

УДК 543.544:543.064

ПРИМЕНЕНИЕ ЛОГИСТИЧЕСКОЙ РЕГРЕССИИ ПРИ КОЛИЧЕСТВЕННОМ АНАЛИЗЕ МЕТОДОМ ПОСЛЕДОВАТЕЛЬНЫХ СТАНДАРТНЫХ ДОБАВОК

Мариничев А.Н., Морозова Т.Е., Зенкевич И.Г.

*Санкт-Петербургский государственный университет, Санкт-Петербург,
e-mail: izenkevich@mail15.com*

Показано, что для обработки результатов количественного хроматографического анализа методом последовательных стандартных добавок с экстраполяцией результатов на нулевую или бесконечно большую величины добавок можно использовать логистическую регрессию, характеризующуюся существованием двух пределов функции при стремящихся как к нулю, так и к бесконечности значениях аргумента. Результаты анализа различных образцов, полученные с применением логистической регрессии, сопоставлены со значениями, экстраполированными с использованием ранее предложенных линейной и гиперболической функций.

Ключевые слова: хроматография, количественный анализ, метод последовательных стандартных добавок, обработка результатов, логистическая регрессия

APPLICATING THE LOGISTIC REGRESSION FOR QUANTITATION BY METHOD OF SUCCESSIVE STANDARD ADDITIONS

Marinichev A.N., Morozova T.E., Zenkevich I.G.

St. Petersburg State University, St. Petersburg, e-mail: izenkevich@mail15.com

The method of successive standard additions for chromatographic quantitation often implies the extrapolation of results to zero or infinite values of addition. It is shown that processing of results is possible using logistic regression having two limits for arguments tending to zero or to infinity. The results of analyses of different samples are compared with values based on the extrapolation using previously proposed linear and hyperbolic functions.

Keywords: chromatography, quantitation, successive standard additions, processing of results, logistic regression

Среди известных методов количественного газохроматографического анализа [9] метод стандартной добавки (СД) характеризуется уникальными возможностями. Для

определения суммарного количества аналита (M_x) анализируют исходный образец до и после добавки известного количества образца сравнения ($M_{доб}$):

$$M_x = M_{доб} S_x / (S_{x+доб} - S_x) = M_{доб} / (S_{x+доб} / S_x - S_x) \quad (1)$$

где S_x и $S_{x+доб}$ – площади пиков определяемого компонента до и после добавки.

Только этот способ позволяет определять суммарное количество целевых аналитов в гетерофазных системах по результатам анализа только одной из сосуществующих фаз (обычно выбирают ту, которая содержит меньшее количество мешающих веществ) [2, 3, 6]. Кроме того, как было недавно показано, он применим

в условиях нелинейности детектирования, когда возможности других методов ограничены [4, 6-8], а также при недостаточной инертности хроматографических систем по отношению к определяемым компонентам. Стандартный вариант метода, основанный на соотношении (1), в подобных случаях неприменим; необходимо использование нескольких последовательных стандартных

добавок, $M_{доб(i)} = \sum_{доб(1 \leq k \leq i)}$:

$$M_{x(i)} = M_{доб(i)} S_x / (S_{x+доб(i)} - S_x) = M_{доб(i)} / (S_{x+доб(i)} / S_x - S_x) \quad (2)$$

В результате получаем совокупность нескольких значений определяемой величины $M_{x(i)}$, соответствующих числу введенных добавок. Если все они одинаковы в пределах стандартных отклонений,

$$M_{x(1)} \approx M_{x(2)} \approx \dots \approx M_{x(i)}$$

то результат определений может быть получен их простым усреднением [5]. Если же наблюдается отчетливо выраженная за-

висимость определяемого количества от массы добавки, $M_x = f(M_{доб})$, то необходима дополнительная обработка величин $M_{x(i)}$, заключающаяся в оценке значений M_{x_2} экстраполированных либо на нулевую добавку ($M_{доб} \rightarrow 0$), либо, как впервые показано в работах [4, 6-8], на ее бесконечно большую величину ($M_{доб} \rightarrow \infty$). Следовательно, расширение областей применения метода СД связано с решением двух достаточно «не-

обычных» для химии и хроматографии математических задач:

1. Располагая набором данных $\{S_{x \pm \text{доб}(i)} \div M_{\text{доб}(i)}\}$ при условии $M_{x(i)} \neq \text{const}$ выбрать «направление» экстраполяции ($M_{\text{доб}} \rightarrow 0$ или $M_{\text{доб}} \rightarrow \infty$);

2. Выбрать функцию для вычисления экстраполированных значений.

Возможности решения первой задачи проанализированы в статье [7]. Настоящая работа посвящена обсуждению второй проблемы. Обсуждаемые ниже примеры определения различных соединений в разнообразных матрицах методом последовательных СД подробно охарактеризованы в публикациях [4, 6-8].

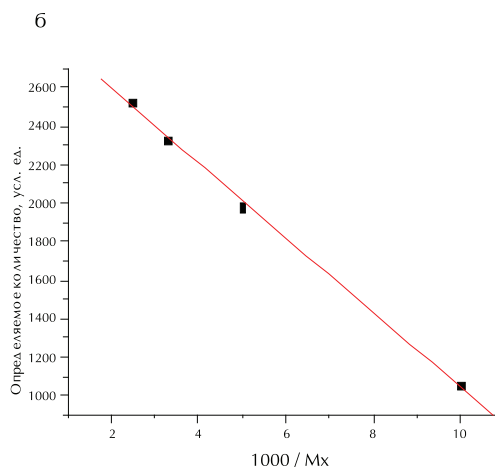
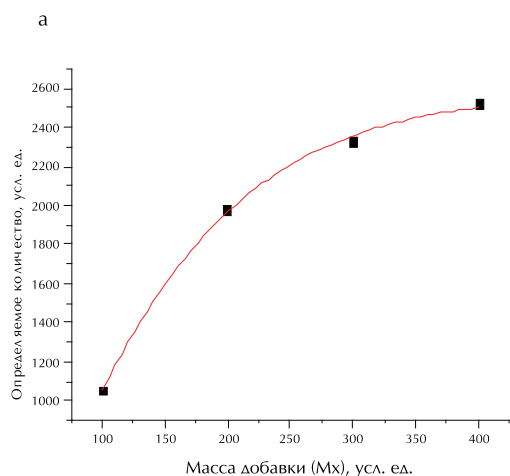
Результаты и их обсуждение. Наиболее часто встречающимся в аналитической практике простейшим вариантом обработки результатов, полученных методом последовательных СД, является их экстраполяция на нулевые величины добавок в соответствии с уравнением линейной регрессии:

$$M_x = aM_{\text{доб}} + b, \lim(M_x)|_{(M_{\text{доб}} \rightarrow 0)} = b \quad (3)$$

Однако при решении различных аналитических задач встречается необходимость «альтернативной» обработки данных с их экстраполяцией на бесконечно большие величины добавок. В этом случае использование уравнения (3) невозможно, так как аппроксимирующая функция должна иметь предел при $M_{\text{доб}} \rightarrow \infty$. Простейшей функцией, удовлетворяющей этому требованию, является двухпараметровая гиперболическая зависимость, выбранная для этих целей в работах [4, 6-8]:

$$M_x = a/M_{\text{доб}} + b, \lim(M_x)|_{(M_{\text{доб}} \rightarrow \infty)} = b \quad (4)$$

Графическая иллюстрация этого варианта экстраполяции приведена на рисунке, а результаты вычислений экстраполированных значений сопоставлены с заданными содержаниями аналитов в таблице. В 12 случаях из 14 линейная [уравнение (1)] либо гиперболическая [уравнение (2)] экстраполяция обеспечивает точность в пределах $\pm 15\%$ отн.



Схематическая иллюстрация экстраполяции результатов определений методом последовательных стандартных добавок на бесконечно большую величину добавки ($M_{\text{доб}} \rightarrow \infty$):

*a – исходная нелинейная зависимость $M_x = f(M_{\text{доб}})$;
 б – зависимость $M_x = a(1000/M_{\text{доб}}) + b, \lim(M_x)/(M_{\text{доб}} \rightarrow \infty) = b$*

Сравнение результатов количественного анализа методом последовательных стандартных добавок с использованием различных способов экстраполяции

Анализируемый образец	Исходные данные, $M_x(M_{доб})$	Экстраполирующая функция	$M_{экстраполир}^*$	Заданное содержание аналита**	Логистическая регрессия***
1. Тест-смеси для области отсутствия инертности хроматографических систем	17.94(17.5), 18.89(35)****, 18.26(52.5), 18.08(70)	Гиперболическая ($\rightarrow \infty$)	17.9	17.5 мкг/мл	18.3(3); 18.4(4) [0.1]
2. То же	26.51(20.5), 25.69(41), 24.76(61.5), 24.12(82)	Гиперболическая ($\rightarrow \infty$)	23.7	25 мкг/мл	25.3(3); 25.3(4) [0.1]
3. То же	43.0(35), 40.9(52.5), 33.5(70), 31.8(87.5)	Гиперболическая ($\rightarrow \infty$)	25.0	17.5 мкг/мл	37.3(3); 37.3(4) [0.1]
4. То же	79.79(50), 73.03(75), 58.51(100), 42.69(125)	Гиперболическая ($\rightarrow \infty$)	27.0	25 мкг/мл	63.7(3); 63.7(4) [0.001]
5. То же	16.97(19), 17.07(38), 17.09(57), 15.7(76), 15.48(95)	Линейная ($\rightarrow 0$)	17.8	19 мкг/мл	4.8(3); 16.5(4) [0.1]
6. То же	15.77(12.5), 13.1(25), 12.63(37.5), 11.72(50), 13.57(62.5)	Гиперболическая ($\rightarrow \infty$)	10.6	12.5 мкг/мл	13.4(3); 13.4(4) [0.1]
7. Гидрофобный аналит в гидрофобной матрице [3]	5.8(8.73), 6.2(17.46), 6.2(26.19), 6.4(34.92)	Гиперболическая ($\rightarrow \infty$)	6.5	8.8 мг	6.4(3); 8.3(4) [0.01]
8. Гидрофильный аналит в гидрофильной матрице [3]	4.52(8.22), 5.43(16.44), 5.54(24.66), 5.46(32.88)	Гиперболическая ($\rightarrow \infty$)	6.0	6.5 мг	5.5(3); 5.5(4) [0.01]
9. 3-(2,2,2-Триметилгидразиний)пропионовая кислота (ТНР) в моче [3, 7]	70.2(200), 69.3(300), 57.9(400), 57.3(500)	Гиперболическая ($\rightarrow \infty$)	48.8	50 мкг/мл	63.7(3); 63.7(4) [0.01]
10. То же	726(100), 634(200), 623(300)	Гиперболическая ($\rightarrow \infty$)	563	600 мкг/мл	660(3); 779(4) [1]
11. То же	672(100), 563(200), 500(300)	Гиперболическая ($\rightarrow \infty$)	427	400 мкг/мл	747(3); 681(4) [1]
12. Камфора в мазях [6]	76(27), 50(55), 47(85)	Линейная ($\rightarrow 0$)	85.2	90 мг	2.1(3); 58(4) [0.01]
13. Моноэтаноламин в водных растворах [4]	0.23(0.375), 0.35(0.875)	Линейная ($\rightarrow 0$)	0.14	0.125 мкг/мл	0.126(3); 0.111(4) [1]
14. То же	1.88(1.0), 1.84(2.0)	Линейная ($\rightarrow 0$)	1.92	2.0 мкг/мл	1.86(3); 1.86(4) [1]

*) Жирным шрифтом выделены экстраполированные значения, совпадающие с заданным содержанием аналитов в пределах $\pm 15\%$; **) масса (мг) или концентрация (мкг/мл); ***) в круглых скобках – число параметров логистической регрессии (3 или 4), в квадратных – коэффициент масштабирования независимой переменной; ****) курсивом выделены значения, исключенные при линейной или гиперболической экстраполяции.

Тем не менее, для лучшего понимания и обоснования возможностей метода последовательных СД целесообразно рассмотрение вопроса, можно ли описать оба варианта экстраполяции данных ($M_{доб} \rightarrow 0$ и $M_{доб} \rightarrow \infty$) одним и тем же регрессионным уравнением. При этом число параметров уравнения не должно быть чрезмерно большим, чтобы не увеличивать объем экспериментальной работы (число добавок). В связи с этим представляет интерес анализ возможностей логистической регрессии.

Логистическая регрессия может быть описана следующим известным нелинейным четырехпараметровым соотношением:

$$y = \frac{a}{1 + b \exp(-kx)} + c, \quad (5)$$

В уравнении трехпараметрового варианта логистической регрессии отсутствует свободный член (коэффициент c). Коэффициенты c и a иногда называют параметрами смещения и масштаба соответственно. Если они известны, то можно ввести новую безразмерную переменную $u = \frac{y-c}{a}$, продифференцировать ее по x и, после исключения параметра b , получить однопараметровое нелинейное дифференциальное уравнение первого порядка:

$$\frac{du}{dx} = k u(1-u). \quad (6)$$

Подобное уравнение, известное как одно из уравнений химической кинетики, было использовано для описания автокаталитических реакций (конечный продукт является катализатором) [1]. В таких случаях в качестве независимой переменной x рассматривают время t , а зависимая переменная u представляет собой отношение концентрации конечного продукта в момент времени t к начальной концентрации (при $t = 0$) исходного реагента. Из уравнения (6) следует, что значение $k/4$ отвечает максимальной скорости процесса в момент времени t^* , соответствующий точке перегиба зависимости $u(x)$. Значение t^* равно отношению $\ln(b)/k$ и поэтому параметры b и k уравнения (5) должны быть большими нуля.

В настоящей работе соотношение (5) и некоторые его частные формы, являющи-

ся решениями уравнения (6), использованы для обработки данных количественного газохроматографического анализа методом последовательных СД. Независимой переменной x , принимающей только положительные значения, будем считать массу стандартной добавки ($M_{доб}$), в качестве зависимой переменной y можно рассматривать инструментальный отклик в виде площадей хроматографических пиков или, нагляднее, вычисленных на их основе по уравнению (2) содержаний аналита в исходной пробе (M_x). Массив экспериментальных данных можно представить как совокупность n пар переменных: (x_i, y_i) , $1 \leq i \leq n$. С использованием этого массива для нахождения значений параметров a , b , c и k соотношения (5) минимизировали целевую функцию W , представляющую собой сумму квадратов отклонений:

$$W = \sum_{i=1}^n \left(y_i - \frac{a}{1 + b \exp(-kx_i)} - c \right)^2. \quad (7)$$

Кроме четырехпараметрового соотношения (5) использовали трех и двухпараметровые уравнения с дополнительными ограничениями ($c = 0$) и ($c = 0, b = 1$), соответственно.

Особый интерес для оценки возможностей применения приведенных выше уравнений для обработки данных, полученных методом последовательных СД, представляют предельные значения

$$y(x = 0) = a / (1 + b) + c. \text{ и } y(x = \infty) = a + c.$$

Именно в интервал между этими значениями попадают экспериментально определяемые величины $M_{x(i)} = f[M_{доб(i)}]$, которые необходимо аппроксимировать обсуждаемой нелинейной зависимостью, характеризующейся точкой перегиба с координатами

$$y(x = \ln(b)/k) = a/2 + c.$$

Вычисления проводили в среде Maple 14. Проверку полученных результатов осуществляли программным средством «Подбор параметров» электронных таблиц Microsoft Excel 2007. В качестве примера приведем фрагмент вычислений параметров a, b, c, k для примера № 7 в среде Maple 14 (после символа # приведены комментарии):

```
>M[доб]:=[8.73,17.46,26.19,34.92]:M:= [5.82,6.2,6.2,6.4]:
#исходные данные
> m:=map(x->x/100,M[доб]):#масштабирование переменной m[доб]
> zip((x,y)->(y-a/(1+b*exp(-k*x))-c)^2,m,M):W4:=add(i,i=%):
#четырёхпараметровая целевая функция
```

```
> with(Optimization):#библиотека команд оптимизации
> Minimize(W4,assume=nonnegative);#наименьшее значение целевой
функции W4 и искомые значения параметров
W4 = 0.024, [a = 5.34, b = 0.96, c = 2.99, k = 1.56]
```

Предельное значение M_x при $M_{\text{доб}} \rightarrow \infty$ равно $a + c = 8.33$ (при заданном количестве анализата 8.8).

Общая сводка результатов оценки содержания анализатов в различных образцах методом последовательных СД с использованием логистической регрессии в сравнении с результатами, полученными линейной и гиперболической экстраполяциями, также приведена в Табл. 1. Соответствие вычисленных и экспериментальных данных в пределах $\pm 15\%$ наблюдается в девяти из 14 примеров, что несколько меньше, чем двумя другими методами. Однако важно отметить, что факты получения неудовлетворительных результатов (примеры № 3, 4, 9, 11 и 12) не только объяснимы, но и позволяют сделать важные заключения об изменении условий анализа методом последовательных СД. Математическим критерием неудовлетворительной логистической регрессии являются высокие значения целевой функции W [соотношение (7)] (примеры 9-12). Кроме того, невысокую точность экстраполяции можно ожидать в тех случаях, когда диапазоны вариаций значений $M_{x(i)}$ чрезмерно велики [в примерах № 3, 4, 11 и 12 отношения $(M_{x(i)\text{макс}} / M_{x(i)\text{мин}})$ превышают 1.35]. В подобных случаях целесообразно изменить экспериментальные условия количественного анализа методом последовательных СД так, чтобы избежать столь существенных вариаций определяемых содержаний анализатов.

Применение метода стандартной добавки и его модификаций (вариант последовательных стандартных добавок) для количественного определения анализатов обеспечивает высокую точность результатов для образцов, отличающихся сложным характером матриц, в условиях нелинейности детектирования, а также в областях

недостаточной инертности хроматографических систем. Однако реализация таких возможностей требует использования «нестандартных» алгоритмов интерпретации результатов, в том числе экстраполяции промежуточных содержаний анализатов на нулевую или бесконечно большую величины добавок. Показано, что для этих целей можно применять логистическую регрессию, характеризующуюся существованием двух пределов функции при стремящихся как к нулю, так и к бесконечности значений аргумента.

Список литературы

1. Денисов Е.Т. Кинетика гомогенных реакций. – М.: Высшая школа, – 1988. – 391 с.
2. Зенкевич И.Г., Климова И.О. Применение метода стандартных добавок для количественного хроматографического анализа. // Журн. аналит. химии. – 2006. – Т. 61. № 10. – С. 1048-1054.
3. Зенкевич И.Г., Морозова Т.Е. Особенности метода стандартных добавок для количественного определения анализатов в сложных матрицах, обладающих сорбционными свойствами // Аналитика и контроль. – 2010. – Т. 14. № 3. – С. 164-171.
4. Зенкевич И.Г., Морозова Т.Е. Особенности ВЭЖХ-МС определения моноэтаноламина в водных растворах методом стандартной добавки. // Аналитика и контроль. – 2012. – Т. 16. № 2. – С. 181-187.
5. Зенкевич И.Г., Рагозина Т.Н. Количественный газохроматографический анализ компонентов гетерофазных систем методом стандартной добавки. // Журн. прикладн. химии. – 1998. – Т. 71. № 5. – С. 763-767.
6. Морозова Т.Е., Зенкевич И.Г. Новые варианты метода стандартных добавок. Газохроматографическое определение камфоры в фармацевтических препаратах. // Вестн. СПбГУ. Сер. физ.-хим. – 2011. – Вып. 4. – С. 61-68.
7. Морозова Т.Е., Каракашев Г.А., Сорокоумов П.Н., Савельева Е.И., Зенкевич И.Г. Сравнение точности метода абсолютной градуировки и модифицированного метода последовательных стандартных добавок в условиях нелинейности детектирования // Аналитика и контроль. – 2013. – Т. 17. № 2. – С. 184-189.
8. Морозова Т.Е., Мариничев А.Н., Зенкевич И.Г. Количественный анализ методом стандартной добавки в условиях нелинейности детектирования. // Вестн. СПбГУ. Сер. физ.-хим. – 2012. – Вып. 4. – С. 126-132.
9. Poole C.P. Gas Chromatography. – Amsterdam: Elsevier, – 2012. – 743 p.