed zastosowaniem właściwej analizy chemometrycznej widma NIR poddaje się obróbce wstępnej. Polega ona na transformacji każdego z widm w celu ograniczenia wpływu zmian kształtu linii bazowej na kształt widma oraz ograniczeniu wpływu rozpraszania na intensywność widma

1.3. Porównanie spektroskopii NIR i IR

Spektroskopia podczerwieni oraz bliskiej podczerwieni wykorzystuje charakterystykę chemiczną badanej substancji. Każdy typ wiązania chemicznego absorbuje w charakterystycznym zakresie w widmie promieniowania elektromagnetycznego. Spektroskopia IR oraz NIR różnią się przede wszystkim zakresem widma oraz sposobem przygotowania próbki. Analiza NIR wykorzystuje promieniowanie w zakresie 10 000-4 000 cm⁻¹ dostarczając informacji na temat właściwości fizyko-chemicznych próbki oraz o stanie fizycznym badanej próbki. W spektroskopii IR wykorzystuje się natomiast promieniowanie w zakresie 4000-650cm⁻¹. Widmo IR daje nam więcej informacji na temat budowy chemicznej badanej substancji (charakterystyczne grupy funkcyjne absorbują w odpowiednim miejscu w widmie), dlatego też spektroskopia IR wykorzystywana jest głównie w analizie tożsamościowej. Pasma absorpcji IR pochodzą od drgań fundamentalnych rozciągających i deformacyjnych wiązań C-H, O-H, N-H (4 000-2 500 cm⁻¹), C≡N (2500-2000 cm⁻¹), C=O (1650-1750 cm⁻¹), C=N, C=C, N=O (1650-1550cm⁻¹),

C-O, C-N, C-C ($1550-650\text{ cm}^{-1}$) [16-18]. W zakresie NIR widoczne są natomiast pierwsze i drugie nadtony oraz pasma kombinacyjne wiązań C-H, N-H, O-H. Wszystkie pasma absorpcji NIR są szerokie. Dodatkowo pasma te wykazują mniejszą intensywność niż pasma podstawowe widoczne w obszarze średniej podczerwieni. Dlatego też analiza widm NIR klasycznymi metodami jest utrudniona, a nawet niemożliwa. Poniżej przedstawiono zakres pasm absorpcji, charakterystycznych dla związków widocznych w widmie NIR [19-20]:

Węglowodory alifatyczne - C-H Pierwszy nadton drgań rozciągających grupy C-H obserwuje się pomiędzy $5\ 555$ a $5\ 852\text{ cm}^{-1}$, drugi nadton pomiędzy $8\ 264-8\ 696\text{ cm}^{-1}$, trzeci nadton w zakresie od $11\ 364-10\ 929\text{ cm}^{-1}$. Najbardziej interesujący zakres drgań kombinacyjnych widoczny jest w zakresie od $6\ 666-7\ 690\text{ cm}^{-1}$ oraz $4\ 540-4\ 500\text{ cm}^{-1}$.

Związki aromatyczne – Główne pasma absorpcji związków aromatycznych widoczne w widmie NIR to: $8\ 772\text{ cm}^{-1}$, $5\ 988\text{ cm}^{-1}$, $4\ 651-4\ 587\text{ cm}^{-1}$ oraz $6\ 065\text{ cm}^{-1}$. Przy $5\ 988\text{ cm}^{-1}$ obserwuje się pierwszy nadton drgań rozciągających wiązania C-H, przy $8\ 772\text{ cm}^{-1}$ obserwuje się drugi nadton oraz drgania kombinacyjne wiązań C-H,

Aminy i amidy - Pierwszy nadton drgania rozciągającego grupy N-H obserwuje się przy $6\ 897\text{ cm}^{-1}$ oraz przy $6\ 452\text{ cm}^{-1}$. W zakresie od $971-962$ obserwuje się pasma dla drugiego nadtonu N-H, a trzeci nadton występuje w zakresie $12\ 407-12\ 387\text{ cm}^{-1}$. Natomiast pasma widoczne przy $5\ 070\text{ cm}^{-1}$ pochodzą od drgań kombinacyjnych N-H rozciągających oraz zginających.

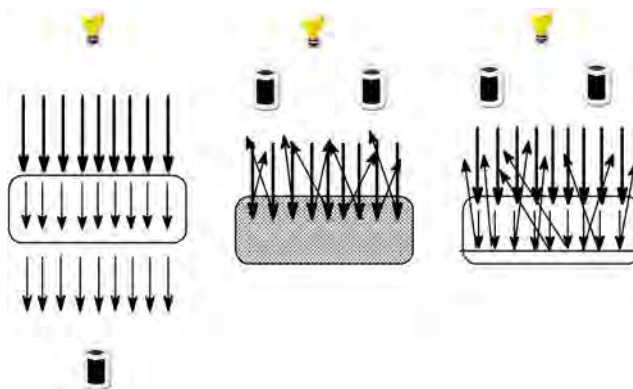
Alkohole – Alifatyczne alkohole posiadają charakterystyczne piki należące do pierwszego nadtonu przy $7\ 090\text{ cm}^{-1}$. Drugi nadton cząsteczki alkoholi alifatycznych O-H pojawia się w okolicy $10\ 400\text{ cm}^{-1}$ a trzeci w $13\ 500\text{ cm}^{-1}$.

Woda – Pasma absorpcji należące do cząsteczki wody obserwowane są w zakresie około $6\ 900\text{ cm}^{-1}$ oraz $10\ 300\text{ cm}^{-1}$. Pasma te należą do pierwszego i drugiego nadtonu drgań rozciągających O-H. Natomiast obserwowany silny pik w zakresie $5\ 150\text{ cm}^{-1}$ pochodzi od rozciągających i zginających drgań kombinacyjnych.

Ostatnio pojawiły się doniesienia o możliwości wykorzystania klasycznych widm IR również do analizy tożsamościowej i oceny jakości surowców. Podobnie jak w przypadku widm NIR wymaga to zastosowania chemometrycznej obróbki widm. Zastosowanie w spektroskopii IR oraz NIR metod chemometrycznych pozwala na pełniejsze wykorzystanie informacji zawartych w analizowanych widmach, analizę struktury wewnętrznej badanej próbki oraz analizę relacji występujących między próbkami.

1.4. Metody pomiaru

W zależności od rodzaju analizowanej próbki (jej właściwości optycznych i stanu skupienia) wykorzystuje się różne tryby rejestracji widm: tryb transmisyjny (technika transmisyjna), odbiciowy (pomiar reflektanci) i transflektancji [21-24].



Rys.4 Przedstawiający różne tryby pomiarowe: a) transmisyjny, b) odbiciowy, c) transflektancji

- **Tryb transmisyjny**

Transmitancja (Tr) jest miarą spadku natężenia promieniowania o danej długości fali w skutek przejścia promieniowania przez próbkę umieszczoną pomiędzy źródłem promieniowania a detektorem. W trybie transmisyjnym bada się najczęściej: ciecze, folie, tabletki, żele i zawiesiny.

- **Tryb odbiciowy**

Pomiar w odbiciu rozproszonym daje w wyniku refleksję (R) czyli stosunek natężenia światła odbitego od próbki (I) do natężenia światła odbitego od tła lub od wzorcowej powierzchni odbijającej (I_r). Promieniowanie NIR przenika znaczącą grubość próbki badanej, w której jest absorbowane przez związek analizowany, pobudzony do drgań. Promieniowanie niezaabsorbowane jest odbijane z powrotem w kierunku detektora. Innym sposobem pomiaru widm w trybie odbiciowym jest metoda całkowitego wewnętrznego odbicia (ATR), która wykorzystywana jest zarówno w spektrofotometrach IR oraz NIR. W metodzie ATR badana substancja przylega do tzw. kryształu ATR o dużej gęstości optycznej. Na powierzchni graniczną kieruje się promieniowanie podczerwone, następnie wiązka promieniowania podczerwonego odbija się od przylegającego materiału na zasadzie całkowitego wewnętrznego odbicia.

Technika odbiciowa wykorzystywana jest głównie do badania substancji stałych (proszków, tabletek), a także past i zawiesin.

- **Tryb transflektancji**

Transflektancja jest połączeniem techniki odbiciowej z techniką transmisyjną. Ten tryb pomiarowy stosowany jest do pomiarów substancji ciekłych oraz zawiesin. W pomiarze transflektancji (T^*) promieniowanie transmitowane przez próbkę jest odbite za pomocą zwierciadła lub powierzchni odbijającej dyfuzyjnie oraz zawrócone do próbki. W ten sposób uzyskuje się podwójną drogę optyczną.

Pomiary odbywają się na przystawce rozproszeniowej lub z wykorzystaniem sondy. Podczas zbierania widm substancji ciekłych na przystawce wykorzystuje się dodatkową powierzchnie odbiciową. W tym celu najczęściej stosuje się aluminiowy sześciokątny dyfuzyjny reflektor. Reflektor powoduje dodatkowe odbicie przechodzącego przez próbkę promieniowania podwajając drogę optyczną. Nie zaabsorbowane przez próbkę promieniowanie odbijane jest od reflektora i zawracane do próbki.

Podczas zbierania widm substancji badanych bardzo ważny jest proces zbierania widm. W szczególności ważne jest to w przypadku pomiarów substancji stałych, gdzie efekt rozpraszania, stopień upakowania substancji oraz położenie próbki na oknie pomiarowym mają istotny wpływ na jakość otrzymanych widm. W celu ułatwienia procesu pomiarów różnych substancji producenci oferują szereg akcesoriów. Bardzo często wykorzystywane są sondy optyczne. Mają one swoje zastosowanie w różnych etapach produkcyjnych: do potwierdzania tożsamości samych surowców w naważalnicach, w trakcie procesu produkcyjnego, a także przy kontroli produktów końcowych.

2. Metody chemometryczne stosowane w spektroskopii

Chemometria jest dziedziną nauki i techniki wykorzystującą metody statystyczne, matematyczne do wydobywania użytecznych informacji z wielowymiarowych danych. Metody chemometryczne wykorzystywane są nie tylko do analizy danych chemicznych, ale także medycznych, fizycznych i innych. Ta interdyscyplinarna dziedzina nauki posiada wiele zalet. Do najważniejszych należą [25-26]:

- wysoką jakość uzyskanych informacji z wielowymiarowych danych,
- niski koszt nakładów finansowych związany ze stosowaniem metod chemometrycznych,
- rozwój nowych szybkich technik analitycznych.

Surowe widma NIR uzyskane bezpośrednio ze spektrofotometru trudno jest interpretować. Na pierwszy rzut oka ciężko jest wyciągnąć pożądaną informację

zwłaszcza, że widma uzyskane tą techniką nie analizuje się w klasyczny sposób. Aby uzyskać jak najwięcej informacji zawartych w widmie stosuje się zaawansowane techniki chemometryczne. Spektroskopia bliskiej podczerwieni wraz z chemometrią są ze sobą ściśle powiązane. Technika NIR nie istniałaby bez chemometrii, a chemometria uzyskała wiele nowych narzędzi opracowanych początkowo dla potrzeb spektroskopii NIR, a obecnie znajdujących również zastosowanie w innych dziedzinach Techniki chemometryczne wykorzystywane są do: 1) usunięcia szumu, gdyż surowe dane są mieszaniną informacji i szumu, 2) wyciągnięcia jak najwięcej informacji zawartych w danych, 3) usunięcia zbędnych informacji [27-29].

Spektroskopia NIR w powiązaniu z wielowymiarowymi technikami chemometrycznymi wykorzystywana jest do analizy jakościowej i ilościowej. Przez analizę jakościową rozumie się identyfikację badanego materiału. Poddając analizie jakościowej w pierwszym etapie badana substancja jest identyfikowana a następnie klasyfikowana. Klasyfikacja widm jest techniką chemometryczną pozwalającą na identyfikację badanego materiału. Analiza ilościowa natomiast służy do wyznaczenia stężenia głównego składnika w mieszaninie wieloskładnikowej [16].

2.1. Transformacja widm

Proces transformacji widm jest nierozłącznym, wręcz najważniejszym etapem tworzenia modeli jakościowych i ilościowych. Wstępna transformacja widm ma trzy podstawowe cele: a) poprawę stosunku sygnału do szumu („wygładzanie” widm), b) ograniczenie wpływu zmienności i nieliniowości linii bazowej oraz c) wyeliminowanie wpływu rozpraszania światła przez próbkę na natężenie widma. Stosując różnego rodzaju techniki chemometryczne należy posługiwać się nimi rozsądnie, gdyż może dojść do sytuacji gdzie przez przypadek zostanie usunięta zbyt duża ilość właściwych informacji. Dlatego też znajomość tych technik jak i doświadczenie są bardzo cenne i wymagane.

Najczęściej stosowanymi technikami obróbki wstępnej są:

- Transformacja SNV (ang. Standard Normal Variate)
- Algorytm MSC (ang. Multiplicative Scatter Correction)
- Transformacja „offset correction”
- Transformacja „detrend correction”
- Metoda „De-trend”
- Algorytm Savitzky-Golay

Transformacja SNV

Transformacja SNV jest matematyczną transformacją widm wykorzystywaną w celu usunięcia efektów różnicy rozpraszania pomiędzy próbkami. Każde widmo jest centrowane a następnie skalowane przez odchylenie standardowe. Centrowanie polega na tym, że dla każdej długości fali wyznacza się wartość średnią z np. refraktancji otrzymując w ten sposób średnie widmo dla danego zestawu widm. Średnie widmo następnie jest odejmowane od każdego widma z zestawu. W kolejnym kroku widma te są dzielone przez wartość odchylenia standardowego każdego widma [30,31]

$$Z_{i_k} = \frac{(x_{i_k} - \bar{x}_k)}{S_k} \quad (6)$$

gdzie:

x_i → intensywność k-tego widma dla i-tej liczby falowej

\bar{x}_k → średnia wartość k-tego widma

S_k → odchylenie standardowe k-tego widma

Algorytm MSC

Algorytm MSC używany jest podobnie jak SNV do usunięcia efektów związanych z rozpraszaniem światła pomiędzy różnymi próbkami. Transformacji MSC dokonuje się według poniższego wzoru:

$$z_{ik}(new) = \frac{(x_{ik}(old) - a_i)}{b_k} \quad (7)$$

gdzie:

$x_{ik}(new), x_{ik}(old)$ → wartości reflektancji przed i po transformacji przy i-tej długości fali

a_k, b_k → współczynniki regresji dla k-tego surowego widma

Wartości a_k i b_k oszacowuje się stosując regresyjną metodą najmniejszych kwadratów polegającą na znalezieniu zależności, między k-tym widmem x_{ik} a średnim widmem z danego zbioru \bar{x}_k [32-33]

$$x_{ik} = a_i + b_i \bar{x}_i + e_{ik} \quad (8)$$

gdzie:

e_{ik} → przedstawia różnicę między k-tym widmem x_{ik} a średnim widmem z danego zbioru \bar{x}_i

Transformacja „offset correction”

Zadaniem transformacji „offset correction” jest korekcja równoległego przesunięcia linii bazowej. Metoda ta polega na odjęciu wartości średniej z kilku pierwszych punktów widma (np. 5) od każdego punktu widma [34]

$$z_{ik} = x_{ik} - \bar{x}_{1-5,k} \quad (9)$$

Transformacja „detrend correction”

Transformacja ta stosowana jest w celu usunięcia zakrzywienia i przesunięcia linii bazowej. Linia bazowa w widmach NIR wykazuje tendencję do przesunięcia w zakresie od 8000 do 4000cm⁻¹. W celu usunięcia tego efektu linia bazowa jest dopasowywana wielomianem drugiego stopnia, względem długości fali, a następnie odejmowana od widma [34]

$$x_{ik,d} = x_{ik} - b_{ik} \quad (10)$$

gdzie:

b_{ik} → jest to wartość linii bazowej k-tego widma dopasowywanej wielomianem drugiego stopnia przy i-tej długości fali

Metoda ta często stosowana jest wraz z transformacją SNV żeby uzyskać jeszcze lepsze efekty wygładzenia linii bazowej [35].

Algorytm Savitzky-Golay

Do korekcji linii bazowej stosuje się również pochodne widm. Pierwsza pochodna stosowana jest do usuwania przesunięć linii bazowej, a druga pochodna natomiast do korekcji przesunięcia oraz nachylenia linii podstawowej. Transformacja ta polega na tym, że metodą najmniejszych kwadratów aproksymuje się fragment widma wielomianem odpowiedniego stopnia tworząc tzw. okno wielomianu. Następnie oblicza się analitycznie wartość odpowiedniej pochodnej wielomianu aproksymującego w danym punkcie. Po tej operacji okno wielomianu przesuwa się o jeden punkt i proces jest powtarzany aż wartość ostatniego okna zostanie wyznaczona. Wartość pochodnej wyznacza się dla środkowego punktu okna. Stopień wielomianu oraz szerokość okna dobierana jest eksperymentalnie [36,37]. Posługując się algorytmem Savitzky-Golay należy być ostrożnym, aby nie wygładzić zbyt wiele widm tracąc przy tym pożądaną informację. Stosując coraz większe okno wielomianu uzyskuje się lepsze wygładzenie, lecz doprowadza to do utraty pożądanego informacji.

2.2. Techniki chemometryczne stosowane w analizie jakościowej

Wielowymiarowe techniki chemometryczne można podzielić na: metody z nadzorem „*Supervised methods*” zwane też jako metody uczenia z nauczycielem oraz bez nadzoru „*Non-supervised methods*”, czyli metody uczenia bez nauczyciela. Metody z nadzorem służą do rozwiązywania zagadnień klasyfikacji, a metody bez nadzoru do analizy podobieństwa. Do analizy jakościowej w spektroskopii NIR stosuje się zarówno metody klasyfikacji jak i analizy podobieństwa [38].

2.2.1. Analiza podobieństwa

Analiza bez nauczyciela nie wymaga znajomości wiedzy na temat struktury danych, gdyż wiedzę na ten temat uzyskuje się stosując odpowiednie metody chemometryczne. Do najczęściej stosowanych metod analizy danych bez nauczyciela należy analiza głównych składowych, PCA.

Analiza głównych składowych

W zestawie danych spektralnych występuje silna współliniowość zmiennych a tym samym duża redundancja informacji. Jednocześnie liczba zmiennych (długości fali lub liczb falowych) jest wielokrotnie większa niż liczba analizowanych obiektów (widm). Redundancja informacji i współliniowość zmiennych przejawia się wzajemną korelacją poszczególnych zmiennych. Przy analizie takich danych stosuje się najczęściej metodę zwaną analizą głównych składowych (PCA, ang. *Principal Component Analysis*) [39].

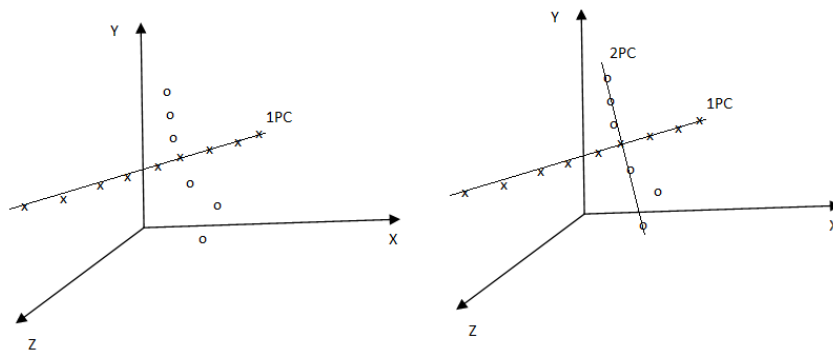
Metoda głównych składowych polega na przekształceniu zestawu m skorelowanych zmiennych objaśniających (zmiennych pomiarowych) w zestaw nowych, nieskorelowanych zmiennych zwanych głównymi składowymi, PC. Każda z głównych składowych jest liniową kombinacją wszystkich zmiennych objaśniających i jest ortogonalna w stosunku do wszystkich pozostałych głównych składowych. Główne składowe są wyznaczone w kolejności malejącego zasobu zmienności (informacji). Zatem pierwsza główna składowa wyjaśnia największą część zmienności danych i zawiera najwięcej informacji, **Rys.4**. Druga składowa jest ortogonalna (prostopadła) do pierwszej i reprezentuje największą zmienność wokół pierwszej głównej składowej. Trzecia główna składowa jest ortogonalna do pierwszej i drugiej i opisuje największą zmienność nie wyjaśnianą przez te składowe. Dla całkowitego odtworzenia macierzy danych pierwotnych potrzeba tyle głównych składowych ile jest zmiennych pierwotnych, równanie (11a). Jednakże ze względu na redundancję informacji i obecność w danych niepewności pomiarowej (szumu) możliwe jest znaczne ograniczenie liczby uwzględnianych głównych

składowych bez utraty istotnych informacji o analizowanych obiektach. W metodzie tej zakłada się, że zmienność właściwa wynikająca z charakteru obiektów jest dużo większa od zmienności niepożądanej. Z tej przyczyny uwzględnia się jedynie k-pierwszych głównych składowych, równanie (11b).

$$[X]_{nm} = [T]_{nm} [P]_{mm}^T \quad (11a)$$

$$[X]_{nm} = [T]_{nk} [P]_{km}^T + [E]_{nm} \quad (11b)$$

W analizie głównych składowych macierz danych X o n-objektach i m-zmiennych opisywana jest przy pomocy dwóch mniejszych macierzy (*wzór 11*), zwanych macierzą wyników T i macierzą wag P. Macierz E przedstawia różnice między wartościami macierzy danych a wartościami otrzymanymi przez przemnożenie T z P^T . Macierz T zawiera informację o relacjach pomiędzy próbkami, macierz P natomiast zawiera informację o korelacji między zmiennymi [39-42].



Rys.4 Konstrukcja głównych składowych

Kluczowym etapem w metodzie głównych składowych jest dobór odpowiedniej liczby istotnych składowych, k. W analizie chemometrycznej istnieje kilka kryteriów wyboru liczby istotnych składowych:

- Kryterium pogładowości
- Kryterium zasobu zmienności
- Kryterium spadku wartości własnej
- Kryterium osypiska
- Zredukowane własności własne
- Kryterium kompozytowe
- Kryterium IND

Przed przystąpieniem do analizy głównych składowych przeprowadza się pewien rodzaj transformacji zmiennych w celu eliminacji różnic między zmiennymi. Taki rodzaj transformacji nazywa się autoskalowaniem i przeprowadza się go według poniższego wzoru [39]

$$Z_{ji} = \frac{X_{ji} - \bar{X}_i}{S_i} \quad (12)$$

gdzie: x_{ji} → pierwotna wartość i-tej zmiennej dla j-tego obiektu, \bar{x}_i → średnia wartość i-tej zmiennej, s_i → odchylenie standardowe i-tej zmiennej

Analiza głównych składowych umożliwia rozrzut obiektów, próbek w n-wymiarowej przestrzeni na dwu lub trójwymiarową podprzestrzeń tzw. mapę liniową. Mapa liniowa dostarcza cennych informacji na temat relacji, podobieństw występujących pomiędzy badanymi klasami.

Metody analizy struktury danych bazujące na przestrzennych relacjach między obiektami

Metody te stosowane są alternatywnymi metodami dla PCA. Stosuje się je w celu szybkiego porównania próbki z widmem wzorca. Używane są ze względu na prostotę i łatwość interpretacji. Głównym narzędziem przy ocenie podobieństw różnic widm jest współczynnik korelacji oraz odległość euklidesowa [43].

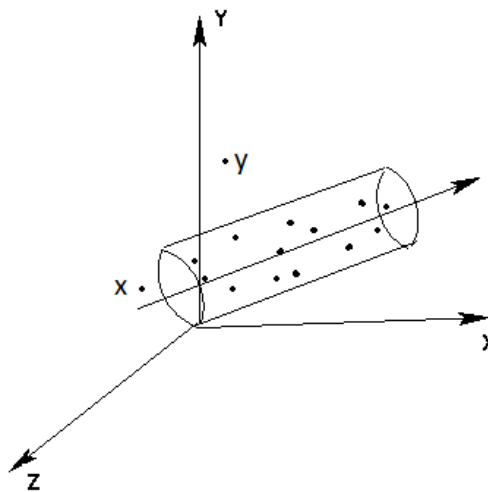
- współczynnik korelacji – definiowany jako $\cos\alpha$ kąta pomiędzy wektorem widma próbki a wektorem widma średniego klasy. Stosując współczynnik korelacji porównuje się podobieństwo kształtu widm
- odległość euklidesowa – metoda ta polega na jednoczesnym porównaniu natężenia i kształtu widm.

2.2.2. Klasyfikacja – rozpoznawanie obrazów z nauczycielem

Analiza z nadzorem wymaga znajomości przynależności badanych substancji do poszczególnych klas oraz liczbę klas. Celem tej metody jest opracowanie reguł umożliwiających przewidzenie dopasowania jakościowego badanej substancji w oparciu o znane wartości ilościowe zmiennych objaśniających. Jest kilka klasyfikatorów stosowanych do klasyfikacji obiektów np. LDA (ang. *Linear Discriminant analysis* [44-45]), KNN (ang. *K-nearest neighbours*) [46-47]. Najczęściej w spektroskopii stosuje się klasyfikator SIMCA.

SIMCA

SIMCA (ang. *Soft Independent Modeling of Class Analogies*) jest klasyfikatorem służącym do określenia przynależności widma badanej substancji do określonej populacji danych. W klasyfikatorze SIMCA każda klasa modelowana jest na podstawie podobieństwa obiektów w obrębie klasy. Model dla każdej klasy tworzony jest niezależnie w oparciu o liczbę istotnych głównych składowych w danej klasie. Po utworzeniu modelu tworzona jest koperta ufności wokół modelu, w której znaleźć się powinny z określonym prawdopodobieństwem wszystkie obiekty danej klasy. Liczbę istotnych głównych składowych dobiera się na podstawie kryterium zasobu zmienności wnoszonych przez główne składowe. Do modelu wykorzystywane są te główne składowe, które łącznie wyjaśnia ustalony wcześniej procent zmienności. Jeśli okaże się, że wcześniej ustalony procent zmienności wyjaśnia tylko jedna główna składowa to wtedy modelem klasy będzie wartość średnia tej składowej oraz wektor określający tę składową. W takim przypadku granica przynależności obiektów do modelu będzie wyznaczona cylindrycznie, **Rys.6**. W przypadku dwóch istotnych głównych składowych modelem będzie położenie środka klasy oraz płaszczyzna wyznaczona przez wektory tych dwóch składowych [40,48-50].



Rys.6 Wygląd modelu SIMCA zbudowanego w oparciu o jedną główną składową [13]

Każdy obiekt można opisać modelem danej klasy

$$X_A = \bar{X}_A + \sum_{j=1}^S T_A V_A^T (rxp) + E_A(nxp) \quad (13)$$

gdzie: $X_A \rightarrow$ średnia wartość macierzy centrowanej, $T_A(nxr) \rightarrow$ "macierz wyników" wartość macierzy uzyskanej dla n obiektów i r głównych składowych,

$V_A^T(rxp) \rightarrow$ "macierz wag" wartość macierzy uzyskanej dla r głównych składowych i p zmiennych

W celu sprawdzenia czy badany obiekt należy do modelu wyznacza się wartość statystyki F przy określonym stopniu istotności. W tym celu potrzebna jest znajomość wariancji resztowej klasy, s^2_K , oraz wariancji resztowej obiektu, s^2_j :

- wariancję resztową klasy oblicza się według wzoru (14)

$$s^2_K = \frac{\sum_{j=1}^n \sum_{k=1}^m (e_{jk})^2}{(m-k)(n_k - k - 1)} \quad (14)$$

gdzie: m - liczba zmienny w klasie A , k - liczba istotnych głównych składowych w klasie k , n_k - liczba obiektów w modelu w klasie k , e_{ki}^2 - różnica między rzeczywistą wartością i -tej zmiennej k -tego obiektu a wartością rzeczywistą obliczoną na podstawie modelu

- wariancję resztową obiektu oblicza się według wzoru (15)

$$s^2_j = \frac{\sum_{j=1}^m e_{jk}^2}{(m-k)} \quad (15)$$

Posiadając wartość wariancji resztowej dla klasy oraz dla j -tego obiektu wyznacza się wartość statystyki F

$$F = \frac{S_K^2}{S_j^2} \quad (16)$$

Po wyznaczeniu wartości F porównuje się ją z wartością F_{kryt} przy poziomie istotności α oraz $(n-r-1)$ i $(m-r)$ stopniach swobody. Jeżeli $F \geq F_{kryt}$ to obiekt nie należy do modelu klasy K . W przypadku gdy j -ty obiekt nie należy do żadnej z klasy wówczas jest to tzw. obiekt odosobniony.

2.3. Techniki chemometryczne stosowane w analizie ilościowej

Ilościowe wyznaczenie zawartości określonej substancji na podstawie widma polega na zbudowaniu matematycznego modelu zależności między określoną własnością lub własnościami badanej próbki a wartościami absorpcji promieniowania przy

określonych długościach fali. Skonstruowany model kalibracyjny umożliwia zastąpienie drogich i czasochłonnych pomiarów, pomiarami tańszymi i szybszymi. Zależność tą wyznacza się najczęściej metodami kalibracji liniowej. Do budowania metod ilościowych najczęściej stosuje się wielokrotną regresję liniową (MLR), regresję głównych składowych (PCR) oraz cząstkową regresję najmniejszych kwadratów (PLS).

2.3.1. Kalibracja jednowymiarowa

Najprostszym modelem matematyczny przedstawiającym zależność między pożądaną własnością badanej próbki (stężeniem) a absorbancją widma jest regresja jednowymiarowa i przyjmuje ona postać [4,51]:

$$[y]_n = [X]_{n \times 2} [b]_2 + [e]_n \quad (17)$$

gdzie:

- y** – wektor macierzy danych referencyjnych zawierający wartości kalibrowanego parametru wyznaczonego dla poszczególnych próbek
- X** – macierz danych (wartości absorbancji dla wybranej długości fali pochodzące od kolejnych próbek)
- b** – wektor współczynników regresji
- e** – wektor reszt modelu

$$\begin{array}{|c|} \hline y_1 \\ \hline y_2 \\ \hline \cdot \\ \hline \cdot \\ \hline \cdot \\ \hline y_n \\ \hline \end{array} = \begin{array}{|c|c|} \hline 1 & x_1 \\ \hline 1 & x_2 \\ \hline \cdot & \cdot \\ \hline \cdot & \cdot \\ \hline \cdot & \cdot \\ \hline 1 & x_n \\ \hline \end{array} \begin{array}{|c|} \hline b_0 \\ \hline b_1 \\ \hline \end{array} + \begin{array}{|c|} \hline e \\ \hline \end{array}$$

Rys. 7 Kalibracja jednowymiarowa

Wartości znajdujące się w macierzach **y** i **X** wykorzystywane są do wyznaczenia wektora współczynników regresji **b** metodą najmniejszych kwadratów. Wartości wektora współczynników regresji oblicza się według poniższego wzoru

$$\mathbf{b} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (18)$$

Opracowany model kalibracyjny (**Rys.7**) wykorzystywany jest do wyznaczenia nieznannej zawartości składnika, y_{unk} w nowych próbkach

$$y_{\text{unk}} = [1 \ x_{\text{unk}}] \mathbf{b} \quad (19)$$

Kalibracja jednowymiarowa stosowana jest w sytuacji, gdy oznacza się zawartość pojedynczego składnika przy użyciu pojedynczej długości fali. Wadą tej metody jest to, że nie może być ona stosowana w sytuacji gdy widmo pochodzące od danego komponenta nakłada się z widmami pochodzącymi od innych komponentów próbki. W spektroskopii bliskiej podczerwieni pasma absorpcyjne są zwykle szerokie i nakładają się na siebie, więc kalibracja jednowymiarowa nie znajduje praktycznego zastosowania.

2.3.2. Kalibracja wielowymiarowa

Kalibracja wielowymiarowa w przeciwieństwie do jednowymiarowej jest w stanie wyodrębnić nakładające się na siebie widma pochodzące od różnych składników badanej próbki. Jednocześnie jest w stanie wykorzystać więcej dostępnych w widmie informacji oraz zredukować wpływ szumu zakłócającego pomiar. Do budowy metod ilościowych najczęściej stosuje się wielokrotną regresję liniową (MLR), regresję głównych składowych (PCR) oraz cząstkową regresję najmniejszych kwadratów (PLS).

Ogólny wzór modeli kalibracji wielowymiarowej jest praktycznie taki sam jak w przypadku kalibracji jednowymiarowej. Z matematycznego punktu widzenia kalibracja wielowymiarowa różni się od kalibracji jednowymiarowej liczbą kolumn macierzy \mathbf{X} i wektora \mathbf{b} , **Rys.7**. Zatem model kalibracji wielowymiarowej przyjmuje następującą postać [4,52-53]

$$[y]_n = [\mathbf{X}]_{n,m+1} [\mathbf{b}]_{m+1} + [e]_{n,m} \quad (20)$$

a wektor współczynników \mathbf{b} oblicza się wg poniższego wzoru

$$\mathbf{b} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (21)$$

Warunkiem rozwiązania równania (21) jest możliwość odwrócenia macierzy $(\mathbf{X}^T \mathbf{X})$. Z zasad rachunku macierzowego wynika, że jest to możliwe wtedy i tylko wtedy gdy:

- liczba wierszy macierzy \mathbf{X} jest równa lub większa od liczby jej kolumn
- żadna z kolumn macierzy \mathbf{X} nie jest kombinacją liniową innych kolumn

Poszczególne metody kalibracji wielowymiarowej różnią się zawartością macierzy danych \mathbf{X} . W regresji wielokrotnej są to wartości absorbancji przy odpowiednio dobranych liczbach falowych, w metodzie PCR główne składowe, a w metodzie PLS czynniki najsilniej skorelowane z zawartością wyznaczanego składnika próbek.

$$\begin{array}{|c|} \hline y_1 \\ \hline y_2 \\ \hline \cdot \\ \hline \cdot \\ \hline \cdot \\ \hline y_n \\ \hline \end{array} = \begin{array}{|c|} \hline 1 \ x_{1,1} \ x_{1,2} \ \dots \ x_{1,m} \\ \hline 1 \ x_{2,1} \ x_{2,2} \ \dots \ x_{2,m} \\ \hline \cdot \ \cdot \ \cdot \ \dots \ \cdot \\ \hline \cdot \ \cdot \ \cdot \ \dots \ \cdot \\ \hline \cdot \ \cdot \ \cdot \ \dots \ \cdot \\ \hline 1 \ x_{n,1} \ x_{n,2} \ \dots \ x_{n,m} \\ \hline \end{array} \begin{array}{|c|} \hline b_0 \\ \hline b_1 \\ \hline b_2 \\ \hline \cdot \\ \hline \cdot \\ \hline \cdot \\ \hline b_m \\ \hline \end{array}$$

Rys.8 Kalibracja wielowymiarowa

2.3.2.1. Wielokrotna regresja liniowa (MLR-Multiple Linear Regression)

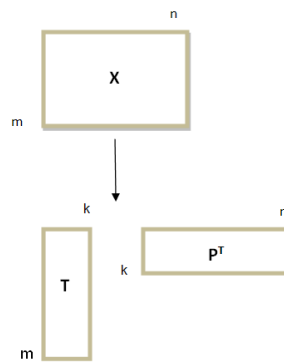
Metoda MLR jest klasyczną metodą tworzącą liniową zależność między zbiorem kilku zmiennych objaśniających \mathbf{X} a pojedynczą zmienną zależną y . Metoda ta jest stosowana w sytuacji, gdy liczba zmiennych jest stosunkowo mała i gdy zmienne te nie są ze sobą zbyt silnie skorelowane. W takiej sytuacji z równania (21) można wyznaczyć współczynniki regresji b_i . Jednak gdy liczba zmiennych przewyższa liczbę próbek i jednocześnie zmienne te są ze sobą bardzo silnie skorelowane wówczas wyznaczenie współczynnika regresji jest niemożliwe. Zgodnie z regułami algebry liniowej macierz odwrotna $(\mathbf{X}^T\mathbf{X})^{-1}$ nie istnieje, gdyż wyznacznik tej macierzy równy jest zero. Czasami z liczego zestawu zmiennych można wybrać taki ich podzbiór, że równanie (20) ma swoje rozwiązanie jednakże model taki będzie charakteryzował się bardzo złymi zdolnościami predykcyjnymi dla nowych próbek a doskonałym dopasowaniem do zbioru kalibracyjnego [54,55].

Metoda MLR ze względu na problem korelacji zmiennych ma ograniczone zastosowanie w spektroskopii, gdyż dane spektralne przeważnie zawierają silnie skorelowane zmienne. Z tym problemem radzą sobie inne metody kalibracji wielowymiarowej takie jak regresja głównych składowych PCR, czy też metoda częściowych najmniejszych kwadratów PLS. W tych metodach problem skorelowanych zmiennych przewyciężony został przez zastąpienie zmiennych pierwotnych nowymi

zmiennymi tzw. czynnikami, które z definicji nie są ze sobą skorelowane, a ich liczba jest zdecydowanie mniejsza niż liczba pierwotnych zmiennych objaśniających.

2.3.2.2. Regresja głównych składowych

Ideą regresji głównych składowych (PCR, ang. *Principal Component Regression*) jest zbudowanie zależności między poszukiwaną własnością próbki, na przykład zawartością danego komponenta a głównymi składowymi - PCs. PCR różni się od MLR tym że macierz \mathbf{X} zamiast skorelowanych zmiennych pierwotnych zawiera wzajemnie ortogonalne główne składowe PCs. Regresja głównych składowych łączy w sobie metodę MLR oraz PCA (*Principal Component Analysis*). Pierwszym etapem analizy jest wyznaczenie głównych składowych (p. 2.2.1) w celu uzyskania nieskorelowanych kolumn macierzy \mathbf{X} .



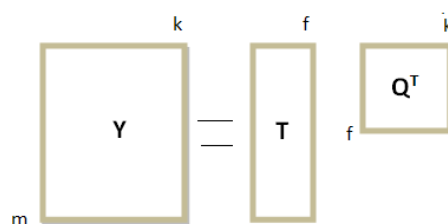
Rys.9 Model PCA

W dalszym etapie analizy macierz \mathbf{T} wykorzystana zostaje do konstrukcji modelu MLR.

$$[y]_n = [T]_{nk} [q^T]_{k+1} + [e]_n \quad (22)$$

gdzie q to wektor współczynników regresji modelu i wyznacza się go metodą najmniejszych kwadratów.

$$[q]_{f,k} = ([T^T]_{f,m} [T]_{f,k})^{-1} T_{f,m}^T [y]_{m,f} \quad (23)$$



Rys.10 Konstrukcja modelu kalibracyjnego używając macierz T

Ortogonalność głównych składowych pozwala wyznaczyć współczynniki regresji metodą najmniejszych kwadratów, gdyż zawsze istnieje macierz odwrotna $(T^T T)^{-1}$, co w metodzie MLR nie zawsze było możliwe. Podczas tworzenia modelu kalibracyjnego bardzo ważny jest wybór liczby głównych składowych, k . Zbyt duża liczba PCs może sprawić, że do modelu zostanie włączona niepożądana informacja jaką są błędy eksperymentalne czy też szumy. Z drugiej strony natomiast jeśli użyjemy zbyt małą liczbę PCs model nie będzie zawierał informacji niezbędnej aby uzyskać poprawny model kalibracyjny posiadający zdolności prognostyczne. Ważne jest również aby przy ustalaniu liczby głównych składowych uwzględnić wszystkie możliwe źródła zmienności chemicznej [56-58]. Zazwyczaj nie posiadamy takich informacji na temat naszych danych dlatego też należy być ostrożnym w doborze liczby składowych. Najczęściej liczbę głównych składowych modelu PCR dobiera się metodą prób i błędów dążąc do zbudowania modelu o najlepszych zdolnościach prognostycznych.

2.3.2.3. Częstkowa metoda najmniejszych kwadratów

Częstkowa metoda najmniejszych kwadratów (PLS, ang. *Partial Least Square*) jest najczęściej stosowaną metodą w konstruowaniu wielowymiarowych modeli kalibracyjnych. Zadaniem metody PLS, podobnie jak PCR, jest konstrukcja modelu w oparciu o wzajemnie ortogonalne czynniki. W metodzie PLS czynniki takie tworzone są jednak inaczej niż w metodzie PCR. W metodzie PCR czynnikami tymi są główne składowe tworzone wyłącznie na podstawie macierzy \mathbf{X} , podczas gdy przy tworzenia czynników w metodzie PLS brana jest pod uwagę zależność pomiędzy macierzami \mathbf{X} a \mathbf{y} . Każdy czynnik metody PLS objaśnia maksymalną kowariancję pomiędzy czynnikami w przestrzeni x a czynnikami w przestrzeni y . Kowariancja łączy wysoką wariancję macierzy X z wysoką korelacją interesującej nas własności [59-61].

PLS 1

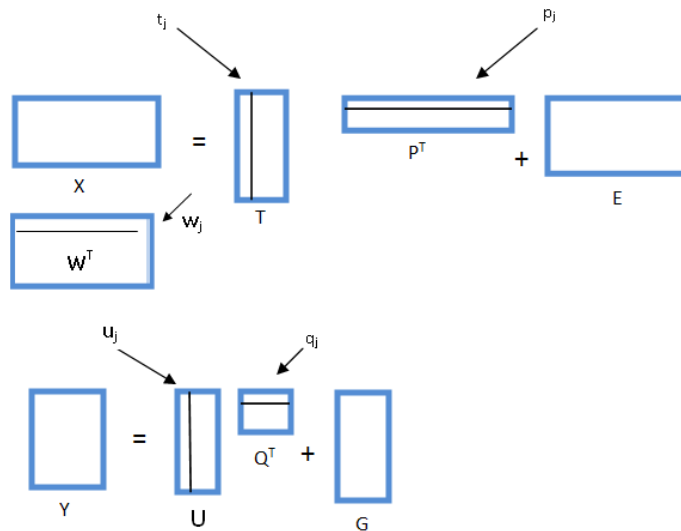
Metoda regresyjna PLS 1 polega na zbudowaniu zależności między X a Y dla każdej zmiennej osobno. Matematyczny zapis tej zależności ma następującą postać:

$$[X]_{m,n} = [T]_{m,f} [P]_{f,n}^T + [E]_{m,n} \quad (24)$$

$$[Y]_{m,k} = [U]_{m,f} [Q]_{f,k}^T + [G]_{m,k} \quad (25)$$

W trakcie konstrukcji modelu dla zmiennych X uzyskuje się macierze wyników T, wag P oraz macierz wag PLS W. Macierz wag P zawiera informację o zależności pomiędzy wynikami a oryginalnymi zmiennymi macierzy X. Wagi PLS W natomiast opisują relację jaka istnieje pomiędzy Y a oryginalnymi zmiennymi.

Pierwszy czynnik w metodzie PLS objaśnia maksymalną kowariancję i oznaczany jest przez jednostkowy wektor wag w_1 . Wyniki wzdłuż tej osi wyznaczone są jako $t_1 = Xw_1$. Wszystkie zmienne z macierzy X są regresyjnie przetwarzane do t_1 w celu otrzymania wektora wag p_1 . Współczynnik regresyjny q_1 jest przetwarzany regresyjnie na t_1 . Iloczyn t_1 i p_1 jest następnie odejmowany z macierzy X a t_1 i q_1 z Y. Kolejny czynnik jest wyznaczany w ten sam sposób, tyle że wyznacza się go z macierzy X powstałej po odjęciu czynnika pierwszego. Dla macierzy Y otrzymuje się również macierz wag i wyników U i macierz wag Q. Wkład poszczególnych zmiennych do modelowania Y określają ich współczynniki korelacji B [51,62].



Rys.11 Schemat metody PLS1

PLS 2

Metoda PLS2 jest rozszerzoną wersją metody PLS1. Istnieje tylko niewielka różnica między obiema tymi modelami. Różnica ta polega na tym, że metoda PLS2 objaśnia kilka zmiennych (wszystkie dostępne zmienne) jednocześnie, **Rys.11**. Użycie metody PLS2 często prowadzi do uzyskania gorszych wyników niż w przypadku metody PLS1. Korzyści stosowania modelu PLS2 można zaobserwować jedynie, gdy pomiędzy zmiennymi Y istnieje silna korelacja [53,62].

3.Procedura postępowania podczas opracowywania metod jakościowych i ilościowych

Spektroskopia w podczerwieni wraz z nowoczesnymi technikami chemometrycznymi wykorzystywana jest w analizie jakościowej oraz ilościowej. Przez analizę jakościową rozumie się identyfikację badanego materiału przez porównanie jego widma z referencyjną biblioteką widm za pomocą miary statystycznej.

Ilościowe wyznaczenie zawartości określonej substancji na podstawie widma próbki zawierającej tą substancję nosi nazwę analizy ilościowej. Polega ona na zbudowaniu matematycznego modelu zależności między wartościami absorpcji promieniowania przy określonych długościach fali a zawartością oznaczanej substancji w próbce. Zależność tą wyznacza się najczęściej metodami kalibracji liniowej. Analiza ilościowa dotyczyć może nie tylko zawartości substancji w próbce, ale również wybranej właściwości próbki. W obu przypadkach skonstruowany model kalibracyjny umożliwia zastąpienie drogich i czasochłonnych pomiarów, pomiarami tańszymi i szybszymi.

3.1. Metoda jakościowa

Procedura opracowywania metod jakościowych składa się z kilku etapów, **Rys.12**. Pierwszym etapem jest optymalizacja procesu pomiarowego. Na tym etapie dobiera się odpowiednio tryb pomiarowy, ilość skanów, rozdzielczość, zakres widmowy. Wyklucza się substancje które nie absorbują promieniowania podczerwonego (są to najczęściej substancje nieorganiczne nie posiadające wiązania O-H). Następnie zbiera się widma substancji tworząc tzw. bibliotekę widm. Etap tworzenia biblioteki widm jest procesem iteracyjnym. Próbkę wykorzystywane do tworzenia biblioteki widm powinny pochodzić z różnych serii i od różnych dostawców, jednocześnie wszystkie widma zarejestrowanych prób powinny mieć potwierdzoną tożsamość metodą referencyjną. W kolejnym etapie widma substancji dla której tworzy się model jakościowych dzieli się na dwa zbiory: zbiór

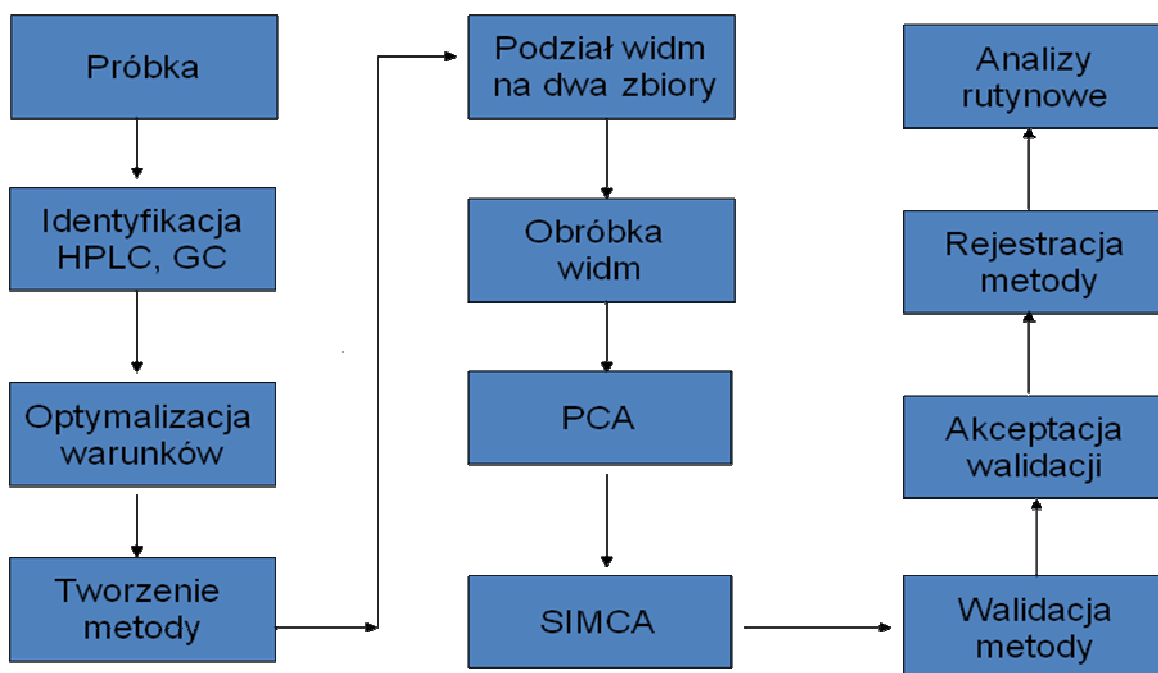
uczący (modelowy) oraz zbiór testowy. Zbiór uczący powinien zawierać próbki reprezentujące wszystkie możliwe źródła wariacji takie jak:

- składniki chemiczne (wielkości cząstek, zmiany składu chemicznego wynikające ze starzenia się materiału)
- czynniki fizyczne – temperatura, zmętnienie, wilgotność, pH
- zmienności pochodzące od spektrometru – nieliniowość detektora, starzenie lampy i instrumentu

Następnie przeprowadza się obróbkę matematyczną surowych widm ze zbioru uczącego, analizę głównych składowych, na tym etapie dokonuje się wyboru liczby istotnych głównych składowych a następnie przeprowadza się klasyfikację. Liczba próbek w zbiorze uczącym zależy od złożoności matrycy badanego materiału, a także od zastosowanego algorytmu. Algorytmy korelacyjne wymagają najmniejszej liczby widm. W przypadku algorytmów wykorzystujących odległości konieczna jest większa liczba prób.

Dobierając widma prób czy to do zbioru uczącego czy testowego należy dokonać oceny widm. Ocena widm ma na celu wyeliminowanie czy też zaobserwowanie prób odbiegających. Próby odbiegające świadczą, albo o tym że dana próba posiada nietypową cechę, albo o niepoprawnym zarejestrowaniu widma badanego materiału. W końcowym etapie dokonuje się oceny predykcyjnej stworzonej metody za pomocą zbioru testowego. Jeśli wszystkie próbki ze zbioru testowego poprawnie zostały zidentyfikowane przeprowadza się proces walidacji zewnętrznej za pomocą prób innych serii niż tych, które zostały użyte w zbiorze uczącym i testowym. Dodatkowo do walidacji modeli jakościowych wykorzystuje się próbki różnych substancji podobnych pod względem chemicznym, do tych które są identyfikowane.

Raz opracowane metody jakościowe należy co jakiś czas rewalidować, uwzględniając dodatkowe pojawiające się parametry np. zmianę dostawcy surowców, pojawiająca się dodatkową zmienność.



Rys.12 Schemat opracowywania metod jakościowych

3.2. Metoda ilościowa

Procedura opracowywania metod ilościowych pod kilkoma względami jest podobna do metod jakościowych. Konstrukcja modeli ilościowych podobnie jak jakościowych składa się z kilku etapów. Bardzo ważnymi etapami w konstruowaniu metody ilościowej jest: a) dobór prób kalibracyjnych, b) matematyczna obróbka danych, c) wybór metody kalibracyjnej, d) optymalizacja i kompleksowość modeli kalibracyjnych.

Zbiór kalibracyjny powinien zawierać próby reprezentujące wszystkie możliwe źródła wariacji (składniki chemiczne, czynniki fizyczne). Ponadto w zbiorze tym powinny znajdować się próby, w których wartość oznaczanego składnika powinna być poniżej jak i powyżej wartości deklarowanej. Przy czym zawartość pozostałych składników nie powinna ulegać zmianie. Reprezentatywność zbioru kalibracyjnego można zapewnić na kilka sposobów. Można przygotować serię próbek o określonej charakterystyce. Takie próbki można przygotować w laboratorium, uzyskując pokrycie całego zakresu zmienności zawartości oznaczanej substancji, jednakże bez pokrycia zmienności fizykochemicznej typowej dla prób produkcyjnych. Próbki zbioru kalibracyjnego można również pozyskać bezpośrednio z prób produkcyjnych. Próbki takie będą wówczas reprezentowały zmienność prób później badanych [63-68]. Kolejnym bardzo ważnym etapem jest obróbka wstępna surowych widm. Na tym etapie konstrukcji modeli ilościowych stosuje się takie same techniki pre-processingu jak w przypadku metod

jakościowych. Najczęściej stosuje się : transformację SNV, MSV, transformację „offset correction”, transformację „detrend correction”, transformację widm do ich drugich pochodnych algorytmem Savitzky’ego-Golaya oraz ważenie szumów, ważenie atmosferyczne, wygładzanie. Po zakończeniu tego etapu dokonuje się wyboru algorytmu kalibracyjnego. Wybór odpowiedniego algorytmu kalibracyjnego zależy od ilości zmiennych oraz od występowania korelacji między zmiennymi. W sytuacji gdy brak jest korelacji między zmiennymi i liczba zmiennych jest stosunkowo mała można wtedy zastosować wielokrotną regresję liniową (MLR). Jeżeli jednak liczba zmiennych jest bardzo duża i jednocześnie występuje silna korelacja między zmiennymi wówczas korzysta się z innych algorytmów (PCA- analiza głównych składowych, PCR – regresja głównych składowych, PLS- cząstkowa metoda najmniejszych kwadratów). W końcowym etapie dokonuje się optymalizacji oraz oceny kompleksowości utworzonego modelu. Na tym etapie dokonuje się wyboru liczby istotnych czynników oraz sprawdza się zdolności predykcyjne opracowanego modelu za pomocą zbioru testowego oraz zbioru walidacyjnego. Wybór liczby czynników modelu kalibracyjnego jest bardzo ważnym zagadnieniem. Modele zbudowane w oparciu o zbyt małą liczbę czynników są źle dopasowane do danych. Natomiast modele o zbyt dużej liczbie czynników, charakteryzują się małym błędem dla obiektów zbioru kalibracyjnego, ale dużym błędem predykcyjnym. Wybór optymalnej liczby czynników do konstrukcji modelu pozwala na osiągnięcie kompromisu między zadawalającym dopasowaniem modelu do danych a dobrymi własnościami predykcyjnymi. W sytuacji gdy liczba dostępnych próbek jest ograniczona i ich podział na 3 zbiory jest niemożliwy wówczas do wyboru liczby czynników, a także walidacji stosuje się tylko dwa zbiory: zbiór kalibracyjny oraz testowy.

Do oceny kompleksowości modelu oraz oszacowania błędu predykcyjnego opracowanego modelu wykorzystuje się również metody walidacji krzyżowej (ang. *cross-validation*). Ze względu na sposób tworzenia zbioru walidacyjnego wyróżnia się walidację Monte Carlo oraz walidację typu „wyrzuć jeden obiekt”[65]. Ogólnie metoda *cross validation* polega na chwilowym usunięciu ze zbioru kalibracyjnego odpowiednio jednej lub kilku próbek. Próbki te usuwane są losowo i bez powtórzeń. Nowo powstały pomniejszony zbiór służy do konstrukcji modelu kalibracyjnego na podstawie którego przewiduje się wartości dla usuniętych próbek. Dla próbek kolejno usuniętych zbiera się wartości reszt modelu o danej liczbie czynników. Następnie na ich podstawie oblicza się średni błąd *cross validation* RMSECV [69] (ang. *Root Mean Square Error of Cross*

Validation). Za optymalną liczbę czynników uznaje się taką liczbę przy której wartość RMSECV osiąga minimum globalne lub pierwsze minimum lokalne.

$$RMSECV = \sqrt{\left(\sum_{i=1}^n y_i - \hat{y}_i \right)^2} / N \quad (28)$$

W sytuacji gdy ze zbioru usuwany jest jeden obiekt mamy do czynienia z *cross validation* typu „wyrzuc jeden obiekt”. Natomiast jeśli zbiór losowo zostanie podzielony na dwa, włączając za każdym razem do zbioru walidacyjnego k próbek wówczas mamy do czynienia z walidacją Monte-Carlo. Spośród tych dwóch metod najczęściej stosuje się walidację typu „wyrzuc jeden obiekt”. Wybór liczby czynników na podstawie *cross validation* zmniejsza ryzyko przeuczenia modeli [69-71]. Zdolność predykcyjną modeli można również sprawdzić za pomocą innych parametrów statystycznych, takich jak: Średni błąd kwadratowy walidacji RMSEP (*Root mean Square of prediction*) oraz Q – współczynnik walidacji

Walidacja wewnętrzna

Przez pojęcie walidacji opracowanego modelu rozumie się ocenę jego zdolności predykcyjnych dla zbioru próbek nie użytych podczas jego konstrukcji. Zbiorem sprawdzającym zdolność predykcyjną modelu kalibracyjnego jest zbiór testowy. W sytuacji gdy liczba dostępnych próbek jest mała zdolności predykcyjne modeli oblicza się stosując metodę walidacji krzyżowej typu „wyrzuc jeden obiekt” uzyskując wartość RMSECV [66].

Zdolność predykcyjną modeli można również sprawdzić za pomocą innych parametrów statystycznych, takich jak:

- Średni błąd kwadratowy walidacji RMSEP – (*Root Mean Square of Errors of Prediction*)
- Q_w – współczynnik walidacji

RMSEP opisuje zdolność predykcyjną modelu kalibracyjnego dla próbek zbioru testowego

$$RMSEP = \sqrt{\left(\sum_{i=1}^n y'_i - y_i \right)^2} / N \quad (29)$$

gdzie y'_i i y_i oznaczają odpowiednio wartość przewidzianą próbki zbioru walidacyjnego i wartość otrzymaną metoda referencyjną.

Współczynnik walidacji opisuje stopień wyjaśnienia przez model wariancji zmiennej y w zbiorze kalibracyjnym. Współczynnik korelacji wyznaczany jest między wartościami przewidzianymi a referencyjnymi zbioru kalibracyjnego i oblicza się go z poniższego wzoru

$$Q_w = \sqrt{1 - \frac{\sum_{i=1}^n (y'_i - y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (30)$$

Kiedy wartość współczynnika korelacji jest bliska 1 lub równa 1 oznacza to że model doskonale opisuje wariancję zmiennej y w próbkach zbioru kalibracyjnego, jednak gdy wartość ta jest bliska zero wówczas świadczy to o tym, że model źle opisuje wariancję zmiennej y a zatem jest źle dopasowany.

Odpowiednikiem współczynnika korelacji dla próbek zbioru testowego jest tzw. współczynnik walidacji Q obliczany również zgodnie ze wzorem (30), lecz dla próbek zbioru testowego.

O dobrej zdolności predykcyjnej opracowanego modelu świadczą: mała wartość RMSEP, wysoka wartość współczynnika walidacji a także wykres reszt na którym próbki zbioru kalibracyjnego oraz testowego powinny być rozmieszczone symetrycznie wzdłuż linii prostej o nachyleniu 45° [72-77].

Inną użyteczną miarą oceny zdolności prognostycznej modelu jest parametr SEP definiowany jako odchylenie standardowe różnicy między wartością przewidzianą a referencyjną.

$$SEP = \sqrt{\sum_{i=1}^N (y'_i - y_i - BIAS)^2 / (N - 1)} \quad (31)$$

gdzie

$$BIAS = \sum_{i=1}^N (y'_i - y_i) / N \quad (32)$$

który interpretowany jest jako średnie odchylenie między y' a y w zbiorze predykcyjnym.

4. Punkty odbiegające

Podczas opracowywania modelu ważne jest aby zwrócić uwagę na obecność tzw. punktów odbiegających. Punkty takie mogą się pojawić zarówno w zbiorze kalibracyjnym ale także walidacyjnym. Zdarza się również sytuacja że mogą się pojawić na etapie testowym czy też w trakcie już rutynowej analizy.

4.1. Rodzaje punktów odbiegających

Punkty odbiegające występujące w zbiorze kalibracyjnym oraz walidacyjnym mają inną naturę. Punkty odbiegające w zbiorze kalibracyjnym mają istotny wpływ na jakość opracowywanego modelu. Ich występowanie świadczy o niejednorodności zbioru kalibracyjnego. Model opracowany na podstawie zbioru kalibracyjnego z punktami odbiegającymi zawiera zwykle dużą liczbę zmiennych objaśniających i jest „przeuczony”.

Punkty odbiegające pojawiają się w zbiorze testowym, gdy w zbiorze kalibracyjnym nie zostały uwzględnione wszystkie źródła zmienności występujące w badanych próbkach. W takiej sytuacji należy rozszerzyć zbiór kalibracyjny o próbki zawierające takie źródło zmienności.

Punkty odbiegające mogą się pojawić w zbiorze testowym w sytuacji gdy:

- Stężenie oznaczanej substancji w próbce testowej przekracza zakres stężeń próbek zbioru kalibracyjnego
- Próbka testowa posiada dodatkowe źródła zmienności nie uwzględnione w zbiorze kalibracyjnym
- Próbka badana należy do innej populacji danych
- Podczas zbierania widma próbki testowej pojawiły się problemy natury sprzętowej
- Wartość referencyjna próbki testowej została źle wyznaczona

Punkty odbiegające mogą się również pojawić w trakcie walidacji zewnętrznej lub też w trakcie rutynowej analizy. Użyteczne jest aby w takich przypadkach rozróżnić między punktami odbiegającymi typu x oraz y. Punkty odbiegające typu x charakteryzują się wartościami zmiennych objaśniających, które w jakiś sposób odbiegają od wartości danych X wykorzystywanych na etapie kalibracji. Natomiast punkty odbiegające typu y definiowane są jako te, które posiadają inną zależność pomiędzy zmiennymi objaśniającymi x_i a zmienną zależną y, niż uwzględnioną w modelu [72,78-80]. Na etapie rutynowej analizy możliwe są do wykrycia jedynie punkty odbiegające typu x.

4.2. Wykrywanie punktów odbiegających typu \mathbf{x}

Przy wykrywaniu punktów odbiegających typu \mathbf{x} przydatne są różne miary statystyczne oraz sposoby wizualizacji. Jedną z takich miar jest odległość Mahalanobisa, MD^2 , która opisuje odległość między i -tym punktem zbioru X a środkiem tego zbioru

$$MD^2(i) = (\mathbf{x}_i - \bar{\mathbf{x}})\mathbf{S}^{-1}(\mathbf{x}_i - \bar{\mathbf{x}})^T \quad (32)$$

gdzie \mathbf{S} jest macierzą kowariancji

$$\mathbf{S} = \frac{1}{n-1}(\mathbf{X}_i^T \mathbf{X}_i) \quad (33)$$

Użycie odległości Mahalanobisa wymaga przyjęcia dodatkowego założenie, że punkty zbioru X mają wielowymiarowy rozkład normalny. Odległość Mahalanobisa, w odróżnieniu od zwykłej odległości euklidesowej, bierze pod uwagę korelacje występujące pomiędzy zmiennymi objaśniającymi. Jest więc szczególnie przydatna, gdy zmiennymi objaśniającymi są silnie skorelowane zmienne spektroskopowe.

Inną wielkością pozwalającą na wykrycie punktów odbiegających typu \mathbf{x} jest tzw. wartość dźwigni (ang. *leverage*). Wartość dźwigni i -tego punktu zbioru X definiuje się jako wartość i -tego elementu diagonalnej macierzy H określonej następująco

$$H = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T, \quad (34)$$

Można również określić wartość dźwigni dla pojedynczego punktu \mathbf{x}_i (należącego lub nie należącego do zbioru X). Wynosi ona:

$$h_i = \mathbf{x}_i (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i^T \quad (35)$$

Wartości h_i dla elementów jednorodnego zbioru o rozkładzie zbliżonym do wielowymiarowego rozkładu normalnego przyjmują zazwyczaj wartości z przedziału od 0 do 1. Próbkki, których wartości \mathbf{x} są dalekie od środka zbioru X posiadają zazwyczaj dużo wyższe wartości h_i , co pozwala je łatwo zidentyfikować jako punkty odbiegające.

Istnieje liniowa zależność między wartością dźwigni punktu oraz jego odległością Mahalanobisa [78,81]

$$h_i = \frac{1}{n-1} MD^2(i) + \frac{1}{n} \quad (36)$$

5. Zastosowanie spektroskopii NIR w przemyśle.

Połączenie metod spektroskopii w bliskiej i klasycznej podczerwieni oraz wielowymiarowych technik chemometrycznych umożliwia uzyskanie efektywnych metod służących do szybkiej identyfikacji surowców i półproduktów oraz określenia składu mieszaniny na analitycznie akceptowalnym poziomie dokładności, precyzji i wydajności. Technika ta ze względu na szybkie oraz nieniszczące pomiary ilościowe i jakościowe stanowi interesującą alternatywę dla tradycyjnych metod analiz przemysłowych. Ze względu na wiele zalet spektroskopia w podczerwieni znalazła zainteresowanie w wielu gałęziach przemysłu, między innymi w:

- przemyśle farmaceutycznym
- przemyśle kosmetycznym
- przemyśle petrochemicznym
- przemyśle spożywczym
- oraz w analityce klinicznej i ochronie środowiska

Ze względu na szerokie spektrum zastosowania spektroskopii w podczerwieni w różnych gałęziach przemysłowych poniżej opisano wykorzystanie spektroskopii NIR w dwóch głównych sektorach: przemyśle farmaceutycznym i spożywczym.

5.1. Przemysł farmaceutyczny

Przemysł farmaceutyczny posiada bardzo ostre i wymagające kryteria kontroli jakości zarówno surowców farmaceutycznych ale także samych produktów końcowych. Ze względu na to że tradycyjne metody analityczne są dosyć pracochłonne i czasochłonne coraz więcej firm farmaceutycznych decyduje się na wykorzystanie spektroskopii bliskiej podczerwieni w analizie jakościowej oraz ilościowej. Zastosowanie spektroskopii bliskiej podczerwieni w przemyśle farmaceutycznym jest bardzo szerokie. Spektroskopia NIR najczęściej wykorzystywana jest w: identyfikacji substancji aktywnych oraz substancji pomocniczych w produktach leczniczych takich jak: tabletki, pastylki, kapsułki, blistry, tabletki powlekane i ich rdzenie zarówno w trakcie procesu otrzymywania jak i w samym produkcie końcowym. Technika NIR stosowana jest również w analizie ilościowej do

oznaczania poziomu wilgotności, postaci polimorficznej, twardości tabletek, homogeniczności i wielkości cząstek [4].

5.1.1. Identyfikacja substancji

Identyfikacja surowców farmaceutycznych jest bardzo ważnym etapem kontroli jakości. Identyfikacja surowców powinna potwierdzać, że surowiec spełnia szereg parametrów takich jak: wilgotność, wielkość cząstek, morfologię oraz źródło pochodzenia. Tradycyjne metody stosowane do potwierdzania tożsamości oraz kontroli jakości wymagają licznych, pracochłonnych i kosztownych analiz [82]. Wykorzystanie spektroskopii bliskiej podczerwieni w analizie tożsamościowej skraca czas analizy przyspieszając tym samym produkcję produktów leczniczych. Identyfikacja badanych surowców przy pomocy spektroskopii NIR może odbywać się na kilka sposobów: a) porównanie kształtu i intensywności widma substancji z widmem standardu odniesienia zakupionego lub otrzymanego w podobnych warunkach, b) porównanie intensywności absorbancji przy określonej długości fali, c) identyfikacja za pomocą klasyfikatora SIMCA.

Po raz pierwszy spektroskopia NIR została wykorzystana w analizie tożsamościowej w 1982 przez Johna Rose, który przedstawił możliwości jej wykorzystania w analizie jakościowej penicyliny. Obecnie wiele doniesień literaturowych przedstawia możliwości wykorzystania spektroskopii NIR w identyfikacji surowców bardzo podobnych do siebie pod względem chemicznym np.: różne typy ampicyliny, czy też laktozy [83]. Spektroskopia NIR zastosowana została również do identyfikacji substancji powleczonych oraz niepowleczonych i zblistrowanych tabletek. W tym celu wykorzystano sondę do zbierania widm. Do budowy metody wykorzystano klasyfikator SIMCA, dystans Mahalanobiasa oraz metodę resztkowej wariancji. Wyniki uzyskane wykazały, że metody rozpoznawania obrazu sprawdzają się w identyfikacji tabletek powleczonych i zapakowanych w takich samych warunkach. Natomiast zmiany w formulacji lub składzie foli opakowań wymagają ponownego procesu rewalidacji metody [84].

W celu identyfikacji surowców farmaceutycznych Blanco opracował bibliotekę widm. Biblioteka ta zawiera widma substancji stosowanych do produkcji substancji leczniczych wytwarzanych w fabryce Laboratoris Menarinni w Barcelonie. Biblioteka opracowana przez Blanco umożliwiła przyspieszenie procesu identyfikacji substancji farmaceutycznych, wykluczając tradycyjne techniki identyfikacyjne [61].

5.1.2. Analiza ilościowa

Spektroskopia w bliskiej podczerwieni ma obecnie bardzo duże spektrum zastosowania w analizie ilościowej w przemyśle farmaceutycznym. Zainteresowanie spektroskopią w podczerwieni i bliskiej podczerwieni do celów analizy ilościowej wzrastało wraz z rozwojem komputerów oraz samej techniki IR wykorzystującej transformację Fouriera. Zainteresowanie to wynikało z szeregu zalet jakie niesie ze sobą połączenie spektroskopii w podczerwieni wraz z technikami chemometrycznymi.

Spektroskopia NIR oraz IR umożliwia w szybki sposób określenie zawartości substancji aktywnych oraz pomocniczych w preparatach farmaceutycznych np. przy oznaczaniu zawartości paracetamolu w tabletkach. Prosty a zarazem szybki pomiar umożliwia przeprowadzenie jednocześnie analizy tożsamościowej oraz ilościowej. Jedynym bardzo ważnym problemem jaki pojawia się podczas analizy ilościowej jest zawartości substancji oznaczanej. Powinna ona być w przeważającej ilości w stosunku do pozostałych substancji lub charakteryzować się intensywnym i unikalnym widmem.

5.1.3. Polimorfizm i izomery

Spektroskopia bliskiej podczerwieni znalazła zastosowanie w oznaczaniu stopnia polimorficzności czy też pseudopolimorficzności wielu substancji takich jak: kofeina, laktoza, mannitol, rafinoza [58]. Oznaczanie stopnia polimorficzności jest bardzo ważne ponieważ ma on istotny wpływ na właściwości fizykochemiczne oraz biofarmaceutyczne materiałów farmaceutycznych. Spektroskopia NIR stała się alternatywną metodą do oznaczania zawartości poszczególnych form polimorficznych w stosunku do tradycyjnych metod takich jak DSC czy rentgenowskiej dyfrakcji proszkowej (X-ray powder diffraction). Palet wraz ze współpracownikami wykazał że spektroskopia NIR może być wykorzystywana do określania zawartości form polimorficznych sulfatiazolu na poziomie 0,3% w substancji dwuskładnikowej [85-86].

Technika NIR znalazła również zastosowanie w identyfikacji i oznaczaniu izomerów. Duża wrażliwość widm NIR na obecność i liczbę międzycząsteczkowych wiązań wodorowych umożliwia wykorzystanie tej techniki do rozróżniania izomerów takich jak efedryna i pseudoefedryna [4].

5.1.4. Opakowania

W przemyśle farmaceutycznym oprócz potwierdzania tożsamości substancji aktywnych oraz pomocniczych bardzo ważna jest również kontrola jakości opakowań w których znajdują się substancje lecznicze. W 1985 Shintai-Young oraz Ciurczak

wykazali możliwość zastosowania widm otrzymanych ze spektroskopu NIR wraz z metodą LDA w identyfikacji polimerów użytych w opakowaniach (butelki plastikowe, blistry) [4].

5.1.5. Wilgotność

Woda absorbuje bardzo silnie w regionie NIR, szczególnie charakterystycznych jest 5 pasm absorpcji pochodzących od cząsteczek wody 760, 970, 1190, 1450, 1940 nm. Pasma te pochodzą od pierwszego nadtonu O-H wiązań rozciągających oraz od kombinacyjnych wiązań rozciągających oraz zginających. Zakres w którym absorbuje woda wykorzystywany jest do oznaczania zawartości wody w wielu produktach farmaceutycznych. Blanco jako pierwszy przedstawił potencjał i możliwość zastosowania spektroskopii NIR w oznaczaniu zawartości wody [87]. Następnie Ciurczak i jego współpracownicy wykazali możliwość zastosowania NIR w rozróżnianiu różnych postaci wody w farmaceutykach. Dziki z kolei wykazał możliwości zastosowania NIR w rozróżnianiu materiałów krystalicznych ze względu na orientację wypełniających je wody [88].

Oznaczanie zawartości wody jest bardzo ważne w przypadku żelatynowych kapsułek. Fizyczny charakter tej formy leku silnie związany jest z odpowiednią zawartością wody. W przypadku zbyt małej zawartości wody, kapsułki stają się łamliwe i na zewnątrz może wydostawać się substancja aktywna. Natomiast zbyt duża zawartość wody powoduje, że kapsułki zaczynają się kleić. Dlatego też ważna jest kontrola poziomu wody w tej postaci leku. Jednakże tradycyjne metody, takie jak oznaczanie straty wody w trakcie suszenia czy też metoda Karla Fishera, są długie i pracochłonne. Zastosowanie spektroskopii NIR w oznaczaniu wilgotności kapsułek żelowych skróciło czas analizy z 16 godzin do 2 minut [89].

5.1.6. Twardość

Nie inwazyjna natura spektroskopii NIR sprawiła, że metoda ta z powodzeniem może być wykorzystywana w oznaczaniu poziomu twardości tabletek. Badanie „twardości” jest bardzo ważnym procesem kontroli farmaceutycznej gdyż dostarcza ono informacji na temat struktury otrzymanego farmaceutyku. Istnieje szereg dostępnych artykułów przedstawiających zastosowanie techniki NIR w oznaczaniu tego parametru [90-91]. Pierwsze zastosowanie NIR w oznaczaniu „twardości” zostało opublikowane przez Drenna, Ciurczaka i Loddera w XX wieku. Autorzy wykazali, że podczas formowania tabletek w zakresie ciśnienia od 1 do 7 kP, wraz ze wzrostem twardości tabletek następuje przesunięcie linii bazowej w kierunku wyższych absorpcji, co jest

związane z efektem rozpraszania promieniowania. Autorzy wykazali również, że wraz ze wzrostem ciśnienia użytego do formowania tabletki występuje inny zależny od ciśnienia efekt wynikający ze zmian w oddziaływaniach międzycząsteczkowych, który przejawia się przesunięciem pików wraz ze wzrostem ciśnienia. W oznaczaniu twardości tabletek autorzy wykorzystali algorytmy PCR i PCA oraz zakres ciśnień od 6 do 12kP [2,90].

5.1.7. Wielkość cząstek

Na proces formułacji końcowych produktów farmaceutycznych ma wpływ wielkość cząstek surowców. Wielkość tych cząstek wpływa na stopień mieszania, kompleksowość, gęstość oraz rozpuszczalność farmaceutyków. Dlatego też bardzo ważna jest kontrola wielkości cząstek surowców przed przystąpieniem do procesu otrzymywania formy leku. W określaniu tego parametru przydatna okazała się spektroskopia bliskiej podczerwieni. W określaniu odpowiedniej wielkości cząstek wykorzystano zjawisko przesunięcia linii bazowej, która dryfuje wraz ze wzrostem cząstek surowca [55].

5.2. Przemysł spożywczy

Przemysł spożywczy jako pierwszy wykorzystał potencjalne możliwości wykorzystania spektroskopii w bliskiej podczerwieni w analizie ilościowej oraz jakościowej. Pierwszym sukcesem jaki odniesiono w wykorzystaniu spektroskopii NIR było oznaczanie zawartości białka w ziarnach pszenicy [92]. Kolejnym krokiem były pomiary innych parametrów ziaren pszenicy takich jak twardość, granulacja [93]. Obecnie spektroskopia bliskiej podczerwieni wykorzystywana jest w analizie jakościowej oraz ilościowej w przemyśle mięsnym, mleczarskim, owocowo-warzywnym i wielu innych.

5.2.1. Mięso i ryby

Po raz pierwszy spektroskopia bliskiej podczerwieni została wykorzystana w przemyśle mięsnym w 1996 przez Isakssona. Autor wykorzystał odbiciową spektroskopię NIR do oznaczania zawartości wody, tłuszczu, oraz białek w wołowym mięsie mielonym. Do budowy modelu Issakson wykorzystał algorytm MLR [94]. Praca Issaksona wzbudziła zainteresowanie kolejnych naukowców tą techniką i w 1999 TØgersen wraz ze współpracownikami opracowali modele ilościowe do oznaczania zawartości wody, białek i tłuszczu zarówno w wołowinie jak i wieprzowinie, a w 2003 model został rozszerzony o zamrożone produkty mięsne [95]. Spektroskopia NIR znalazła również zastosowanie w identyfikacji oraz klasyfikacji różnych gatunków mięsa [96].

Spektroskopia bliskiej podczerwieni znalazła także zastosowanie w analizie mięsa ryb. Gjerde i Martens wykorzystali technikę NIR do oznaczania zawartości wody, białek i tłuszczu w homogenizowanych zamrożonych płatach rybnych [97]. Błąd predykcji opracowanych modeli wynosił odpowiednio: 0,45% dla tłuszczu, 0,35% dla wody oraz 0,50% dla białek. Solberg wraz z kolegami określił zawartość tłuszczu w atlantyckim łososiu. Cozzolino ze współpracownikami wykazał możliwość wykorzystania spektroskopii w podczerwieni do potwierdzania autentyczności gatunków ryb wykorzystywanych w przygotowanych posiłkach rybnych [90]. Autorzy skanowali próbki w zakresie 1100-2500 nm i następnie za pomocą PCA, DPLC, LDA sprawdzali autentyczność zadeklarowanych gatunków ryb.

5.2.2. Produkty mleczne

Spektroskopia w podczerwieni jest kluczową techniką wykorzystywaną w analizie on-line produktów mlecznych. Produkty mleczne analizowane są za pomocą spektroskopii w podczerwieni od 1980 roku [98]. Wraz z rozwojem komputerów oraz aparatury do spektroskopii NIR/IR zwiększyło się zastosowanie tej techniki w przemyśle mleczarskim. Obecnie technika ta wykorzystywana jest w oznaczaniu zawartości wody, białek, tłuszczu i laktozy w prawie wszystkich produktach mlecznych. Możliwość pomiarów on-line najważniejszych składników produktów mlecznych przyspieszyło cały proces kontroli jakości [99-102]. W 2003 roku Kawamura wraz z współpracownikami opracował system on-line do określania zawartości trzech najważniejszych składników mleka (białka, tłuszczu oraz laktozy). W tym celu wykorzystał miernik przepływu oraz próbnik. Widma mleka zbierał w trybie transmisyjnym w zakresie 600-1050 nm. System ten umożliwił pomiar zawartość wyżej opisanych parametrów w trakcie dojenia krów z dostateczną precyzją oraz dokładnością [101]. Blazquez wykorzystał spektroskopię bliskiej podczerwieni do oznaczania zawartości wody, tłuszczu i soli nieorganicznych w serach [103]. Widma próbek serów zbierane były w trybie odbiciowym w zakresie 1100-2498 nm. Autor opracował trzy modele uwzględniając różne warunki preprocesingu: a) wykorzystując surowe widma, b) stosując transformację SNV+de-trend, c) samą transformację SNV, d) stosując kombinację SNV z de-trendem oraz z drugą pochodną. Do budowy modelu wykorzystał algorytm mPLS (zmodyfikowany algorytm PLS). W przypadku oznaczania zawartości wody oraz tłuszczu najlepsze wyniki otrzymał stosując transformację SNV z de-trendem oraz z drugą pochodną, natomiast w przypadku

określania zawartości soli najlepsze wyniki uzyskał stosując samą transformację SNV [103].

5.2.3. Owoce i warzywa

Również w przetwórstwie owoców oraz warzyw spektroskopia NIR znalazła swoje zastosowanie, głównie w systemie on-line w sortowaniu owoców i warzyw pod względem zawartości wody, cukrów, koloru czy genotypu [104]. Kawano wraz ze współpracownikami opracował model ilościowy w systemie on-line do sortowania mandarynek oraz brzoskwiń pod względem zawartości cukru w tych owocach. Ich praca później została rozszerzona przez Choi [67]. Clark wraz z pomocnikami przedstawił użyteczność tej techniki w wykrywaniu wewnętrznych zbrunatnień jabłek, dzięki czemu od razu podczas sortowania owoców można wykluczyć ładnie wyglądające owoce na zewnątrz lecz gnijące wewnątrz [105]. Inny autor, Xie, zastosował spektroskopię Vis/NIR do klasyfikacji pomidorów transgenicznych od nie transgenicznych [105]. Xie opracował 3 modele wykorzystując klasyfikator SIMCA, algorytm PCA oraz PLSR. Najlepsze wyniki dające 100% poprawną klasyfikację pomidorów uzyskał stosując algorytm PLSR.

III CEL PRACY

Na pierwszym roku studiów doktorancki została nawiązana współpraca między Katerą Technologii Leków i Biochemii a Zakładem produkującym Kosmetyki i Farmaceutyki Ziaja Ltd. W ramach tej współpracy rozpoczęto pracę nad opracowaniem nowych metod analitycznych stosując spektroskopię bliskiej podczerwieni (NIR).

Kosmetyki to grupa produktów służąca do pielęgnacji, oczyszczania, ochrony, upiększania i poprawiania kondycji ludzkiego ciała. Ze względu na ich zastosowanie przybierają one różne formy. Na rynku dostępne są w postaci, emulsji, kremów, oliwek, maści itp. Kosmetyki są chemicznymi mieszaninami substancji bardzo do siebie podobnych, co czyni ich analizę dość trudną. Kontrola jakości kosmetyków obejmuje przede wszystkim metody klasyczne, które jednak są dość czasochłonne i wymagają użycia dużej ilości reagentów.

Spektroskopia bliskiej podczerwieni wraz z chemometrią odgrywa coraz większą rolę w analizie ilościowej i jakościowej wykorzystywanej w wielu gałęziach przemysłu. Podstawowym problemem podczas opracowywania modeli ilościowych i jakościowych jest transformacja surowych widm. Proces transformacji widm jest nierozłącznym, wręcz najważniejszym etapem tworzenia modeli jakościowych i ilościowych. Surowe widma otrzymane bezpośrednio ze spektrofotometru obarczone są szumem fotometrycznym. Szumy charakterystyczne dla pomiarów NIR mogą pochodzić od wielu czynników takich jak: interferencja różnych czynników fizycznych i/lub chemicznych, niedoskonałości aparatury wynikającej ze starzeniem się sprzętu oraz wielu losowych czynników. Zbyt duży szum powoduje zaburzenia stosunku szumu do sygnału, rozdzielczości widm oraz dokładności i precyzji końcowych modeli. Inne niepożądane czynniki niekorzystnie wpływające na widma NIR często dotyczą złożonego tła pomiarowego oraz linii bazowych, które pomimo, że nie są związane ze składem, stężeniem badanej substancji posiadają duży wpływ na dane widmowe. Dotyczy to wielkości ziaren substancji badanych w wyniku czego następuje przesunięcie oraz zakrzywienie linii bazowej. Na intensywność oraz kształt widma ma również wpływ wiele czynników nieswoistych takich jak temperatura, wilgotność, grubość i własności optyczne próbki np. zdolność do rozpraszania światła, wiek próbki, a nawet zawartość CO₂ w powietrzu otaczającym próbkę. Tego typu zmienność widmowe wariacje nie tylko prowadzi do niepotrzebnie skomplikowanych modeli kalibracyjnych, ale także do uzyskania niepoprawnych wyników końcowych.

Stosowanie jednak różnych algorytmów transformujących ma zarówno swoje zalety jak i wady. Z jednej strony poprawiają one jakość i uwidaczniają pewne cechy znajdujące się w widmie, z drugiej jednak strony mogą obniżyć jakość innych istotnych właściwości tych widm, np. wygładzanie stosowane jest głównie w celu usunięcia części losowego szumu obecnego w sygnale i w poprawieniu stosunku sygnału do szumu, jednak gdy użyje się tej funkcji nieodpowiednio można zaburzyć oryginalny sygnał i obniżyć jego rozdzielczość. Pochodne mogą być użyte do uwypuklenia różnic między widmami, ale również często prowadzą do obniżenia stosunku sygnału do szumu. Ponadto stosowanie najbardziej popularnych metod takich jak wygładzanie czy pochodne wciąż mają pewne parametry które należy zoptymalizować by osiągnąć zadowalające rezultaty. Te parametry to szerokość okna pomiarowego, stopień wielomianu czy też rodzaj pochodnej. W praktyce, zazwyczaj metody transformujące z różnymi parametrami, lub łączące podstawowe metody (np. wygładzanie + pochodne, wygładzanie + MSC, etc.) mogą w różny sposób podkreślać, wzmacniać różne parametry widm i stąd prowadzić do różnych wyników końcowych modeli jakościowych czy też ilościowych. Często nawet dobrze zoptymalizowane metody obróbki wstępnej mogą powodować utratę części informacji zawartych w widmach a interesujących z punktu widzenia analizy tożsamościowej i kalibracyjnej i nie wykluczone, że mniej „optymalne” metody pre-processingu pozostawią część uzupełniających informacji widmowych. Oprócz samej wiedzy dotyczącej obróbki surowych widm bardzo ważna jest także umiejętność a co najważniejsze doświadczenie. W chwili obecnej w literaturze przedmiotu brak jest ogólnie przyjętych zasad doboru metod transformacyjnych i większość badaczy stosuje metodę prób i błędów. Jest to podejście bardzo pracochłonne. Wymaga ponadto dogłębnej znajomości zasad działania stosowanych metod transformacyjnych i dużego doświadczenia z zakresu chemometrii i spektroskopii NIR. Jest to obecnie jeden z istotnych czynników ograniczających zastosowanie tej wartościowej metody analitycznej w laboratoriach przemysłowych.

Celem mojej pracy doktorskiej jest opracowanie procedury postępowania podczas konstruowania nowych metod ilościowych oraz jakościowych tak, aby w szybki i łatwy sposób uzyskać zadowalające końcowe wyniki. Opracowanie racjonalnych podstaw doboru metod transformacji widm NIR ułatwi rozpowszechnienie się tej techniki analitycznej w laboratoriach przemysłowych i ośrodkach badawczych nie posiadających zaplecza naukowego z zakresu chemometrii i obróbki danych. Ograniczy również prawdopodobieństwo błędnego użycia metod transformacji surowych widm NIR. W tym

celu opracowałam szereg przykładowych modeli, które pokazują wpływ poszczególnych etapów analizy (matematyczna obróbka surowych widm, dobór liczby istotnych czynników) na końcowe wyniki i wskazują na co należy zwracać szczególną uwagę podczas tworzenia nowych metod ilościowych oraz jakościowych. Opracowując przykładowe modele zmieniałam różne parametry analizy, kontrolując tym samym różne źródła zmienności: a) wpływ otoczenia środowiska, b) wpływ pory roku, c) powtarzalność pomiarów jako funkcje położenia próbki na okienku pomiarowym, d) liczbę skanów.

Plan pracy doktorskiej przewidywał następujące zagadnienia:

- Opracowanie biblioteki widm w zakresie klasycznej oraz bliskiej podczerwieni
- Optymalizacja warunków zbierania widm (czas analizy-liczb skanów, rozdzielczość)
- Opracowanie procedury tworzenia modeli ilościowych i jakościowych dla surowców i produktów kosmetycznych i farmaceutycznych
- Zbadanie wpływu czynników chemiczno-fizycznych na jakość samych widm (temperatura otoczenia, powtarzalność pomiarów)
- Określenie wpływu transformacji widm surowych na efekt końcowy opracowywanych metod jakościowych
- Ocena wpływu liczby istotnych czynników na zdolności predykcyjne opracowanych modeli.
- Wykazanie możliwości stosowania spektroskopii podczerwieni (NIR oraz IR) w analizie ilościowej półproduktów i produktów kosmetycznych i farmaceutycznych

IV METODY I MATERIAŁY

4.1. Aparatura

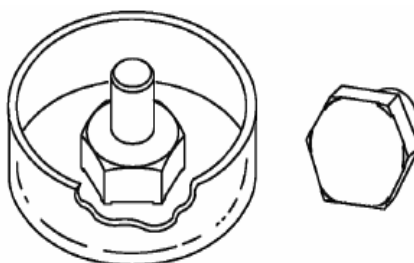
Pomiary absorpcji wykonano w przy użyciu Spektrofotometru „Spectrum 400 FT-IR/FT-NIR Spectrometer Perkin Elmer”. Do pomiarów wykorzystano spektrofotometr IR oraz przystawkę NIRa, w trybie rozproszonego światła i trasfleksacji.

Do pomiaru próbek w zakresie NIR zastosowano specjalne szklane naczynka: wiale do pomiaru absorpcji surowców stałych-proszków- oraz małe płytki Petriego do pomiaru absorpcji surowców płynnych. W przypadku ciał stałych surowiec nakładano do wial w ilości $\frac{3}{4}$ objętości wiali przy pomocy metalowej szpatułki. Po zamknięciu wiali badaną substancję ubijano uderzając delikatnie dnem wiali o blat, a następnie przecierano dno wiali papierowym ręcznikiem. Następnie wypełnioną badanym surowcem wiałę przykładano na szafirowe okienko znajdujące się na przystawce NIRa.



Rys13. Rysunek przedstawiający pomiar substancji stałych w trybie odbiciowym w zakresie spektroskopii bliskiej podczerwieni

W przypadku ciał płynnych surowiec „nakładano” do płaskiego szklanego talerza (mała płytka Petriego) a następnie nałożony surowiec dociskano aluminiowym sześciokątnym dyfuzyjnym reflektorem tak aby w przestrzeni między cieczą a reflektorem nie było żadnych bąbelków powietrza [105].



Rys.14 Szklany talerz, reflektor służące do pomiaru substancji płynnych w trybie transfleksacji w zakresie NIR

Mierząc próbki w zakresie klasycznej podczerwieni badaną substancję bezpośrednio nakładano na metalowy talerz-źródło promieniowania, za pomocą metalowej szpatułki. Między poszczególnymi pomiarami metalowy talerz oczyszczano skażonym alkoholem etylowym.

Ponadto wykorzystano następujące urządzenia:

- Waga elektroniczna
- Waga analityczna
- Ultra Wysokosprawna Chromatografia Cieczowa
- Łaźnia ultradźwiękowa

4.1.1. Szkło laboratoryjne

- Kolba miarowa
- Biureta szklana

4.1.2. Oprogramowanie

W pracy wykorzystano następujące oprogramowanie firmy PerkinElmer:

- AssureID MetodExplorer,
- AssureID Analyser
- Quant Plus

Dodatkowo wykorzystano program PCAMAX oraz Unscramble 10.1.

4.1.3. Ustawienia sprzętowe

Pomiary przeprowadzono stosując następujące parametry sprzętowe zgodnie z wymaganiami farmakopealnymi :

NIR

- liczba skanów 32
- detektor 2
- rozdzielczość 16 cm^{-1}
- szybkość skanowania $1\text{ cm}^{-1}/\text{s}$
- odległość punktów pomiarowych 2 cm^{-1}
- zakres pomiarów $10000 - 4000\text{ cm}^{-1}$
- filtr wody i dwutlenku węgla: włączony
- AVI: włączone

IR

- liczba skanów 4
- detektor 2
- rozdzielczość 4 cm^{-1}
- odległość punktów pomiarowych 1 cm^{-1}
- szybkość skanowania $0,2\text{ cm}^{-1}/\text{s}$
- zakres skanowania $4000\text{-}650\text{ cm}^{-1}$

Każdego dnia przed rozpoczęciem analizy przeprowadzano testy SST mające na celu sprawdzić poprawność działania aparatury. Jednocześnie przynajmniej raz w miesiącu przeprowadzano kalibrację sprzętu.

Testy SST sprawdzające poprawność działania aparatury:

- Noise Check – szum fotometryczny w zakresie $6000\text{-}6200\text{ cm}^{-1}$ (Kryterium akceptacji RMS (%T) max 0,0242 Pik do pik (%T) max 0,1056, Trend (%T) max 0,0088,
- Abscissa Check – dokładność długości fali dla $5669,3\text{ cm}^{-1}$ i $4571,6\text{ cm}^{-1}$ przez porównanie z widmem wbudowanego wzorca polistyrenu, (Kryterium akceptacji $5672,00\text{ – }5667,00\text{ cm}^{-1}$ $4574,00\text{ – }4569,00\text{ cm}^{-1}$),
- Throughput Check – przepustowość w stosunku do widma referencyjnego zapisanego w bibliotece tła dla $10000,00\text{ cm}^{-1}$; $6000,00\text{ cm}^{-1}$; $4000,00\text{ cm}^{-1}$ (Kryterium akceptacji (%T) min 80,00),
- Contamination Check – zanieczyszczenia w zakresie od $5300,00$ do $5000,00\text{ cm}^{-1}$ w stosunku do referencyjnego widma tła (Kryterium akceptacji Nominalna wartość (A) 0,050 Limitowa wartość (A) 0,100.

4.2. Materiały, odczynniki

W pracy zastosowano następujące odczynniki:

- Hyamine 1622
- Roztwór fenoloftaleiny
- Bromk dimidowy (brome 3,5-diamino-5-metylo-6-fenylofenantredyniowy)
- Błękit disulfiniowy (sól di-sodowa 2,4-disulfo-4,4-diaminodietylotrifenylostanu)
- Alkohol etylowy

- Kwas siarkowy stężony
- Wodorotlenek sodu
- Kwas octowy
- Metanol
- Acetonitryl
- Woda

4.3. Procedura pomiarowa

4.3.1. Tworzenie metody identyfikacyjnej

Tworzenie metod identyfikacyjnych przeprowadzono w kilku etapach. W pierwszym etapie zebrano widma wybranych surowców w zakresie klasycznej podczerwieni oraz bliskiej podczerwieni. Następnie w oparciu o wykresy *Leverage Ratio* dokonano wyboru odpowiedniego zakresu analizy. W kolejnym etapie zebrane widma podzielono na dwa zestawy danych, na zbiór uczący i zbiór testowy. W następnym kroku przeprowadzono transformacje widm ze zbioru uczącego. Po czym zastosowano analizę głównych składowych i klasyfikator SIMCA. Po stworzeniu modelu przeprowadzono walidację wewnętrzną sprawdzając poprawność działania modelu stosując zbiór testowy. Walidację przeprowadzono za pomocą oprogramowania dostarczonego przez producenta PerkinElmer ID Validation w AssureID. W kolejnym etapie przeprowadzono walidację zewnętrzną stosując program AnalyzerID na niezależnych próbkach, oraz walidację zewnętrzną.

4.3.1.1. Transformacja widm

Filtr

- Wazenie szumów
- Wazenie atmosferyczne – opcja ta umożliwia usunięcie efektu związanego z obecnością CO₂ i H₂O w widmie
- Wygładzanie - opcja ta redukuje poziom szumu w widmie oraz obniża rozdzielczość widma
- Blank - opcja ta umożliwia eliminację niechcianych cech widocznych w widmie poprzez usunięcie danego fragmentu widma

Korekcja linii bazowej

- Pochodne 1 i 2 – algorytm Stavitzky-Golay
- Offset Correction

- Slope – funkcja ta koryguje nachylenie linii bazowej nie zmieniając kształtu pików i nie zmniejsza stosunku sygnału do szumu
- Curve – koryguje zakrzywienie linii bazowej
- Resolution Weight – ważenie rozdzielczości

Normalizacja

- SNV
- MSC

4.31.2. Analiza głównych składowych

Podczas opracowywania metod jakościowych w oparciu o oprogramowanie dostarczone przez producenta sprzętu po wykonaniu transformacji widm przeprowadzono analizę głównych składowych. Początkowo liczbę istotnych głównych składowych wybierano na podstawie kryterium zmienności ustalonym przez producenta. Jednak ze względu na to, że zaobserwowano przeuczenie modeli dobór liczby istotnych czynników dokonano manualnie w oparciu o wykresy *Residual Ratio* oraz *Model Distance Ratio*.

4.3.1.3. Klasyfikacja

Klasyfikację przeprowadzono stosując klasyfikator SIMCA. Wyniki uzyskane przedstawiono graficznie oraz tabelarycznie. Kryterium poprawności klasyfikacji oceniono na podstawie parametru „total distance ratio”. Jest to stosunek wariancji resztowej widma do wartości wariancji resztowej widma średniego danej klasy. Wartość tego parametru została ustalona przez producenta oprogramowania i mieści się w granicach od 0 do 1.

4.3.1.4. Walidacja

Zanim opracowana metoda zostanie dopuszczona do rutynowych analiz przeprowadzana jest tzw. walidacja zewnętrzna w oparciu o zbiór testowy. Zbiór testowy zawiera próbki niezależne, nie użyte w zbiorze kalibracyjnym oraz walidacyjnym. Walidacja zewnętrzna miała na celu sprawdzić specyficzność, precyzję oraz odporność metody zgodnie z wytycznymi farmakopealnymi.

- Odporności modelu na niewielkie zmiany – do oceny odporności modelu zbadano wpływ temperatury na jakość klasyfikacji modelu. Próbki ze zbioru testowego rozdzielono na dwie grupy. Jedną próbkę wstawiono do lodówki a druga do ciepłarki na okres 1 godziny. Po upływie tego czasu przeprowadzono analizę. Dodatkowo zbadano wpływ wymiany lampy oraz osuszaczy na opracowany model

- Specyficzność modelu - do oceny specyficzności modelu poddano analizie surowce inne niż znajdujące się w opracowanym modelu. Surowce te charakteryzowały się podobieństwem postaci oraz składu chemicznego surowców znajdujących się w modelu
- Powtarzalność badano mierząc badaną substancję kilkakrotnie

Kryteriami akceptacji poprawności działania opracowanych modeli były następujące parametry

- Odległość danej próbki od środka klasy („*Specified material distance ratio*”) - wartość nie przekraczająca 1
- Identyfikacja badanego materiału– wynik „*pass*” lub nazwa identyfikowanego materiału w przypadku substancji znajdującej się w opracowanym modelu
- Brak potwierdzenia tożsamości próbki nie znajdującej się w opracowanym modelu – wynik negatywny „*fail*” lub „*unidentified*”
- 100% poprawnej klasyfikacji
- 100% stopień odrzucenia

4.4. Tworzenie metod ilościowych

Opracowanie metody ilościowej przeprowadzono również w kilku etapach. W pierwszym etapie zebrane próbki poddano analizie w celu uzyskania wartości referencyjnej badanego parametru, po czym zebrano ich widma w zakresie od 10000-4000 cm^{-1} -NIR oraz 4000-650 cm^{-1} IR. W kolejnym etapie dokonano wyboru algorytmu i przeprowadzono transformację widm. Po czym rozpoczęto optymalizację i dobór liczby czynników. W końcowym etapie analizy przeprowadzono walidację w celu sprawdzenia poprawności działania opracowanego modelu wykorzystując walidację typu „ wyrzuc jeden” (*Cross-Validation*), oraz walidację niezależną stosując niezależny zbiór danych.

Modele ilościowe opracowano dla następujących parametrów: w dwóch produktach firmy Ziaja Ltd.:

- Oznaczanie zawartości Ketorprofenu w żelu Ketorprofen,
- Oznaczanie zawartości filtrów ESCALOL 557 w emulsjach do opalania.

oraz w półprodukcie gdzie oznaczano zawartość substancji anionowo czynnych w roztworze TEXAPON.

Modele ilościowe nazwano odpowiednio : Model K- do oznaczania zawartości Ketoprofenu, Model E – do oznaczania filtrów oraz Model T - do oznaczania substancji anionowo czynnych.

4.4.1. Wyznaczenie wartości referencyjnej badanego parametru

Wartość parametru badanego została oznaczona za pomocą metody referencyjnej. Poniżej przedstawiono sposób oznaczania wartości referencyjnych w poszczególnych modelach.

Model K

Metodą referencyjną na oznaczanie zawartości substancji aktywnej (ketoprofen) w żelu Ketoprofen była Ultra Wysokosprawna Chromatografia Cieczowa (*Ultra High Performance Liquid Chromatograph*).

Procedura analizy miała następujący przebieg: Około 1g produktu (żelu) umieszczano w 25 ml kolbie pomiarowej i dodawano 5 ml metanolu. Wymieszano delikatnie w celu uzyskania zawiesiny, uzupełniono metanolem i wstawiono do łaźni ultradźwiękowej na czas 15 minut. Do dalszej analizy wykorzystano supernatant. Zawartość ketoprofenu w produkcie farmaceutyczny wyznaczano według następującego wzoru:

$$x = 0,2 \cdot \frac{A_p \cdot m_s}{A_s \cdot m_p} \left[\frac{mg}{g} \right] \quad (37)$$

gdzie:

m_s - średnia wartość [mg] z dwóch prób wzorca

m_p – masa badanego produktu leczniczego [g]

A_p – powierzchnia piku w chromatogramie pochodzącego od leku

A_s – powierzchnia piku w chromatogramie pochodzącego od wzorca

Model E

W przypadku modelu ilościowego Model E, zawartość filtrów w emulsjach do opalania nie oznaczano metodą referencyjną. Wartość tego parametru uznano za zgodną zgodnie z recepturą opracowaną w zakładzie.

Model T

Zawartość substancji anionowo czynnych oznaczano według instrukcji znajdującej się w firmie Ziaja Ltd. Procedura analizy miała następujący przebieg:

W zlewce odważono 0,5 g surowca i rozpuszczono w wodzie, następnie dodano 5 kropli 1% roztworu fenoloftaleiny. Otrzymany roztwór przeniesiono ilościowo do kolby miarowej o pojemności 250 ml i uzupełniono wodą do kreski. Odmierzono następnie pipetą 10 ml otrzymanego roztworu, dodano 10 ml wody, 15 ml chloroformu, 10 ml wskaźnika mieszanego i miareczkowano roztworem Hyamine 1622. Punkt końcowy ustalono w momencie całkowitego zniknięcia różowego zabarwienia dolnej warstwy chloroformowej i pojawienie się niebieskiego zabarwienia szaroniebieskiego.

Wartość substancji anionowo czynnej X (%) w próbce obliczono wg wzoru:

$$X = \frac{2,5 \cdot C \cdot V \cdot 380}{m} \quad (38)$$

gdzie:

m – masa próbki [g]

C – stężenie molowe roztworu Hyamine 1622 [mol/l]

V – objętość roztworu Hyamine 1622 zużytego do miareczkowania [ml]

4.4.2. Skład badawczy serii kalibracyjnych i walidacyjnych

Model K

W zbiorze kalibracyjnym znalazły się próby laboratoryjne, przygotowane w taki sposób aby zawartość substancji oznaczanej była zarówno poniżej jak i powyżej wartości deklarowanej (50-150%). Ponadto do zbioru kalibracyjnego dodano próby produkcyjne.

Model E

W przypadku modelu E na oznaczanie filtrów w emulsjach do opalania, w zbiorze uczącym znalazły się próby zarówno laboratoryjne jak i produkcyjne otrzymane bezpośrednio z produkcji. W **tab.1** zestawiono zakres wartości oznaczanego filtru w próbach laboratoryjnych

Tabela 1: Zakres wartości filtra Escalol 557 użytych do budowy modelu ilościowego

Nazwa serii	Zawartość Escalolu
spf 3	1,25%
spf 6	2,5%
spf 8	3,5%
spf 10	4,5%
spf 12	5,25%
spf 15	6%
spf 20	6%

Model T

W modelu T podobnie jak w przypadku poprzednich modelach w zbiorze kalibracyjnym znalazły się próby zarówno poniżej jak i powyżej wartości oznaczanej. Zakres stężeń oznaczanego parametru przedstawiono w zakresie od 80 - 120%.

4.4.3. Wybór metody regresji

W pracy wykorzystano następujące algorytmy:

- PLS1
- PLS2
- PCR

4.4.4. Transformacja widm

Korekcja linii bazowej

- Offset correction
- Pochodna 1 i 2 (Stavitzky-Golay)

Normalizacja

- brak
- SNV+trend

Ponadto przeprowadzono wygładzanie widm, centrowanie oraz autoskalowanie.

4.4.5. Optymalizacja i kompleksowość modeli

Wybór liczby istotnych czynników dokonano w oparciu o wykres zależności wartości własnej od liczby czynników, wskaźnik dźwigni (*Leverage ratio*), oraz wskaźnik reszt (*Residual ratio*).

4.4. 6. Ocena jakości skonstruowanych modeli kalibracyjnych

Do oceny jakości skonstruowanych modeli posłużyły wykresy zależności wartości przewidzianej względem wartości referencyjnej oraz wartości SEC, SEP, R^2 , RMSEP.

4.4.7 Walidacja modeli ilościowych

Kontrola zestawu kalibracyjnego

Wstępna analizę modelu przeprowadzono na zestawie próbek kalibracyjnych sprawdzając [106-114]:

- Współczynnik korelacji $R \geq 0.95$
- Dokładność metody analitycznej, która wyraża bliskość zgodności pomiędzy wartością uzyskaną metodą referencyjną a wartością oznaczoną metodą NIR. W tym celu obliczono standardowy błąd kalibracji SEC oraz średnią wartość reszkowych oznaczonego składnika (ang. Bias)

$$SEC = \sqrt{\frac{RSS}{n-p}} \quad RSS = \sum_{i=1}^n (y_i - Y_i)^2 \quad (39)$$

$$Bias = \frac{\sum_{i=1}^n (y_i - Y_i)}{n} \quad (40)$$

gdzie y jest „prawdziwą” zawartością ketoprofenu w żelu oznaczoną metodą referencyjną, Y jest zawartością wyznaczoną metodą NIR, n liczbą próbek a p liczbą istotnych czynników.

Dokładność metody została również zbadana za pomocą sparowanego t-testu

$$t = \frac{\left| \bar{d} \right|}{\frac{S_d}{\sqrt{n}}} \quad (41)$$

gdzie $[d]$ wartość bezwzględna średniej wartości resztowej pomiędzy wynikami NIR a wynikami referencyjnymi, S_d - odchylenie standardowe wartości reszkowych n wyników. Jeśli $t_{obl.} > t_{kryt.}$. Dla wybranego poziomu prawdopodobieństwa ($P=95\%$) to odrzucamy hipotezę zerową, że dokładność obu metod jest

porównywalna. Jeśli $t_{obl.} < t_{kryt}$ to nie ma podstaw do odrzucenia tej hipotezy, zatem obie metody nie różnią się pod względem dokładności

- Liniowość - procedury analitycznej jest jej zdolnością do otrzymania wyników, które są wprost proporcjonalne do stężenia (ilości) analitu w próbce.

Liniowość opracowanego modelu została obliczona na podstawie równania liniowej zależności:

$$Y = ax + b \quad (42)$$

Gdzie y jest „prawdziwą” zawartością, Y przewidywaną zawartością, a współczynnikiem kierunkowym (nachyleniem) prostej, b przesunięciem prostej. Równanie to zostało zweryfikowane statystycznie wykorzystaniem poniższych zależności:

$$b \pm t \cdot \frac{SD_r}{\sqrt{S_{xx}}} \quad a \pm t \cdot SD_r \cdot \sqrt{\frac{1}{n} + \frac{x^2}{S_{xx}}} \quad (43)$$

$$SD_r = \sqrt{\frac{RSS}{n-2}} \quad S_{xx} = \sum x^2 - \frac{(\sum x)^2}{n} \quad (44)$$

gdzie t , jest współczynnikiem Studenta przy poziomie prawdopodobieństwa 95%; $n-2$ stopniami swobody; SD_r jest reszkowym odchyleniem standardowym a S_{xx} jest sumą kwadratów; n ilością próbek. Przedział ufności dla współczynnika kierunkowego powinien obejmować 0 a dla przesunięcia prostej 1.

Kontrola zestawu walidacyjnego

Po kontroli zestawu kalibracyjnego przeprowadzono kontrolę zestawu walidacyjnego sprawdzając następujące parametry.

- Współczynnik korelacji $r \geq 0.95$
- Dokładność, obliczając standardowy błąd predykcji SEP, średnią wartość reszkowych oznaczonego składnika Bias oraz wartość t studenta
- Liniowość jak w przypadku kontroli zestawu kalibracyjnego
- Błąd standardowy oznaczonych wartości SEP musi być porównywalny z SEC a obliczona wartość SEP/SEC powinna mieścić się w zakresie 0,5-1,5

Kontrola zestawu testowego

Po skontrolowaniu zestawu kalibracyjnego i walidacyjnego przystąpiono do kontroli zestawu testowego. Kontrolę przeprowadzono sprawdzając następujące parametry:

- Dokładność metody zbioru testowego definiowaną jako różnicę między wartością oczekiwaną a prawdziwą zawartością oznaczonego składnika

$$\Delta x_{\text{met}} = E_x - \mu_x \quad (45)$$

Wartość tej różnicy stanowi jedynie błąd systematyczny tej metody. Miarą dokładności metody jest więc wielkość jej błędu.

- Odzysk, procentowy odzysk wartości oznaczonych metoda NIR w odniesieniu do wartości oznaczonych metoda referencyjną obliczono według poniższego wzoru

$$\text{Odzysk} = \frac{\text{NIR}}{\text{Ref}} \cdot 100\% \quad (46)$$

Wartość odzysku przypadku metody na oznaczanie zawartości ketoprofenu w żelu nie powinna być większa niż $100 \pm 3\%$ w przypadku pozostałych modeli $100 \pm$

- Precyzję jako powtarzalność sprawdzano przez 6-krotny pomiar zawartości badanego parametru w próbce w tym samym dniu przez tego samego analityka według wzoru:

$$\text{RSD} = \left(\frac{\text{SD}}{\text{średnia}} \right) \cdot 100\% \quad (47)$$

Wyznaczona wartość RSD z 6 pomiarów nie powinna być większa niż 0,85%.

- Precyzję pośrednią określono za pomocą 6-krotnego pomiaru zawartości parametru badanego w danej serii produktu przez 6 kolejnych dni, przez różnych analityków. Wartość RSD z 6 pomiarów nie powinna być większa niż 2%.
- Specyficzność – jako zdolność jednoznacznego oznaczania analitu w obecności innych składników, które mogą znajdować się w próbce, oznaczono poprzez wykazanie charakterystycznych pasm w widmach oznaczanego składnika
- Miary statystyczne: *M-distance ratio* oraz *Residual ratio*

M-distance opisuje stosunek między odległością Mahalanobiasa, dla próbki i wszystkich współczynników w modelu, oraz wartością graniczną dla tych współczynników. *Residual Ratio* w programie Quant + opisuje stosunek resztkowej wariancji próbki do średniej resztkowej wariancji wzorców.

W sytuacji gdy:

- M-distance ratio >1 a residual ratio >3 wówczas oznacza to, że występują dodatkowe czynniki obecne w widmie w zakresie oznaczanego składnika, parametru
- M-distance ratio >1 a residual ratio <3 któryś z oznaczonych składników występuje z zwiększoną intensywnością niż w modelu kalibracyjnym
- M-distance ratio <1 a residual ratio >3 w zbiorze modelowym nie zostały uwzględnione cechy obecne a próbce badanej
- M-distance ratio <1 i residual ratio <3 badane widmo jest reprezentatywne dla zestawu widm zbioru kalibracyjnego

4.5. Analiza chemometryczna danych

W celu zbadania wpływu czynników chemiczno-fizycznych na jakość samych widm oraz określenie wpływu transformacji widm surowych na efekt końcowy opracowywanych metod jakościowych wykorzystano 6 przykładowych substancji

- Benzoesan sodu
- Bronopol
- Ekstrakt glikolowy z ogórka
- Ac.Net
- Alkohol lanolinowy
- Alkohol cetearylowy

Widma powyższych substancji zbierano w różnych warunkach uwzględniając wpływ otoczenia, środowiska, ilość skanów (widma zbierano w 9 różnych warunkach skanowania : odpowiednio 1, 3, 5, 10, 15, 23, 32, 45 i 62 skanów), powtarzalność pomiarów. Po zebraniu widm surowców przeprowadzono analizę chemometryczną i opracowano modele jakościowe.

Do przeprowadzenia analizy chemometrycznej wykorzystano program PCAMAX oraz program Unscramble 10.1. Zadaniem analizy głównych składowych (PCA) jest przedstawienie danych X (o m obiektach i n zmiennych) jako iloczyn dwóch nowych macierzy wyników T oraz macierzy wag P . Kolumny T (ortogonalne) i P (ortonormalne) zawierają współrzędne obiektów i parametrów na nowych ukrytych zmiennych zwanych czynnikami głównymi. Każdy z głównych składowych jest liniową kombinacją zmiennych i jest ortogonalna w stosunku do wszystkich pozostałych głównych składowych. Pierwsza główna składowa wyjaśnia największą zmienność danych i zawiera najwięcej informacji. Druga składowa jest ortogonalna (prostopadła) do pierwszej i reprezentuje największą

zmienność wokół pierwszej głównej składowej. Dane spektralne po zastosowaniu analizy głównych składowych reprezentowane są przez zbiór składowych [41-42]. Ilość takich nowych zmiennych mówi o kompleksowości zestawu danych. Analiza głównych składowych umożliwia:

- a) zaobserwowanie podobieństwa próbek w przestrzeni mierzonych parametrów
- b) zbadanie zależności między próbkami
- c) zbadanie różnic między próbkami
- d) zbadanie kompleksowości badanego układu

Na przebieg analizy głównych składowych istotny wpływ ma wstępne przygotowanie danych. Transformacja taka ma na celu podnieść jakość danych oraz poprawić interpretację danych. Metody przygotowania danych do dalszej analizy można podzielić na trzy grupy:

- a) pierwsza z nich obejmuje metody stosowane do indywidualnych obiektów macierzy danych (metody eliminacji szumu i linii bazowej różnego rodzaju procedury normalizacyjne, pochodne)
- b) druga grupa to metody, których zadaniem jest modyfikacja indywidualnych zmiennych (metody centrowania do usunięcia efektu związanego z przesunięciem linii bazowej, autoskalowania – w celu usunięcia efektu rozpraszania oraz przesunięcia linii bazowej, standaryzacja)
- c) trzecia grupa to metody stosowane do nakładania sygnałów instrumentalnych.

Wybór odpowiedniej metody wstępnego przygotowania danych nie jest oczywisty i wywiera wpływ na konstrukcję czynników głównych oraz na ich późniejszą interpretację [115].

Program PCAMAX wykorzystano do zbadania podobieństwa próbek zebranych w różnych warunkach. W tym celu za pomocą powyższego programu wyznaczono liczbę istotnych głównych składowych. Dane uzyskane z programu PCAMAX automatycznie transformowane były za pomocą metody autoskalowania. Program Unscramble 10.0 wykorzystano do modyfikacji surowych danych różnymi metodami preprocessingu w celu sprawdzenia wpływu transformacji danych na konstrukcję czynników głównych.

V WYNIKI I DYSKUSJA

5.1. Analiza chemometryczna

5.1.1. Ocena zmienności warunków pomiarowych

W celu oceny zmienności wynikającej z warunków pomiarowych przeprowadzono badania mające na celu sprawdzenie:

- a) Wpływu ilości skanów na jakość końcowych modeli
- b) Powtarzalności pomiarów
- c) Wpływu położenia próbki na oknie pomiarowym
- d) Wpływu różnych warunkach środowiskowych
- e) Wpływu ilości głównych składowych na zdolności predykcyjne opracowanych modeli
- f) Wpływu punktów odbiegających na konstrukcje czynników głównych

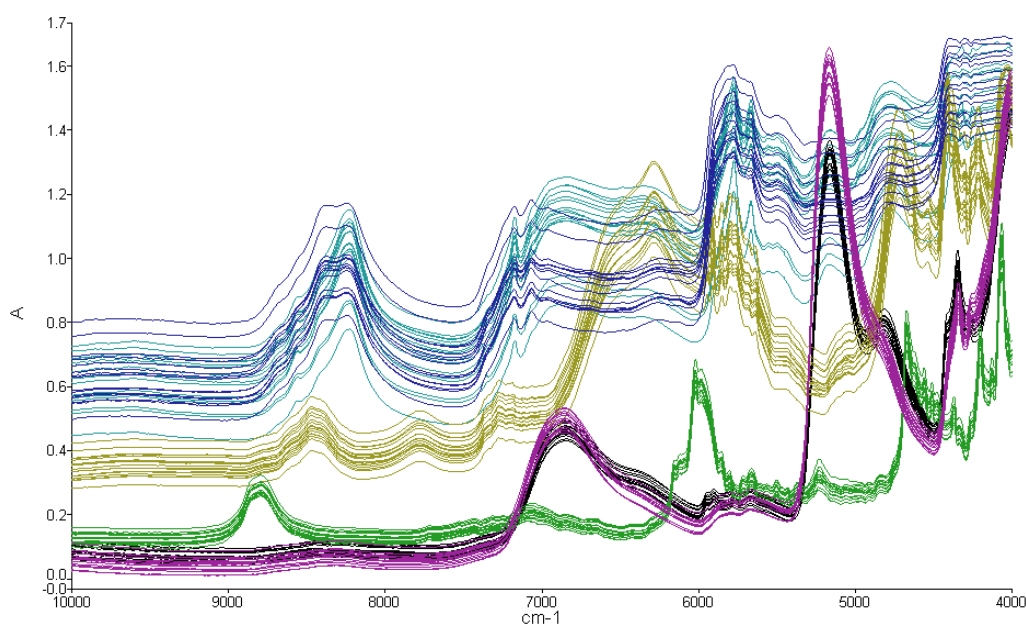
Do oceny wpływu powyższych warunków wykorzystano 6 przykładowych surowców stosowanych w firmie Ziaja:

- benzoesan sodu
- glikol butylenowy
- ekstrakt z ogórka
- alkohol lanolinowy
- alkohol cetearylowy
- 2-bromo-2-nitropropano-1,3-diol

5.1.1.1. Ocena wpływu ilości skanów na jakość opracowanego modelu jakościowego

W związku z tym że w zakładach produkcyjnych bardzo ważny jest czas analizy opracowano 9 modeli w celu sprawdzenia wpływu ilości zarejestrowanych skanów próbki na jakość opracowanego modelu.

Widma zbierano w 9 różnych trybach skanowania odpowiednio dla: 1, 3, 5, 10, 15, 23, 32, 45, 62 skanów. Na **Rys.15** przedstawione zostały surowe widma analizowanych surowców uzyskane przy różnej liczbie skanów.



Rys.15 Surowe widma surowców zarejestrowane przy różnej liczbie skanów: zielony - benzoesan sodu, żółty – 2-bromo-2-nitropropano-1,3-diol, czarny - ekstrakt z ogórka, różowy - glikol butylenowy, niebieski - alkohol lanolinowy, turkusowy - alkohol cetearylowy

Tabela2. Warunki rejestracji widm

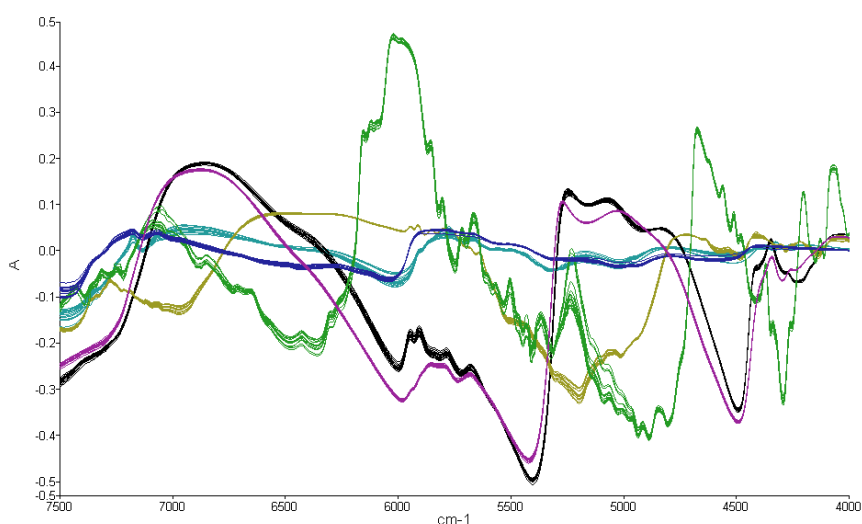
Warunki pomiarów	Model 1-9
liczba skanów	1,3,5,10,15,23,32,45,62
detektor	2
rozdzielczość cm^{-1}	16
szybkość skanowania cm/s	1
Odległość punktów pomiarowych cm^{-1}	2
zakres pomiarów cm^{-1}	10000-4000
filtry	włączone

Jak widać z **Tab.2** warunki pomiarowe między wszystkimi modelami różniły się jedynie ilością skanów. Wszystkie pozostałe parametry były identyczne.

Po zebraniu widm surowców w powyższych warunkach pomiarowych dokonano podziału widm na dwa zbiory: zbiór uczący i walidacyjny. Oba zbiory we wszystkich modelach zawierały po tyle samo próbek każdego z surowców: w zbiorze uczącym użyto 18 widm benzoesu sodu, 16 alkoholu lanolinowego, 18 2-bromo-2-nitropropano-1,3-diolu, 16 glikolu butylenowego, 17 alkoholu cetearylowego, 18 widm ekstraktu z ogórka. W zbiorze walidacyjnym wykorzystano łącznie 21 widm. Widma ze zbioru uczącego poddano transformacjom. Poniżej przedstawiono rodzaj zastosowanej transformacji.

Tabela.3 Typ zastosowanej transformacji

	<i>Model1,2,3</i>
Zakres analizy (cm^{-1})	7500-4000
Filtr	ważenie szumów
Korekcja linii bazowej	nachylenie-curve
Normalizacja	SNV



Rys.16 Zestawy widm 6 surowców uzyskane po transformacji. Kolory jak na Rys. 15

Po wybraniu odpowiednich algorytmów transformujących przeprowadzono klasyfikację widm surowców za pomocą klasyfikatora SIMCA. Następnie przeprowadzono proces sprawdzania poprawności działania powyższych modeli. Parametrami branymi pod uwagę były:

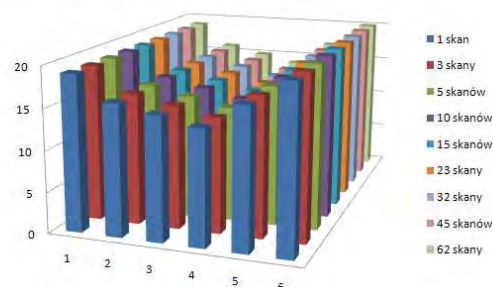
- wskaźniki programowe: „Recognition rate” (wskaźnik rozpoznania) oraz „Rejection rate” (wskaźnik odrzucenia) wartość poprawna 100% oraz Specified Material Total Distance Ratio (Odległość danej próbki od środka klasy) w granicach 0-1 (wartości wygenerowane przez program)
- Wynik tożsamości każdej próbki. Potwierdzenie tożsamości (Pass)
- Nazwa zidentyfikowanego materiału
- Walidacja programowa (wewnętrzna) – parametry „% Classification Rate” (wskaźnik klasyfikacji) oraz „Rejection Rate” (wskaźnik odrzucenia) wartość poprawna 100

Tabela 4. Wyniki walidacji próbek ze zbioru uczącego i testowego: liczba próbek poprawnie zidentyfikowanych

Surowiec	Liczba próbek		Liczba próbek poprawnie zaklasyfikowanych we wszystkich modelach	
	Z. uczący	Z. testowy	Z. uczący	Z. testowy
alkohol lanolinowy	16	3	16	3
alkohol cetearylowy	17	3	17	3
Glikol butylenowy	15	3	15	3
ekstrakt z ogórka	14	3	14	3
2-brom0-2nitropropano-1,3-diol	19	3	19	3
benozesan sodu	19	4	19	4

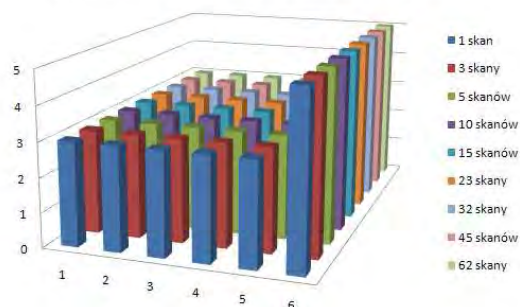
a)

Liczba poprawnie sklasyfikowanych surowców zbioru uczący



b)

Liczba poprawnie sklasyfikowanych surowców zbioru testowy



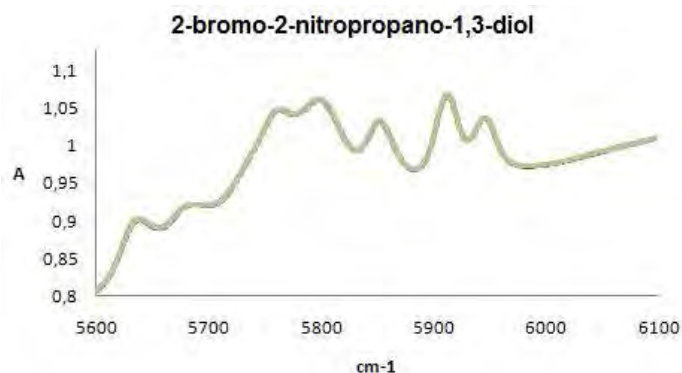
Rys.17 Liczba poprawnie sklasyfikowanych surowców a) zbioru uczący, b) zbioru testowy

Z uzyskanych wyników można stwierdzić, że ilość skanów nie ma istotnego wpływu na jakość opracowywanych modeli. Można więc opracowywać modele jakościowe zmniejszając ilość skanów, a tym samym skracając czas analizy bez utraty jakości analizowanych widm.

Widma surowców zebrane w 9 trybach pomiarowych (różne liczby skanów) poddano również analizie chemometrycznej w celu sprawdzenia podobieństwa kształtu widm zebranych przy różnych liczbach skanów, a także oceny wpływu ilości skanów próbki na jakość otrzymanych modeli. W tym celu przeprowadzono analizę głównych składowych wykorzystując program PCAMAX. Za pomocą tego programu wyznaczono liczbę istotnych głównych składowych w oparciu o a) % informacji wyjaśnianej przez PC b) widma resztowe. Poniżej przedstawiono jedynie przykładowe wyniki uzyskane dla trzech wybranych surowców. Wyniki uzyskane dla pozostałych trzech badanych surowców były analogiczne.

2-bromo-2-nitropropano-1,3-diol

Wpływ ilości skanów zbadano najpierw na przykładzie widm uzyskanych dla jednej próbki przy 9 różnych liczbach skanów. Do tego celu wybrano 2-bromo-2-nitropropano-1,3-diol (substancja stała). **Rys. 18** przedstawia uzyskane widma.



Rys.18 Surowe widm jednej próbki uzyskane dla 9 różnych liczb skanów (tab.1) w zakresie 5500-6100 cm^{-1}

Widma analizowano w zakresie 5500-6100 cm^{-1} (Zakres analizy wybrano na podstawie parametru Variable Leverage). Po wprowadzeniu danych do programu PCAMAX wykonano wykresy widm resztowych (**Rys.19**), a następnie wyznaczono liczbę istotnych głównych składowych.

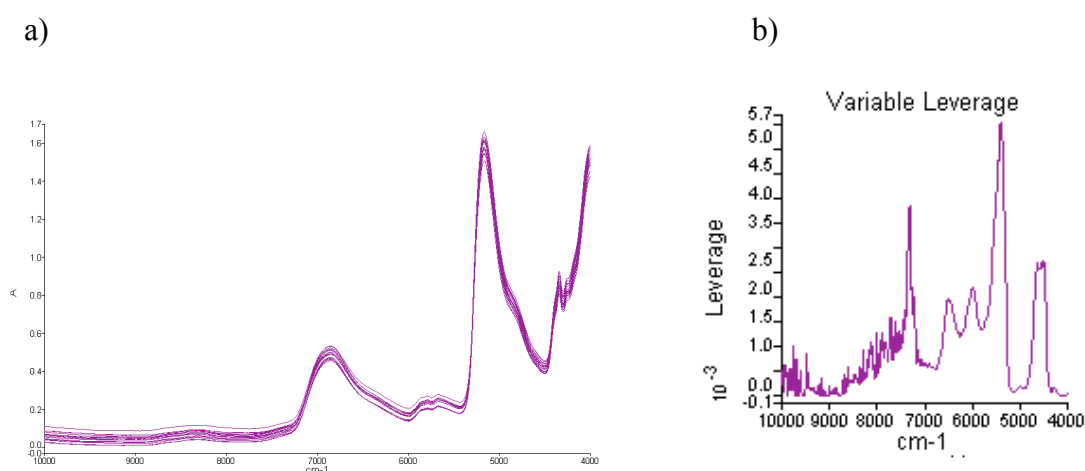


Rys.19 Widma resztowe zestawu widm 2-bromo-2-nitropropano-1,3-diolu z Rys.18

Widma resztowe I-, II- i III-rzędu nie posiadają punktów przecięcia. W widmach tych przeważa obecność szumu losowego o zróżnicowanej częstotliwości. Zatem, na tej podstawie można uznać, że tylko jedna główna składowa jest istotna.

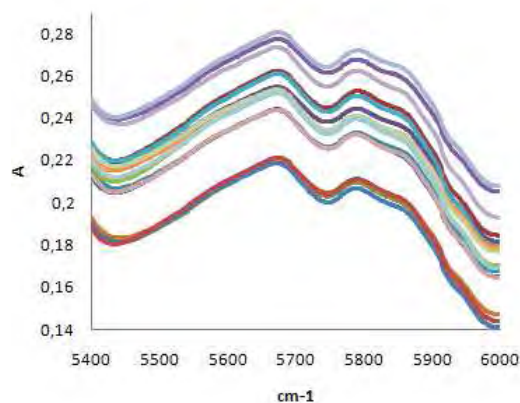
Glikol butylenowy

Analiza podobieństwa kształtu widm glikolu butylenowego (surowiec ciekły) odbyła się w zakresie $6000\text{-}5400\text{ cm}^{-1}$, na przykładzie 19 widm zarejestrowanych jako pojedyncze skany. Zakres analizy został wybrany na podstawie parametru: "Variable Leverage" (sposób doboru zakresu analizy w oparciu o ten parametr zostanie opisany szerzej w rozdziale 6), oraz w oparciu o widmo badanego surowca, **Rys.20**.



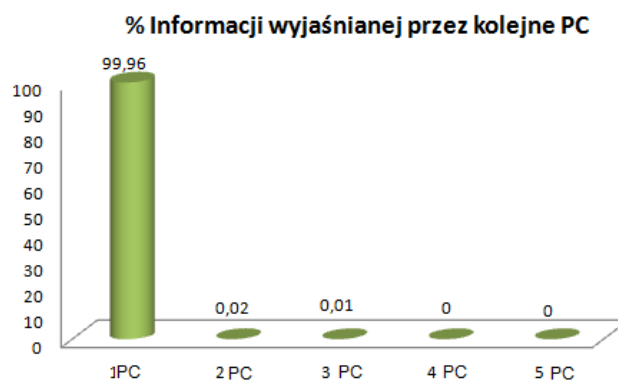
Rys. 20 Zestaw widm surowych glikolu butylenowego zarejestrowanych przy pojedynczym skanowaniu: **a)** surowe widma, **b)** wykres „Variable Leverage” dla tego zestawu widm

Analizując wykres Variable Leverage obserwuje się, że interesujące własności badanego surowca znajdują się w zakresie od $7500\text{-}4000\text{ cm}^{-1}$. W tym zakresie obserwuje się wyraźne piki. Dodatkowo analizując widma surowe glikolu można stwierdzić, że interesującym zakresem do dalszej analizy jest zakres od $6000\text{-}5400\text{ cm}^{-1}$. Zatem na tej podstawie do dalszej analizy wybrano zakres $6000\text{-}5400\text{ cm}^{-1}$, **Rys.21**.



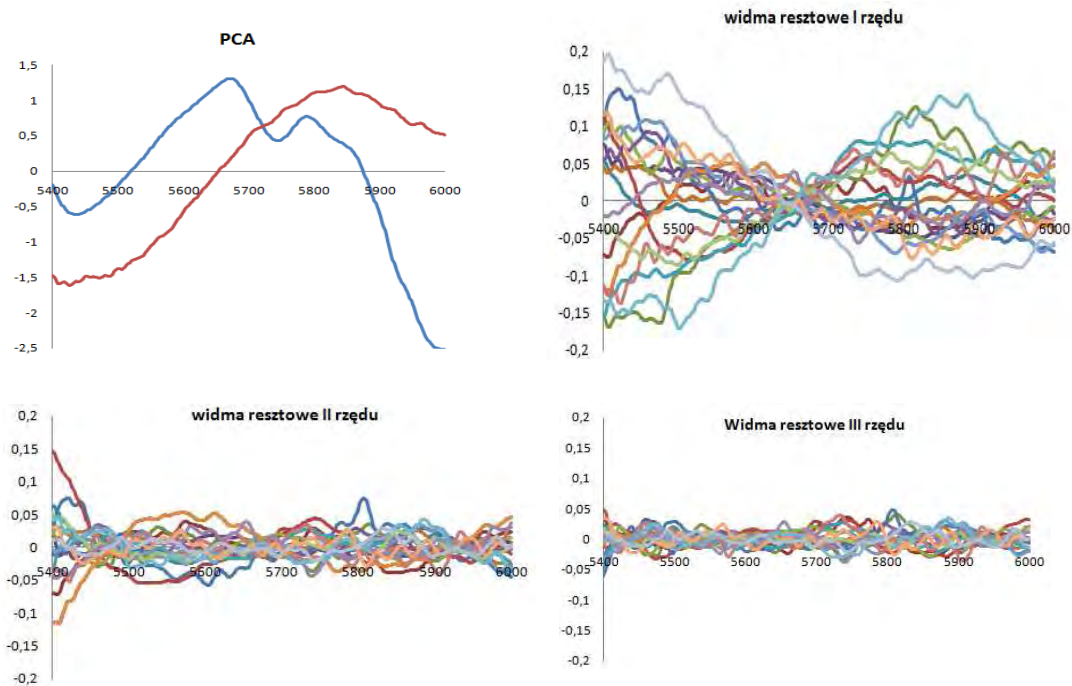
Rys.21 Widma surowe dla glikolu butylenowego w zakresie $5400-6000\text{ cm}^{-1}$ dla pojedynczych skanów

Za pomocą programu PCAMAX wyznaczono wartości własne składowych oraz otrzymano widma resztowe badanego surowca. Liczbę istotnych głównych składowych (PC) można wyznaczać w oparciu o wartości własne lub w oparciu o procent wariacji danych lub za pomocą widm resztowych. **Rys.22** przedstawia wykres % wariacji wyjaśnionej przez pierwsze 5 składowych dla widm zebranych przy 1 skanie



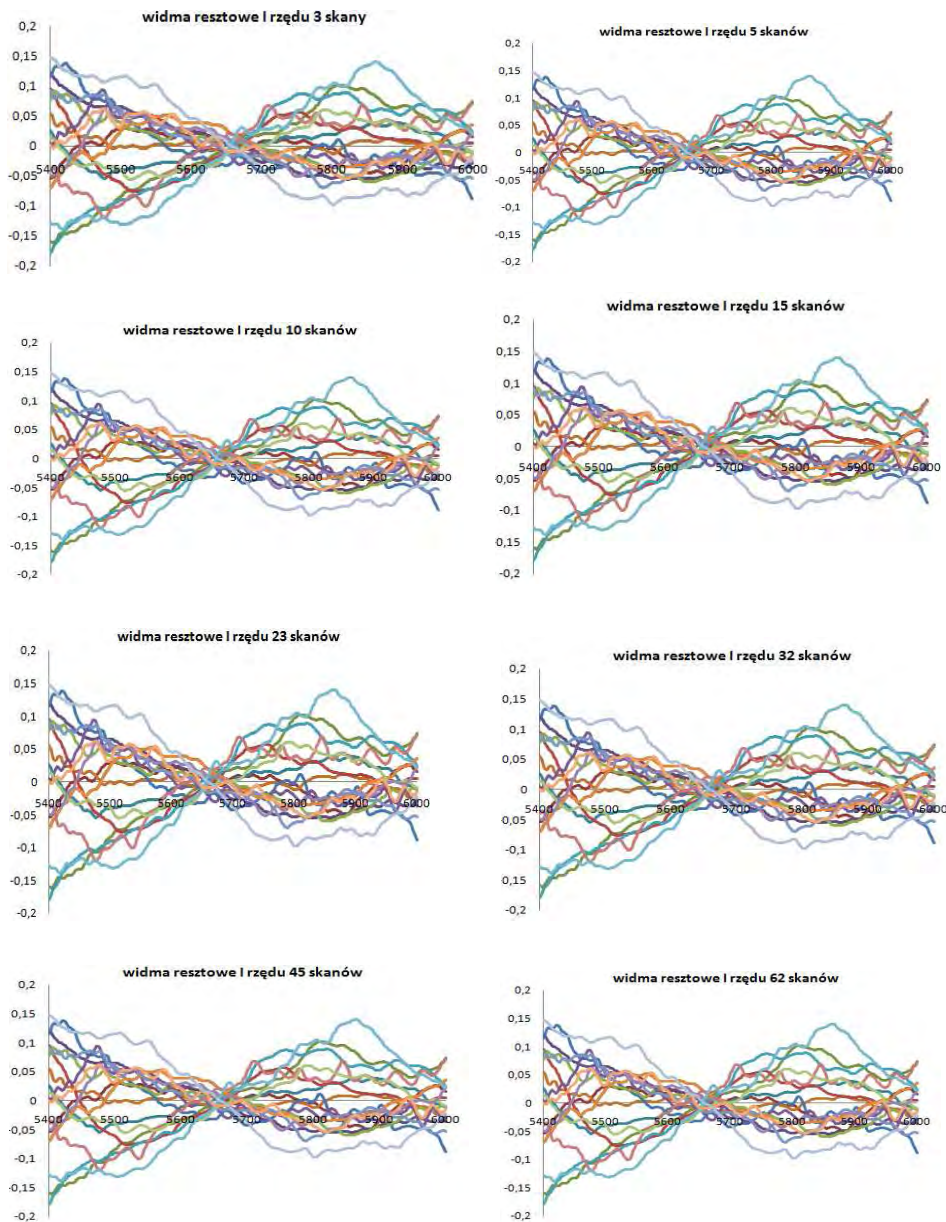
Rys.22 Sumaryczny procent zmienności wyjaśniane przez kolejne główne składowe

Analizując wykres na **Rys.22** można wywnioskować że tylko jedna główna składowa jest istotna. Wyjaśnia ona bowiem aż 99,96% zmienności danych. Jednakże analizując widmo resztowe I rzędu, rys.23, obserwuje się punkt przecięcia wszystkich widm co sugeruje istotność dwóch składowych. Zatem na tej podstawie można stwierdzić, że w analizowanym zestawie danych do wyjaśnienia wariacji systematycznej potrzebne są dwie składowe.



Rys.23 Widma dwóch pierwszych składowych oraz widma resztowe dla zestawu widm z Rys.21

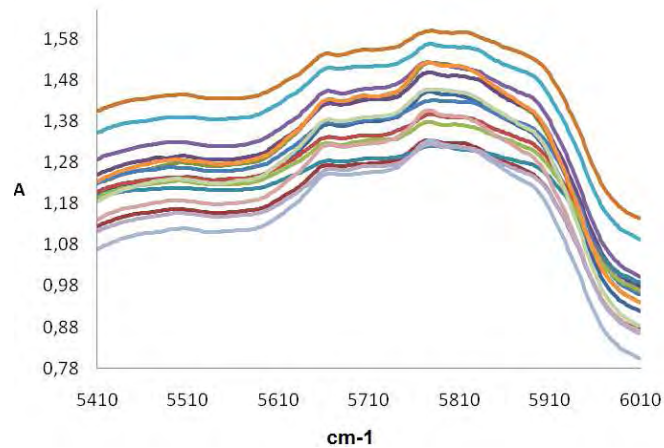
Analizując zestawy widm uzyskanych dla glikolu butylenowego zarejestrowanych przy wzrastającej liczbie skanów (3,5,10,15,23,32,45,62) obserwuje się podobną sytuację. Analizując procentowy udział zmienności danych w kolejnych głównych składowych uzyskuje się sugestię, że istotna jest jedna główna składowa. Jednakże podczas analizy widm resztowych, **rys.24**, okazuje się, że w widmie resztowym I rzędu widoczny jest wyraźny punkt przecięcia, a poszczególne widma mają charakter widm różnicowych. W widmach resztowych wyższych rzędów przeważa nasilający się szumu losowy. Tym samym dla pełnego opisu zmienności systematycznej zestawów widm glikolu butylenowego uzyskanych przy różnej liczbie skanów niezbędne są dwie główne składowe.



Rys.24 Zestaw widm resztowych uzyskanych dla glikol butylenowego przy różnej liczbie skanów

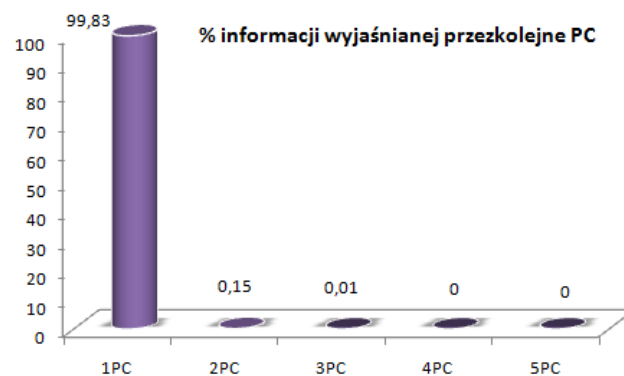
Alkohol lanolinowy

Kolejnym analizowanym surowcem był alkohol lanolinowy (substancja stała). Widma NIR analizowano w zakresie od $5410\text{--}6010\text{ cm}^{-1}$. Zakres analizy został wybrany również na podstawie parametru „Variable Leverage” oraz w oparciu o surowe widmo badanego surowca. **Rys.25** przedstawia surowe widma tego surowca zarejestrowane przy pojedynczym skanowaniu.



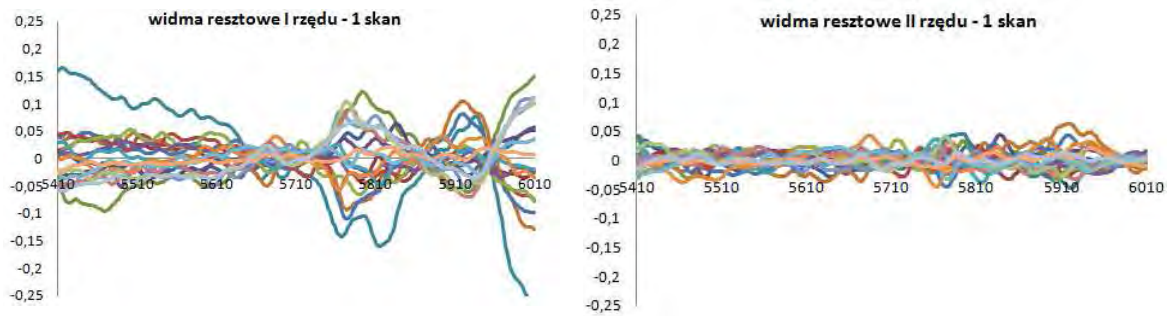
Rys.25 Widma surowe dla zestawu próbek alkoholu lanolinowego w zakresie 5410- 6010 cm^{-1} zarejestrowane jako pojedyncze skany

Po wprowadzeniu danych do programu PCAMAX uzyskano wyniki przedstawione na **Rys.26**. Analizując % zmienności wyjaśnianej przez kolejne czynniki główne można dojść do wniosku, że istotna jest tylko jedna główna składowa.



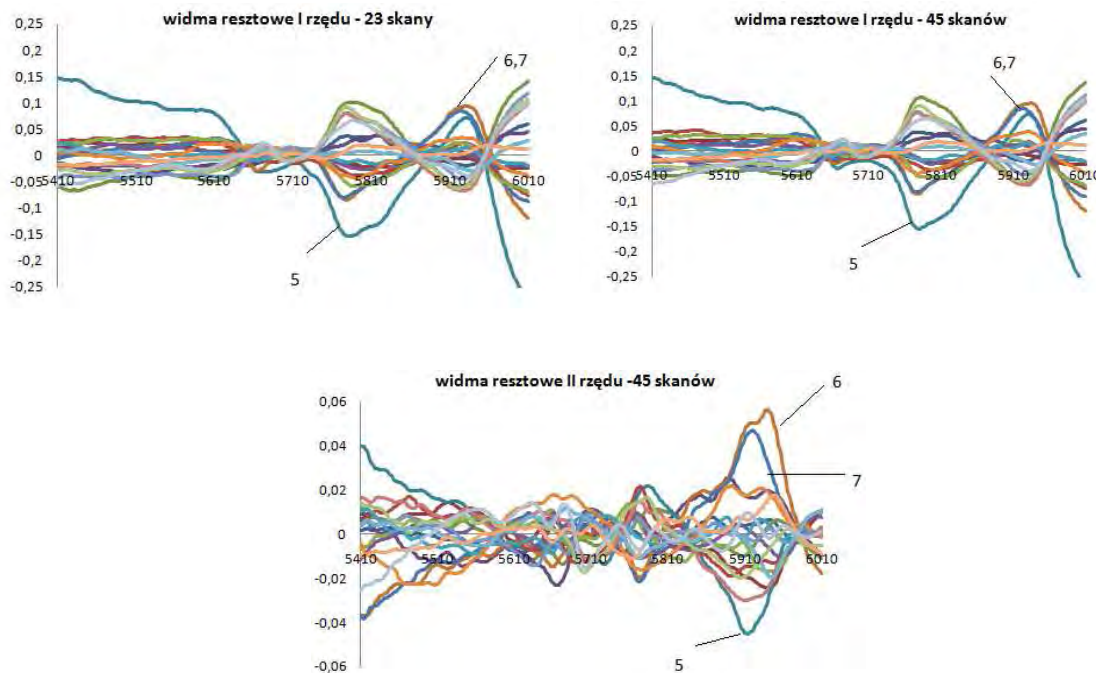
Rys.26 Procent zmienności wyjaśniane przez kolejne główne składowe dla zestawu widm z **Rys.25**

Analizując natomiast widma resztowe widać, **Rys.27**, że punkty przecięcia występują tylko w widmie I rzędu. W pozostałych widmach resztowych widoczny jest jedynie szum. Za istotne należy więc uznać dwie pierwsze główne składowe.



Rys.27 Widma resztowe dla zestawu widm z Rys.25 (pojedyncze skany)

W widmach resztowych zestawów widm rejestrowanych odpowiednio dla 3, 5, 10, 15, 23, 32, 45 i 62 skanów obserwuje się jednak punkty przecięcia oraz pasma absorpcji zarówno w widmach resztowych I rzędu jak i II rzędu, **Rys.28**.



Rys.28 Przykładowe widma resztowe zestawów widm alkoholu lanolinowego zarejestrowanych przy a) 23 i b) 45 skanach

We wszystkich widmach resztowych pierwszego rzędu wyraźnie odbiegają próbki 5, 6 oraz 7. Pasma pochodzące od tych próbek wyraźne są również w zestawach widm resztowych II rzędu. W kolejnych widmach resztowych widać już sam szum. Zatem na podstawie tych wyników należy uznać za istotne 3 czynniki główne.

Brak punktów przecięcia i pasm w zestawie widm resztowych II rzędu dla widm zarejestrowanych jako pojedyncze skany może świadczyć o tym, że przy tym trybie rejestracji poziom szumu losowego jest porównywalny do swoistej zmienności zawartej w próbkach 5, 6 i 7. Zatem dla zestawu widm zarejestrowanych jako pojedyncze skany należy uznać za istotne 2 składowe, natomiast dla pozostałych 3 składowe.

W przypadku pozostałych surowców również nie zaobserwowano wpływu ilości skanów na podobieństwo widm próbek rejestrowanych w postaci 3 lub więcej skanów.

Wyniki te pokazują, że w celu ograniczenia czasu analizy można opracowywać metody jakościowe nawet już dla widm rejestrowanych jako kilkakrotne skany. Wskazują również, że nie zbierać widm stosując tylko jeden skan pomiarowy. Istnieje bowiem ryzyko, że w pojedynczym trybie skanowania może dojść do zaburzenia stosunku sygnału do szumu i zbierane widma mogą być obciążone silnym szumem losowym, co może się przyczynić do zafałszowania wyników końcowych.

5.1.1.2. Ocena wpływu ilość głównych składowych na jakość końcowych wyników modeli tożsamościowych

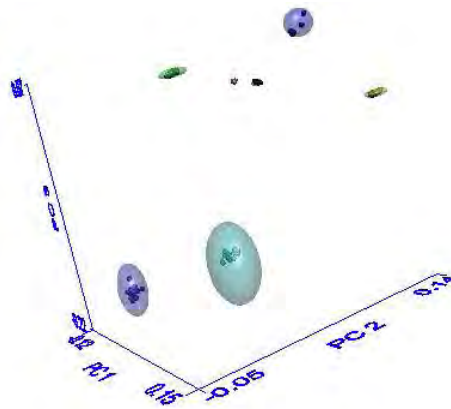
Spektrometr NIR dostępny w firmie Ziaja Ltd. wyposażony jest w firmowe oprogramowanie AssureID-MethodExplorer pozwalające na dokonanie analizy jakościowej (tożsamościowej) próbek z wykorzystaniem klasyfikatora SIMCA. Program ten wyznacza m.in. liczbę głównych składowych używaną do identyfikacji próbek. Zaobserwowałam, że wyznaczana przez program liczba składowych jest zdecydowanie większa niż wynika to z analizy chemometrycznej przeprowadzonej w p.5.1.1.1, Tab.5.

Tabela 5. Porównie liczby istotnych głównych składowych wyznaczonych za pomocą programu PCAMAX oraz AssureID MethodExplorer

<u>PCAMAX</u>		<u>AssureID MetodExplorer</u>		<u>AssureID MetodExplorer</u>	
<i>32 skany</i>	<i>PC</i>	<i>32 skany</i>	<i>PC</i>	<i>62 skany</i>	<i>PC</i>
alkohol lanolinowy	3	alkohol lanolinowy	7	alkohol lanolinowy	7
glikol butylenowy	2	glikol butylenowy	5	glikol butylenowy	4
benzoesan sodu	2	benzoesan sodu	6	benzoesan sodu	6
2-bromo-2-nitropropano-1,3-diol	1	2-bromo-2-nitropropano-1,3-diol	5	2-bromo-2-nitropropano-1,3-diol	6
ekstrakt z ogórka	2	ekstrakt z ogórka	7	ekstrakt z ogórka	5
alkohol cetearylowy	2	alkohol cetearylowy	7	alkohol cetearylowy	7

Dlatego też postanowiłam porównać jakość modeli tożsamościowych uzyskanych dla narzuconej, mniejszej liczby głównych składowych.

W pierwszym etapie za maksymalną liczbę głównych składowych przyjąłm cztery czynniki. Uzyskane wyniki identyfikacji pokazały, że zmniejszenie liczby czynników nie pogorszyło jakości opracowanych modeli. Odległości między poszczególnymi klasami okazały się zadawalające, **Rys.29**. Klasyfikator SIMCA w oparciu o zmniejszoną liczbę głównych składowych poprawnie zaklasyfikował próbki ze zbioru uczącego do odpowiednich klas.



Rys.29 Rzut przedstawiający próbki surowców ze zbioru uczącego w przestrzeni głównych składowych dla modeli klas zawierających nie więcej niż 4 główne składowe

Również wszystkie próbki ze zbioru testowego zostały poprawnie sklasyfikowane. Wartości *Specified Material Total Distance Ratio* (odległość danej próbki od środka klasy – poprawna wartość w granicach 0-1) okazały się być w dolnej granicy kryterium akceptacji, podczas gdy dla modeli z automatycznie wygenerowaną liczbą składowych były dużo wyższe. Świadczy to, że generowane automatycznie modele tożsamościowe wykazują cechy modeli „przeuczonych”.

W przypadku 3 czynników głównych również nie zaobserwowano znaczących zmian. Przy 3 istotnych głównych składowych wszystkie próbki ze zbioru uczącego oraz testowego zostały poprawnie sklasyfikowane. Natomiast po kolejnym zmniejszeniu liczby składowych z 3 na 2 czynniki zaobserwowano, że dwa czynniki główne nie wyjaśniły całkowitej zmienności występującej między badanymi surowcami. Próbkami ze zbioru uczącego zostały poprawnie sklasyfikowane, jednakże nie wszystkie próbki ze zbioru testowego zostały poprawnie zakwalifikowane do swoich klas. Wartości *Specified Material Total Distance Ratio* dla tych próbek przekraczały nieznacznie górną granicę

kryterium akceptacji klasyfikacji. Zbyt mała liczba głównych składowych okazała się zatem być nie wystarczająca do poprawnego opisu zmienności występującej w obrębie i pomiędzy klasami.

Różnica w liczbie czynników wyznaczonych przez oba programy wynika z użycia różnych kryteriów doboru. W programie dostarczonym przez producenta sprzętu jako istotne uznaje się główne składowe wyjaśniające łącznie ponad 99,999 % zmienności zawartej w danych. Tym samym bardzo często brana jest pod uwagę nie tylko zmienność systematyczna, ale także niepożądana zmienność losowa wynikająca z szumów pomiarowych. Modele oparte o takie „losowe” główne składowe mają tendencję do „przeuczania się”: są bardzo dobrze dopasowane do zbioru uczącego, ale źle klasyfikują niektóre próby ze zbioru testowego. Program PCAMAX pozwala na zastosowanie analizy widm resztowych jako kryterium doboru liczby składowych. Pozwala to uwzględnić jedynie te składowe, które zawierają w sobie znaczący udział zmienności systematycznej (wspólnej).

Dlatego też bardzo ważny jest poprawny dobór liczby istotnych składowych. Modele zbudowane w oparciu o zbyt mało czynników są źle dopasowane do danych, natomiast modele o zbyt dużej liczbie czynników charakteryzują się dużym błędem predykcyjnym. Wybór optymalnej liczby czynników do konstrukcji modeli pozwala osiągnąć kompromis pomiędzy zadowalającym dopasowaniem modelu do danych a dobrymi własnościami predykcyjnymi modelu.

5.1.1.3. Powtarzalność pomiarów

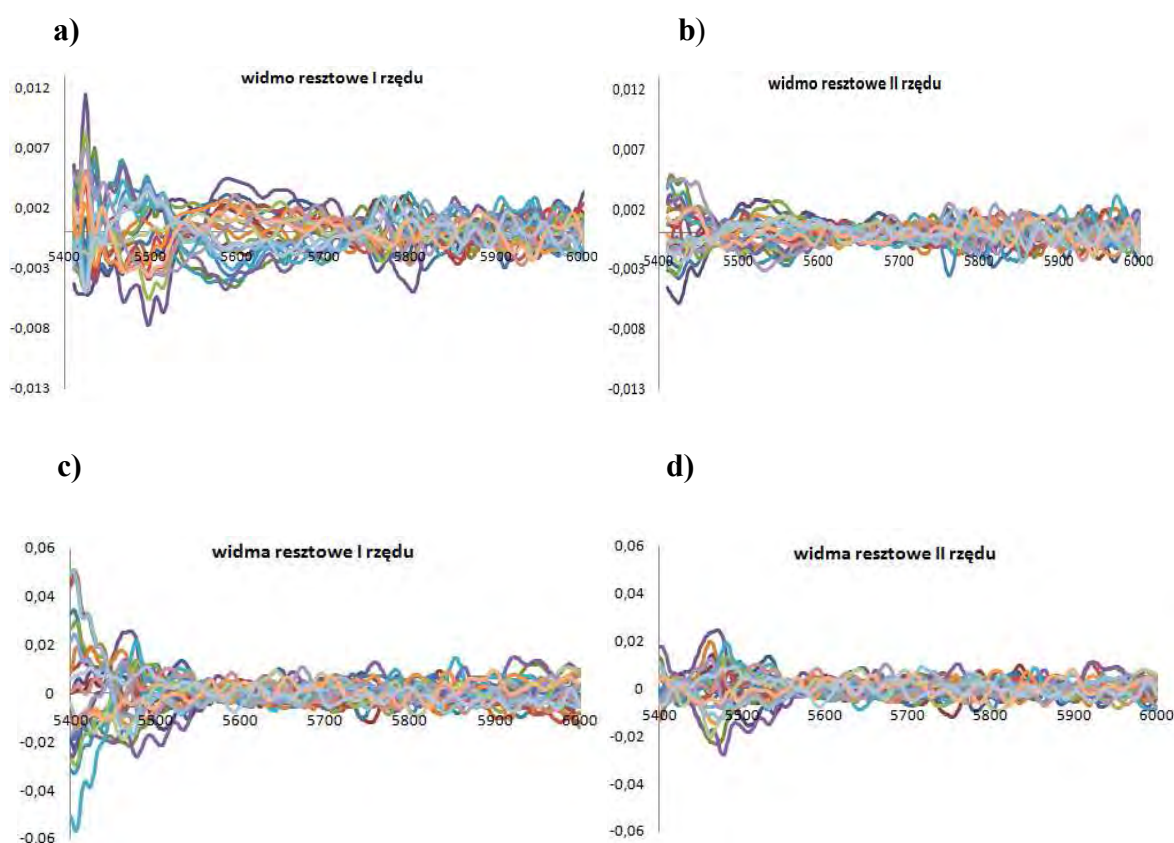
W celu zbadania powtarzalności wyników wykonano pomiary jednej próbki bez zmiany jej położenia pomiędzy pomiarami. Wybraną próbkę mierzono 25 razy. Poniżej przedstawiono warunki pomiarowe.

Tabela 6. Warunki pomiarów podczas analizy powtarzalności pomiarów

Warunki pomiarów	Model
Liczba skanów cm^{-1}	32
Detektor	2
Rozdzielczość cm^{-1}	16
Szybkość skanowania cm/s	1
Odległość punktów pomiarowych cm^{-1}	2
Zakres pomiarów cm^{-1}	10000-4000
Filtry	włączone

Pomiary przeprowadzono dla wszystkich badanych substancji (p.5.1.1). Poniżej przedstawiono przykładowe wyniki uzyskane dla: glikolu butylenowego, 2-bromo-2-nitropropano-1,3-diolu oraz alkoholu cetearylowego.

W przypadku dwóch pierwszych substancji za pomocą programu PCAMAX stwierdzono, że pierwsza główna składowa wyjaśnia ponad 99,99% zmienności zestawu widm.

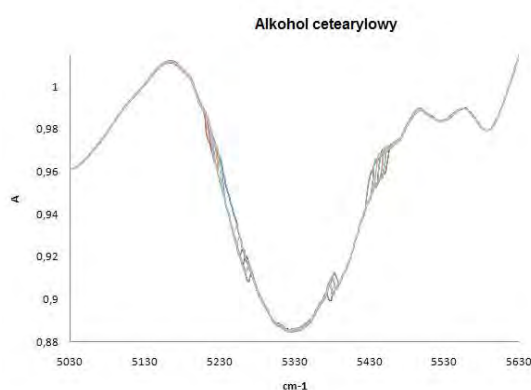


Rys.30 Widma resztowe: a) i b) glikol butylenowy, c) i d) 2-bromo-2-nitropropan-1,3-diol. Zestawy widm rejestrowane były bez zmiany położenia próbki na okienku pomiarowym

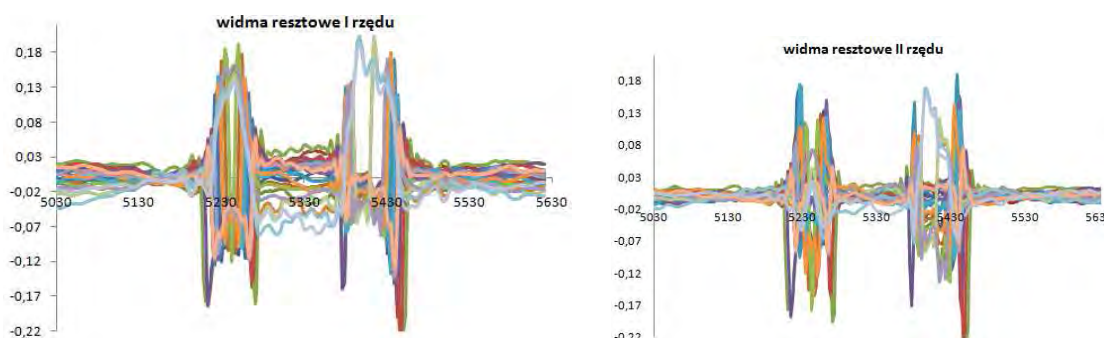
W widmach resztowych I i II rzędu, **Rys.30**, nie zaobserwowano wyraźnych pasm absorpcji oraz punktów przecięcia – zatem za istotną można przyjąć tylko jedną główną składową. Świadczy to o dobrej powtarzalności pomiarów – ponawianie tego samego eksperymentu nie generuje dodatkowej zmienności. Analogiczne wyniki uzyskano również dla zestawów widm benzoesanu sodu, alkoholu lanolinowego i ekstraktu z ogórka.

Ciekawe zjawisko zaobserwowano dla trzeciego surowca: alkoholu cetearylowego. Widma w zakresie $5030\text{--}5630\text{ cm}^{-1}$ zebrane w wyniku wielokrotnego prześwietlenia tej samej próbki wykazały obecność kilku obszarów nieciągłości, **Rys.31**. Zjawisko takie

pojawia się czasami, zwłaszcza w przypadku substancji stałych o dużych, niejednorodnych cząstkach. Analiza widm resztowych wykazuje, **Rys.32**, że nieciągłości te mają charakter losowy, a generowana przez nie zmienność widm nie jest uwzględniana w trzech pierwszych głównych składowych. Prawdopodobnie, dodatkowa zmienność wynika z dużej niejednorodności próbki, która składała się z granulek o średnicy kilku milimetrów. Istotny wpływ na otrzymane widma może mieć efekt rozpraszania światła laserowego na niejednorodnym materiale. Pod wpływem drgań przyrządu lub podłoża orientacja cząstek próbki względem padającego światła może ulegać przypadkowym zmianom.



Rys.31 Widma surowe otrzymane dla alkoholu cetearylowego w zakresie 5030-5630 cm^{-1} bez zmiany położenia próbki



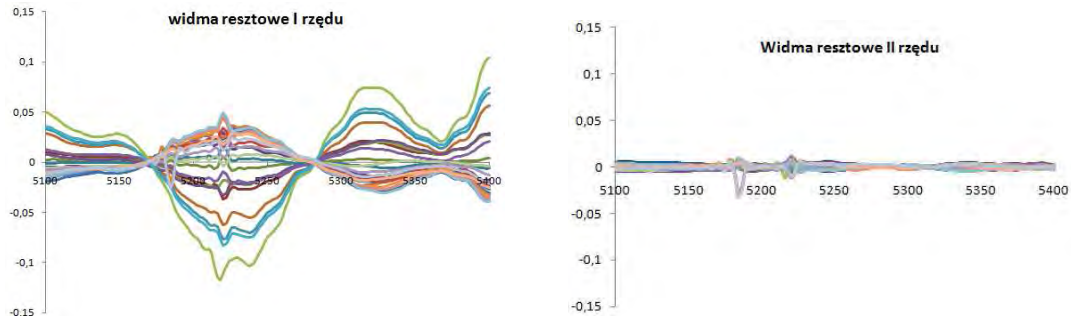
Rys.32 Widma resztowe zestawu widm alkoholu cetearylowego z Rys.31

5.1.1.4. Ocena wpływu położenia próbki na okienku pomiarowym

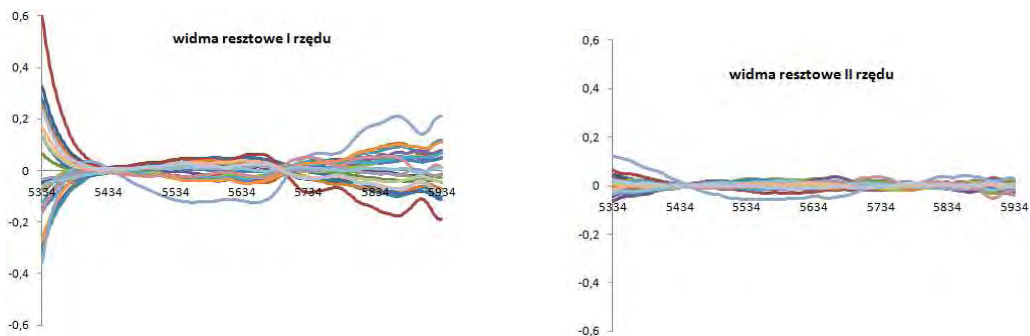
Ze względu na to, że na efekty końcowe opracowywanych modeli może mieć wpływ wiele czynników, postanowiono również zbadać czy położenia próbki na okienku pomiarowym ma wpływ na kształt widm NIR. Często bowiem zdarza się tak, że analityk nie zwraca uwagi na poprawne położenie próbki na oknie pomiarowym. W tym celu zebrano widma jednej próbki 25 razy przy różnych położeniach badanej próbki na okienku pomiarowym. Próbki mierzono w 32 skanowym trybie pomiarowym. Próbkę ekspozowano

na oknie pomiarowym w sposób losowy, tak aby badana substancja w różnym stopniu przykrywała źródło promieniowania.

a)



b)



Rys.33 Widma resztowe a) benzoianu sodu b) ekstraktu z ogórka. Każdy zestaw zawierał 25 widm tej samej próbki zarejestrowanych przy różnym położeniu próbki na okienku pomiarowym. W niektórych przypadkach tylko część promieniowania NIR przechodziła przez próbkę

Na **Rys.33** przedstawiono przykłady analizy widm resztowych dla zestawów widm benzoianu sodu i ekstraktu z ogórka. Na podstawie powyższych widm można wywnioskować, że położenia próbki ma istotny wpływ na kształt rejestrowanego widma. Widma odbiegające od reszty na **Rys.33** pochodzą od próbek, które nie przykrywały w całości okna pomiarowego. Zatem podczas opracowywania modeli, a także w czasie przeprowadzania analiz rutynowych należy starannie i dokładnie stawiać badaną substancję na okno pomiarowe. Niedokładne przykrycie okna pomiarowego może doprowadzić do sytuacji gdy próbka nie zostanie poprawnie zidentyfikowana.

5.1.1.5. Ocena wpływu środowiska na jakość opracowywanych modeli

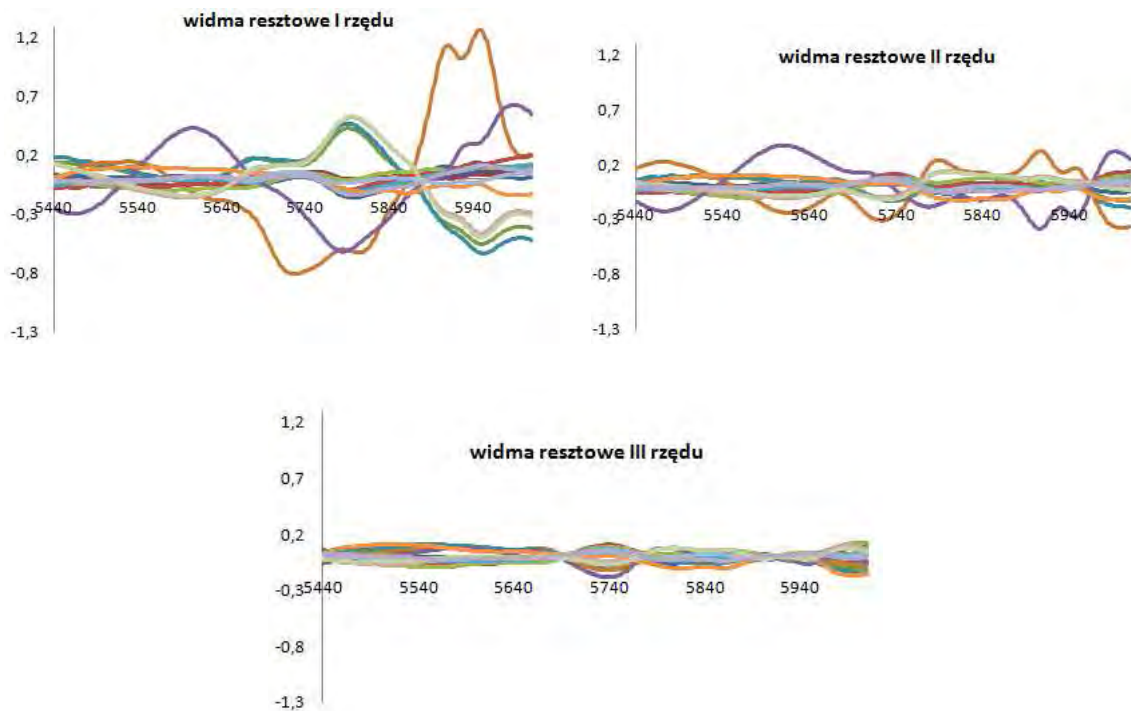
Widma NIR są wrażliwe na środowisko w jakim znajduje się próbka podczas pomiaru. Przede wszystkim dotyczy to takich parametrów środowiska jak temperatura czy wilgotność. Nie można jednak wykluczyć, że również inne, mniej oczywiste parametry środowiska mogą wpływać na kształt i intensywność widma.

W celu sprawdzenia łącznego wpływu otoczenia na odtwarzalność widm, zebrano widma modelowych surowców w różnych dniach i o różnych porach dnia. Do tego doświadczenia wybrano po 1 próbce z każdego surowca i następnie rejestrowano widma wybranych próbek surowców w 32 skanowym trybie pomiarowym. Ze względu na wyniki przedstawione w p.5.1.1.3 podczas wykonywania widm starannie kontrolowano położenie próbki na oknie pomiarowym przyrządu.

Tabela 7. Liczba istotnych głównych składowych dla zestawów widm modelowych surowców. Widma rejestrowane były w przypadkowych dniach w okresie 12 miesięcy

<u>Surowce</u>	
	PC
alkohol lanolinowy	3
glikol butylenowy	4
benzoesan sodu	2
2-bromo-2nitropropano-1,3-diol	3
ekstrakt z ogórka	4
alkohol cetearylowy	2

W widmach wszystkich surowców zaobserwowano znaczący wpływ warunków środowiska na powtarzalność rejestrowanych widm. Mimo że za każdym razem rejestrowana była ta sama próbka danego surowca na widmach resztowych widać punkty izobestyczne oraz wyraźne pasma absorpcji. Zatem do wyjaśnienia zmienności wynikającej ze zmian otoczenia potrzebne jest aż kilka głównych składowych.



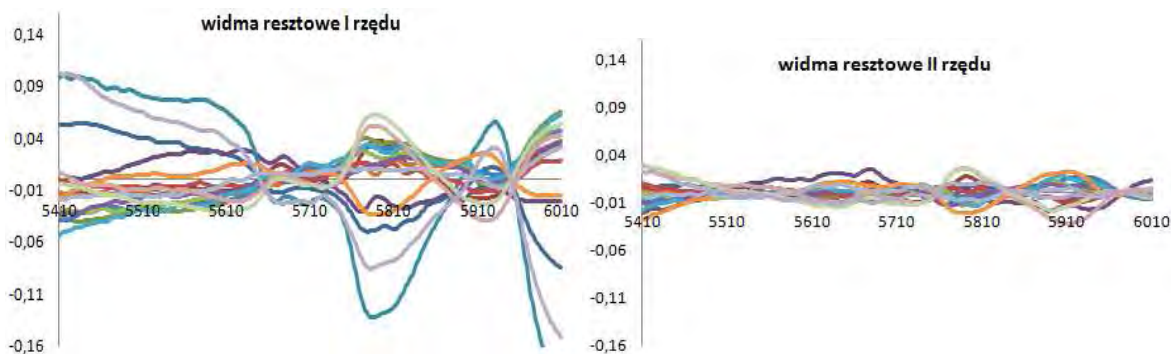
Rys.34 Widma resztowe zestawu widm próbki ekstraktu z ogórka mierzone w różnych warunkach w ciągu 12 miesięcy

Ekstrakt z ogórka

Wyniki uzyskane dla ekstraktu z ogórka, **Rys.34**, pokazują, że do wyjaśnienia systematycznej zmienności w tym zestawie widm potrzebne są aż 4 główne składowe. W widmach resztowych I. i II. rzędu widoczne są wyraźne pasma absorpcji, szczególnie silne dla próbek 10 oraz 6 (odpowiednio linia fioletowa i brązowa). Ponadto obserwuje się punkty przecięcia. W przypadku widma resztowego III. rzędu również można jeszcze zaobserwować punkty przecięcia.

Alkohol lanolinowy

W przypadku zestawu widm zarejestrowanych dla próbki alkoholu lanolinowego w zróżnicowanych warunkach otoczenia analiza widm resztowych, **Rys.35**, również prowadzi do wniosku, że zmienność systematyczna wynika z kilku źródeł zmienności. W widmach resztowych I oraz II rzędu obserwuje się występowanie wyraźnych pasm absorpcji oraz punktów przecięcia. W widmie resztowym III. rzędu widoczny jest już głównie szum losowy. Zatem na tej podstawie można stwierdzić, że dla tego zestawu widm istotne są 3 główne składowe.

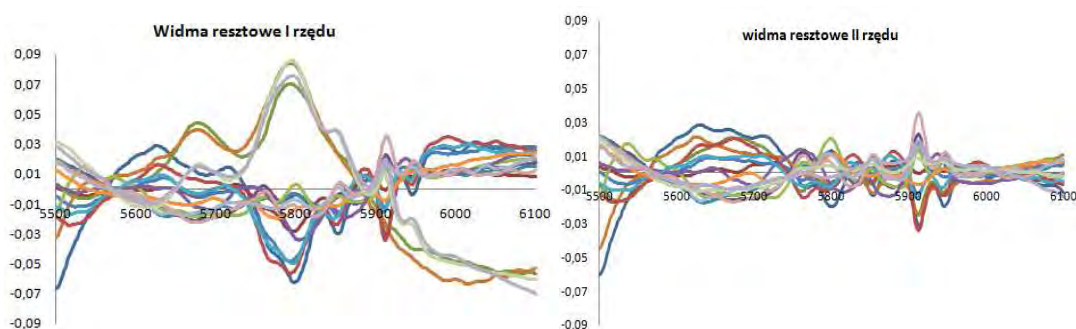


Rys.35 Widma resztowe zestawu widm próbki alkoholu lanolinowego mierzone w różnych warunkach otoczenia w czasie 12 miesięcy

2-bromo-2-nitropropan-1,3-ol

Również dla tego surowca stwierdzono wpływ warunków otoczenia na kształt widm próbki. Analiza widm resztowych sugeruje, **Rys.36**, że dla wyjaśnienia zmienności systematycznej w zarejestrowanym zestawie widm potrzeba 3 głównych składowych.

Mimo, że widma próbek badanych surowców mierzone były w tych samych dniach (podobne warunki otoczenia), to obserwuje się różną liczbę istotnych czynników dla poszczególnych surowców, **Tab.7**. Wydaje się, że liczba istotnych składowych związana jest ze stanem skupienia surowca. W przypadku surowców ciekłych (ekstrakt z ogórka oraz glikol butylenowy) za istotne uznać należy 4 czynniki, a w przypadku surowców stałych 2 lub 3 czynniki. Na tym etapie badań trudno jest znaleźć racjonalne uzasadnienie obserwowanej zależności. Nie jest ponadto jasne, czy zależność ma charakter ogólny czy też dotyczy tylko badanego zestawu surowców.

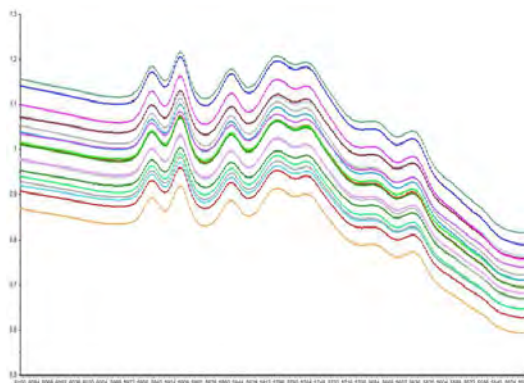


Rys.36 Widma resztowe zestawu widm próbki 2-bromo-2-nitropropan-1,3-diolu mierzone w różnych dniach w ciągu 12 miesięcy

Podsumowując można stwierdzić, że warunki otoczenia mają znaczący wpływ na zbieranych widm. Zatem w opracowywanych modelach należy uwzględnić zmienność wynikającą z różnorodnych warunków panujących podczas pomiaru. Widma próbek nie należy zbierać jednego dnia, lecz w możliwie długim okresie czasu. Wskazane jest ponadto kilkakrotne zarejestrowanie widm tej samej próbki w różnych porach roku. Pozwoli to uwzględnić zarówno wpływ czynników przypadkowych jak i wynikających ze zmian sezonowych. Modele zbudowane z wykorzystaniem takiego zróżnicowanego zbioru uczącego powinny działać poprawnie niezależnie od zmian warunków otoczenia.

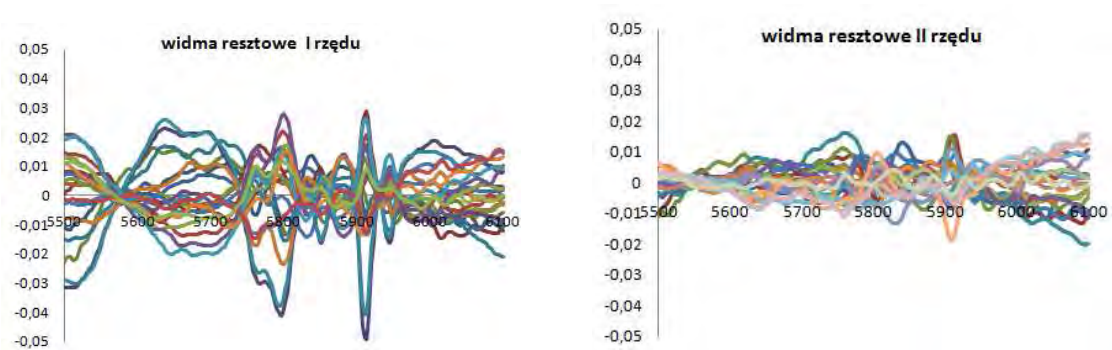
5.1.1.6. Wpływ wstępnego przygotowania danych na konstrukcję czynników głównych.

Wpływ technik wstępnego przygotowania widm na liczbę istotnych czynników głównych przebadano na przykładzie 21 widm 2-bromo-2-nitropropan-1,3-diolu.



Rys.37 Surowe widma próbek 2-bromo-2-nitropropan-1,3-diolu zarejestrowane w zakresie $5500\text{-}6100\text{ cm}^{-1}$ przy 32 skanach

W pierwszym etapie badań wykonano analizę głównych składowych dla surowych danych, **Rys.37**. W tym celu posłużono się programem PCAMAX oraz Unscramble 10.1. Analiza widm resztowych wykazała, że dla tego zestawie widm istotne są dwie główne składowe, **Rys.38**. Na widmie resztowym pierwszego rzędu obserwuje się wyraźne pasma absorpcji oraz punkty przecięcia. Na widmie resztowym drugiego rzędu natomiast widoczny jest już tylko szum.

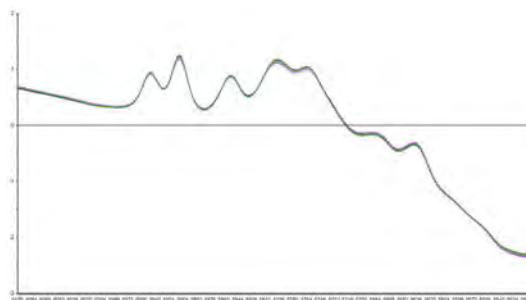


Rys.38 Widma resztowe zestawu widm z Rys.37

W dalszym etapie analizy przeprowadzono różnorodną transformację widm surowych. Poniżej przedstawiono kilka typowych przykładów wpływu zastosowanej transformacji na konstrukcję czynników głównych.

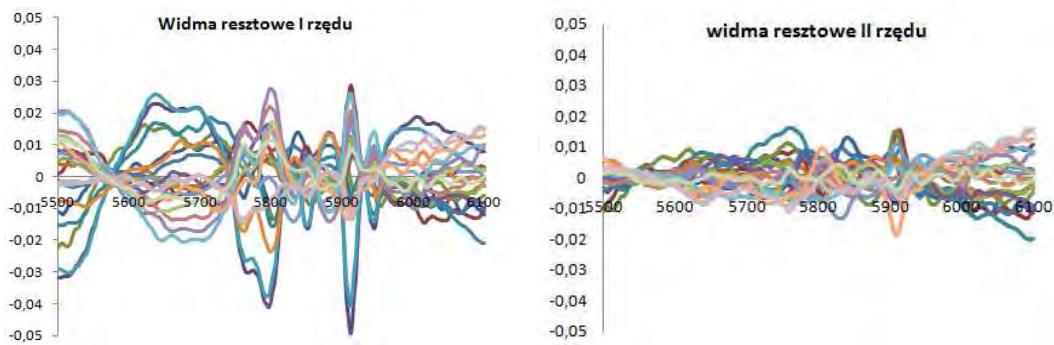
a) Normalizacja I – transformacja SNV

Transformacja SNV ma na celu usunięcie różnic w intensywności widm wynikających z różnego nasilenia efektów interferencji bądź rozpraszania promieniowania elektromagnetycznego dla poszczególnych próbek. Porównując widma badanego surowca przed i po transformacji (odpowiednio **Rys.37** i **Rys.39**) można stwierdzić, że w wyniku transformacji doszło do zdecydowanego zmniejszenia różnic w intensywności widm poszczególnych próbek.



Rys.39 Widma 2-bromo-2-nitropropan-1,3-diolu z rys.37 otrzymane po zastosowaniu transformacji SNV

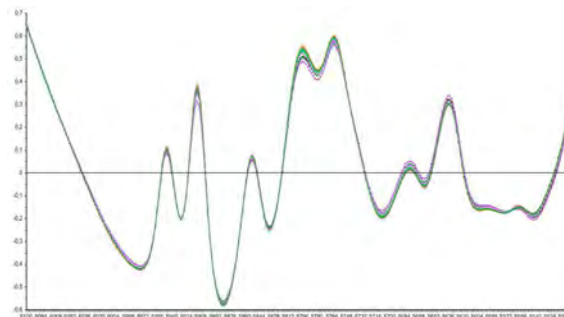
Jednakże nie wpłynęło to na liczbę istotnych głównych składowych. W dalszym ciągu za istotne należy uznać dwie pierwsze główne składowe. Widoczne jest to wyraźnie na widmach resztowych przed i po zastosowaniu transformacji SNV (odpowiednio **Rys.38** i **Rys.41**).



Rys.40 Widma resztowe zestawu widm próbek 2-bromo-2-nitropropan-1,3-diolu po zastosowaniu transformacji SNV

b) Normalizacja II – transformacja SNV + transformacja Detrend

Często normalizację SNV łączy się z transformacją Detrend. Ma to na celu nie tylko ujednoczenie intensywności widm, ale również wyeliminowanie wpływu linii bazowej na kształt poszczególnych widm.

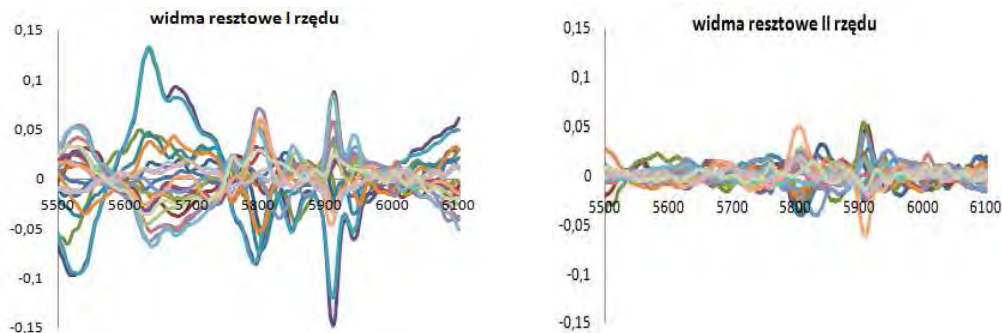


Rys.41 Widma uzyskane po zastosowaniu transformacji SNV wraz z transformacją Detrend do zestawu widm z Rys.37

Analizując zestaw widm uzyskany po takiej łącznej transformacji obserwuje się wyraźniejsze pasma, szczególnie w zakresie od 5700-5500 cm^{-1} . Pasma te w widmach surowych (**Rys.37**) oraz w widmach po zastosowaniu samej transformacji SNV (**Rys.39**) nie były tak wyraźnie widoczne.

Po przeprowadzeniu analizy widm resztowych dla tak transformowanych widm nie stwierdzono zmiany liczby istotnych czynników, **Rys.42**. Jednakże zaobserwowano, że widma resztowe I rzędu dla zestawu po transformacji łączonej mają zdecydowanie większą intensywność niż dla zestawu po samej transformacji SNV. Dzięki temu wyraźniejsze stały się występujące w nich pasma absorpcji. Ponadto, w zakresie od 5600-5700 cm^{-1} obserwuje się dwa widma, które wyraźnie różnią się od pozostałych. Są to

widma próbek 11 i 4.. Zatem mimo, że zastosowanie łączonej transformacji SNV wraz z Detrend nie przyczyniło się do zmiany liczby istotnych czynników, to transformacja taka ułatwia wykrycie widm odbiegających i w niektórych przypadkach może być korzystna.

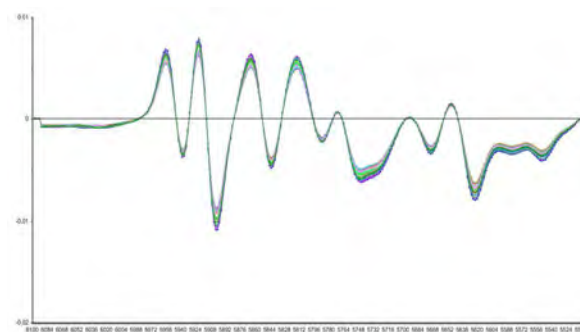


Rys.42 Widma resztowe uzyskane po zastosowaniu transformacji łączonej SNV+Detrend. Odbiegające widma resztowe I. rzędu (linie fioletowa i niebieska) odpowiadają próbkom 11 i 4

c) Transformacja widm do ich pierwszych pochodnych

Stosowanie pochodnych jako funkcji transformujących ma na celu usunięcie efektów związanych z przesunięciem oraz zakrzywieniem linii bazowej, a także w celu rozróżnienia nakładających się na siebie pasm absorpcji i uwidocznienia zmienności nie widocznej gołym okiem.

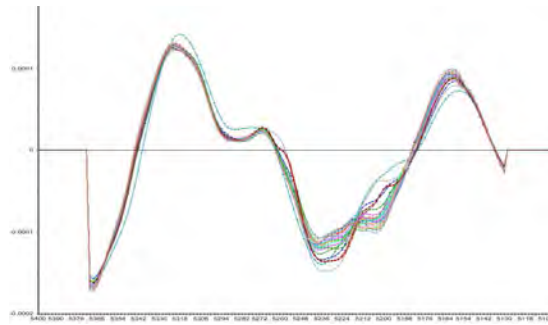
Do transformacji widm do ich pierwszych pochodnych zastosowano algorytm Stavickego-Golay'a. Podczas transformacji użyto wielomianu drugiego stopnia z 9-elementowym oknem. Uzyskane wyniki przedstawia **Rys.43**.



Rys.43 Wynik transformacji zestawu widm z Rys.37 do ich pierwszej pochodnej

d) Transformacja widm do ich drugich pochodnych

Podczas wyznaczania 2. pochodnych widm zastosowano algorytm Stavickego-Golay'a z wielomianem stopnia 3. Wynik transformacji widm do ich drugich pochodnych przedstawia **Rys.44**.



Rys.44 Wynik transformacji zestawu widm po zastosowaniu transformacji do ich drugich pochodnych

Analiza widm resztowych zestawu 2. pochodnych widm wykazała, że istotna jest tylko jedna główna składowa. Zaobserwowano więc spadek liczby istotnych czynników. W widmie resztowym pierwszego rzędu dla tego zestawu, **Rys.45**, nie obserwuje się żadnych punktów przecięcia ani wyraźnych pasm absorpcji. Charakterystyczne jest przy tym, że w widmach resztowych 1. rzędu widma próbek odbiegających (próbki 11 i 4) maksymalnie tylko dwukrotnie przekraczają poziom szumu.



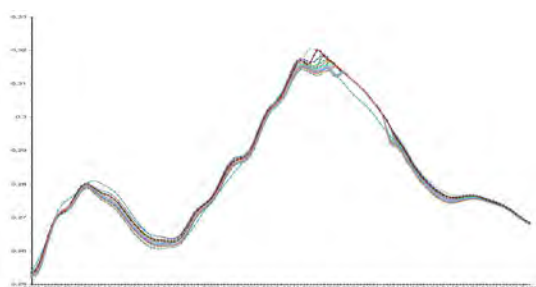
Rys.45 Widma resztowe I. rzędu dla zestawu widm po zastosowaniu transformacji do ich drugich pochodnych. Próbkom odbiegającym 11 i 4 odpowiadają linie niebieska i fioletowa

Uzyskane wyniki pokazują, że rodzaj zastosowanej transformacji ma wpływ na liczbę istotnych czynników. Stosując różnego rodzaju techniki transformujące zmienia się kształt widm a tym samym wpływ poszczególnych punktów na liczbę czynników głównych. Wyraźnie jest to widoczne przy zastosowaniu drugiej pochodnej. Stosując pierwszą pochodną nie wyeliminowano wpływu dwóch próbek na zestaw danych. Wpływ ten zmniejszył się, jednak nie uległ całkowicie i dlatego program za istotne wyznaczył dwie główne składowe. Efekt ten został zlikwidowano po zastosowaniu 2 pochodnej i dlatego za istotną uznano tylko jedna główną składową.

5.1.1.7. Wpływ punktów odbiegających na konstrukcję czynników głównych

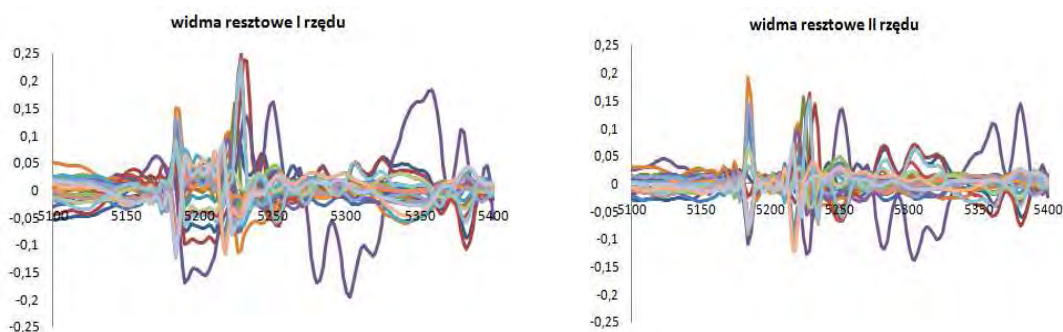
Obecność punktów odbiegających w analizowanym zestawie danych ma bardzo istotny wpływ na liczbę istotnych czynników. Wpływ takich punktów może być pozytywny wynikający z dodatkowej zmienności wnoszonej przez dane pomiarowe. Efekt wpływu takich próbek może zmaleć w momencie gdy do modelu zostaną dodane punkty o podobnej zmienności. Jednak punkty takie mogą również świadczyć o jakimś błędzie wynikającym np. z błędów powstałych podczas pomiaru danej próbki bądź z niepoprawnych zmienności zawartych w takich danych. Punkty takie należy cały czas kontrolować i obserwować sprawdzając czy ich wpływ maleje czy rośnie.

Na przykładzie widm surowca benzoesan sodu zostanie pokazany wpływ punktów odbiegających na konstrukcję czynników głównych. Badany zestaw danych został poddany różnorakiej transformacji widm.



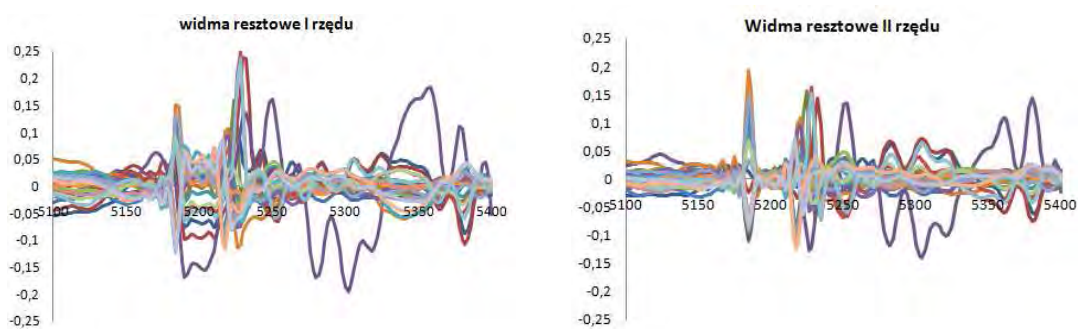
Rys.46 Surowe widma surowca benzoesan sodu zarejestrowane w zakresie $5100-5400\text{ cm}^{-1}$

W pierwszym etapie przeprowadzanych badań wykonano analizę głównych składowych na surowych danych. Podczas analizy widm resztowych zaobserwowano wyraźny wpływ punktu 10 **Rys.47** (linia fioletowa). Widmo to odbiega znacznie od pozostałych.



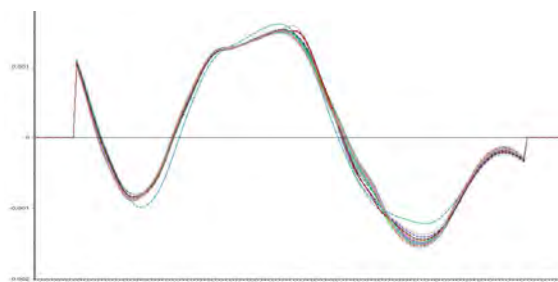
Rys.47 Widma resztowe surowca benzoesan sodu

Dodatkowo w widmach resztowych brak jest wyraźnych punktów przecięcia. Pasma absorpcji przybierają dziwny szpiczasty kształt. Taki efekt może świadczyć o konieczności przeprowadzenia transformacji danych w celu eliminacji negatywnej zmienności wpływającej na kształt danych. Efekt widoczny w widmach resztowych może pochodzić od wielu czynników : warunków środowiska panującego w laboratorium (zmiennej temperatury i wilgotności), z błędów pochodzących od sprzętu bądź też z różnicy wielkości cząstek substancji badanej lub z nowej zmienności wniesionej do danych. W pierwszym etapie badań przeprowadzono transformację SNV. Po zastosowaniu transformacji SNV został zmniejszony efekt różnicy rozpraszania między próbkami, ostre pasma uległy lekkiemu wygładzeniu ale tym samym zwiększył się wpływ punktu odbiegającego na konstrukcję głównych składowych. W widmach resztowych nadal nie obserwuje się wyraźnych punktów przecięcia dlatego też wyznaczenie liczby istotnych czynników jest nie możliwe. Zatem zastosowanie transformacji SNV nie przyniosło pożądanego efektu. **Rys.48.**



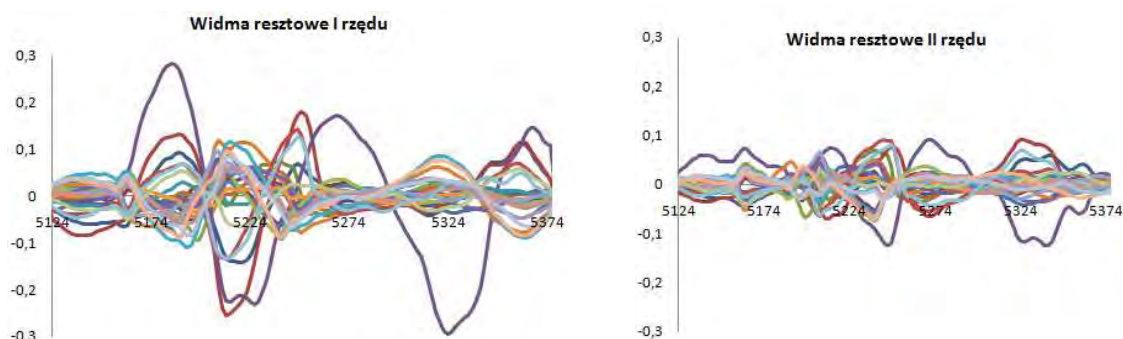
Rys.48 Widma resztowe otrzymane po zastosowaniu transformacji SNV

W kolejnym etapie badań zastosowano transformację widm do ich pierwszej pochodnej stosując algorytm Stavickego-Golaya. Do transformacji użyto wielomianu drugiego stopnia z 12- elementowym oknem.



Rys.49 Wykres przedstawiający widma surowca benzoesan sodu po zastosowaniu 1 pochodnej

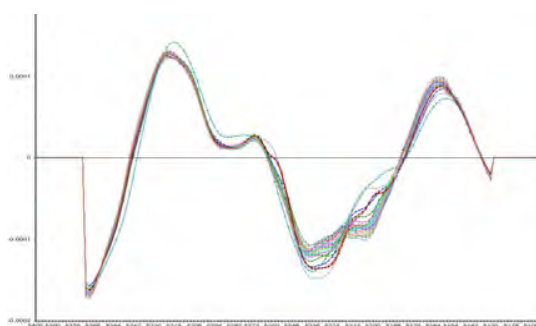
Porównując widma surowe z widmami po zastosowaniu 1 pochodnej widać wyraźne różnice w kształcie widm. Uwidocznione zostały mało widoczne punkty przegięcia, maksima natomiast w widmie widoczne są jako punkty o zerowej wartości rzędnych.



Rys.50 Widma resztowe uzyskane po zastosowaniu 1 pochodnej

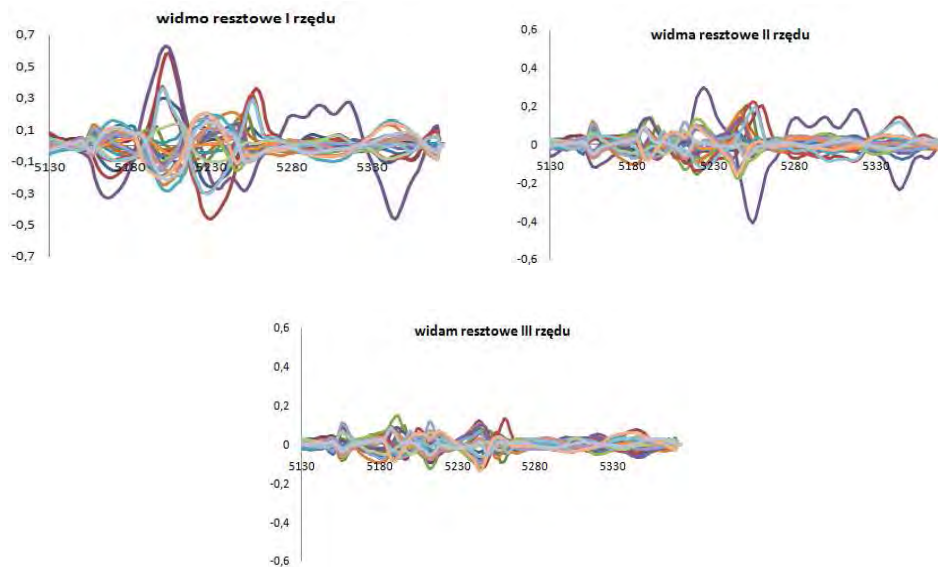
Analizując natomiast widma resztowe I i II (**Rys.50**) rzędu obserwuje się pojawienie się punktów przecięcia wcześniej nie widocznych. Na tej podstawie uznano za istotne 3 czynniki główne. Zatem uwidocznienie słabo widocznych punktów przegięcia na wykresie absorbancji niesie ze sobą tak samo ważne informacje co obecność maksimów oraz minimów.

W dalszym etapie badań również zastosowano algorytm Stavickgo-Golaya. Tym razem do wyznaczenia 2 pochodnej. Do transformacji użyto wielomian 2 stopnia z 15 elementowym oknem **Rys 51**.



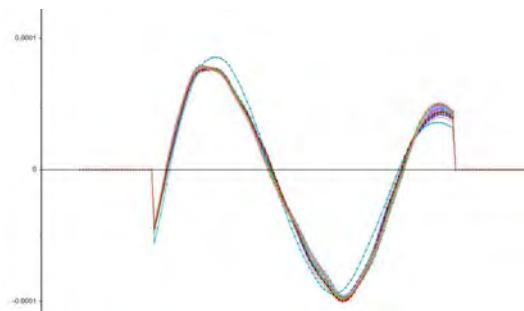
Rys.51 Wykres przedstawiający widma benzoesan sodu po zastosowaniu 2 pochodnej(okno 15-elementowe)

Zastosowanie drugiej pochodnej wzmocniło szумы widoczne w zakresie 5248 - 5188 cm^{-1} . Ponadto, odpowiednie wykresy resztowe (**Rys.52**) wykazały mniejszą liczbę punktów przecięcia. W widmie resztowym II rzędu nie obserwuje się punktów przecięcia, co wynika ze zmniejszenia się stosunku sygnału do szumu, co jest charakterystyczne dla transformacji różniczkowania.



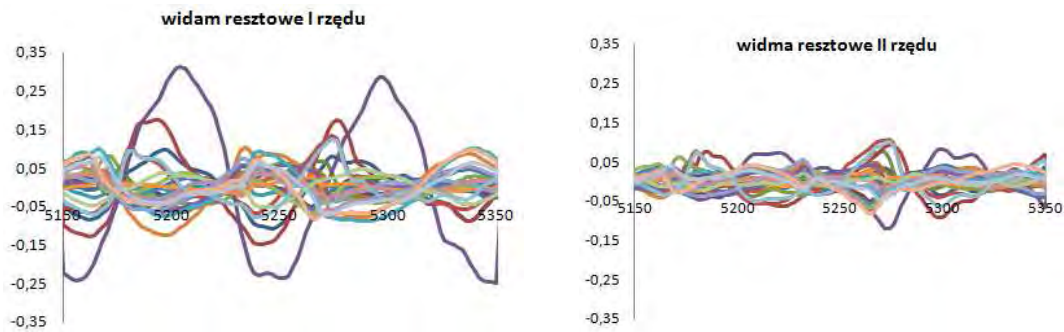
Rys.52 Widma resztowe otrzymane po zastosowaniu 2 pochodnej

Zwiększając szerokość okna wielomianu z 15 elementów na 25 zaobserwowano wygładzenie szumów **rys.53** w zakresie $5248 - 5188 \text{ cm}^{-1}$. Świadczyć to może o eliminacji informacji zarówno niepożądanych ale także pożądaných.



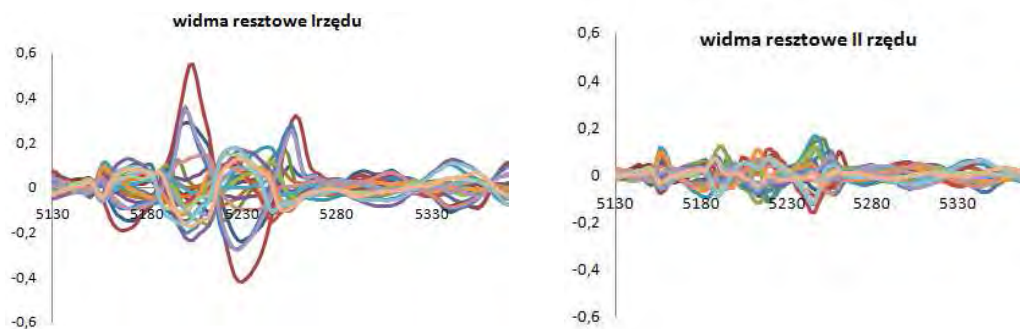
Rys.53 Widma resztowe otrzymane po zastosowaniu 2 pochodnej (25 elementowe okno wielomianu)

Analizując widma resztowe dla tak transformowanych widm widoczny jest tylko jeden punkt przecięcia w widmie resztowym I rzędu **Rys.54**. W pozostałych widać wyraźnie odbiegający punkt oraz nasilenie się szumu losowego. Zatem za istotne główne składowe uznano dwa czynniki. Zbyt duże okno wielomianu doprowadziło nie tylko do zmniejszenia liczby istotnych głównych składowych ale także do usunięcia informacji pożądaney.



Rys.54 Widma resztowe uzyskane dla danych po zastosowaniu drugiej pochodnej

W celu potwierdzenia wpływu punktu odbiegającego na konstrukcje czynników głównych, z danych usunięto punkt odbiegający i przeprowadzono transformację danych do ich drugich pochodnych z 12-elementowym oknem wielomianu stosując algorytm Stavickego-Golaya. Uzyskane wyniki przedstawiono na **Rys.55**.



Rys.55 Widma resztowe uzyskane dla danych pomniejszonych o jeden punkt po zastosowaniu 2. pochodnej.

Otrzymane wyniki potwierdzają duży wpływ punktu odbiegającego na liczbę istotnych głównych czynników. Po usunięciu punktu odbiegającego i zastosowaniu drugiej pochodnej istotne okazały się dwa czynniki główne. Zatem z powyższych wyników widać jak duży wpływ miał taki punkt na konstrukcje głównych składowych.

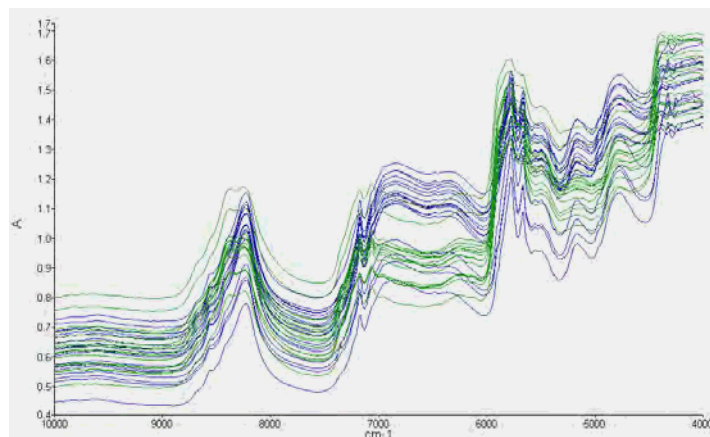
Opracowując czy to modele ilościowe czy jakościowe należy uważnie obserwować punkty odbiegające. Punkty takie nie zawsze są widoczne na pierwszy rzut oka. Dopiero po przeprowadzeniu analizy wstępnej można zaobserwować obecność takich punktów. Wpływ takich punktów można poprzez różnorakie metody transformacji usunąć bądź zniwelować. Niekiedy jednak wpływ takich punktów na opracowywany model jest tak duży, że jedynym wyjściem jest go usunięcie.

5.2. Analiza jakościowa

Po zbudowaniu biblioteki widm dla surowców stosowanych w zakładzie Ziaja Ltd. rozpoczęto pracę nad opracowywaniem metod jakościowych. Procedura tworzenia nowych metod jakościowych składa się z kilku etapów. Każdy etap opracowywania takich metod ma istotny wpływ na wynik końcowy – zdolność predykcyjną modelu. W niniejszym rozdziale przedstawione zostaną problemy związane z opracowywaniem metod jakościowych. Jednocześnie przedstawiona zostanie procedura postępowania podczas tworzenia takich metod.

5.2.1. Procedura opracowywania metod jakościowych

Procedura opracowywania nowych metod jakościowych zostanie przedstawiona na przykładzie dwóch surowców stałych : alkoholu cetearylowego, alkoholu lanolinowego. W opracowywanym modelu łącznie wykorzystano 41 widm: 20 widm próbek alkoholu cetearylowego i 21 widm próbek alkoholu lanolinowego. Tożsamość każdej próbki potwierdzono metodą referencyjną – chromatografia gazowa.

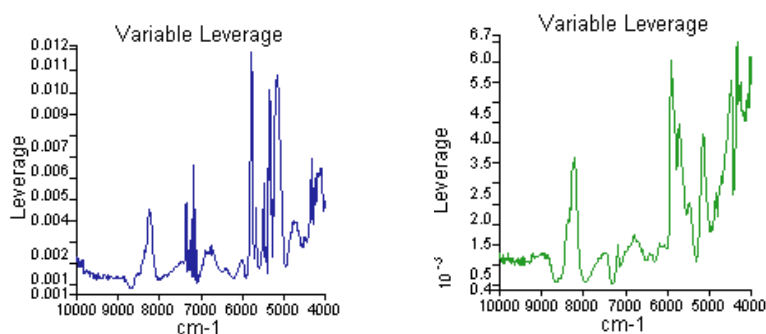


Rys.56 Widma surowe dwóch surowców: niebieski – alkohol cetearylowy, zielony- alkohol lanolinowy

5.2.1.1. Wybór zakresu analizy

Wybór zakresu analizy dokonano na podstawie tzw. „wykresu dźwigni” (ang. *Variable Leverage V-L*). Wykres ten przedstawia wpływ absorbancji przy danej liczbie falowej na zmienność zawartą w głównych składowych dla widm badanego materiału i pozwala określić zakres liczb falowych odpowiadających największej zmienności reprezentującej przez widma danej substancji. Kształt i stopień komplikacji wykresów dźwigni zależy przy tym od liczby uwzględnionych głównych składowych. Na **Rys.57**

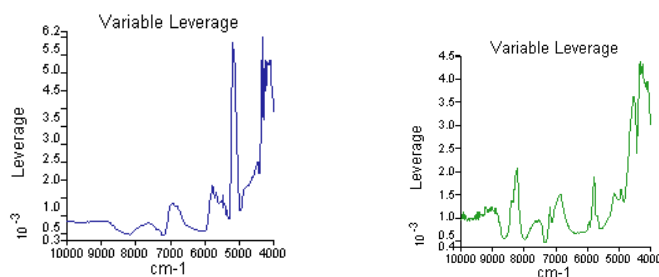
przedstawione zostały wykresy dźwigni dla obu surowców dla 6 istotnych czynników (liczba składowych wyznaczonych przez program dostarczony przez producenta sprzętu).



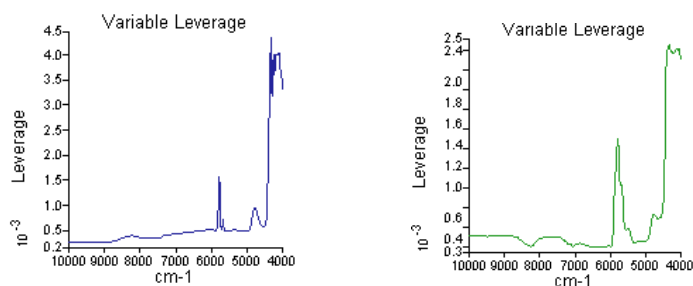
Rys.57 Wykresy V-L dla surowców : niebieski – alkohol cetearylowy, zielony- alkohol lanolinowy, otrzymane przy 6 istotnych czynnikach

Analizując te wykresy można stwierdzić, że najsilniejsze piki (pasma) widoczne są w zakresie od $6000\text{-}5000\text{ cm}^{-1}$ oraz $9000\text{-}8000\text{ cm}^{-1}$. Jednakże, jak wykazałam w p.5.1.1.2 program ten ma tendencję do uwzględniania zbyt dużej liczby czynników. Prowadzi to do uwzględnienia na wykresie dźwigni nie tylko danych spektralnych, ale również zmienności niepożądaną - szumu fotometrycznego. Wykonano więc również wykresy dźwigni dla mniejszej liczby czynników: 4 (**Rys.58a**) i 2 (**Rys.58b**).

a)



b)



Rys.58 Wykresy V-L otrzymane dla dwóch surowców badanych a) przy 4 istotnych czynnikach b) przy 2 istotnych czynnikach. Kolory jak na Rys.57

Z wykresów tych wyznaczono optymalny dla analizy jakościowej zakres liczb falowych ($6000\text{-}5500\text{ cm}^{-1}$).

5.2.1.2. Podział widm

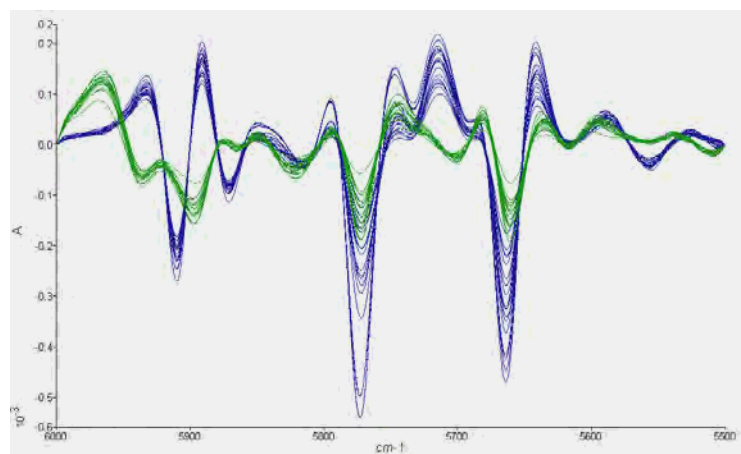
Na tym etapie analizy widma badanych surowców zostały podzielone na trzy rozłączne zbiory: zbiór uczący, walidacyjny oraz testowy. Do zbioru uczącego należy wybrać około 70-75% całkowitej liczby próbek z każdej klasy, a pozostałe dostępne próbki zaliczyć do zbioru walidacyjnego. Bardzo ważne jest, aby w zbiorze uczącym znalazły się próbki opisujące maksymalną wariancję w danej klasie. W zbiorze testowym natomiast powinny znaleźć się próbki nie znajdujące się w zbiorze uczącym i walidacyjnym. Zbiory walidacyjny oraz testowy służą do przeprowadzenia walidacji opracowanego modelu. Zbiór walidacyjny wykorzystywany jest do tzw. walidacji wewnętrznej, natomiast zbiór testowy do walidacji zewnętrznej. Na etapie walidacji wewnętrznej istnieje jeszcze możliwość zmiany warunków opracowywanych modeli (zmiana liczby głównych składowych, liczby próbek oraz metod transformacji).

W prezentowanym przykładzie do utworzenia zbioru uczącego wykorzystano 33 próbki (17 alkoholu lanolinowego, 16 alkoholu cetearylowego). W zbiorze walidacyjnym znalazło się więc 8 próbek (po 4 dla każdego surowca).

Dodatkowo przeprowadzono walidację wewnętrzną w oparciu o zestaw testowy.

5.2.1.3. Wstępna obróbka widm

Ze względu na to, że surowe widma są trudne w interpretacji poddano je wstępnej obróbce polegającej na obliczeniu ich 2. pochodnych. W **Tab.9** przedstawiono warunki obróbki wstępnej, a na **Rys.59** widma uzyskane po przeprowadzeniu transformacji danych.



Rys.59 Widma surowców po obliczeniu 2. pochodnych: zielony –alkohol lanolinowy, niebieski – alkohol cetearylowy

Tabela 9. Warunki analizy

Zakres analizy (cm^{-1})	6000-5500
Filtr	Ważenie szumów
Korekcja linii bazowej	2 pochodna 13 punktowe okno wielomianu
Normalizacja	brak

5.2.1.4. Analiza głównych składowych

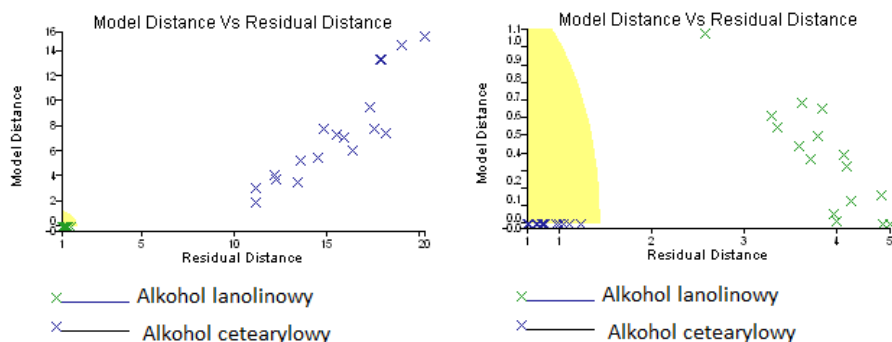
Na tym etapie dokonano wyboru liczby głównych składowych widm obu surowców w oparciu o parametry *Residual Distance* oraz *Model Distance*.

„*Residual Distance*” (RD) to parametr umożliwiający ocenę odległości widma od modelu danej klasy do której należy. Widma dla których wartość tego parametru przekracza tzw. granicę identyfikacji wyznaczoną przez klasyfikator SIMCA są niepoprawnie identyfikowane.

Parametr „*Model Distance*” (MD) podaje odległość widma od modelu danej klasy i wyznaczany jest dla wszystkich widm w zbiorze uczącym. Na podstawie wykresu zależności MD w funkcji RD widma należące do danej klasy powinny znaleźć się możliwie blisko początku układu. Wartość krytyczną odległości obrazuje żółty wycinek koła. Widma należące do innych klas powinny się znaleźć poza tym obszarem.

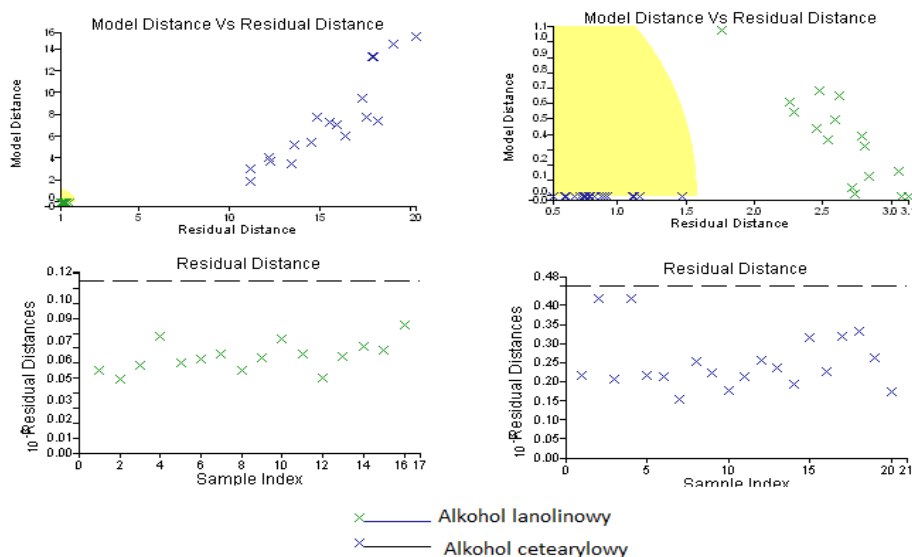
Program dostarczony przez producenta (AssureID-MethodExplorer) generuje zbyt dużą liczbę czynników, przyczyniając się do tego, że opracowywane modele tożsamościowe mogą działać niepoprawnie. Dobór liczby istotnych czynników należy więc dokonać „ręcznie” w oparciu o wykresy parametrów *RD* oraz *MD*.

Program uznał za istotne 3 główne składowe dla alkoholu lanolinowego oraz 4 składowe dla alkoholu cetearylowego. Na wykresach MD w funkcji RD (**Rys.60**) widać zdecydowane zróżnicowanie obu klas.



Rys. 60 Wykresy MD w funkcji RD otrzymane dla dwóch badanych surowców a) 3 czynniki dla alkoholu lanolinowego b) 4 czynniki dla alkoholu cetearylowego

Ponieważ nadmierna liczba czynników w modelach klas może prowadzić do utraty zdolności predykcyjnych modeli, więc postanowiono sprawdzić, czy zmniejszenie liczby czynników do dwóch zapewni w dalszym ciągu poprawną klasyfikację próbek ze zbioru uczącego. Wyniki przedstawione na **Rys. 61** pokazują że odległości między klasami nieznacznie zmalały, jednak nadal obserwuje się wyraźne rozdzielenie badanych surowców. Ponadto wartości parametru RD dla próbek z obu klasy ze zbioru uczącego nie przekraczają obszaru identyfikacji. Zatem na tej podstawie można uznać, że dla poprawnej klasyfikacji wystarczą modele wykorzystujące tylko dwa istotne czynniki.



Rys. 61 Wykresy MD w funkcji RD (górny wiersz) oraz wykresy RD (dolny wiersz) uzyskane dla modeli alkoholu lanolinowego (lewa kolumna) i alkoholu cetearylowego (prawa kolumna) przy dwóch czynnikach w modelach obu klas

5.2.1.5. Klasyfikacja SIMCA

Klasyfikator SIMCA w oparciu o 3 wyznaczone główne składowe stworzył indywidualne modele klas dla każdego surowca. Wyniki klasyfikacji przedstawiono poniżej w postaci **Tab.10**.

Tabela 10. Ocena poprawności klasyfikacji próbek ze zbioru uczącego dla modeli o 3 składowych

	Wynik klasyfikacji	
	<i>alkohol lanolinowy</i>	<i>alkohol cetearylowy</i>
<i>alkohol lanolinowy</i>	16	0
<i>alkohol cetearylowy</i>	0	17

Wszystkie próbki składające się na zbiór uczący zostały poprawnie sklasyfikowane. Dla pełnej oceny poprawności działania opracowanego modelu klasyfikacji należy potwierdzić jego zdolności prognostyczne. W tym celu wykorzystano zbiór walidacyjny.

5.2.1.6. Walidacja wewnętrzna oraz zewnętrzna

Opracowany na podstawie zbioru uczącego model klasyfikacji SIMCA wykorzystujący 3 główne składowe poprawnie zaklasyfikował również próbki należące do zbioru testowego i zbioru walidacyjnego, **Tab.11**. Zatem opracowany model może być stosowany w rutynowej analizie w celu identyfikacji i klasyfikacji dwóch badanych surowców.

Tabela 11. Ocena poprawności klasyfikacji a) na zbiorze walidacyjnym, b) zbiorze testowym

a)

Surowiec	Liczba próbek	Wynik klasyfikacji	
		<i>alk. lanolinowy</i>	<i>alk. cetearylowy</i>
<i>alkohol lanolinowy</i>	4	4	0
<i>alkohol cetearylowy</i>	4	0	4

b)

Surowiec	Liczba próbek	Wynik klasyfikacji	
		<i>alk. lanolinowy</i>	<i>alk. cetearylowy</i>
<i>alkohol lanolinowy</i>	5	5	0
<i>alkohol cetearylowy</i>	4	0	4

5.2.2. Wpływ transformacji danych oraz liczby istotnych czynników na zdolności predykcyjne opracowanych modeli

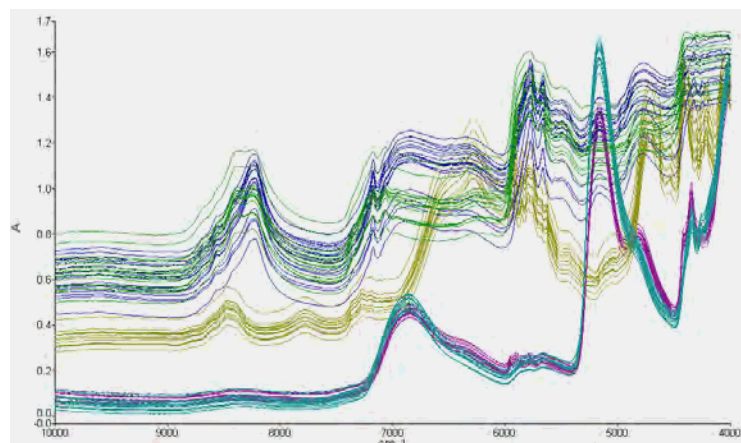
Wpływ transformacji danych oraz liczby istotnych czynników na zdolności predykcyjne modelu klasyfikacji został przebadany na modelu opracowanym dla 6 surowców stosowanych w firmie Ziaja Ltd.

Surowce wykorzystane do opracowania modelu

- glikol butylenowy
- ekstrakt glikolowy z ogórka
- alkohol lanolinowy
- alkohol cetearylowy
- 2-bromo-2-nitropropano-1,3-diol

Do opracowania powyższego modelu wykorzystano łącznie widma 107 próbek tych surowców. 87 próbek przypisano do zbioru uczącego, a 20 utworzyło zbiór walidacyjny.

Rys.62 przedstawia surowe widma tych próbek zarejestrowane w zakresie od 10 000 do 4 000 cm^{-1} .



Rys.62 Widma surowe 6 badanych surowców: zielony - alkohol lanolinowy, niebieski - alkohol cetearylowy, żółty - 2-bromo-2-nitropropano-1,3diol, różowy - ekstrakt z ogórka, morski - glikol butylenowy

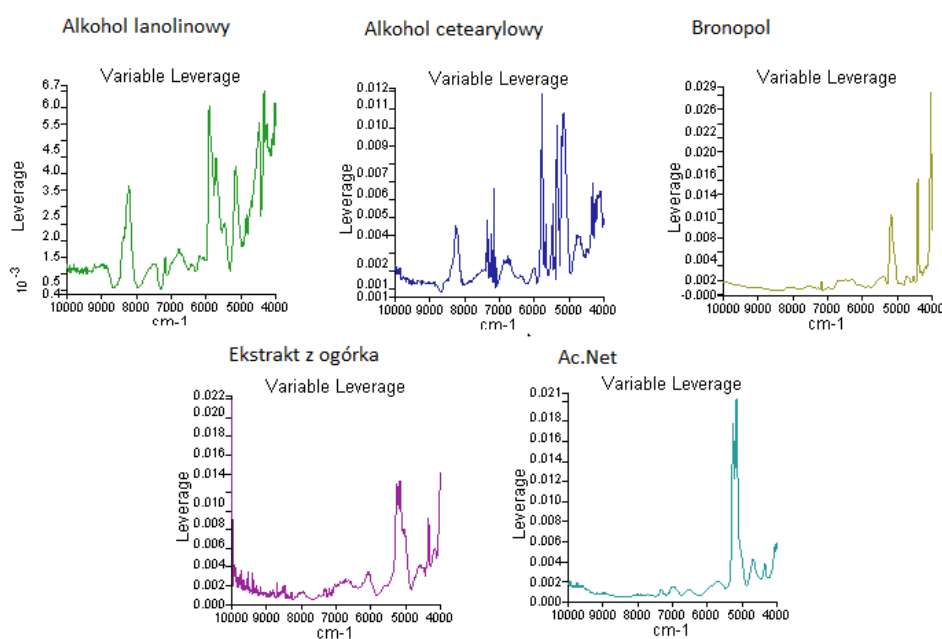
Na początku opracowywania modelu tożsamościowego. Wybór odpowiedniego zakresu analizy jest pierwszym etapem w opracowywaniu metod jakościowych. Źle dobrany zakres analizy może spowodować trudności w identyfikacji badanych surowców. Nawet bardzo dobrze dobrane metody pre-processingu nie dadzą wtedy zadawalających wyników. W niniejszym przykładzie dokonano wyboru zakresu analizy w oparciu o wykresy dźwigniowe.

Program AssureID-MethodExplorer wyznaczył dla surowych danych liczby istotnych głównych składowych w poszczególnych klasach przedstawione w **Tab.12**.

Tabela 12. Liczba głównych składowych wyznaczona przez program AssureID-MethodExplorer

Nazwa surowca	Liczba głównych składowych
<i>alkohol lanolinowy</i>	6
<i>alkohol cetearylowy</i>	8
<i>2-bromo-2-nitropropano 1,3 diol</i>	6
<i>ekstrakt z ogórka</i>	8
<i>glikol butylenowy</i>	6

Rys.63 przedstawia wykresy dźwigniowe dla badanych surowców otrzymane w zakresie od 1 0000 – 4 000 cm^{-1} przy liczbie głównych składowych z **Tab.12**.

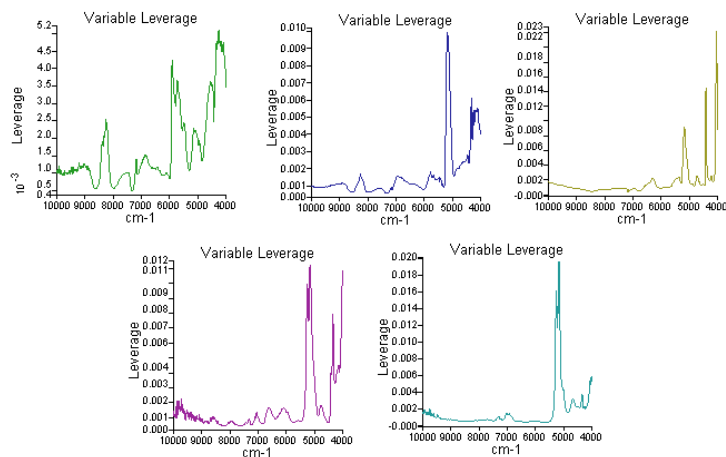


Rys.63 Wykresy dźwigniowe (V-L) uzyskane dla poszczególnych surowców : zielony - alkohol lanolinowy, niebieski - alkohol cetearylowy, żółty - 2-bromo-2-nitropropano-1,3-diol (bronopol), różowy – ekstrakt z ogórka, morski – glikol butylenowy przy liczbie czynników z **Tab.12**

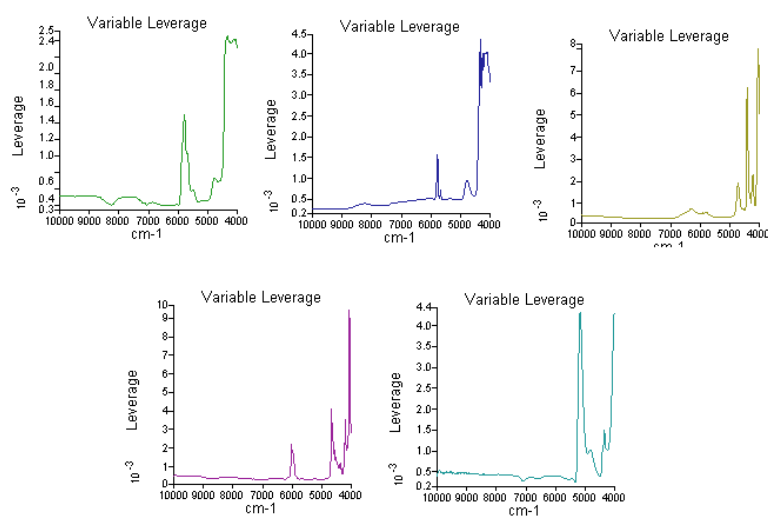
Analiza wykresów dźwigniowych z **Rys.63** jest trudna, ale jak wykazano w p.5.2.1.1 staje się dużo łatwiejsza po uwzględnieniu tylko kilku pierwszych składowych. Na **Rys.64** pokazano wykresy dźwigniowe dla 5 i 2 głównych składowych. Przy uwzględnieniu tylko 2 składowych można stwierdzić, że wyraźne piki występują dla alkoholu lanolinowego tylko w zakresie 6 000-4 500 cm^{-1} , natomiast w przypadku pozostałych surowców ostre

pasma widoczne są w zakresie poniżej $5\ 500\ \text{cm}^{-1}$. Zatem niższy model tożsamościowy opracowany będzie w zakresie od $6\ 000\text{--}4\ 500\ \text{cm}^{-1}$

5PC



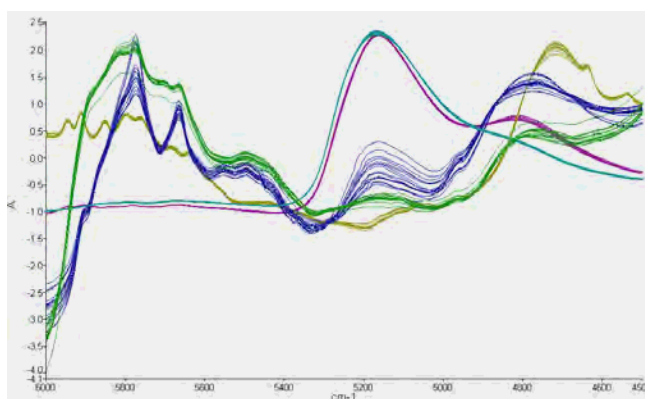
2PC



Rys.64 Wykresy dźwigniowe (V-L) uzyskane dla 5 surowców : zielony - alkohol lanolinowy, niebieski - alkohol cetearylowy, żółty – 2-bromo-2-nitroporopano-1,- diol, różowy - ekstrakt z ogórka, morski - glikol butylenowy przy 5 oraz 2 głównych składowych

Po wyborze optymalnego zakresu analizy rozpoczęto pracę nad wyborem najlepszej metody transformacji surowych widm oraz optymalnej liczby głównych składowych w modelach poszczególnych klas. W pierwszym etapie widma zostały poddane normalizacji SNV oraz MSC.

a) Normalizacja SNV



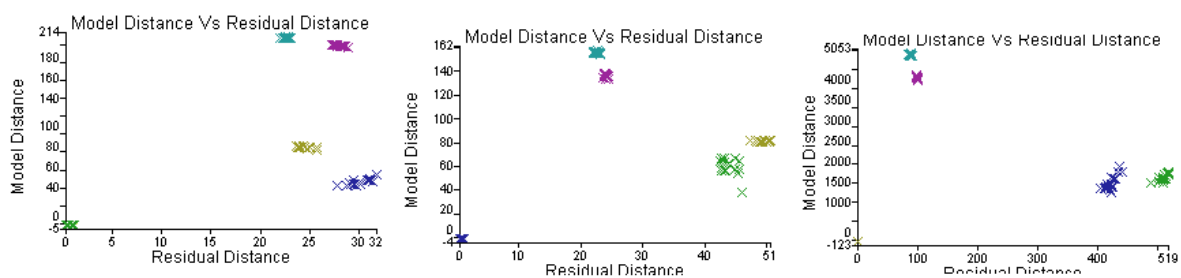
Rys. 65 Widma NIR 5 badanych surowców w zakresie od 6 000 do 4500 cm^{-1} otrzymane po transformacji SNV. Kolory jak na Rys. 62

Liczba głównych składowych wyznaczona przez program dla 5 badanych substancji została przedstawiona w **Tab.13**.

Tabela 13. Liczba głównych składowych wyznaczona przez program AssureID-MethodExplorer dla widm poddanych transformacji SNV

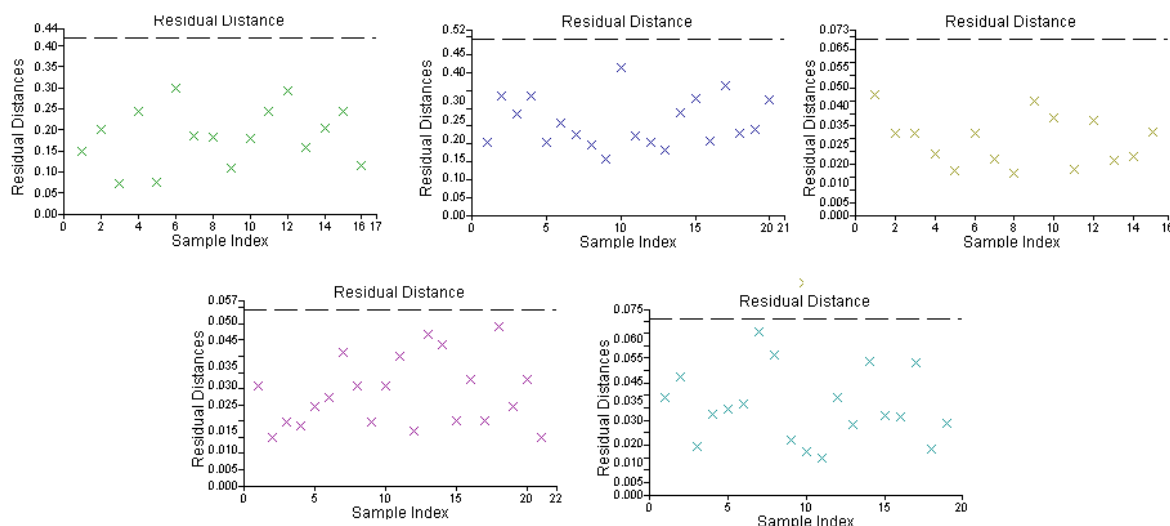
Nazwa surowca	Liczba głównych składowych
alkohol lanolinowy	7
alkohol cetearylowy	8
2-bromo-2-nitropropano-1,3-diol	6
ekstrakt z ogórka	9
glikol butylenowy	9

Na podstawie wykresów zależności MD w funkcji RD dokonano wstępnej oceny zdolności klasyfikacyjnych opracowanego modelu, **Rys.66**.



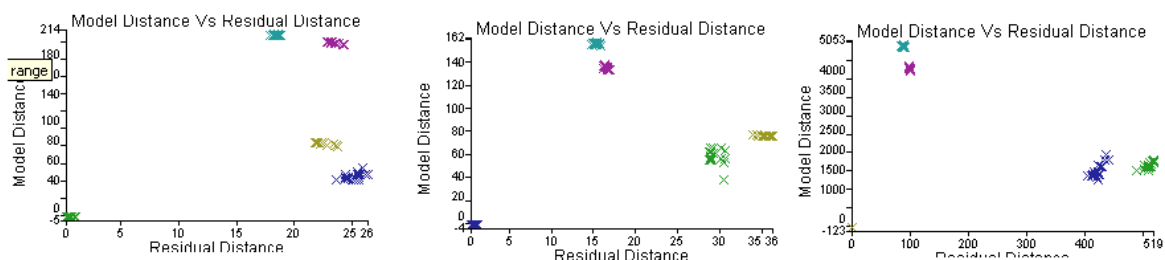
Rys. 66 Wykresy MD w funkcji RD dla 3 pierwszych badanych surowców na podstawie liczby głównych składowych z Tab.13

Dla wszystkich surowców próbki wszystkich klas mieszczą się w zakresie identyfikacji. Jednakże duża liczbą głównych składowych w modelach klas stwarza ryzyko „przeuczenia” modelu. Zatem podjęto próbę wyznaczenia minimalnej liczby czynników w każdej z klas pozwalającą na poprawną klasyfikację. W pierwszym kroku zmniejszono liczbę głównych składowych do 6 w każdej klasie.



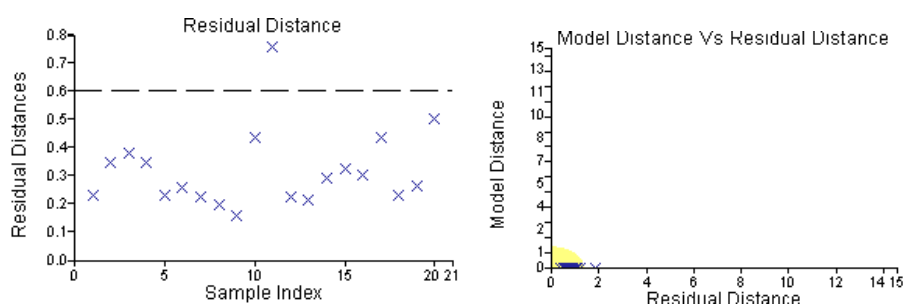
Rys. 67 Wykresy RD dla modelu zawierającego po 6 głównych składowych w każdej klasie

Analizując wykresy RD przedstawione na **Rys. 67** można stwierdzić, że wszystkie widma mieszczą się w przedziale identyfikacji. Jednocześnie z **Rys. 68** wynika, że również odległości między klasami są bardzo dobre.



Rys. 68 Wykresy MD vs. RD dla wybranych klas. Od lewej: alkohol lanolinowy, alkohol cetearyłowy i 2-bromo-2-nitropropano-1,3-diol

Przy 5 głównych składowych w każdej klasie zaobserwowano, że dla alkoholu cetearyłowego jedna próbka lokuje się poza granicą identyfikacji.



Rys.69 Wykresy RD oraz MD vs. RD otrzymane dla alkoholu cetearylowego przy 5składowych

Natomiast przy 4 czynnikach obserwuje się zwiększenie liczby surowców dla których niektóre próbki przekraczają granicę identyfikacji do trzech (alkohol lanolinowy, alkohol cetearylowy, ekstrakt z ogórka).

Powyższe wyniki potwierdzają wniosek z p.5.2.1.4: do poprawnej klasyfikacji zbioru uczącego mogą być z powodzeniem zastosowane modele klas o mniejszym stopniu komplikacji (mniejszej liczbie głównych składowych) niż sugeruje to program AssureID-MethodExplorer. Podjęto również próbę oceny zdolności prognostycznych modelu przy różnej liczbie uwzględnianych składowych. Użyto w tym celu zbioru walidacyjnego, Tab.14.

Tabela 14. Wyniki klasyfikacji próbek ze zbioru walidacyjnego przy różnej liczbie uwzględnianych składowych na podstawie widm NIR poddanych transformacji SNV. W polach tabeli podano liczbę próbek poprawnie sklasyfikowanych/ogólną liczbę próbek w klasie

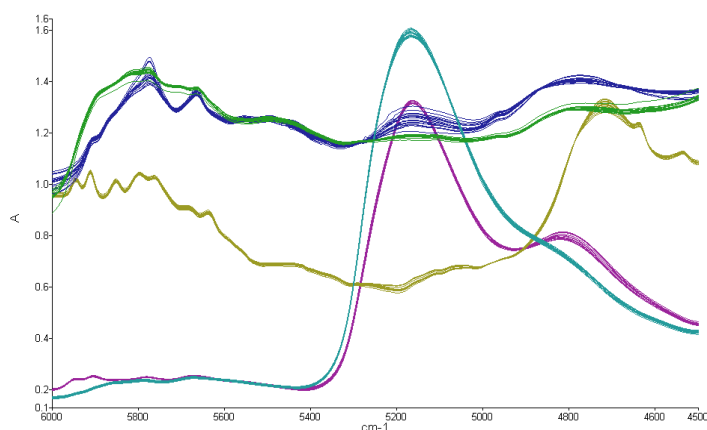
Nazwa surowca	Liczba składowych w modelu							
	8/9PC	7PC	6PC	5PC	4PC	3PC	2PC	1PC
alkohol lanolinowy	-	4/4	4/4	4/4	4/4	3/4	3/4	3/4
alkohol cetearylowy	4/4	-	4/4	4/4	2/4	2/4	2/4	3/4
2-bromo-2-nitropropan-1,3-diol		-	6/6	6/6	6/6	6/6	4/6	4/6
ekstrakt z ogórka	3/3	3/3	3/3	3/3	3/3	2/3	2/3	1/3
glikol butylenowy	4/4	4/4	4/4	4/4	4/4	4/4	3/4	3/4

Podsumowując powyższe wyniki można stwierdzić, że program AssureID-MethodExplorer wyznacza zbyt dużą liczbę czynników dla klasyfikatora SIMCA. W takiej sytuacji w opracowywanym modelu uwzględniana jest nie tylko zmienność systematyczna wynikająca ze zróżnicowania próbek badanych substancji, ale także zmienność niepożądana (szумы losowe). W prezentowanym przykładzie nie zaobserwowano

„przeuczenia” modelu, ale nie można tego wykluczyć w przypadku innego doboru próbek do zbioru walidacyjnego.

b) Normalizacja MSC

Rys.70 przedstawia widma 5 badanych surowców poddane transformacji SNV. W efekcie zastosowania tej transformacji uzyskano dużą wizualną zgodność kształtu i intensywności widm próbek w obrębie poszczególnych surowców. Jednakże program AssureID-MethodExplorer również w przypadku tej transformacji sugeruje użycie nadspodziewanie dużą liczbę istotnych składowych, **Tab.15**.



Rys.70 Widma NIR 5 badanych surowców w zakresie od 6 000 do 4500 cm^{-1} otrzymane po transformacji MSC. Kolory jak na Rys.62

Tabela 15. Liczba głównych składowych wyznaczona przez program AssureID-MethodExplorer dla widm poddanych transformacji MSC

Nazwa surowca	Liczba głównych składowych
<i>alkohol lanolinowy</i>	7
<i>alkohol cetearylowy</i>	8
<i>2-bromo-2-nitropropan-1,3-diol</i>	6
<i>ekstrakt z ogórka</i>	9
<i>glikol butylenowy</i>	9

Przy tym stopniu komplikacji modelu (liczbie głównych składowych w modelach poszczególnych klas) nie występują punkty poza granicami identyfikacji, a wszystkie próbki zarówno ze zbioru uczącego jak i walidacyjnego zostają poprawnie sklasyfikowane. Podobnie jak w przypadku transformacji SNV również w tym przypadku po zmniejszeniu liczby czynników zaczęły się pojawiać próbki zbioru uczącego przekraczające granicę identyfikacji. Ponadto zaobserwowano, że przy 6 składowych mimo że próbki ze zbioru

uczącego nie przekraczały granicy identyfikacji, to niektóre próbki alkoholu lanolinowego ze zbioru walidacyjnego nie zostały poprawnie sklasyfikowane. Dalsze upraszczanie modelu znacząco obniżało jego zdolności prognostyczne, **Tab.16**. Na podstawie uzyskanych wyników można stwierdzić, że transformacja MSC nie przyniósł pożądanego efektu, tzn. nie obniżyła w znaczący sposób stopnia komplikacji modelu w porównaniu z transformacją SNV.

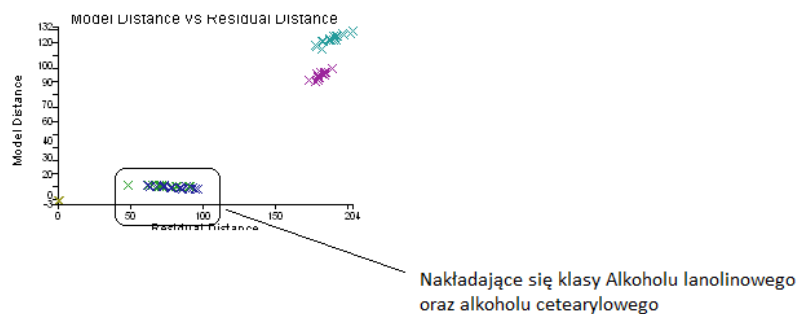
Tabela 16. Wyniki klasyfikacji próbek ze zbioru walidacyjnego przy różnej liczbie uwzględnianych składowych na podstawie widm NIR poddanych transformacji SNV. Zawartość pól tabeli analogiczna jak w Tab.13

Nazwa surowca	Liczba składowych w modelu							
	8/9PC	7PC	6PC	5PC	4PC	3PC	2PC	1PC
alkohol lanolinowy	-	4/4	4/4	3/4	4/4	3/4	3/4	3/4
alkohol cetearylowy	4/4	-	4/4	4/4	2/4	3/4	2/4	3/4
2-bromo-2-nitropropan- 1,3 diol	-	-	6/6	6/6	6/6	6/6	4/6	4/6
ekstrakt z ogórka	3/3	3/3	3/3	3/3	3/3	2/3	2/3	2/3
glikol butylenowy	4/4	4/4	4/4	4/4	4/4	4/4	3/4	3/4

Po przeprowadzeniu normalizacji na surowych danych rozpoczęto badanie wpływu korekty linii bazowej na zdolności predykcyjne modeli tożsamościowych.

c) Transformacja offset

Dla widm poddanych transformacji offset uzyskano z wykorzystaniem programu AssureID-MethodExplorer model o analogicznym stopniu komplikacji jak po transformacji MSC. Model ten poprawnie sklasyfikował próbki ze zbioru uczącego i walidacyjnego. Jednakże próba obniżenia stopnia komplikacji modelu wykazała, że już dla 6 składowych klasyfikator SIMCA nie jest w stanie poprawnie sklasyfikować próbek alkoholu lanolinowego oraz alkoholu cetearylowego ze zbioru uczącego, **Rys.71**.



Rys. 71 Wykres MD vs. RD obrazuje nakładanie się próbek alkoholu lanolinowego i cetearylowego po zastosowaniu transformacji offset dla modelu o 6 składowych

Również próbki tych dwóch surowców ze zbioru walidacyjnego były niepoprawnie klasyfikowane, **Tab.17**.

Tabela 17. Wyniki klasyfikacji próbek ze zbioru walidacyjnego przy różnej liczbie uwzględnianych składowych na podstawie widm NIR poddanych transformacji offset. Zawartość pól tabeli analogiczna jak w Tab.15

Nazwa surowca	Liczba składowych w modelu							
	8/9PC	7PC	6PC	5PC	4PC	3PC	2PC	1PC
alkohol lanolinowy	-	4/4	3/4	3/4	4/4	3/4	3/4	3/4
alkohol cetearylowy	4/4	-	2/4	1/4	2/4	3/4	2/4	3/4
2-bromo2-nitropropan- 1,3diol	-	-	6/6	6/6	6/6	6/6	4/6	4/6
ekstrakt z ogórka	3/3	3/3	3/3	3/3	3/3	2/3	2/3	2/3
glikol butylenowy	4/4	4/4	4/4	4/4	4/4	4/4	3/4	3/4

d) Transformacja Curve

Po zastosowaniu transformacji Curve zaobserwowano zmniejszenie się liczby istotnych czynników wyznaczonych przez program AssureID-MethodExplorer jedynie dla 2-bromo-2-nitropropano-1,3-diolu (**tab. 18**).

Tabela 18. Liczba głównych składowych wyznaczona przez program AssureID-MethodExplorer dla widm poddanych transformacji Curve

Nazwa surowca	Liczba głównych składowych
alkohol lanolinowy	6
alkohol cetearylowy	7
2-bromo-2- nitropropan- 1,3- diol	6
ekstrakt z ogórka	7
glikol butylenowy	9

Przy próbach „ręcznego” obniżenia stopnia komplikacji modelu zaobserwowano problemy z klasyfikacją próbek alkoholu lanolinowego oraz alkoholu cetearylowego zarówno ze zbioru uczącego jak i walidacyjnego. Otrzymane wyniki pokazują zatem, że ten typ transformacji nie przynosi znaczącego obniżenia stopnia komplikacji modelu

Tabela 19. Wyniki klasyfikacji próbek ze zbioru walidacyjnego przy różnej liczbie uwzględnianych składowych na podstawie widm NIR poddanych transformacji Curve. Zawartość pól tabeli analogiczna jak w **tab.15**

Nazwa surowca	Liczba składowych w modelu							
	8/9PC	7PC	6PC	5PC	4PC	3PC	2PC	1PC
alkohol lanolinowy	-	4/4	3/4	3/4	4/4	3/4	3/4	3/4
alkohol cetearylowy	4/4	-	2/4	1/4	2/4	3/4	2/4	3/4
2-bromo-2-nitropropan- 1,3- diol	-	-	6/6	6/6	6/6	6/6	6/6	6/6
ekstrakt z ogórka	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3
glikol butylenowy	4/4	4/4	4/4	4/4	4/4	4/4	3/4	4/4

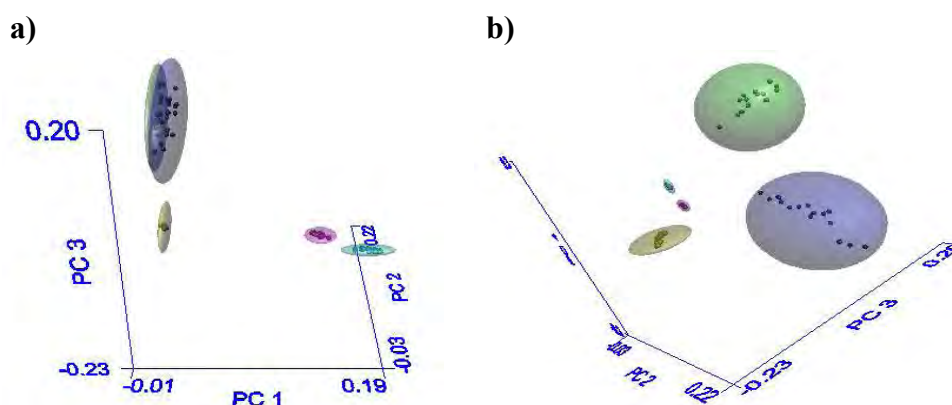
e) 1 pochodna

Po transformacji widm do ich pierwszych pochodnych uzyskano zdecydowanie mniejsze liczby głównych składowych wyznaczonych przez program AssureID-MethodExplorer niż w przypadku poprzednio omówionych transformacji, **Tab.20**.

Tabela 20. Liczba głównych składowych wyznaczona przez program AssureID-MethodExplorer dla widm przekształconych w 1. pochodne

Nazwa surowca	Liczba głównych składowych
alkohol lanolinowy	3
alkohol cetearylowy	5
2-bromo-2- nitropropan- 1,3 -diol	4
ekstrakt z ogórka	5
glikol butylenowy	5

Po „ręcznym” zmniejszeniu dla alkoholu cetearylowego, ekstraktu z ogórka oraz glikolu butylenowego liczby składowych do 4 nie zaobserwowano pogorszenia jakości modelu: odległości między wszystkimi klasami badanych surowców były zadawalające, **Rys.73**, nie zaobserwowano próbek ze zbioru uczącego będących poza granicami identyfikacji, a wszystkie próbki ze zbioru walidacyjnego zostały poprawnie sklasyfikowane.



Rys.72 Rzut przedstawiający rozmieszczenie widm próbek badanych surowców po ich przeliczeniu na 1. pochodne w przestrzeniach: a) 3 pierwszych głównych składowych b) 2, 3 i 4 składowej

Dalsze uproszczenie modelu doprowadziło jednak do utraty zdolności prognostycznej: klasyfikator nie był w stanie poprawnie sklasyfikować wszystkich próbek alkoholu lanolinowego i cetearylowego, **Tab.21**.

Tabela 21. Wyniki klasyfikacji próbek ze zbioru walidacyjnego przy różnej liczbie uwzględnianych składowych na podstawie widm NIR przeliczonych na ich 1. pochodne. Zawartość pól tabeli analogiczna jak w Tab.15

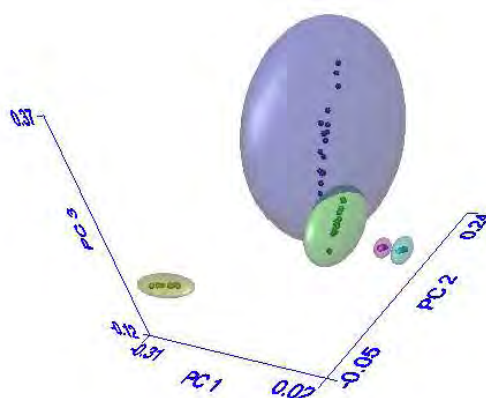
Nazwa surowca	Liczba składowych w modelu				
	5PC	4PC	3PC	2PC	1PC
alkohol lanolinowy	-	4/4	3/4	3/4	3/4
alkohol cetearylowy	4/4	4/4	3/4	2/4	2/4
2-bromo-2- nitropropan - 1,3- diol	-	6/6	6/6	5/6	6/6
ekstrakt z ogórka	3/3	3/3	3/3	3/3	3/3
glikol butylenowy	4/4	4/4	4/4	3/4	4/4

f) 2 pochodna

Po transformacji widm do ich drugich pochodnych program AssureID-MethodExplorer wyznaczył jeszcze mniejszą liczbę istotnych czynników, **Tab.22**. Jednak analizując rozmieszczenie próbek badanych surowców w przestrzeni 3 pierwszych składowych obserwuje się nakładanie modeli klas alkoholu lanolinowego oraz cetearylowego, **Rys.73**.

Tabela 22. Liczba głównych składowych wyznaczona przez program AssureID-MethodExplorer dla widm przekształconych w 2. pochodne

Nazwa surowca	Liczba głównych składowych
alkohol lanolinowy	2
alkohol cetearylowy	4
2-bromo-2- nitropropan- 1,3- diol	2
ekstrakt z ogórka	3
glikol butylenowy	4

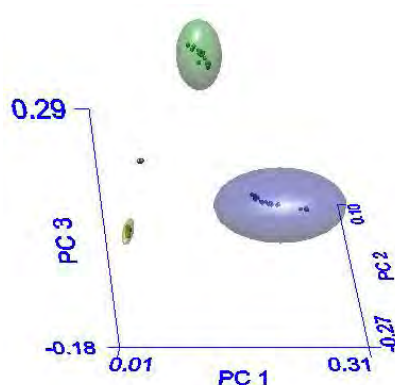


Rys.73 Rzut przedstawiający rozmieszczenie widm próbek badanych surowców po ich przeliczeniu na 2. pochodne w przestrzeni 3 pierwszych głównych składowych

Ze względu na to że ten typ transformacji wydaje się być najlepszym z wszystkich dotychczas przebadanych postanowiono połączyć ten rodzaj transformacji z normalizacją SNV. Uzyskane wyniki pokazują, **Tab.23**, że wcześniej już mała liczba istotnych czynników uległa dalszemu zmniejszeniu. Ponadto zaobserwowano lepsze rozdzielanie próbek alkoholu lanolinowego oraz cetarylowego, **Rys.74**. Zaobserwowano jednak zmniejszenie się odległości pomiędzy próbkami surowców płynnych (ekstrakt z ogórka, glikol butylenowy).

Tabela 23. Liczba głównych składowych wyznaczona przez program AssureID-MethodExplorer dla widm przekształconych w 2. pochodne i poddanych normalizacji SNV

Nazwa surowca	Liczba głównych składowych
alkohol lanolinowy	2
alkohol cetearylowy	4
2-bromo-2- nitropropan- 1,3- diol	2
ekstrakt z ogórka	3
glikol butylenowy	4



Rys74. Rzut przedstawiający rozmieszczenie widm próbek badanych surowców po ich przeliczeniu na 2. pochodne i normalizacji SNV w przestrzeni 3 pierwszych głównych składowych

W **Tab.24** przedstawiono wyniki klasyfikacji próbek zbioru walidacyjnego. Pokazują one, że połączenie 2. pochodnej i normalizacji SNV prowadzi do modeli o niskim stopniu komplikacji i dobrych zdolnościach prognostycznych. Na uwagę zasługuje fakt, że po takiej łączonej transformacji w przypadku niektórych surowców (alkohole lanolinowy i cetarylowy oraz 2-bromo-2-nitropropan-1,3-diol) dobre zdolności klasyfikacyjne i predykcyjne wykazują modele klas wykorzystujące tylko pierwszą główną składową.

Tabela 24. Wyniki klasyfikacji próbek ze zbioru walidacyjnego przy różnej liczbie uwzględnianych składowych na podstawie widm NIR przeliczonych na ich 2. pochodne i poddanych normalizacji SNV. Zawartość pól tabeli analogiczna jak w Tab.13

Nazwa surowca	Liczba składowych w modelu			
	4PC	3PC	2PC	1PC
alkohol lanolinowy	-	-	4/4	4/4
alkohol cetarylowy	4/4	4/4	4/4	4/4
2-bromo-2- nitropropan- 1,3- diol	-	-	6/6	6/6
ekstrakt z ogórka	-	-	3/3	2/3
glikol butylenowy	-	4/4	3/4	2/4

Otrzymane wyniki pokazują wyraźnie, że rodzaj zastosowanej transformacji widm surowych ma zdecydowany wpływ na stopień komplikacji modeli klasyfikacji (liczbę składowych niezbędnych dla zdefiniowania poszczególnych klas), a tym samym na zdolności predykcyjne opracowywanych modeli.

Opracowując model tożsamościowy z reguły należy dążyć do uzyskania modelu o możliwie niskim stopniu komplikacji. Modele takie charakteryzują się większą stabilnością i mniejszą wrażliwością na przypadkowe zmiany widm. Konieczność

zastosowania modelu o wysokim stopniu komplikacji świadczy, że etap transformacji widm surowych nie został poprawnie przeprowadzony. Dla zapewnienia stabilności opracowywanego modelu wskazany jest w takim przypadku powrót do etapu transformacji widm surowych i dobranie lepszego sposobu transformacji. W wielu przypadkach skuteczne jest kolejne przeprowadzenie dwóch odmiennych typów transformacji (p.5.2.3).

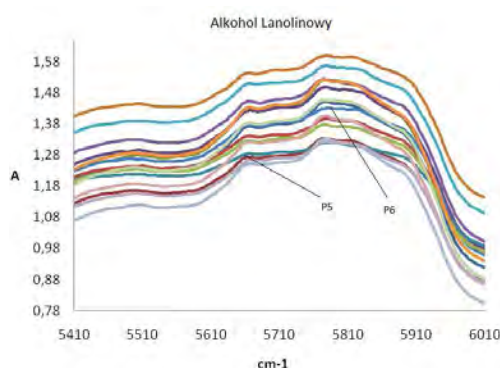
Dobór poprawnego typu transformacji jest żmudnym i pracochłonnym procesem. Wymaga od osoby opracowującej modele tożsamościowe wiedzy i doświadczenia. W rozdziale 5.2.3 zostanie przedstawiona ocena przydatności różnych technik transformacji widm do minimalizacji zmienności pochodzącej z różnych źródeł.

5.2.3. Określenie przydatności poszczególnych technik transformacji do minimalizacji zmienności pochodzącej z różnych źródeł

W chwili obecnej w literaturze przedmiotu brak jest ogólnie przyjętych zasad doboru metod transformacyjnych i większość badaczy stosuje metodę prób i błędów. Jest to podejście bardzo pracochłonne. Wymaga ponadto dogłębnej znajomości zasad działania stosowanych metod transformacyjnych i dużego doświadczenia z zakresu chemometrii i spektroskopii NIR. Jest to obecnie jeden z istotnych czynników ograniczających zastosowanie tej wartościowej metody analitycznej w laboratoriach przemysłowych. W niniejszym podrozdziale przedstawiona zostanie propozycja procedury doboru zestawu technik chemometrycznych zapewniających możliwie optymalną obróbkę wstępną widm NIR

5.2.3.1. Dobór metod technik chemometrycznych stosowanych podczas transformacji surowych widm

Dobór możliwie najlepszych technik chemometrycznych zostanie przedstawiony na przykładzie modelu tożsamościowego opracowanego dla jednego surowca – Alkoholu lanolinowego.



Rys. 75 Widma surowe próbek alkoholu lanolinowego

Do opracowania modelu tożsamościowego wykorzystano 16 widm badanego surowca zarejestrowanych w zakresie 6 010-5 410 cm^{-1} . W oparciu o analizę chemometryczną przedstawioną w rozdziale 5 (Analiza głównych składowych, transformacja surowych widm), próbki oznaczone nr 5 (linia jasnoniebieska) oraz 6 (linia ciemnoniebieska) (**Rys.75**) uznane zostały za punkty odbiegające i umieszczono je w zbiorze walidacyjnym. Pozostałe 14 próbek umieszczono w zbiorze uczącym i przeprowadzono różnorodne transformacje surowych danych.

Tabela 25. Wyniki klasyfikacji próbek P5 i P6 ze zbioru walidacyjnego: *FAIL* – próbka nie należy do klasy, *PASS* – próbka należy do klasy

Rodzaj zastosowanej transformacji	Liczba składowych w modelu							
	4PC		3PC		2PC		1PC	
	P5	P6	P5	P6	P5	P6	P5	P6
<i>Transformacja SNV</i>	FAIL	FAIL	FAIL	FAIL	FAIL	FAIL	FAIL	PASS
<i>Transformacja MSC</i>	FAIL	FAIL	FAIL	FAIL	FAIL	FAIL	FAIL	PASS
<i>Transformacja Offset</i>	FAIL	FAIL	FAIL	FAIL	FAIL	FAIL	FAIL	PASS
<i>Transformacja Curve</i>	-	-	FAIL	PASS	FAIL	PASS	FAIL	PASS
<i>Transformacja Slope</i>	FAIL	PASS	FAIL	PASS	FAIL	PASS	PASS	PASS
<i>1 pochodna</i>	-	-	FAIL	PASS	FAIL	PASS	PASS	PASS
<i>2 pochodna</i>	-	-	FAIL	PASS	FAIL	PASS	PASS	PASS

Wyniki klasyfikacji próbek P5 i P6 tworzących zbiór walidacyjny przedstawiono w **tab.25**. Pokazują one odmienne zachowanie się obu próbek. Próbka P6 po zastosowaniu transformacji Curve, i Slope oraz po przekształceniu widm w ich 1. lub 2. pochodne jest poprawnie klasyfikowana przez modele zbudowane na podstawie od 4 do 1 głównej składowej. Wskazuje to, że widmo próbki P6 różni się od widm próbek ze zbioru uczącego przede wszystkim innym globalnym kształtem linii bazowej, a rozbieżność ta może zostać bez trudu zniwelowana przez zastosowanie odpowiedniej transformacji usuwającej efekty związane ze zmiennością linii bazowej.

Natomiast w przypadku próbki P5 obserwuje się jej poprawną identyfikację tylko w modelu zawierającym 1. główną składową i po zastosowaniu transformacji Slope lub po przekształceniu widm w 1. lub 2. pochodną. Wyniki te wskazują, że widmo próbki P5 zawiera dodatkowe elementy o lokalnym charakterze nieusuwalne przez żadną z zastosowanych metod transformacji. Próbkę tą należy więc uznać za rzeczywisty punkt odbiegający.

W celu potwierdzenia niniejszych wniosków widma próbek P5 i P6 zdublowano i umieszczono zarówno w zbiorze uczącym jak i walidacyjnym.

Tabela 26. Wyniki klasyfikacji próbek P5 i P6 ze zbioru walidacyjnego za pomocą modelu otrzymanego na podstawie zbioru uczącego zawierającego widma próbek P5 i P6

Rodzaj zastosowanej transformacji	Liczba składowych w modelu									
	6/5PC		4PC		3PC		2PC		1PC	
	P5	P6	P5	P6	P5	P6	P5	P6	P5	P6
<i>T. SNV</i>	PASS	PASS	FAIL	PASS	PASS	PASS	PASS	FAIL	FAIL	PASS
<i>T. MSC</i>	PASS	PASS	PASS	PASS	PASS	FAIL	PASS	FAIL	FAIL	PASS
<i>T. Offset</i>	PASS	PASS	PASS	FAIL	PASS	PASS	FAIL	PASS	PASS	PASS
<i>T. Curve</i>	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS
<i>T. Slope</i>	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS
<i>1 pochodna</i>	-	-	-	-	PASS	PASS	PASS	PASS	PASS	PASS
<i>2 pochodna</i>	-	-	-	-	PASS	PASS	PASS	PASS	PASS	PASS

Po dodaniu próbek P5 i P6 do zbioru uczącego zauważono, że program AssureID-MethodExplorer dla większości zastosowanych funkcji transformujących zbudował modele klasyfikacyjne zawierające zdecydowanie większą liczbę istotnych składowych (porównaj **Tab.25** i **Tab.26**). Wskazuje to wyraźnie, że zmienność niesiona przez te próbki na tyle różni się od zmienności zawartej w widmach pozostałych próbek, że jej uwzględnienie wymaga użycia dodatkowych składowych. Jedynie w przypadku przekształcenia widm w ich 1. lub 2. pochodne obserwuje się taką samą liczbę istotnych czynników. Techniki transformacji widm korygujące efekty związane z kształtem i przesunięciem linii bazowej (transformacje offset, Curve i Slope) mają efekt globalny, podczas gdy użycie pochodnych pozwala również na eliminację efektów lokalnych. Wyniki przedstawione w **Tab. 26** potwierdzają przypuszczenie, że widmo co najmniej jednej z dodanych próbek zawierało dodatkową zmienność o charakterze lokalnym.

Analizując wykresy RD uzyskane po wprowadzeniu do zbioru uczącego widma próbki P6 (dane nie zamieszczone w tej pracy) można stwierdzić, że najlepsze wyniki uzyskuje się po przekształceniu widm w ich 2. pochodne przy wszystkich analizowanych modelach o różnej liczbie składowych. po wprowadzeniu do zbioru uczącego widma próbki P5 najlepsze wyniki uzyskuje się również dla 2. pochodnej, ale tylko dla modelu opartego na dwóch istotnych czynnikach.

Pokazany przykład dobrze ilustruje zasadę mówiącą, że warunkiem poprawności modelu klasyfikacji jest umieszczenie w zbiorze uczącym widm zawierających wszystkie

spotykane w praktyce źródła zmienności. Przykład ten sugeruje również tryb postępowania z próbkami o nietypowych widmach. Próbki takie należy kontrolować i po potwierdzeniu ich tożsamości odpowiadające im widma umieszczać w zbiorze uczącym. Mogą one nieść bardzo cenne informacje i jednocześnie uchronić opracowywany model przed niepoprawnym klasyfikowaniem próbek nieznanymi.

5.2.3.2. Ocena wpływu normalizacji na zdolności predykcyjne modeli opracowanych dla surowców stałych oraz surowców płynnych

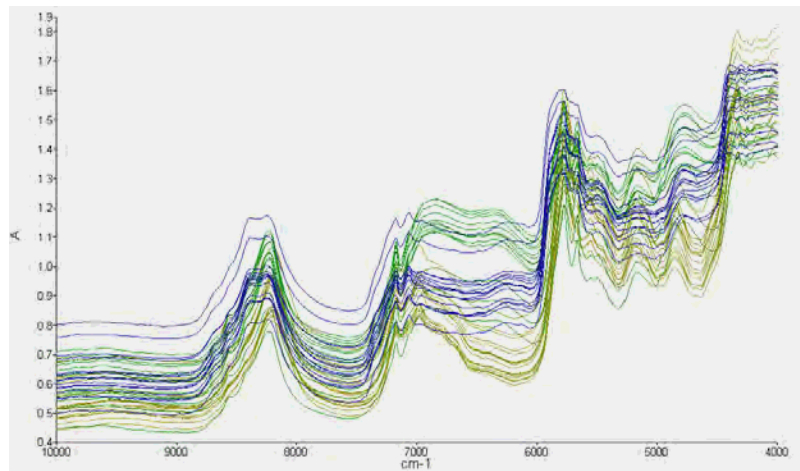
Intensywność pasm absorpcji jest wprost proporcjonalna do długości drogi optycznej jaką pokonuje światło przechodząc przez próbkę. Różnice w długości drogi optycznej mogą pochodzić od wielkości cząstek substancji badanych oraz stopnia upakowania naczynka pomiarowego. Różnice te można zniwelować poprzez zastosowanie transformacji SNV lub MSC.

Stosowanie normalizacji jako jednej z technik transformujących jest bardzo często przydatna w przypadku surowców stałych. Natomiast w przypadku surowców płynnych nie zawsze udaje się uzyskać lepsze wyniki. Niekiedy po zastosowaniu transformacji SNV czy też MSC dla surowców płynnych można uzyskać pogorszenie zdolności predykcyjnych modeli,(przypuszczenia takie zostały wyciągnięte na podstawie wyników własnych).

Poniżej zostanie przedstawiony wpływ normalizacji na zdolności predykcyjne dwóch modeli tożsamościowych: jednego dla surowców stałych (**Model 1 - Rys.76**) i drugiego dla surowców płynnych (**Model 2 – Rys.77**).

Opracowane modele**Model 1:**

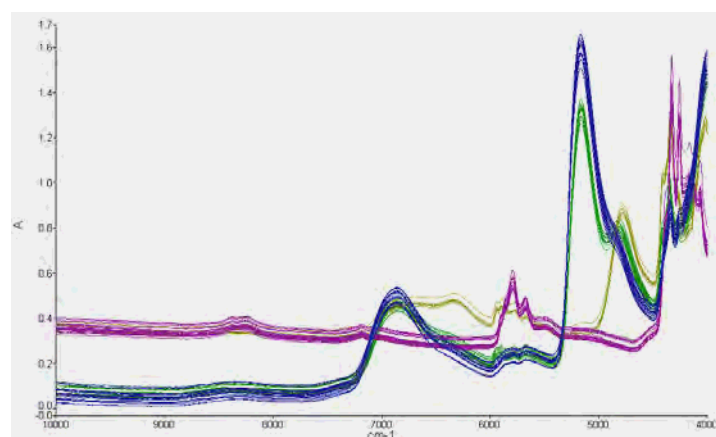
- alkohol lanolinowy
- alkohol cetearylowy
- Cetareth 20



Rys. 76 Widma surowe surowców stałych użytych w modelu 1: zielony - alkohol lanolinowy, granatowy – alkohol cetearylowy, żółty- Cetareth 20

Model 2:

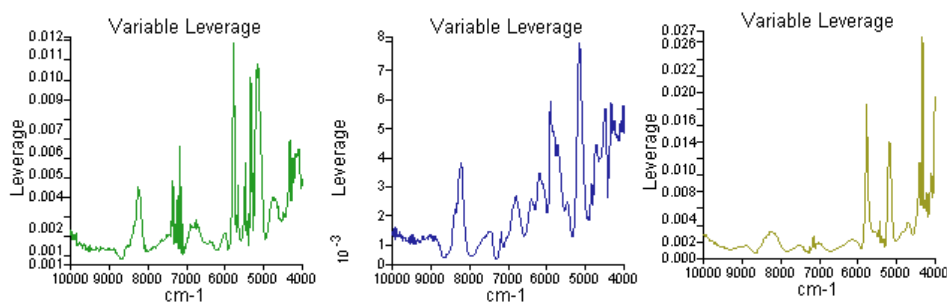
- ekstrakt z ogórka
- parafina
- glikol propylenowy
- glikol butylenowy



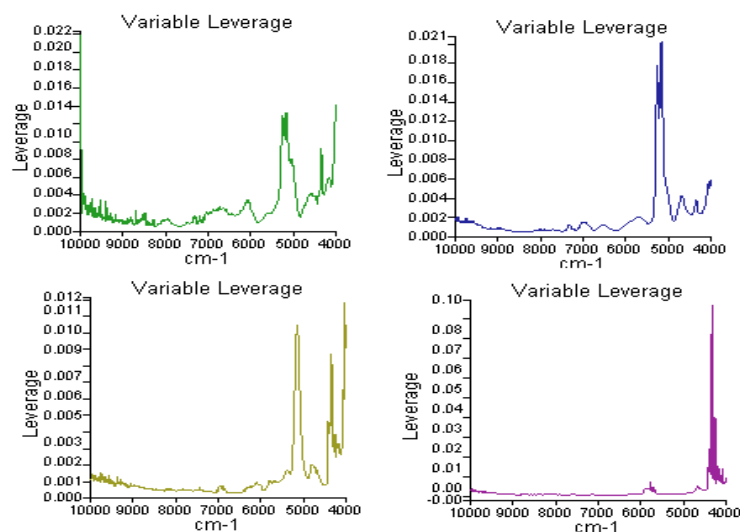
Rys. 77 Widma surowe surowców płynnych użytych w modelu 2: zielony-ekstrakt z ogórka, niebieski-glikol butylenowy, różowy-parafina, żółty-glikol propylenowy

W pierwszym etapie opracowywania powyższych modeli dokonano wyboru zakresu analizy w oparciu o wykresy dźwigni, **Rys.78**. Dla **Modelu 1** wybrano zakres 7 000-4 000 cm^{-1} , a dla **Modelu 2** od 5 500 do 4 000 cm^{-1}

a)



b)

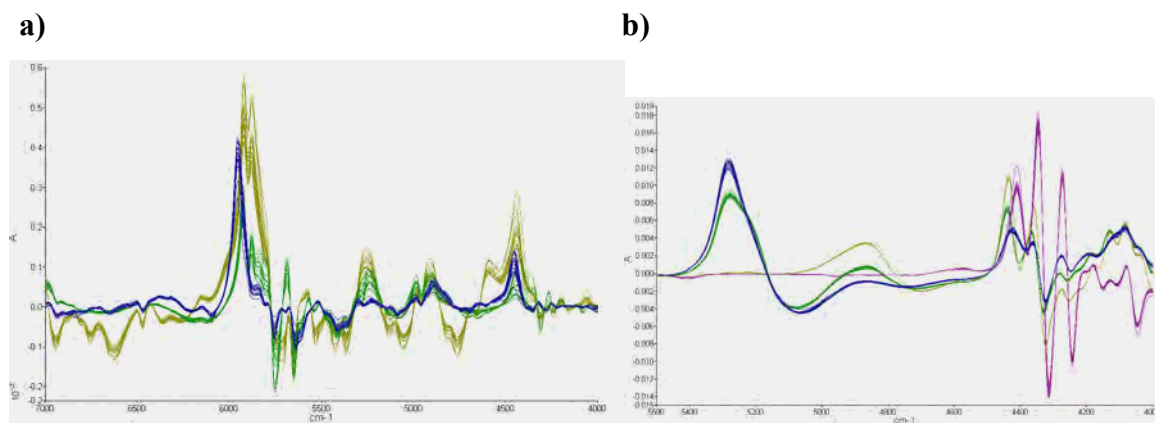


Rys. 78 Wykresy dźwigni otrzymane dla a) surowców Modelu 1, b) surowców Modelu 2. Kolory jak na Rys. 76 i 77

Do opracowania **Modelu 1** łącznie wykorzystano 66 widm. 52 widma umieszczono w zbiorze uczącym (16 widm alkoholu lanolinowego, 17 alkoholu cetearylowego, 19 dla Cetareth20), a pozostałe widma użyto do zbadania zdolności predykcyjnych modeli (zbiór walidacyjny). W **Modelu 2** natomiast użyto zbioru uczącego zawierającego 17 widm ekstraktu z ogórka, 15 widm glikolu butylenowego, 13 widm glikolu propylenowego oraz 17 widm parafiny. W zbiorze walidacyjnym dla tego modelu znalazło się 16 widm, po 4 dla każdego surowca. W kolejnym etapie przeprowadzono transformację widm do ich 1. i 2. pochodnych i zbadano zdolności predykcyjne obu modeli. Następnie do pochodnych widm zastosowano dodatkowo transformację.

I 1 Pochodna

Rys. 79 przedstawia 1. pochodne widm zawartych w zbiorach uczących obu modeli. Na panelu a) dotyczącym próbek surowców stałych wyraźnie widać większe zróżnicowanie intensywności widm.



Rys. 79 Widma uzyskane po zastosowaniu 1. pochodnej a) **Model 1**, b) **Model 2**. Kolory jak na Rys. 76 i 77

Tabela 27. Wyniki klasyfikacji próbek ze zbiorów walidacyjnych przy pomocy modeli o różnym stopniu komplikacji. W polach tabeli podano liczbę próbek poprawnie sklasyfikowanych do ogólnej liczby próbek danej klasy

a) Model 1

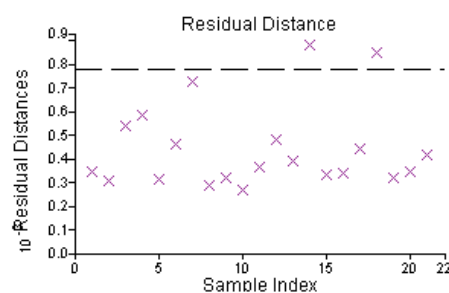
Nazwa surowca	Liczba składowych w modelu				
	5PC	4PC	3PC	2PC	1PC
alkohol cetearylowy	3/4	4/4	4/4	4/4	4/4
alkohol lanolinowy	-	4/4	4/4	4/4	3/4
ceteareth20	6/6	6/6	6/6	6/6	6/6

b) Model 2

Nazwa surowca	Liczba składowych w modelu					
	6PC	5PC	4PC	3PC	2PC	1PC
ekstrakt z ogórka	-	4/4	4/4	3/4	4/4	4/4
glikol butylenowy	-	-	-	4/4	4/4	4/4
glikol propylenowy	-	-	1/4	1/4	4/4	4/4
parafina	3/4	3/4	3/4	4/4	4/4	4/4

Wyniki klasyfikacji próbek ze zbioru uczącego dla **Modelu 1**, wykazały że optymalną jakość klasyfikacji uzyskuje się przy 4 głównych składowych. Przy tym stopniu komplikacji modelu żadna próbka ze zbioru uczącego nie przekracza granicy identyfikacji, odległości między klasami są zadawalające i wszystkie próbki ze zbioru walidacyjnego są prawidłowo sklasyfikowane, **Tab.27a**. W przypadku 3 składowych zaobserwowano pojawienie się próbek ze zbioru uczącego poza granicami identyfikacji mimo że klasyfikator sklasyfikował poprawnie wszystkie próbki ze zbioru walidacyjnego. Taka sytuacja pokazuje, że model nie uwzględnia już wszystkich źródeł zmienności występujących w próbkach zbioru uczącego.

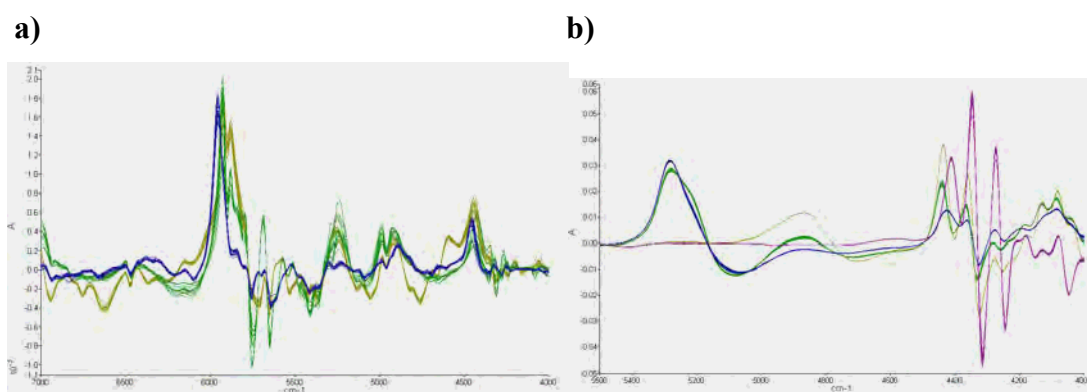
W przypadku **Modelu 2** uzyskano niejednoznaczne wyniki. Przy dwóch czynnikach wszystkie próbki ze zbioru walidacyjnego zostały poprawnie sklasyfikowane, **Tab.27b**. Analizując jednak wykresy RD dla poszczególnych surowców, okazuje się że 2 próbki parafiny znajdują się poza granicami identyfikacji, **Rys.88**. Zwiększenie liczby składowych modelu do 3 zapewnia wprawdzie poprawną klasyfikację wszystkich próbek zbioru uczącego, ale 4 próbki ze zbioru walidacyjnego są klasyfikowane niepoprawnie, **Tab. 27b**.



Rys.80 Wykres RD otrzymany dla parafiny na podstawie modelu zawierającego 2 główne składowe

II 1. pochodna +SNV

W kolejnym kroku 1. pochodne widm poddano dodatkowo normalizacji SNV. Uzyskane widma przedstawia **Rys.81**. Porównując ten rysunek z **Rys.79** widać zdecydowane zmniejszenie zróżnicowania intensywności widm próbek surowców stałych (panele a).



Rys.82 Widma uzyskane po przeprowadzeniu transformacji łączonej (1. pochodna + SNV) a) Model 1, b) Model 2. Kolory widm jak na rys. 76 i 77

Dla **Model 1** zastosowanie transformacji łączonej przyniosło pożądane efekty. Zdecydowanie wzrosły odległości pomiędzy poszczególnymi klasami. Umożliwiło to zmniejszenie stopnia komplikacji modelu. Najlepsze wyniki uzyskano dla modelu zawierającego 4 główne składowe dla cetareth 20 oraz alkoholu cetearylowego oraz przy 3 składowe dla alkoholu lanolinowego, **Tab.28**.

Tabela 28. Wyniki klasyfikacji próbek ze zbioru walidacyjnego dla Modelu 1 po transformacji łączonej. W polach tabeli podano liczbę próbek poprawnie sklasyfikowanych do ogólnej liczby próbek danej klasy

Nazwa surowca	Liczba składowych w modelu				
	5PC	4PC	3PC	2PC	1PC
<i>alkohol cetearylowy</i>	3/4	4/4	4/4	4/4	4/4
<i>alkohol lanolinowy</i>	-	-	4/4	4/4	3/4
<i>cetareth20</i>	6/6	6/6	6/6	6/6	6/6

W przypadku Model 2 zastosowanie transformacji łączonej doprowadziło do pogorszenia zdolności predykcyjnych modelu. Przy żadnej liczbie głównych składowych nie uzyskano poprawnej klasyfikacji wszystkich próbek ze zbioru walidacyjnego, **Tab.29**. Dodatkowo, w porównaniu do modelu bez transformacji SNV wzrosła liczba próbek przekraczających granice identyfikacji.

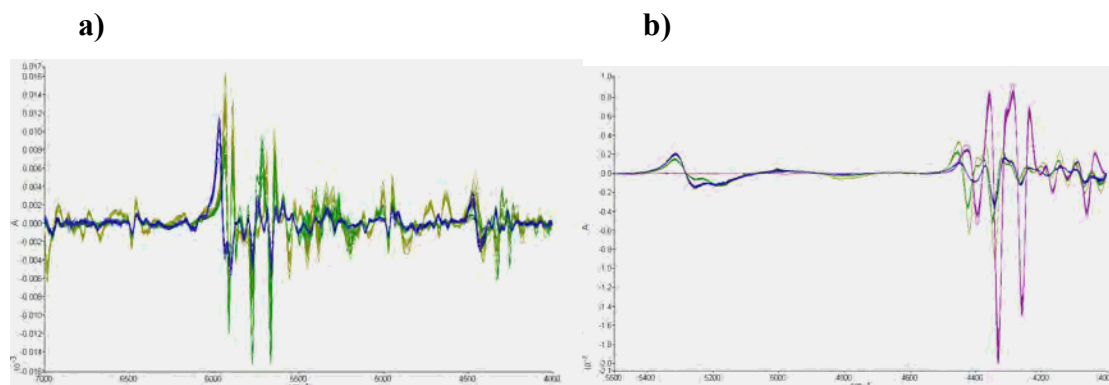
Tabela 29. Wyniki klasyfikacji próbek ze zbioru walidacyjnego dla Modelu 2 po transformacji łączonej. W polach tabeli podano liczbę próbek poprawnie sklasyfikowanych do ogólnej liczby próbek danej klasy

Nazwa surowca	Liczba składowych w modelu					
	6PC	5PC	4PC	3PC	2PC	1PC
ekstrakt z ogórka	-	4/4	4/4	3/4	4/4	4/4
glikol butylenowy	-	-	-	4/4	4/4	4/4
glikol propylenowy	-	-	1/4	1/4	2/4	3/4
parafina	3/4	3/4	3/4	4/4	4/4	4/4

Podsumowując można stwierdzić, że transformacja łączona (1. pochodna + SNV) polepszyła wyniki w przypadku modelu dla surowców stałych. W modelu opracowanym dla surowców płynnych jakość klasyfikacji nie uległa poprawie, a zdolności predykcyjne uległy pogorszeniu.

I 2. pochodna

Na **Rys.81** pokazano 2. pochodne widm zbiorów uczących dla **Modelu 1** (panel a) i **Modelu 2** (panel b).

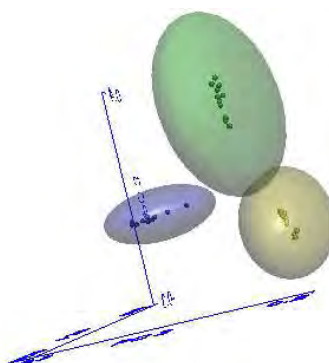


Rys.82 Widma uzyskane po przekształceniu widm do ich 2. pochodnych a) Model 1, b) Model 2. Kolory widm jak na Rys. 76 i 77

Tabela 30. Wyniki klasyfikacji próbek ze zbioru walidacyjnego **Modelu 1** po przekształceniu widm do ich 2. pochodnych przy różnej liczbie składowych

Nazwa surowca	Liczba składowych w modelu			
	4PC	3PC	2PC	1PC
alkohol cetearylowy	-	4/4	4/4	4/4
alkohol lanolinowy	-	-	4/4	3/4
ceteareth 20	6/6	6/6	6/6	6/6

Po przeprowadzeniu transformacji widm zbioru uczącego **Modelu 1** do ich drugich pochodnych uzyskano niepokojące wyniki. Mimo, że nie zaobserwowano próbek ze zbioru uczącego będących poza granicami identyfikacji oraz wszystkie próbki ze zbioru walidacyjnego zostały poprawnie sklasyfikowane nawet przy modelu opartym na jednej głównej składowej, **Tab.30**, to jednak odległości pomiędzy klasami okazały się być niepokojąco małe, **Rys.83**. Model o takiej charakterystyce może działać niepoprawnie w przypadku próbek roboczych nawet nieznacznie odbiegających od zbioru uczącego.



Rys.83 Rzut przedstawiający rozkład klas badanych surowców z **Modelu 1** w przestrzeni 3 pierwszych głównych składowych po przekształceniu widm w ich 2. pochodne

W przypadku zastosowania 2. pochodnych do próbek z **Modelu 2** uzyskano zdecydowanie lepsze wyniki niż w przypadku zastosowania 1. pochodnych. Przede wszystkim możliwe jest uzyskanie poprawnego modelu o niskim stopniu komplikacji. Już dla modelu wykorzystującego tylko 2 pierwsze główne składowe nie zaobserwowano żadnych próbek ze zbioru uczącego poza granicami identyfikacji. Model ten posiada również dobre zdolności predykcyjne: wszystkie próbki ze zbioru walidacyjnego zostały poprawnie sklasyfikowane, **Tab.31**.

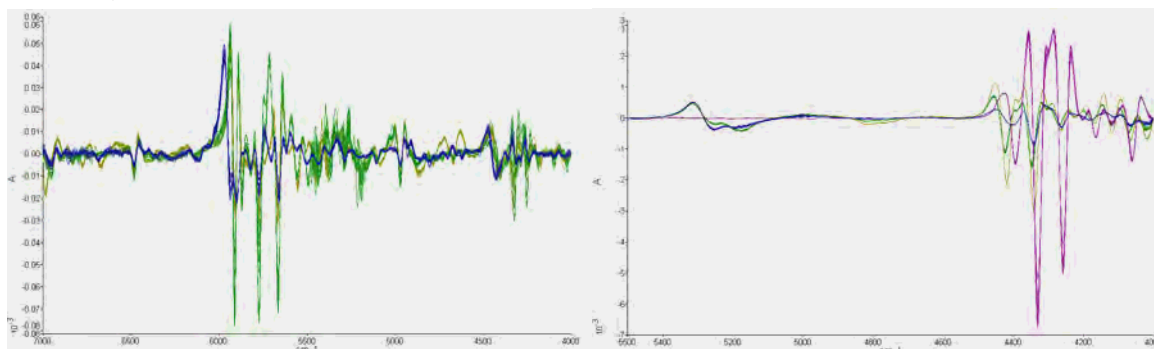
Tabela 31. Wyniki klasyfikacji próbek ze zbioru walidacyjnego **Modelu 2** po przekształceniu widm do ich 2. pochodnych przy różnej liczbie składowych

Nazwa surowca	Liczba składowych w modelu				
	5PC	4PC	3PC	2PC	1PC
<i>ekstrakt z ogórka</i>	-	-	4/4	4/4	4/4
<i>glikol butylenowy</i>	-	-	-	4/4	4/4
<i>glikol propylenowy</i>	-	-	-	4/4	4/4
<i>parafina</i>	3/4	4/4	4/4	4/4	4/4

I 2 pochodna + SNV

W 2. pochodne widm poddano dodatkowo normalizacji SNV. Uzyskane widma przedstawia **Rys.84**. Porównując ten rysunek z **Rys.79** widać zdecydowane zmniejszenie zróżnicowania intensywności widm próbek surowców stałych (panele a).

25 b)



Rys.84 Widma próbek obu modeli uzyskane po zastosowaniu transformacji łącznej (2. Pochodna+SNV) a) Model 1, b) Model 2

Po zastosowaniu łącznej transformacji (2. Pochodna + SNV) zaobserwowano polepszenie jakości **Modelu 1**. Odległości między klasami zwiększyły się, a zdolności predykcyjne nie uległy pogorszeniu, **Tab.32a**. Klasyfikator o takiej charakterystyce powinien już być odporny na wzrost zróżnicowania próbek obserwowany zwykle w zbiorach roboczych.

W przypadku **Modelu 2** przeprowadzenie normalizacji SNV 2. Pochodnych widm doprowadziło do uzyskania gorszego klasyfikatora. Aby uzyskać poprawne przyporządkowanie próbek ze zbioru walidacyjnego należy zastosować tylko 1. Główną składową, **Tab.32b**. Jednakże w przypadku klasyfikatora o tak niskim stopniu komplikacji niektóre próbki ze zbioru uczącego lokują się poza granicami identyfikacji. Potwierdza to jeszcze raz przypuszczenie, że stosowanie normalizacji w przypadku surowców ciekłych nie przynosi pożądanego efektu.

Tabela 30. Wyniki klasyfikacji próbek ze zbioru walidacyjnego a) Model 1, b) Model 2 po zastosowaniu transformacji łączonej (2. Pochodna+SNV) dla różnej liczby użytych składowych

25 Model 1

Nazwa surowca	Liczba składowych w modelu			
	4PC	3PC	2PC	1PC
<i>alkohol cetearylowy</i>	-	4/4	4/4	4/4
<i>alkohol lanolinowy</i>	-	-	4/4	3/4
<i>cetareth 20</i>	6/6	6/6	6/6	6/6

25 Model 2

Nazwa surowca	Liczba składowych w modelu				
	5PC	4PC	3PC	2PC	1PC
<i>ekstrakt z ogórka</i>	-	-	4/4	4/4	4/4
<i>glikol butylenowy</i>	-	-	-	4/4	4/4
<i>glikol propylenowy</i>	-	-	-	3/4	4/4
<i>parafina</i>	3/4	3/4	4/4	4/4	4/4

Zamieszczone w niniejszym rozdziale wyniki pokazują, że normalizacja widm wpływa znacząco na poprawę jakości uzyskiwanych modeli tożsamościowych, zwłaszcza w przypadku surowców stałych. Jeśli chodzi o surowce płynne normalizacja może w pewnych przypadkach pogarszać jakość uzyskiwanych modeli, tak jak to zostało pokazane na przykładzie **Modelu 2**. Nie oznacza to jednak, że we wszystkich przypadkach normalizacja przyczynia się do pogorszenia wyników. Z doświadczenia wiadomo, że często w przypadku substancji półpłynnych, żelowych stosowanie transformacji SNV może polepszyć wyniki.

5.2.4. Wpływ doboru widm do zbioru uczącego na zdolność predykcyjną modelu jakościowego

W rozdziale 5 (p.5.1.1.2-5.1.1.6) za pomocą analizy chemometrycznej został przedstawiony wpływ różnych czynników na efekty końcowe opracowywanych metod jakościowych. W niniejszym podrozdziale na przykładzie modelu tożsamościowego zostanie pokazany wpływ podziału widm na zbiór uczący i walidacyjny na zdolności predykcyjne modeli jakościowych. Do tego eksperymentu wykorzystano surowce płynne użyte wcześniej (p.5.2.3.2) do wykazania wpływu normalizacji na zdolności predykcyjne modeli tożsamościowych. W opracowywanym modelu wykorzystano próbki zmierzone w różnych warunkach pomiarowych (temperatura, wilgotność, czas od wymiana lampy).

W pierwszym etapie eksperymentu w zbiorze uczącym i walidacyjnym łącznie użyto 56 widm, 26 w każdym zbiorze (9 widm zarejestrowanych dla ekstraktu z ogórka, 5 widm dla surowca glikol butylenowy, 4 dla parafiny oraz 7 dla glikolu propylenowego), ale mierzone w zupełnie innych warunkach pomiarowych. Na tak dobranych zbiorach następnie przeprowadzono transformację widm do ich 1. I 2. Pochodnej wykorzystując algorytm Stavickego – Golaya. Do stworzenia modelu identyfikacji zastosowano klasyfikator SIMCA. Uzyskane zdolności predykcyjne modelu dla różnej liczby głównych składowych przedstawiono poniżej.

a) 1 pochodna

Tabela 33. Wyniki klasyfikacji widm ze zbioru walidacyjnego dla różnej liczby użytych składowych

Nazwa surowca	Liczba składowych w modelu				
	5PC	4PC	3PC	2PC	1PC
<i>ekstrakt z ogórka</i>	0/9	0/9	0/9	0/9	3/9
<i>glikol butylenowy</i>	-	0/5	0/5	0/5	0/5
<i>glikol propylenowy</i>	-	0/4	0/4	0/4	0/4
<i>parafina</i>	-	0/7	0/7	0/7	3/7

W oparciu o wykres RD dla zbioru uczącego uznano, że optymalna liczba czynników to 3. Jednak patrząc na wyniki zdolności predykcyjnej zbioru walidacyjnego, **Tab.33**, można stwierdzić, że powyższy model nie działa poprawnie. Nie obserwuje się bowiem poprawnej klasyfikacji widm znajdujących się w zbiorze walidacyjnym.

b) 2 pochodna

Tabela 34. Wyniki klasyfikacji widm ze zbioru walidacyjnego dla różnej liczby użytych składowych

Nazwa surowca	Liczba składowych w modelu				
	5PC	4PC	3PC	2PC	1PC
<i>ekstrakt z ogórka</i>	-	-	0/9	0/9	8/9
<i>glikol butylenowy</i>	-	-	0/5	0/5	4/5
<i>glikol propylenowy</i>	-	-	0/4	0/4	0/4
<i>parafina</i>	0/7	0/7	0/7	0/7	5/7

Najlepsze wyniki po transformacji widm do 2. Pochodnych uzyskano przy zastosowaniu w modelu tylko jednego istotnego czynnika: nie zaobserwowano bowiem w zbiorze uczącym widm poza granicami identyfikacji, jednakże klasyfikacja widm w zbiorze walidacyjnym przy jednej głównej składowej okazała się niepoprawna, **Tab.34**.

Wyniki uzyskane pokazują, że opracowany model tożsamościowy niezależnie od rodzaju zastosowanej transformacji działa niepoprawnie. Wynika to prawdopodobnie z tego, że widma ze zbioru walidacyjnego zawierają dodatkowe źródła zmienności nie występujące w zbiorze uczącym. W celu potwierdzenia tego przypuszczenia opracowano nowy model tożsamościowy dla powyższych surowców. Różnica polegała na tym, że zbiór uczący stanowił mieszaninę prób ze starego zbioru uczącego oraz walidacyjnego.

a) 1 pochodna

Tabela 35 Wyniki walidacji próbek ze zbioru walidacyjnego: liczba próbek poprawnie sklasyfikowana

Nazwa surowca	Liczba składowych w modelu							
	8PC	7PC	6PC	5PC	4PC	3PC	2PC	1PC
<i>ekstrakt z ogórka</i>	-	-	8/9	8/9	6/9	8/9	7/9	8/9
<i>glikol butylenowy</i>	-	-	1/5	2/5	2/5	4/8	4/8	4/8
<i>glikol propylenowy</i>	-	-	-	-	2/4	3/4	3/4	2/4
<i>parafina</i>	5/7	7/7	6/7	6/7	5/7	5/7	6/7	7/7

Analizując powyższą tabelę obserwuje się zwiększenie liczby czynników potrzebnych do wyjaśnienia zmienności w nowym zestawie danych. Jednocześnie zwiększyła się liczba prób poprawnie identyfikowanych ze zbioru walidacyjnego.

b) 2 pochodna

W przypadku 2 pochodne najlepsze wyniki uzyskano przy 3 czynnikach. Wszystkie próbki ze zbioru walidacyjnego zostały poprawnie sklasyfikowane. Do tego nie zaobserwowano żadnych prób będących poza granicami identyfikacji. Zatem można stwierdzić że ten typ transformacji daje zadawalające wyniki.

Tabela 36 Wyniki walidacji próbek ze zbioru walidacyjnego: liczba próbek poprawnie sklasyfikowana

Nazwa surowca	Liczba składowych w modelu					
	6PC	5PC	4PC	3PC	2PC	1PC
<i>ekstrakt z ogórka</i>	-	-	-	8/9	8/9	9/9
<i>glikol Butylenowy</i>	-	-	1/5	5/5	4/5	5/5
<i>glikol propylenowy</i>	-	-	-	4/4	4/4	2/4
<i>parafina</i>	6/7	7/7	6/7	7/7	6/7	7/7

Uzyskane wyniki potwierdzają że dobór prób do zbioru uczącego jest istotny. W zbiorze uczącym powinny znaleźć się próbki, które będą wyjaśniać wariancję całego zestawu danych. W sytuacji gdy tak nie będzie mimo, że próby ze zbioru uczącego nie będą przekraczać granicy identyfikacji to zdolności predykcyjne modeli będą złe.

5.2.5. Porównanie zastosowania spektroskopii NIR oraz IR w analizie tożsamościowej surowców kosmetycznych i farmaceutycznych

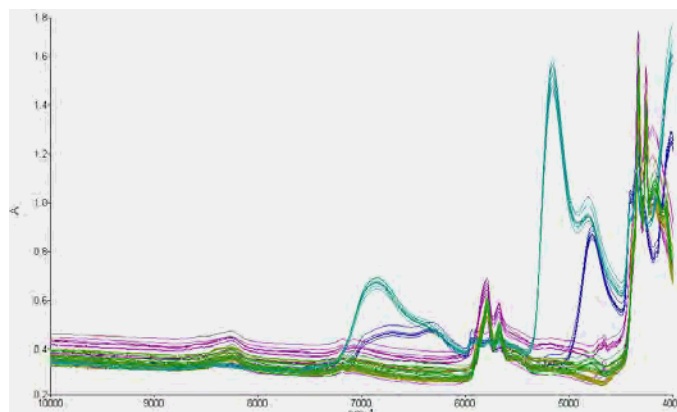
Ze względu na obecność w Zakładzie Ziąja Ltd. zarówno spektrofotometru klasycznej podczerwieni jak i przystawki NIR postanowiono sprawdzić zdolności predykcyjne modeli tożsamościowych opracowanych w dwóch zakresach spektroskopii podczerwieni. W tym celu w oparciu o widma zarejestrowane na bieżąco dla surowców dostarczonych do Zakładu, a następnie archiwizowanych, opracowano dwie procedury analizy tożsamościowej (modele). Pierwszy model opracowano wykorzystując spektroskopię bliskiej podczerwieni drugi natomiast wykorzystując spektroskopię IR. Oba modele opracowano dla tych samych surowców.

Surowce wykorzystane do porównania modeli IR oraz NIR

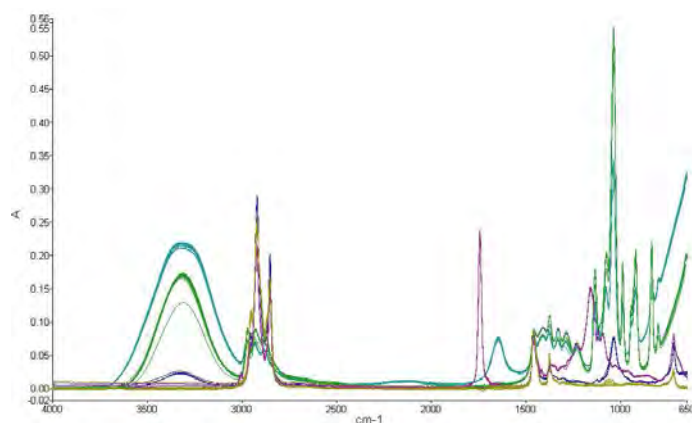
- oktylodekanol
- 1,2-dihydroksypropan
- parafina
- olej z oliwek
- ekstrakt glikolowy z ogórka

Na **Rys.85** przedstawiono widma surowe użyte przy budowie modeli.

a)



b)



Rys.85 Surowe widma **a)** NIR, **b)** IR surowców użytych do budowy obu modeli: zielony-oktylodekanol, niebieski-1,2-dihydroksypropan, żółty –parafina, różowy – olej z oliwki, morski-ekstrakt glikolowy z ogórka

5.2.5.1. Podział widm na trzy zbiory

Widma surowców użytych do budowy powyższych modeli podzielono na trzy zbiory: zbiór uczący, zbiór testowy i zbiór walidacyjny. W zbiorze uczącym użyto widma o największej zmienności. Zbiór walidacyjny posłużył do oceny zdolności prognostycznych modeli (walidacja wewnętrzna), a zbiór testowy do oceny ich jakości (walidacja zewnętrzna). We wszystkich zbiorach zastosowano widma tych samych substancji z różnych partii.

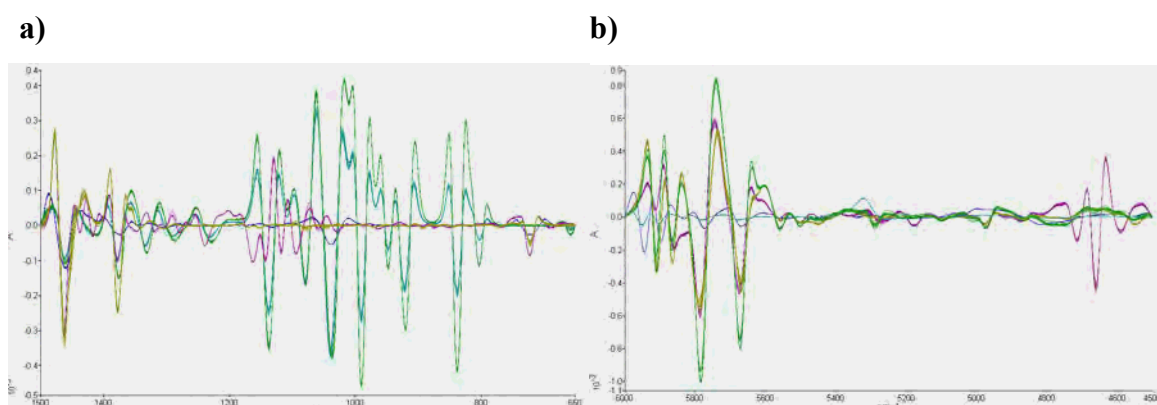
W obu modelach łącznie wykorzystano 124 widm, 78 widm zaliczono do zbioru uczącego, 26 widm do zbioru testowego oraz 20 widm do zbioru walidacyjnego. W zbiorze uczącym w obu metodach użyto: 15 widm oktylodekanolu, 15 widm 1,2 dihydroksy propanu, 20 widm parafiny, i po 14 widm oleju z oliwek oraz ekstraktu glikolowego z ogórka. Modele odpowiednio nazwano – Model NIR oraz Model IR. Tożsamość każdej próbki potwierdzono metodą referencyjną – chromatografia gazowa.

5.2.5.2 Wstępna obróbka danych

Zebrane widma znajdujące się w bibliotece przechowywane są w stanie surowym. Aby wykorzystać je do dalszej analizy wymagana jest obróbka wstępna tych widm. Na podstawie wcześniejszych wyników (p. 5.2.3.2) wybrałam dla obu modeli transformację widm do ich 2. Pochodnych i normalizację. W **Tab.37** podsumowano informacje o zastosowanych transformacjach dla każdej z metod, a na **Rys.86** widma otrzymane po przeprowadzeniu transformacji.

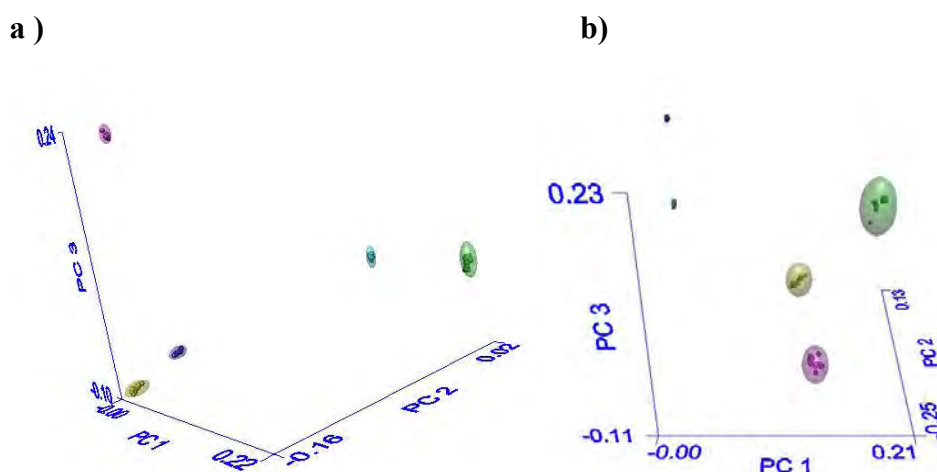
Tabela 37. Rodzaj transformacji użytej podczas opracowywania modeli IR oraz NIR

	<i>Metoda IR</i>	<i>Metoda NIR</i>
<i>Zakres analizy (cm⁻¹)</i>	1500-650	6000-4500
<i>Filtr</i>	Ważenie szumów	Ważenie szumów
<i>Korekcja linii bazowej</i>	2 pochodna 25 punktowe okno wielomianu	2 pochodna 25 punktowe okno wielomianu
<i>Normalizacja</i>	MSC	SNV

Rys.86 Widma 5 surowców użytych do budowy obu modeli po przeprowadzeniu transformacji. **a)** widma IR, **b)** widma NIR

5.2.5.3. Klasyfikacja widm zbioru uczącego

Do oceny przydatności zgromadzonych danych dla analizy tożsamościowej zastosowano analizę głównych składowych (PCA). Widma poszczególnych surowców tworzą na mapie liniowej wyraźnie oddzielone skupienia, **R.87**. Wskazuje to, że widma zbioru uczącego zawierają ilość informacji wystarczającą do zbudowania poprawnych modeli tożsamościowych. Zwraca uwagę większy rozrzut widm NIR w obrębie każdego ze skupień. Do ustalenia reguł identyfikacji zastosowano klasyfikator SIMCA. Klasyfikator SIMCA wyznaczył odpowiednio następującą liczbę głównych składowych dla **Modelu IR**: 4PC oktylodeanol, 2PC 1,2-dihydroxy-propan, 3PC parafina, 5PC oliwa oliwek, 3PC Ekstrakt z ogórka, dla **Modelu NIR**: 3PC oktylodekanol, 2PC 1,2-dihydroxy-propan, 3PC parafina, 3PC oliwa z oliwek, 3PC ekstrakt z ogórka.



Rys.87 Rzut przedstawiający widma surowców obu modeli w przestrzeni głównych składowych
a) Model NIR, b) Model IR

Tabela 38 .Ocena poprawności klasyfikacji próbek ze zbioru uczącego. Wyniki klasyfikacji dla Metody IR i Metody NIR były takie same

Rzeczywista przynależność do klasy	Wynik klasyfikacji				
	oktylodekanol 1	1,2- dihydrokso- propan	parafina	olej z oliwek	Ekstrakt z ogórka
oktylodekanol	15	0	0	0	0
1,2 dihydrokso - propan	0	15	0	0	0
parafina	0	0	20	0	0
olej z oliwek	0	0	0	14	0
Ekstrakt z ogórka	0	0	0	0	14

Z danych przedstawionych w **Tab. 38** wynika, że opracowane modele działają poprawnie. Wszystkie próbki ze zbioru uczącego zostały poprawnie sklasyfikowane. Odległości między klasami są zadawalające, **Rys.87**, dlatego też istnieje szansa, że powyższe modele działają poprawnie również dla próbek spoza zbioru uczącego. Jednak aby potwierdzić to przypuszczenie przeprowadzono walidację metod przy pomocy zbioru testowego (walidacja zewnętrzna) oraz walidacyjnego (walidacja wewnętrzna).

5.2.5.4. Walidacja zewnętrzna oraz wewnętrzna

Do oceny prognostycznej opracowanych modeli wykorzystano zbiór testowy (przeprowadzając walidację wewnętrzną natomiast do walidacji metod użyto zbiór walidacyjny, uzyskując 100% poprawną klasyfikację próbek należących do zbioru testowego oraz walidacyjnego w modelu IR oraz NIR, **Tab.39**.

Tabela 39. Ocena poprawności klasyfikacji próbek ze zbioru a) walidacyjnego b) testowego

a)

<i>Surowiec</i>	<i>Liczba próbek</i>	<i>Liczba próbek poprawnie zaklasyfikowanych</i>	
		<i>Metoda IR</i>	<i>Metoda NIR</i>
oktylodekanol	5	5	5
1,2 dihydroksy propan	6	6	6
parafina	5	5	5
olej z oliwki	4	4	4
ekstrakt ogórkowy	4	4	4

b)

<i>Surowiec</i>	<i>Liczba próbek</i>	<i>Próbki poprawnie zaklasyfikowane</i>	
		<i>Metoda IR</i>	<i>Metoda NIR</i>
oktylodekanol	5	5	5
1,2 dihydroksy propan	4	4	4
parafina	5	5	5
olej z oliwki	3	3	3
ekstrakt ogórkowy	3	3	3

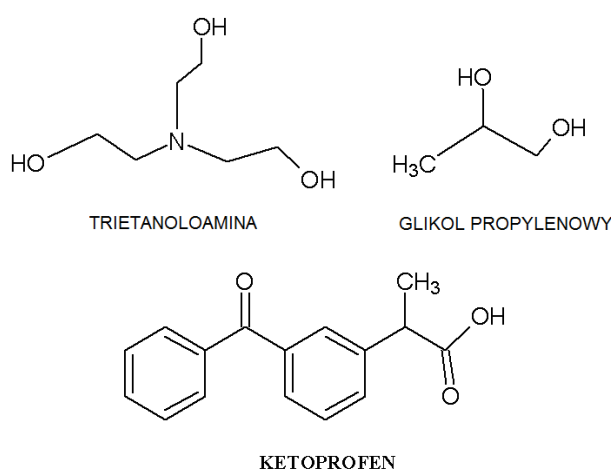
Przedstawione w **Tab.39** wyniki potwierdzają, że zarówno widma NIR jak i IR mogą, po zastosowaniu odpowiednich technik chemometrycznych, być wykorzystywane w analizie tożsamościowej analizowanych surowców. Zastosowanie widm IR wraz z odpowiednimi technikami chemometrycznymi umożliwia skrócenie czasu analiz oraz daleko posuniętą ich automatyzację. Opracowując modele tożsamościowe na podstawie widm NIR i IR należy zadbać o:

- a) powtarzalność warunków przechowywania próbek i rejestracji widm,
- b) właściwy podział widm na zbiór uczący, walidacyjny oraz testowy,
- c) właściwy dobór procedur wstępnej transformacji widm
- d) właściwy dobór liczby istotnych czynników

5.3. Analiza ilościowa

5.3.1. Oznaczanie w mg/g zawartości substancji aktywnej – ketoprofen w żelu Ketoprofen

Ketoprofen jest stosowany w leczeniu objawowym zmian zapalnych i zwyrodnieniowych w schorzeniach reumatycznych, czasem stosowany jest również w łagodzeniu niektórych zespołów bólowych. Podstawowymi składnikami żelu są: trietanolamina, glikol propylenowy, ketoprofen (substancja aktywna), etanol, karbopol (kwas poliakrylowy, karbomer substancja pomocnicza stosowana w nowoczesnej preparatyce farmaceutycznej, stosowana do zagęszczenia analizowanego żelu) oraz woda.

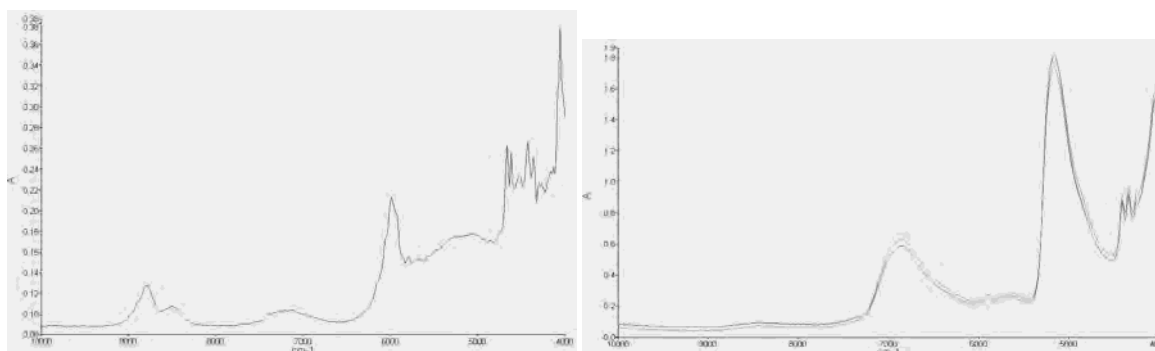


Rys.88 Wzory strukturalne surowców używanych przy produkcji żelu Ketoprofen

Widma żelu zbierane były w zakresie bliskiej podczerwieni w trybie transflektancji. Do opracowania modelu wykorzystano łącznie widma 45 próbek. 30 z nich umieszczono w zbiorze kalibracyjnym (12 próbek laboratoryjnych i 18 próbek produkcyjnych), a w zbiorze walidacyjnym (zbiór niezależny) 11 próbek. Opracowany model testowano dodatkowo za pomocą zbioru testowego – 4 próbki produkcyjne.

5.3.1.1. Analiza widm, wybór zakresu analizy

Pierwszym krokiem w opracowywaniu modelu ilościowego dla oznaczania zawartości ketopofenu w żelu Ketoprofen była analiza widmowa czystego surowca Ketoprofen oraz produktu końcowego – żelu. Na podstawie analizy widmowej wybrano zakres umożliwiający opracowanie modelu ilościowego.



Rys.89 Widma otrzymane w zakresie od 10000-4000 cm^{-1} dla **a)** czystego surowca ketoprofen, **b)** dla żelu Ketoprofen

Widmo żelu, **Rys.89b**, posiada dwa wyraźne pasma przy długości fali 6 800 oraz 5 200 cm^{-1} . Pasma te pochodzą od pierwszego nadtonu wiązania O-H oraz drgań kombinacyjnych cząsteczki wody. Na widmie widoczne są również pasma absorpcji przy długości fali 4348 cm^{-1} . Obszar ten charakterystyczny jest dla etanolu. Dla nas interesującym obszarem jest zakres 6 800-5 856 cm^{-1} . W tym zakresie obserwuje się pasma należące do substancji aktywnej dla której opracowany jest model ilościowy (**Rys. 89a**) [112-114].

5.3.1.2. Zastosowanie algorytmu PCR do oznaczania zawartości ketoprofenu

Korzystając z algorytmu PCR opracowano kilka modeli kalibracyjnych, których celem ma być przewidywanie zawartości substancji aktywnej – ketoprofen w żelu Ketoprofen. Modele różniły się względem siebie zakresem analizy, liczbą użytych głównych składowych oraz rodzajem zastosowanej transformacji. Liczba głównych składowych była dobierana automatycznie przez program oraz ręcznie w oparciu o krzywe zależności własnych próbek od ilości faktorów dla zbioru kalibracyjnego oraz walidacyjnego. W **Tab.40** przedstawiono szczegóły opracowanych modeli.

Tabela 40. Parametry 8 przykładowych modeli kalibracyjnych opracowanych dla oznaczania zawartości ketoprofenu w żelu Ketoprofen - algorytm PCR

Nazwa modelu	Normalizacja	Korekta linii bazowej	skalowanie	Zakres (cm^{-1})	L. czynników
KETPCR1	-	1 pochodna (13)	A	6100-5856	3
KETPCR2	SNV+detrend	1 pochodna (13)	A	6100-5856	7
KETPCR3	-	1 pochodna (13)	A	6800-6100	1
KETPCR4	-	1 pochodna (13)	C	6800-6100	1
KETPCR5	SNV+detrend	1 pochodna (13)	C	6800-6100	1
KETPCR6	-	1 pochodna (13)	A	6200-5956	9
KETPCR7	SNV+detrend	1 pochodna (13)	A	6200-5956	1
KETPCR8	-	2 pochodna (13)	A	6200-5956	8

A – autoskalowanie, C- centrowanie

PC – liczba głównych składowych użyta w modelu

W **Tab.41** przedstawiono wyniki walidacji otrzymane dla 8 modeli kalibracyjnych na zbiorach kalibracyjnym i walidacyjnym.

Tabela 41. Wyniki identyfikacji i walidacji modeli z tab.38 uzyskanych za pomocą algorytm PCR

Nazwa modelu	Zbiór kalibracyjny			Zbiór walidacyjny			SEC/SEP
	R	SEC	Bias	R	SEP	Bias	
KETPCR1	0,997	0,70	0,0096	0,992	0,92	0,22	0,76
KETPCR2	0,998	0,57	0,0089	0,996	0,51	0,067	1,38
KETPCR3	0,25	8,04	0,0093	0,35	5,79	1,63	1,38
KETPCR4	0,26	7,74	0,0093	0,062	6,01	2,43	1,3
KETPCR5	0,29	7,9	0,0087	0,01	6,9	2,82	1,14
KETPCR6	0,998	0,5	0,0096	0,996	0,69	0,0093	0,72
KETPCR7	0,97	1,84	0,0098	0,992	0,873	0,23	2,1
KETPCR8	0,997	0,74	0,009	0,999	0,52	0,35	1,42

Wyniki uzyskane pokazują, że najlepsze rezultaty otrzymano dla modeli opracowanych w zakresie od 6 100-5 856 cm^{-1} (modele KETPCR1 i KETPCR2). Dla modeli opracowanych w tym zakresie uzyskano zbliżone wartości SEC oraz SEP, a wartości parametru Bias są stosunkowo niskie. W przypadku modeli opracowanych w zakresie 6 100-6 800 cm^{-1} (KETPCR3 – KETPCR5) obserwuje się duże wartości SEC, SEP oraz małe wartości współczynników korelacji. Ponadto, wartości pozostałych parametrów statystycznych również nie są zadowalające. Natomiast w przypadku 3 ostatnich modeli

(KETPCR6 – KETPCR8), wykorzystujących zakres 6200-5956 cm^{-1} uzyskano ponownie zadawalające wyniki.

Dla wszystkich testowanych modeli zbadano liniowość zależności pomiędzy zawartością ketoprofenu wyznaczoną doświadczalnie i oszacowaną na podstawie widma NIR. W **Tab.42** przedstawiono wartości współczynników regresyjnych oraz wartości statystyki t Student oceniającą istotność otrzymane dla wszystkich modeli, a na **Rys.91** przedstawiono graficznie wyniki uzyskane dla 3 wybranych modeli.

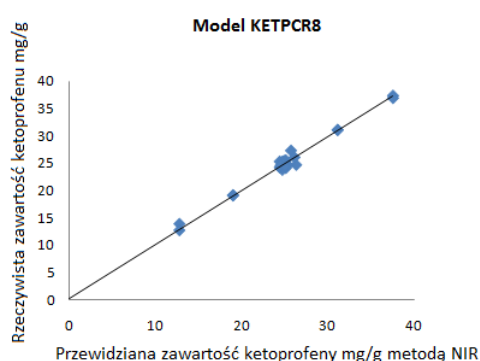
Tabela 42. Liniowość zależności pomiędzy rzeczywistą (y) i oszacowaną na podstawie widm NIR (x) zawartością ketoprofenu zgodnie z zależnością $y = ax + b$

Nazwa modelu	Zbiór kalibracyjny			Zbiór walidacyjny		
	Wsp. regresji		Statystyka t^*	Wsp. regresji		Statystyka t^{**}
	a	b		a	b	
KETPCR1	0,994±0,031	0,14±0,77	0,08	1,069±0,1	1,94±2,87	0,78
KETPCR2	0,997±0,023	0,07±0,58	0,099	0,996±0,067	0,05±1,7	0,42
KETPCR3	0,065±1,46	21,9±34,2	0,006	0,075±3,54	21,5±82,9	0,92
KETPCR4	0,07±1,4	21,7±32,8	0,007	-0,012±4,5	22,9±103	1,26
KETPCR5	0,086±1,25	21,4±29	0,006	-0,0025±3,6	22,25±79,4	1,47
KETPCR6	0,998±0,02	0,042±0,49	0,13	1,07±0,085	1,67±2,19	0,04
KETPCR7	0,95±0,09	1,13±2,17	0,03	1,05±0,1	-1,01±2,8	0,85
KETPCR8	0,995±0,03	0,12±0,74	0,078	1,05±0,065	-1,58±1,66	2,89

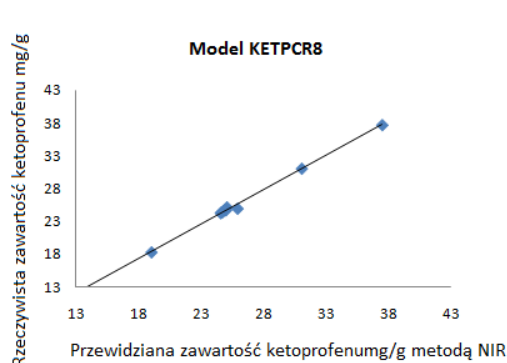
* obliczona wartość statystyki t w teście istotności różnic dla par związanych; wartość krytyczna 2,042

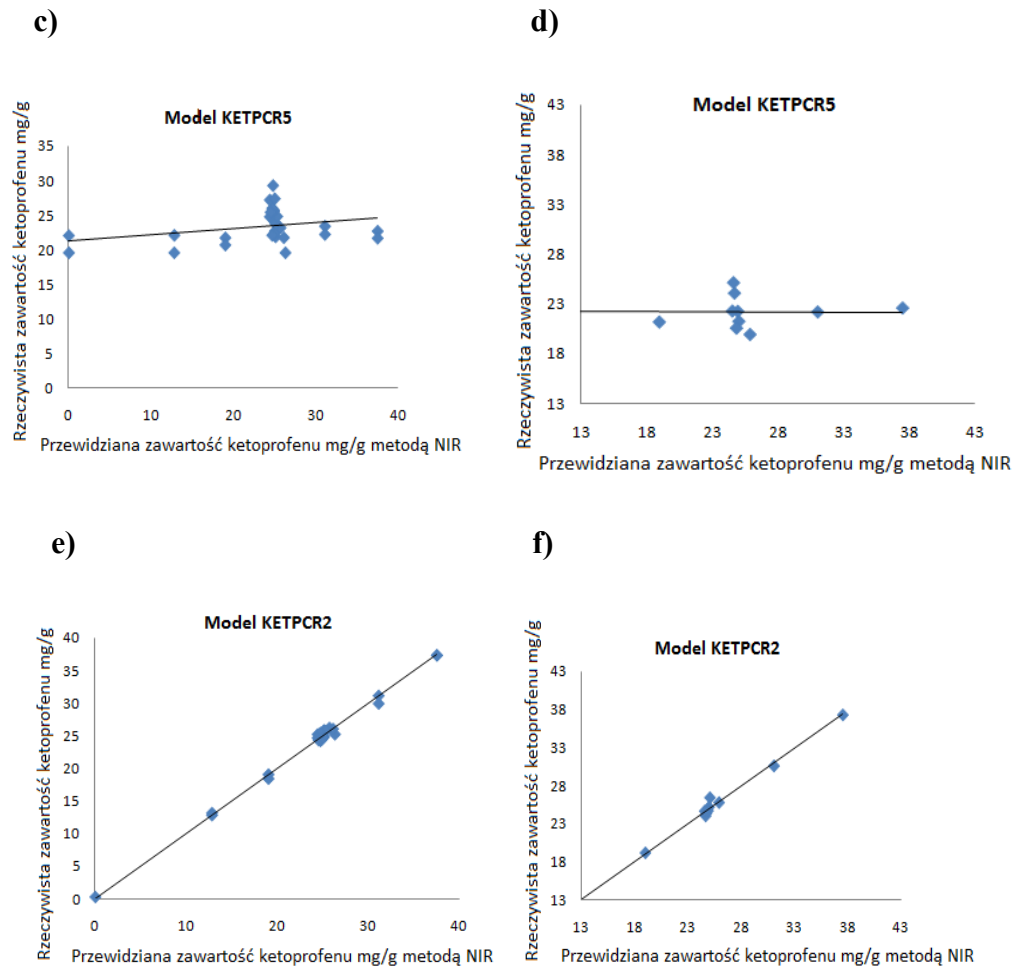
** obliczona wartość statystyki t w teście istotności różnic dla par związanych; wartość krytyczna 2,2

a)



b)





Rys. 90 Wykresy liniowości odpowiedzi modelu uzyskane dla modeli KETPCR2, KETPCR5, KETPCR8; panele a, c, e) - zbiór kalibracyjny, panele b, d, f - zbiór walidacyjny

Z przedstawionych powyżej ocen modeli kalibracyjnych uzyskanych metodą PCR wynika jednoznacznie, że modele KETPCR1, KETPCR2, KETPCR6, KETPCR7 i KETPCR8 spełniają wstępne warunki wymagane w analityce przemysłowej.

Dla 3 najlepszych modeli wykonano dodatkowo walidację zewnętrzną na zbiorze testowym. Wyniki przedstawiono w **Tab.43**.

Tabela 43. Wyniki walidacji zewnętrznej najlepszych modeli kalibracyjnych dla oznaczania zawartości ketoprofenu w żelu Ketoprofen – algorytm PCR

Model	Nazwa próbki	Odzysk%	Precyzja RSD%	Precyzja pośrednia	MD*	RR**
KETPCR2	P29	101,2	0,52	1,2	0,4	2,3
	P58	100,4	0,68	1,7	0,45	10,9
	P55	102	0,78	1,55	0,41	3,1
	P49	100,2	0,80	1,73	0,68	8,4
KETPCR6	P29	102,2	0,75	1,42	0,25	0,86
	P58	100	0,63	1,7	1,33	4,9
	P55	101,8	0,47	1,5	0,68	1,86
	P49	99,6	0,55	1,45	0,62	3,1
KETPCR8	P29	98,9	1,4	2,1	0,27	1,07
	P58	102	0,90	1,6	2,03	4,5
	P55	106,5	0,87	1,95	0,49	1,22
	P49	102,6	0,75	1,8	0,56	4,8

***MD** (M-distance)- opisuje stosunek między odległością Mahalanobiasa, dla próbki i wszystkich współczynników w modelu, oraz wartością graniczną dla tych współczynników, **MD<1**

****RR** (Residual Ratio) opisuje stosunek resztkowej wariancji próbki do średniej resztkowej wariancji wzorców **RR<3**

Analizując uzyskane wartości stopnia odzysku dla próbek ze zbioru testowego, **Tab.43**, można stwierdzić, że powszechnie stosowane kryterium $100 \pm 2\%$ jest spełnione tylko w przypadku modeli KETPCR2 i KETPCR6. Obydwa te modele mają również dla widm ze zbioru testowego poprawne wartości parametru MD (*M-Distance*). Jednak wartości parametru statystycznego RR (*Residual Ratio*) okazały się przekraczać wartość graniczną. Zatem na tej podstawie można stwierdzić, że wszystkie opracowane przeze mnie modele dla oznaczanie zawartości ketoprofenu w żelu Ketoprofen przy użyciu algorytmu PCR są niepoprawne z punktu widzenia analityki przemysłowej.

5.3.1.3. Zastosowanie algorytmu PLS1 do oznaczania zawartości ketoprofenu

Podobnie jak w przypadku algorytmu PCR opracowano kilka wariantów modelu kalibracyjnego na oznaczanie zawartości ketoprofenu w żelu Ketoprofen. W **Tab.44** przedstawiono szczegóły 8 testowanych modeli.

Tabela 44. Parametry 8 modeli kalibracyjnych opracowanych dla oznaczania zawartości ketoprofenu w żelu Ketoprofen - algorytm PLSI

Nazwa modelu	Normalizacja	Korekta linii bazowej	skalowanie	Zakres (cm ⁻¹)	L. czynników
KETPLS1	-	1 pochodna (13)	A *	6100-5856	4
KETPLS2	SNV+ detrend	1 pochodna (13)	A *	6100-5856	5
KETPLS3	-	1 pochodna (13)	A *	6100-6800	5
KETPLS4	SNV+ detrend	1 pochodna (13)	A *	6100-6800	4
KETPLS5	-	2 pochodna (13)	A *	6100-6800	4
KETPLS6	-	1 pochodna (13)	A *	6200-5956	5
KETPLS7	SNV+ detrend	1 pochodna (13)	A *	6200-5956	3
KETPLS8	-	2 pochodna (13)	A *	6200-5956	2

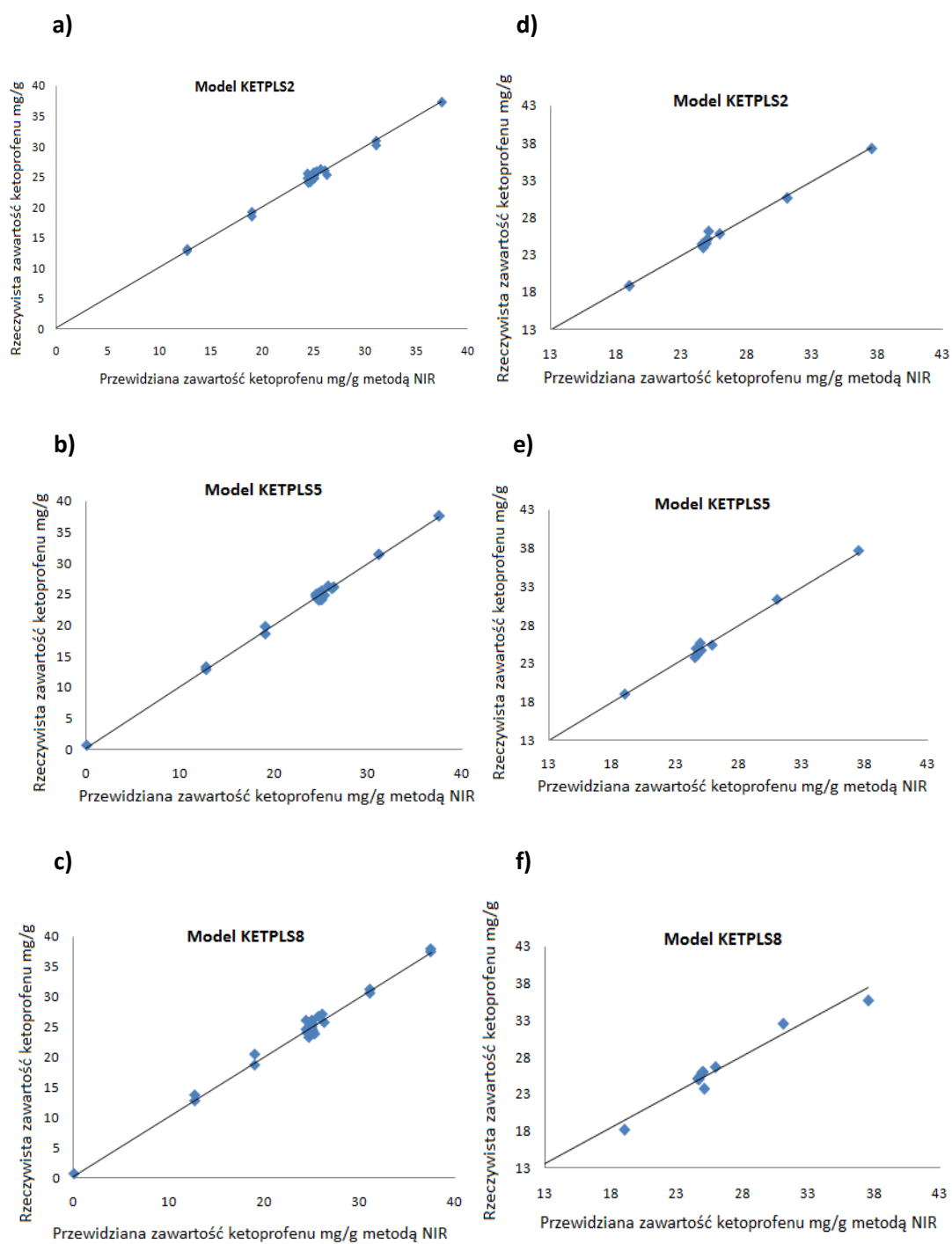
*A- autoskalowanie,

Wstępna ocena jakości analizowanych modeli dla oznaczania zawartości ketoprofenu w żelu Ketoprofen wykazała, że uzyskane wyniki są zadawalające, **Tab.45**. Dla wszystkich modeli uzyskano zadawalające wartości współczynników korelacji, wartości SEP oraz SEC, a także wartości średniego błędu pomiędzy wartościami oznaczonymi metodą NIR a wartościami oznaczonymi metodą referencyjną

Tabela 45. Liniowość zależności pomiędzy rzeczywistą (y) i oszacowaną na podstawie widm NIR(x) zawartością ketoprofenu zgodnie z zależnością $y = ax + b$

Nazwa modelu	Zbiór kalibracyjny			Zbiór walidacyjny			
	R	SEC	Bias	R	SEP	Bias	SEC/SEP
KETPLS1	0,997	0,67	0,0022	0,992	0,91	0,0027	0,74
KETPLS2	0,998	0,51	0,0037	0,997	0,46	0,126	1,11
KETPLS3	0,997	0,64	0,0027	0,993	0,9	0,25	0,71
KETPLS4	0,998	0,59	0,0035	0,988	0,56	0,21	1,04
KETPLS5	0,998	0,54	0,0024	0,997	0,46	0,082	1,18
KETPLS6	0,997	0,61	0,0032	0,997	0,59	0,056	1,03
KETPLS7	0,992	1,053	0,0033	0,993	0,84	0,09	1,25
KETPLS8	0,995	0,85	0,00019	0,985	0,95	0,198	0,89

Dla wszystkich modeli zbadano ich liniowość, oraz sprawdzono dokładność metod za pomocą testu t Studenta. Uzyskane wyniki umieszczono w **Tab.46**. Dodatkowo, na **Rys.92** zamieszczono kilka wykresów dla wybranych metod przedstawiających zależność pomiędzy zawartością ketoprofenu w mg/g wyznaczoną doświadczalnie i oszacowaną na podstawie modelu.



Rys. 91 Wykresy liniowości odpowiedzi modelu uzyskane dla modeli KETPCR2, KETPCR5, KETPCR8; panele a, c, e) - zbiór kalibracyjny, panele b, d, f) - zbiór walidacyjny

Tabela 46. Liniowość zależności pomiędzy rzeczywistą (y) i oszacowaną na podstawie widm NIR(x) zawartością ketoprofenu zgodnie z zależnością $y = ax + b$

Nazwa metody	Zbiór kalibracyjny			Zbiór walidacyjny		
	Liniowość		Statystyka t^*	Liniowość		Statystyka t^{**}
	a	b		a	b	
KETPLS1	0,99±0,03	0,15±0,72	0,02	1,07±0,1	1,96±0,97	0,88
KETPLS2	0,996±0,02	0,089±0,55	0,04	0,995±0,06	0,0007±1,5	0,9
KETPLS3	0,994±0,03	0,135±0,7	0,025	1,075±0,11	2,13±28	0,91
KETPLS4	0,995±0,02	0,113±0,63	0,035	1,026±0,12	-0,86±31	0,7
KETPLS5	0,996±0,02	0,103±0,59	0,026	0,994±0,06	0,068±1,53	0,57
KETPLS6	0,994±0,03	0,135±0,65	0,031	1,065±0,073	1,57±1,89	0,29
KETPLS7	0,984±0,05	0,36±1,17	0,018	0,99±0,18	0,64±1,25	0,31
KETPLS8	0,989±0,04	0,251±0,97	0,013	0,974±0,14	0,848±3,53	0,62

* obliczona wartość statystyki t w teście istotności różnic dla par związanych; wartość krytyczna 2,042

** obliczona wartość statystyki t w teście istotności różnic dla par związanych; wartość krytyczna 2,2

Na tym etapie analizy stwierdzono poprawność wszystkich modeli, poddano je więc walidacji zewnętrznej wykorzystując niezależny zestaw danych. Wyniki uzyskane dla trzech najlepszych modeli zamieszczono w **tabTab.47**.

Tabela 47 Wyniki walidacji zewnętrznej najlepszych modeli kalibracyjnych dla oznaczania zawartości ketoprofenu w żelu Ketoprofen – algorytm PLS

Metoda	Nazwa próbki	Odzysk	Precyzja RSD%	Precyzja pośrednia%	MD*	RR**
KETPLS2	P29	99,45	0,18	0,90	0,28	0,46
	P58	99,43	0,21	0,79	0,21	0,20
	P55	97,99	0,37	1,02	0,22	0,48
	P49	100,04	0,38	1,3	0,20	0,49
KETPLS5	P29	101,6	0,85	1,78	0,45	0,67
	P58	98,3	0,74	1,64	0,62	0,76
	P55	98,9	0,68	1,22	0,34	0,34
	P49	97	0,77	0,98	0,24	0,99
KETPLS8	P29	103,1	0,64	1,3	0,86	0,83
	P58	102,4	0,79	2,1	0,34	1,87
	P55	99,6	0,72	1,85	0,79	2,9
	P49	97,4	0,46	1,55	0,88	1,77

*MD (M-distance)- opisuje stosunek między odległością Mahalanobiasa, dla próbki i wszystkich współczynników w modelu, oraz wartością graniczną dla tych współczynników, $MD < 1$

**RR (Residual Ratio) opisuje stosunek resztkowej wariancji próbki do średniej resztkowej wariancji wzorców $RR < 3$

Uzyskane wyniki pokazały, że najlepsze rezultaty otrzymano dla modelu KETPLS2. Wartości odzysku nie przekroczyły progu granicznego $100\pm 2\%$, jednocześnie w granicach błędu uzyskano wartości MD oraz RR.

Porównując opracowane modele przy użyciu algorytmu PCR oraz PLS można stwierdzić, że uzyskano porównywalne wyniki. Dla obu algorytmów otrzymano przynajmniej jeden model spełniający kryteria akceptacji. Jednakże nieco lepsze wartości parametrów statystycznych (MD, RR) uzyskano dla modeli po zastosowaniu algorytmu PLS. Dlatego też można stwierdzić, że nieco lepsze wyniki uzyskano dla modeli po zastosowaniu cząstkowej metody najmniejszych kwadratów.

5.3.1.4. Zastosowanie algorytmu PLS2 do oznaczania zawartości w mg/g ketoprofenu

W przypadku modeli z wykorzystaniem algorytmu PLS2 przeanalizowano również szereg modeli. Wyniki uzyskane dla 4 modeli przedstawiono w **Tab.48**.

Tabela 48. Parametry 4 przykładowych modeli kalibracyjnych opracowanych dla oznaczania zawartości ketoprofenu w żelu Ketoprofen - algorytm PLS2

Nazwa metody	Normalizacja	Korekta linii bazowej	skalowanie	Zakres (cm^{-1})	L. czynników
KETPLS21	-	1 pochodna (13)	A *	6100-5856	4
KETPLS22	SNV+detrend	1 pochodna (13)	A *	6100-5856	5
KETPLS23	-	1 pochodna (13)	A *	6100-6800	12
KETPLS24	-	1 pochodna (13)	A *	6100-6800	5

*A- autoskalowanie,

Po przeprowadzeniu wstępnej oceny jakości modeli na zbiorze kalibracyjnym i walidacyjnym zaobserwowano, **Tab.49**, że w przypadku modelu KETPLS23 uzyskano najniższą wartość współczynnika korelacji. Wartość ta dla zbioru walidacyjnego jest mniejsza niż założona wartość 0,95. Na tej podstawie już na tym etapie oceny można model KETPLS23 odrzucić.

Tabela 49 Wyniki identyfikacji i walidacji modeli z tab.38 uzyskanych za pomocą algorytm PLS2

Nazwa metody	Zbiór kalibracyjny			Zbiór walidacyjny			
	R	SEC	Bias	R	SEP	Bias	SEC/SEP
KETPLS21	0,997	0,71	0,0023	0,992	0,92	0,26	0,77
KETPLS22	0,998	0,52	0,00366	0,997	0,46	0,13	1,54
KETPLS23	0,996	2,95	0,0035	0,82	3,38	0,46	0,87
KETPLS24	0,997	0,61	0,0032	0,997	0,59	0,056	0,88

Dla pozostałych trzech modeli następnie zbadano liniowość i dokładność t-testem. Uzyskane wyniki pokazały, że liniowość dla wszystkich 3 modeli jest poprawna. Przedział ufności dla współczynnika kierunkowego a jest bliska zeru, natomiast wartość współczynnika przesunięcia b 1. Zatem opracowane modele nie zawierają błędu systematycznego.

Tabela 50. Liniowość zależności pomiędzy rzeczywistą (y) i oszacowaną na podstawie widm NIR(x) zawartością ketoprofenu zgodnie z zależnością $y = ax + b$

Nazwa metody	Zbiór kalibracyjny			Zbiór walidacyjny		
	Liniowość		t-test	Liniowość		t-test
	a	b		a	b	
KETPLS21	0,993±0,03	0,165±0,77	0,021	1,07±0,11	-1,94±2,87	0,92
KETPLS22	0,996±0,02	0,089±0,55	0,042	0,995±0,06	0,0007±1,55	0,9
KETPLS24	0,995±0,03	0,123±0,65	0,031	1,06±0,07	-1,57±1,88	0,29

* obliczona wartość statystyki t w teście istotności różnic dla par związanych; wartość krytyczna 2,042

** obliczona wartość statystyki t w teście istotności różnic dla par związanych; wartość krytyczna 2,2

W kolejnym etapie oceny jakości modeli wykorzystujących algorytm PLS2 przeprowadzono ich walidację zewnętrzną w oparciu o zestaw 4 prób niezależnych. Uzyskane wyniki wykazały, że najlepszym modelem dla oznaczanie procentowej zawartości ketoprofenu w żelu Ketoprofen jest model KETPLS22. Wartości odzysku w tym modelu mieszczą się w zakresie $100 \pm 2\%$ oraz wartości parametrów statystycznych nie przekraczają powszechnie przyjętych w analityce przemysłowej wartości granicznych.

Tabela 51 Wyniki walidacji zewnętrznej najlepszych modeli kalibracyjnych dla oznaczania zawartości ketoprofenu w żelu Ketoprofen – algorytm PLS2

Metoda	Nazwa próbki	Odzysk	Precyzja RSD%	Precyzja pośrednia	M-D*	R- R**
KETPLS2 2	P29	99,45	0,18%	0,95%	0,28	0,46
	P58	98,43	0,21%	0,80%	0,21	0,20
	P55	98	0,37%	1,2%	0,22	0,48
	P49	100,04	0,38%	1,5%	0,20	0,49

***M-D** (M-distance)- opisuje stosunek między odległością Mahalanobiasa, dla próbki i wszystkich współczynników w modelu, oraz wartością graniczną dla tych współczynników, **M-D**<1

****R-R** (Residual Ratio) opisuje stosunek resztkowej wariancji próbki do średniej resztkowej wariancji wzorców **R-R**<3

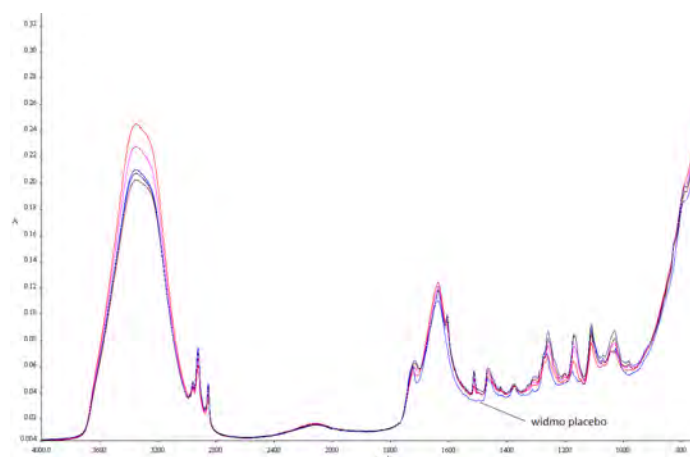
Porównując ze sobą najlepsze modele kalibracyjne uzyskane przy użyciu 3 różnych algorytmów można stwierdzić, że algorytmy PLS1 oraz PLS2 dały bardzo podobne wyniki, zdecydowanie lepsze niż algorytm PCR. Wartości parametrów walidacyjnych dla modeli KETPLS12 oraz KETPLS22 są praktycznie identyczne. Zatem oznaczanie procentowej zawartości ketoprofenu w żelu Ketoprofen należy zastosować algorytm PLS1 lub PLS2 do widm NIRw zakresie od 6 100 do 5 866 cm^{-1} po przeprowadzeniu mieszanej transformacji widm surowych (1. pochodna + SNV + detrend).

5.3.2. Oznaczanie zawartości filtrów UV w emulsjach do opalania

Model ilościowy dla oznaczanie zawartości filtrów ESCALOL 557 w emulsjach do opalania opracowano w zakresie klasycznej podczerwieni (4 000 - 650 cm^{-1}). Opracowana technika bazuje na iteracyjnym kalibrowaniu modelu chemometrycznego. Kolejne iteracje wykorzystują jako wartości referencyjne wartości stężeń uzyskane w poprzedniej iteracji. Parametrami akceptacji finalnej iteracji są miary statystyczne: *Residual Ratio* oraz *M-Total Distance Ratio* (Odległość Mahalanobiasa, analiza wariancji).

5.3.2.1 Dobór zakresu analizy, analiza widmowa

Zakres analizy wybrano w oparciu o porównanie widm produktu bez zawartości filtrów (placebo) oraz widm produktów końcowych z odpowiednią zawartością filtru. Z **Rys.93** wynika, że w zakresie od 1 600-1 000 cm^{-1} obserwuje się wyraźne różnice wynikające z braku obecności filtru Escalol 557. Ostatecznie model ilościowy opracowano w zakresie od 1 525-1 000 cm^{-1} .



Rys. 92 Surowe widma IR emulsji do opalania z filtrami (czerwone) i bez filtrów (niebieskie)

5.3.2.2. Transformacja surowych widm

Po wyborze analitycznego zakresu liczb falowych dobrano odpowiednie metody transformacji widm. Najlepsze wyniki uzyskano po zastosowaniu zestawu transformacji przedstawionych w **Tab.52**.

Tabela 52. Warunki przeprowadzonej analizy

Zakres analizy (cm^{-1})	1525-1000
Wygładzenie	brak
Korekcja linii bazowej	2 pochodna 9p
Normalizacja	SNV+trend
skalowanie(widm/własności)	auto /średnie

W modelu jako algorytm kalibracyjny wybrano cząstkową metodę najmniejszych kwadratów, PLS1.

5.3.2.3. Optymalizacja i kompleksowość modeli

W zbiorze uczącym wykorzystano widma 107 próbek pochodzących z różnych serii produkcyjnych i laboratoryjnych. Poprawność działania modelu sprawdzano stosując metodę walidacji krzyżowej (ang. *cross validation*). Na tym etapie dokonano również wyboru optymalnej liczby istotnych czynników w modelu PLS. W trakcie optymalizacji wykorzystano wartości wskaźnika dźwigni (ang. *Leverage ratio*) oraz wskaźnika reszt (ang. *Residual ratio*) opisującego stosunek resztkowej wariancji próbki do średniej resztkowej wariancji wzorców. W modelu brano pod uwagę od 3 do 5 czynników.

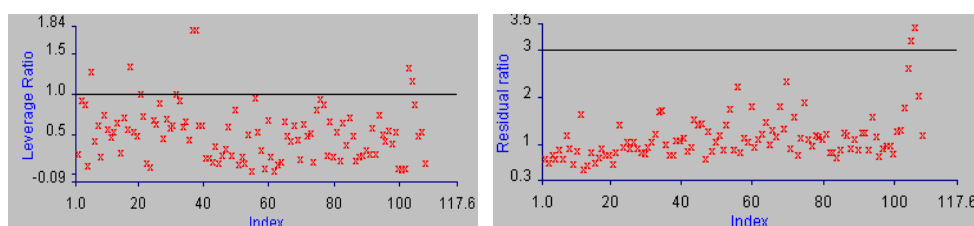
Uzyskane wyniki identyfikacji i walidacji krzyżowej przeprowadzone przedstawiono w **Tab.53**. Na **Rys.93** pokazano wartości.

Tabela 53 Wyniki identyfikacji i walidacji krzyżowej modeli PLS przy różnej liczbie czynników – oznaczanie zawartości filtru Escalol 557

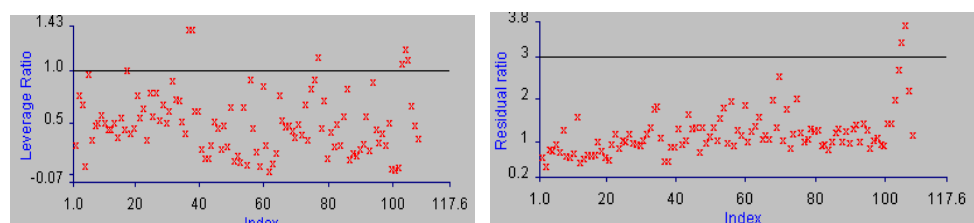
Liczba <i>LV</i>	Zbiór kalibracyjny			Walidacja krzyżowa			
	<i>R</i>	<i>SEC</i>	<i>Bias</i>	<i>R</i>	<i>SEP</i>	<i>Bias</i>	<i>SEC/SEP</i>
3	0,996	0,124	0,00011	0,994	0,151	0,0031	0,82
4	0,997	0,1096	0,011	0,995	0,1446	0,045	0,76
5	0,997	0,1006	0,00006	0,994	0,1472	0,007	0,68

Wyniki otrzymane przy różnej liczbie czynników pokazują, że w miarę zwiększania liczby czynników w modelu zwiększa się wartości *R* i jednocześnie zmniejsza się odległość próbek przekraczających wartości graniczne wskaźnika *dźwigni* oraz wskaźnika reszt od tych wartości, **Rys.93**.

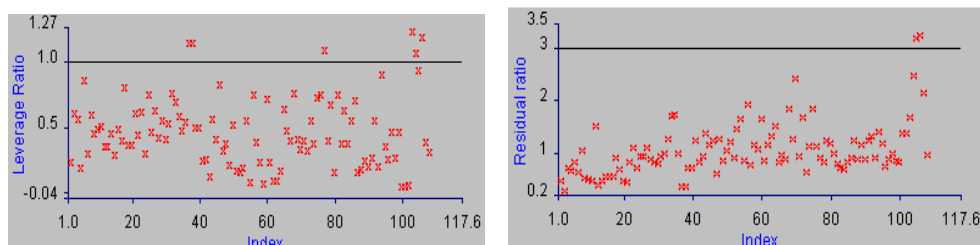
a)



b)



c)



Rys. 93 Wykresy wskaźnika *dźwigni* (panele lewe) oraz wskaźnika reszt (panele prawe) otrzymane przy a) 3, b) 4 i c) 5 czynnikach w modelu PLS

Tabela 54. Liniowość zależności pomiędzy rzeczywistą (y) i oszacowaną na podstawie widm NIR(x) zawartością filtru zgodnie z zależnością $y = ax + b$

Liczba czynników	Zbiór kalibracyjny			Walidacja krzyżowa		
	Liniowość		t -test*	Liniowość		t -test*
	a	b		a	b	
3	0,992±0,017	0,036±0,08	0,009	0,98±0,02	0,082±0,1	0,008
4	0,994±0,015	0,029±0,07	0,011	0,98±0,07	0,071±0,4	0,079
5	0,995±0,014	0,023±0,06	0,012	0,97±0,05	0,078±0,2	0,067

* obliczona wartość statystyki t w teście istotności różnic dla par związanych; wartość krytyczna 1,98

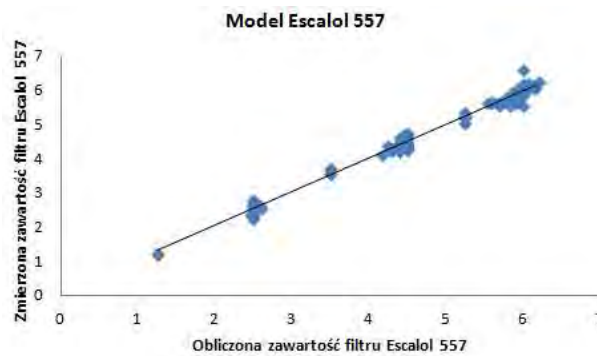
Po zbadaniu liniowości pomiędzy zawartością filtru wyznaczoną metodą referencyjną i oszacowaną na podstawie modelu dla 3 wariantów modelu PLS można uznać te modele za poprawne modele kalibracyjne. Dla wszystkich wariantów przeprowadzono następnie walidację zewnętrzną wykorzystując niezależny zbiór testowy, **tab.54**. Otrzymane wyniki pokazały, że najlepszym wariantem modelu kalibracyjnego dla oznaczania zawartości filtru Escalol 557 w emulsjach do opalania jest model z 3 czynnikami. Wartości odzysku dla tego wariantu nie przekroczyły zakresu $100 \pm 6\%$. Również wartości parametrów statystycznych M -distance oraz $Residual Distance$ nie zostały przekroczone. Tym samym model ten spełnia kryteria wymagane od poprawnych modeli kalibracyjnych w analizie przemysłowej.

Tabela 55 Wyniki walidacji zewnętrznej dla oznaczania zawartości filtrów w emulsjach do opalania

Liczba LV	Odzysk%		Precyzja RSD%	Precyzja pośrednia%	M -D*	R - R **
3	94,2	106	0,79	0,78	0,86	1,38
4	9,2	106	0,83	0,83	0,79	1,63
5	93,9	105,5	0,81	0,81	1,14	0,68

* M -D (M -distance)- opisuje stosunek między odległością Mahalanobiasa, dla próbki i wszystkich współczynników w modelu, oraz wartością graniczną dla tych współczynników, M -D<1

** R -R (Residual Ratio) opisuje stosunek resztkowej wariancji próbki do średniej resztkowej wariancji wzorców R -R<3



Rys. 94 Wykres liniowości odpowiedzi modelu kalibracyjnego do oznaczania zawartości filtru Escalol 557

Warto w tym miejscu podkreślić, że uzyskanie zwalidowanego modelu kalibracyjnego wykorzystującego widma NIR lub IR nie kończy problemów występujących w analityce przemysłowej. Model będzie działał poprawnie dla próbek ze zbioru roboczego tylko wówczas, gdy w widmach tych próbek występować będą te same źródła zmienności co w widmach próbek ze zbioru uczącego i dodatkowo z podobnym natężeniem. Można to sprawdzić obliczając przynajmniej dla niektórych nowych próbek wartości odpowiednich wskaźników takich jak wskaźnik dźwigni i wskaźnik reszt. Wartości tych wskaźników nie powinny systematycznie przekraczać pewnych wartości granicznych. W przypadku wskaźnika dźwigni, wartością graniczną jest 1. Gdy pojawią się próbki o wskaźniku dźwigni znacznie większym niż 1, opracowany model nie będzie dla nich działał poprawnie. W takiej sytuacji należy zwiększyć w modelu liczbę czynników. Jeżeli zmiana ta nie obniży wartości wskaźnika, oznacza to zwykle, że w widmie próbki pojawiły się nowe źródła zmienności, które nie były uwzględnione w zbiorze uczącym. W takiej sytuacji należy wzbogacić zbiór uczący o nowe próby i ponownie przeprowadzić identyfikację i walidację modelu. Podobnie jest w przypadku wskaźnika reszt: próbki robocze o wskaźniku reszt przekraczającym znacznie wartość 3, zwykle zawierają zmienność nie uwzględnioną w modelu kalibracyjnym.

Dobór odpowiedniej liczby istotnych czynników jest również bardzo ważny podczas opracowywania modeli ilościowych. Zbyt duża liczba czynników może doprowadzić do niepoprawnych oczekiwanych wyników a tym samym doprowadzić do wypuszczenia na rynek produktu z zawyżoną bądź obniżoną wartością substancji aktywnej, co mogłoby skutkować niepożądanymi efektami. Zatem opracowując modele ilościowe wykorzystujące algorytmy PCR i PLS należy bacznie obserwować i kontrolować liczbę czynników.

5.3.3. Oznaczanie zawartości anionowych związków powierzchniowo czynnych w składnikach środków myjących

Związki powierzchniowo czynne (ZPC) są to substancje wykazujące aktywność powierzchniową tzn. rozpuszczone w cieczy, nawet przy bardzo niskich stężeniach, obniżają jej napięcie powierzchniowe lub międzyfazowe wskutek adsorpcji na granicy faz. Związki powierzchniowo czynne klasyfikuje się ze względu na rodzaj ich hydrofilowej grupy funkcyjnej, która w dużej mierze decyduje o ich właściwościach i zastosowaniu.

Anionowe związki powierzchniowo czynne, AZPC, to substancje w których część powierzchniowo czynną stanowi ujemnie naładowana grupa hydrofilowa (anion). Do tej grupy związków należą między innymi: siarczany alkilowe, saponiny. Stosowane są one w substancjach myjących takich jak: szampony, mydła i żele pod prysznic.

Ze względu na szybki wzrost produkcji w firmie Ziaja Ltd. środków higieny i kosmetyków zawierających AZPC wystąpiła konieczność opracowania szybkiej i taniej metody ilościowego oznaczania tych związków za pomocą spektroskopii IR. W mojej pracy przedstawiam próbę budowy modelu kalibracyjnego do oznaczania zawartości AZPC w 25% półprodukcie - roztworze dodawanym do środków myjących. Doświadczalnie oznaczono zawartość tych związków w próbkach zbioru uczącego za pomocą tradycyjnej metody miareczkowej. Do budowy modelu wykorzystano widma 46 próbek stanowiące zbiór uczący. Do oceny zdolności predykcyjnych opracowanego modelu wykorzystano metodę walidacji krzyżowej. Opracowany model został sprawdzony i zwalidowany przy pomocy zbioru testowego w skład którego weszły widma 15 próbek produkcyjnych.

Przebadano szereg wariantów wstępnej obróbki widm oraz algorytmów identyfikacji budowanego modelu kalibracyjnego. Parametry wytypowanego ostatecznie modelu przedstawia **Tab.56**.

Tabela 56. Parametry modelu kalibracyjnego dla oznaczanie anionowych związków powierzchniowo czynnych

<i>Algorytm</i>	PCR
<i>Zakres analizy (cm^{-1})</i>	1300-1138
<i>Wyglądzenie</i>	brak
<i>Korekcja linii bazowej</i>	1 pochodna 9p
<i>Normalizacja</i>	brak
<i>Skalowanie</i>	autoskalowanie
<i>Liczba składowych PC</i>	3

Dla tak otrzymanego modelu przeprowadzono walidację wewnętrzną, wykorzystując metodę walidacji krzyżowej. Wyniki otrzymane przedstawiono w **Tab.57.** i **Tab.58**

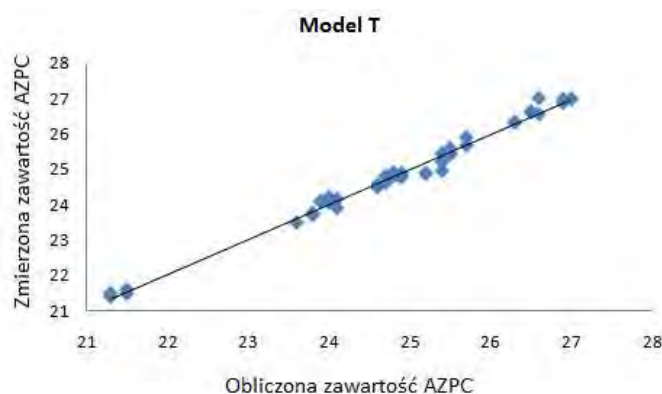
Tabela 57. Wyniki walidacji wewnętrznej modelu kalibracyjnego do oznaczania zawartości AZPC

Zbiór kalibracyjny			Walidacja krzyżowa			
<i>R</i>	<i>SEC</i>	<i>Bias</i>	<i>R</i>	<i>SEP</i>	<i>Bias</i>	<i>SEC/SEP</i>
0,994	0,29	0,0036	0,992	0,31	0,014	0,95

Tabela 58 Liniowość zależności pomiędzy rzeczywistą (*y*) i oszacowaną na podstawie widm NIR(*x*) zawartością AZPC zgodnie z zależnością $y = ax + b$

Zbiór kalibracyjny			Walidacja krzyżowa		
Liniowość		<i>t-test</i> *	Liniowość		<i>t-test</i> *
<i>a</i>	<i>b</i>		<i>a</i>	<i>b</i>	
0,988±0,032	0,28±0,81	0,024	0,97±0,036	0,63±0,9	0,081

* obliczona wartość statystyki t w teście istotności różnic dla par związanych; wartość krytyczna 2,013



Rys. 95 Wykres liniowości odpowiedzi modelu kalibracyjnego do oznaczania zawartości AZPC

Wyniki zamieszczone w **Tab. 57** i **Tab. 58** pokazują, że opracowany model jest dobrze dopasowany i posiada zdolności prognostyczne. Wartości współczynników korelacji dla zbioru uczącego oraz otrzymane podczas walidacji krzyżowej są wysokie, do tego wartości SEC oraz SEP są do siebie zbliżone. Dobre dopasowanie modelu do danych eksperymentalnych potwierdza również wykres na **Rys.95**.

Po zakończeniu walidacji wewnętrznej przystąpiono do przeprowadzenia walidacji zewnętrznej. Wyniki przedstawiono poniżej w **Tab.59**.

Tabela 59. Wyniki walidacji zewnętrznej przeprowadzonej za pomocą zbioru testowego

<i>Liczba LV</i>	<i>Odzysk%</i>		<i>Precyzja</i>	<i>Precyzja</i>
	<i>MIN</i>	<i>MAX</i>	<i>RSD%</i>	<i>pośrednia%</i>
3	97,9	102,4	0,69	1,32

Otrzymane wyniki pokazują, że opracowany model kalibracyjny do oznaczanie anionowych związków powierzchniowo czynnych działa poprawnie i może być stosowany w rutynowej analityce przemysłowej. Wartości oszacowane przez model są zbliżone do wartości wyznaczonych metodą referencyjną, a wartości odzysku mieszczą się w założonym przedziale $100\pm 3\%$.

VI Podsumowanie

Pomimo wielu zalet, spektroskopia w zakresie bliskiej podczerwieni wciąż jest w Polsce mało popularną metodą w analityce przemysłowej. Bardzo wiele osób ma duże uprzedzenia do tej nowej metody analitycznej. Podstawową przyczyną takiego stanu rzeczy jest przywiązanie wielu analityków do tradycyjnych metod analitycznych, opartych na prostej statystyce i zależnościach między zawartością oznaczanego składnika a intensywnością absorpcji promieniowania przy pojedynczej długości fali (liczbie falowej). Metody analityczne w zakresie NIR oraz IR wymagają odmiennego podejścia do zarejestrowanych widm. Metody te wymagają zastosowania technik chemometrycznych – narzędzia nowego dla większości analityków.

W literaturze przedmiotu znaleźć można ponadto stwierdzenia, że uzyskanie prawidłowych modeli do analizy jakościowej i ilościowej z wykorzystaniem widm NIR lub IR wymaga spełnienia szeregu specyficznych warunków. Przy czym lista tych warunków nie jest jednoznacznie ustalona, podobnie jak występujące w nich wartości krytyczne. Stanowi to dodatkowy czynnik zwiększający nieufność w stosunku do omawianych metod. Przedmiotem mojej pracy doktorskiej jest ustalenie tych warunków na etapach:

- doboru próbek,
- warunków wykonania rejestracji widm,
- wstępnej obróbki widm, oraz
- decyzji o stopniu komplikacji stosowanych modeli chemometrycznych.

Nie rozwiązuje to oczywiście wszystkich problemów pojawiających się przy praktycznym wykorzystaniu spektroskopii NIR/IR w analityce przemysłowej.

Wykazałam, że na kształt i intensywność rejestrowanych widm NIR wpływa nie tylko postać i skład próbki, ale również warunki przeprowadzenia pomiaru. Warunki jakie należy spełnić dla uzyskania poprawnych modeli chemometrycznych można podsumować w postaci następujących zaleceń:

- ważne jest aby badaną próbkę umieszczać na okienku pomiarowym przyrządu w taki sposób, aby wiązka promieniowania w całości przechodziła przez badaną próbkę. Niestaranne umieszczone próbki na oknie pomiarowym wprowadza do modelu dodatkową zmienność, która ma duży wpływ na jakość uzyskanych modeli.
- warunki panujące w pomieszczeniu podczas rejestracji widm, takie jak wilgotność,

temperatura czy zawartość CO₂, mają znaczący wpływ na kształt rejestrowanych widm. Wpływu zmienności tych czynników nie da się w praktyce wyeliminować należy je więc uwzględnić przy tworzeniu modelu. Widma wykorzystywane przy tworzeniu modeli należy zatem rejestrować w różnych porach dnia oraz roku, gdy temperatura i wilgotność powietrza w laboratorium są zróżnicowane. Modele oparte na widmach zarejestrowanych np. jednego dnia będą działać fałszywie negatywnie w przypadku widm zarejestrowanych przy innej pogodzie lub w innych porach dnia czy roku.

- wykazałam, że liczba skanów przy rejestracji widma nie ma praktycznie wpływu na kształt widma. W celu skrócenia czasu analizy można rejestrować widma próbek już przy 3 skanach bez pogorszenia jakości uzyskiwanych na ich podstawie modeli. Należy jednak zdawać sobie sprawę, że w takich przypadkach duży wpływ na jakość modelu ma prawidłowy dobór zakresu widma: w wybranym zakresie powinny się znajdować możliwie silne pasma charakterystyczne dla analizowanej substancji.

Warunkiem uzyskania poprawnych modeli jakościowych i ilościowych bazujących na widmach NIR/IR jest odpowiednia obróbka wstępna tych widm (preprocessing). Na podstawie przeprowadzonych przeze mnie badań można sformułować kilka zaleceń, których przestrzeganie powinno znacząco zwiększyć prawdopodobieństwo uzyskania poprawnego i użytecznego modelu. Dotyczą one:

- *umiejętnego doboru zakresu widma wykorzystywanego do tworzenia modelu.* Zły dobór zakresu widma, nawet przy odpowiedniej transformacji widm surowych, nie przyniesie zadowalających rezultatów. Wybrany zakres widma powinien w miarę możliwości zawierać silne i charakterystyczne dla danej substancji pasma absorpcji. Dużą pomocą przy wyborze zakresu analitycznego mogą być tzw. wykresy dźwigni. Szczególnie wartościowe są wykresy dźwigni uzyskane dla małej liczby uwzględnianych głównych składowych. Skupienie uwagi na kilku pierwszych głównych składowych pozwala wyeliminować wpływ przypadkowych różnic w kształcie i intensywności widm surowych i skupić uwagę na głównych źródłach zmienności.
- *wyboru algorytmu transformującego.* Uzyskane przeze mnie wyniki wskazują, że dobór algorytmu transformującego ma kluczowy wpływ na jakość opracowywanych modeli tożsamościowych i kalibracyjnych. Najlepsze wyniki

uzyskuje się zwykle stosując dwie funkcje transformujące. Przy tym jedna z nich powinna mieć charakter transformacji normalizującej natężenie widm (np. transformacje SNV lub MSC), a druga powinna ograniczać wpływ zmiany kształtu i położenia linii bazowej (np. transformacja de trend) lub numeryczne wyznaczenie 1. lub 2. pochodnej). Wykazałam, że zastosowanie normalizacji ma szczególne znaczenie w przypadku próbek stałych o różnym stopniu granulacji. W przypadku surowców płynnych normalizacja nie przynosi zwykle poprawy jakości modelu. Użycie widm NIR z umiejętnie dobranymi algorytmami transformującymi umożliwia otrzymanie zaskakująco użytecznych modeli.

Przedstawione w pracy wyniki wskazują wyraźnie, że jednym z najważniejszych czynników mających wpływ na jakość uzyskiwanych modeli chemometrycznych jest właściwy dobór próbek.

W przypadku modeli tożsamościowych dotyczy to dwóch zagadnień:

- podziału zestawu widm na zbiór uczący, walidacyjny i testowy
- postępowania z widmami mającymi charakter punktów odbiegających.

Wykazałam w swojej pracy, że widma tworzące zbiór uczący powinny zawierać wszystkie źródła zmienności, które mogą wystąpić w zbiorze roboczym. Powinny się więc charakteryzować możliwie dużą różnorodnością. Warunek ten można spełnić uwzględniając w zbiorze uczącym widma próbek z różnych szarż zarejestrowane w różnych warunkach otoczenia (w różnych porach dnia i roku).

Istotnym zagadnieniem może też być problem pojedynczych widm odbiegających kształtem od pozostałych, tzw. punktów odbiegających. Wykazałam, że widma takie mogą mieć zarówno korzystny jak i niekorzystny wpływ na jakość opracowywanego modelu. Widm takich nie należy od razu usuwać ze zbioru danych. Należy najpierw dokładnie sprawdzić czy nie są one wynikiem błędu operatora lub złego funkcjonowania aparatu. Po wyeliminowaniu tych przyczyn należy sprawdzić w jakim stopniu wpływają one na stopień komplikacji modelu i na podział zbioru danych na klasy. Dopiero stwierdzenie negatywnego wpływu takiego widma na jakość modelu stanowić może przesłankę do jego wykluczenia ze zbioru danych.

Podczas użytkowania modelu należy też zwracać uwagę na występowanie widm odbiegających w zbiorze roboczym. Jeżeli kilkakrotnie stwierdzimy występowanie widm odbiegających o podobnym charakterze, to należy opracować nowy model tożsamościowy uwzględniając te widma w zbiorze uczącym.

W przypadku modeli ilościowych (kalibracyjnych) zagadnienie sprowadza się zwykle do problemu odpowiedniego przygotowania próbek o dostatecznym zróżnicowaniu oznaczanej substancji. Możliwe są tu trzy podejścia:

- 1 analizowane próbki ze swej natury charakteryzują się dostatecznie dużą zmiennością
- 2 próbki naturalne charakteryzują się małą zmiennością
- 3 do zbudowania modelu kalibracyjnego wykorzystujemy próbki laboratoryjne o z góry zaplanowanym składzie

Każde z tych podejść posiada swoje wady i zalety. Podejście 1. prowadzi zwykle do zbioru uczącego w którym większość próbek posiada zbliżoną wartość oznaczanej wielkości, a jedynie nieliczne mają wartości zdecydowanie mniejsze lub większe. Model kalibracyjny zbudowany na takich danych jest więc bardzo uzależniony od pojedynczych, skrajnych próbek. Podejście 2. nie tylko może być stosowane przy małej zmienności próbek naturalnych, ale również umożliwia skorygowanie rozkładu zróżnicowanych próbek naturalnych. Wydaje się, że jest ono najbardziej właściwe i dlatego wykorzystałam je w swojej pracy. Z kolei podejście 3. pozwala na staranne, optymalne zaplanowanie zróżnicowania próbek. Jego podstawową wadą jest wykorzystanie jedynie próbek laboratoryjnych, co grozi eliminacją źródeł zmienności powstających podczas rzeczywistego procesu technologicznego.

Praktyczne zastosowanie w analityce przemysłowej znajdują jedynie modele chemometryczne charakteryzujące się dobrymi właściwościami prognostycznymi. Przyjmuje się powszechnie, że właściwości takie wykazują przede wszystkim modele o niskim stopniu komplikacji, czyli bazujące na możliwie małej liczbie głównych składowych (modele tożsamościowe SIMCA) lub czynników (modele kalibracyjne PLS). Uzyskane przeze mnie wyniki pokazały, że bardzo ważny jest poprawny dobór liczby istotnych czynników. Zarówno zbyt mała jak i zbyt duża liczba istotnych czynników prowadzi do modeli o niskich zdolnościach predykcyjnych.

Wykazałam również, że na stopień komplikacji modelu istotny wpływ ma dobór funkcji transformujących widma surowe. Wyeliminowanie na tym etapie zmienności wynikającej ze zmienności linii bazowej pozwala na znaczące obniżenie stopnia komplikacji modelu.

W niniejszej pracy wykazałam ponadto, że metody chemometryczne opracowane dla potrzeb spektroskopii NIR można z powodzeniem zastosować do widm w zakresie

klasycznej podczerwieni (IR). Dotyczy to w szczególności modeli kalibracyjnych. Jedyne ograniczenie wynika z natury samych widm: widma w zakresie IR zawierają przede wszystkim informacje o występowaniu w badanej próbce określonych grup funkcyjnych i ewentualnie ich względnej zawartości podczas gdy widma NIR dostarczają głównie informacji o postaci w jakiej występują składniki próbki. Dlatego modele kalibracyjne IR można z powodzeniem stosować, gdy postać oznaczanej substancji nie ma istotnego znaczenia, a więc przede wszystkim do próbek w stanie płynnym. W przypadku gdy w trakcie analizy kontrolowana ma być zarówno postać produktu jak i stężenie wybranej substancji należy zastosować model w zakresie NIR.

Uzyskane przeze mnie wyniki wskazują, że spektroskopia w zakresie podczerwieni (zarówno NIR jak i IR) wraz z odpowiednio dobranymi technikami chemometrycznymi może być z powodzeniem stosowana w analizie tożsamościowej i ilościowej surowców, półproduktów i produktów kosmetycznych oraz farmaceutycznych. Wykorzystanie tych technik analitycznych wraz z metodami chemometrycznymi w laboratorium firmy Ziaja Ltd. znacznie skróciło czas rutynowych analiz oraz ograniczyło ilość zużytych odczynników.

VII LITERATURA

1. M.Kolbuszewski "Spektroskopia bliskiej podczerwieni w laboratorium" Laboratorium 11-12, 2009
2. C. Pasquini „Near Infrared Spectroscopy: Fundamentals, Practical Aspects and Analytical Applications” J.Braz.Chem.Soc.14, 198-219, 2003
3. P. H. Hindle “Towards 2000 The past, present and future on-line NIR sensing”, Process control and quality 9, 105-115, 1997
4. E.W.Ciurczak, J.K.Drennen III “Pharmaceutical and Medical application of Near-Infrared Spectroscopy” Marcel Dekker, New York, 2002
5. <http://www.wiw.pl/astrologia/a-herschel.asp>
6. P.H. Hindle “The past, present and future of on-line NIR sensing” *Process Control Qual.* 9, 105-115, 1997
7. L. Piela „Idee chemii kwantowej” Wydawnictwo naukowe PWN, Warszawa 2006
8. Materiały konferencyjne, “NIR-zastosowanie w analizie chemicznej” Poznań, 10-11 maj 2006
9. W. Szczepaniak „Metody instrumentalne w analizie chemicznej” Wydawnictwo naukowe PWN, Warszawa 1996
10. Zieliński „Metody spektroskopowe i ich zastosowanie do identyfikacji związków organicznych” Wydawnictwo Naukowo-Techniczne, Warszawa 2000
11. <http://www.mif.pg.gda.pl/kfze/wyklady/IM3rozdzial3.pdf>
12. J. Workman, Jr.L. Weyer “ Practical guide to interpretive Near-Infrared Spectroscopy”
13. E. Barton II “Theory and principles of near infrared spectroscopy”
14. J. Luypert, D.L. Massart, Y. V. Heyden “Near-Infrared spectroscopy application in pharmaceutical analysis” Tal. 72, 865-883, 2007
15. Y. Roggo, P.Chalus, L.Maurer, C.Lema-Martinez, A.Edmond, N.Jent “A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies” *J.Pharm.Biomed.Anal.* 44,683-700,2007
16. R. Kramer “Chemometric techniques for quantitative analysis” 1998
17. Szafran “Spektroskopia FT-IR w określaniu struktury oraz analizie jakościowo-ilościowej związków organicznych : NIR-zastosowanie w analizie chemicznej” Ogólnopolska Konferencja Naukowa, materiały konferencyjne, Poznań 2006

18. K. Buhłak "Spektroskopia w bliskiej podczerwieni (NIR) jako szybka i skuteczna metoda oceny jakości surowców i produktów" *Towaroznawcze Problemy Jakości* 3(8), 33-43, 2006
19. N. Broad, P. Graham, P. Hailey, A. Hardy, S. Holland, S. Hughes, D. Lee, K. Prebble, N. Salton, P. Warren "Guidelines for the Development and validation of Near-infrared Spectroscopic Methods in the Pharmaceutical Industry" *Handbook of Vibrational Spectroscopy*, John Wiley & Sons Ltd. Chichester 2002
20. P. Turner, „FT- Near Infrared Spectroscopic Determination of Hydroxyl number using a fiber optic probe”, Bruker Spectroscopy Ltd., Coventry UK 1999
21. M. Strzelińska „Oznaczanie kwasu askorbinowego w preparatach farmaceutycznych za pomocą spektroskopii w bliskiej podczerwieni” *Praca doktorska*, Poznań 2008
22. *Farmakopea Polska* wyd. VII Tom I, Rozdział 2.2.40, 151-155, 2006
23. M. Watari „Applications of Near Infrared Spectroscopy to Process Analysis Using Fourier Transform Spectrometer” *Optical Review* 17, 317-322, 2010
24. H. Owen-Reece, M. Smith, C.E. Elwell, J.C. Goldstone "Near infrared spectroscopy" *Br. J. Anaesth* 82, 418-26, 1999
25. J. Workman "The state multivariate thinking for scientist in industry:1980-200" *Chem.Intel.Lab.Syst.* 60, 13-23, 2002
26. P.K. Hopke "The evolution of chemometrics" *Anal. Chim.Acta* 500, 365-377, 2003
27. W. Wang, J. Paliwa „Near-infrared spectroscopy and imaging in food quality and safety" *Food Qual.* 11, 193-207, 2007
28. Pharmaceutical Analytical Science Group "Guidelines for the development and validation of near infrared (NIR) spectroscopic methods", October 2001
29. The European Agency for the evaluation of Medicinal Products "Note for Guidance on the use of near infrared spectroscopy by the pharmaceutical industry and the data requirements for new submissions and variations" February 2003
30. A. Candolfi, D.L. Massart, S. Heureding "Investigation of sources of variance which contribute to NIR-spectroscopic measurement of pharmaceutical formulation" *Anal. Chim. Acta* 345, 185-196, 1997
31. A. Rinnan, F van den Berg, S.B. Engelsen "Review of the most common pre-processing techniques for near – infrared spectra" *Trends in Anal. Chem.* 28,1201-1222, 2009

32. Y.Dou, T.Zou, T.Liu, N.Qu, Y.Ren, "Calibration in non-linear NIR spectroscopy using principal component artificial neural networks" *Spectrochimica Acta Part A*, 68, 1201-1206, 2007
33. Q. Chen, J. Zhao, H. Zhang, X. Wang "Feasibility study on qualitative and quantitative analysis in tea by near infrared spectroscopy with multivariate calibration" *Analytic Chimica Acta*, 572, 77-84, 2006
34. M.R.Maleki, A.M.Mouazen, H.Ramon, J.De Baerdemaeker "Multiplicative Scatter Corection during On-line Measurement with Near Infrared Spectroscopy" *Biosys. Eng.* 96(3), 427-433, 2003
35. P.Chalus, Y.Roggo, S.Walter, M.Ulmschneider "Near-infrared determination of active substance content in intact low-dosage tablets" *Tal.* 66, 1294-1304, 2005
36. J.Luyper, S.Heuerding, Y.V.Heyden, D.L.Massart "The effect of preprocessing methods in reducing interfering variability from near-infrared measurements of creams" *J. Pharm. Biomed. Anal.* 36, 495-503, 2004
37. M.Zeaiter, J.M.Roger, V.Bellon-Maurel, "Robustness of models developed by multivariate calibration. Part II: The influence of pre-processing methods" *Tren. Anal. Chem.* 24, 437-445, 2005
38. J.Madan, A.K.Diwivedi, S.Singh "Estimation of antitubercular drugs combination in pharmaceutical formulations using multivariate calibration" *Anal. Chim. Acta* 538, 345-353, 2005
39. H.Mark, J.Workman "Chemometrics in spectroscopy" *Wydawnictwo Elsevier*, Part 3, 363-369,
40. P.Geladi "Chemometrics in spectroscopy. Part B. Classical chemometrics" *Spectrochim. Acta Part B* 58, 767-782, 2003
41. S.K.Setarehdan, J.J.Soraghan, D.Littlejohn, D.A.Sadler "Maintenance of calibration model for near infrared spectrometry by combined principal component analysis-partial least squares approach" *Anal. Chim. Acta*, 452, 35-45, 2002
42. J.Mazerski "Podstawy chemometrii" *Wydawnictwo Politechniki Gdańskiej*, Gdańsk 2004
43. W.Chenglin, W.Tianzhen, H. Jing "Relative principle component and relative principle component analysis algorithm" *J. Electronics* 24, 108-111, 2007
44. T.Skov, R.Bro "A New approach for modeling sensor based data" *Sensor and Actuators B*, 106, 719-729, 2005

45. W.Wu, D.L.Massart, S.de Jong, "Kernel-PCA algorithms for wide data. Part II:Fast cross-validation and application in classification of NIR data", *Chemom. Intel. Lab.Syst.* 37, 271-280
46. Y.Roggo, P.Chalus, L.Maurer, C.Lema-Martinez, A.Edmond, N.Jent "A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies" *J. Pharm. and Biomed.Anal.* 44, 683-700, 2007
47. T.Li, S. Zhu. M.Ogihara "Using discriminant analysis for multi-class classification: an experimental investigation" *Knowl. Inf. Syst.* 10, 453-472, 2006
48. M. Casale, N. Sinelli , P. Oliveri , V. Di Egidio, S. Lanteri "Chemometrical strategies for feature selection and data compression applied to NIR and MIR spectra of extra virgin olive oils for cultivar identification" *Talanta* 80, 1832-1837, 2010
49. R. M. Peña, aS. García, R. Iglesias, S. Barro, C. Herrero "Authentication of Galician (N.W. Spain) quality brand potatoes using metal analysis. Classical pattern recognition techniques versus a new vector quantization-based classification procedure" *Analyst* 126, 2186-2193, 2001
50. S. Tsuchikawa, · K. Yamato, · K. Inoue "Discriminant analysis of wood-based materials using near-infrared spectroscopy" *J. Wood Sci.* 49, 275-280, 2003
51. R.De Maesschalck, A.Candolfi, D.L.Massart, S.Heuerding "Decision criteria for sort independent modeling of class analogy applied to near infrared data" *Chemom. Intel. Lab.Syst.* 47, 65-77, 1999
52. A.Candolfi, R.De Maesschalck, D.L.Massart, P.A, Hailey "Identification of pharmaceutical excipients using NIR spectroscopy and SIMCA" *J. Pharm. Biomed. Anal.* 19, 923-935, 1999
53. M.daszykowski, K.Kaczmarek, I.Stanimirova, Y.V. Heyden, B.Walczak "Robust SIMCA-bounding influence of outliers" *Chemom. Intel. Lab.Syst.*
54. H. Huang, H. Yu, H. Xu, Y. Ying "Near infrared spectroscopy for on/in-line of quality in foods and beverages:A review" *J. Food Eng.* 87, 303-313, 2008
55. S.Kawano, H. Watanabe, M.Iwamoto, „Determination of Sugar Content In Intact Peaches by Near Infrared Spectroscopy with Fibre-Optics in Interactance Probe" *J. Japan Soc.Hort.Sci* 61, 445-451, 1992
56. W.Wang, J. Paliwal "Near-infrared spectroscopy and imaging in food quality and safety" *Since Instr. Food Qual.* 1, 193-207, 2007

57. Dae-weon Lee, K. W. Seo, B. R. Min, Discrimination between Cancer and Normal Tissue using Near Infrared Spectroscopy” World Congress on Medical Physics and Biomedical Engineering 2006, August 27 – September 1, 2006 COEX Seoul, Korea
58. G.Reich “Near-infrared spectroscopy and imaginig: Basic principles and pharmaceutical application” *Advanced drug delivery* 57, 1109-1143, 2005
59. M.A. Arnold, J.J. Burmeister H.Chung “Phantoms for noninvasive blood glucose sensing with near infrared transmission spectroscopy” *Pphotochem. Photobiol.* 67,50,1998
60. F. Malin, T. L. Ruchti, T. B. Blank, S. N. Thennadil, S. L. Monfre “Noninvasive Prediction of Glucose by Near-Infrared Diffuse Reflectance Spectroscopy” *Clinical Chem.* 45, 1651-1658, 1999
61. M. Blanco, M.A. Romero, „Near-infrared libraries in the pharmaceutical industry: a solution for identity confirmation” *Analyst* 126, 2212, 2001
62. C.Demir, R. G. Bereton „Multivariate calibration on designed mixtures of four pharmaceuticals” *Analyst* 123, 181-189, 1998
63. P. Geladi „Some recent trends in calibration literature ”*Chemom. Intel. Lab. Syst.* 60,211-224, 2002
64. M.Forina, M. C. Casolino, C. de la P. Martinez „Multivariate calibration: application to pharmaceutical analysis” *J. Pharm. Biomed. Anal.* 18, 21-33, 1998
65. M.Zeaiter, J.M.Roger, V.Bellon-Maurel, “Robustness of models developed by multivariate calibration. Part II:The influence of pre-processing methods” *Trend.Anal.Chem.* 24, 437-445(2005)
66. M. Blanco, A. Peguro “Influence of physical factors on accuracy of calibration models for NIR spectroscopy” *J. Pharm. Biomed. Anal.* 52, 59-65, 2010
67. A.Fischer “Torzenie metod ilościowych i ich walidacja: NIR-zastosowanie w analizie chemicznej” , Ogólnopolska Konferencja Naukowa, materiały konferencyjne, Poznań 2006
68. D.Jouan-Rimbaud, D.L.Massart, C.A.Saby, C. Puel, “Characterisation of the representativity of selected sets of samples in multivariate calibration and patter recognition” *Anal. Chim. Acta* 350, 149-161,1997
69. D.Jouan-Rimbaud, D.L.Massart, C.A.Saby, “Determination of the representativity between two multidimensional data sets by a comparison of their structure” *Chemom. Intel. Lab. Syst.* 40, 129-144, 1998

70. I. Stanimirova, M. Daszykowski, B. Walczak "Metody uczenia z nadzorem- Kalibracja, Dyskryminacja I Klasyfikacja" Chemometria w Analityce, wydawnictwo IES 2008
71. M. Daszykowski, B. Walczak, D.L. Massart "Representative subset selection" Anal. Chim. Acta 468,91-103,2002,
72. P.J. de Groot, G.J. Postma, W.J. Melssen, L.M.C. Buydens "Selecting a representative training set for the classification of demolition waste using remote NIR sensing" Anal. Chim. Acta 392, 67-75, 1999
73. D.Xiang, J.Berry, S.Buntz, P.Gargiulo, J.Cheney, Y.Joshi, B.Wabuye, H.Wu, M.Hamed, A.S.Hussain, M.A.Khan, "Robust calibration design in the pharmaceutical Quantitative measurements with near-infrared (NIR) spectroscopy: Avoiding the chemometric pitfalls" J. Pharm. Scien. 10,1-12,2008
74. Q.S. X, Y.Z Liang "Monte carlo cross validation" Chemom. Intel. Lab. Syst. 56,1-11,2001
75. M.P.Gomez-Carracedo, J.M.Andrade, D.N.Rutledge, N.M. Faber "Selecting the optimum number of partial least squares components for the calibration of attenuated total reflectance-mid-infrared spectra of undesigned kerosene Samales" Anal. Chim. Acta 585, 253-265,2007
76. T.Naes, T.Isaksson, T.Fearn, T.Davies "A user friendly guide to multivariate calibration and classification" NIR publication Chester 2004
77. R.G.Brereton „Introduction to multivariate calibration in analytical chemistry" Analyst 125, 2125-2154, 2000
78. D.J.Rimbaud, B.Walczak, D.L Massart, I.R. Last, K.A. Prebble "Comparison of multivariate methods based on latent vectors and methods based on wavelength selection for the analysis of near infrared spectroscopic data" Anal. Chim. Acta 304, 285-295, 1995
79. K.Varmuza, P.Filzmoser, „Introduction to multivariate statistical analysis in chemometrics" CRC Press, 2009
80. B.Walczak „Outlier detection in multivariate calibration" Chemom. Intel. Lab. Syst. 28, 259-272, 1995
81. D.J.Rimbaud, E.Bouveresse, D.L. Massart, O.E.de Noord "Detection of prediction outliers and inliers in multivariate calibration" Anal. Chim. Acta 388, 283-301, 1999

82. M.J. Vredenburg, P.W.J. Caspers, R. Hoogerbrugge, D.M. Barends, „Choice and validation of near infrared spectroscopic application for the identity control of starting materials. Practical experience with the EU draft Note for Guidance on the use of near infrared spectroscopy by the pharmaceutical industry and the data to be forwarded in part II of dossier for marketing authorization”, *Eu. J. Pharm. Biopharm.*, 56, 489, 2003
83. S.Kawano T.Fijiwara, M.Iwamoto “Nondestructive Determination of Sugar Content in Satsuma Mandarin using Near Infrared (NIR) Transmittance” *J.Japan Soc. Hort.Sci* 62, 465-470,1993
84. M.A. Dempster, B.F. MacDonald, P.J. Gemperline, N.R. Boyer, „A near-infrared reflectance analysis method for the noninvasive identification of film-coated and non-film-coated, blister-packed tablets”, *Anal. Chim. Acta*, 310, 43, 1995
85. A.D. Patel, P.E Luner, M.S. Kemper, Low-level determination of polymorph composition in physical mixtures of near-infrared reflectance spectroscopy” *J. Pharm. Sci* 90, 360-370, 2001
86. A.D.Patel, P.E. Luner, M.S. Kemper “Quantitative analysis of polymorphs in binary and multi-component powder mixtures by near-infrared reflectance spectroscopy” *Int. J. Pharm* 206, 63-74, 2000
87. T. Norris, P. K. Aldridge, S. S. Sekulic “Determination of End-points for Polymorph Conversions of Crystalline Organic Compounds Using On-line Near-infrared Spectroscopy” *Analyst* 122, 549-552, 1997
88. W.Dziki, J.F. Bauer, J.J. Szpylman, J.E.Quick, B.C. Nichols”The use of near-infrared spectroscopy to monitor the mobility of water within the sarafloxacin crystal lattice” *J. Pharm.Biomed. Anal.* 22, 829, 2000
89. O.Brentsen, G. Zackrisson, G. Ostling, “Determination of moisture in hard gelatin pharmaceuticals” *J.Pharma. Biomed. Anal.* 15, 895, 1997
90. J.D. Kirsch, J.K Drennen “Nondestructive tablet hardness testing by near-infrared spectroscopy: a new and robust spectral best-fit algorithm” *J. Pharm. Biomed, Anal.* 19, 351-362, 1999
91. Y.Chen, S.S Thosar. R.A Forbess, M.S Kemper, R.L Rubinowitz, A.J. Shukla “Prediction of drug content and hardnes of intact tablets using artificial neural network and near-infrared spectroscopy” *Drug Dev. Ind. Pharm.* 27, 623-631, 2001

92. M. Blanco, I. Villarroya "NIR spectroscopy: a rapid – response analytical tool" Trends in Anal. Chem. 21, 240-250, 2002
93. H. Huang, H. Yu, H. Xu, Y. Ying, "Near infrared spectroscopy for on/in-line monitoring of quality in foods and beverages :A review" *J. Food Eng.* 87, 303-313, 2008
94. T. Isaksson, B.N. Nilsen, G. Tøgersen, R. P. Hammond, K.I. Hildrum, "On-line proximate analysis of ground beef directly at a meat grinder outlet" *Meat Sci.* 43, 245-253, 1996
95. G. Tøgersen, J.F. Arnsen, B.N. Nilsen, K.I. Hildrum, "On-line prediction of chemical composition of semi-frozen ground beef by noninvasive NIR spectroscopy" *Meat Sci.* 63, 515-523, 2003
96. H. Ding, R.J. Xu, D.K.O. Chan "Identification of broiler chicken meat using a visible/near-infrared spectroscopic technique" *J. Sci. Food Arg.* 79, 1382-1388, 1999
97. F. Tang, H. Tao "Fast linear discriminant analysis using binary bases" *Pattern recognition Letters* 28, 2209-2218 2007
98. J.F. Kennedy, C.A. White, A.J. Browne "Application of near infrared reflectance spectroscopy to the analysis of milk and dairy products" *Food Chem.* 16, 115-131, 1985
99. B.G Osborne "Near infrared Spectroscopy in Food Analysis" *Encyclopedia of analytical Chemistry*
100. S. Kawamura, M. Tsukahara, M. Natsuga, K. Itoh, "Online near infrared spectroscopic sensing technique for assessing milk quality Turing milk" Transactions of the ASAE
101. R. Jankovska, K. Sustova "Analysis of cow milk by near-infrared spectroscopy" *Czech J. Food. Sci* 21, 123-128, 2003
102. C. Blazquez, G. Downey, C. O'Donnell, D. O'Callaghan, V. Howard, "Prediction of moisture, fat and inorganic salts in processed cheese by near infrared reflectance spectroscopy and multivariate data analysis" *J. Near Infrared Spectrosc.* 12, 149, 2004.
103. S. Kawano, H. Watanabe, M. Iwamoto, "Determination of Sugar Content In Intact Peaches by Near Infrared Spectroscopy with Fibre-Optics in Interactance Probe" *J. Japan Soc. Hort. Sci* 61, 445-451, 19

104. W. Wang, J. Paliwal “Near-infrared spectroscopy and imaging in food quality and safety” *Since Instr. Food Qual.* 1, 193-207, 2007
105. M. Blanco, M.A. Romero “Near infrared transreflectance spectroscopy determination of dexketoprofen in hydrogel” *J. Pharm. Biomed. Anal.* 30, 497-472, 2002
106. C. Bodson, E. Rozet, E. Ziemons, B. Eurard, Ph. Hubert, L. Delattre „Validation of manufacturing process of Diltazen HCl tablets by NIR spectrophotometry (NIR)” *J. of Pharmaceutical and Biomedical Analysis* 45 (2007) 356 – 361
107. M. Blaneo, J. Coelle, H. Iturriaga, S. Maspah, N. Pou “Development and validation of a near infrared methods for the analytical control of pharmaceutical preparation in three steps of the manufacturing process” *J. Anal. Chem.* (2000) 368: 534-539
108. EMEA The European Agency for the Evaluation of Medicinal Product “Note for guidance on the use of Near infrared spectroscopy by the pharmaceutical industry and the data requirements for new submissions and variations”, London 20 February 2003 EMEA / CV MP/ 961 / 01
109. EMEA European Medicines Agency “Guideline on the use of Near infrared spectroscopy by the pharmaceutical industry and the data requirements for new submissions and variations” London, 16 February 2009
110. Pharmaceutical Analytical Science Group Guidelines for develop and validation of near infrared Spectroscopy, October 2001
111. M. Howard, Gary E. Ritchie, R.W. Roller, E.W. Ciurczak, C. Tso, S.A. MacDonald “Validation of near – infrared transmission spectroscopic procedure, part A: validation protocols” *J. of Pharm. And Biomed. Anal.* 28 (2002) 251 – 260
112. M. Blanco, M. Alcalá, M. Bautista “Pharmaceutical gel analysis by NIR spectroscopy. Determination of active principle and low concentration of preservatives” *J. Pharm. Sciences* 33, 409-414, 2008
113. http://www.thermo.com/eThermo/CMA/PDFs/Articles/articlesFile_1193.pdf
114. M. S. Kemper, E. J. Magnuson, S. R. Lowry, W. J. McCarthy “Use of FT-NIR Transmission Spectroscopy for the Quantitative Analysis of an Active Ingredient in a Translucent Pharmaceutical Topical Gel Formulation” *AAPS Pharm. Sci.* 3, 1-5, 2001
115. M. Daszykowski, B. Walczak „Analiza czynników głównych i inne metody eksploracji danych” *Chemometria w Analityce*, pod redakcją D. Zuby, A. Parczewskiego, Kraków 2008

VIII STRESZCZENIE

W ostatnich latach nastąpił znaczący wzrost wymagań dotyczących kontroli jakości we wszystkich sektorach przemysłowych. Coraz większa liczba analiz sprawia, że tradycyjne, czasochłonne metody wymagają użycia drogich i szkodliwych odczynników, czyniąc je niekorzystnymi.

Istnieje jednak bardzo szybka metoda analityczna coraz częściej wykorzystywana w analizie jakościowej i ilościowej. Jest to spektroskopia bliskiej podczerwieni (NIR). Spektroskopia NIR opiera się na absorpcji promieniowania elektromagnetycznego w zakresie 780-2500 nm. Absorpcja NIR pochodzi od nadtonów i drgań kombinacyjnych wiązań C-H, N-H, O-H. Zaletami wykorzystywania technologii NIR są: 1) nieniszczący, nieinwazyjny charakter analizy, 2) brak konieczności przygotowania próbki przed analizą, 3) prostota i szybkość wykonania rutynowej analizy, 4) wysoka czułość 5) niski koszt 6) uzyskanie kompleksowego obrazu badanego materiału, 7) zwiększenie bezpieczeństwa pracy 8) ograniczenie zużycia rozpuszczalników do przygotowania próbek przed analizą, co ma korzystny efekt proekologiczny. Oprócz samych zalet technika ta również posiada pewne wady. Surowe widma są dość trudne do interpretacji i wymagają znajomości zaawansowanych technik chemometrycznych.

Celem i przedmiotem badań było opracowanie procedury postępowania podczas konstruowania nowych metod ilościowych oraz jakościowych tak, aby w szybki i łatwy sposób uzyskać zadowalające końcowe wyniki.

Plan pracy doktorskiej przewidywał następujące zagadnienia:

- Optymalizacja warunków zbierania widm (czas analizy-liczb skanów, rozdzielczość)
- Opracowanie procedury tworzenia modeli ilościowych i jakościowych dla surowców i produktów kosmetycznych i farmaceutycznych
- Zbadanie wpływu czynników chemiczno-fizycznych na jakość samych widm (temperatura otoczenia, powtarzalność pomiarów)
- Określenie wpływu transformacji widm surowych na efekt końcowy opracowywanych metod jakościowych
- Ocena wpływu liczby istotnych czynników na zdolności predykcyjne opracowanych modeli

- Wykazanie możliwości stosowania spektroskopii podczerwieni (NIR oraz IR) w analizie ilościowej półproduktów i produktów kosmetycznych i farmaceutycznych.

IX ABSTRACT

In recent years there has been a significant increase in the quality control requirements for all industrial sectors. A growing number of studies that makes the traditional, time-consuming methods require use of expensive and harmful chemicals, making them unfavorable.

However, there is a very fast analytical method increasingly used in the analysis of qualitative and quantitative. It is the near-infrared spectroscopy (NIR). NIR spectroscopy is based on the absorption of electromagnetic radiation in the range 780-2500 nm. NIR absorption comes from nadton and combinational vibrations of bonds CH, NH, OH. The advantages of using NIR technology are: 1) non-destructive, non-invasive nature of the analysis, 2) no sample preparation before analysis, 3) simplicity and speed of execution routine analysis, 4) high sensitivity 5) low cost 6) obtain a comprehensive picture of the test material, 7) improving safety at work 8) limiting the use of solvents for sample preparation before analysis, which has a positive effect environmentally friendly. Besides the advantages of this technique also has some disadvantages. Raw spectra are quite difficult to interpret and require knowledge of advanced chemometric techniques.

The aim of my work was to develop procedures to be followed during the development of new methods of quantitative and qualitative, so that a quick and easy way to get a satisfactory final results

The plan the doctoral thesis include the following issues:

- Develop a library of spectra of classical and near-infrared
- Optimization of conditions for the collection of spectra (the time of analysis, the number of scans, resolution)
- Develop procedures for the creation of qualitative and quantitative models for raw materials and cosmetic and pharmaceutical products
- Examine the effects of chemical and physical factors on the quality of the same spectra (ambient temperature, repeatability of measurements)
- Determine the impact of the transformation of raw spectra of the final result of qualitative

X DOROBEK NAUKOWY

Publikacje:

1. Dopierała Anita, **Wrosz Patrycja**, Mazerski Jan „*Akrydyny jako związki aktywne przeciwnowotworowo*” Postępy Higieny i Medycyny Doświadczalnej, (online) 65,263-269, **2011**
2. **Wrosz Patrycja**, Dopierała Anita, Mazerski Jan „*Rola i dobór wstępnej obróbki surowych widm w analizie jakościowej i ilościowej z wykorzystaniem techniki spektroskopii bliskiej podczerwieni (NIR)*”3/2011

Rozdział w książce:

1. **Ogonowska- Wrosz Patrycja**, Jan Mazerski „*Zastosowanie techniki NIR w analizie jakościowej surowców kosmetycznych i farmaceutycznych*” Nauka i Przemysł metody spektroskopowe w praktyce nowe wyzwania i możliwości, Praca zbiorowa pod redakcją prof.dr hab. Zbigniewa Hubickiego, Lublin **2009**
2. **Wrosz Patrycja**, Dopierała Anita, Mazerski Jan, „*Porównanie spektroskopii IR oraz NIR w analizie tożsamościowej surowców kosmetycznych i farmaceutycznych*”, III. Ogólnopolskie Sympozjum „Nauka i przemysł – metody spektroskopowe w praktyce, nowe wyzwania i możliwości”, Lublin 8-10.06. **2010**
3. **Wrosz Patrycja**, Mazerski Jan, „*Theory and application of Near-infrared spectroscopy*” W “13th International Symposium of Students and Young Mechanical Engineers. Advances in Chemical and Mechanical Engineering”; Wydawnictwo Politechniki Gdańskiej, Gdańsk, 20-22.05.**2010**
4. **Wrosz Patrycja**, Dopierała Anita, Mazerski Jan „*Spektroskopia bliskiej podczerwieni (NIR)*” Biomediteh Badania i Innowacje, Gdańsk 9-13.05.**2011**
5. **Wrosz Patrycja**, Dopierała Anita, Mazerski Jan „*Wpływ różnych źródeł zmienności pojawiających się podczas zbierania widm w spektroskopii bliskiej podczerwieni*” Biomediteh Badania i Innowacje, Gdańsk 9-13.05.**2011**
6. Dopierała Anita, **Wrosz Patrycja**, Mazerski Jan „*Przeciwnowotworowe działanie akrydyn*” Biomediteh Badania i Innowacje, Gdańsk 9-13.05.**2011**
7. Dopierała Anita, **Wrosz Patrycja**, Mazerski Jan „*Termodynamika niekowalencyjnego oddziaływania małowcząsteczkowych ligandów z DNA*” Biomediteh Badania i Innowacje, Gdańsk 9-13.05.**2011**
8. **Wrosz Patrycja**, Dopierała Anita „*Determination of the active ingredient in pharmaceutical gel formulation by NIR spectroscopy*” W “14th International

Symposium of Students and Young Mechanical Engineers. Advances in Chemical and Mechanical Engineering”; Wydawnictwo Politechniki Gdańskiej, Gdańsk, 5-7.05.2011

9. Dopierała Anita, **Wrosz Patrycja**, Mazerski Jan „*Aggregation of imidazoacridinones spectroscopic and microcalorimetric approach*” W “14th International Symposium of Students and Young Mechanical Engineers. Advances in Chemical and Mechanical Engineering”; Wydawnictwo Politechniki Gdańskiej, Gdańsk, 5-7.05.2011

Doniesienia konferencyjne:

1. **Wrosz Patrycja**, Jan Mazerski „Zastosowanie techniki NIR w analizie jakościowej surowców kosmetycznych i farmaceutycznych” Lublin 2009
2. Dopierała Anita, **Wrosz Patrycja**, Mazerski Jan, „*Wpływ warunków reakcji na zjawisko agregacji barwników używanych w biochemii*”, VI Ogólnopolskie Studenckie Seminarium Naukowo-Technologiczne BIOMEDITECH, Poznań, 12-17.04.2010
3. **Wrosz Patrycja**, Dopierała Anita, Mazerski Jan, „*Przykład zastosowania spektroskopii NIR w analizie jakościowej surowców kosmetycznych i farmaceutycznych*”, VI Ogólnopolskie Studenckie Seminarium Naukowo-Technologiczne BIOMEDITECH, Poznań, 12-17.04.2010
4. **Wrosz Patrycja**, Dopierała Anita, Mazerski Jan, „*Porównanie spektroskopii IR oraz NIR w analizie tożsamościowej surowców kosmetycznych i farmaceutycznych*”, III. Ogólnopolskie Sympozjum „Nauka i przemysł – metody spektroskopowe w praktyce, nowe wyzwania i możliwości”, Lublin 8-10.06. 2010
5. Dopierała Anita, **Wrosz Patrycja**, Mazerski Jan. „*Nowe podejście chemometryczne do analizy zestawów widm UV/VIS mieszanin dwuskładnikowych*”, VIII Polska Konferencja Chemii Analitycznej, Kraków 4-9.07.2010
6. **Wrosz Patrycja**, Dopierała Anita, Mazerski Jan, „*Wpływ transformacji danych na jakość opracowanych metod tożsamościowych w spektroskopii bliskiej podczerwieni*”, VIII Polska Konferencja Chemii Analitycznej, Kraków 4-9.07.2010
7. Dopierała Anita, **Wrosz Patrycja**, Mazerski Jan “*Problem of aggregation in dye-DNA interaction, calorimetry studies*”, XLV Zjazd Polskiego Towarzystwa Biochemicznego, Wisła 20-23.09.2010
8. **Wrosz Patrycja**, Dopierała Anita, Mazerski Jan „*Wpływ różnych źródeł zmienności pojawiających się podczas zbierania widm w spektroskopii bliskiej podczerwieni*” Biomediteh Badania i Innowacje, Gdańsk 9-13.05.2011
9. Dopierała Anita, **Wrosz Patrycja**, Mazerski Jan „*Przeciwnowotworowe działanie akrydyn*” Biomediteh Badania i Innowacje, Gdańsk 9-13.05.2011

Wystąpienie ustne:

1. **Ogonowska Patrycja** „Zastosowanie techniki NIR w analizie jakościowej surowców i produktów kosmetycznych“, Sesja Sprawozdawcza Studium Doktoranckiego przy Wydziale Chemicznym Politechniki Gdańskiej, **Gdańsk 2008** (streszczenie w materiałach konferencyjnych s. 99)
2. **Wrosz Patrycja** „Zastosowanie techniki IR i NIR w analizie ilościowej surowców kosmetycznych“, Sesja Sprawozdawcza Studium Doktoranckiego przy Wydziale Chemicznym Politechniki Gdańskiej, **Gdańsk 2009** (streszczenie w materiałach konferencyjnych s.110,
3. **Wrosz Patrycja** „Near infrared pectroscopy” W “13th International Symposium of Students and Young Mechanical Engineers. Advances in Chemical and Mechanical Engineering”; Wydawnictwo Politechniki Gdańskiej, Gdańsk, 20-22.05.2010
4. **Wrosz Patrycja** „Determination of the active ingredient in pharmaceutical gel formulation by NIR spectroscopy” W “14th International Symposium of Students and Young Mechanical Engineers. Advances in Chemical and Mechanical Engineering” Gdańsk 5-7.05.2011
5. **Wrosz Patrycja** “Spektroskopia bliskiej podczerwieni (NIR)” Biomeditech Badania i Innowacje, Gdańsk 9-13.05.2011

Stypendia

Stypendium InnoDoktorant – stypendia dla doktorantów, III edycja”. Projekt współfinansowany przez Unię Europejską z Europejskiego Funduszu Społecznego (Program Operacyjny Kapitał Ludzki, Priorytet VIII, Działanie 8.2, Poddziałanie 8.2.2: „Regionalne Strategie Innowacji”)