

**АВТОМАТИЗАЦИЯ ПОСТРОЕНИЯ УРАВНЕНИЯ  
РЕГРЕССИИ ДЛЯ ПРОВЕДЕНИЯ КАЧЕСТВЕННОГО АНАЛИЗА  
ЭНЕРГЕТИЧЕСКИХ СПЕКТРОВ**

**Самарина А.В.**

**Научный руководитель – к.т.н. Сиротин Э.Е.**

*Сибирский федеральный университет*

В современной аналитической химии рентгеноспектральные методы определения химического состава вещества относятся к наиболее динамично развивающимся. Одно из основных направлений рентгеноспектрального анализа - рентгенофлуоресцентный анализ (РФА). Метод РФА относится к группе неразрушающих методов анализа, дает возможность определять содержания большинства элементов Периодической системы в широком динамическом диапазоне. Анализируемые пробы не требуют особой подготовки, могут быть монолитными, порошковыми или жидкими. Продолжительность анализа одной пробы составляет от 5-10 секунд до 1-5 мин.

Метод РФА применяется в установках, разрабатываемых на предприятии ООО «Радос». Эти установки предназначены для применения в технологии поточного бесконтактного экспрессного опробования состава сырья на транспортном носителе (конвейерная лента, тарельчатый питатель) в режиме реального времени, дают возможность крупнопорционной сортировки по измеренным содержаниям анализируемого элемента, а также возможность управления технологическими процессами подготовки и переработки сырья.

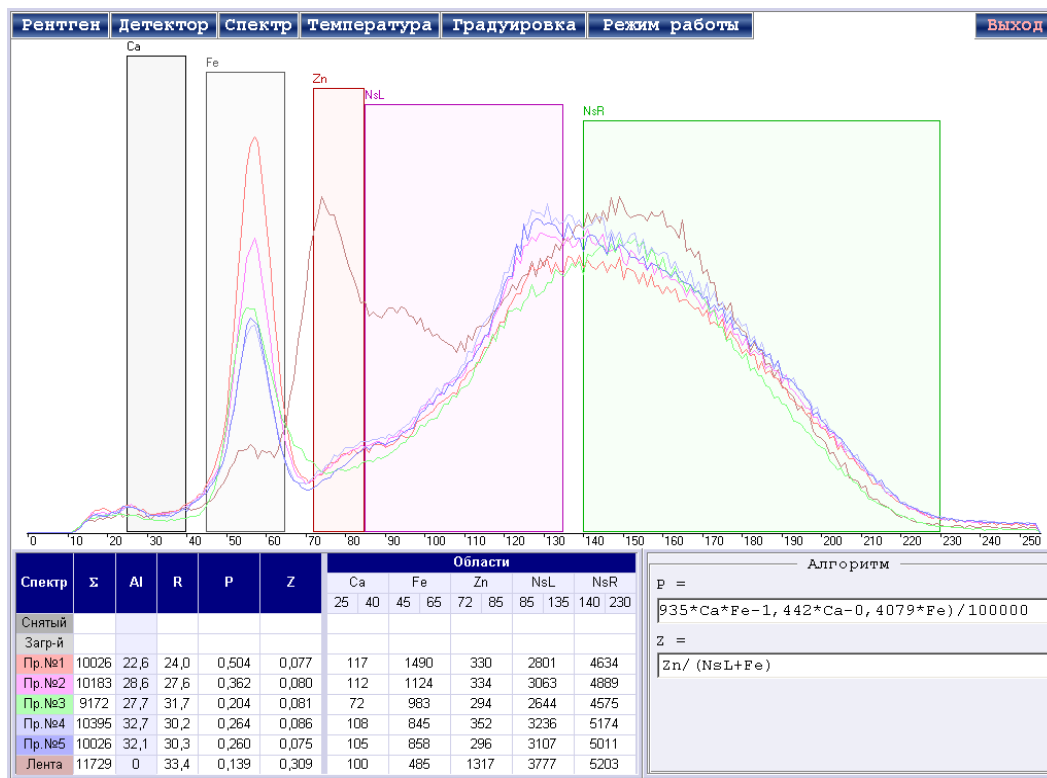
При проведении анализа руда подвергается рентгеновскому облучению, в результате чего происходит возбуждение атомов, входящих в состав руды. Вторичное излучение улавливается детектором и измеряется энергия квантов, входящих в состав повторного излучения. Спектр представляет собой гистограмму распределения квантов от их энергий, зарегистрированных за единицу времени. Спектр может включать в себя 256, 512 или 1024 каналов. По результатам измерений определяется качество измеряемого продукта (тип, сорт) и процентное содержание каждого из химических элементов и определить химический состав (рис. 1). Основная задача РФА - нахождение неизвестной концентрации элемента в анализируемом образце по его энергетическому спектру

Число рентгеновских фотонов, излучаемых атомами в единицу времени, пропорционально количеству атомов, но регистрируемый аналитический сигнал  $i$ -го элемента от анализируемой пробы зависит от многих факторов (поглощения, рассеяния, дополнительного возбуждения и других видов взаимодействия рентгеновского излучения с веществом, влияние аппаратурных условий анализа), в общем случае может быть представлен как

$$I = F(\vec{C}, \vec{X}, \vec{Y})$$

(1),

где  $\vec{C}$  — вектор концентраций всех элементов, составляющих анализируемую пробу;  $\vec{X}$  — вектор фундаментальных характеристик отдельных элементов (атомных



номеров, длин волн излучения и т. д.);  $\vec{Y}$  — вектор, определяющий аппаратные условия возбуждения и регистрации аналитических сигналов.

Рис. 1. Энергетические спектры нифелиновых руд АГК с различным содержанием  $Al_2O_3$ .

В основе физической модели возбуждения рентгеновской флуоресценции лежат предположения об идеально гомогенных образцах с идеально гладкой поверхностью, а также о возможности полного устранения аппаратных эффектов. На практике из-за перечисленных выше факторов реальную физическую модель построить невозможно, поэтому целесообразно заменить ее формальной математической моделью. Правильно построенная математическая модель не должна увеличивать «шумы» исходных данных. При построении математической модели выражение (1) рассматривается как ограниченная непрерывная нелинейная функция состава анализируемого образца, которую по теореме Вейерштрасса с любой наперед заданной точностью можно представить в виде действительного степенного ряда

$$I_i = \alpha_0 + \sum_i \alpha_i c_i + \sum_{i \neq j} \alpha_{ij} c_i c_j + \sum_i \alpha_{ii} c_i^2 + \dots$$

или в записи через интенсивности:

$$c_i = \alpha'_0 + \sum_i \alpha'_i l_i + \sum_{i \neq j} \alpha'_{ij} l_i l_j + \sum_i \alpha'_{iii} l_i^3 + \dots$$

Эти уравнения известны как феноменологические, регрессионные или статистические.

Регрессионный анализ - раздел математической статистики, объединяющий практические методы исследования регрессионной зависимости между величинами по статистическим данным. В регрессионном анализе рассматривается связь между одной переменной, которая называется зависимой переменной, или признаком, и несколькими другими, которые называются независимыми переменными.

На данный момент построение уравнения регрессии на предприятии ООО «Радос» проводится после проведения анализа полученных энергетических спектров от образцов руды определенного металла с известным содержанием интересующих элементов и определения каналов спектров, значения которых наиболее связаны с величиной содержания интересующих элементов - наиболее значимых каналов. Это делается для того, чтобы сократить количество исходных данных для решения поставленной задачи. Первым этапом построения уравнения регрессии является нахождение алгоритма расчета корреляционного параметра как функции от нескольких переменных, представляющих собой интегральные суммы значений значимых каналов спектра. На втором этапе исследования при помощи методики регрессионного анализа осуществляется поиск аналитической формы уравнения регрессии между значениями корреляционного параметра и содержанием интересующих химических элементов.

Эти этапы выполняются до настоящего времени исключительно эмпирическим способом, на основе многолетнего опыта экспертов ООО «Радос». В моем исследовании производится попытка автоматизировать этапы построения уравнения регрессии с целью увеличения скорости, надежности обработки получаемых вторичных спектров руд металлов (качественного анализа), а так же снижения стоимости анализа руды за счет уменьшения нагрузки специалистов, осуществляющих эти этап исследования (увеличение эффективности исследований).

Для произведения расчетов были взяты данные с Зырянского рудника, они содержат информацию о самих пробах : имя пробы (с высотой), загрузка по всему спектру, время накопления спектра, загрузка по областям, исходное содержание элементов (показания химанализа), значения спектров по каждому из 2048 каналов.. Было произведено 26 проб на 4 разных высотах.

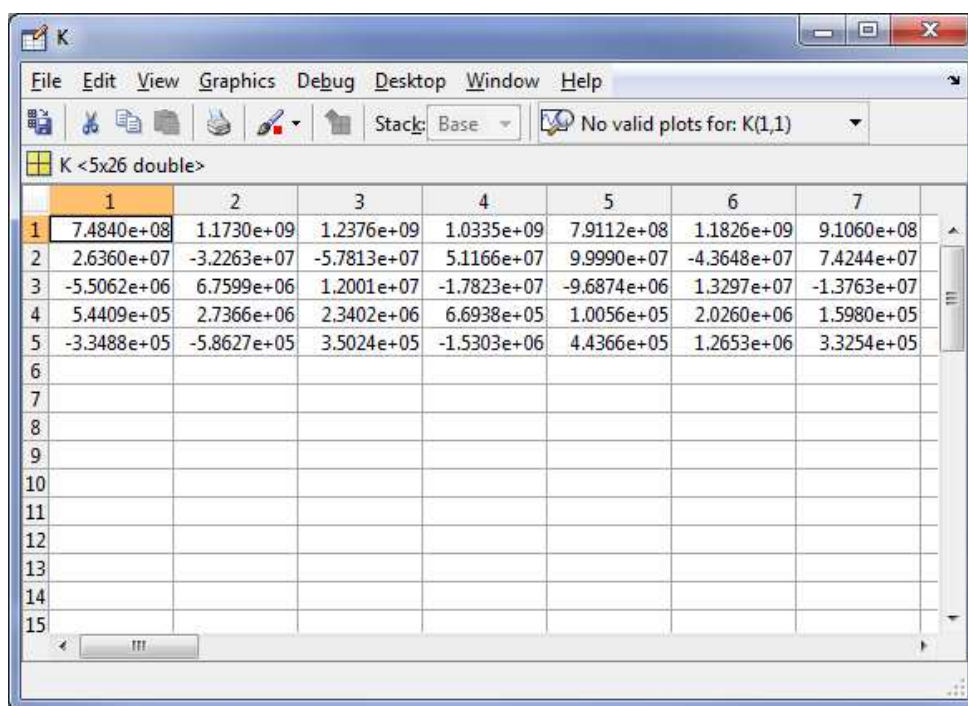
На первом этапе производилась обработка исходных данных: было выбрано 26 проб на одной высоте - 40 см. Таким образом, была получена матрица X размерностью 2048x26 (значения 26 проб по энергетическим каналам). Далее на основании исследований, проводимых студенткой ИКИиТ СФУ Ф.Садреевой в дипломной работе «Применение факторного анализа для исследования энергетических спектров руд различных металлов» (2010г.), был совершен переход от 2048 к 5 значимым переменным с помощью метода главных компонент (РСА). С математической точки

зрения PCA — это декомпозиция исходной матрицы  $X$ , т.е. представление ее в виде произведения двух матриц  $T$  и  $P$ :

$$X = TP^t + E (2),$$

матрица  $T$  называется матрицей счетов (scores), матрица  $P$  — матрицей нагрузок (loadings), а  $E$  — матрицей остатков. Матрицы  $T$  и  $P$  — были найдены с помощью [SVD разложения](#) через стандартную функцию MATLAB, называемую `svd`. Перед применением функции `svd` исходная матрица  $X$  была центрирована (`centering(X)`). Из матрицы счетов  $T$  размерностью  $2048 \times 26$  получаем новую матрицу  $F$  размерностью  $2048 \times 5$ , состоящую из первых пяти столбцов матрицы  $T$ .

После этого была рассчитана линейная комбинация для каждого из исходных спектров (матрица  $X$ ) с матрицей  $F$  (векторов-столбцов), где каждая строка — это значение факторов для соответствующего содержания элементов. Результатом является матрица  $K$  (рис.2) размерностью  $5 \times 26$ , которая будет использована в дальнейшем для построения уравнения регрессии.



|    | 1           | 2           | 3           | 4           | 5           | 6           | 7           |
|----|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| 1  | 7.4840e+08  | 1.1730e+09  | 1.2376e+09  | 1.0335e+09  | 7.9112e+08  | 1.1826e+09  | 9.1060e+08  |
| 2  | 2.6360e+07  | -3.2263e+07 | -5.7813e+07 | 5.1166e+07  | 9.9990e+07  | -4.3648e+07 | 7.4244e+07  |
| 3  | -5.5062e+06 | 6.7599e+06  | 1.2001e+07  | -1.7823e+07 | -9.6874e+06 | 1.3297e+07  | -1.3763e+07 |
| 4  | 5.4409e+05  | 2.7366e+06  | 2.3402e+06  | 6.6938e+05  | 1.0056e+05  | 2.0260e+06  | 1.5980e+05  |
| 5  | -3.3488e+05 | -5.8627e+05 | 3.5024e+05  | -1.5303e+06 | 4.4366e+05  | 1.2653e+06  | 3.3254e+05  |
| 6  |             |             |             |             |             |             |             |
| 7  |             |             |             |             |             |             |             |
| 8  |             |             |             |             |             |             |             |
| 9  |             |             |             |             |             |             |             |
| 10 |             |             |             |             |             |             |             |
| 11 |             |             |             |             |             |             |             |
| 12 |             |             |             |             |             |             |             |
| 13 |             |             |             |             |             |             |             |
| 14 |             |             |             |             |             |             |             |
| 15 |             |             |             |             |             |             |             |

Рис.2. Полученная матрица  $K$

Для завершения исследования будут выполнены следующие шаги: определение коэффициентов уравнения регрессии для расчета аналитического параметра определения интересующих элементов с помощью метода наименьших квадратов (МНК); построение различных уравнения регрессии (уравнение Лукаса-Туса и Прайса, уравнение Белова—Дуймакаева, уравнение Мэсю—Джонсона); оценивание каждого уравнения регрессии в целом через показателями качества (коэффициент детерминации, значение  $F$ -статистики, коэффициент корреляции, сумма квадратов остатков, стандартная ошибка регрессии). И на основе полученных данных будет проводится разработка алгоритма построения уравнений регрессии для проведения качественного анализа руды по ее энергетическому спектру вторичного излучения.