

ПРОГНОЗИРОВАНИЕ ПОГРЕШНОСТЕЙ ФОТОМЕТРИИ С ИСПОЛЬЗОВАНИЕМ ЗАКОНА НАКОПЛЕНИЯ ОШИБОК И МЕТОДА МОНТЕ-КАРЛО

В.И. Голованов, Е.И. Данилина

В вычислительном эксперименте, при сочетании закона распространения ошибок и метода Монте-Карло, исследовано влияние погрешностей приготовления растворов, погрешностей холостого опыта и погрешностей измерения пропускания на метрологические характеристики фотометрического анализа. Найдено, что результаты прогнозирования погрешностей аналитическим и статистическим методами взаимосогласованы. Показано, что особенностью метода Монте-Карло является возможность прогнозирования закона распределения погрешностей в фотометрии. На примере сценария рутинного анализа рассмотрено влияние гетероскедастичности разброса вдоль градуировочного графика на качество анализа.

Ключевые слова: фотометрический анализ, закон накопления ошибок, градуировочный график, метрологические характеристики, метод Монте-Карло, стохастическое моделирование.

Введение

Прогнозирование погрешностей фотометрического анализа в основном строят на использовании закона накопления ошибок (ЗНО) [1, 2]. Для случая линейной формы закона светопоглощения: $-\lg T = A = \epsilon l c$, ЗНО обычно записывают уравнением:

$$\frac{s_A}{A} = \frac{s_c}{c} = -\frac{0,434 \cdot 10^A}{A} s_T. \quad (1)$$

При этом стандартное отклонение измерения степени пропускания предполагается постоянным во всем динамическом интервале фотометра. Вместе с тем, как отмечают в [1], кроме аппаратурных погрешностей на точность анализа влияют погрешность холостого опыта, погрешность настройки пределов шкалы прибора, кюветная погрешность, химические факторы, погрешность установки аналитической длины волны. Эти факторы считают основными источниками погрешности результата анализа. Вкладами в накопленную погрешность точности приготовления градуировочных растворов обычно пренебрегают.

Отсюда видим, что уравнение (1) не имеет существенной прогностической силы, поскольку учитывает влияние только одного фактора. Кроме того, уравнение (1) является следствием приближенного разложения закона светопоглощения в ряд Тейлора. Отсюда возникает вопрос о его точности, обусловленной пренебрежением членами разложения выше первого порядка. Математический анализ остатков разложения сопряжен с вычислительными трудностями и в практике химического анализа не применяется.

Целью данной работы является изучение возможности применения метода Монте-Карло (метода статистических испытаний) [3] в качестве независимого метода для изучения и прогнозирования накопления погрешностей фотометрического анализа, дополняющего и углубляющего возможности ЗНО.

Теоретическая часть

В этой работе будем полагать, что итоговая случайная погрешность градуировочной функции обусловлена не только инструментальными погрешностями измерения оптической плотности, но и погрешностями настройки шкалы прибора на 0 и 100 % пропускания (погрешность хо-

лостого опыта), а также погрешностями приготовления градуировочных растворов. Другими, названными выше, источниками погрешностей пренебрегаем. Тогда перепишем уравнение закона Бугера-Ламберта-Бэра в удобной для дальнейшего построения форме:

$$A = \varepsilon c_{st} \frac{V_a}{V_{fl}} + A_{bl} = k \frac{V_a}{V_{fl}} + A_{bl} = kc' + A_{bl}. \quad (2)$$

В этом уравнении c_{st} - концентрации головного стандартного раствора окрашенного вещества, аликвоты (V_a) которого разбавляют в колбах с номинальным объёмом V_{fl} для получения градуировочной серии растворов, A_{bl} - оптическая плотность раствора холостого опыта. Поскольку при фотометрировании оптическую плотность испытуемых растворов измеряют относительно холостого раствора, т. е. A_{bl} принимают за условный ноль, то $A_{bl} = 0$. (Заметим, что измеренное при этом значение оптической плотности можно называть условной экстинкцией.) В уравнении (2) безразмерная величина c' имеет смысл концентрации рабочего раствора, выраженной в единицах концентрации головного стандарта. Коэффициент k назовем экстинкцией стандарта, поскольку $A_{st} = \varepsilon c_{st}$ при $c' = 1$.

Применим к выражению (2) оператор закона накопления случайных ошибок, полагая V_a , V_{fl} и A_{bl} случайными величинами. Получаем:

$$s_{A'}^2 = s_{A_{bl}}^2 + \left(\frac{k}{V_{fl}}\right)^2 s_{V_a}^2 + \left(\frac{kV_a}{V_{fl}^2}\right)^2 s_{V_{fl}}^2. \quad (3)$$

Еще одной независимой случайной величиной, влияющей на разброс значений A , является степень пропускания, поскольку

$$A = -\lg T, \quad (4)$$

поэтому к дисперсиям в левой части уравнения (3) добавляем еще одно слагаемое:

$$s_A^2 = (0,434 \cdot 10^A)^2 s_T^2 + s_{A_{bl}}^2 + \left(\frac{kV_a}{V_{fl}}\right)^2 s_{r,V_a}^2 + \left(\frac{kV_a}{V_{fl}^2}\right)^2 s_{V_{fl}}^2. \quad (5)$$

В этой окончательной записи закона накопления ошибок постоянны абсолютные стандартные отклонения T , A_{bl} и V_{fl} , а для V_a постоянна относительная стандартная погрешность.

При построении стохастической модели градуировочной функции на основе метода Монте-Карло считаем, что возможные значения x_i случайных величин T , A_{bl} , V_a и V_{fl} , распределены по нормальному закону. Согласно принципу Монте-Карло, возможные значения будем разыгрывать по методу обратной функции [3]:

$$x_i = M(x_i) + F^{-1}(r_j) \cdot s_{x_i}, \quad (6)$$

где $M(x)$ - математическое ожидание (действительное значение) переменной, $F(ij)$ - функция Лапласа-Гаусса, ij - возможные значения равномерно распределенной на интервале (0,1) случайной величины R , т. е. случайные числа, s_x - стандартное отклонение соответствующей переменной, $i = 1 \dots m$ - порядковый номер независимой случайной величины. После подстановки выражения (6) в уравнения (4) и (2) имеем:

$$A' = -\lg x_1 = -\lg 10^{-A'} + F^{-1}(r_j) s_T, \quad (7)$$

$$\text{где } A'' = k \frac{x_3}{x_4} + x_2.$$

Вычисления по уравнению (7) возвращают отдельную реализацию градуировочной функции, т.е. зависимость A' от математического ожидания $M(c')$ (номинального значения c'). Поэтому запись (7) является аналитическим выражением случайной функции [3]. Сечения этой функции получают при многократном разыгрывании случайных чисел в каждой точке градуировочной зависимости. Выборочную совокупность реализаций обрабатывают методами математической статистики с целью оценивания генеральных параметров градуировки и проверки гипотез о свойствах генеральной совокупности.

Очевидно, что рассматриваемые нами два подхода к проблеме прогнозирования метрологических характеристик в фотометрии - на основе ЗНО, с одной стороны, и на основе метода Монте-Карло, с другой, должны дополнять друг друга. В частности, из уравнения (5) можно получить результат при гораздо меньшем, по сравнению с (7), объёме вычислений, а также проранжиро-

вать случайные величины по значимости их вкладов в результирующую погрешность. Ранжирование позволяет отказаться от отсеивающего эксперимента при статистических испытаниях и априори исключить из рассмотрения малозначимые переменные. Уравнение (5) несложно проанализировать математически для того, чтобы судить о характере вкладов факторов в общую дисперсию. Частные вклады факторов можно подразделить на независимые от A , либо возрастающие с увеличением оптической плотности. Поэтому s_A как функция A должна быть монотонно возрастающей зависимостью, лишенной минимума. При аппроксимации экспериментальных данных уравнением (5) частные вклады одинакового характера будут смешиваться, например, кюветная погрешность может смешиваться с погрешностью холостого опыта. С другой стороны, при статистических испытаниях модели методом Монте-Карло можно выявить такие важные свойства градуировочного графика как закон (законы) распределения погрешностей, а также оценить быстроту сходимости выборочных оценок к генеральным. На основе ЗНО такой анализ невозможен.

Описание вычислительного эксперимента

При построении имитационной модели градуирования полагаем, что градуировочная серия растворов приготовлена в мерных колбах с номинальной вместимостью 50 мл и предельной погрешностью $\pm 0,05$ мл. В серию колб добавляют от 1 до 17 мл головного стандартного раствора с погрешностью пипетирования $\geq 1\%$. Погрешности измерения объёмов оценивали по справочнику [4]. Аликвоты вносят с равномерным шагом 1 мл. Всего в серии 17 растворов, оптическая плотность которых охватывает интервал от 0,1 до 1,7 ед. Тогда в уравнении (2) коэффициент $k = 5$. Погрешность холостого опыта принимаем на уровне 0,01 ед. оптической плотности. Погрешности измерения степени пропускания, согласно [1], зависят только от класса прибора и находятся в интервале от 0,1 до 0,5 % T .

Для большей привязки условий вычислительного эксперимента к лабораторному эксперименту воспользовались данными [1] по воспроизводимости измерений оптических плотностей растворов $K_2Cr_2O_7$ в присутствии 0,05 М H_2SO_4 спектрофотометре СФ-26. Авторы [1] аппроксимируют экспериментальные данные на интервале $A = 0,1 \dots 1,5$ уравнением параболы:

$$s_{\text{восп}} \cdot 10^3 = 7,9 - 3,53A + 10,3A^2. \quad (8)$$

Нам удалось подогнать расчеты по теоретическому уравнению (5) к расчетам по эмпирическому уравнению (8) с использованием оптимизационного метода Ньютона. Нашли, что уравнение (5) удовлетворительно описывает эксперимент при $s(T) = 0,12\%$, $s(A_{bl}) = 0,007$ и $s_r(V_a) = 1,1\%$.

Приведенные в предыдущем абзаце независимые оценки погрешностей хорошо согласуются с найденными при подгонке. Для вычислений по уравнению (7) создана программа в виде формы листа электронных таблиц MS Excel. Наиболее существенной особенностью нашей Excel-программы является использование выражения НОРМСТОБР(СЛЧИС()) для генерирования нормально распределенных погрешностей, см. уравнение (6). В специальной литературе по статистическим вычислениям в Excel [5] подробно описана утилита «Генерация случайных чисел», которую во многих случаях предпочтительно заменять на функции типа НОРМСТОБР(СЛЧИС()). Такая замена особенно удобна при создании собственных программ для моделирования методом Монте-Карло.

Результаты и их обсуждение

Прежде чем приступить к статистическим испытаниям, оценим вклады слагаемых в левой части уравнения (5) в общую дисперсию оптической плотности. Для этого каждое слагаемое нормируют на общую дисперсию. Расчеты выполнены при $s(T) = 0,12\%$, $s(A_{bl}) = 0,007$, $s_r(V_a) = 1,1\%$ и $s(V_{fl}) = 0,05$. Результаты вычислений показаны на рис. 1. Видим, что вкладами в общую дисперсию погрешностей измерения V_{fl} можно пренебречь.

Тогда как вклады другой, влияющей на погрешности приготовления растворов, величины V_a доминируют в интервале оптических плотностей 0,8... 1,2. Однако этот вывод не имеет общего характера, поскольку при измерениях на фотометре с $s(T) = 0,5\%$ погрешности градуировки, согласно расчету, определяются главным образом, разбросом A_{bl} и разбросом T . На рис. 2 сравниваются относительные ошибки измерений оптических плотностей, прогнозируемых на основе ЗНО (сплошная линия) и метода Монте-Карло (значки). При статистических испытаниях кривую

ошибок восстанавливали по 100 реализациям градуировочной зависимости (1700 значениям оптических плотностей). Видим, что оба прогноза взаимно согласованы. Точки равномерно группируются возле теоретической кривой. Однако даже при таком, довольно внушительном, статистическом материале полной сходимости не наблюдается. Во всяком случае, разброс не позволяет выявить приближенный характер ЗНО, см. введение.

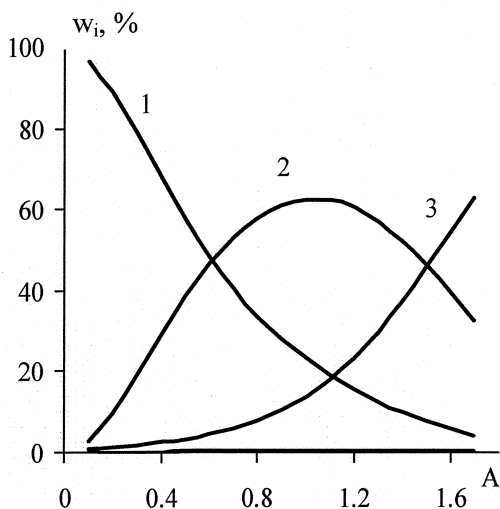


Рис. 1. Весовые вклады слагаемых уравнения (5) в дисперсию A : 1 – для $A_{\text{в}}$; 2 – для V_a ; 3 – для T ; 4 – для $V_{\text{н}}$

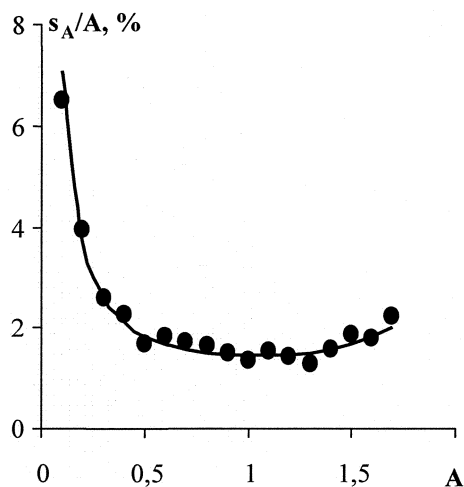


Рис. 2. Кривая погрешностей градуировочного графика

Из теории математической статистики известно [3], что при интервальном оценивании математического ожидания случайной величины надежность оценивания повышается, если известен закон распределения для этой величины. Кроме того, в случае нормального распределения оценка является наиболее эффективной. Поэтому исследование закона распределения погрешностей градуировочного графика является важной задачей. При таком исследовании, прежде всего, проверяют гипотезу нормальности разброса оптических плотностей в отдельных точках графика.

Простым способом проверки основной гипотезы являются вычисления коэффициентов асимметрии (a) и коэффициентов эксцесса (e) эмпирических распределений, а также их сравнение с критериальными значениями. Надежность статистических выводов повышается при увеличении объема выборочных данных. На рис. 3 приведены последовательности коэффициентов для 17 сечений градуировочной функции. Коэффициенты вычислены по результатам 100 испытаний в каждой точке. Критические значения коэффициентов для нашего примера равны $|a| = 0,72$ и $|e| = 0,23$.

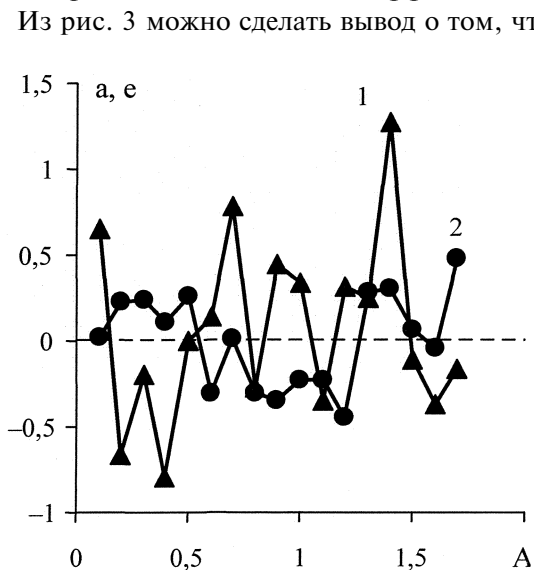


Рис. 3. Коэффициенты эксцесса (1) и коэффициенты асимметрии (2) в точках градуировочного графика

Из рис. 3 можно сделать вывод о том, что рассеяние значений в точках графика, в целом, не противоречит гипотезе нормальности, поскольку последовательности коэффициентов почти не имеют предпочтительной направленности. Коэффициенты случайным образом локализуются вблизи нулевой линии (показана пунктиром). Для нормального распределения, как известно, математическим ожиданием коэффициента асимметрии и коэффициента эксцесса является нуль. Судя по тому, что при всех сечениях коэффициенты асимметрии существенно ниже критического значения, можно уверенно говорить о симметричности распределения погрешностей градуирования. Возможно, что распределения погрешностей обладают небольшой остроконечностью по сравнению с нормальной кривой распределения. Этот вывод следует из наблюдающегося на рис. 3 небольшого поло-

жительного смещения центральной линии рассеяния коэффициентов эксцесса. Таким образом, из исследования модели обобщенной градуировочной функции фотометрического анализа методом Монте-Карло (2), можно сделать вывод о близком к нормальному распределению погрешностей градуирования. Поэтому вычисления доверительных интервалов для результатов фотометрического анализа с использованием коэффициентов Стьюдента можно считать вполне оправданными.

При выполнении стохастического моделирования была оценена быстрота сходимости выборочных кривых погрешностей (см. рис. 2) к математическому ожиданию кривой. За математическое ожидание кривой погрешностей примем рассчитанную из ЗНО кривую. Близость результатов статистических испытаний с различным числом реализаций градуировки n к теоретической кривой оценим коэффициентом неопределенности $1 - R^2$. Этот коэффициент характеризует долю вариации в выборке, которую не удалось описать теоретически. Нами установлено, что зависимость коэффициента неопределенности от числа реализаций градуировочной функции можно описать эмпирическим уравнением $1 - R^2 = -2,3n^{-1} + 1,6n^{-1/2} - 0,1$. Из уравнения получаем, что при $n = 213$ следует ожидать практически полного совпадения теоретической и эмпирической кривых погрешностей. Таким образом, состоятельную оценку погрешностей фотометрического анализа можно получить только на довольно большом статистическом материале.

Рассмотрим возможности метода статистических испытаний для прогнозирования результатов регрессионного анализа градуировочного графика и использования графика при определении концентраций фотометрируемых растворов. Для этого в качестве сценария выберем измерительную ситуацию рутинного анализа. Построение графика осуществляют при однократных измерениях оптических плотностей серии стандартных растворов. Концентрацию анализируемого раствора находят из графика по 3-4 результатам параллельных измерений. При выборе регрессионной модели следует принять во внимание то, что разброс оптических плотностей в различных точках градуировочного графика неодинаков, см. уравнение (8). В случае гетероскедастичного разброса рекомендуют [1] использовать схему взвешенного метода наименьших квадратов (ВМНК). Однако в литературе мы не встретили четких указаний на причины, по которым классическая схема МНК, одним из условий применимости которой является требование гомоскедастичности разброса, менее предпочтительна. Эти причины можно установить при обработке одного и того же статистического материала, полученного методом Монте-Карло по сценарию рутинного анализа, двумя вариантами МНК - классическим и взвешенным.

В результате регрессионного анализа только одной реализации градуировочной функции получены следующие МНК-оценки: $k = 4,979$ при $S_k = 0,023$. При оценке тех же характеристик ВМНК получаем $k = 5,000$ при $s_k = 0,016$. Регрессии восстанавливали по 17 стандартным растворам. Концентрации в градуировочной серии возрастали в арифметической прогрессии, а оптические плотности изменялись столь же равномерно на интервале от 0,1 до 1,7 ед. В случае ВМНК статистические веса точек градуировочного графика находили с использованием рассчитанных по уравнению (5) дисперсий.

Дисперсии оценок k тем и другим методом статистически неразличимы по критерию Фишера при 1 %-ном уровне значимости. Однако при том же уровне значимости МНК-оценка k отличается от ВМНК-оценки по t -критерию. МНК-оценка коэффициента градуировочного графика смещена относительно действительного значения $M(k) = 5,000$, судя по t -тесту при 5 %-ном уровне значимости. Тогда как взвешенный МНК дает оценку, которая не содержит систематической погрешности.

Теперь выясним, каким образом пренебрежение гетероскедастичностью может отразиться на качестве химического анализа. В таблице приведены результаты имитационного эксперимента по анализу 17 контрольных проб окрашенного вещества с различной концентрацией. Причем каждая аналитическая серия включала четыре раствора, т.е. для каждой пробы выполнено по четыре параллельных определения. Для обработки результатов использовали две разные градуировочные зависимости: одна была восстановлена простым МНК, а вторая - взвешенным. Полагаем, что контрольные растворы готовились для анализа точно так же, как градуировочные.

Из таблицы видим, что действительные значения концентраций контрольных растворов как в случае ВМНК, так и в случае МНК не выходят за пределы доверительных интервалов, т.е. результаты анализа не содержат значимых систематических погрешностей. Предельные погрешности того и другого метода статистически не различаются, иными словами, и та и другая оценка

Сравнение результатов определения концентраций контрольных растворов двумя методами

№ п/п	с', задано	с', найдено (P = 95 %)	
		МНК	ВМНК
1	0,020	0,021±0,002	0,021±0,002
2	0,040	0,041±0,001	0,041±0,001
3	0,060	0,061±0,003	0,061±0,003
4	0,080	0,080±0,004	0,080±0,004
5	0,100	0,098±0,004	0,098±0,004
6	0,120	0,122±0,006	0,121±0,006
7	0,140	0,140±0,006	0,139±0,006
8	0,160	0,163±0,003	0,162±0,003
9	0,180	0,181±0,006	0,180±0,006
10	0,200	0,201±0,002	0,200±0,002
11	0,220	0,219±0,008	0,218±0,008
12	0,240	0,242±0,002	0,241±0,002
13	0,260	0,262±0,008	0,261±0,008
14	0,280	0,281±0,010	0,280±0,010
15	0,300	0,307±0,015	0,306±0,015
16	0,320	0,325±0,013	0,323±0,013
17	0,340	0,340±0,026	0,339±0,026

ках градуировки статистические веса различаются более чем в 20 раз. Однако обратим внимание на то, что градуировочные функции восстанавливали по 17 точкам графика, тогда как при анализе выполняли только 4 параллельных определения. Поэтому обнаруженное нами значимое различие МНК и ВМНК градуировочных функций и незначительное различие результатов анализа с использованием этих функций можно объяснить существенно различным числом степеней свободы, которыми располагали при построении статистических выводов.

Заключение

1. Предложен новый подход к стохастическому моделированию в фотометрическом анализе на основе метода Монте-Карло и закона накопления ошибок с использованием табличного процессора Excel.

2. По 100 реализациям градуировочной зависимости показано, что прогнозирование погрешностей аналитическим и статистическим методом взаимосогласованы.

3. Изучены коэффициенты асимметрии и эксцесса вдоль градуировочного графика. Найдено, что вариации погрешностей градуирования подчиняются закону распределения, близкому к нормальному.

4. Рассмотрено влияние гетероскедастичности разброса оптических плотностей при градуировании на качество анализа. Обнаружено, что при рутинных анализах использование простой схемы невзвешенного МНК не приводит к заметному снижению точности результатов анализа.

Литература

1. Бернштейн, И.Я. Спектрофотометрический анализ в органической химии / И.Я. Бернштейн, Ю.Л. Каминский. - Л.: Химия, 1986. - 200 с.
2. Булатов, М.И. Практическое руководство по фотометрическим методам анализа / М.И. Булатов, И.П. Калинин. - Л.: Химия, 1986. - 432 с.
3. Гмурман, В.Е. Теория вероятностей и математическая статистика / В.Е. Гмурман. - М.: Высшая школа, 1977. - 470 с.

4. Правдин, П.В. Лабораторные приборы и оборудование из стекла / П.В. Правдин. - М.: Химия, 1988. - 336 с.

5. Макарова, Н.В. Статистика в Excel / Н.В. Макарова, В.Я. Трофимец. - М.: Финансы и статистика, 2002. - 368 с.

Поступила в редакцию 27 января 2010 г.

PREDICTION OF ERRORS IN PHOTOMETRY WITH THE USE OF ACCUMULATION OF ERRORS LAW AND MONTE CARLO METHOD

During computing experiment, in combination of the accumulation of errors law and Monte Carlo method, the influence of solution-making errors, blank experiment errors and optical transmission measurement errors upon metrological performance of photometrical analysis has been studied. It has been shown that the results of prediction by analytical and statistical methods are interconsistent. The unique feature of Monte Carlo method has been found to enable prediction of the accumulation of errors law in photometry. For the version of routine analysis the influence of heteroscedasticity of dispersion along calibration curve upon analysis quality has been studied.

Keywords: photometric analysis, accumulation of errors law, calibration curve, metrological performance, Monte Carlo method, stochastic modeling.

Golovanov Vladimir Ivanovich - Dr. Sc. (Chemistry), Professor, Head of the Analytical Chemistry Subdepartment, South Ural State University.

Голованов Владимир Иванович — доктор химических наук, профессор, заведующий кафедрой Аналитической химии, Южно-Уральский государственный университет.

E-mail: vigchel@hotmail.ru

Danilina Elena Ivanovna - PhD (Chemistry), Associate Professor, Analytical Chemistry Subdepartment, South Ural State University.

Данилина Елена Ивановна - кандидат химических наук, доцент, кафедра «Аналитическая химия», Южно-Уральский государственный университет.

E-mail: deicu@mail.ru