



## ОЦЕНКА ВЕЛИЧИНЫ ШУМА И ЕЕ ИСПОЛЬЗОВАНИЕ ПРИ ОБРАБОТКЕ ХРОМАТОГРАФИЧЕСКОГО СИГНАЛА

Описан алгоритм оценки уровня шума произвольного хроматографического сигнала (в том числе и в присутствии пиков), используемый при обработке хроматографических данных программой «МультиХром». Рассмотрена эффективность использования специализированных алгоритмов обработки сигнала для уменьшения влияния шумов на погрешность измерения.

**Ю. А. Каламбет**

к.ф.-м.н., генеральный директор  
ЗАО «Амперсанд»

**К. В. Михайлова**

инженер ЗАО «Амперсанд»

**ЗАО «Амперсанд», Москва**

support@ampersand.ru

Одним из важнейших параметров процесса измерения, определяющих его метрологические характеристики, является уровень шумов. В хроматографическом процессе именно шумами определяется предельное значение такого важного параметра, как предел обнаружения. Значительный вклад вносят шумы, наряду с погрешностями ввода пробы, и в суммарную погрешность измерения концентрации, особенно при низком уровне содержания компонентов.

В качестве меры уровня шумов в хроматографии традиционно используется *размах*  $2\Delta X$  (*peak-to-peak* – «пик к пику») – разность максимального и минимального измерения уровня сигнала базовой линии за заданное время измерения, рассчитанная за вычетом *дрейфа* (рис. 1). Выбор этой характеристики связан с простотой ее измерения при регистрации сигнала с помощью интегратора и самописца [1].

Переход к приему сигнала с помощью аналого-цифровых преобразователей (АЦП) с последующей обработкой его с помощью компьютерной техники позволяет рассчитывать другую характеристику шумового процесса – среднее квадратичное отклонение  $\sigma X$  (СКО). При автоматической обработке сигнала использование этого параметра предпочтительно, так как его величина малочувствительна к одиночным выбросам большой амплитуды, что было особенно важно на ранних этапах использования АЦП, когда появление артефактов подобного рода имело довольно высокую вероятность (для современных АЦП этой проблемы практически не существует).

### Расчет шумовых характеристик в программе «МультиХром»

В отсутствие хроматографических пиков расчет СКО для базовой линии представляет собой элементарную задачу. При наличии пиков формально рассчитанная величина СКО будет много больше величины шума из-за вклада измерений, соответствующих пикам. В программе «МультиХром» используется оригинальная процедура оценки величины шума, позволяющая исключить пики. Она состоит в следующем. Рассматриваются значения разности сигналов (приращение) для двух соседних точек хроматограммы. Если три идущие подряд приращения имеют один и тот же знак, это считается признаком того, что точки находятся на склоне пика. Такие приращения из дальнейшего рассмотрения исключаются. Далее рассчитываются параметры распределения для оставшихся приращений. Поскольку в их числе могут оказаться величины, относящиеся-

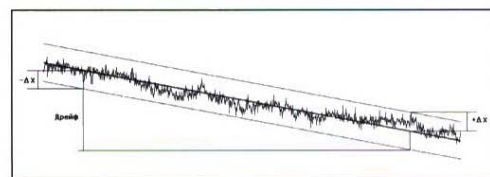


Рис. 1 Шумы и дрейф базовой линии

Автоматизация,  
хроматография, погрешность  
измерения, шум хроматогра-  
фического сигнала,  
«МультиХром»

ся к вершинам пиков, а также случайные одно- или двухточечные выбросы большой амплитуды, далее исключаются все приращения, отклонение которых от среднего значения превышает 5-кратную величину среднеквадратичного отклонения для распределения приращений. Такая процедура пересчета с отбраковкой выпадающих приращений повторяется до тех пор, пока их число не стабилизируется.

Расчет величины шума производится автоматически по окончании хроматограммы, поскольку в программе «МультиХром» эта процедура выполняется, в первую очередь, для решения «внутренней» задачи: его величина используется при определении точек начала и конца хроматографического пика. Алгоритм детектирования пиков основан на определении первой производной (наклона) хроматографической кривой и сравнении его с уровнем шума. Для того чтобы решить, является ли наклон в некоторой точке значимым, величина первой производной делится на значение шума базовой линии. Если полученная величина превышает величину порога, заданного пользователем, точка считается началом пика. Конец пика определяется в точке, в которой указанное отношение становится меньше порога (величины порога для начала и конца пика могут не совпадать).

Дополнительный вопрос возникает при разметке многоканальных хроматограмм: какой канал выбрать для этой цели, если ни в одном из них не представлены все пики? В этом случае в программе «МультиХром» создается синтетический суммарный канал, причем суммируемые сигналы предварительно нормируются на величину шума канала. Такая процедура позволяет получить канал, содержащий все пики, и одновременно избежать ухудшения отношения *сигнал/шум* [2].

По окончании приема хроматограммы пользователь имеет возможность увидеть величину шума в одном из окон. Значения собственно СКО, а также размаха сигнала могут выводиться при создании отчета дополнительно. Для сигнала без пиков измеренная с использованием описанной процедуры величина шума будет мало отличаться от рассчитанной для него величины СКО, а соотношение между СКО и размахом составит приблизительно  $2\Delta X = 6\sigma X$  (для «белого» шума при количестве точек изме-

рения  $10^3-10^4$ ). Если в соответствии с требованиями методик или стандартов необходимо использовать для вычисления тех или иных хроматографических параметров величину размаха, можно рекомендовать использовать его расчетное значение по величине автоматически определяемого шума, однако предварительно необходимо убедиться в выполнении указанного соотношения на хроматограммах без пиков.

### Источники шумов

Шумовой сигнал, который видит пользователь в окне хроматограммы, как правило, совместно создают АЦП, наводки на соединительные провода, детектор, поток элюента. Можно экспериментально оценить вклад каждого источника, считая их независимыми, и последовательно измеряя вначале шум одного АЦП с замкнутыми контактами на входе; затем АЦП вместе с кабелями, подсоединенными к выключенному детектору; далее — то же, но с включенным детектором без потока элюента, и, наконец, шум при реальных условиях регистрации хроматограммы. В случае недопустимо высокого уровня шумов подобный анализ поможет правильно определить наиболее слабое звено всего измерительного тракта.

Как правило, вклад современного АЦП в суммарный шум является наименьшим. Для используемого в системе «МультиХром» 24-разрядного выносного модуля Е-24 он составляет менее 1,3 мкВ, или 2-3 младших двоичных разряда (при скорости сбора данных 10 Пц). При большей скорости приема данных уровень шума несколько возрастает (при 50 Пц эффективная разрядность АЦП соответствует 19).

Для уменьшения вклада наводок АЦП рекомендуется устанавливать в непосредственной близости от хроматографа, так чтобы длина кабелей, соединяющих детекторы с входами АЦП, не превышала 3-5 м. Самое пристальное внимание необходимо уделять вопросам заземления — предпочтительно иметь отдельное заземление корпусов хроматографа и компьютера непосредственным подсоединением к одной и той же заземляющей шине.

При соблюдении указанных требований основные источники шумов, как правило, оказываются вне системы «МультиХром», поэтому далее в настоящей публикации не рассматриваются.

### Использование специализированных алгоритмов обработки сигнала — наиболее эффективный метод борьбы с шумами

Традиционно для уменьшения влияния шумов используются различные методы фильтрации. Некоторые такие фильтры были включены в программу «МультиХром». Но развитие специализированных алгоритмов обработки сигнала, применяемых на отдельных участках хроматограммы привело к тому, что выигрыш от предварительной фильтрации сигнала практически перестал давать дополнительный выигрыш в сглаживании шумов. Однако само рассмотрение достоинств и недостатков фильтров шумов в контексте современных методов обработки хроматографического сигнала представляется весьма поучительным.

В дальнейшем изложении используется определение «точка хроматограммы», — под этим следует понимать пару чисел (отклик по ординате, время измерения).

Фильтрация шумов (сглаживание) может рассматриваться как применение к каждой точке хроматограммы некоторой процедуры, заменяющей значение ординаты на новое, предположительно более точно соответствующее действительному значению сигнала во время измерения. Перечислим наиболее популярные процедуры:

*Медианная фильтрация*

*Фильтрация одиночных выбросов*

Замена центральной точки окна на сумму соседних по времени точек нефильтрованного массива с некоторыми коэффициентами. В частности, *метод скользящего среднего, Гауссово сглаживание, метод Савицкого-Голея* [3].

Кратко опишем каждый метод, обращая внимание на его достоинства и недостатки.

При *медианной фильтрации* значения внутри окна сортируются в порядке возрастания. Отклик, соответствующий середине окна, заменяется другим значением, попадающим в центр отсортированного массива. Этот метод влияет на хроматографические пики в наименьшей степени, хорошо сглаживает базовую линию, не меняет форму пика на склонах и очень эффективно устраняет отдельные выбросы (в этом случае вы-



брос заменяется на одну из соседних точек). Однако он слегка «приглаживает» вершины пиков и ложбины между пиками и может изменять как высоту, так и площадь хроматографических пиков.

**Фильтрация одиночных выбросов** — уникальный метод фильтрации, основанный на знании шумовых параметров сигнала. Применяется в программе «МультиХром» для выявления только точек с ошибочной ординатой (выбросов). Имеет скорее историческую ценность, поскольку для современных АЦП причины выбросов — сбои измерения и передачи данных — редки. После применения этого метода с небольшой вероятностью может измениться как площадь, так и высота пика, однако в подавляющем большинстве случаев ничего не изменяется.

**Метод скользящего среднего** состоит в замене точки на среднее значение в окне. В программе «МультиХром» применяется при вычислении производной в процессе разметки хроматограммы на пики и, возможно (в зависимости от предпочтений пользователя), при отображении хроматограммы на экране. Отличается очень высокой скоростью выполнения алгоритма, пропорциональной только числу точек в хроматограмме. Площадь пика сохраняется, высота изменяется.

**Гауссово сглаживание** состоит в замене значения на сумму точек окна с весами, распределенными по Гауссу. Площадь пика сохраняется, высота и полуширина изменяются.

**Метод Савицкого-Голея** состоит в замене ординаты точки на значение, полученное путем аппроксимации хроматографической зависимости в пределах окна полиномом третьей степени. В большинстве случаев не должны изменяться ни высота, ни полуширина, ни площадь хроматографического пика. Возможны артефакты: при избыточной ширине окна интерполяции возникают характерные провалы перед началом пика и, реже, после окончания. При отсутствии артефактов — весьма эффективный метод, артефакты могут приводить к существенному изменению как площади, так и высоты пика.

После знакомства с методами фильтрации шумов зададимся вопросом: для чего фильтруются шумы?

Чаще всего экспериментаторы дают следующие обоснование.

1. Облегчение работы алгоритмов обработки данных (поиска пиков).
2. Более точный расчет площадей, высот и полуширин пиков.
3. Улучшение визуального представления данных.
4. Исправление сбоев измерения.

Внимательно рассмотрим каждую из причин.

### 1. Облегчение работы алгоритмов обработки данных (поиска пиков).

Это так, только если алгоритм поиска пиков разработан без учета присутствия шумов. Нормальный алгоритм должен давать одинаковые результаты на исходных и фильтрованных данных.

### 2. Более точный расчет площадей, высот и полуширин пиков.

Для всех методов фильтрации существуют артефакты, поэтому корректность измерения высоты и площади может вызывать сомнения.

**Высота.** Для корректного измерения высоты пика нужно очень точно знать положение максимума пика и базовой линии, т. е. аппроксимация в большинстве точек не нужна, — она нужна только в точке максимума и в точках, определяющих положение базовой линии. При этом избежать аппроксимации даже отфильтрованных данных не удастся: максимальная высота достигается обычно не в одной из этих точек, а между ними, и вершина аппроксимирующей параболы может быть как выше, так и ниже самой высокой точки пика. Поэтому при определении высоты по отфильтрованным данным нет преимущества ни по точности, ни по сложности процедуры. Более того, фильтрация обычно ведется одинаково для всех точек и неоптимально для пиков разной ширины. В программе «МультиХром» применен адаптивный способ вычисления окна квадратичного полинома, аппроксимирующего вершину пика — это окно зависит от полуширины конкретного пика. Для более широких пиков это окно будет больше, чем для узких, что дает дополнительный выигрыш в точности.

**Площадь.** Для точного вычисления площади необходимо знать только положение базовой линии. В самом деле, поскольку сглаживающие функции конструируются так, чтобы не менять площадь пика в условиях «идеальной» базовой линии, изменение площади, подсчитанной

по сглаженным и несглаженным данным, зависит в основном от способа проведения базовой линии (подробнее см. следующий раздел). Более того, при аппроксимации точки начала базовой линии по исходным данным можно избежать артефактов, присущих методу Савицкого-Голея. Артефакты вызваны неадекватностью модели аппроксимации (полиномиальной с определенным окном) и реального характера изменения сигнала. Заметим, что "плохая" интерполяция возникает при наличии хотя бы одного из экстремумов аппроксимирующего полинома в окне интерполяции. Если начало пика аппроксимируется полиномом, обладающим таким свойством, следует проверить качество аппроксимации. Для проверки можно, к примеру, найти максимальное отклонение аппроксимирующего полинома от исходных точек и сравнить его с величиной шума. Если модель неадекватна, можно уменьшить ширину окна интерполяции или вовсе отказаться от нее.

**Полуширина.** В программе «МультиХром» полуширина вычисляется путем аппроксимации склона пика вблизи нужной точки кубическим полиномом с адаптивным окном, и выигрыша применение предварительного сглаживания данных не дает. То же касается и ширины, вычисленной по базовой линии.

### 3. Улучшение визуального представления данных.

а) На экране и принтере гораздо лучше иметь реальную картину процесса, а не «приглаженный» вариант. б) Данные, представленные на экране, в любом случае отличаются от исходных данных. По горизонтали число точек на экране обычно в десятки — сотни раз меньше числа точек в хроматограмме, разрешение же экрана по вертикали более чем в тысячу раз меньше разрешения современных АЦП. Если мы рисуем заведомо не совсем то, нам приходится регулиро-

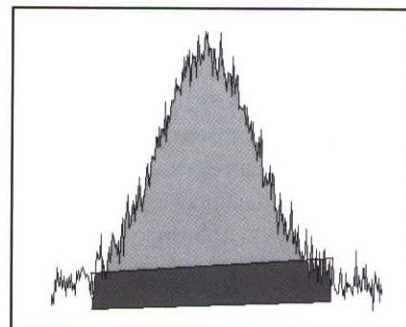


Рис. 2 Влияние проведения базовой линии на площадь пика

вать процесс формирования изображения по исходным данным, и элементы сглаживания в него бывают внесены.

#### 4. Исправление сбоев измерения.

С современными АЦП одиночные сбои измерения крайне редки.

##### Недостатки сглаживания:

Изменение исходных («сырых») данных. Это противоречит ряду законов о Хорошей Лабораторной Практике (Good Laboratory Practice, GLP). Это действительно огромный недостаток, сводящий на нет все достоинства.

**Вывод.** Лучше пользоваться несглаженными данными и алгоритмами, приспособленными для работы с ними. Отключайте фильтры шумов!

#### Использование алгоритма интерполяции базовой линии для уменьшения погрешности измерения низких концентраций

Существенный вклад в погрешность измерения параметров хроматографических пиков малой высоты вносит погрешность проведения базовой линии — это относится как к высоте пика так и, в особенности, к его площади. Значение абсциссы начала и конца пика определяется с помощью описанной выше процедуры. В качестве ординаты в простейшем случае берутся ординаты соответствующих точек хроматограммы, однако, при высоком уровне шумов (то есть, при небольшом соотношении *сигнал/шум* для вершины пика), сдвиг начала или конца пика всего на одну точку может привести к заметному изменению результата измерения площади.

В программе «МультиХром» предусмотрена возможность интерполяции участка базовой линии в окрестности точек начала и конца пика, обеспечивающая вычисление усредненных значений ординаты этих точек. Эффективность этой процедуры продемонстрирована с помощью вычислительного эксперимента, в котором для одних и тех же пиков рассчитывалась величина площади при разных значениях отношения *сигнал/шум* без интерполяции базовой линии и с ее использованием (рис. 3, рис. 4).

Расчеты проводились для трех серий хроматограмм с высоким (I), средним (II) и низким (III) значением отношения *сигнал/шум* (S/N). Каждая серия состояла из 9 хроматограмм, полученных пу-

Серия	Пик	S/N	RSD	Delta	RSD-i	Delta-i
I	B	39	2.62	10	1.38	1.1
	A	49	1.68	8	0.832	0.7
	D	19.5	2.97	16	2.95	0.02
	C	30.8	3.53	11	1.48	-0.2
II	B	13	7.43	30	4.01	0.04
	A	16.4	5.64	20	4.04	1.2
	D	6.5	15.4	33	11.4	-1.4
	C	10.3	8.90	30	5.12	0.4
III	B	5.2	10.1	53	7.24	-1
	A	7.3	6.49	43	6.52	6
	D	2.6	47.5	38	20.4	-10
	C	4.1	20.1	57	13.2	-1

Таблица 1.

тем сложения одной и той же хроматограммы с пренебрежимо низким уровнем шумов (0) с хроматограммами, содержащими записи шумовых процессов. Для каждой серии рассчитывалось среднее значение и СКО площади для каждого из 4 пиков (B, A, D, C), а также разность между средним и истинным значением площади пика, рассчитанным для пиков без шумов. Расчеты выполнялись при разметке хроматограммы без интерполяции базовой линии и с использованием этой процедуры.

Результаты расчетов представлены в

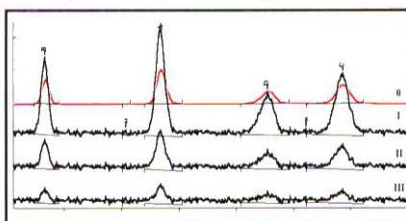


Рис. 3 Разметка хроматограмм без интерполяции базовой линии.

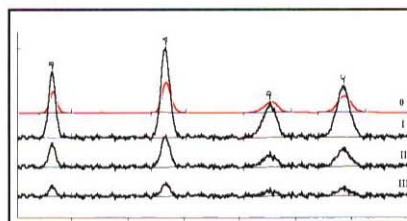


Рис. 4 Разметка хроматограмм с интерполяцией базовой линии.

#### Литература

1. IUPAC Chromatography] International Union of Pure and Applied Chemistry, Nomenclature for chromatography, L. S. Ettre, editor, Pure and Applied Chemistry 65: 819-872, 1993. 450+ definitions.
2. Yu. A. Kalambet, Yu. P. Kozmin, M. P. Perelroyzen // Computer spectrochromatography. Principles and practice of multi-channel chromatographic data processing, J. Chromatography (1991), v. 542, 247-261.
3. Savitsky, A. and Golay, M. J. E., "Smoothing and Differentiation of Data by Simplified Least Squares Procedures", Analytical Chemistry, v. 36, no. 8, July 1964, pp. 1627-1639.

таблице 1 и на рисунке 5. Используемые обозначения: S/N — отношение высоты пика к шуму; RSD и RSD-i — СКО площади пика; Delta и Delta-i — разность среднего значения измеренной и истинной площади пика (величины -i относятся к измерениям с использованием интерполяции базовой линии).

Как видно из приведенных данных, использование интерполяции базовой линии позволяет производить измерения площади пиков даже при очень низком отношении *сигнал/шум* ( $< 3$  для пика D в серии III) с погрешностью около 20%, тогда как без интерполяции она достигает почти 50%. Кроме того, без интерполяции измеренная площадь пика всегда превышает истинную, и эта систематическая погрешность для всех пиков серии составляет около 50%. ■

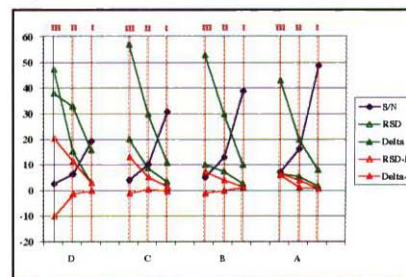


Рис. 5 Погрешности расчета площади пиков без интерполяции базовой линии и с использованием интерполяции (обозначения см. в тексте).