

Witold ILEWICZ
Instytut Automatyki, Zakład Systemów Pomiarowych
Politechnika Śląska

ALGORYTM ESTYMACJI PARAMETRÓW ZŁOŻONYCH SYGNAŁÓW CHROMATOGRAFICZNYCH Z ZASTOSOWANIEM WZORCA ZDEFINIOWANEGO NUMERYCZNIE

Streszczenie. W wyniku analizy chromatograficznej uzyskuje się sygnał interpretowany jako suma pików, z których każdy odpowiada jednemu ze składników analizowanej mieszaniny. Na podstawie parametrów pików, takich jak położenie pików, wysokość pików czy też powierzchnia pików wyznacza się skład ilościowy i jakościowy analizowanej mieszaniny. Często jednakże występuje sytuacja, gdy piki składowe sygnału nachodzą na siebie i niemożliwa jest bezpośrednia analiza. Mówi się wtedy, że sygnał jest złożony. Jedną z metod postępowania w takim przypadku jest zastosowanie regresji nieliniowej, przy czym zadanie estymacji rozwiązuje się z zastosowaniem metody najmniejszych kwadratów. Metoda wymaga określenia parametrycznego modelu matematycznego pojedynczego pików. W artykule przedstawiono sposób budowy modelu pików w oparciu o zapis cyfrowy rzeczywistego sygnału chromatograficznego w postaci pojedynczego pików. Zapis taki nazywa się wzorcem numerycznym. Zaprezentowano algorytm, w którym wykorzystanie wzorca numerycznego prowadzi do poprawy dokładności modelu pików PMG (Polynomially Modified Gaussian) w zastosowaniu do analizy złożonego sygnału chromatograficznego metodą regresji.

AN ALGORITHM WITH NUMERICAL SHAPE STANDARD FOR ESTIMATION PARAMETERS OF COMPLEX CHROMATOGRAPHIC SIGNALS

Summary. Chromatographic signals are most often interpreted as a sum of the peaks. Each peak corresponds to one component of the analyzed mixture. Sometimes, however, peaks are more or less overlapping each other due to the bad chromatographic separation. Then there is no way to estimate parameters of the peaks directly from the signal. We can define such signals as complex. The complex signal can be examined by nonlinear regression analysis. In this method there is assumption about knowledge of the mathematical parametric model of the single peak. There are described many models of chromatographic peaks in the literature and proper choice of the model is the most important factor that determine the accuracy of analysis of complex chromatographic signals.

In this paper there is presented a method of construction of peak model on the basis of the real signal of single peak obtained during a chromatographic experiment. Such a signal is named as numerical shape standard. An application of numerical shape standard to improve

accuracy of PMG (Polynomially Modified Gaussian) peak model used for purpose of analysis of complex chromatographic signals by means of the nonlinear regression is shown.

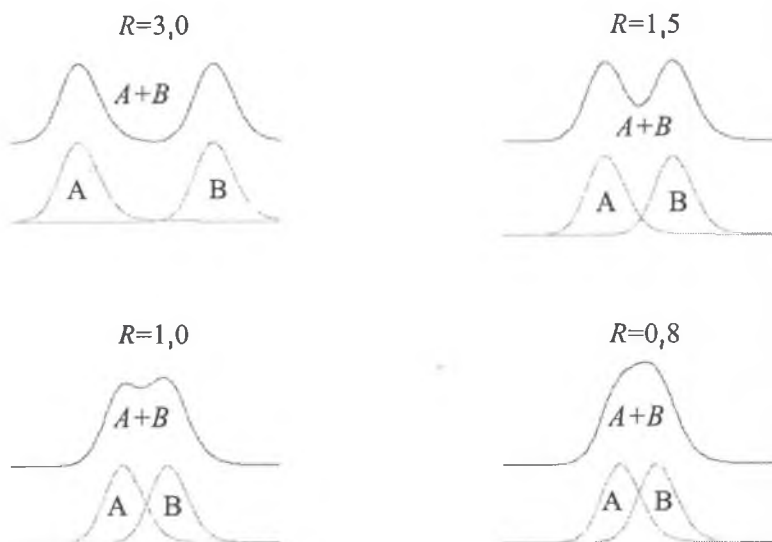
1. WSTĘP

W technikach analitycznych, takich jak chromatografia i spektroskopia uzyskiwany sygnał jest sumą pików składowych. W idealnym przypadku wszystkie piki w sygnale są rozdzielone i analiza takiego sygnału nie stanowi problemu. Na podstawie położenia pików składowych identyfikuje się substancje, a na podstawie powierzchni pików wyznacza się skład ilościowy analizowanej mieszaniny.

Niska rozdzielczość aparatury pomiarowej, której przyczyny mogą być różne, doprowadza do uzyskiwania sygnałów złożonych, to znaczy takich, w których piki składowe nakładają się na siebie. Złożoność sygnału można także dopuścić w celu uzyskania innych korzyści. Na przykład, jeżeli wynik analizy chromatograficznej stanowi informację wykorzystywaną do sterowania procesem, to można tak dobrać parametry pracy chromatografu, żeby znacząco skrócić czas analizy (który stanowi czas martwy w pętli sprzężenia zwrotnego) kosztem dopuszczenia nałożenia pików składowych. W literaturze podaje się, że w ten sposób można uzyskać skrócenie czasu analizy, nawet rzędu kilkudziesięciu razy. W takim przypadku należy jednakże dysponować wiarygodną metodą estymacji parametrów pików w złożonym sygnale. Często stosowaną miarą złożoności jest rozdzielczość R definiowana jako:

$$R = 2 \frac{t_2 - t_1}{w_1 + w_2}, \quad (1)$$

czyli różnica położenia maksimów dwóch sąsiadujących pików odniesiona do średniej szerokości pików w połowie wysokości. Przykładowe sygnały oraz ich piki składowe dla różnych wartości R przedstawiono na rys. 1.



Rys. 1. Sygnały złożone z dwóch pików dla różnych stopni rozdzielienia R
 Fig. 1. Examples of complex signals for different levels of separation R

Podstawową metodą analizy takich sygnałów jest regresja nieliniowa [12,14,3]. Aby stosować metodę, należy znać postać modelu matematycznego pojedynczego pików $f(\mathbf{b},x)$ i na jego podstawie formuluje się model obserwowanego sygnału:

$$F(\mathbf{b},x) = \sum_{k=1}^q f(\mathbf{b}_k, x) + \eta, \quad (2)$$

gdzie: $F(\mathbf{b},x)$ – model obserwowanego sygnału, $f(\mathbf{b}_k, x)$ – funkcja opisująca pojedynczy pik w sygnale, q – liczba pików składowych w sygnale, $\mathbf{b}=[\mathbf{b}_1, \dots, \mathbf{b}_q]^T$ – wektor parametrów modelu, \mathbf{b}_k – wektor parametrów k -tego pików składowego, x – zmienna stanowiąca dziedzinę obserwowanego sygnału (np. w chromatografii x oznacza czas, w spektroskopii x może oznaczać długość fali lub też energię), η – zakłócenia losowe.

Zadanie regresji rozwiązuje się znajdując minimum sumy kwadratów różnic pomiędzy analizowanym sygnałem $y(x_i)$ zarejestrowanym eksperymentalnie a modelem matematycznym sygnału, przy czym minimalizacja odbywa się po wartościach parametrów modelu matematycznego sygnału. Dla minimalizowanego wskaźnika postaci

$$J(\mathbf{b}) = \sum_{i=1}^M (y_i - F(\mathbf{b}, i))^2 \quad (3)$$

poszukuje się takiego wektora parametrów $\hat{\mathbf{b}}$, że zachodzi

$$J(\hat{\mathbf{b}}) = \min_{\mathbf{b}} (J(\mathbf{b})). \quad (4)$$

Na podstawie wektora parametrów $\hat{\mathbf{b}}=[\hat{\mathbf{b}}_1, \hat{\mathbf{b}}_2, \dots, \hat{\mathbf{b}}_q]^T$ minimalizującego wskaźnik J oraz modelu pojedynczego pików odtwarza się kształty pików składowych i wyznacza się ich powierzchnię i położenie.

W świetle studiów literaturowych jednym z najważniejszych czynników warunkujących dokładność analizy złożonego sygnału analitycznego metodą regresji jest dokładność modelu matematycznego pojedynczego pików, czyli zgodność kształtu modelu z kształtem rzeczywistych pików, stąd trwają ciągłe prace nad poszukiwaniem coraz dokładniejszych modeli matematycznych pojedynczych pików [15,11,8,9,10], a liczba modeli matematycznych pików chromatograficznych praktycznie stosowanych wynosi ponad 100 [2]. Dokładność kształtu modelu pojedynczego pików ma znaczenie zwłaszcza dla wyników analizy ilościowej w warunkach dużego nałożenia pików składowych – nawet niewielkie błędy systematyczne kształtu modelu mogą prowadzić do całkowicie fałszywych wyników analizy ilościowej. Na przykład, błędy wyznaczenia powierzchni pików składowych sygnału mogą sięgać kilkudziesięciu procent [5].

2. MODELE MATEMATYCZNE PIKÓW W OPARCIU O WZORZEC KSZTAŁTU ZDEFINIOWANY NUMERYCZNIE

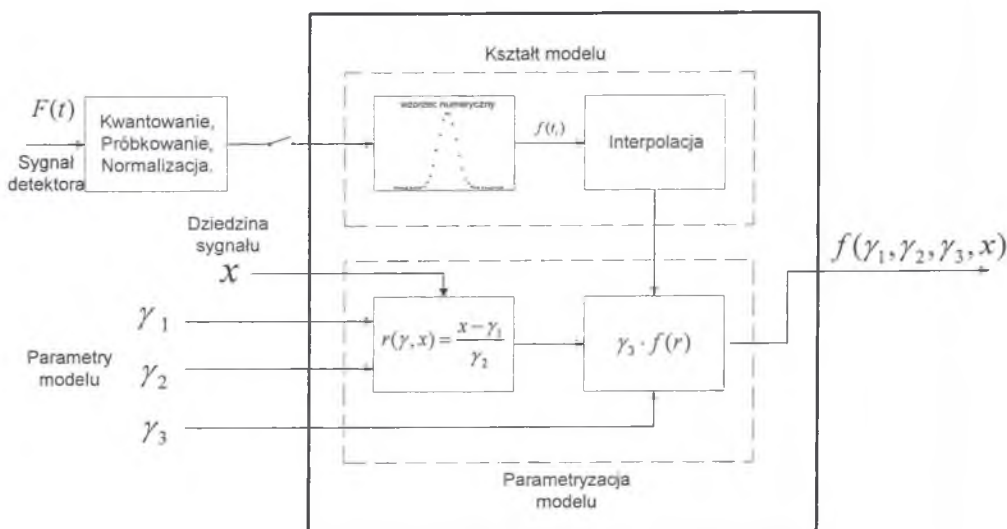
Keller i Lusebrink [6], Anderson i inni [1], Westerberg [16] pierwsi przedstawili metodę modelowania złożonych sygnałów analitycznych w postaci sumy pików składowych w oparciu o wzorzec numeryczny. Metodę stosował także Ilewicz [4] do analizy mieszanin węglowodorów z zastosowaniem chromatografii gazowej. Metoda polega na budowie modelu pojedynczego pików na podstawie cyfrowego zapisu kształtu dla pojedynczego pików uzyskanego w eksperymencie. Pik taki może być generowany w osobnym eksperymencie analitycznym bądź też można wykorzystać pojedynczy pik wprost z analizowanego sygnału. Taki

pojedynczy pik nazywa się wzorcem numerycznym. Keller i Lusebrink [6] dla celów analizy sygnałów NMR stosowali pojedynczy pik brany wprost z analizowanego sygnału. Anderson i współpracownicy stosowali jako piki wzorcowe zarówno piki pojedyncze z analizowanego sygnału, jak i piki wygenerowane w specjalnie przeprowadzonym eksperymencie, polegającym na chromatografowaniu czystej substancji. Podstawowym założeniem metody wzorca numerycznego z [1] jest niezmienność kształtu pojedynczego piku. Prowadzi to do prostego modelu piku, mianowicie do opisu dowolnego piku w sygnale wystarczają 3 parametry:

$$y_k = h_k \cdot f\left\{\frac{(t-t_k)}{w_k}\right\}, \quad (5)$$

gdzie $[h_k \ t_k \ w_k]^T$ stanowią wektor parametrów tak skonstruowanego modelu piku, a jego elementy to odpowiednio wysokość piku, położenie piku oraz szerokość piku. f oznacza funkcję opisującą kształt piku wzorcowego. Ponieważ wartości tej funkcji znane są tylko w chwilach próbkowania, obliczanie wartości modelu wymaga zastosowania odpowiedniej metody interpolacji bądź aproksymacji.

Strukturę modelu matematycznego piku z kształtem piku zadany numerycznie przedstawiono na rys. 2. Na schemacie tym uwzględniono wszystkie istotne czynniki, mające wpływ na jakość modelu matematycznego piku, a więc czynności podejmowane po wygenerowaniu sygnału wzorca: próbkowanie, kwantowanie, wygładzanie, normalizacja wzorca. Przełącznik umieszczony na strzałce łączącej zapis wzorca symbolizuje możliwość uaktualniania wzorca, na przykład jeżeli zmienia się własności aparatury analitycznej, wpływając na kształt pików składowych w sygnale.

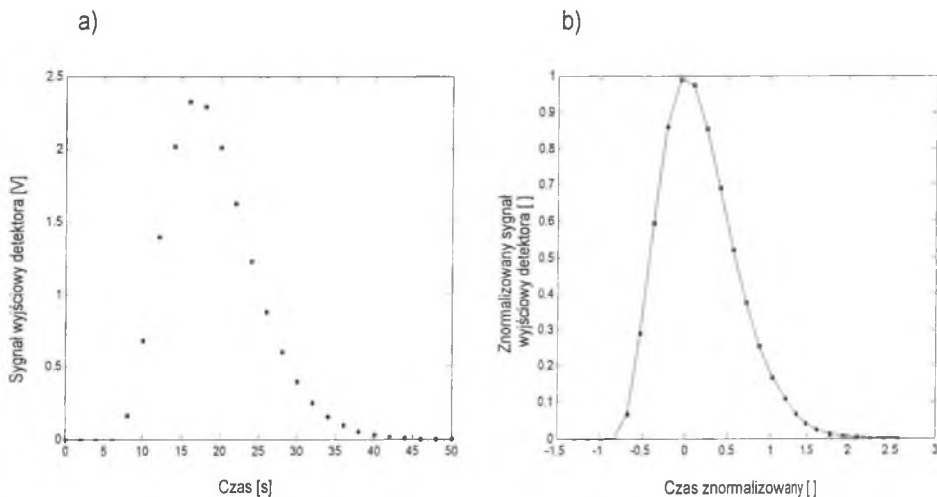


Rys. 2. Procedura wyznaczania wartości modelu matematycznego piku w oparciu o wzorec numeryczny, przy założeniu niezmienności kształtu pików składowych; $r(\gamma, x)$ – transformacja osi x ; γ_1 – parametr położenia piku modelu; γ_2 – parametr szerokości piku modelu; γ_3 – parametr wysokości piku modelu; x – dziedzina analizowanego sygnału

Fig. 2. Procedure of calculating peak model with numerical shape standard.

Wzorce normalizuje się tak, żeby wysokość wzorca oraz szerokość wzorca w połowie wysokości były równe jedności, natomiast położenie maksimum piku wynosiło 0. Normalizacja przeprowadzona w ten sposób pozwala bezpośrednio graficznie porównywać

kształty wzorców między sobą, a wyznaczone wartości parametrów modelu matematycznego bezpośrednio opisują piki analizowanego sygnału bez potrzeby dalszych przeliczeń.



Rys. 3. Normalizacja wzorca numerycznego: a) zarejestrowany pik „surowy” po estymacji i odjęciu linii bazowej, b) pik wzorcowy znormalizowany, wysokość równa 1, szerokość w połowie wysokości równa 1. Linia ciągłą przedstawiono wartości wyznaczone na podstawie interpolacji

Fig. 3. Normalization and interpolation of numerical shape standard

Podczas obliczania modelu na podstawie wzorca dla zadanych wartości chwil czasu niezbędne jest zastosowanie interpolacji, gdyż wzorec, jako cyfrowy zapis sygnału, zadany jest tylko w dyskretnych chwilach czasu. W pracy przyjęto metodę polegającą w zasadzie na aproksymacji splajnami kubicznymi wygładzającymi. Zamiast interpolacji zastosowano aproksymację ze względu na fakt obecności zakłóceń losowych w danych wzorca. Inną drogą jest wygładzenie wzorca i stosowanie interpolacji, jednakże należy mieć na uwadze, że procedury wygładzające wprowadzają własne błędy systematyczne, modyfikując kształt piku [13].

3. ZASTOSOWANIE WZORCA NUMERYCZNEGO DO POPRAWY DOKŁADNOŚCI MODELU PMG

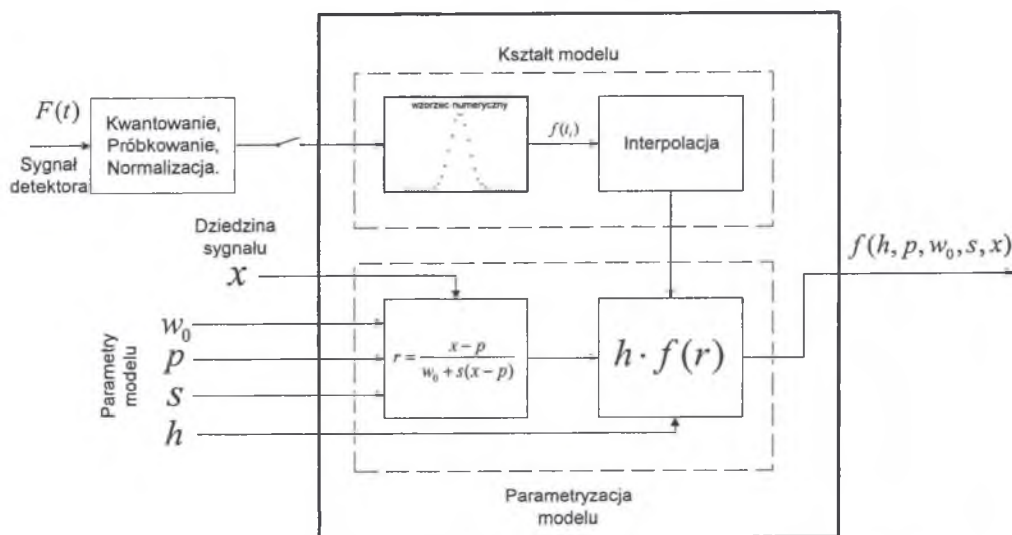
Funkcja Gaussa stanowi podstawę budowy wielu modeli matematycznych pojedynczych pików. Dobrym przykładem tego typu modeli jest funkcja gaussowska o odchyleniu standardowym σ w postaci wielomianu [15]:

$$f(t) = h \cdot \exp \left[-\frac{1}{2} \left(\frac{t - p}{w_0 + s_1(t - t_0) + s_2(t - t_0)^2 + \dots} \right)^2 \right], \quad (6)$$

Funkcja nosi nazwę PMG (Polynomially Modified Gaussian). Z zastosowaniem owej funkcji można, w przypadku ogólnym, opisać dowolny kształt piku. W praktycznych zastosowaniach należy ograniczyć stopień wielomianu w mianowniku. Dla $s_i = 0$ model jest po prostu funkcją Gaussa. Model postaci [2] :

$$f(t) = h \cdot \exp\left(-\frac{(t-p)^2}{2(w_0 + s(t-t_0))^2}\right), \quad (7)$$

często określany w literaturze jako model Levy-Martina, powstaje przez odrzucenie wszystkich członów wielomianu o $i \geq 2$ i jest najprostszym przypadkiem modelu PMG. W dalszej części tekstu model będzie określany jako model L-M. Model ten charakteryzuje się czterema parametrami, przy czym parametr s interpretuje się jako współczynnik asymetrii pików, wobec czego model można stosować do wyznaczania parametrów pików asymetrycznych metodą regresji. Niewielka liczba parametrów i ich oczywista interpretacja oraz możliwość modelowania pików asymetrycznych stanowią zalety modelu. Z drugiej strony, niewielka liczba parametrów może być przyczyną stosunkowo dużych błędów systematycznych modelu, co może prowadzić do dużych błędów estymacji w warunkach słabego rozdzielania pików składowych (dużej złożoności sygnału). Dokładność ową można poprawić, zastępując gaussowską funkcję bazową modelu, definiującą kształt pików: $f(x)=\exp(-x^2)$, wzorcem numerycznym i zachowując strukturę parametryczną modelu L-M. Model tak zmodyfikowany można przedstawić schematycznie jak na rys. 4.



Rys. 4. Procedura obliczania modelu matematycznego L-M z kształtem bazowym zdefiniowanym numerycznie

Fig. 4. Procedure of calculating L-M model with numerical shape standard

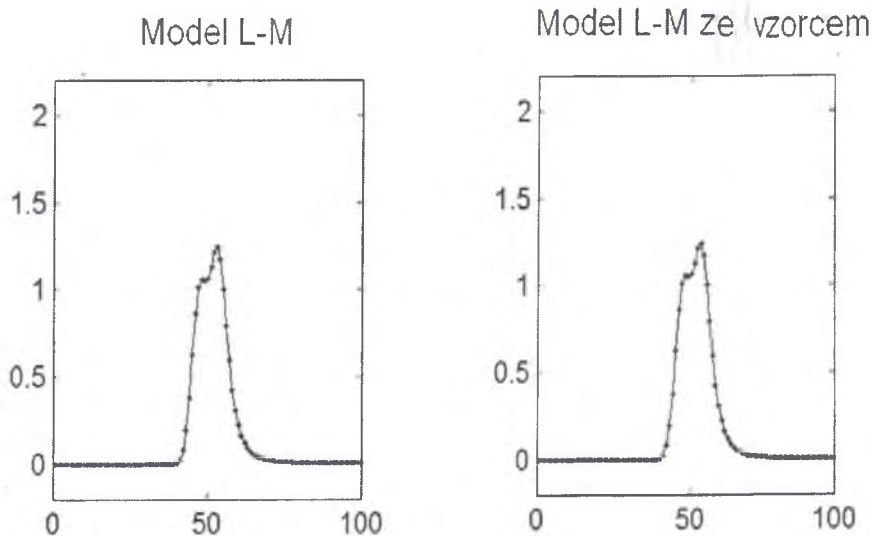
Zdefiniowany, jak wyżej, model pojedynczego pików ma taką samą liczbę parametrów jak model Levy-Martina i interpretacja tych parametrów jest podobna. Różnica polega na zastąpieniu funkcji gaussowskiej określającej bazy kształt pików wzorcem numerycznym uzyskanym w eksperymencie, a więc uwzględniającym bieżące właściwości aparatury chromatograficznej, a tym samym bardziej dokładnym. Model ten będzie nazywany w dalszej części tekstu jako NL-M.

4. PORÓWNANIE DOKŁADNOŚCI WYNIKÓW UZYSKIWANYCH Z ZASTOSOWANIEM MODELU L-M I NL-M

Opisany algorytm wyznaczania modelu piku ze wzorcem numerycznym zaimplementowano w środowisku programowania MATLAB. Analizowane sygnały chromatograficzne uzyskano z chromatografu gazowego CHROM5, wyposażonego w detektor płomieniowo-jonizacyjny. Chromatografowane substancje stanowiły alkohole

Badania symulacyjne polegały na budowie złożonych sygnałów z zapisów cyfrowych rzeczywistych sygnałów chromatograficznych, a następnie estymacji parametrów tych złożonych sygnałów metodą najmniejszych kwadratów z zastosowaniem modeli L-M i NL-M. Modele dopasowywano do sygnałów z zastosowaniem procedury 'nlfitt' środowiska MATLAB. Dla symulowanych sygnałów zmieniano złożoność poprzez zmianę stopnia rozdzielania R .

Na rys. 5. przedstawiono wynik dopasowania modeli LM i NLM do sygnału składającego się z dwóch pików składowych o zbliżonej wysokości i powierzchni, asymetrycznych. W procedurze minimalizacji estymowano wartości 8 parametrów – po 4 na każdy pik składowy sygnału. Stopień rozdzielania pików wynosi $R=0,9$, a więc analizowany sygnał charakteryzuje się wysokim stopniem złożoności. Za pomocą przebiegu kropkowanego przedstawiono próbki analizowanego sygnału, a linią ciągłą przebieg modelu dla optymalnego dopasowania.



Rys. 5. Wynik dopasowania modeli L-M i NL-M (L-M ze wzorcem) do sygnału złożonego z dwóch pików rzeczywistych, stopień rozdzielania $R = 0,9$

Fig. 5. Result of fitting procedure for $R=0.9$. Signal consists of two peaks

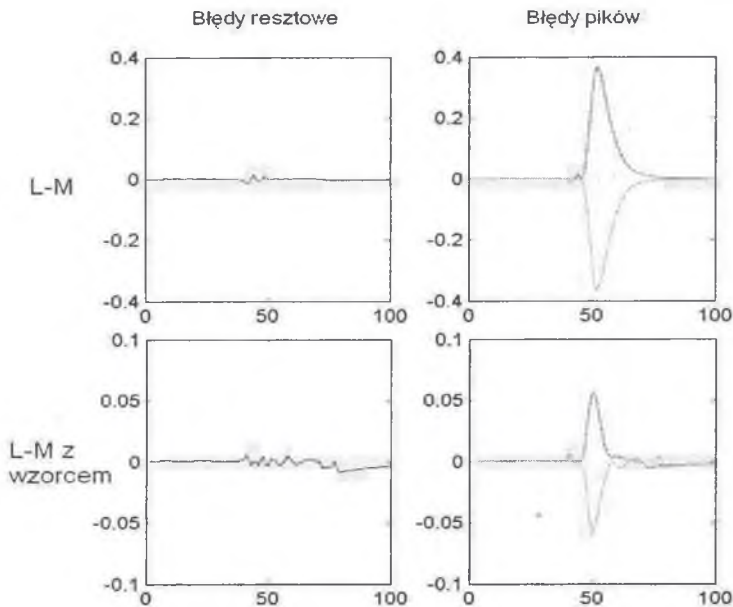
Wizualnie jakość dopasowania obu modeli do sygnału nie różni się istotnie. Wartości wskaźnika jakości dopasowania, czyli sumy kwadratów błędów resztowych, mają zbliżone wartości, które przedstawiono w tab.1, a przebieg błędów resztowych dla obu modeli pokazano na rys. 6. na wykresach z lewej strony.

Tabela 1

Porównanie wyników estymacji powierzchni pików sygnału złożonego z rys. 5 dla modeli L-M i NL-M.
 $\Delta P_1, \Delta P_2$ - błędy odtworzenia powierzchni pików składowych sygnału

Model	$J(\mathbf{b})$	ΔP_1	ΔP_2
L-M	$12,9 \cdot 10^{-4}$	45,9%	-47,6%
NL-M	$10,9 \cdot 10^{-4}$	3,70%	-5,27%

Dopiero analiza dokładności odtworzenia pików składowych pokazuje przewagę modelu NL-M ze wzorcem numerycznym. Na rys. 6 z lewej strony przedstawiono przebiegi błędów dopasowania modeli do sygnałów (błędów resztowych), a z prawej błędy odtworzenia krzywych pików składowych sygnału dla modeli L-M i NL-M (L-M ze wzorcem). Ponieważ przebiegi błędów odtworzenia pików sygnału są prawie symetryczne względem siebie (bo dopasowana do sygnału krzywa modelu zachowuje powierzchnię sumaryczną pod sygnałem), więc duże błędy odtworzenia kształtu pików składowych nie uwidaczniają się w przebiegu błędu resztowego (dopasowania modelu do sygnału). Dla modelu L-M maksymalne błędy odtworzenia krzywych pików składowych (rys. 6) sięgają 30% ($\approx 0,4/1,2$) wysokości sygnału (z rys. 5). Dla modelu NL-M te same błędy są mniejsze od 5% wysokości sygnału ($\approx 0,05/1,2$).

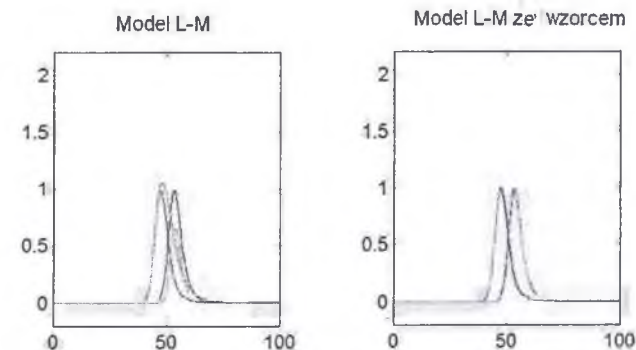


Rys. 6. Błędy resztowe oraz błędy odtworzenia pików składowych w złożonym sygnale z rys. 5 dla modeli L-M oraz NL-M (L-M ze wzorcem)

Fig. 6. Fitting errors and resolved peaks errors for L-M and NL-M models

Przedstawiony przykład ilustruje niebezpieczeństwa związane ze stosowaniem metody regresji do estymacji parametrów złożonych sygnałów chromatograficznych. Okazuje się, że mimo

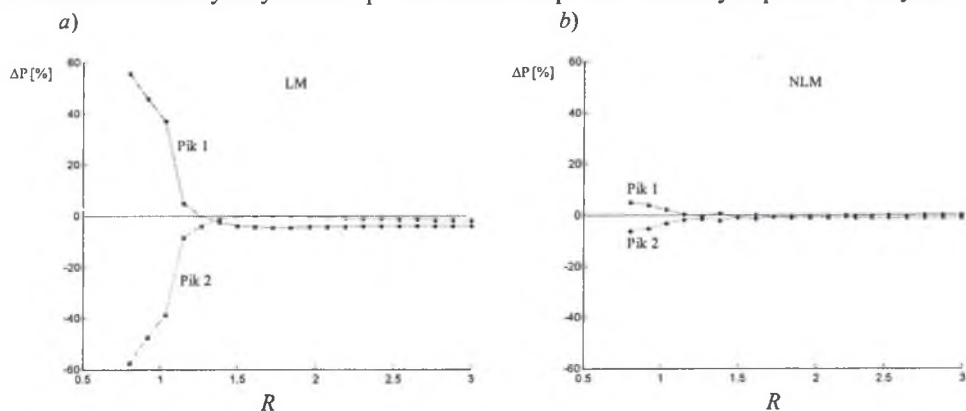
'dobrej' wartości wskaźnika jakości dopasowania błędy analizy mogą sięgać kilkudziesięciu procent, a niewielkie błędy systematyczne kształtu modelu, nieistotne dla osiągniętej dokładności w przypadku dobrego rozdzielania pików składowych są przyczyną dużych błędów w warunkach słabego rozdzielania, przy czym te duże błędy nie odzwierciedlają się w wartościach wskaźnika jakości. W pokazanym przykładzie, gdy stopień rozdzielania pików składowych $R = 0,9$, błędy odtworzenia powierzchni pików składowych dla modelu L-M wynoszą około 45% (tab.1.) podczas gdy dla modelu L-M ze wzorcem numerycznym błędy są na poziomie (3-5)% przy zbliżonych wartościach wskaźnika jakości dopasowania, tzn. różniących się mniej niż o 20%.



Rys. 7. Odtworzone na podstawie modeli L-M i NL-M piki składowe sygnał złożonego z rys. 5. Stopień rozdzielania pików $R=0,9$

Fig. 7. True peaks and peaks resolved with L-M and NL-M models for $R=0$.

Zależność błędów estymacji powierzchni pików z zastosowaniem modeli pików L-M oraz L-M ze wzorcem numerycznym od stopnia rozdzielania pików składowych pokazano na rys. 8.



Rys. 8. Porównanie błędów wyznaczenia powierzchni pików składowych dla modeli L-M i NL-M; Ośie odciętych na wszystkich wykresach oznaczają stopień rozdzielania pików R ; a) błędy wyznaczenia powierzchni pików składowych dla modelu L-M; b) błędy wyznaczenia powierzchni pików składowych dla modelu NL-M

Fig. 8. Area estimation errors vs. R for L-M and NL-M models

Stopień rozdzielania R zmieniano w zakresie od 3 (całkowite rozdzielenie) do 0,8 (niezauważalne rozdzielenie), przeprowadzono procedurę wyznaczania parametrów pików dla obu modeli i na podstawie tych parametrów odtwarzano powierzchnie pików. Otrzymane zależności wykreślono na rys. 8a) i 8b).

Widać, że w szerokim zakresie zmienności R błędy estymacji powierzchni nie zależą od R . Dla modelu L-M graniczną wartością jest $R = 1,3$. Dla $R < 1,3$ błędy odtworzenia pików składowych i ich powierzchni, w miarę zbliżania się R do wartości zerowej, gwałtownie rosną, osiągając poziom 60% dla $R = 0,8$. W przypadku modelu NL-M błędy także nie zależą od R w zakresie (1,2, 3), a w miarę dążenia R do 0 wzrastają tylko nieznacznie, o rząd wielkości mniej niż błędy modelu L-M. Dla stopnia rozdzielania $R = 0,8$ błędy te są na poziomie 6%, a więc jeszcze akceptowalnym.

5. PODSUMOWANIE I WNIOSKI

W chromatografii nie opracowano do tej pory ogólnego modelu matematycznego pojedynczego pików, a liczba modeli opisanych w literaturze i stosowanych w praktyce jest rzędu 100. Stąd koncepcja wykorzystania do modelowania pików wzorca numerycznego jest zgodna ze specyfiką metody pomiarowej, jaką charakteryzuje się chromatografia, a eksperymentalne wyznaczenie wzorca można traktować jako kalibrację kształtu modelu pików.

W artykule przedstawiono sposób zastosowania wzorca numerycznego do poprawy dokładności modelu matematycznego pików i jego zastosowanie w konkretnym przypadku modelu pików – tzw. wielomianowo modyfikowanej funkcji Gaussa PMG (Polynomially Modified Gaussian). W najprostszej postaci model taki charakteryzuje pik za pomocą 4 parametrów, w tym parametr wyrażający zmianę stopnia asymetrii pików. Idea polega na wprowadzeniu do modelu informacji o kształcie rzeczywistych pików chromatograficznych poprzez zastąpienie bazowej dla modelu funkcji Gaussa kształtem pików uzyskanym eksperymentalnie. Liczba parametrów modelu nie zmienia się.

Jak pokazano, zaprezentowana metoda pozwala zwiększyć kilkakrotnie dokładność analizy ilościowej złożonego sygnału chromatograficznego metodą regresji przy zastosowaniu modelu Levy-Martina, zwłaszcza w warunkach bardzo słabego rozdzielania składników analizowanej mieszaniny w kolumnie chromatograficznej. Zwiększenie dokładności okupione jest koniecznością przeprowadzenia dodatkowych eksperymentów w celu wygenerowania wzorca, a więc aby stosować metodę, trzeba ponieść dodatkowe koszty. Mimo generacji dodatkowych kosztów metoda jest łatwa do stosowania i w prosty sposób można zautomatyzować czynności konieczne do jej przeprowadzenia.

Przedstawioną metodykę można wprost stosować do poprawy dokładności czy też 'kalibracji' wielu innych modeli pików wyrażonych za pomocą funkcji elementarnych, o strukturze podobnej do tej przedstawionej na rys. 2 lub 4. A więc metoda nosi znamiona pewnej ogólności.

Potencjalne zastosowania metody to, na przykład, skrócenie czasu analizy w chromatografii procesowej, co prowadzi do zmniejszenia czasu opóźnienia (czasu martwego) w pętli sprzężenia zwrotnego przy sterowaniu procesem chemicznym. Dopuszczając niecałkowite rozdzielanie analizowanej mieszaniny w procesie chromatograficznym, a tym samym złożony (składający się z niecałkowicie rozdzielonych pików) sygnał na wyjściu chromatografu, można uzyskać skrócenie czasu analizy, nawet rzędu kilkudziesięciu razy.

Zaprezentowana metoda pozwala w prosty sposób na uzyskanie bardzo dokładnego modelu pików o niewielkiej liczbie parametrów. Model taki ma wobec tego bardzo korzystne

własności z punktu widzenia metody regresji nieliniowej, zapewniając dokładność konieczną do analizy sygnałów chromatograficznych, charakteryzujących się dużym stopniem nałożenia pików składowych.

Literatura

1. Anderson A.H., Gibb T.C., Littlewood A.B.: *Computer Analysis of Unresolved Non-Gaussian Gas Chromatograms by Curve Fitting*; *Anal.Chem.*, vol.42, no. 4, April 1970.
2. Di Marco V.B., Bombi G.G.: *Mathematical functions for representation of chromatographic peaks*, *Journal of Chromatography A*, 931 (2001), 1–30.
3. Hartley H.O.: *The Modified Gauss-Newton Method for the Fitting of Non-Linear Regression Function by Least Squares*, *Technometrics*, vol.3, no.2, 269-280, May 1961.
4. Ilewicz W.: *Estymacja parametrów sygnałów za pomocą wzorca zadanego numerycznie*. Materiały konferencyjne Konferencji Analityków Hutnictwa, październik 1996.
5. Ilewicz W.: *Problemy analizy złożonych sygnałów na przykładach sygnałów chromatograficznych i NMR*. Konferencja: Podstawowe Problemy Metrologii, Ustroń 2001.
6. Keller W.D., Lusenbrink T.R., Sederholm C.H.: *NMR Spectrum of 3-Chlorothietane, an Example of Complete NMR Spectral Analysis by Computer Techniques*, *Journal of Chemical Physics*, vol. 44, nr 2, January 1966.
7. Kirmse D.W., Westerberg A.W.: *Resolution Enhancement of Chromatograph Peaks*, *Anal.Chem.*, vol. 43, no. 8, July 1971.
8. Li J.: *Comparison of the capability of peak functions in describing real chromatographic peaks*, *Journal of Chromatography A*, 952 (2002), 63-70.
9. Nikitas P., Pappa-Louisi A., Papageorgiou A.: *On the equations describing chromatographic peaks and the problem of deconvolution of overlapped peaks*, *Journal of Chromatography A*, 912 (2001), 13-29.
10. Pap T.L., Papai Zs.: *Application of a new mathematical function for describing chromatographic peaks*; *Journal of Chromatography A*, 930 (2001), 53–60.
11. Papoff P., Ceccarini A., Lanza F., Fanelli N.: *Enhancing the quality of information obtained by a comparison between experimental and deconvolved peak parameters in ion chromatography*, *Journal of Chromatography A*, 789 (1997), 51-65.
12. Reh E.: *Peak-shape analysis for unresolved peaks in chromatography: comparison of algorithms*, *Trends In Analytical Chemistry*, vol.14, no.1, 1995.
13. Savitzky A., Golay M.J.E.: *Smoothing and Differentiation of Data by Simplified Least Squares Procedures*, *Anal.Chem.* , vol. 36, NO. 8, July 1964.
14. Scheeren P.J.H., Barna P., Smit H.C.: *A Software Package for the Evaluation of Peak Parameters in an Analytical Signal Based on a Non-Linear Regression Method*, *Analytica Chimica Acta*, 167 (1985), 65-80.
15. Torres-Lapasio J.R., Garcia-Alvarez-Couque M.C., J.J. Baeza-Baeza *Global treatment of chromatographic data with MICROCHROM*, *Analitica Chimica Acta* 348 (1997), 187-196.
16. Westerberg A.W.: *Detection and Resolution of Overlapped Peaks for An On-Line Computer System for Gas Chromatographs*, *Anal. Chem.* 41, 1770 (1969).

Recenzent: Dr hab.inż. Anna Domańska

Wpłynęło do Redakcji dnia 7 grudnia 2004

Abstract

The main goal of chromatography is separation of mixtures and its qualitative and quantitative analysis. Output signal from a chromatographic detector is recorded and a graphical presentation of the detector response vs. time is a chromatogram. The chromatogram contains of peaks and each of them corresponds to one component of the analyzed mixture. If peaks on the chromatogram are well separated parameters of each peak (position, width, height and area) are calculated directly from the chromatogram using simple mathematical routines. Next, on the basis of parameter values qualitative and quantitative analysis of the mixture is performed.

If the peaks on the chromatogram overlap each other we say that the chromatographic signal is complex. There is no direct method of determination of values of peak parameters in complex signal and analysis such a signal needs more sophisticated mathematical method. One of the most popular method of analyzing such a signals is nonlinear regression analysis performed by means of the least squares. In this method one has to assume a form of a mathematical parametric model of single peak. On the basis of the model of the single chromatographic peak a model of the signal is constructed as a sum of functions and each of them describes one peak in the complex signal. Then the parameters of the peaks are determined by fitting the assumed model to the real chromatographic data. The parameters that gives an optimal fit are used for purpose of resolving overlapped peaks and then qualitative and quantitative analysis.

There are described many models of chromatographic peaks in the literature based on elementary functions and proper choice of the function is the most important factor that determine the accuracy of analysis of the complex chromatographic signals with a nonlinear regression. The mathematical function for purpose of describing a shape of the single chromatographic peak should as close as possible represents the shape of the real signal with respect to the minimal number of parameters. In mostly practical cases the number of parameters is equal or less than 5, so the models contribute a systematic error of shape. This error could be ignored if peaks are well separated but if the signal is complex, even small error of the model can result in large errors of analysis.

Precision of peak model can be improved without increasing number of model parameters by using recorded signal of one chromatographic peak obtained experimentally That record is named numerical shape standard and it defines exactly the real shape of chromatographic peaks.

In the paper there is presented an algorithm of accuracy improvement of model named PMG (Polynomially Modified Gaussian). This model is based on Gauss function and can be used to describe skewed peaks. Improvement of model accuracy is achieved without changing number of model parameters by replacing Gaussian representing basic shape of peak in the PMG model with numerical shape standard. There is shown, that if peaks in complex signal are badly overlapped, improvement in accuracy is about order of magnitude comparing to PMG model. Presented procedure could be easy adopted to improve accuracy of other mathematical models of the chromatographic peak.