

# ООО Амперсенд


КОМПЬЮТЕРНАЯ АВТОМАТИЗАЦИЯ ХРОМАТОГРАФИИ



[www.multichrom.ru](http://www.multichrom.ru)

[support@ampersand.ru](mailto:support@ampersand.ru)

+7(499)196-52-90



## Сколько нужно точек, чтобы правильно измерить площадь пика?

Юрий Каламбет, ООО «Амперсенд», Москва

Юрий Козьмин, Институт биоорганической химии, Москва

Андрей Самохин, Московский Государственный Университет, Химический факультет

[Kalambet@ampersand.ru](mailto:Kalambet@ampersand.ru)

## Откуда берутся узкие пики?

- ▶ «Быстрые» сканирующие детекторы: Милихром, ГХ-МС, ЖХ-МС
- ▶ Быстрая хроматография (например, второе D в 2D-хроматографии)
- ▶ Жидкостная хроматография сверхвысокого давления (UPLC)
- ▶ Почему мы? Опыт написания программ для хроматографии 35+ лет

Появились новые технологии, продуцирующие узкие пики. ООО «Амперсенд» разрабатывает и распространяет программное обеспечение для хроматографии «МультиХром». Команда «Амперсенд»а разрабатывает программы для хроматографии с начала 1980х.

## Оценка погрешности интегрирования согласно работе Нормана Дайсона (N.Dyson, J.Chromatography A, 1999, 842, 321-340)

- ▶ Погрешность интегрирования методом трапеций может быть оценена как

$$I_{true} - I_{meas} = \left( \frac{W_b^3}{12n^2} \right) |h''(t)| = (\epsilon^3 N / 12) \cdot |f''(t)| \quad (1)$$

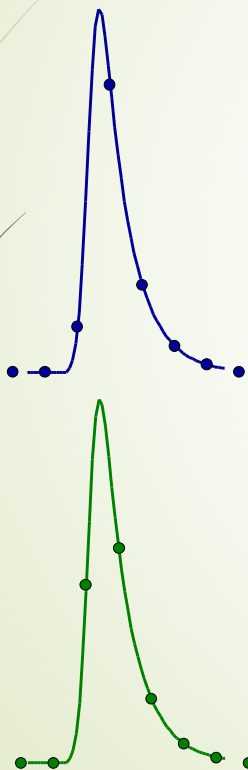
- ▶ Погрешность интегрирования по методу Симпсона меньше, чем по методу прямоугольников или трапеций
- ▶ Для надежного измерения площади требуется от 40 до 100 точек на пик
- ▶ Асимметричные ( $\tau/\sigma=3$ ) пики требуют гораздо большего числа точек, чем пики Гауссовой формы (до 2.5 раз) при одинаковой погрешности интегрирования

Норман Дайсон написал очень полезную и популярную книгу "Chromatographic Integration Methods" 1996, 1998. К сожалению, он также написал статью "Peak distortion, data sampling errors and the integrator in the measurement of very narrow chromatographic peak", некоторые утверждения из которой приведены на слайде. Все утверждения, кроме последнего, неверны.

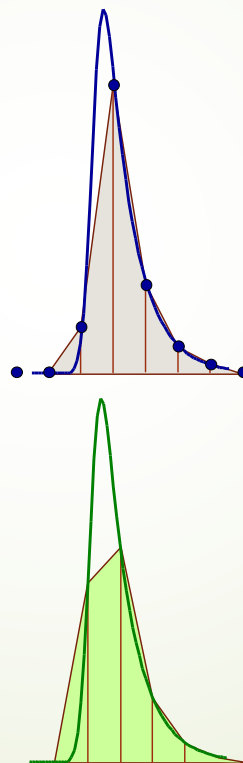
# Правила интегрирования

4

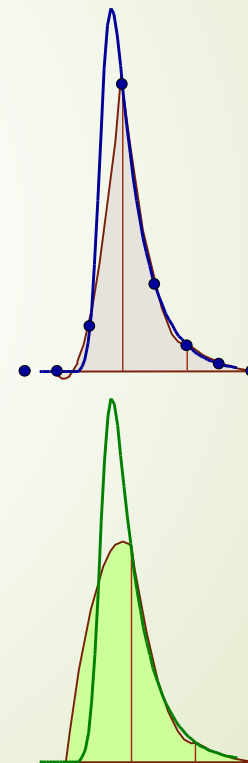
Узлы сетки оцифровки



Правило трапеций



Правило Симпсона



Графическая иллюстрация сетки оцифровки и некоторых правил интегрирования.

# Большинство схем для отдельно стоящего пика даёт один ответ

## = **правило прямоугольника**

Ниже приведены наборы весов индивидуальных точек, получающиеся при интегрировании по разным правилам

- ▶ Правило прямоугольника: [001111...111100]
- ▶ Правило трапеций: [01222...22210]/2
- ▶ Правило Симпсона, усредненное по двум вариантам начала интегрирования: 
$$\frac{([014242...242410]/3 + [0014242...242410]/3)/2 = [0156666...666510]/6}$$
- ▶ Разница между правилами интегрирования сводится к разному учету периферийных точек, значения в которых равны нулю
- ▶ Погрешность интегрирования пика НЕ МОЖЕТ зависеть от использованного правила, поскольку большинство правил дают один и тот же результат: **площадь по правилу прямоугольника**

## «Пикоподобная» функция $f(x)$

- ▶ Непрерывно и бесконечно дифференцируемая
- ▶ Функция и все ее производные равны нулю вне пределов ограниченного интервала значений аргумента  $x$
- ▶ *Свойство:* Определенный интеграл любой производной по всей области пика равен нулю. (Интеграл производной равен производной предыдущего порядка; все производные по краям пика обращаются в ноль.)
- ▶ *Хроматографический пик* – «пикоподобная» функция

Это определения некоторых свойств пика. Они потребуются позже.

# Оценка погрешности интегрирования

7

- Ряд Тейлора

$$f(x + \tau) = f(x_i) + f'(x_i) \cdot \tau + \frac{1}{2} f''(x_i) \cdot \tau^2 + \dots + \frac{1}{n!} f^{(n)}(x_i) \cdot \tau^n + \dots; \quad -\frac{\varepsilon}{2} < \tau < \frac{\varepsilon}{2}$$

- Для оценки площади  $A$  пика интегрируем ряд Тейлора в  $\varepsilon/2$ -окрестности узла и суммируем оценки всех узлов; координаты узлов  $x_i = \frac{\varepsilon}{2} + i \cdot \varepsilon$

$$A = A_0 + \Delta A_2 + \Delta A_4 + \dots$$

$$A_0 = \varepsilon \cdot \sum_{i=1}^N f(x_i)$$

$$\Delta A_2 = \frac{2}{3!} \left(\frac{\varepsilon}{2}\right)^3 \sum_{i=1}^N f''(x_i)$$

...

$$\Delta A_{2k} = \frac{2}{(2k+1)!} \left(\frac{\varepsilon}{2}\right)^{2k+1} \cdot \sum_{i=1}^N f^{(2k)}(x_i) = K_{2k} \cdot \sum_{i=1}^N f^{(2k)}(x_i)$$

- Попробуем оценить средний (по положению начальной точки сетки) вклад каждого члена суммы

Можно попытаться получить оценку площади с помощью полиномиальной аппроксимации функции в окрестности каждого узла рядом Тейлора, число членов ряда бесконечно. Нечетные степени имеют интеграл, равный нулю и потому исчезают из суммы после интегрирования (симметричный интеграл нечетной функции равен нулю). Интегрированный ряд Тейлора сходится к «истинному» интегралу функции.

## Средний (по положению начальной точки сетки) вклад $2k$ -слагаемого

- Среднее значение вклада  $2k$ -й производной

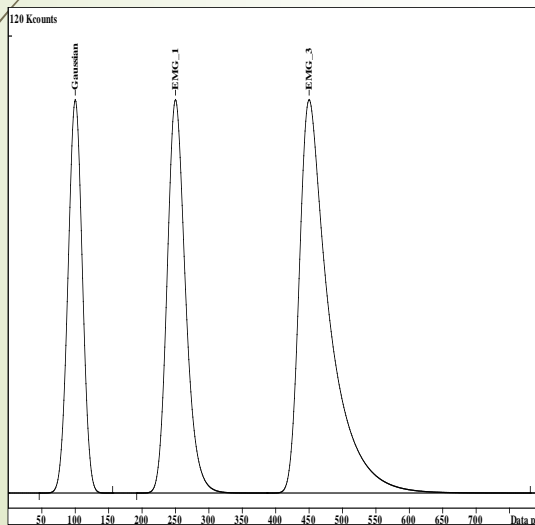
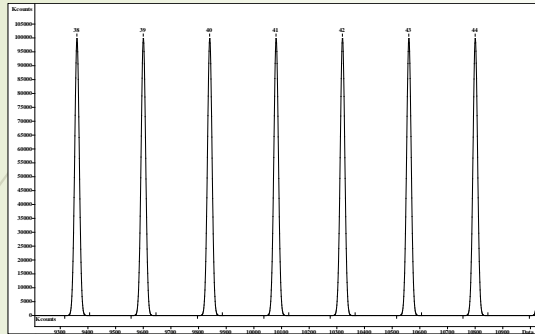
$$\begin{aligned} \Delta A_{2k} &= \frac{1}{\varepsilon} \int_{-\frac{\varepsilon}{2}}^{\frac{\varepsilon}{2}} \Delta A_{2k}(\tau) d\tau = \frac{1}{\varepsilon} K_{2k} \int_{-\frac{\varepsilon}{2}}^{\frac{\varepsilon}{2}} \sum_{i=1}^N f^{(2k)}(x_i + \tau) d\tau \\ \sum_{i=1}^N \int_{-\frac{\varepsilon}{2}}^{\frac{\varepsilon}{2}} f^{(2k)}(x_i + \tau) d\tau &= \int_0^{(N+1)\varepsilon} f^{(2k)}(\tau) d\tau \\ &= f^{(2k-1)}(0) - f^{(2k-1)}((N+1) \cdot \varepsilon) = \mathbf{0} \end{aligned}$$

- Среднее значение вклада каждой производной равно нулю
- Следовательно,  $A_0$  (площадь по методу прямоугольников) является несмещенной оценкой площади пика**

Интеграл и сумму в данном случае можно менять местами, а сумма интегралов производной в окрестностях точек равна интегралу этой производной по всему пику, который равен нулю (интеграл любой производной по пику равен нулю).



# Моделирование пиков



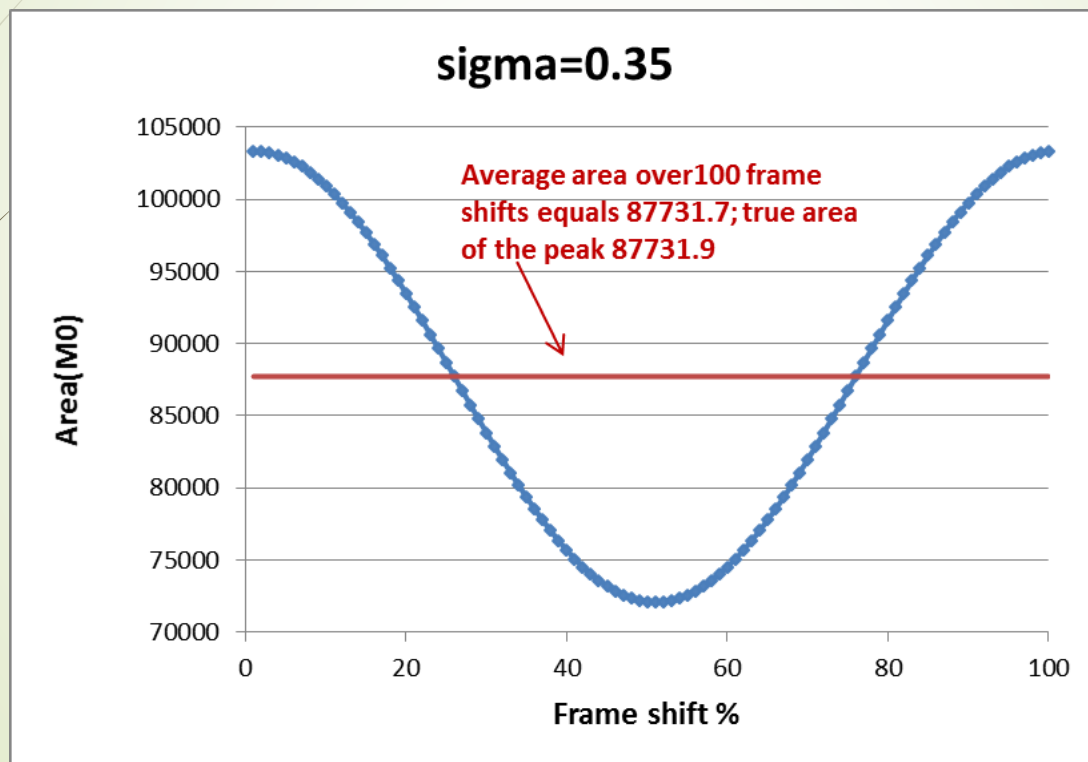
100 пиков  
 Без шума  
 Экспоненциально модифицированная гауссиана  
 Высота=100000 ед.  
 Отклик в каждой точке округлен до целого числа  
 $\sigma_G = 0.35 \dots 8$  точек  
 $t/\sigma_G = 0$  (Гауссиана); 1 (ЭМГ-1); 3 (ЭМГ-3)

Шаг =  $\max(20 \cdot \sigma + 0.01, 30.01)$

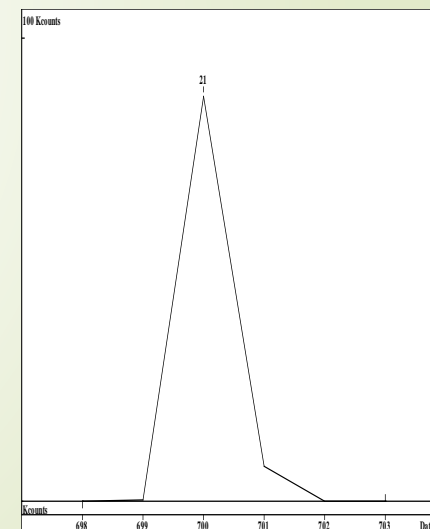
Свойство: Дисперсия ЭМГ  $\sigma^2 = \sigma_G^2 + t^2$

Теорию несложно проверить простым модельным экспериментом. Делаем 100 неперекрывающихся пиков, у каждого следующего вершина сдвинута относительно предыдущего на целое число плюс 1/100 точки, и меряем их площади. У нас должна получиться средняя площадь, равная «истинной» площади пика. Истинная площадь гауссианы известна из ее формулы.

# ЗАВИСИМОСТЬ ПЛОЩАДИ ОТ СДВИГА СЕТКИ



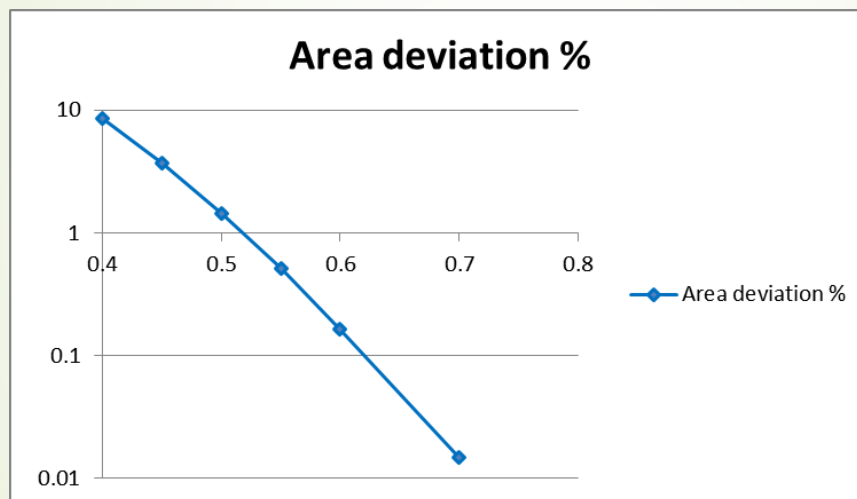
Гауссиана;  
Площадь vs сдвиг сетки. Сдвиг на 100% эквивалентен сдвигу на 1 точку.



На графике приведена зависимость площади пика от сдвига сетки оцифровки. Сдвиг на 100% соответствует одной точке сетки. Генерировались пики Гауссовой формы, частота оцифровки предполагалась равной 0.35 точек на сигму. Форма одного из пиков приведена справа.

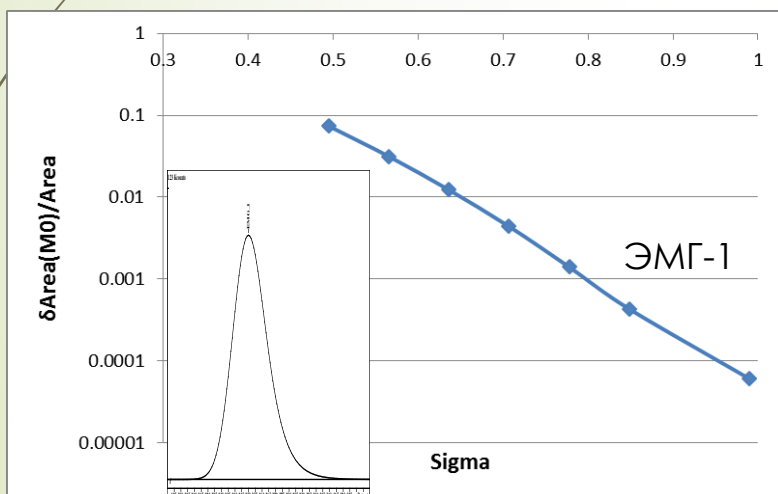
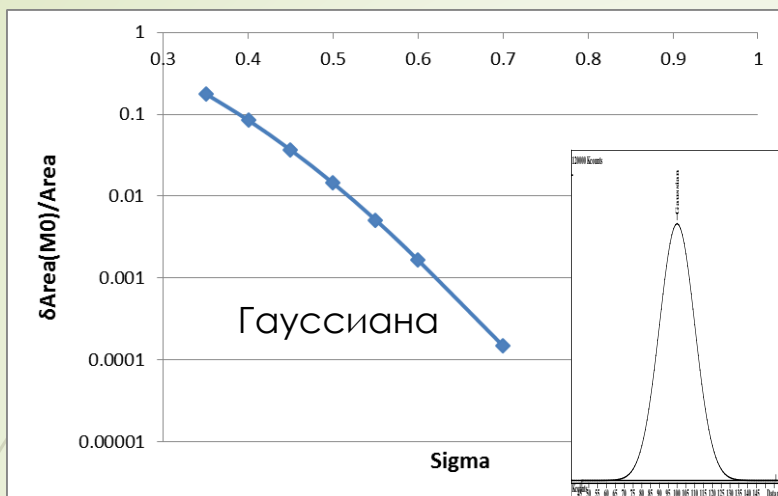
Как и ожидалось, среднее значение площади по 100 точкам равно ожидаемой площади пика. Разница в последнем (шестом) знаке вызвана округлением отклика до целых чисел. График отклонений от среднего очень напоминает синусоиду.

## Гауссиана: максимальная погрешность в зависимости от частоты оцифровки

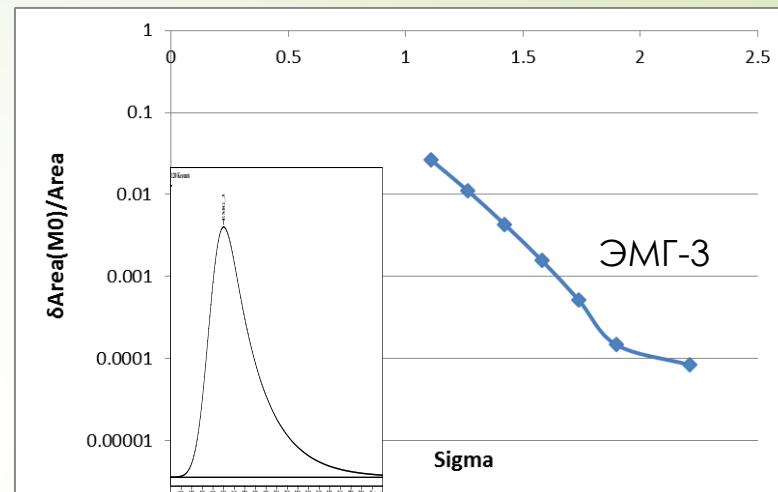


Для измерения площади Гауссианы с точностью до 0.1% требуется всего 0.63 точек на  $\sigma$  (2.5 точки на ширину по базовой линии, 5 точек на весь пик)!

Зависимость амплитуды «синусоиды» от частоты оцифровки. Для Гауссианы погрешность площади 0.1% достигается при частоте оцифровки 0.63 точки на сигму.



## Гауссиана vs ЭМГ:

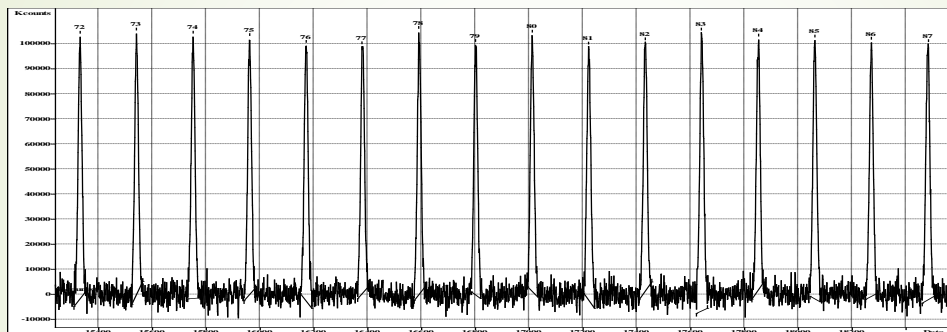


Частота измерений, при которой погрешность не превышает заданных пределов, для пиков разной формы (измерений на  $\sigma$  пика,  $\sigma^2 = \sigma_G^2 + \tau^2$ )

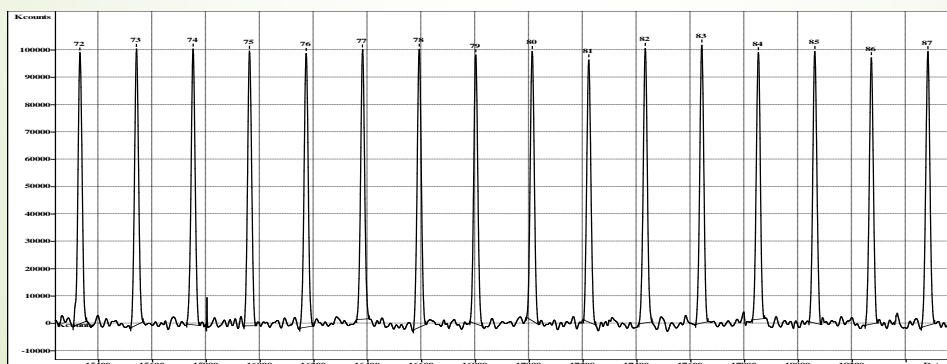
	1%	0.1%
Гауссиана	0.52	0.63
ЭМГ-1	0.65	0.8
ЭМГ-3	1.25	1.6

Зависимость амплитуды «синусоиды» от частоты оцифровки для асимметричных пиков. Форма модельных пиков приведена на каждом рисунке.

$$S/N = (\text{Высота пика } h) / 3\sigma_{\text{noise}} = 10$$
$$h = 100000; \sigma_{\text{noise}} = 3333$$



Нефильтрованные  
данные



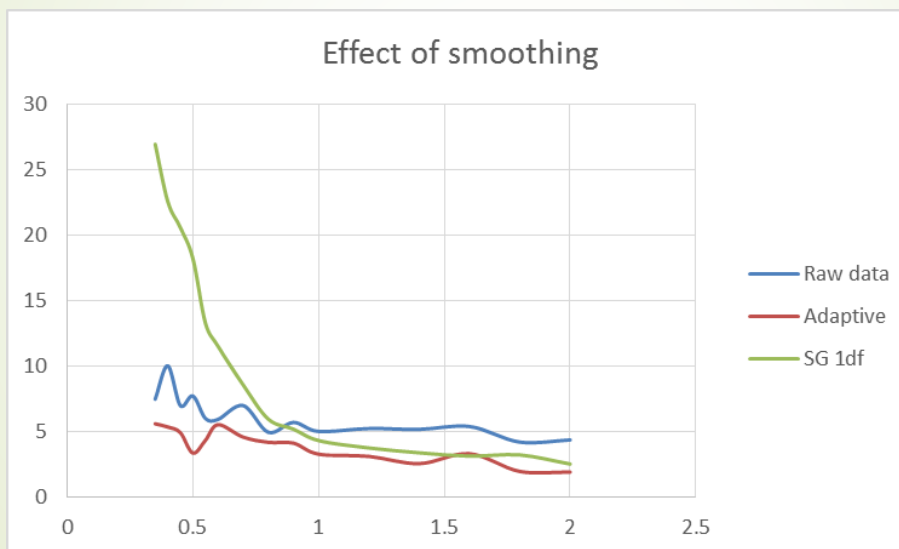
Применен Адаптивный  
фильтр

Влияние шумов на параметры пика и возможность применения цифровой фильтрации. На хроматограмму наложен шум, равный 1/10 его высоты. Это максимальный шум, при котором согласно правилам Фармакопей допустим количественный расчет.

На нижнем рисунке – хроматограмма после фильтрации Адаптивным фильтром (Каламбет Ю.А., Мальцев С.А., Козьмин Ю.П. Фильтрация шумов: окончательное решение проблемы. – Аналитика, №1 (1), 2011, с.50–55; <http://www.multichrom.ru/Docs/50-56Analyt-journ.pdf>)

# Фильтрация шумов

Погрешность измерения площади, %



$100 \cdot \frac{\text{измеренная-ожидаемая}}{\text{ожидаемая}}$   
 медиана по 100 пикам

Метод Савицкого-Голея увеличивает ошибку интегрирования для узких ( $<0.8$  тчк/σ) пиков

Адаптивный фильтр эффективен для всех пиков

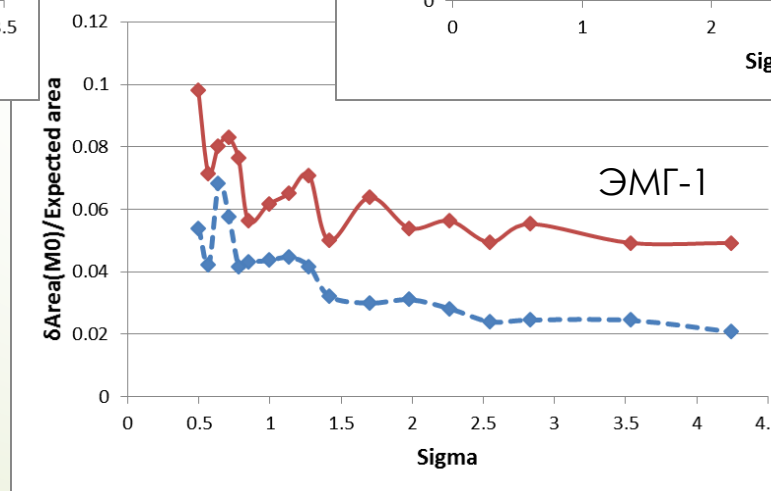
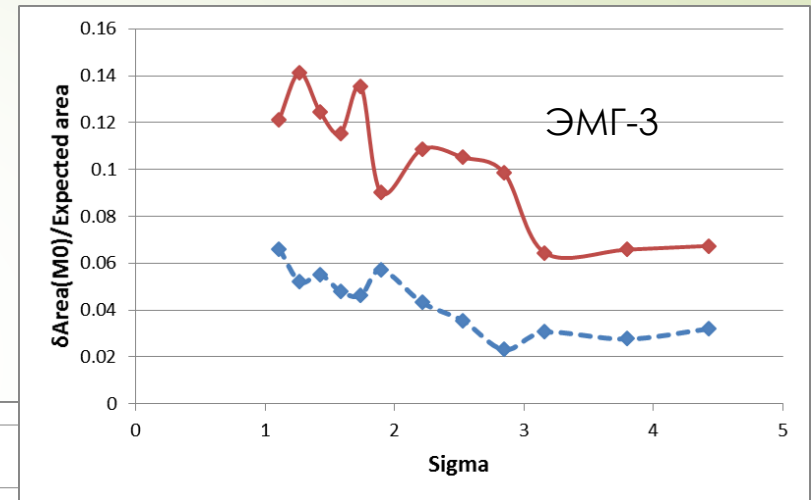
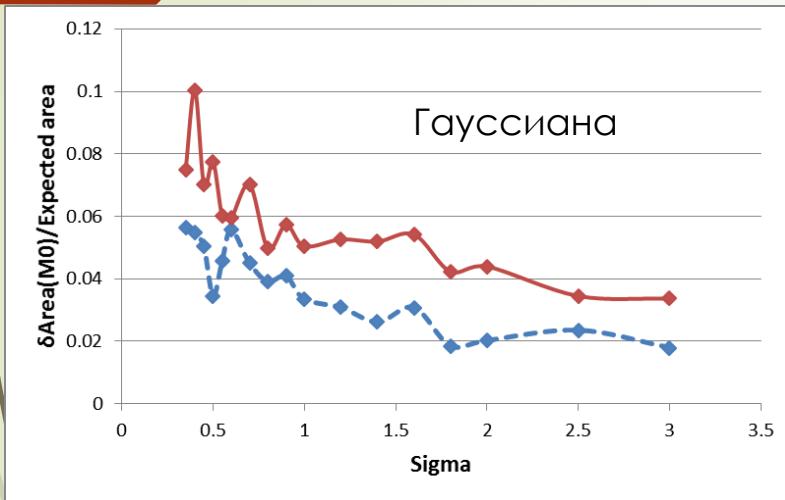
Адаптивный фильтр всегда лучше Савицкого-Голея

Большинство фильтров не способны сглаживать шумы при наличии узких пиков. Выбраны фильтры Савицкого-Голея и Адаптивный, поскольку они меняют форму пика в наименьшей степени. Для метода Савицкого-Голея применен фильтр с минимальным возможным окном, равным 5 точкам, по которым, согласно правилам фильтра, проводится полином 3 степени (одна степень свободы). Отфильтрованным значением является значение полинома в середине окна. Выбор окна, степени и положения аппроксимирующего полинома в Адаптивном фильтре производится автоматически.

На графике приведена зависимость средней медианной погрешности вычисленной площади пика от частоты оцифровки. Поиск пиков в каждой хроматограмме производился программой МультиХром заново. Результаты в случае применения Адаптивного фильтра всегда лучшие. Преимущество Адаптивного фильтра в данной ситуации состоит в том, что он в случае узких пиков фильтрует точки, относящиеся к базовой линии, и не трогает сами пики.

Далее везде под фильтрованным сигналом понимаются данные, в которых проведено сглаживание Адаптивным фильтром.

# Ошибки измерения площади, вызванные шумом нефильтрованные и фильтрованные данные



Фильтрация шумов уменьшает погрешность определения площади в 2-3 раза, при этом погрешность определения площади отфильтрованных данных находится в пределах от 2 до 7 процентов.

Во-первых, для того, чтобы погрешность от частоты квантования для Гауссовых пиков была пренебрежимо мала по сравнению с погрешностью, происходящей от шума, можно в данном случае ориентироваться на погрешность от частоты оцифровки не 0.1%, а 1%.

Во-вторых, это значит, что порог количественного определения должен учитывать как частоту оцифровки, так и фильтрацию шумов, и во многих случаях может быть существенно понижен.

## ВЫВОДЫ

- **Метод прямоугольников (или трапеций) является лучшим методом измерения площади пиков, обеспечивая ее несмещенную оценку**
- **Площадь пика может быть надежно (0.1%) измерена для Гауссовых пиков при 0.63 точек на  $\sigma$ , для несимметричных пиков эта величина может вырасти до 1.6 точек на  $\sigma$**
- **Шумы оказывают на оценку параметров узких пиков большое влияние, зависящее от интенсивности шума**
- **Адаптивная фильтрация шумов может заметно уменьшить погрешность интегрирования**
- **Предел определения может быть существенно понижен**