

# МОДЕЛИРОВАНИЕ СТРУКТУРЫ ДАННЫХ ПРИ ИСПОЛЬЗОВАНИИ МНОГОМЕРНОЙ ГРАДУИРОВКИ В АТОМНО-ЭМИССИОННОЙ СПЕКТРОМЕТРИИ

**И.Е. Васильева, Е.В. Шабанова**

*Институт геохимии им. А.П. Виноградова СО РАН,  
Иркутск, 664033, ул. Фаворского 1А  
e-mail: [vasira@igc.irk.ru](mailto:vasira@igc.irk.ru)*

Поступила в редакцию 13 января 2009 г.

Применение многомерного статистического анализа данных (МСА) в атомно-эмиссионной спектроскопии (АЭС) является перспективным, так как позволяет учесть изменения параметров плазмы, влияющих на интенсивность спектральных линий при получении дуговых атомно-эмиссионных спектров и ухудшающих точность результатов анализа. С целью учета спектральных и неспектральных эффектов в АЭС изучены условия применения  $n$ -мерных градуировок, использующих регрессию на главных компонентах (РГК), при моделировании структуры таблицы экспериментальных данных в методах определения золота и платины в нерастворимом углеродистом веществе и примесей бора в кристаллическом кремнии и кварце. Для выбора наилучшей градуировочной зависимости из рассматриваемых вариантов  $n$ -мерных градуировок использована оценка "размах относительной систематической погрешности на интервале определяемых содержаний".

**Ключевые слова:** атомно-эмиссионная спектроскопия с дуговым разрядом, многомерная градуировка, регрессия на главных компонентах

**Васильева Ирина Евгеньевна** – заведующая лабораторией оптического спектрального анализа и стандартных образцов ИГХ СО РАН, доктор технических наук  
**Область научных интересов:** аналитическая химия, спектральные методы анализа

**Шабанова Елена Владимировна** – научный сотрудник ИГХ СО РАН, кандидат технических наук

**Область научных интересов:** многоэлементные спектральные методы анализа, хемометрика

Рекомендации по применению многомерного статистического анализа данных (МСА) для конкретных задач во многих предметных областях, в том числе в атомно-эмиссионной спектроскопии (АЭС), отсутствуют. В то же время, использование методов МСА для градуировки методик в АЭС является перспективным, так как не требует устранения факторов, влияющих на интенсивность спектральных линий при получении атомно-эмиссионных спектров и ухудшающих точность результатов анализа. В настоящей работе с целью учета спектральных и неспектральных эффектов в АЭС изучены условия применения  $n$ -мерных градуировок, использующих регрессию на главных компонентах (РГК), при моделировании структуры таблицы экспериментальных данных, и предложен критерий выбора лучшей градуировочной зависимости.

## ТЕОРИЯ

Основу атомно-эмиссионной спектроскопии при использовании дугового разряда как

источника возбуждения спектра составляет однозначная связь между входной информацией, то есть химическим составом пробы, и выходными аналитическими сигналами или интенсивностями спектральных линий элементов, которая описывается эмпирической формулой Ломакина-Шейбе [1, 2]

$$I = aC^b, \quad (1)$$

где  $I$  – интенсивность спектральной линии,  $C$  – содержание элемента,  $a$  и  $b$  – некоторые постоянные, зависящие от свойств излучающих атомов и способа их возбуждения. Преобразование уравнения через логарифмирование описывает линейную связь между переменными на интервале содержания определяемого элемента в 1,5-2,5 порядка. Известно, что интенсивность спектральной линии аналита зависит от:

- его содержания в образце;
- общего химического состава образца, то есть содержаний макроэлементов, влияющих на параметры плазмы дугового разряда;

- содержаний интерферирующих элементов;
- наличие молекулярных полос в спектре.

Рассматривая атомно-эмиссионный спектр как многомерное наблюдение, можно на качественном уровне описать влияние перечисленных признаков на интенсивность линий аналита в спектрах градуировочных образцов, набор которых рассматривается как обучающая выборка. Однако теоретически невозможно рассчитать с удовлетворительной точностью степень влияния этих факторов на интенсивность линий в спектрах новых проб [2].

В методиках атомно-эмиссионного анализа градуировочная зависимость (**ГЗ**) задается как некоторая регрессионная функция для массива аналитических сигналов ( $I - Y$ ) и содержаний элементов ( $C - X$ ) в спектрах обучающей выборки. Коэффициенты ГЗ вычисляются и оцениваются обычно с помощью метода наименьших квадратов (**МНК**). Естественно, что любая функция отклика включает случайные шумы измерительной системы (пробоподготовки и спектрометра) и систематическую погрешность. Благодаря использованию регистраторов спектра на основе твердотельных детекторов изображения и высокостабильных дуговых генераторов [3, с. 4-40] современные спектрометры вносят минимальные по величине случайные погрешности при измерении аналитических сигналов (в  $Y$ -переменных). Следовательно, главными причинами изменения интенсивности линий элемента в спектрах различных проб с его одинаковым содержанием являются в основном матричные (неспектральные) и интерференционные (спектральные) эффекты, которые проявляются как систематическая погрешность в аналитических результатах. Поэтому, задав вид функции отклика и вычислив коэффициенты ГЗ, можно рассчитать содержание аналита в образцах обучающей выборки и оценить долю вклада систематической погрешности в аналитический результат. Чтобы учесть зависимость величины погрешности от содержания определяемого элемента, в качестве оценки используем относительную систематическую погрешность (**ОСП**) результата определения аналита в каждом  $j$ -том образце обучающей выборки, рассчитываемую по формуле (2)

$$R_j = \frac{C_{выч} - C_{ам}}{C_{ам}} \times 100\%, \quad (2)$$

где  $C_{выч}$  и  $C_{ам}$  – это вычисленное и аттестованное содержания соответственно. Для всего набора градуировочных образцов на интервале  $\Delta C_{train} = C_{max} - C_{min}$  по вычисленным значениям  $R_j$  рассчитывается размах ОСП (**РОСП**) по формуле (3)

$$\Delta R = R_{max} - R_{min}. \quad (3)$$

Чем меньше значение  $\Delta R$ , тем лучше модельная функция аппроксимирует экспериментальные данные обучающей выборки на интервале обучения, то есть

$$\Delta R_{train} \rightarrow \min. \quad (4)$$

В данной интерпретации мы рассматриваем размах ОСП как характеристику всего диапазона содержаний элемента  $\Delta C$ , а не отдельной точки, описывающей массовую долю аналита в каждом градуировочном образце  $C_j$  [4, 5]. Поэтому предлагаемая оценка (РОСП –  $\Delta R$ ) позволяет исследовать влияние структуры признаков, включенных в обучающую выборку, на качество градуировки в интервале  $\Delta C_{train}$ . С этой точки зрения размах ОСП можно считать характеристикой влияния различий химического состава образцов на предсказательную способность ГЗ (функции отклика).

Предсказательную способность дополнительно можно оценить по данным тестовой выборки. Тестирующая выборка, по спектрам которой рассчитываются аналитические сигналы, а затем содержания аналита с использованием модельной ГЗ, также состоит из образцов с известными содержаниями аналита. В неё включаются спектры образцов, которые не были использованы в обучающей выборке, или новые спектры зашифрованных образцов из градуировочного набора (обучающей выборки). В интервале  $\Delta C_{test} = C_{max} - C_{min}$  тестовая выборка показывает РОСП равный  $\Delta R_{test}$  (формула (3)).

В аналитической химии качество результата оценивается величинами случайной и систематической погрешностей, которые минимизируются при выборе методических условий выполнения анализа. Поэтому при одинаковых случайных погрешностях, которые определены операционными условиями методики, хорошее качество градуировочной модели описывается общим условием минимизации систематической погрешности для обучающей и тестирующей выборок

$$\begin{cases} \Delta R_{train} \rightarrow \min, \\ \Delta R_{test} \rightarrow \min. \end{cases} \quad (5)$$

Величины размахов ОСП указывают на способность градуировочной модели учесть разнообразие макросоставов стандартных образцов (**СО**) в наборах образцов обучающей и тестирующей выборок. При этом желательно выполнение условия равенства диапазонов  $\Delta C_{train} \sim \Delta C_{test}$ , то есть обучающая и тестирующая выборки должны содержать идентичные по макросоставу наборы образцов одинакового количества. Выполнение этого условия часто практически невозможно из-за отсутствия достаточного количества стандартных образцов и экономически невыгодно из-за высокой стоимости

многоэлементных СО и труднодоступности их материала. Поэтому тестовый набор обычно меньше, чем обучающий, и может включать новые спектры образцов из обучающей выборки. Это не противоречит стратегии сегментирования набора данных, хотя существует точка зрения, что "можно обойтись и без проверки прогнозных возможностей модели, если модель строится для того, чтобы понять взаимосвязи в рассматриваемой системе, или для изучения возможностей использования многомерного моделирования для будущих косвенных измерений" [5, с. 104]. Мы считаем полезным вычисление размаха ОСП для  $X$ -переменной обучающей и тестирующей выборок и их сопоставление для оценки, как качества градуировки, так и прогнозных возможностей модели и её структуры (формулы (2)-(5)).

## ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

Спектры образцов обучающей и тестирующей выборок были получены при полном испарении 10 мг вещества из канала графитового электрода в вертикальной дуге постоянного тока. Была использована спектральная установка, которая включает дифракционный спектрограф ДФС-458, электродуговой генератор "Везувий" и многоканальный анализатор эмиссионных спектров МАЭС [3]. Прибор МАЭС укомплектован коммерческим программным обеспечением АТОМ для регистрации и обработки спектров. Программа АТОМ обеспечивает аппроксимацию по экспериментальным данным одномерных регрессионных зависимостей 1-ой (**МНК-1**) и 2-ой (**МНК-2**) степеней и вычисляет их статистические характеристики. Опытным путем установлено, что наблюдаемые значения относительного стандартного отклонения для измерений спектральной интенсивности не превышают 5 % отн.

Метод регрессии на главных компонентах (**РГК**) [5-7] используется для аппроксимации экспериментальных данных  $n$ -мерной градуировочной моделью. Ограничением РГК является хорошее описание только линейных связей. Градуировка с помощью РГК требует большого количества образцов, в которых определены содержания всех участвующих в расчете элементов, т.е. в таблице данных отсутствуют "пустые" клетки. Расчеты выполняются в специальной программе PCR-Calibration, которая входит в модуль построения градуировочных зависимостей и подсчета содержаний в образцах [8]. Для исследования предсказательной способности РГК-градуировки в зависимости от структуры данных в программе АТОМ создавались таблицы с различным числом спектральных линий тех или иных микро- и макроэлементов. После измерения в программе АТОМ интенсивностей заданных спектральных линий каждая таблица с новой структурой данных экспортировалась в программу PCR-Calibration

для расчета коэффициентов ГЗ и содержаний аналитов в образцах обучающей и тестовой выборок. РГК-градуировки, полученные для каждой структуры данных, сравнивались между собой по величине размахов ОСП.

Для учета вклада случайных погрешностей (шумов) в РОСП было введено условие, что отличие величины РОСП двух или более градуировок не более чем на 5 % обусловлено случайными погрешностями.

## ЧИСЛЕННЫЙ ЭКСПЕРИМЕНТ И ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Моделирование структуры спектральных данных и построение  $n$ -мерных градуировок было выполнено для двух аналитических методик:

1. Определение золота и платины в нерастворимом углеродистом веществе.
2. Определение примесей в кристаллическом кремнии и кварце (на примере бора).

### 1. Определение золота и платины в нерастворимом углеродистом веществе

Нерастворимое углеродистое вещество (**НУВ**) содержит более 50 % углерода и концентрирует благородные металлы в черных сланцах [9, 10]. Для снижения несектрального (матричного) влияния градуировочные образцы были приготовлены на основе графитового порошка. Градуировочные образцы (отмечены буквой **S**) приготовлены последовательным разбавлением образца с содержанием золота и платины 0,5 мас. %. Образцы с маркировкой **GP** приготовлены накопыванием стандартных растворов золота и платины на графитовый порошок. Образцы СОП-Аu, СНР- $Pt^+$ , ВП-2 являются стандартными образцами руд.

В табл. 1 показаны независимые признаки обучающих и тестирующих выборок. Строки таблицы – это названия образцов. Столбцы таблицы ( $X$ -переменные) – это концентрации золота и платины. Как зависимые признаки выбраны наиболее чувствительные линии аналитов, имеющие близкие пределы обнаружения. Для золота это линии Au 267,595 нм и Au 242,795 нм, обозначенные 1Au и 2Au соответственно (табл. 2). В табл. 3 показаны результаты моделирования при использовании в градуировках линий платины: Pt 265,945 нм (1Pt); Pt 306,471 нм (2Pt); Pt 299,797 нм (3Pt). В таблицу экспериментальных данных мы не включили столбцы ( $Y$ -переменные), отвечающие за влияние основы и интерферентов. Это объясняется двумя причинами. Во-первых, аналитические линии золота и платины свободны от наложений линий интерферентов и молекулярных полос (рис. 1). Во-вторых, матричное влияние углерода не рассматривается, так как его содержание в градуировочных и образцах НУВ примерно одинаковое.

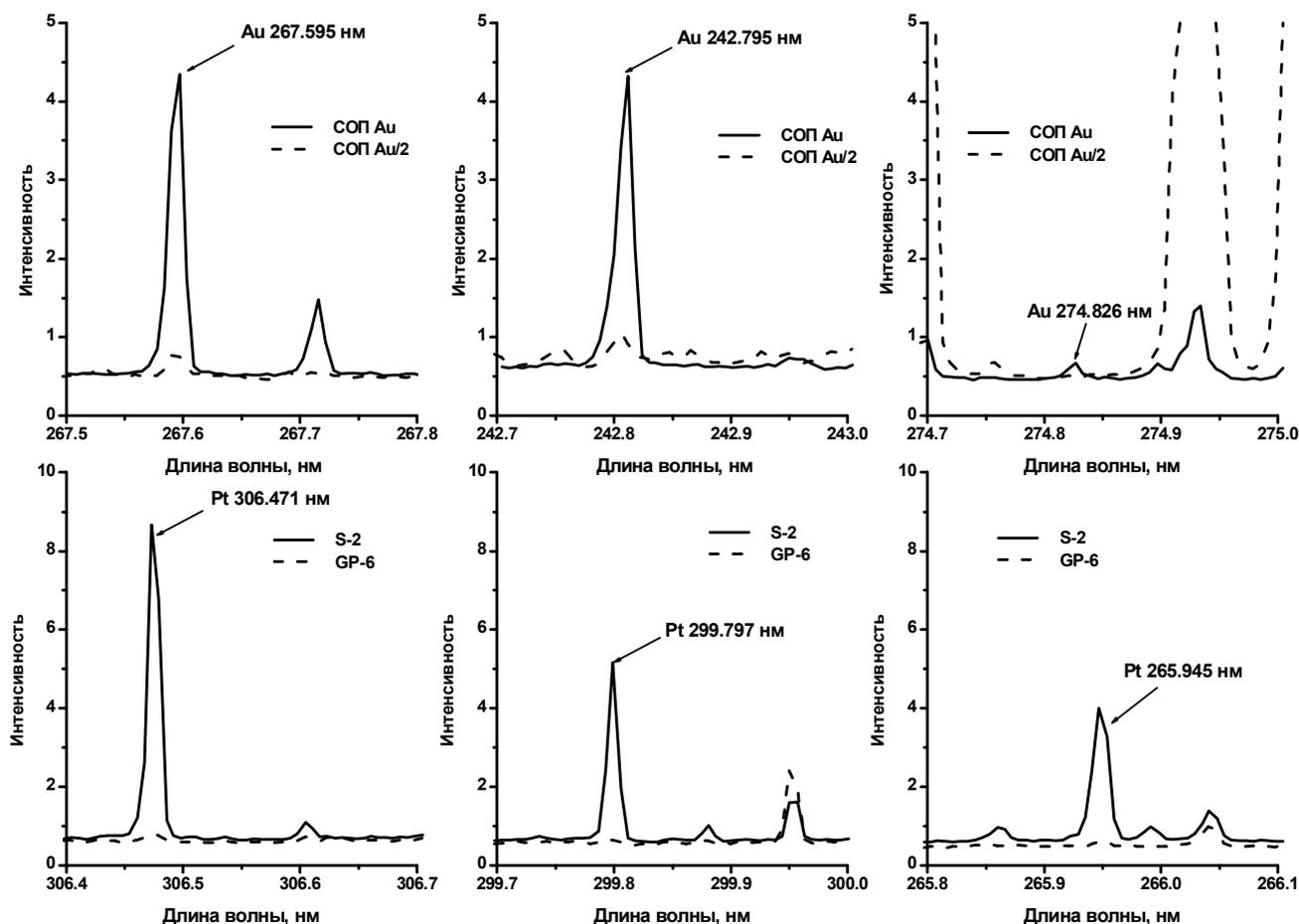


Рис. 1. Участки спектров градуировочных образцов с различными линиями золота и платины.

Таблица 1

Обучающая и тестирующая выборки для определения золота и платины в нерастворимом углеродистом веществе

Выборка	Образец	Au, мас. %	Образец	Pt, мас. %
Обучающая	S/2	0,25	S	0,5
	S-1	0,25	S/2	0,25
	S-3	0,05	S-1	0,25
	S-4	0,025	S-3	0,05
	GP-5	0,001	S-4	0,025
	СОП Au	0,00418	CHR-Pt <sup>+</sup>	0,0058
	СОП Au /2	0,00209	GP-6	0,003
	GP-3	0,0002	GP-4	0,0005
Тестирующая	S-2	0,125	GP-3	0,0002
	GP-4	0,0005	S-2	0,125
	CHR-Pt <sup>+</sup>	0,00043	CHR-Pt <sup>+</sup>	0,0058
			GP-5	0,001
		ВП-2 (ГСО № 927-76)	0,000264	

Интенсивности линий аналитов в спектрах градуировочных образцов меняются от предела обнаружения до максимальных значений интенсивностей этих же линий, регистрируемых в спектрах изучаемых НУВ. Поэтому диапазоны содержаний обучающих наборов составили для

золота 0,0002–0,25 % и для платины 0,0002–0,5 %. Диапазон  $\Delta C_{train}$  включает  $\Delta C_{test}$  (табл. 1), значительно превышая область линейной связи между переменными (формула (1)).

Численный эксперимент был выполнен отдельно для линий золота и платины. Для

каждой линии аналитов были построены градуировки МНК-1 и МНК-2. Для РГК-градуировки структура данных менялась последовательным добавлением зависимых признаков, т.е. интенсивностей следующей линии аналита. В табл. 2 и 3 показаны статистические характеристики градуировок, полученные при изменении структуры данных: для Y-переменной были вычислены дисперсия адекватности (ДиспАд), СКО градуировки (СКО град.) и размах ОСП (РОСП) для X-переменной. Для градуировки МНК-2(1Au) имеют наименьшие значения статистические характеристики ДиспАд, СКО град. и РОСП (выделены жирным шрифтом в таблицах). Проверка тестовым набором показывает, что эта модель также имеет минимальный размах ОСП. Нелинейные градуировки оказываются наилучшими для определения золота, хорошо описывая процесс получения аналитического сигнала золота, имеющего низкую температуру плавления и высокую скорость поступления паров в плазму дугового электрического разряда. Выполнение условия (5) выделяет модель МНК-2(1Au) как наилучшую градуировку. РГК-модель, удовлетворительно описывающая только линейные связи, дает максимальные значения  $\Delta R$  как для обучающего, так и для тестового наборов.

Статистические характеристики моделей МНК-1 и МНК-2 для линий платины при обучении имеют близкие значения, т.е. процесс получения

аналитического сигнала удовлетворительно описывается линейной функцией отклика. Все модели РГК-градуировок дают хорошие результаты при обучении  $\Delta R_{train} = 40 \pm 5 \%$ . При тестировании только градуировка РГК(1Pt & 3Pt) дает минимальный размах ОСП. Условие (5) примерно одинаково реализуется для моделей МНК-2(1Pt), МНК-1(1Pt) и РГК(1Pt&3Pt). Образец хромита  $\text{CHR-Pt}^+$  не соответствует по составу НУВ, но включен в обучающий и тестовый наборы. Его присутствие в тестовой выборке резко увеличивает  $\Delta R$  для всех моделей с линией 2Pt, и указывает на высокую чувствительность этого признака к наличию в выборках образцов, отличающихся по составу макрокомпонентов.

Таким образом, численный эксперимент показал, что:

- оценка "размах относительной систематической погрешности – РОСП" (формула (3)) является чувствительной к изменению общего химического состава образцов, который отражается в структуре экспериментальных данных изменением числа спектральных линий и диапазона их интенсивностей;
- величина РОСП отражает степень учета внутренних связей при обработке спектральной информации;
- выполнение условия (5) обеспечивает выбор структуры градуировочной модели, которая наиболее полно описывает аналитический процесс.

Таблица 2

Статистические характеристики градуировок, использующих линии золота

Градуировочная модель	Структура выборки	Обучающая			Тестирующая
		ДиспАд	СКО град.	Размах ОСП	Размах ОСП
МНК-1	1Au	0,0176	0,1014	74	5
	2Au	0,0492	0,1603	108	44
МНК-2	1Au	<b>0,0091</b>	<b>0,0076</b>	<b>52</b>	<b>3</b>
	2Au	0,0284	0,1200	68	40
РГК	1Au & 2Au			129	97

Таблица 3

Статистические характеристики градуировок, использующих линии платины

Модель градуировки	Структура выборки	Обучающая			Тестирующая	
		ДиспАд	СКО град.	РОСП	РОСП	РОСП без CHR-Pt <sup>+</sup>
МНК-1	1Pt	<b>0,0046</b>	<b>0,0788</b>	<b>40</b>	<b>23</b>	<b>14</b>
	2Pt	0,0098	0,0932	49	1206	40
	3Pt	0,0116	0,1075	71	327	6
МНК-2	1Pt	<b>0,0046</b>	<b>0,0794</b>	35	<b>21</b>	<b>1</b>
	2Pt	0,0096	0,0925	29	1256	46
	3Pt	0,0100	0,1043	63	378	3
РГК	1Pt & 2Pt	-	-	36	610	610
	1Pt & 3Pt	-	-	45	<b>23</b>	<b>13</b>
	2Pt & 3Pt	-	-	40	699	699
	1Pt & 2Pt & 3Pt	-	-	39	317	317

## 2. Определение примеси бора в кристаллическом кремнии и кварце

Моделирование структуры данных выполнено с целью изучения возможности учета неспектральных и спектральных помех с помощью многомерной РГК-градуировки.

В обучающем и тестовом наборах образцов (табл. 4) анализ представлен линиями бора 249,772 и 249,677 нм, соответственно 1В и 2В. Область определяемых содержаний составляет менее двух порядков от 0,00015 до 0,0059 мас. % и хорошо описывается линейной градуировкой (формула (1)). Обе линии бора испытывают наложение кантов молекулярной полосы SiO. На линию В 249,677 нм налагается также линия Fe 249,653 нм. Это хорошо видно на спектрах кварца БС-17 и образцов кристаллического кремния (рис. 2). Варианты одномерных (МНК) и многомерных (РГК) градуировок и их статистические характеристики приведены в таблицах 5-7.

Учет вклада молекулярной полосы в интенсивность аналитических сигналов был проведен путем вычитания контрольного опыта (ко) из спектров обучающего и тестирующего наборов. Контрольный опыт – это спектр высокочистого обогатенного кварца IOTA 4 (99,99 мас. % SiO<sub>2</sub>) с содержанием бора

( $4 \cdot 10^{-6}$  мас. %), что значительно ниже предела определения бора для прямого атомно-эмиссионного анализа ( $7 \cdot 10^{-5}$  мас. %) [11]. Влияние интерференционной помехи на интенсивность линии бора мы оценили по градуировкам, построенным без учета и с учетом контрольного опыта (табл. 5). При обучении минимальные величины статистических характеристик градуировок были получены для МНК-1 и МНК-2 с учетом контрольного опыта. Об этом свидетельствуют как минимальные значения дисперсий адекватности и СКО градуировок, так и размахи ОСП для обучающей и тестовой выборок.

Так как число градуировочных образцов кварца и чистого кристаллического кремния крайне ограничено, тестовый набор повторил ряд образцов из обучающего набора. Однако в тестовую выборку мы включили только новые спектры зашифрованных образцов. В этом случае, исходя из теоретических предположений, оценка предсказательной способности по тестовому набору может оказаться одинаковой или лучше, чем для обучающей выборки. Экспериментальные оценки, полученные при обучении и тестировании (табл. 5), оказались примерно одинаковыми и показали, что учет контрольного опыта улучшает предсказательную способность градуировок.

**Таблица 4**

Обучающая и тестирующая выборки при определении примесей в кристаллическом кремнии и кварце

Выборка	Образец		Содержание, мас. %				
	Название и номер	Вещество	B	Al	Ca	Fe	Si
Обучающая (Training)	Кв -1	кварц	0,00024	0,0079	0,0021	0,322	46,4
	БС-17	кварцит	<i>0,00015*</i>	0,00055	0,00055	0,00055	46,5
	Кварц "Лисма"	обогащенный высокочистый кварц	<i>0,0002</i>	<i>0,0054</i>	<i>0,004</i>	<i>0,03</i>	46,5
	PV Silicon Standard № 6106-56, США	кристаллический кремний	0,0019	0,141	0,0104	0,319	99,4
	СОП-1 (Si № 1)		0,0059	1,14	0,68	1,02	97,2
	СОП-3 (Si № 3)		0,0047	0,73	0,34	0,68	98,1
	СОП-4 (Si № 4)		0,0011	0,07	0,017	0,45	99,2
	СОП-12 (Si №12)		0,0015	0,056	0,0082	0,185	99,8
	КП 1	смесь Кв-1+ БС-17 (1:1)	<i>0,000165</i>	<i>0,0042</i>	<i>0,00132</i>	<i>0,161</i>	46,5
	КП 3	смесь СОП-4 + БС -17 (1:1)	<i>0,000745</i>	<i>0,0353</i>	<i>0,00878</i>	<i>0,225</i>	72,8
	СДПС-1 (ГСО № 2498-83)	почва	0,0030	1,78	0,19	0,69	42,6
Тестирующая (Test)	СКв-1 (ГСО № 2577-83)	кварцит	<i>0,0004</i>	0,063	0,071	0,44	46,2
	БС-17		<i>0,00015</i>	0,00055	0,00055	0,00055	46,5
	Кв-1	кварц	0,00024	0,0079	0,0021	0,322	46,4
	СОП-1 (Si № 1)	кристаллический кремний	0,0059	1,14	0,68	1,02	97,2
	СОП-4 (Si № 4)		0,0011	0,07	0,017	0,45	99,2
	СОП-12 (Si №12)		0,0015	0,056	0,0082	0,185	99,8
	СДПС-1 (ГСО № 2498-83)	почва	0,0030	1,78	0,19	0,69	42,6

Примечание: \* - рекомендованные содержания указаны курсивом

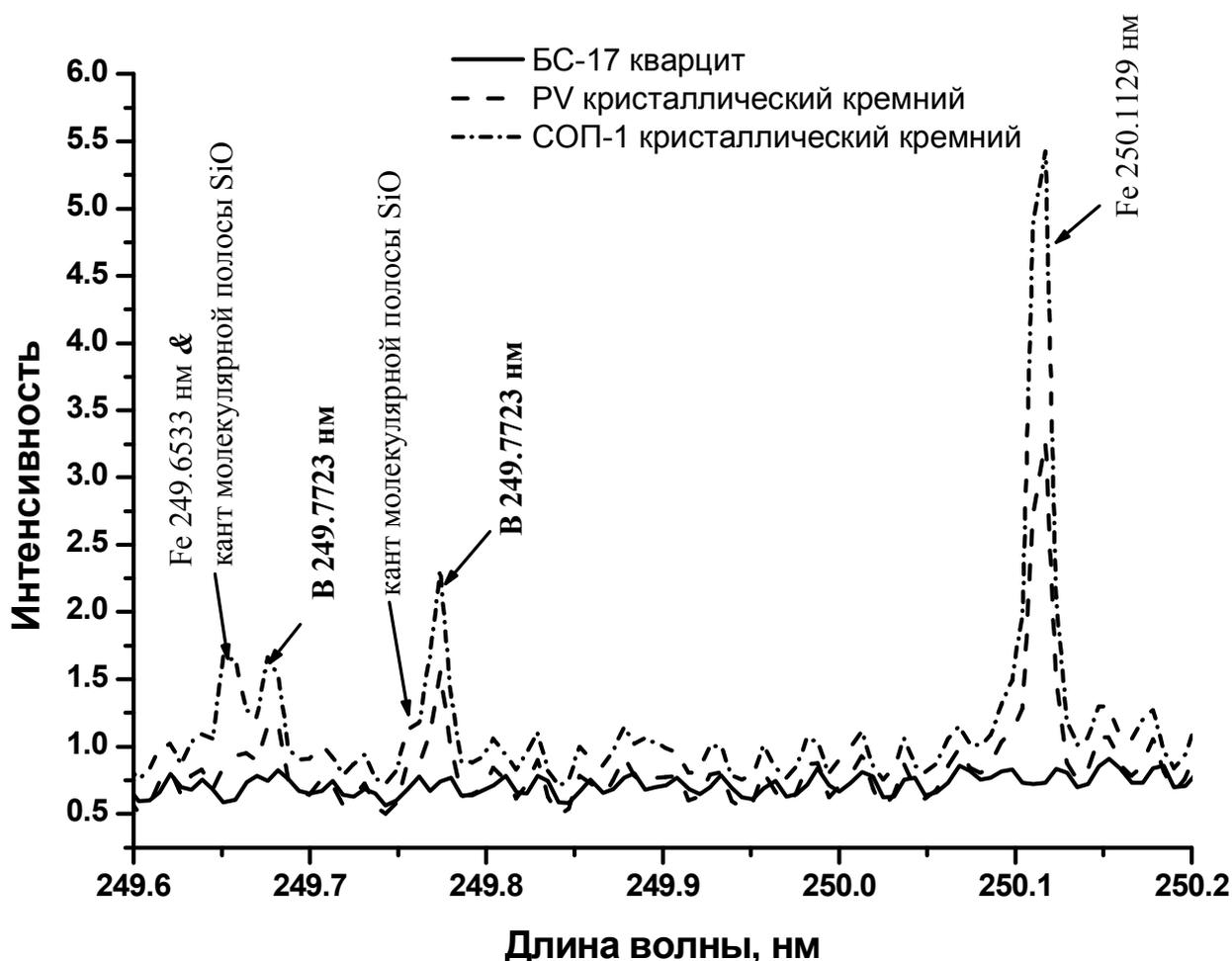


Рис. 2. Участок спектра с линиями бора, железа и молекулярной полосой SiO.

Таблица 5

Статистические характеристики МНК-градуировок для линий бора

Градуировка	Структура выборки		Обучающая			Тестирующая
			ДиспАд	СКО град.	$\Delta R_{train}$	$\Delta R_{test}$
МНК-1	1 В	без учета ко	0,0887	0,1569	139	109
			0,1387	0,1916	198	420
	2 В	с учетом ко	0,0436	0,1201	68	114
			0,0877	0,1543	108	67
МНК -2	1 В	без учета ко	0,097	0,1578	109	170
			0,1574	0,1960	260	494
	2 В	с учетом ко	0,0307	0,1042	76	147
			0,0627	0,1276	145	102

Для учета наложения линии Fe 249,653 нм на линию бора В 249,677 нм мы использовали линию Fe 246,887 нм (2Fe). Эти линии железа являются аналогами по типу электронного перехода и спектральной яркости. Измеренная интенсивность линии 2Fe была включена в таблицу данных для РГК-градуировки (табл. 6). Мы сравнили по условиям (4) и (5) все варианты линейных градуировок из табл. 5 и 6 для одинаковых обучающих и тестовых наборов СО (табл. 7). Наилучшие размахи ОСП при обучении и

тестировании получены для модели РГК(1В-ко; 2В-ко) или РГК(1В; 2В), которая выделена в таблице жирным шрифтом. Так как далее в названиях структур моделей для многомерной градуировки используются только полученные после учета контрольного опыта интенсивности линий бора, указание на вычет контрольного опыта не повторяется в названиях структуры моделей (табл. 6).

Таблица 6

Особенности модельных структур и размахи ОСП для обучающей и тестирующей выборок при определении бора с использованием РГК-градуировок

Влияние	Структура модели		$\Delta R_{train}$	$\Delta R_{test}$
Спектральные помехи	РГК (1В; 2В)	без учета ко	98	100
	<b>РГК (1В; 2В)</b>	линии бора с учетом ко	<b>35</b>	<b>43</b>
	РГК (2В; 2Fe)		192	113
	РГК (1В; 2В; 2Fe)		51	80
Матричные эффекты	<b>РГК (1В; 2В; Si)</b>	линии бора с учетом ко	<b>37</b>	<b>42</b>
	<b>РГК (1В; 2В; 1Fe)</b>		<b>31</b>	<b>51</b>
	РГК (1В; 2В; Ca)		154	122
	РГК (1В; 2В; Al)		59	62
Спектральные помехи + матричные эффекты	<b>РГК (1В; 2В; 1Fe; 2Fe)</b>	линии бора с учетом ко	<b>27</b>	<b>55</b>
	<b>РГК (1В; 2В; Si; 1Fe; 2Fe)</b>		<b>27</b>	<b>54</b>
	<b>РГК (1В; 2В; Si; 1Fe)</b>		33	51
	РГК (1В; 2В; 1Fe; 2 Fe; Al)		70	68
	РГК (1В; 2В; Si; 1Fe; 2Fe; Al)		70	68
	РГК (1В; 2В; Si; 1Fe; Ca)		147	163
	РГК (1В; 2В; Si; 1Fe; 2Fe; Ca; Al)		149	137

Выполнение условия (5) указывает на то, что молекулярная полоса SiO оказывает более сильное влияние на качество аналитических результатов, чем помеха линии железа. Тем не менее, учет обоих наложений необходим, если мы определяем бор в образцах неизвестного состава. Использование в градуировке двух линий бора с близкими пределами обнаружения улучшает качество обучения и прогноза, также как в хроматографическом анализе [6].

Таблица 7

Варианты линейных градуировок для учета спектральных наложений

Вариант градуировки	$\Delta R_{train}$	$\Delta R_{test}$
МНК-1(1В без учета ко)	139	109
МНК-1(2В без учета ко)	198	420
МНК-1(1В-ко)	68	114
МНК-1(2В-ко)	108	67
РГК(1В без учета ко; 2В без учета ко)	98	100
РГК(2В-ко; 2Fe)	192	113
<b>РГК(1В-ко; 2В-ко)</b>	<b>35</b>	<b>43</b>
РГК(1В-ко; 2В-ко; 2Fe)	51	80

Рассмотрим неспектральные помехи, моделируя структуру табличных данных для РГК-градуировок (табл. 6). Известно, что изменение состава основы может привести к изменениям интенсивности линий бора из-за изменений параметров плазмы разряда, так как его атомы имеют высокий потенциал ионизации (8,298 эВ) [1, 2]. Рассматриваемые аналитические объекты (кварц и высокочистый металлургический кремний) разнообразны по валовому химическому составу и концентрациям примесей [11]. В них содержание элемента-основы, кремния, меняется в обучающей выборке от 42,6 до 99,8; железа –

до 1; кальция – до 0,68; Al – до 1,78 мас. %. Для того, чтобы оценить и учесть влияние матричных эффектов, в таблицу данных мы добавили столбцы измеренных интенсивностей линий Si 302,0005 нм – (Si), Fe 299,77 нм – (1 Fe), Al 265,247нм – (Al), Ca 317,933 нм – (Ca). Интенсивности перечисленных линий меняются в 500-1000 раз по сравнению с фоновым значением. Моделирование структуры экспериментальных данных обучающего и тестового наборов мы проводили, добавляя в таблицу к линиям бора последовательно по одной линии макроэлемента. Для каждой модельной структуры был вычислен размах ОСП обучающей и тестирующей выборки. Полученные значения размахов ОСП ( $\Delta R_{train}$  и  $\Delta R_{test}$  в табл. 6) для градуировок РГК(1В; 2В) без учета матричного эффекта и с последовательным учетом влияния макроэлементов показывают, что степень матричного влияния на интенсивность линий бора уменьшается от кремния к кальцию в рассматриваемых диапазонах изменения их содержаний Fe ~ Si > Al > Ca. Несмотря на то, что содержание железа в образцах меньше, чем кремния, влияние его присутствия оказывается не менее значительным.

В РГК-градуировках также возможен одновременный учет матричных влияний нескольких макроэлементов и спектральных помех (табл. 6). Рост значений  $\Delta R_{train}$  и  $\Delta R_{test}$  указывает, на то, что одновременный учет матричного влияния алюминия и кальция вносит дополнительный шум и не приводит к улучшению точности результатов определения бора. Для получения удовлетворительных аналитических результатов достаточно в градуировке учитывать матричные эффекты, обусловленные влиянием кремния и железа.

Для моделей РГК(1В; 2В; 2Fe; 1Fe) и РГК(1В; 2В; Si; 1Fe; 2Fe) мы получили одинаковые

значения  $\Delta R_{train}$  и  $\Delta R_{test}$ , равные 27 % и 54-55 % соответственно (табл. 6). РГК(1B; 2B; Si; 1Fe; 2Fe), кроме спектральных помех SiO и 2Fe и матричного влияния железа через 1Fe, учитывает и матричное влияние кремния. Для этих градуировок выполняются условия (4) и (5), и они оцениваются как наилучшие. Также необходимо отметить для модели РГК(1B; 2B; Si; 1Fe) близкие к предыдущим моделям (в пределах точности измерений) значения  $\Delta R_{train}$  и  $\Delta R_{test}$ , составляющие 33 % и 51 % соответственно. Используя только линию железа 1Fe, удается учесть как спектральное наложение линии Fe 249,653 нм на линию бора B 249,677 нм, так и матричный эффект присутствия меняющегося в пробах содержания железа, влияющее на интенсивность линий бора.

В ходе численного эксперимента в условиях конкретной аналитической методики выделен класс неразличимых  $n$ -мерных РГК-градуировок, которые в рамках точности измерений обеспечивают одинаковый учет спектральных наложений и матричных влияний на интенсивности линий аналита при различной структуре данных. В табл. 6 такие градуировки выделены жирным шрифтом. Нахождение не одной, а множества градуировочных моделей, неразличимых по величине оценок, но различимо лучших по сравнению с остальными, естественно в условиях неопределенности, связанной со случайными погрешностями измерений [12]. Однако правила, сформулированные при составлении модельных структур выделенного множества неразличимых градуировок, относятся только к изученным диапазонам содержания элементов. Их изменение потребует проверки достаточности выбранной структуры.

## ЗАКЛЮЧЕНИЕ

Показано, что применение в дуговой атомно-эмиссионной спектрометрии многомерных градуировок обеспечивает одновременный учет спектральных и неспектральных эффектов. При

## ЛИТЕРАТУРА

1. Русанов А.К. Основы количественного спектрального анализа руд и минералов. М.: Недра, 1978. 400 с.
2. Райхбаум Я.Д. Физические основы спектрального анализа. М.: Наука, 1980. 158 с.
3. Заводская лаборатория. Диагностика материалов. 2007. Т. 73. Специальный выпуск.
4. Эренберг А. Анализ и интерпретация статистических данных. М.: Финансы и статистика, 1981. 406 с.
5. Esbensen K.H. Multivariate data analysis – in practice. Camo ASA, 2001. 598 p.
6. Шараф М.А., Иллмен Д.Л., Ковальский Б.Р. Хемометрика. / Пер. с англ. Л.: Химия, 1989. 272 с.

автоматическом сборе спектральной информации, который обеспечивается современными многоканальными приборами, моделирование структуры спектральных данных при построении  $n$ -мерных градуировок и их сравнение между собой с помощью оценки "размах относительной систематической погрешности" является удобным способом исследования возможностей аналитических методик. Для этого в структуру обучающей выборки данных для набора градуировочных образцов должны включаться как частные признаки, отражающие спектральные и неспектральные эффекты:

- интенсивности нескольких линий аналита, откорректированные на присутствие молекулярных полос;
- линии интерферирующих микроэлементов;
- линии макроэлементов, которые дают основные матричные эффекты.

Предложенная оценка "размах относительной систематической погрешности" достаточно чувствительна к модельным изменениям в структуре данных и позволяет выявить линии элементов, участие которых в вычислении градуировочной зависимости необходимо, чтобы минимизировать погрешности результатов при выполнении методики анализа на конкретной спектральной установке. Выполнение условий, когда размах ОСП стремится к минимуму только для обучающей выборки или одновременно для обучающей и тестовой выборок, обеспечивает выбор класса градуировочных моделей, которые с удовлетворительной точностью описывают процедуры обработки спектральной информации.

Численный эксперимент, выполненный на примере двух методик прямого атомно-эмиссионного анализа, показал, что многомерные модели РГК-градуировки позволяют преобразовать неградулируемые влияния в градулируемые и имеют преимущества по сравнению с одномерными зависимостями для диапазонов содержания аналита, не превышающих два порядка.

7. Faber K., Kovalski B.R. // Journal of Chemometrics. 1997. № 11. P. 181-238.
8. Васильева И.Е., Шабанова Е.В., Васильев И.Л. // Аналитика и контроль. 2005. Т. 9. № 2. С.150-156.
9. Развозжаева Э.А. [и др]. // Геология и геофизика. 1999. Т. 40. № 9. С. 1324-1330.
10. Развозжаева Э.А. [и др]. // Геология и геофизика. 2002. Т. 43. № 3. С. 286-296.
11. Васильева И.Е. [и др]. // Аналитика и контроль. 2001. Т. 5. № 1. С.24-34.
12. Воцинин А.П. // Заводская лаборатория. 1981. Т. 47. № 7. С. 50-55.

## MODELING OF DATA STRUCTURE FOR MULTIVARIATE CALIBRATION IN ATOMIC EMISSION SPECTROMETRY

*I.E. Vasilyeva and E.V. Shabanova*

Vinogradov Institute of Geochemistry, SB RAS  
1A Favorsky Str., Irkutsk, 664033 Russia

The multivariate statistical data analysis (MDA) applied in atomic emission spectrometry (AES) has a great potential as it considers the changes of plasma parameters affecting spectral line intensities observed in the direct current arc and deteriorating accuracy of analytical results. The conditions of application of  $n$ -dimensional calibrations with a principal component regression (PCR) and data structure modeling are investigated to take into account the spectral and non-spectral effects in gold and platinum determination in insoluble carbonaceous substance and boron impurity in crystalline silicon and quartz. The best calibration dependence, out of the options considered in  $n$ -dimensional calibrations, has been selected based on "the range of relative systematic error within the interval of contents to be determined".

**Key words:** atomic emission spectrometry with the direct current arc, multivariate calibration, principal component regression