

**CHEMOMETRYCZNE METODY MODELOWANIA
DANYCH. STUDIUM PORÓWNAWCZE**
CHEMOMETRIC METHODS OF DATA MODELLING.
A COMPARATIVE STUDY

Andrzej J. Kałka

*Uniwersytet Jagielloński, Wydział Chemii
ul. Gronostajowa 2, 30-387, Kraków
e-mail: andrzej.kalka@student.uj.edu.pl*

*Praca wyróżniona w konkursie Krakowskiego Oddziału PTChem na konferencji
studenckiej Horyzonty Nauki – Forum prac Dyplomowych 2017*

Abstract

Wprowadzenie

1. Podstawy metod modelowania danych
2. „Twarde” metody modelowania danych
3. „Miękkie” metody modelowania danych
4. „Twardo-miękkie” metody modelowania danych

Uwagi końcowe

Podziękowania

Piśmiennictwo cytowane

Andrzej J. Kałka, student chemii stopni magisterskiego na Wydziale Chemii Uniwersytetu Jagiellońskiego w Krakowie, pod opieką naukową Zespołu Badań Fotochemicznych i Luminescencyjnych. Zainteresowania: natura, wykorzystanie i podstawy teoretyczne zjawisk fotochemicznych, chemia teoretyczna i obliczeniowa.

ABSTRACT

The chemometric methods of data analysis allow to resolve complex multi-component systems by decomposing a measured signal into the contributions of pure substances. Mathematical procedure of such decomposition is called empirical data modelling. The main aim and subject of this article is to provide some basic information on the chemometric analysis.

The chemometric techniques are divided into three categories, resulting from the assumed premises. A base of hard type of modelling is an assumption, that the measured dataset can be a priori described by generally accepted physical or chemical laws, expressed by analytical forms of mathematical functions, however with unknown values of parameters [1]. Numerical values of those constants are optimised by using procedures such as the least squares curve fitting [1, 2]. When explicit form of equations are found, the whole system of data can be resolved. Therefore, the white types of data modelling are often used for kinetic measurements [3–8] and analysis of fluorescence quenching [9–13].

Completely different approach to data modelling is offered by so called soft chemometric methods [14–20]. Those techniques do not require any presumptions; solutions obtained for the considered system are thus far much more unconstrained. The black types of analysis make use not only of the least squares fitting procedures [18, 19], but also some other geometrical optimisation algorithms [16, 17]. The results of that approach suffer however from one main drawback: the final outcome is not unique – system is described by a set of feasible solutions. As a consequence, soft data modelling is generally applied to resolve empirical data, which cannot be easily expressed by an explicit form of a function. Such measurement techniques are for example chromatography or volumetry.

However, if some conjecture could be made about the measurement system and the obtained data, it is possible to stiffen the black methods by applying white constraints [20]. These types of the chemometric analysis are called grey or hard-soft, and are a practical combination of model-free optimisation with the limitations of feasible solutions, resulting from conformity with physical or chemical laws.

Due to the fact, that data modelling provides an opportunity for simultaneous identification of several components of the analysed mixture, the chemometric procedures, although not so popular yet, are extremely powerful research tools.

Keywords: chemometrics, hard data modelling, hard-soft data modelling, soft data modelling

Słowa kluczowe: chemometria, „twarde” metody modelowania danych, „twardo-miękkie” metody modelowania danych, „miękkie” metody modelowania danych

WPROWADZENIE

Od niepamiętnych czasów, naukowcy zgłębiający naturę otaczającego ich świata zmierzyć musieli się z niezwykle złożonością badanych układów. Duża liczba zmiennych i parametrów znacząco utrudnia, bądź wręcz uniemożliwia jednoznaczny opis obserwowanych zjawisk. Naturalną zatem wydaje się tendencja do upraszczania skomplikowanych układów poprzez eliminację lub marginalizację pewnych mniej istotnych czynników. Utworzone w ten sposób układy nazywa się modelowymi.

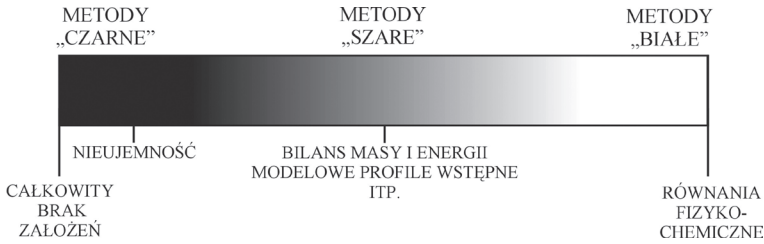
Wraz z postępującym udoskonaleniem aparatury pomiarowej możliwe stało się uwzględnienie i opis co raz większej ilości zmiennych mających wpływ na badane zjawiska i procesy. Nadal jednak problem stanowiły układy wieloskładnikowe, które w niektórych przypadkach należało sprowadzić do kilku układów jednoskładnikowych. Remedium w tym wypadku stanowiło zastosowanie rozwijających się równolegle procedur matematycznych. Tę swoistego rodzaju hybrydę empirycznych metod badawczych z teoretycznymi algorytmami obliczeniowymi nazwano chemometrią. Rozwiązywanie dużej ilości różnorodnych równań było jednak czasochłonne, dlatego pełny rozwój analizy chemometrycznej przyniosła popularyzacja urządzeń obliczeniowych i ostatecznie komputeryzacja. Modelowanie metodami chemometrycznymi tysięcy danych, otrzymanych najczęściej przez pomiary chromatograficzne i spektroskopowe, stało się pełnoprawną dziedziną chemii o ogromnych możliwościach, która na dzień dzisiejszy pozostaje jednak z niewiadomych przyczyn względnie mało znana.

1. PODSTAWY METOD MODELOWANIA DANYCH

1.1. PODZIAŁ METOD CHEMOMETRYCZNYCH

Główny podział metod chemometrycznych opiera się na ilości przyjętych założeń wstępnych, sprowadzających badany układ do układu modelowego. Wyróżnić zatem można dwa skrajne podejścia. Pierwszym jest narzucenie pełnej zgodności z modelem chemicznym bądź fizycznym, wyrażonym przez równania i zależności matematyczne – metody takie nazywane są „twardymi”, ze względu na małą dowolność przyjmowanych w czasie analizy możliwości, lub „białymi”, gdyż postać modelu je opisującego jest jawna. Przeciwnym tokiem myślenia jest całkowity brak wprowadzania jakichkolwiek uprzednich założeń – metody zwane są wówczas „miękkimi” bądź „czarnymi”, a cechuje je duża swoboda rozwiązań i „przyciemniony” obraz wzajemnych powiązań matematycznych. Trzecią, najbardziej obszerną grupę stanowią metody „szare”, zwane także „twardo-miękkimi”, które dopełniają obszar pomiędzy dwoma powyższymi granicznymi podejściami, łącząc w sobie rozwiązania zarówno „miękkie”, jak i „twarde”. Schematycznie, wszystkie trzy typy ana-

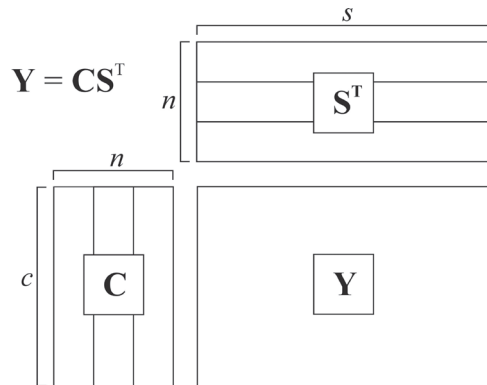
lizi chemometrycznej, zwanej także modelowaniem danych, przedstawiono poniżej (Rys. 1.1.).



Rysunek 1.1. Umowny schemat podziału metod chemometrycznych
Figure 1.1. The chemometric methods types – conventional scheme

1.2. NOTACJA MACIERZOWA I CEL MODELOWANIA

Analiza chemometryczna badanych układów zakłada, iż wszystkie dane opisać można za pomocą równania macierzowego [1] (Rys. 2.1):



Rysunek 1.2. Notacja macierzowa analizowanych danych
Figure 1.2. Matrix form of the analysed data

Każda z trzech macierzy ma swoją konkretną interpretację i wymiar, odpowiadający następującym parametrom układu: n określa liczbę składników, czyli faktorów układu, s – określa liczbę punktów pomiarowych pojedynczej próbki (np. długość/częstotliwość fali w przypadku spektroskopii), natomiast c jest liczbą próbek związanych ze zmianą stężenia składników, czasem pomiaru, ilością dodanej do układu substancji, etc. Wszystkie zarejestrowane wartości pomiarowe zbierane są w macierzy Y , zwanej macierzą danych – każdy jej wiersz odpowiada kolejnej próbce, natomiast kolumna – punktowi pomiarowemu. Transponowana macierz S^T zawiera profile widmowe czystych substancji, natomiast macierz stężeniowa C określa zmiany intensywności sygnałów pochodzących od pojedynczych składników

W przypadku niemal każdego modelowania, elementy macierzy Y zawsze pozostają znane, natomiast poszczególne wartości macierzy C i S^T pozostają niewiadomymi - celem analiz jest zatem znalezienie takich ich wartości, aby po wykonaniu działania:

$$R = Y - CS^T$$

elementy macierzy błędu R były jak najbliższe zeru, co jest tożsame z optymalnym dopasowaniem danych obliczonych i rzeczywistych.

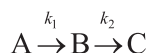
2. „TWARDE” METODY MODELOWANIA DANYCH

2.1. ALGORYTM OGÓLNY

Podstawą „białych” metod chemometrycznych jest założenie, iż empiryczne dane można opisać stosując model fizyko-chemiczny. Jest to pewne uproszczenie, które pozwala jednak na wstępne oszacowanie, a następnie optymalizację elementów macierzy C – poszczególne jej elementy buduje się w oparciu o znaną analityczną postać funkcyjną, której parametry pozostają jednak nieznanne. W celu ich wyznaczenia wprowadza się szacunkowe wartości do macierzy stężeniowej, na postawie których obliczana jest wstępna macierz profili S^T . Reprodukując macierz danych poprzez iloczyn CS^T , a następnie porównując ją z macierzą oryginalną Y , wyznaczona zostaje macierz błędu, która w „twardym” podejściu zależna jest tylko i wyłącznie od parametrów funkcji opisującej model. Metodami algorytmów optymalizacyjnych najmniejszych kwadratów [1, 2] można zatem wyznaczyć taką ich wartość, aby reprodukowane „białe” dane jak najlepiej oddawały rzeczywistą zmienność układu.

2.2. KINETYKA REAKCJI

Jednymi z najpowszechniejszych zastosowań „twardych” metod chemometrycznych są badania związane z kinetyką reakcji chemicznych [3–8]. Stężenia poszczególnych substratów, jak i produktów są bowiem związane zależnością funkcyjną z czasem pomiaru. Wykorzystując zatem techniki spektroskopii, najczęściej absorpcyjnej UV-Vis lub IR, śledzić można zmiany intensywności sygnału w czasie, poprzez rejestrację widm mieszanin, zebranych następnie w macierz Y . Macierz C buduje się na podstawie równań opisujących zależności stężeń poszczególnych składników przy określonym mechanizmie i rzędzie reakcji. Podając przykład pierwszorzędowej reakcji następczej [1]:



poszczególne elementy „białej” macierzy stężeniowej będą równe:

$$C(t, A) = [A_0] \cdot \exp(-k_1 \cdot t)$$

$$C(t, B) = [A_0] \cdot \frac{k_1}{k_2 - k_1} [\exp(-k_1 \cdot t) - \exp(-k_2 \cdot t)]$$

$$C(t, C) = [A_0] - C(t, A) - C(t, B)$$

przy czym t – czas, $[A_0]$ – stężenie początkowe substancji A, k_1 i k_2 – stałe szybkości reakcji. Wprowadzając zatem szacunkowe wartości stałych szybkości, możliwa jest taka ich optymalizacja, aby jednocześnie otrzymać wszystkie wartości maksymalnie zbliżone do spodziewanych. W wyniku działania algorytmu otrzymuje się też profile spektralne pojedynczych składników. W przypadku znanych widm substratu i produktu, zweryfikować zatem można, który z zakładanych modeli kinetycznych jest zgodny z rzeczywistością. Metoda pozwala też na wyznaczanie przebiegu profili widmowych produktów przejściowych, co jest niejednokrotnie niemożliwe do zrealizowania klasycznymi technikami pomiarowymi.

2.2. WIDMA WYGASZANIA FLUORESCENCJI

Metody „twarde” wykorzystuje się także w przypadku analizy widm emisyjnych, gdyż intensywność fluorescencji bądź fosforescencji jest związana z ilością dodanego do układu w zakresie niskich stężeń wygaszacza, przez liniową zależność Sterna-Volmera:

$$\frac{F_0}{F} = 1 + K_{SV} \cdot Q$$

gdzie F_0 – intensywność emisji w nieobecności wygaszacza, F – intensywność emisji przy stężeniu wygaszacza równym Q , K_{SV} – stała Sterna-Volmera. W przypadku mieszanin kilku fluoroforów, całkowita intensywność emisji jest sumą natężeń fluorescencji każdego ze składników. Kolumny macierzy stężeniowej, odpowiadające i -temu składnikowi, buduje się zatem w następujący sposób:

W wyniku modelowania otrzymuje się zatem rozdzielone widma emisji pojedynczych składników w postaci macierzy S^T [9].

Warto w tym miejscu także wspomnieć o historycznie uzasadnionych punktowych metodach analizy danych, operujących na jednej, bądź kilku wybranych liniach emisji, a nie na pełnym widmie fluorescencji. Zastosowane po raz pierwszy dla układów dwuskładnikowych [10, 11], umożliwiały określenie stosunku frakcji ulegających

w różnej mierze wygaszeniu, poprzez wprowadzenie pojęcia ułamka intensywnościowego i -tego składnika f_i :

$$C(t, i) = \frac{1}{1 + K_{SV}^i \cdot Q}$$

gdzie F_0^i – natężenie punktowej (przy jednej linii emisji) niewygaszanej fluorescencji i -tego składnika, F_0 – natężenie punktowej fluorescencji mieszaniny fluoroforów w nieobecności wygaszacza. Stosując zatem powyższą notację, wyprowadzić można następujące modyfikacje równania Sterna-Volmera:

$$\frac{F}{F_0} = \sum_i^n \frac{f_i}{1 + K_{SV}^i \cdot Q} \quad \frac{F_0 - F}{F_0} = \sum_i^n \frac{f_i \cdot K_{SV}^i \cdot Q}{1 + K_{SV}^i \cdot Q}$$

które poprzez zastosowanie algorytmów optymalizacyjnych dla analizowanych danych, pozwalają na wyznaczenie wkładu danego składnika do emisji całej próbki. Zastosowanie metod punktowych zostało wykorzystane między innymi do badań układów białkowych i poszczególnych grup fluoroforów w nich zawartych [12, 13]. Traktując poszczególne ugrupowania jako pojedyncze składniki, poprzez obserwację stałych Sterna-Volmera im odpowiadającym, określić można, do których części złożonych związków wygaszacza ma większy „dostęp”, a które fragmenty poprzez zawadę steryczną wygaszane są w mniejszym stopniu.

Zastosowanie jednakże metod punktowych pozwala raczej na analizę jakościową niż ilościową zachodzących w układzie procesów, gdyż wyznaczone optymalne wartości stałych wygaszania Sterna-Volmera wykazują znaczny rozrzut w zależności od badanej linii emisji.

2.3. „TWARDE” METODY A INNE TECHNIKI POMIAROWE

Niezbędnym i teoretycznie jedynym warunkiem możliwości wykorzystania „białych” metod chemometrycznych są jawne zależności funkcyjne opisujące badany układ. Modelowanie „twarde” może nie przynieść spodziewanych rezultatów, jeżeli takowe nie istnieją lub mają bardzo zawiłą i złożoną postać, chociaż w drugim przypadku można próbować zastosować pewne uproszczenia. Dobrym przykładem jest analiza danych chromatograficznych – poszczególne chromatogramy mogą być aproksymowane za pomocą obwiedni gaussowskich [1], których parametry są optymalizowane z wykorzystaniem metody najmniejszych kwadratów. W przypadku zastosowania modeli przybliżonych należy jednak zachować szczególną ostrożność interpretując otrzymane rezultaty rozdzielania chemometrycznego.

3. „MIĘKKIE” METODY MODELOWANIA DANYCH

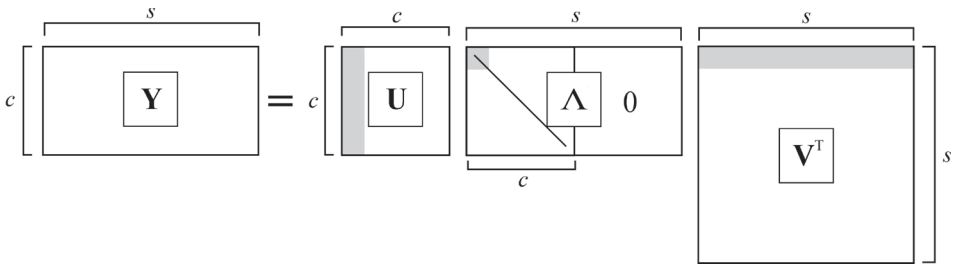
3.1. ALGORYTM OGÓLNY

„Miękkie” metody chemometryczne wykorzystują model „czarnej” skrzynki, o której zawartości nic a priori nie wiadomo – w podejściu tym analizowanych danych nie wiążą zatem w ogólności żadne analityczne zależności funkcyjne, z wyjątkiem dopuszczonego przez konwencję więzu nieujemności. Postać zarówno macierzy stężeniowej, jak i macierzy profili widmowych składników, pozostaje nieznana – aby podjąć jakąkolwiek procedurę optymalizacyjną, należy jednak wyznaczyć najpierw szacunkową postać jednej z nich.

W tym celu wykorzystać można algorytm rozkładu macierzy według wartości szczególnych SVD (ang. *Singular Value Decomposition*) [1]. Macierz danych \mathbf{Y} przedstawia się w postaci iloczynu trzech macierzy:

$$\mathbf{Y} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T = \mathbf{C}\mathbf{S}^T$$

Graficznie, operację tę przedstawiono na Rysunku 3.1. Macierz $\mathbf{\Lambda}$, jest macierzą diagonalną o wymiarze $c \times c$, zawierającą w malejącej kolejności wartości szczególne macierzy kowariancji $\mathbf{Y}\mathbf{Y}^T$; ze względów jednak na specyfikę działań na macierzach, „dopełnieniem” $\mathbf{\Lambda}$ do rozmiaru $c \times s$ jest macierz zerowa. Macierze \mathbf{U} i \mathbf{V}^T (lub \mathbf{V} w zależności od notacji) są macierzami ortonormalnych wektorów, będących wektorami własnymi odpowiednio macierzy $\mathbf{Y}\mathbf{Y}^T$ i $\mathbf{Y}^T\mathbf{Y}$.



Rysunek 3.1 Macierzowe przedstawienie rozkładu według wartości szczególnych (SVD); kolorem szarym zaznaczono wektory odpowiadające istotnym faktorom układu

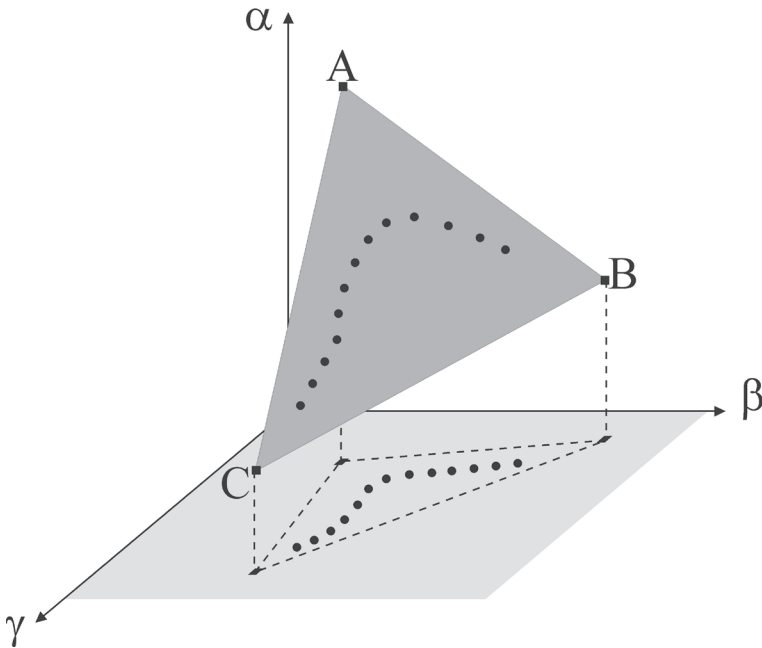
Figure 3.1. Singular Value Decomposition (SVD) in matrix notation; vectors referring to principal factor(s) are marked in gray

Znając ilość składników – istotnych faktorów f – w mieszaninie, możliwe jest tzw. „odszumienie”, czyli redukcja niedoskonałości danych pomiarowych. Poprzez rozkład macierzy \mathbf{Y} przy pomocy algorytmu SVD, a następnie jej reprodukcję przy użyciu ograniczonej liczby f wektorów macierzy \mathbf{U} , $\mathbf{\Lambda}$, i \mathbf{V}^T (szare obszary na rysunku 3.1), eliminuje się wkład innych czynników wpływających na wariancję układu, często przejawiających się właśnie jako „szum”, zawarty w macierzy „błędów” \mathbf{R} .

Głównym celem chemometrycznego rozdzielania danych jest jednak odtworzenie przebiegu macierzy stężeniowej i profili widmowych czystych składników. Korzystając z ortogonalnych własności macierzy \bar{V}^T :

$$Y = \bar{U}\bar{A}\bar{V}^T + R$$

Powyższe wyrażenie prowadzi do następujących wniosków: kolejne wartości danych pomiarowych można zapisać w układzie utworzonym przez wektory zredukowanej macierzy \bar{V}^T , o współrzędnych równych wierszom macierzy $\bar{U}\bar{A}$. Ponadto, wszystkie wektory pomiarowe powinny zawierać się w zależności od liczby czynników w płaszczyźnie lub hiperpłaszczyźnie utworzonej i ograniczonej przez wektory własne czystych składników (ilustracja – Rys. 3.2).



Rysunek 3.2. Graficzna interpretacja punktów pomiarowych (kropki) trójskładnikowego układu (A, B, C) w przestrzeni rozpiętej na wektorach własnych pochodzących z rozkładu SVD (α, β, γ)
 Figure 3.2. Graphical depiction of the measured three-component (A, B, C) data (dots), represented in a SVD eigenvector space (α, β, γ)

Graficzna interpretacja wyników wydaje się jednak być mało praktyczną – zdecydowanie bardziej pożądane byłyby jawne postaci macierzy stężeniowej i macierzy widm pojedynczych składników. Przejście pomiędzy układem abstrakcyjnym $\bar{U}\bar{A}\bar{V}^T$ a przewidywanym układem rzeczywistym $\hat{C}\hat{S}^T$ odbywa się przez kwadratową macierz rotacji (transformacji) T o wymiarach $f \times f$ [17, 19]:

$$Y \approx \bar{U}\bar{A}\bar{V}^T = \bar{U}T T^{-1}\bar{A}\bar{V}^T = \hat{C}\hat{S}^T$$

$$\hat{\mathbf{C}} = \bar{\mathbf{U}}\mathbf{T}$$

$$\hat{\mathbf{S}}^T = \mathbf{T}^{-1}\bar{\mathbf{A}}\bar{\mathbf{V}}^T$$

Znalezienie poszczególnych jej elementów jest zatem kluczowym celem „miękkiej” analizy chemometrycznej.

3.2. WYBRANE „MIĘKKIE” ALGORYTMY ANALIZY

Kamieniem milowym „czarnych” metod chemometrycznych było opracowanie algorytmu Lawtona-Sylvestra [14], pozwalającego na rozdział dwuskładnikowych układów spektrofotometrycznych. Jego trójskładnikowym uogólnieniem jest metoda zaproponowana pierwotnie przez Borgena i Kowalskiego [15], a udoskonalona przez Rajkó [16]. W tym podejściu dane pomiarowe leżą na płaszczyźnie ograniczonej trójkątem, którego wierzchołki odpowiadają czystym składnikom (analogicznie do Rys. 3.2). Ich położenia pozwalają wyznaczyć poszczególne elementy macierzy \mathbf{T} [17]. Normując jeden z wektorów macierzy transformacji, można sprowadzić badany układ do współrzędnych dwuwymiarowych, co znacząco ułatwia wyznaczenie poszczególnych wartości macierzy rotacji.

Nieco innym podejściem cechuje się metoda sympleksów [17]. Wprowadzając normalizację jednego z wektorów macierzy rotacji \mathbf{T} , otrzymuje się dla układów trójskładnikowych:

$$\mathbf{T} = \begin{vmatrix} 1 & t_{12} & t_{13} \\ 1 & t_{22} & t_{23} \\ 1 & t_{32} & t_{33} \end{vmatrix} = \begin{vmatrix} 1 & \mathbf{t} \\ \mathbf{1} & \mathbf{M} \end{vmatrix} \quad \mathbf{1} = \begin{vmatrix} 1 \\ 1 \end{vmatrix} \quad \mathbf{t} = \begin{vmatrix} t_{12} & t_{13} \end{vmatrix} \quad \mathbf{M} = \begin{vmatrix} t_{22} & t_{23} \\ t_{32} & t_{33} \end{vmatrix}$$

Dla każdej wprowadzonej pary liczb \mathbf{t} optymalizuje się wartości macierzy \mathbf{M} metodą sympleksów geometrycznych. Następnie, znając wszystkie elementy macierzy \mathbf{T} , odtwarza się macierz stężeniową i macierz profili widmowych pojedynczych składników, nakładając przy tym więzy nieujemności. Reprodukując macierz danych i porównując z oryginalną macierzą \mathbf{Y} , znaleźć można rozwiązania wykazujące największą zgodność z wartościami spodziewanymi. Optymalizacja macierzy \mathbf{M} może odbyć się także poprzez algorytmy geometryczne [16, 17], dopasowujące obwiednię badanego obszaru w układzie wektorowym wyznaczoną przez czyste składniki.

Ostatecznie należy tutaj wspomnieć o leżącej na pograniczu metod „miękkich” metodzie MCR-ALS (ang. *Multivariate Curve Resolution – Alternating Least Squares*) [18, 19]. Wykorzystuje ona macierze stężeniowe oraz widmowe pojedynczych składników, jednak w ogólności nie opisane żadną zależnością funkcyjną. Poszczególne elementy \mathbf{C} oraz \mathbf{S}^T przedstawić zatem można przy pomocy pewnych krzywych o niekoniecznie znanej formule matematycznej. Algorytm tej metody oparty jest na naprzemiennej optymalizacji metodą najmniejszych kwadratów przebiegu krzy-

wych stężeniowych oraz profili widmowych tak, aby otrzymana przy ich użyciu macierz reprodukowanych danych wykazywała największą zgodność z oryginalną macierzą Y . Metoda wymaga jednak podania wstępnych krzywych C – można w tym wypadku posłużyć się procedurami szeroko pojętej analizy faktorowej [1] lub zastosować wstępny model „biały”, co jednak kwalifikuje metodę do „twardszej” kategorii.

„Czarne” metody chemometryczne szczególnie dobrze nadają się do analizy układów, w których nie można wprost zastosować jawnych funkcji analitycznych. Dobrym przykładem mogą być tutaj krzywe miareczkowania oraz chromatogramy. Należy jednak brać pod uwagę to, iż otrzymane rezultaty mogą być niekiedy wieloznaczne lub stanowić pewien zbiór odpowiedzi, zamiast jednego konkretnego rozwiązania.

4. „TWARDO-MIĘKKIE” METODY MODELOWANIA DANYCH

Kategoria „szarych” metod chemometrycznych jest zdecydowanie najobszerniejsza – zawiera bowiem połączenia algorytmów „białych” i „czarnych” w dowolnej niemalże konfiguracji, co stwarza możliwości wprowadzania niezliczonych ilości wersji i modyfikacji algorytmów. W rozdziale tym Autor zdecydował się posłużyć przykładem dwóch często stosowanych procedur „twardo-miękkich”.

Pierwszą z nich jest wspomniana już metoda MCR-ALS [18, 19]. Narzucając odpowiednie więzy na wstępną macierz stężeniową, budowaną zgodnie z jawną analityczną formułą, oczekiwać można optymalizacji w kierunku zgodnym z przyjętą zależnością. Rozdzielenie chemometryczne odbywa się więc w kolejności: model „biały”, dopasowanie „czarne”. Oczywiście, każdą z kolejnych otrzymywanych na drodze optymalizacji macierzy C można sprowadzić do spodziewanej postaci funkcyjnej, jednakże analiza przesuwana się wówczas znacznie w stronę metod „białych”.

Drugim podejściem [20] może być odwrócenie kolejności stosowanych algorytmów. Za pomocą „miękkiego” rozdziału i rozkładu SVD wyznacza się poszczególne elementy abstrakcyjnych macierzy stężeniowych i profili widmowych. Poprzez macierz rotacji oblicza się rzeczywiste wartości wyżej wymienionych przewidywanych macierzy \hat{C} i \hat{S}^T , a następnie metodą „białą” odtwarza się przebieg pierwszej z nich zgodnie z przyjętym równaniem fizyko-chemicznym. Na podstawie tak zmodyfikowanej macierzy \hat{C} reprodukuje się macierz danych widmowych \hat{Y} i porównuje z oryginalnymi punktami pomiarowymi Y .

Jak zauważono na wstępie, ilość możliwych rozwiązań, mających na celu usprawnienie optymalizacji połączonymi metodami „twardo-miękkimi” jest znaczna, a każde kolejne powstaje w wyniku narzucenia dowolnych więzów na badany układ „czarnej” skrzynki.

UWAGI KOŃCOWE

Korzystając z „twardych” metod, opartych na względnie prostych algorytmach optymalizacyjnych, możliwy jest rozdział danych pomiarowych na składowe pochodzące od pojedynczych składników. Bazujące na jawnych zależnościach fizyko-chemicznych „białe” procedury świetnie nadają się do określania zgodności z teoretycznymi założeniami przyjętymi dla rzeczywistych układów, szczególnie w przypadku rozważań kinetycznych i procesów luminescencyjnych. Analizując otrzymane na drodze rozdziału chemometrycznego widma spektroskopowe, możliwa staje się jakościowa lub nawet ilościowa identyfikacja poszczególnych substancji, co często nie jest możliwe w trakcie interpretacji „suchych” wyników pomiarowych.

W przypadku procesów, których natury nie sposób opisać analitycznymi zależnościami funkcyjnymi, rozsądnym rozwiązaniem wydaje się zastosowanie „czarnych” metod modelowania, nie wymagających żadnych wstępnych informacji odnoszących się do badanego układu. Jeżeli jednak zasadne jest przypuszczenie, iż dany zbiór danych pomiarowych ograniczony jest przez pewne więzy, najlepszym rozwiązaniem jest użycie algorytmów „szarych”, łączących w sobie swobodę „miękkich” metod z ich usztywnieniem przez „twarde” elementy procedur „białych”.

Możliwości oferowane przez analizę chemometryczną wydają się ogromne: od „odszumiania” uzyskanych danych pomiarowych po opis skomplikowanych układów wieloskładnikowych. Autor wyraża zatem szczerą nadzieję zainteresowania czytelników metodami modelowania danych oraz zastosowania tychże procedur w prowadzonych przez nich badaniach.

PODZIĘKOWANIA

Składam serdeczne podziękowania Kierownikowi Zespołu Badań Fotochemicznych i Luminescencyjnych Wydziału Chemii UJ, p. dr hab. Andrzejowi M. Turkowi za udzielone wsparcie merytoryczne.

PIŚMIENICTWO CYTOWANE

- [1] M. Maeder, Y.M. Neuhold, *Practical Data Analysis in Chemistry*, Elsevier, Amsterdam 2007.
- [2] M. Maeder, A.D. Zuberbühler, *Anal. Chem.*, 1990, **62**, 2220.
- [3] K.J. Molloy, M. Maeder, M.M. Schumacher, *Chemom. Intell. Lab. Syst.*, 1999, **46**, 221.
- [4] M. Maeder, Y.M. Neuhold, G. Puxty, P. Gemperline, *Chemom. Intell. Lab. Syst.*, 2006, **82**, 75.
- [5] G. Puxty, M. Maeder, K. Hungerbühler, *Chemom. Intell. Lab. Syst.*, 2006, **81**, 149.
- [6] G. Puxty, Y.M. Neuhold, M. Jecklin, M. Ehly, P. Gemperline, A. Nordon, D. Littlejohn, J.K. Basford, M. De Cecco, K. Hungerbühler, *Chem. Engin. Scien.*, 2008, **63**, 4800.
- [7] J. Billeter, Y.M. Neuhold, L. Simon, G. Puxty, K. Hungerbühler, *Chemom. Intell. Lab. Syst.*, 2008, **93**, 120.

-
- [8] M. Hasani, M. Shariati-Rad, H. Abdollahi, *Anal. Chim. Acta*, 2009, **636**, 175.
- [9] Ł.J. Witek, A.M. Turek, *Chemom. Intell. Lab. Syst.*, 2017, **160**, 77.
- [10] S.S. Lehrer, *Biochem.*, 1971, **10**(17), 3253.
- [11] A.U. Acuña, F.J. Lopez-Hernandez, J.M. Oton, *Biophys. Chem.*, 1982, **16**, 253.
- [12] W. Stryjewski, Z. Wasylewski, *Eur. J. Biochem.*, 1986, **158**, 547.
- [13] S. Matsumoto, E. Nishimoto, H. Soejima, S. Yamashita, *Biosci. Biotechnol. Biochem.*, 2010, **74** (7), 1396.
- [14] W.H. Lawton, E.A. Sylvestre, *Technometrics*, 1971, **13**, 617.
- [15] O.S. Borgen, B.R. Kowalski, *Anal. Chim. Acta*, 1985, **174**, 1.
- [16] R. Rajkó, *J. Chemom.*, 2009, **23**, 265.
- [17] A. Golshan, H. Abdollahi, S. Beyramysoltan, M. Maeder, K. Neymeyr, R. Rajkó, M. Sawall, R. Tauler, *Anal. Chim. Acta*, 2016, **911**, 1.
- [18] J. Diewok, A. Anna de Juan, M. Maeder, R. Tauler, B. Lendl, *Anal. Chem.*, 2003, **75**, 641.
- [19] J. Jaumot, A. de Juan, R. Tauler, *Chemom. Intell. Lab. Syst.*, 2015, **140**, 1.
- [20] A. Golshan, H. Abdollahi, M. Maeder, *Anal. Chim. Acta*, 2012, **709**, 32.

Praca wpłynęła do Redakcji 26 maja 2018