УДК 543.153

ВЛИЯНИЕ ИСКАЖЕНИЯ СПЕКТРА НА ПРАВИЛЬНОСТЬ КОЛИЧЕСТВЕННОГО АНАЛИЗА ПО МЕТОДУ МНОГОМЕРНОЙ КАЛИБРОВКИ

И. М. Дубровкин, В. А. Ананьев

THE IMPACT OF THE SPECTRA DISTORTIONS ON THE TRUENESS OF THE QUANTITATIVE ANALYSIS IN MULTIVARIATE CALIBRATION

I. M. Dubrovkin, V. A. Ananiev

Путем компьютерного моделирования оценена правильность спектрометрического анализа многокомпонентных смесей по методу многомерной калибровки при использовании искаженных спектров. Указаны границы допустимых искажений.

Computer-based evaluation of the trueness of the quantitative analysis of the multicomponent mixtures using multivariate calibration has been performed in the case of the disturbed spectra of the training set. The permissible limits of these distortions have been established.

Ключевые слова: количественный анализ, многомерная калибровка, искажение полос поглощения, правильность.

Keywords: quantitative analysis, multivariate calibration, pure-component spectra distortions, trueness.

В аналитической химии для количественного анализа многокомпонентных смесей применяется многомерная калибровка, которая позволяет установить связь между спектрами эталонных образцов (тренировочное множество) и концентрациями определяемых компонентов [1, с. 302 – 321]. При этом знание параметров спектров значительно улучшает эффективность математических процедур [2, с. 30 – 136]. Параметры могут быть определены или с помощью экспертных систем, основанных на использовании баз спектральных данных, в том числе с привлечением теоретического расчета спектров молекул [3, с. 1053 – 1066; 4], или путем разложения сложных спектров на полосы, обусловленные поглощением индивидуальных компонентов [5, с. 108 – 115]. Очевидно, что параметры индивидуальных полос, полученные математически или экспериментально с использованием первого или второго подхода, и их аналоги в изучаемой смеси будут отличаться из-за неточности теории, погрешности эксперимента и вычислений. Это должно приводить к погрешности в определении концентрации компонентов в смеси. Однако оценить в аналитическом виде возможную погрешность количественного анализа, обусловленную искажением полос индивидуальных компонентов, используя метод многомерной калибровки, не представляется возможным.

Целью настоящей работы являлась численная оценка относительного отклонения вектора расчетных концентраций от истинного значения в смеси (правильность анализа) при использовании, как МНК (*Метод наименьших квадратов*), так и метода PLS (*The Partial Least Squares* или *Проекция на латентные структуры* (ПЛС) [1]). При расчетах предполагалось, что спектры тренировочного множества состоят из полос гауссовой или лоренцевой формы, положения максимумов которых и ширина на полувысоте (полуширина) были возмущены случайными погрешностями. Анализ проводился относительно спектра смеси, который состоял из суммы неискаженных полос.

Теория. В тексте используются общепринятые термины линейной алгебры. Заглавные и строчные буквы, выделенные жирным шрифтом, обозначают матрицы и векторы-столбцы, соответственно, скаляры выделены курсивом. Надстрочные индексы T и –1 используются для обозначения транспонированной и обратной матриц соответственно.

Рассмотрим линейную модель спектра многокомпонентной смеси **y** при соблюдении закона Бугера-Ламберта-Бера:

$$\mathbf{y} = \mathbf{Kc} + \boldsymbol{\varepsilon}, \qquad (1)$$

где $\mathbf{K} = m \times n$ – матрица полос поглощения индивидуальных компонентов; n – число компонентов; m – число точек на оси абсцисс спектра; $\mathbf{c} = n \times 1$ – вектор концентраций компонентов, $\boldsymbol{\varepsilon} = m \times 1$ – вектор нормального шума с нулевым средним и дисперсией σ_y^2.

В общем случае p смесей выражение (1) может быть переписано в классической форме:

$$\mathbf{Y} = \mathbf{CK}^{\mathrm{T}} + \boldsymbol{\Xi}_{\mathbf{y}}, \qquad (2)$$

где $\mathbf{Y} = p \times m$ – матрица спектров смесей тренировочного множества; строки $p \times m$ матрицы **C** составлены из векторов $\mathbf{c}^{\mathrm{T}}$; $\boldsymbol{\Xi}_{\mathbf{y}} = \mathrm{I} \times \sigma_y^2$; I – единичная матрица; σ_y^2 – дисперсия шума.

Предположим, что матрица **K** известна, тогда, согласно классическому методу прямой калибровки [1], оценка матрицы **C** из выражения (2) по МНК [6] может быть получена из выражения:

$$\hat{\mathbf{C}} = \mathbf{Y}\,\hat{B}_{КЛ}, \qquad (3)$$

где $\hat{B}_{КЛ} = \mathbf{K}(\mathbf{K}^{\mathrm{T}}\mathbf{K})^{-1}$ – оценка регрессионной матрицы.

Для обратного метода калибровки (метод **P**-матрицы) [1]:

$$\mathbf{C} = \mathbf{YB}_{\text{обр}} - \boldsymbol{\Xi}_{\mathbf{C}}, \qquad (4)$$

где $\mathbf{B}_{\text{обр}}$ – регрессионная матрица рассчитываемая по ПЛС; $\mathbf{\Xi}_{\mathbf{C}} = \mathrm{I}\sigma_C^2$; σ_C^2 – дисперсия шумов матрицы концентраций тренировочного множества.

Вектор концентрации индивидуальных компонентов в смеси оценивается по ее спектру (1):

$$\hat{\mathbf{c}}^{\mathrm{T}} = \mathbf{y}^{\mathrm{T}} \hat{B}. \quad (5)$$

Пусть столбцы неискаженной матрицы **K** составлены из полос гауссовой (A_{G}) или лоренцевой (A_{L}) формы:

$$A_{\mathrm{Gn}}(x) = \mathrm{A}_{0\mathrm{n}}\exp[-4ln2(x - x_{0\mathrm{n}})^2], \quad (6)$$

$$A_{\mathrm{Ln}}(x) = \mathrm{A}_{0\mathrm{n}}/[1 + 4(x - x_{0\mathrm{n}})^2], \quad (7)$$

где $\mathrm{A}_{0\mathrm{n}}$ – оптическая плотность в максимуме полосы n; $x = i/w_{\mathrm{n}}$; i – целочисленная координата по оси абсцисс спектра; w_{n} – (полуширина); $x_{0\mathrm{n}} = i_{0\mathrm{n}}/w_{\mathrm{n}}$ – положение максимума полосы.

Предположим, что максимум полосы n смещен на величину $\Delta i_{0\mathrm{n}}$, а ее форма искажена за счет изменения полуширины Δw_{n}. Относительные смещения будут равны $\delta_{i_{0n}} = \Delta x_{0\mathrm{n}} = \Delta i/w_{\mathrm{n}}$ и $\delta_{\mathrm{wn}} = \Delta w_{\mathrm{n}}/w_{\mathrm{n}}$. Кроме этого, примем, что каждая полоса искажена случайным шумом. Задавая матрицу концентраций тренировочного множества **C**, из выражения (2) получим матрицу спектров смесей для многомерной калибровки:

$$\mathbf{Y}_{\text{иск}} = (\mathbf{C} + \mathbf{\Xi}_{\mathbf{C}})\mathbf{K}_{иск}^{T} + \mathbf{\Xi}_{\mathbf{y}}, \quad (8)$$

где $\mathbf{K}_{иск}^{T} = \mathbf{K}_{иск}^{сист} + \mathbf{\Xi}_{\mathbf{K}}$ – матрица, составленная из полос, искаженных систематическими погрешностями; $\mathbf{\Xi}_{\mathbf{K}} = \mathrm{I}\sigma_K^2$; σ_K^2 – дисперсия шумов. Из уравнений (4) и (8) находим регрессионную матрицу $\mathbf{B}_{обр}^{иск}$ для обратного метода калибровки, используя ПЛС.

Для классической калибровки регрессионная матрица (3) может быть представлена в виде:

$$\hat{B}_{КЛ}^{ИСК} = \mathbf{K}_{\text{иск}}(\mathbf{K}_{иск}^{T}\mathbf{K}_{\text{иск}})^{-1}.$$

Матрица $\mathbf{K}_{\text{иск}}$ может быть определена из спектров тренировочного множества (8) по МНК (метод **K**-матрицы) [1]:

$$\hat{K}_{ИСК}^{T} = (\mathbf{C}_{\mathrm{T}}\mathbf{C})^{-1}\mathbf{C}_{\mathrm{T}}\mathbf{Y}_{\text{иск}}. \quad (9)$$

Оценка вектора концентрации смеси, обладающей спектром (1), производится с помощью выражения подобного (5):

$$\hat{\mathbf{c}}_{иск}^{T} = \mathbf{y}^{\mathrm{T}}\hat{B}^{иск}.$$

В качестве меры влияния погрешности матрицы $\mathbf{K}_{\text{иск}}$ на вектор определяемых концентраций будем использовать относительную погрешность оценки этого вектора:

$$\delta_c = \| \overline{\hat{c}_{ИСК} - \bar{c}} \| / \|\mathrm{c}_{\text{здн}}\|, \quad (10)$$

где $\|\ \|$ – символ Евклидовой нормы вектора; черта означает усреднение по независимым экспериментам; $\mathrm{C}_{\text{здн}}$ – заданный вектор концентрации смеси.

Поскольку вектор $\hat{\mathbf{c}}$ искажен только случайными погрешностями, а $\hat{\mathbf{c}}_{\text{иск}}$ – как случайными, так и систематическими, то относительная погрешность (10) служит характеристикой относительного отклонения вектора расчетных концентраций от истинного значения (правильность анализа).

Описание численного эксперимента для модельных спектров

Для численного эксперимента использовали три теоретических спектра трехкомпонентных смесей, представленных на рис.1а-в ($\Delta = \mathrm{A}/\mathrm{A}_{\max}$ – относительная оптическая плотность; $\mathrm{A}_{\max}$ – максимальная оптическая плотность спектра смеси). Полоса поглощения каждого компонента (столбец матрицы **K**) имеет гауссову форму. Параметры полос представлены в таблице. Как видно, спектры ***а, б*** состоят из полос одинаковой полуширины, различным образом перекрывающихся между собой, а спектр ***в*** – из полос разной полуширины. Все три полосы имеют одинаковую оптическую плотность в максимуме. Мерой чувствительности расчетной концентрации к возмущениям матрицы **K** может служить ее число обусловленности – cond(**K**) [6] (таблица). При расчетах использовали $m = 100$.

Таблица

Параметры полос и значение cond(K)

Спектр	*Параметры полос*						*cond( )*	
	i_{01}	i_{02}	i_{03}	w_1	w_2	w_3	*Гаусс*	*Лоренц*
1	120	160	180	50	50	50	4	5
2	150	160	165	100	100	100	453	228
3	150	160	170	100	75	35	19	28

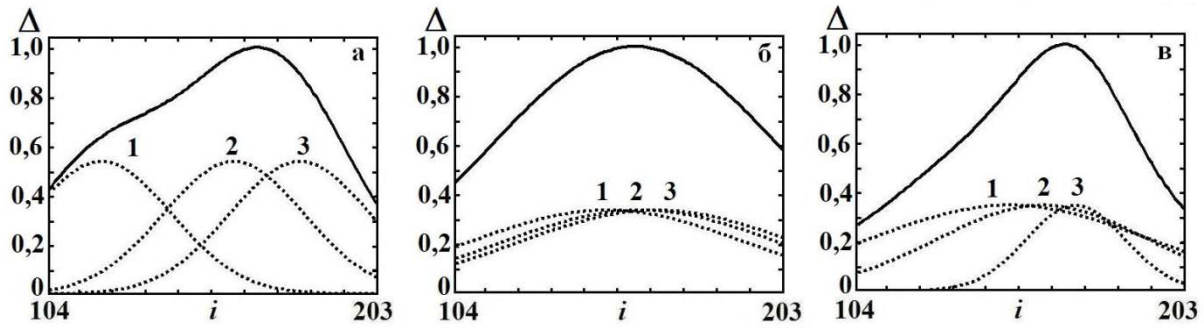

Рис. 1. ***Спектры смесей а-в (—), состоящих из полос 1 – 3 (…)***

Кроме этого, были использованы спектры, аналогичные описанным выше, но состоящие из полос лоренцевой формы с параметрами, совпадающими с теми, которые представлены в таблице.

Относительная погрешность δ_c рассчитывалась как среднее из результатов 10^5 независимых численных экспериментов. Стандартное отклонение для этой величины оценивалось при P = 95 %. В каждом эксперименте относительное смещение максимума полосы δ_{n0} и относительное смещение её полуширины δ_w являлись нормально распределенными случайными величинами с нулевым средним и с фиксированным стандартным отклонением $\sigma_{\delta_{i0}}$ и σ_{δ_w} соответственно.

Матрица концентраций тренировочного множества включала 50 образцов. Каждая её строка была составлена из трёх случайных чисел, равномерно распределенных в области 0-1. В каждой серии численных расчетов генерировались шумы, обладающие следующими стандартными отклонениями: $\sigma_Y = \sigma_Y = \sigma_K = 10^{-3}$; $\sigma_C = 5\times10^{-3}$.

Для вычислений и построения графиков был написана программа на языке MATLAB. При калибровке по методу ПЛС использовалось два компонента.

Результаты и их обсуждение

Результаты численных расчетов верхней границы погрешности δ_C^m для спектров 1 ***а-в*** от ошибок δ_{n0} и δ_w (P = 95 %) представлены на рис. 2, 3 (буквы соответствуют таковым для рисунка 1)**.** Выбор верхней границы обусловлен тем, что в некоторых случаях стандартное отклонение величины δ_c превышало её среднее значение.

Анализ полученных данных позволяет сделать следующие выводы.

1. Использование МНК для сильно перекрывающихся полос неприемлемо, так как уже при смещении максимума полос δ_{n0}, имеющем стандартное отклонение $\sigma_{\delta_{i0}}$ = 10 %, ошибка определения вектора концентрации может превысить 20 % (рис. 2 $б_1$, $б_2$). Кроме этого, значительный разброс рассчитанных величин свидетельствует о большой неустойчивости полученных оценок.

2. Согласно [8], верхняя граница погрешности, определяемая из уравнения (10):

$$\delta_c^{\sup} < \text{или} \approx \text{cond}(\mathbf{K})||\Delta\mathbf{K}||/||\mathbf{K}||, \qquad (11)$$

где $||\Delta\mathbf{K}||$ норма возмущения матрицы **K**. Из наших оценок следует, что $\delta_C^m << \delta_c^{\sup}$ и что cond(**K**) не является доминирующим фактором, определяющим величину δ_C^m. Так, например, значения cond(**K**) для спектров ***а, б*** отличаются в десятки раз (таблица), а погрешности δ_C^m близки между собой (рис.2 $а_1$,$б_1$). В этой связи представляется необоснованной рекомендация по использованию расширенной формы выражения (11) со знаком равенства в качестве универсальной оценки погрешности анализа [9].

3. Как видно из данных, представленных на рис. 2 и 3, переход от МНК к методу ПЛС приводит к существенному уменьшению погрешности. Даже для спектра ***б***, характеризующегося высокой коллинеарностью индивидуальных полос поглощения, максимальное значение δ_C^m не превышает 12 % (рис. 3 $б_1$).

4. Смещение максимумов полос поглощения оказывает значительно больший возмущающий эффект, чем изменение их полуширины. Этим, по всей видимости, объясняется относительно слабое влияние формы полос (гауссовая или лоренцевая) на погрешность δ_C^m (при $\sigma_{\delta_{i0}}$ или σ_{δ_w} < 10 %) при неизменности положения их максимумов. Тем не менее наблюдается зависимость δ_C^m от ширины индивидуальных полос при калибровке по МНК. Например, для спектра ***в***, обусловленного полосами разной ширины, погрешность гораздо меньше при больших значениях $\sigma_{\delta_{i0}}$,чем для спектров ***а, б***, полосы которых имеют равную полуширину. Необходимо отметить, что для ПЛС этот эффект не наблюдается (рис. 3).

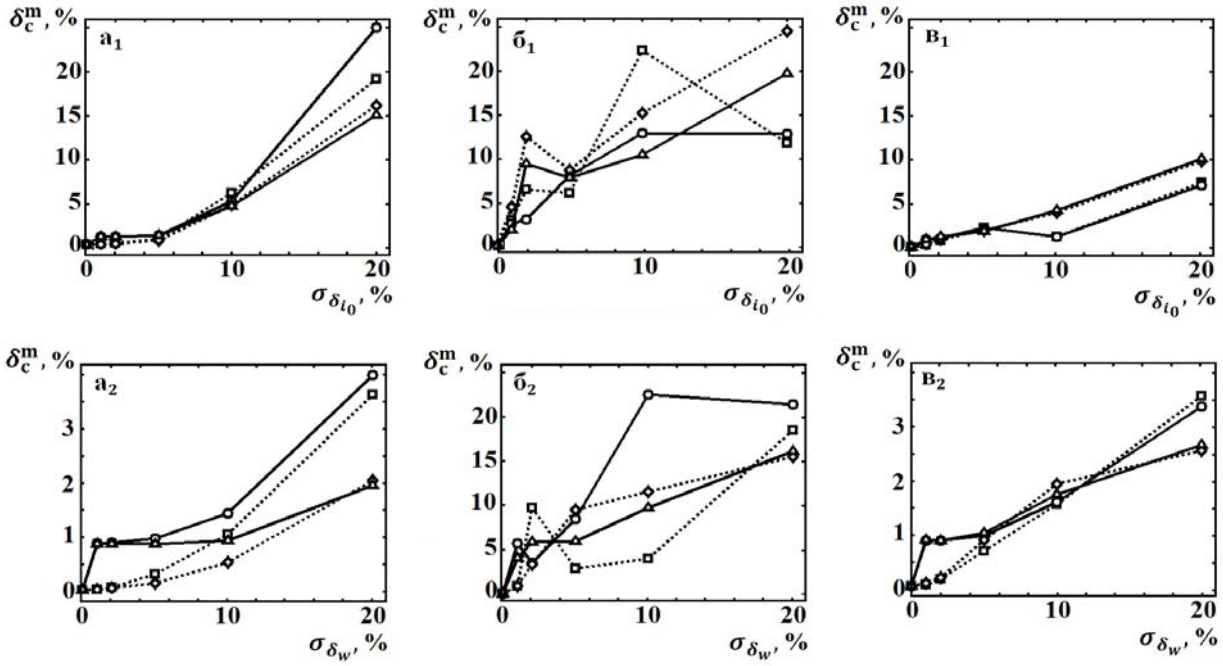

Рис. 2. ***Максимальные погрешности определения вектора концентраций по МНК для спектров 1 а-в без реконструкции полос поглощения индивидуальных компонентов (—) и с реконструкцией (…) для гауссовых (o, □) и лоренцевых (Δ, ◊) полос соответственно***

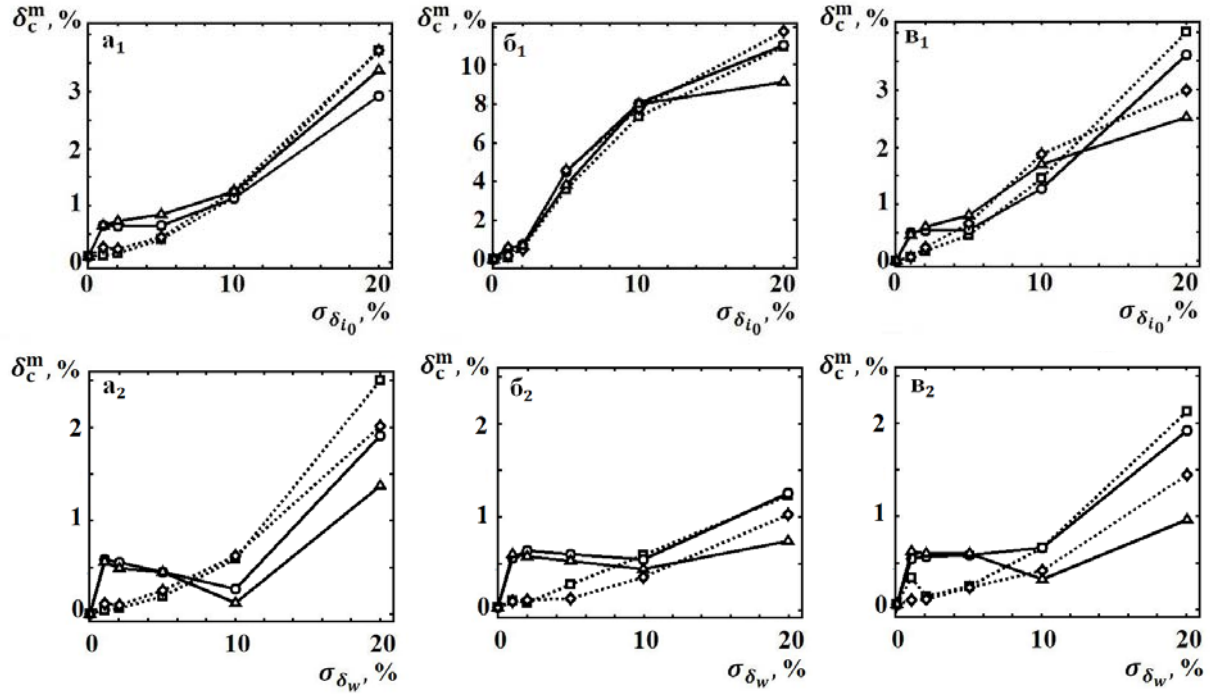

Рис. 3. ***Максимальные погрешности определения вектора концентраций по ПЛС для спектров 1 а-в без реконструкции полос поглощения индивидуальных компонентов (—) и с реконструкцией (…) для гауссовых (o, □) и лоренцевых (Δ, ◊) полос соответственно***

Учет смещений максимумов полос поглощения, обусловленных индивидуальными компонентами, при анализе таблеток Экседрин

В качестве практического приложения предложенной методики нами оценены погрешности определения вектора концентраций при анализе препарата Экседрин. Спектры Экседрина и его ингредиентов представлены на рис. 4а [7].

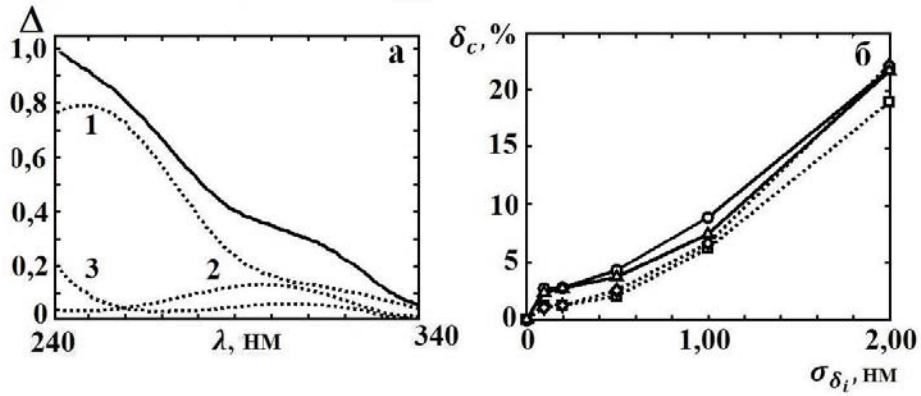

Рис. 4а. ***Спектры препарата Экседрин (—) и его ингредиентов (…) (парацетамол (1), кофеин (2), аспирин(3)), взятых в соотношении 0.4425:0.1150:0.4425 соответственно. Суммарная концентрация – 20 г/ л в смеси вода – метанол – ледяная уксусная кислота (69-28-3). Воспроизводится с разрешения http://www.dissolutiontech.com.***

Рис. 4б. ***Средние погрешности определения вектора концентраций при анализе препарата Экседрин по МНК (o ,□) и ПЛС (Δ, ◊) без реконструкции (—) и с реконструкцией (…) спектров индивидуальных компонентов***

Методом полиномиальной аппроксимации (с помощью многочлена 5-й степени) в области 241-340 нм (шаг 1 нм, m = 100) были получены функциональные выражения для дискретных спектров, позволившие осуществить их контролируемое смещение по оси абсцисс без потери крайних точек. При расчетах принималось, что форма спектров индивидуальных компонентов Экседрина остается постоянной. Спектр, обусловленный индивидуальным компонентом, смещался независимо от других спектров по оси абсцисс на величину δ_i, имеющей нормальное распределение с нулевым средним и стандартным отклонением σ_{δ_i}. Все остальные параметры численных расчетов были тождественны параметрам, используемым в предыдущем разделе. Зависимости средних погрешностей δ_c от стандартного отклонения смещений по оси абсцисс σ_{δ_i} для препарата Экседрин представлены на рис. 4б.

Выбор средних величин обусловлен относительно слабым разбросом значений δ_c при различных численных экспериментах. В отличие от случая модельных спектров, реконструкция индивидуальных спектров, обусловленных различными компонентами, из спектров смесей дала некоторые преимущества по сравнению с непосредственным использованием искаженных спектров. Неожиданным, однако, оказалось, что ПЛС-калибровка не имеет преимуществ по сравнению с МНК. Данный факт указывает на необходимость анализа погрешностей в каждой конкретной ситуации.

Как известно, погрешность спектральной калибровки современных спектрофотометров составляет ±0.2 нм. Исходя из данных, представленных на рис. 4б, можно утверждать, что ошибка определения вектора концентраций при анализе Экседрина не превышает 1.5 – 2 %. Указанная величина согласуется с опубликованными погрешностями спектрометрического анализа таблеток Экседрин [10].

Заключение

Генерирование спектров смесей тренировочного множества, основанное на использовании нескольких полос поглощения, параметры которых отягощены погрешностями, позволяет проводить компьютерное моделирование погрешности, обусловленное использованием многомерной калибровки. При относительных величинах погрешностей до 10 % относительная погрешность определения концентраций не превышала 10 % даже в случае сильно перекрывающихся полос поглощения. Предлагаемая методика позволяет в каждом конкретном случае оценить погрешность определения концентраций при варьировании параметров полос спектров.

Литература

1. Родионова, О. Е. Хемометрика: достижения и перспективы / О. Е. Родионова, А. Л. Померанцев // Успехи химии. – 2006. – Т. 75. – № 4.

2. Boulet, J.-C. Improvement of direct calibration in spectroscopy / J.- C. Boulet, J.- M. Roger // Anal. Chim. Acta. – 2010. – V. 668.

3. Elyashberg, M. E. Fuzzy structure generation: a new efficient tool for computer-aided structure elucidation (CASE) / M. E. Elyashberg, K. A. Blinov, S. G. Molodtsov, A. J. Williams, G. E. Martin // J. Chem. Inf. Model. – 2007. – V. 47.

4. Грибов, Л. А. Безэталонный молекулярный спектральный анализ. Теоретические основы / Л. А. Грибов, В. И. Баранов, М. Е. Эляшберг. – М.: URSS. – 2002. – 320 с.

5. Monakhova, Y. B. Independent components in spectroscopic analysis of complex mixtures / Y. B. Monakhova, S. A. Astakhov, A. Kraskov, S. P. Mushtakova // Chem. Intell. Lab. Syst. – 2010. – V. 103.

6. Себер, Дж. Линейный регрессионный анализ / Дж. Себер. – М.: Мир, 1980. – 456 с.

7. Simonian, M. H. Full spectrum quantitation: an advance UV/Visible spectroscopic method for multi-component dissolution testing / M. H. Simonian // http://www.dissolutiontech.com/DTresour/1995 Articles /DT199505_A02.pdf.

8. Воеводин, В. В. Матрицы и вычисления / В. В. Воеводин, Ю. А. Кузнецов. – М: Наука, 1984. – 320 с.

9. Danzer, K. Analytical chemistry. Theoretical and metrological fundamentals / K. Danzer. – Springer, 2005. – 315 p.

10. Özdemir, A. Utilization of multivariate calibration techniques for the spectrophotometric simultaneous determination of paracetamol, aspirin and caffeine in pharmaceutical formulation / A. Özdemir , E. Dinc, F. Onur // Turk. J. Pharm. Sci**. –** 2004. – V. 1.