

**Witold ILEWICZ**POLITECHNIKA ŚLĄSKA, INSTYTUT AUTOMATYKI,  
ZAKŁAD SYSTEMÓW POMIAROWYCH**Zastosowanie kryterium Durbina-Watsona  
w automatycznej analizie sygnałów chromatograficznych**

Dr inż. Witold ILEWICZ

Ukończył studia na Wydziale Automatyki, Elektroniki i Informatyki Politechniki Śląskiej w 1996 r. Tytuł doktora uzyskał w 2004 r. Pracuje jako adiunkt w Instytucie Automatyki Politechniki Śląskiej. Swoje zainteresowania naukowe koncentruje wokół algorytmów analizy złożonych sygnałów chromatograficznych i ich automatyzacji. Członek Komisji Metrologii Oddziału PAN w Katowicach.



e-mail: witold.ilewicz@polsl.pl

**Streszczenie**

Automatyzacja analizy sygnałów chromatograficznych jest, w ogólnym przypadku, trudnym zadaniem wieloetapowym. Jednym z etapów jest wygładzanie sygnału w celu zmniejszenia wariancji zakłóceń losowych. Od jakości wygładzania zależy efektywność wykrywania pików oraz wyznaczania wstępnych ocen ich parametrów metodą analizy przebiegu pochodnych sygnału chromatograficznego. W pracy do wygładzania sygnału rozważa się procedurę wygładzania wielomianowego wg Savitzky'ego i Golay'a (SG). Optymalny dobór parametrów tej procedury, czyli szerokości okna oraz stopnia wielomianu wygładzającego jest kluczowy dla jakości wygładzania. Automatyzacja wymaga określenia kryterium, wg którego dokonuje się doboru tych parametrów. W tym celu zastosowano i przetestowano kryterium Durbina-Watsona.

**Słowa kluczowe:** wygładzanie wielomianowe, kryterium Durbina-Watsona, automatyzacja analizy sygnału chromatograficznego.

**The use of Durbin-Watson criterion for  
automatic analysis of chromatographic signals****Abstract**

Automation of analysis of chromatographic signals is a difficult task generally. One of the stage of the analysis is signal smoothing for purpose of eliminate of random noise. Efficiency of peaks detection procedure depends on quality of the smoothing when features of first and second order derivatives are used for peak detection. In the article Savitzky-Golay smoothing procedure is considered. Optimal choosing of parameters of the procedure is a main factor of quality of smoothing. Automation of calculation of smoothing parameters needs to define a criterion. One of the possible is a Durbin-Watson criterion of quality of smoothing, tested in the article.

**Keywords:** polynomial smoothing, Durbin-Watson criterion, automation of analysis of chromatographic signal.

**1. Wstęp**

Celem analizy sygnału chromatograficznego jest przyporządkowanie każdemu pikowi na chromatogramie nazwy substancji chemicznej, której odpowiada (identyfikacja składnika) oraz określenie stężeń poszczególnych składników poprzez wyznaczenie powierzchni pików. Sygnały chromatograficzne mogą zawierać ogromną liczbę pików, przy czym piki mogą się nakładać na siebie, co komplikuje analizę i prowadzi do stosowania skomplikowanych algorytmów analizy [1]. Analiza takich sygnałów jest na ogół trudna do automatyzacji. Jakość wygładzania sygnału ma tu podstawowe znaczenie, gdyż podstawie wygładzonego sygnału wyznacza się jego pochodne, a ich przebieg oraz odpowiednio dobrane wartości progów detekcji dla sygnału i jego pochodnych determinują jakość procedury wykrywania pików. Ogólnie procedura wykrywania pików na chromatogramie na podstawie przebiegu pochodnych przedstawia się następująco: miejsca zerowe

pierwszej pochodnej sygnału chromatograficznego przy zmianie znaku pochodnej z dodatniego na ujemny określają położenie wierzchołków pików na osi czasu. W przypadku pików nałożonych procedura oparta na przebiegu pierwszej pochodnej może okazać się zbyt mało selektywna, wtedy liczbę pików i położenie ich maksimów określa się na podstawie przebiegu drugiej pochodnej – położenia minimów lokalnych drugiej pochodnej przy ujemnej wartości tej pochodnej utożsamia się z położeniem wierzchołków pików.

Procedura różniczkowania sygnału cyfrowego w obecności zakłóceń losowych należy do zadań trudnych [2], bowiem przy różniczkowaniu sygnału różniczkuje się również zakłócenia. Związek między gęstościami widmowymi mocy sygnału  $x$  i jego pochodnej  $x'$  jest następujący:

$$S_{x'}(\omega) = \omega^2 \cdot S_x(\omega) \quad (1)$$

Wynika z niego, że moc sygnału po różniczkowaniu wzrasta  $\omega^2$  razy. W przypadku sygnału chromatograficznego, którego składowa użyteczna jest wolnozmienna w stosunku do szumów, stosunek sygnał-szum po różniczkowaniu pogorszy się, a przebieg pochodnej nie będzie gładki. W związku z tym należy zaproponować procedurę, która zmniejszy udział szumów w sygnale, tak aby przebieg pochodnej był odpowiednio gładki.

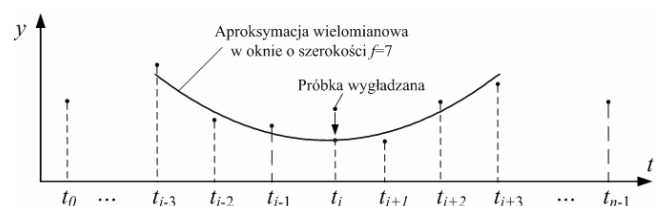
**2. Procedura wygładzania  
Savitzky'ego-Golaya**

W eksperymencie chromatograficznym rejestruje się sygnał wyjściowy  $y$  z detektora chromatograficznego w postaci cyfrowej:

$$y_i = c_i + z_i + \varepsilon_i, \quad i = 0..n-1, \quad (2)$$

gdzie  $y_i = y(t_i)$  – wartości sygnału  $y$  w dyskretnych chwilach czasu  $t_i$ ,  $t_i = t_0 + i \cdot \Delta t$ ,  $t_0$  – chwila rozpoczęcia rejestracji sygnału,  $\Delta t$  – stały okres próbkowania;  $c_i$  – wartości składowej użytecznej sygnału;  $z_i$  – wartości dryftu linii bazowej;  $\varepsilon_i$  – zakłócenia losowe;  $n$  – liczba zarejestrowanych próbek sygnału.

W procedurze wygładzania wielomianowego Savitzky'ego-Golaya elementy wektora danych  $y$  są zastępowane wartościami wielomianu  $k$ -tego stopnia aproksymującego lokalnie sygnał w oknie o szerokości  $f$ , przy czym próbka wygładzana  $y_i$  znajduje się w środku okna. Szerokość okna jest liczbą nieparzystą  $f = 2 \cdot m + 1$ , gdzie  $m$  oznacza liczbę próbek otaczających próbkę wygładzaną z lewej i z prawej strony. Na ogół  $f \ll n$ . Aby uzyskać efekt wygładzenia stopień wielomianu  $k$  musi być przynajmniej o 2 mniejszy od szerokości okna  $f$ . Ilustrację procedury stanowi rys. 1 ( $f=7$ , wielomian 2 stopnia jako linia ciągła). Podobnie postępuje się dla próbek w kolejnych chwilach czasu  $t_{i+1}$ ,  $t_{i+2}$ , ... itd.



Rys. 1. Ilustracja procedury wygładzania wielomianowego dla próbki w chwili  $t_i$   
Fig. 1. Illustration of Savitzky-Golay smoothing procedure at time  $t_i$

W swojej pracy [3] Savitzk'y i Golay pokazali, że obliczanie wygładzonego sygnału przy spełnieniu założenia, że okres próbkowania jest stały, sprowadza się do splotu sygnału wygładzanego z oknem współczynników o stałych wartościach i szerokości  $f$ , a więc procedura, przy swoich zaletach, jest bardzo prosta obliczeniowo.

Matematycznie zagadnienie wygładzania wielomianowego przedstawia się następująco: przy założeniu, że wielomian dobrze aproksymuje dane w oknie o szerokości  $f$  zadanie aproksymacji sprowadza się do wyznaczenia współczynników  $b$  wielomianu założonego stopnia metodą najmniejszych kwadratów dla modelu:

$$y_f = T_f b + \varepsilon_f \quad (3)$$

gdzie  $y_f$  – wektor wartości danych w oknie o rozmiarze  $f \times 1$ ,  $T_f$  – macierz Vandermonda zmiennej niezależnej o rozmiarze  $f \times k+1$ , gdzie  $k$  – stopień wielomianu aproksymującego;  $b$  – wektor współczynników wielomianu o rozmiarze  $(k+1) \times 1$ ;  $\varepsilon_f$  – wektor realizacji błędów o rozmiarze  $f \times 1$ . Rozwiązanie uzyskane metodą najmniejszych kwadratów ma postać:

$$\hat{b} = (T_f^T T_f)^{-1} T_f^T y_f, \quad (4)$$

a wartości wielomianu aproksymującego wyznacza się z wzoru

$$\hat{y}_f = T_f \hat{b} = T_f (T_f^T T_f)^{-1} T_f^T y_f = H y_f, \quad (5)$$

gdzie  $H = T_f (T_f^T T_f)^{-1} T_f^T$  jest dla wybranego stopnia wielomianu i szerokości okna wygładzającego  $f$  macierzą stałą o rozmiarze  $f \times f$ , niezależną od indeksu  $i$  wygładzanej próbki  $y_i$ . Ponieważ wygładzana dana  $y_i$  leży w środku okna, do wygładzania wykorzystuje się tylko środkowy wiersz  $h$  macierzy  $H$ , jednakowy dla wszystkich próbek wygładzanych  $y_i$ , a więc procedura wygładzania sygnału  $y$  w chwili  $t_i$  sprowadza się do wyrażenia:

$$\begin{aligned} \hat{y}(t_i) &= \sum_{j=-m}^m h_j \cdot y(t_{i+j}) = h \cdot y_f, \\ h &= [h_{-m} h_{-m+1} \dots h_0 \dots h_{m-1} h_m], \\ y_f &= [y(t_{i-m}) \dots y(t_i) \dots y(t_{i+m})]^T, \end{aligned} \quad (6)$$

gdzie  $y_f$  jest wektorem wartości próbek sygnału o długości  $f=2 \cdot m+1$ . W swojej pracy Savitzk'y i Golay podali wartości wektorów współczynników wygładzających  $h$  dla różnych szerokości okna i stopni wielomianu. Zaletą procedury jest prostota obliczeniowa. Za pomocą procedury można równocześnie wygładzać i różniczkować sygnał, przyjmując wartości pochodnej sygnału równe wartościom pochodnej wielomianu wygładzającego w punkcie wygładzanym, które bardzo łatwo wyznaczyć. Wartości współczynników dla procedury różniczkowania z wygładzaniem podano dla różnych szerokości okien wygładzających oraz rzędu pochodnej w [3]. Jeśli stopień wielomianu wynosi  $k$ , to w jednym kroku wyznaczyć można co najwyżej  $k$ -tą pochodną sygnału, pochodne wyższych niż  $k$  rzędów będą miały wartości równe 0. Jakość wygładzania metodą S-G zależy od odpowiedniego doboru parametrów procedury, którymi są: stopień wielomianu  $k$  oraz szerokość okna  $f$ . Na ogół operator oprogramowania dobiera te wartości ręcznie. W algorytmie automatycznej analizy sygnału należy podać kryterium, wg którego parametry te byłyby dobierane bez udziału człowieka.

### 3. Kryterium Durbin-Watsona

Automatyczny dobór szerokości okna przeprowadza się za pomocą kryterium Durbin-Watsona. Statystyka ta wyraża się wzorem [4]:

$$DW = \frac{\sum_{i=2}^n [(y_i - y_{w,i}) - (y_{i-1} - y_{w,i-1})]^2}{\sum_{i=1}^n (y_i - y_{w,i})^2} \cdot \left( \frac{n}{n-1} \right), \quad (7)$$

gdzie  $y_i$  – wartość  $i$ -tej próbki zarejestrowanego sygnału chromatograficznego,  $y_{w,i}$  – wartość  $i$ -tej próbki po wygładzeniu sygnału,  $n$  – liczba próbek w chromatogramie. Statystyka DW jest miarą korelacji elementów wektora  $r = y - y_w$  obliczanego jako różnice między sygnałem oryginalnym i wygładzonym. W liczniku statystyki DW występuje suma kwadratów różnic kolejnych residuów sygnałów oryginalnego i wygładzonego:  $r_i$  i  $r_{i-1}$ , a w mianowniku suma kwadratów residuów. Przyjmuje się założenie, że brak korelacji między elementami wektora  $r$  oznacza, że procedura wygładzania ma dobrze dobrane parametry. W przypadku braku korelacji między kolejnymi elementami wektora  $r$  wartość DW jest bliska 2, co wynika z wzoru (7), dla większej jasności zapisanego poniżej dla reszt  $r_i$ :

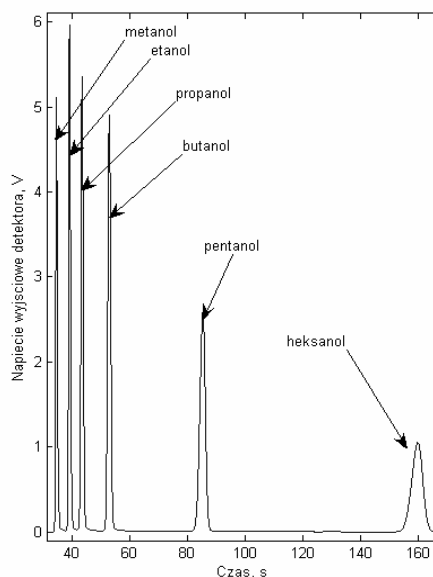
$$\begin{aligned} DW &= \frac{\sum (r_i - r_{i-1})^2}{\sum r_i^2} = \frac{\sum (r_i^2 - 2 \cdot r_i r_{i-1} + r_{i-1}^2)}{\sum r_i^2} \approx \\ &[\text{brak korelacji} \Rightarrow \sum r_i r_{i-1} \approx 0] \approx \frac{\sum (r_i^2 + r_{i-1}^2)}{\sum r_i^2} \approx 2 \end{aligned} \quad (8)$$

Gdy okno wygładzające jest zbyt wąskie, wartości kryterium  $DW > 2$ . Przyjęcie w procedurze wygładzania zbyt dużej szerokości okna wygładzającego  $f$  i zbyt małego stopnia wielomianu prowadzi do wartości  $DW < 2$ .

Zastosowanie kryterium DW do optymalnego doboru szerokości okna polega na przeglądzie wartości DW dla różnych szerokości okna wygładzającego i wybraniu tej szerokości okna, dla której wartość DW jest możliwie najbliższa liczbie 2 [2].

### 4. Zastosowanie procedury do danych chromatograficznych

Procedurę automatycznego doboru optymalnej szerokości okna zastosowano do sygnałów uzyskanych z chromatografu gazowego Varian 3800.



Rys. 2. Chromatogram mieszaniny 6 alkoholi  
Fig. 2. Chromatogram of mixture of 6 alcohols

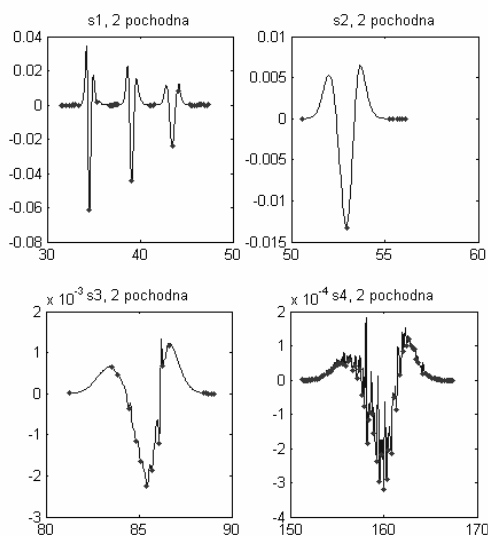
Chromatografowano mieszaninę sześciu alkoholi, od metanolu do heksanolu, w stałej temperaturze 60°C. Zastosowano kapilarną kolumnę chromatograficzną CP-Sil5CB o średnicy 0.53mm dedykowaną do rozdzielania alkoholi. Dane z detektora FID (płomieniowo-jonizacyjny) zarejestrowano z częstotliwością próbkowania 40Hz, wartość kwantu przetwornika A/C  $q=10^{-6}$  V. Uzyskano chromatogram z 6 rozdzielonymi pikami, reprezentującymi w kolejności metanol, etanol, propanol, butanol, pentanol, heksanol. Uzyskany chromatogram przedstawiono na rys. 2.

Odchylenie standardowe zakłóceń losowych wyniosło ok.  $10^{-5}$  V. Rozkład błędów losowych jest zbliżony do rozkładu normalnego.

## 5. Wyniki

Sygnal z rys. 2 podzielono na 4 segmenty: pik 1-3 (s1), pik 4 (s2), pik 5 (s3), pik 6 (s4), grupując pik 5 w segmenty o zbliżonych szerokościach

Stosując kryterium DW do sygnału z rys. 2 wyznaczono optymalną szerokość okna wygładzającego dla stopnia wielomianu 3:  $f=7$  przy wartości kryterium  $DW=2,36$ . Dla okna o szerokości 7 oraz stopnia wielomianu wygładzającego 3 wyznaczono pierwszą i drugą pochodną sygnału w poszczególnych segmentach. Wyniki przedstawiono na rys. 3 4 pierwsze pik charakteryzują się pojedynczymi minimami w ujemnej części drugiej pochodnej, co w połączeniu z progowaniem sygnału pozwala poprawnie zidentyfikować liczbę pików. Dla pików 5 i 6 wygładzenie sygnału pochodnej jest tu niewystarczające i w przebiegu drugiej pochodnej widoczne są wielokrotne minima lokalne. Na skutek tego pojedynczy pik jest interpretowany jako grupa pików mocno nałożonych, a więc analiza drugiej pochodnej w przypadku pików 5 i 6 prowadzi do błędnej liczby wykrytych pików.



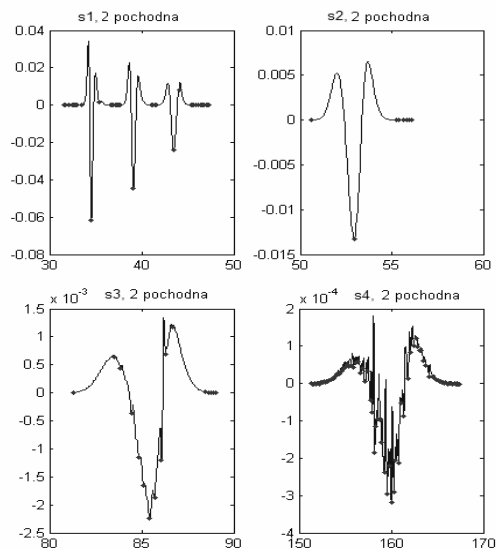
Rys. 3. 2-gie pochodne sygnału wyznaczone metodą SG dla szerokości okna 7 i wielomianu 3 stopnia. Kropkami zaznaczono minima lokalne drugiej pochodnej

Fig. 3. 2-nd derivatives of chromatographic signal calculated by SG method with width of window equal to 7 and 3-rd order polynomial. Dots indicate 2-nd derivatie minimas

Następnie do każdego z segmentów osobno dobrano optymalną wg kryterium DW szerokość okna wygładzającego dla stopnia wielomianu 3. Uzyskane wyniki przedstawiono na rys. 4. Jak widać zmiana nastąpiła tylko dla segmentu s1, gdzie wyznaczono wartość szerokości okna 5. W przypadku segmentów s3 i s4 sytuacja jest podobna jak w przypadku poprzednim – to znaczy, zwłaszcza dla segmentu 4, poprawne wyznaczenie liczby pików

w segmencie jest niemożliwe. Należałoby zwiększyć szerokość okna.

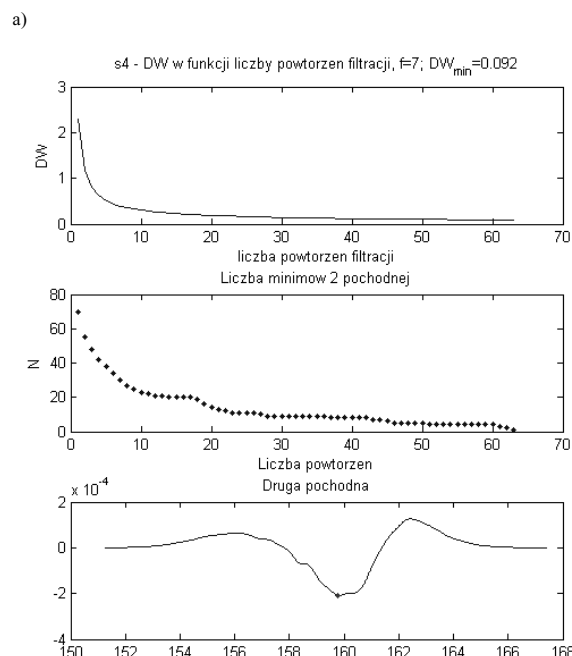
Wyniki pokazują, że kryterium DW nie jest skuteczne jeśli częstotliwość próbkowania oraz szerokość pików są duże. Aby rozwiązać problem proponuje się zastosować wielokrotne powtórzenie wygładzania metodą Savitzky'ego-Golaya dla tej samej próby przy tej samej szerokości okna, przy czym krotność powtórzeń przy ustalonej szerokości okna staje się dodatkowym parametrem, którego odpowiedni dobór może poprawić to efekty wygładzania. Autorzy w [4] nie uwzględnili tego aspektu w swojej pracy.

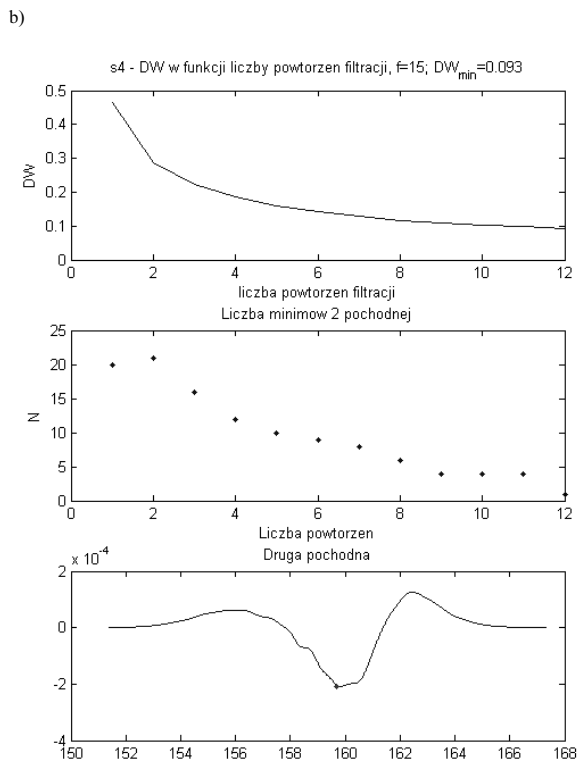


Rys. 4. Wynik obliczenia drugiej pochodnej sygnałów w segmentach. Kryterium DW stosowane osobno do każdego segmentu

Fig. 4. Results of 2-nd derivative calculation for particular segment sof signal DW criterion applied to each segment separately

Wyniki wielokrotnego wygładzania dla segmentu s4, składającego się z najszerszego, pojedynczego pik, przedstawiono na rys. 5.





Rys. 5. Wyniki wyznaczenia drugiej pochodnej przy wielokrotnym wygładzaniu dla segmentu s4; a) szerokość okna  $f=7$ ; liczba powtórzeń filtracji  $n=63$ ;  $DW=0.092$ ; b) szerokość okna  $f=15$ ; liczba powtórzeń filtracji  $n=12$ ;  $DW=0.093$

Fig. 5. Results of 2-nd derivative of s4 segment calculation for repeated smoothing. a) window width  $f=7$ ; number of repetition  $n=63$ ;  $DW=0.092$ ; b) window width  $f=15$ ; number of repetition  $n=12$ ;  $DW=0.093$

## 6. Podsumowanie

Automatyczna analiza złożonych sygnałów chromatograficznych wymaga wyznaczenia pochodnych sygnału chromatograficznego w celu prawidłowej detekcji pików na chromatogramie. Zaproponowane w literaturze wygładzanie wielomianowe wg Savitzky'ego i Golaya przy zastosowaniu kryterium Durbin-Watsona do doboru szerokości okna wygładzającego w pewnych przypadkach prowadzi do uzyskania sygnału drugiej pochodnej z wieloma minimami lokalnymi, co jest przyczyną błędnego wyznaczenia (zawyżenia) liczby pików w sygnale. W pracy pokazano przykłady uzyskania błędnych wyników dla rzeczywistych

sygnałów chromatograficznych, charakteryzujących się dużym stosunkiem sygnał-szum. Kryterium  $DW$  jest w tym przypadku nieczułe i nie daje dobrych rezultatów zwłaszcza przy pikach szerokich, o większych czasach retencji.

W pracy podjęto próbę rozwiązania problemu poprzez zastosowanie wielokrotnego wygładzania sygnału przy oknie o stałej szerokości. W trudnych przypadkach, gdy wygładza się segmenty chromatogramu zawierające szerokie piki, udało się znaleźć liczbę powtórzeń procedury filtracji przy ustalonej szerokości okna, prowadzącą do prawidłowego określenia liczby pików. Dodatkowo przy stosowaniu szerszego okna wygładzającego do uzyskania prawidłowego przebiegu pochodnej wystarczyła mniejsza liczba powtórzeń procedury wygładzania. Optymalny przebieg drugiej pochodnej uzyskano dla wartości  $DW=0.23$  dla pików węższego (segment s3) oraz dla  $DW=0.093$  dla pików szerszego (segment s4). Należy zauważyć, że obie te wartości znacznie odbiegają od wartości teoretycznie najlepszej, czyli  $DW=2$ . Przy tym minimum drugiej pochodnej znajduje się praktycznie w tym samym miejscu, co przy wygładzaniu prowadzącym do wartości  $DW$  bliskiej liczby 2. Okazuje się, że zniekształcenie systematycznie pochodnej (poszerzenie przebiegu drugiej pochodnej w osi odciętych oraz spłaszczenie w osi rzędnych) powstałe na skutek powtarzania procedury wygładzania nie prowadzi do błędów w wyznaczeniu położenia pików. Dodatkowo, im szerszy pik, tym mniejsza wartość  $DW$  daje optymalny przebieg drugiej pochodnej. Gdy celem stosowania wygładzania metodą Savitzky'ego-Golaya jest wyznaczenie przebiegu pochodnych sygnału (zwłaszcza drugiej pochodnej), dopuszcza się stosowanie mniejszych wartości  $DW$  niż 2 jako optymalne, jednakże dobór tych wartości zależy od konkretnego systemu oraz od szerokości pików składowych w sygnale i wydaje się trudny do automatyzacji.

## 7. Literatura

- [1] Andrade L., Manolagos E., Signal Background Estimation and Baseline Correction Algorithms for Accurate DNA Sequencing, J.of VLSI Sig. Proc., 35, 229-243, 2003
- [2] Piotrowski J.: Procedury pomiarowe i estymacja sygnałów. Skrypt uczelniany nr 1889, Politechnika Śląska, Gliwice, 1994
- [3] A. Savitzky, M. Golay, Anal. Chem. 36 (1964) 1627
- [4] Vivo-Truyols et al.: Automatic program for peak detection and deconvolution of multi-overlapped chromatographic signals, J. Chromatogr. A 1096 (2005) 133-145

Artykuł recenzowany

## INFORMACJE

# Zapraszamy do publikacji reklam w czasopiśmie PAK

Redakcja czasopisma POMIARY AUTOMATYKA KONTROLA  
44-100 Gliwice, ul. Akademicka 10, pok. 30b,  
tel./fax: 032 237 19 45, e-mail: wydawnictwo@pak.info.pl