

Электроника, фотоника, приборостроение и связь

УДК 681.391:543/545

КОРРЕЛЯЦИОННАЯ ОБРАБОТКА ДАННЫХ ДЛЯ ОПРЕДЕЛЕНИЯ УСРЕДНЕННОЙ ДИФФЕРЕНЦИАЛЬНОЙ ХРОМАТОГРАММЫ ПРИ МНОГОКРАТНОМ ВВОДЕ ПРОБ

М.Ж. Сайфулин¹

Самарский государственный технический университет
Россия, 443100, г. Самара, ул. Молодогвардейская, 244

E-mail: sayfulinmarsel@yandex.ru

Аннотация. Показана связь входного и выходного сигналов хроматографа. Отмечена целесообразность повышения отношения сигнал/шум и использования для этого корреляционной обработки данных. Приводится алгоритм вычисления хроматограммы, усредненной по множеству вводов пробы некоторого отдельно взятого образца. Далее этот алгоритм модифицируется на случай вычисления дифференциальной усредненной хроматограммы, когда требуется определить различие компонентного состава двух различных образцов. Для формирования входных сигналов используется псевдослучайная последовательность Лежандра. Алфавит этой последовательности 1 или 0: 1 означает ввод пробы, 0 – отсутствие ввода пробы в данный момент времени. В случае анализа двух образцов входные последовательности, управляющие вводом проб первого и второго образцов, являются дополнительными: если проба первого образца не вводится, то вводится проба второго образца, и наоборот. Алгоритм вычисления усредненной хроматограммы сводится к регрессионному анализу. Вводится информационная матрица плана эксперимента. Точность оценивания характеризуется с помощью ковариационной матрицы оценок. Выбор оптимальной входной последовательности ввода проб осуществляется исходя из условия D-оптимальности (условия максимизации определителя информационной матрицы), что обеспечивает минимум обобщенной дисперсии всех оценок ординат усредненной хроматограммы. Приведены результаты численного имитационного эксперимента, доказывающие эффективность использования корреляционного метода обработки данных для усреднения хроматограмм.

Ключевые слова: корреляционная обработка данных, многократный ввод пробы, отношение сигнал/шум, точность оценивания, дифференциальная хроматограмма.



© Автор(ы), 2024

¹ Марсель Жаудатович Сайфулин, аспирант кафедры информационно-измерительной техники.

Качество оценок параметров хроматографического сигнала, полученных на этапе первичной обработки, и качество результатов всего анализа в целом определяется в первую очередь качеством исходной хроматограммы. Поэтому ее улучшению, повышению отношения сигнал/шум необходимо уделять внимание уже в ходе хроматографического эксперимента. Особенно это важно, когда измерения проводятся на пределе обнаружения хроматографа [1]. Снижение предела обнаружения компонентов позволяет проводить эксперименты с малыми объемами пробы и быстрым их вводом в колонку, что позволяет получать хроматограммы с хорошо разделенными пиками. Это особенно важно при обработке быстро элюируемых компонентов. Однако в результате проведения таких анализов часто получаются хроматограммы низкого качества, т. е. с малым отношением сигнал/шум. Таким образом, важнейшая проблема – повышение качества хроматографических сигналов до проведения их первичной и вторичной обработки [2].

Выходной сигнал хроматографа $s(t)$ как реакция на входное воздействие $x(t)$, имеющее форму узкого импульса, описывается интегралом линейной суперпозиции:

$$s(t) = \int_{-\infty}^{+\infty} g(\tau)x(t-\tau)d\tau. \quad (1)$$

Входное воздействие $x(t)$ представляют дельта-функцией $\delta(t)$, тогда интегральное соотношение (1) примет вид

$$s(t) = \int_{-\infty}^{+\infty} g(\tau)\delta(t-\tau)d\tau = g(t). \quad (2)$$

Сигнал $s(t)$ называют хроматограммой (совпадает с приборной функцией). В случае анализа многокомпонентных смесей $s(t)$ может быть представлена в виде аддитивной модели совмещенных хроматографических пиков [3]. Длительность хроматограммы определяется временем элюирования последнего компонента смеси T_s из условия $s(t) = g(t) \cong 0$ при $t \geq T_s$.

Практически выходной сигнал хроматографа $y(t)$ представляет собой аддитивную смесь полезного сигнала $s(t)$, несущего информацию о качественном и количественном составе анализируемых веществ, и флуктуационной помехи (шума) $n(t)$:

$$y(t) = s(t) + n(t). \quad (3)$$

Шум $n(t)$ представляют распределенным по нормальному закону $n(t) \sim N(0, \sigma_n^2)$ с нулевым математическим ожиданием и дисперсией σ_n^2 . Нормальный характер распределения шума в хроматографическом сигнале подтверждается измерениями [3].

В хроматографии именно шумом определяется предел обнаружения (детектирования) компонента: высота пика определяемого вещества должна не менее чем вдвое превышать уровень флуктуации шума (нулевого сигнала). По существующим методикам отношение сигнал/шум (s/n) также часто рекомендуют определять как отношение удвоенной высоты пика на хроматограмме к размаху h флуктуации шума [4]. Этот размах можно оценить в единицах СКО σ_n ,

например принять $h = 6\sigma_n$. Таким образом, для понижения предела обнаружения компонентов необходимо повышать отношение сигнал/шум хроматографического сигнала.

На практике эффект повышения отношения сигнал/шум может достигаться предварительным концентрированием компонентов анализируемой пробы. Однако при количественном анализе ко всем методам предварительной обработки пробы и концентрирования следует относиться осторожно и стремиться выполнить анализ без химического и физического изменения состава пробы. Кроме того, отсутствие предварительной обработки пробы облегчает автоматизацию анализа.

Традиционно используемые методы подавления помехи, основанные на цифровой фильтрации хроматографического сигнала, в большинстве случаев изменяют форму сигнала [5]. Например, на хроматограмме после фильтрации шумов может измениться форма пиков, и чем лучше результат фильтрации шумов, тем более существенно изменение формы пика.

Для уменьшения влияния $n(t)$ и повышения отношения сигнал/шум можно использовать усреднение по ансамблю реализаций, когда осуществляется регулярный ввод проб с интервалами между вводами, большими T_s . Каждая реализация отсчитывается от момента ввода очередной пробы. Если таким образом осуществить p вводов пробы, а затем рассчитать усредненную по множеству полученных реализаций хроматограмму, то отношение сигнал/шум будет возрастать пропорционально $\sqrt{p}$. Однако такой подход к повышению соотношения сигнал/шум требует обеспечения стабильности работы хроматографа в течение длительного времени и приводит к увеличению общего времени проведения эксперимента $T_{\text{экс}} = pT_s$. Если, например, T_s составляет 30 минут, то это может привести к неприемлемо большим значениям $T_{\text{экс}}$.

В этой связи перспективным методом получения усредненных на некотором множестве вводов проб хроматограмм, позволяющим увеличить отношение сигнал/шум при приемлемом суммарном времени проведения эксперимента $T_{\text{экс}}$, является корреляционная обработка данных [6, 7].

Анализ известных результатов в области корреляционной обработки хроматографических сигналов позволил выделить комплекс нерешенных проблем, связанных с созданием эффективных и точных алгоритмов вычислений, позволяющих определять несмещенные оценки усредненной хроматограммы.

Необходимо отметить, что практически во всех работах в качестве входной последовательности используются M -последовательности. Эта разновидность псевдослучайных последовательностей часто применяется в задачах идентификации объектов управления, а алгоритм их получения прост в реализации. Однако использование M -последовательностей сужает возможность корреляционной обработки, поскольку период этой последовательности изменяется большими скачками, что является нежелательным при корреляционной обработке хроматографических сигналов.

Таким образом, важным представляется выбор входных последовательностей, позволяющих упростить вычислительные алгоритмы при сохранении высокой точности результатов, без дополнительных систематических погрешностей. В данной работе в качестве входной последовательности используется псевдослучайная последовательность Лежандра.

В соответствии с этим разработан алгоритм получения оценок усредненного хроматографического сигнала, учитывающий вид входной последовательности, конечность как длины последовательности, так и времени усреднения. Далее этот алгоритм модифицирован для вычисления усредненной дифференциальной хроматограммы, когда интересует различие компонентного состава двух образцов (например, исследуемого и эталонного).

Определение усредненной хроматограммы при многократном вводе проб одного образца. При этом на вход хроматографа подается последовательность проб $x(t)$. Обработка связана с вычислением оценок автокорреляционной функции $R_{xx}(\tau)$ входного сигнала и взаимно-корреляционной функции $R_{xy}(\tau)$ входного $x(t)$ и выходного $y(t)$ сигналов.

Пусть $g(t)$ – искомая хроматограмма, $g(t) = 0$, при $t < 0$ и $t \geq T_s$, где T_s – время элюирования хроматограммы. Представим функции $g(t)$, $y(t)$, $x(t)$ в виде последовательности дискретных отсчетов: g_j , $j = 0, 1, 2, \dots, m$; y_j , $j = 0, 1, 2, \dots, k \geq m$; x_l , $l = -m, -(m-1), \dots, 0, 1, 2, \dots, k$; $m = \frac{T_s}{\Delta}$, $k = \frac{T}{\Delta}$, Δ – интервал дискретизации по времени. Здесь T – время, соответствующее последнему вводу пробы. Таким образом, в дискретном виде отсчеты входного воздействия могут действовать от $-m$ до k .

В дискретной форме для динамической системы справедливо соотношение [8]

$$\hat{R}_{xy}(j) = \sum_{l=0}^m g_l \hat{R}_{xx}(j-l), \quad j = 0, 1, 2, \dots, m, \quad (4)$$

где $\hat{R}_{xy}(j) = \frac{1}{k+1} \sum_{l=0}^k y_l x_{l-j}$; $\hat{R}_{xx}(j) = \frac{1}{k+1} \sum_{l=0}^k y_l x_{l-j}$.

После введения матричных обозначений

$$G = \begin{bmatrix} g_0 \\ g_1 \\ \dots \\ g_m \end{bmatrix}; \quad Y = \begin{bmatrix} y_0 \\ y_1 \\ \dots \\ y_k \end{bmatrix}; \quad X = \begin{bmatrix} x_0 & x_{-1} & x_{-2} & \dots & x_{-m} \\ x_1 & x_0 & x_{-1} & \dots & x_{-m(-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_k & x_{k-1} & x_{k-2} & \dots & x_{-(m-k)} \end{bmatrix} \quad (5)$$

соотношение (4) можно представить в виде

$$X^T Y = (X^T X) G, \quad (6)$$

где T – знак транспонирования матрицы.

Обозначим вектор оценок приборной функции (усредненную хроматограмму) через $\hat{G}$, тогда из (6)

$$\hat{G} = (X^T X)^{-1} X^T Y. \quad (7)$$

Алгоритм вычисления усредненной хроматограммы эквивалентен методу наименьших квадратов, и все результаты регрессионного анализа могут быть использованы при корреляционной обработке хроматографических данных [9–11]. При этом матрицу $C = X^T X$ будем рассматривать как информационную матрицу плана эксперимента [12].

Точность оценивания ординат $\hat{G}$ описывается ковариационной матрицей оценок $\text{cov}(\hat{G})$. При равенстве дисперсий всех отсчетов выходного сигнала и их некоррелированности ковариационная матрица $\text{cov}(\hat{G})$ равна

$$\text{cov}(\hat{G}) = \sigma_n^2 C^{-1}, \quad (8)$$

где C^{-1} – матрица, обратная информационной матрице $C = X^T X$; σ_n^2 – дисперсия помехи.

Тогда дисперсии $\sigma_{\hat{g}_j}^2$ оценок $\hat{g}_j$, $j = 0, 1, 2, \dots, m$ вектора $\hat{G}$ и коэффициенты корреляции r_{jg} между оценками $\hat{g}_j$ и $\hat{g}_g$ будут определяться соотношениями:

$$\sigma_{\hat{g}_j}^2 = \sigma_n^2 c_{jj}^{-1}; \quad (9)$$

$$r_{jg} = \frac{c_{jg}^{-1}}{\sqrt{c_{jj}^{-1} c_{gg}^{-1}}}, \quad (10)$$

где c_{jj}^{-1} , c_{jg}^{-1} , c_{gg}^{-1} – элементы матрицы C^{-1} .

Дисперсии оценок значений усредненной хроматограммы $\sigma_{\hat{g}_j}^2$, $j = 0, 1, 2, \dots, m$ зависят не только от дисперсии помехи σ_n^2 , но и от вида выбранных последовательностей вводов проб (регулярная, псевдослучайная), их длительностей, размещения во времени и т. д.

Хроматография связана с активным экспериментом, поэтому может быть поставлена задача планирования эксперимента как задача выбора входной последовательности, обеспечивающей наибольшую точность оценивания всей совокупности ординат $\hat{g}_j$, $j = 0, 1, 2, \dots, m$. В качестве показателя, интегрально характеризующего точность оценивания этой совокупности, можно выбрать определитель ковариационной матрицы $\det \text{cov}(\hat{G})$. Тогда выбор оптимальной входной последовательности может быть осуществлен исходя из условия минимизации определителя $\det \text{cov}(\hat{G})$ (критерий D -оптимальности) или, что то же самое, из условия максимизации определителя $\det C$. В статистическом смысле D -оптимальность обеспечивает минимум обобщенной дисперсии всех оценок $\hat{g}_j$, $j = 0, 1, 2, \dots, m$ [13].

Входные последовательности ввода проб. Генерирование последовательности Лежандра. Если в некоторый момент l дискретного времени осуществляется ввод пробы, то соответствующее значение x_l принимается равным единице ($x_l = 1$); если нет, то $x_l = 0$. Таким образом, входной сигнал представляет собой последовательность единиц и нулей, соответствующих вводу или отсутствию ввода пробы в данный момент времени. При этом длина последовательности L – это общее количество элементов (нулевых и единичных). Она должна быть равна или незначительно превышать длительность хроматограммы от однократного ввода пробы [7].

Входные последовательности должны позволять гибко изменять их длину, этому требованию удовлетворяют последовательности с одноуровневой корре-

ляционной функцией [14], в частности последовательности Лежандра. Период последовательности Лежандра определяется соотношением $L = 4t + 3$, где L – простое число, t – положительное целое число. Допустимые значения длин L последовательности Лежандра для $L < 200$: 11, 19, 23, 31, 43, 47, 55, 67, 71, 79, 83, 103, 107, 127, 131, 139, 151, 163, 167, 179, 191, 199.

В отличие от M -последовательностей для последовательности Лежандра не существует генератора, легко реализуемого на сдвиговых регистрах и логических элементах, и она требует более специфических вычислений [15].

Пусть задана длина последовательности L . Обозначим через $\{a_l\} = a_0, a_1, a_2, \dots, a_{L-1}$ искомую последовательность Лежандра. Символ $a_l = 1$, если для соответствующего $l = 1, 2, \dots, L-1$ выполняется тождество

$$z^2 = l_z \pmod{L}, \quad (11)$$

где z принимает все возможные значения от 1 до $L-1$; все остальные символы $a_l = 0$, $a_0 = 1$.

Таким образом, синтез последовательности сводится к тому, что, постепенно увеличивая значения z от 1 до $L-1$, находят номера l , удовлетворяющие соотношению (11) (повтор номера не учитывается). Общее количество найденных единичных символов будет равно $p = \frac{L+1}{2}$. Оно определяет общее число вводов

пробы. Конкретные примеры последовательностей Лежандра разной длины приведены в таблице.

В общем случае произвольная входная последовательность может быть задана на интервале $(-m, k)$, причем $k > 0$, $k \geq m$. В зависимости от конкретных значений параметров k , m матрица X будет иметь различный вид.

Анализ входной последовательности относительно интервала усреднения $(0 \div k)$ позволяет выделить вариант, когда входная последовательность $\{x\}$ содержит две части: основную – длительностью L , начинающуюся в момент времени $l=0$ и заканчивающуюся в момент времени $k=L-1$, и предысторию, представляющую собой фрагмент основной части, имеющей длину m , значения которой удовлетворяют условию: $x_{-l} = x_{k-l+1}$; $l=1, 2, \dots, m$. Это позволяет сформировать симметричную теплицевую матрицу $C = X^T X$, что упрощает анализ информационной матрицы и алгоритм вычисления усредненного сигнала.

Результаты анализа псевдослучайных входных воздействий

Входная последовательность*	p	L	$m+1$	c_{jj}^{-1}	$\det C$
$1^5 0 1 0 1^2 0^2 1^2 0 1 0 1^4$	12	23	23	0,1597	$1,8954 \cdot 10^{19}$
$1^3 0 1^2 0^2 1^4 0 1 0^2 1 0 1^6 0 1^2 0^4 1^2 0^3 1 0 1 0 1 0 1 0$ $1^3 0^2 1^4 0^2 1 0^6 1 0 1^2 0 1 0^4 1^2 0^2 1 0^2$	40	79	79	0,0494	$4.8357 \cdot 10^{104}$
$1^2 0 1^3 0^3 1 0^2 1^6 0 1^2 0 1 0^2 1 0 1 0 1 0 1 0^4 1 0^2 1 0^2$ $1^2 0 1^5 0 1^2 0^3 1^3 0^2 1^5 0 1 0^3 1^4 0^2 1^4 0 1 0 1^3 0^3 1 0$ $1 0^4 1^2 0^4 1^3 0 1 0^5 1^2 0^3 1 0^2 1^3 0^2 1 0^5 1 0^2 1 0 1^2$ $0 1^4 0 1 0 1 0 1 0 1^2 0 1 0^2 1 0^6 1^2 0 1^3 0^3 1 0$	90	179	179	0,0221	$1.5157 \cdot 10^{298}$

*Краткая запись последовательности со степенями, показывающими количество следующих подряд единичных или нулевых значений.

Как следует из соотношения (9), дисперсия $\sigma_{\hat{g}_j}^2$ оценки любой ординаты $\hat{g}_j$, $j = 0, 1, \dots, m$ одна и та же и равна

$$\sigma_{\hat{g}_j}^2 = c_{jj}^{-1} \sigma_n^2.$$

Значения коэффициентов c_{jj}^{-1} зависят от длины последовательности L и длительности хроматограммы $(m+1)$. Для конкретных псевдослучайных последовательностей ввода проб значения коэффициентов c_{jj}^{-1} представлены в таблице. При сравнении СКО оценок $\sigma_{\hat{g}_j}$ и помехи σ_n коэффициент $\sqrt{c_{jj}^{-1}}$ можно рассматривать как коэффициент подавления помехи. Например, при $L = (m+1) = 23, 79, 179$ значения этого коэффициента равны соответственно $\sim 0,4$; $\sim 0,22$; $\sim 0,15$.

Вычисление усредненной дифференциальной хроматограммы. Пусть имеются два образца А и В, каждый из которых является смесью некоторых компонентов. Интересует различие компонентного состава этих образцов между собой. Такая задача возникает, например, при необходимости выявления маркерных соединений для идентификации лекарственного сырья.

При получении лекарственного сырья предпочтение отдают определенным сортам. Изучение компонентного состава эфирного масла растений методом газовой хроматографии может подтвердить или опровергнуть принадлежность образца к искомому виду и, соответственно, определить возможность его использования в качестве источника для получения лекарственного растительного сырья [16].

Представим приборные функции (хроматограммы от однократного ввода проб) первого и второго образцов соответственно как векторы отсчетов G_1 и G_2 .

Для дифференциальной хроматограммы входные последовательности, управляющие вводом проб соответственно первого и второго образцов, являются дополнительными: если проба первого образца в данный момент времени не вводится, то вводится проба второго образца, и наоборот.

Рассмотрим алгоритм вычисления усредненной дифференциальной хроматограммы. В качестве входной последовательности используется псевдослучайная последовательность Лежандра с периодом L , выбираемым из условия

$$L \geq \max \{ (m_1 + 1), (m_2 + 1) \},$$

где $(m_1 + 1)$, $(m_2 + 1)$ – длительности хроматограмм первого и второго образцов при их однократном вводе.

Для формирования последовательности вводов проб первого образца используется последовательность $\{x\}$, содержащая основную часть x_l , $l = 0, 1, \dots, k$ длительностью L и предысторию $x_{-l} = x_{k-l+1}$, $l = 1, 2, \dots, m$. На основе последовательности $\{x\}$ формируется дополнительная последовательность $\{1 - x_l\}$, которая служит для управления вводом второй пробы.

С учетом всех введенных проб первого и второго образцов можно записать матричное уравнение

$$Y = XG_1 + (T - X)G_2. \quad (12)$$

Здесь G_1 и G_2 – отсчеты хроматограмм соответственно первого и второго образцов; X – матрица, сформированная из входной последовательности $\{x\}$; T – матрица соответствующей размерности, все элементы которой равны 1.

Из соотношения (12)

$$Y = \bar{X}(G_1 - G_2) + TG_2,$$

где матрица $\bar{X}$ соответствует матрице X , в которой единичные элементы сохраняются, а нулевые элементы заменяются на (-1) . Следовательно,

$$(G_1 - G_2) = (\bar{X}^T \bar{X})^{-1} \bar{X}^T [Y - TG_2] \text{ или } (G_1 - G_2) + B = (\bar{X}^T \bar{X})^{-1} \bar{X}^T Y,$$

где B – это вектор-столбец, все значения которого одинаковы и равны $\sum_{j=0}^k g_j$.

Таким образом, получаем оценку разностной хроматограммы $(G_1 - G_2)$, которая будет приподнята или опущена на некоторый постоянный уровень B . Этот уровень возникает за счет того, что число введенных проб одного из образцов при использовании многократного ввода проб на единицу больше числа вводов другого образца. Он определяется на этапе настройки алгоритма вычислений для конкретного эксперимента.

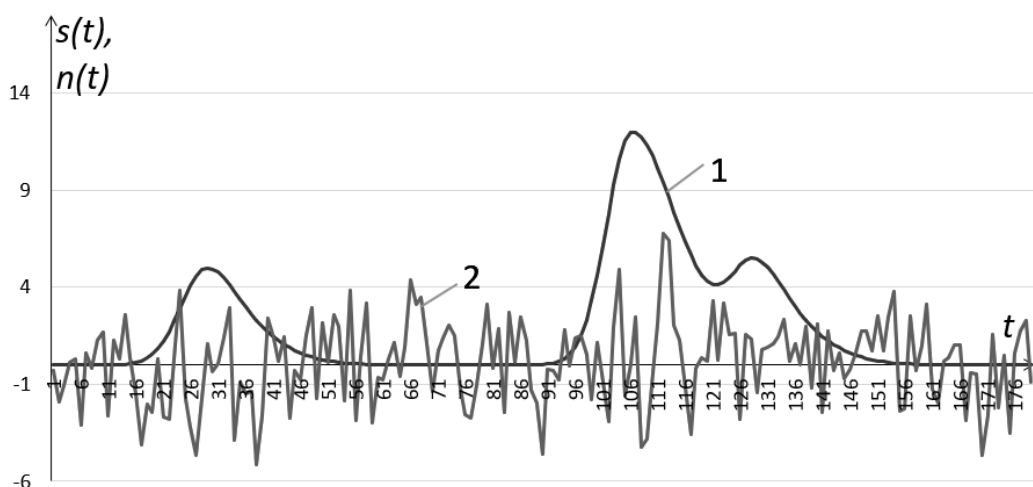
Моделирование. Для проверки работы алгоритма корреляционной обработки хроматографического сигнала, полученного при многократном вводе проб одного образца и двух образцов при сравнении их компонентного состава, проведены численные имитационные эксперименты. Методика проведения эксперимента заключается в моделировании заданных приборных функций хроматографического сигнала при многократном вводе проб одного или двух образцов с наложенной помехой, генерировании соответствующих входных последовательностей, корреляционной обработке данных и сопоставлении результатов обработки с заданной приборной функцией.

Для реализации методики разработан комплекс программ компьютерного моделирования в среде MATLAB. Имитационное моделирование проводилось для разных входных последовательностей $\{x\}$, видов приборной функции $g(t)$ и уровня помехи $n(t)$. Некоторые из результатов моделирования приведены ниже.

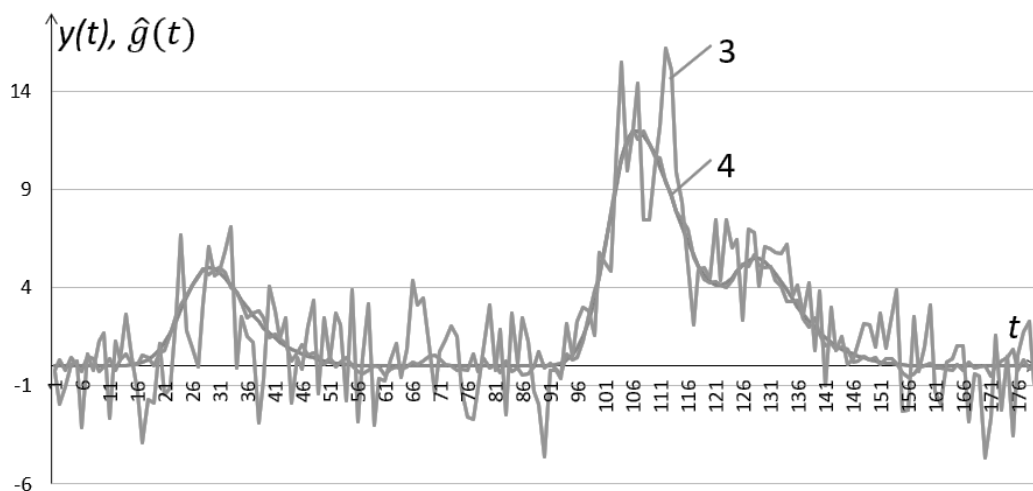
На рис. 1 кривая 4 представляет результат оценки усредненного сигнала на множестве вводов проб. Ввод проб производился с использованием псевдослучайной последовательности Лежандра длительностью 179 отсчетов. Число вводов проб – 90.

Зашумленный сигнал от однократного ввода пробы представлен кривой 3. Кривая 1 соответствует полезной составляющей сигнала с пиками несимметричной формы. В качестве математической модели несимметричных пиков использована комбинированная функция, предложенная в работе [17]. Кривая 2 соответствует высокочастотной помехе, для генерирования которой использовалась стандартная программа MATLAB: $n(t) \sim N(0, \sigma_n^2)$, СКО помехи $\sigma_n = 2$.

Сопоставление сигнала 3 с результатом корреляционной обработки (кривая 4) показывает существенное уменьшение помехи (коэффициент подавления помехи $\sim 0,15$).

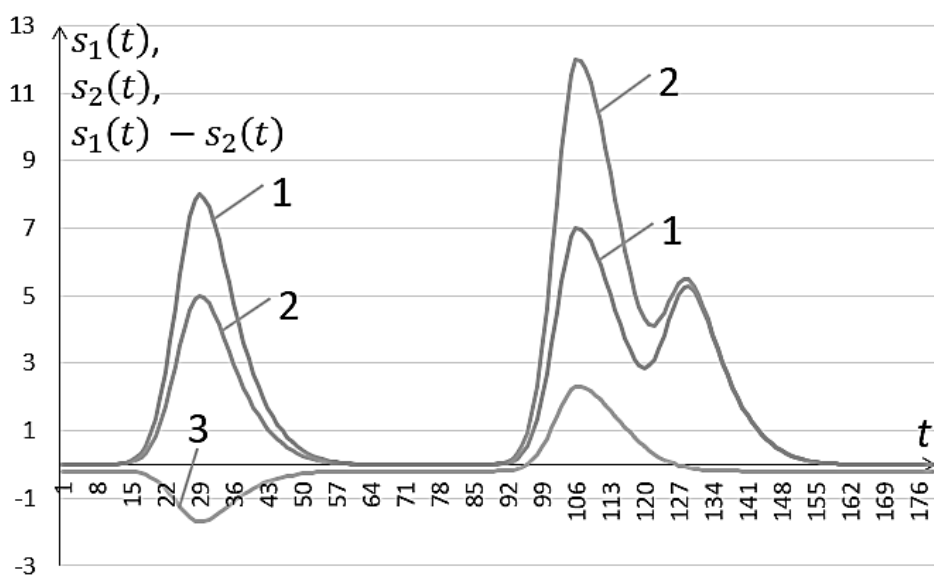


a

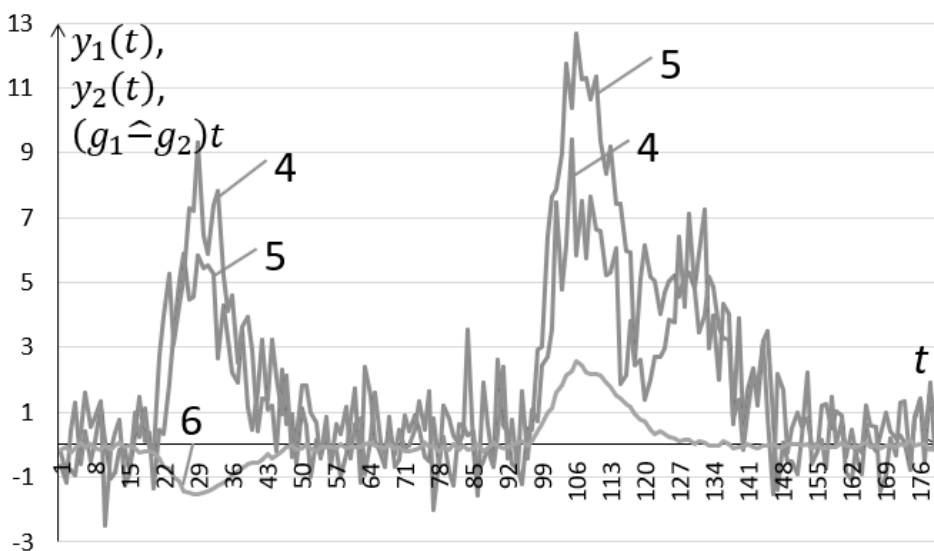


б

Рис. 1. Корреляционная обработка хроматограмм: а – сигнал от однократного ввода пробы: 1 – $s(t)$ без наложенной помехи; 2 – высокочастотная помеха $n(t)$ ($\sigma_n = 2$); б – 3 – сигнал $y(t)$ с наложенной помехой при однократном вводе пробы; 4 – усредненный хроматографический сигнал $\hat{g}(t)$ на множестве вводов проб: $p = 90$



a



б

Рис. 2. Дифференциальная хроматограмма: а – 1 – сигнал без помехи $s_1(t)$ от однократного ввода пробы первого образца; 2 – сигнал без помехи $s_2(t)$ от однократного ввода пробы второго образца; 3 – разностный сигнал $s_1(t) - s_2(t)$; б – 4 – сигнал $y_1(t)$ с наложенной помехой ($\sigma_n = 2$) от однократного ввода пробы первого образца; 5 – сигнал $y_2(t)$ с наложенной помехой ($\sigma_n = 2$) от однократного ввода пробы второго образца; 6 – усредненный разностный сигнал $(g_1 - g_2)t$: $p = 90$ для первого образца, $p = 89$ для второго образца

На рис. 2 представлена дифференциальная хроматограмма (кривая 6). Кривые 1, 2 – хроматограммы без помехи от однократного ввода проб первого и вто-

рого образцов. Кривые 4, 5 – сигналы с наложенной помехой первого и второго образцов при их однократном вводе. Кривая 6 соответствует усредненному разностному сигналу, когда производится 90-кратный ввод проб первого образца и 89-кратный ввод проб второго образца.

Простое сравнение зашумленных хроматограмм первого и второго образцов (кривые 4 и 5 на рис. 2) не позволяет оценить различие компонентного состава этих образцов, однако это можно сделать по усредненному разностному сигналу (кривая 6 на рис. 2).

В целом результаты моделирования подтвердили эффективность использования корреляционной обработки данных для вычисления дифференциальной хроматограммы и правильность предложенных алгоритмов оценок усредненного хроматографического сигнала. Корреляционная обработка позволяет значительно уменьшить помеху, тем самым понизить предел обнаружения компонентов анализируемой пробы и расширить диапазон измеряемых концентраций.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Платонов И.А., Новикова Е.А., Платонов В.И. Хроматографические методы анализа: учебное пособие. Самара: Самарский университет, 2021. 96 с. EDN: JXKOWU.
2. Гуревич А.Л., Русинов Л.А., Сягаев Н.А. Автоматический хроматографический анализ. Л.: Химия, 1980. 192 с.
3. Русинов Л.А. Автоматизация аналитических систем определения состава и качества веществ. Л.: Химия, 1984. 160 с.
4. Эштейн Н.А. О допустимых значениях порога игнорирования примесей и отношение сигнал/шум при проверке чувствительности хроматографической системы // Ведомости НЦЭСМП. 2014. Т. 7, № 2. С. 85–90.
5. Каламбет Ю., Мальцев С., Козьмин Ю. Фильтрация шумов: окончательное решение проблемы // Аналитика. 2011. № 1. С. 50–55. EDN: PIEIFT.
6. Кальюранд М. Анализ нестационарных потоков вещества: возможности компьютерной хроматографии // Химия. 1989. № 28. С. 25–33. DOI: 10.3176/chem.1989.1.06.
7. Филаретов Г.Ф., Сайфуллин Ж.Т. Корреляционный метод измерений в системах автоматизации хроматографического эксперимента // Заводская лаборатория. 1991. № 10. С. 59–63.
8. Сайфуллин Р.Т. Использование корреляционного метода для усреднения хроматограмм // Вестник Самарского государственного технического университета. Серия: Технические науки. 2009. Вып. № 3 (25). С. 109–116. EDN: LALEMB.
9. Вуколов Э.А. Регрессионный анализ. М.: МИЭТ, 2000. 52 с.
10. Гроп Д. Методы идентификации систем. М.: Мир, 1979. 302 с.
11. Эйхгофф П., Ванечек А., Савараги Е. и др. Современные методы идентификации систем. М.: Мир, 1983. 400 с.
12. Мелас В.Б., Шпилев П.В. Планирование и анализ регрессионных моделей. СПб: СПбГУ, 2014. 94 с.
13. Маслов Г.Г., Трубилин Е.И., Цибульский В.В. Моделирование в агроинженерии. Краснодар: КубГАУ, 2019. 304 с.
14. Дейч А.М. Методы идентификации динамических систем. М.: Энергия, 1979. 240 с.
15. Сайфуллин Р.Т. Ввод пробы по случайному закону в корреляционной хроматографии // Информационно-измерительные и управляющие системы: сб. науч. ст. Самара: СамГТУ, 2009. С. 80–84.
16. Маланкина Е.Л., Козловская Л.Н., Кузьменко А.Н., Евграфов А.А. Определение компонентного состава эфирного масла видов тимьяна методом газовой хроматографии // Вестник Моск. ун-та. Серия 2: Химия. 2019. Т. 60, № 6. С. 411–415. EDN: OXNJTQ.
17. Бочкарев А.В. Повышение точности количественного хроматографического анализа с помощью новой модели пика // Автоматизация в промышленности. 2022. № 7. С. 36–41. DOI: 10.25728/avtprom.2022.07.11. EDN: EPPYCW.

CORRELATION DATA PROCESSING FOR DETERMINING THE AVERAGED DIFFERENTIAL CHROMATOGRAM IN THE CASE OF MULTIPLE SAMPLE INJECTIONS

M.Zh. Sayfulin¹

Samara State Technical University
244, Molodogvardeyskaya str., Samara, 443100, Russian Federation

E-mail: sayfulinmarsel@yandex.ru

Abstract. *The relationship between the input and output signals of a chromatograph is demonstrated. The expediency of improving the signal-to-noise ratio and using correlation data processing for this purpose is noted. An algorithm for calculating a chromatogram averaged over multiple injections of a sample from a specific individual specimen is presented. This algorithm is then modified for the case of calculating a differential averaged chromatogram, which is required to determine the difference in the component composition of two different specimens. A Legendre pseudorandom sequence is used to generate the input signals. The alphabet of this sequence is 1 or 0: 1 indicates a sample injection, and 0 indicates the absence of a sample injection at a given time. In the case of analyzing two specimens, the input sequences controlling the injection of the first and second specimens are complementary: if the sample of the first specimen is not injected, the sample of the second specimen is injected, and vice versa. The algorithm for calculating the averaged chromatogram is reduced to regression analysis. An information matrix of the experimental design is introduced. The accuracy of estimation is characterized using the covariance matrix of estimates. The choice of the optimal input sequence for sample injection is based on the D-optimality condition (the condition for maximizing the determinant of the information matrix), which ensures the minimum generalized variance of all estimates of the ordinates of the averaged chromatogram. The results of a numerical simulation experiment are presented, proving the effectiveness of using the correlation data processing method for averaging chromatograms.*

Keywords: *correlation data processing, multiple sample injection, signal-to-noise ratio, estimation accuracy, differential chromatogram.*

REFERENCES

1. Platonov I.A., Novikova E.A., Platonov V.I. Hromatograficheskie metody analiza [Chromatographic Methods of Analysis]: uchebnoe posobie. Samara: Samarskij universitet, 2021. 96 p. (In Russian)
2. Gurevich A.L., Rusinov L.A., Syagaev N.A. Avtomaticheskij hromatograficheskij analiz [Automatic Chromatographic Analysis]. L.: Himiya, 1980. 192 p. (In Russian)
3. Rusinov L.A. Avtomatizaciya analiticheskikh sistem opredeleniya sostava i kachestva veshchestv [Automation of Analytical Systems for Determining the Composition and Quality of Substances]. L.: Himiya, 1984. 160 p. (In Russian)
4. Epshtejn N.A. O dopustimyh znacheniyah poroga ignorirovaniya primesej i otnoshenie signal/shum pri proverke chuvstvitelnosti hromatograficheskoy sistemy [On the Permissible Values of the Impurity Ignoring Threshold and the Signal-to-Noise Ratio in Testing the Sensitivity of a Chromatographic System] // Vedomosti NCESMP. 2014. Vol. 7, № 2. P. 85–90. (In Russian)
5. Kalambet Yu., Malcev S., Kozmin Yu. Filtraciya шумов: okonchatelnoe reshenie problemy [Noise Filtering: The Ultimate Solution to the Problem] // Analitika. 2011. № 1. P. 50–55. (In Russian)



© The Author(s), 2024

¹ Marsel Zh. Sayfulin, Postgraduate Student.

6. *Kalyurand M.* Analiz nestacionarnykh potokov veshchestva: vozmozhnosti kompyuternoy hromatografii [Analysis of Non-Stationary Substance Flows: Capabilities of Computer Chromatography] // *Himiya*. 1989. № 28. P. 25–33. DOI: 10.3176/chem.1989.1.06. (In Russian)
7. *Filaretov G.F., Sajfullin Zh.T.* Korrelyatsionnyy metod izmereniy v sistemah avto-matizatsii hromatograficheskogo eksperimenta [Correlation Method of Measurements in Systems for Automating Chromatographic Experiments] // *Zavodskaya laboratoriya*. 1991. № 10. P. 59–63. (In Russian)
8. *Sajfullin R.T.* Ispolzovanie korrelyatsionnogo metoda dlya usredneniya hromatogramm [Application of the Correlation Method for Averaging Chromatograms] // *Vestnik Samarskogo gos. tehn. un-ta. Seriya: Tehnicheskie nauki*. 2009. Vyp. № 3 (25). P. 109–116. (In Russian)
9. *Vukolov E.A.* Regressionnyy analiz [Regression Analysis]. M.: MIET, 2000. 52 p. (In Russian)
10. *Grop D.* Methods of System Identification. M.: Mir, 1979. 302 p.
11. *Ejkhoff P., Vanechek A., Savaragi E. i dr.* Sovremennye metody identifikatsii sistem [Modern Methods of System Identification]. M.: Mir, 1983. 400 p. (In Russian)
12. *Melas V.B., Shpilev P.V.* Planirovanie i analiz regressionnykh modelej [Planning and analysis of regression models]. SPb: SPbGU, 2014. 94 p. (In Russian)
13. *Maslov G.G., Trubilin E.I., Cibulskiy V.V.* Modelirovanie v agroinzhenerii [Modeling in Agricultural Engineering]. Krasnodar: KubGAU, 2019. 304 p. (In Russian)
14. *Deutsch A.M.* Methods of Dynamic Systems Identification. M.: Energiya, 1979. 240 p.
15. *Sajfullin R.T.* Vvod probe po sluchajnomu zakonu v korrelyatsionnoy hromatografii [Introduction of a sample according to a random law in correlation chromatography] // *Informacionno-izmeritelnye i upravlyayushie sistemy: sb. nauch. st.* Samara: SamGTU, 2009. P. 80–84. (In Russian)
16. *Malankina E.L., Kozlovskaya L.N., Kuzmenko A.N., Evgrafov A.A.* Opredelenie komponentnogo sostava efirnogo masla vidov timyana metodom gazovoy hromatografii [Determination of the component composition of essential oils from thyme species using gas chromatography] // *Vestnik Mosk. un-ta. Seriya 2: Himiya*. 2019. Vol. 60, № 6. P. 411–415. (In Russian)
17. *Bochkarev A.V.* Povyshenie tochnosti kolichestvennogo hromatograficheskogo analiza s pomoshchyu novoy modeli pika [Improving the accuracy of quantitative chromatographic analysis using a new peak model] // *Avtomatizatsiya v promyshlennosti*. 2022. № 7. P. 36–41. (In Russian)

Original article submitted 03.06.2024