

ПРИМЕНЕНИЕ СТАТИСТИЧЕСКИХ МЕТОДОВ ДЛЯ АНАЛИЗА ДАННЫХ В ЦВЕТНОЙ МЕТАЛЛУРГИИ

**Казачковская О.В., Дранишников С.В.,
Сибирский Федеральный университет**

ОАО «Красноярский завод цветных металлов им. В. Н. Гулидова» — крупнейшее в мире предприятие, осуществляющее в промышленных масштабах аффинаж всех металлов платиновой группы, а также золота и серебра.

В процессе аффинажа платиновых металлов нет таких технологических операций, где бы не использовались или не образовывались вредные для окружающей среды и человека химические продукты.

Несмотря на постоянное совершенствование процессов обезблагораживания и утилизации промышленных стоков аффинажного производства концентрация жидких растворов на выходе процесса переработки отходов превышает предельно допустимые концентрации по аммиаку, благородным, цветным металлам, железу и ряду анионов. Это приводит к экологическим проблемам в регионе и экономическим потерям предприятия, в том числе, и за счет штрафных санкций,

Поэтому было принято решение провести анализ работы процесса переработки жидких отходов с целью оценки возможности повышения его эффективности.

Прежде чем начинать анализ данных, необходимо было выбрать инструменты для их обработки. В первую очередь выбирают статистические методы, поскольку они признаются важным условием рентабельного управления качеством.

Из множества статистических методов для широкого применения обычно используют «Семь простых инструментов»: стратификацию, гистограмму, контрольную карту, контрольный листок, диаграмму Исикавы, диаграмму рассеивания, диаграмму Парето. С их помощью выявляют проблемы, устанавливают основные факторы, с которых нужно начинать действовать, и распределять усилия с целью эффективного решения этих проблем. По мнению Каору Исикавы применение перечисленных методов позволяет решить 95% любых проблем, возникающих на производстве.

В нашем случае для анализа процесса была выбрана диаграмма рассеивания, так как необходимо было выявить зависимость выхода процесса от влияющих на него факторов $y = f(x_i)$.

Попытки получения однофакторных зависимостей выхода процесса от его параметров, не показали ни каких зависимостей. Поэтому было принято решение использовать многомерные статистические методы. Например, метод главных компонент, который хорошо подходит для анализа таких данных. Этот метод может легко иметь дело с сотнями или тысячами чрезвычайно коррелированных шумных переменных одновременно.

Метод главных компонент – это классический метод снижения размерности данных путем определения незначительного числа линейных комбинаций исходных признаков, объясняющих большую часть изменчивости данных в целом, дающий однозначное решение [1, 2, 3].

Метод главных компонент имеет дело с множеством неизвестных факторов и связей. При этом не только число таких факторов и характер их взаимосвязи заранее не известны, многие из них латентны и проявляются лишь опосредовано через внешние признаки или их группы. Задача заключается в том, чтобы на основе множества внешних (и существенных, и мало существенных, и многочисленных ничтожных)

признаков найти небольшое, но значимое число их гибридов, которые коротко и ясно объяснят их причинно-следственные связи в пространстве и времени.

Метод главных компонент осуществляет переход к новой системе координат U_1, \dots, U_p в исходном пространстве признаков x_1, \dots, x_p , которая является системой ортонормированных линейных комбинаций. Первая главная компонента – это линейная комбинация, обладающая наибольшей дисперсией. Вторая главная компонента имеет наибольшую дисперсию среди всех оставшихся.

Одним из важнейших преимуществ метода главных компонент, является возможность представить сложные данные физического эксперимента в более простом виде, так чтобы исследователь смог «увидеть» результаты экспериментов в простой графической интерпретации [4].

Один из наиболее мощных инструментов метода главных компонент - *график счетов*. Графики счетов можно рассматривать как двумерные «окна» в пространство главных компонент, в которых можно увидеть взаимное расположение образцов.

Наиболее часто при анализе многомерных данных используется график, на котором счета первой главной компоненты (ГК1) изображаются относительно второй (ГК2). Это и понятно, так как именно в этих двух направлениях происходят наибольшие изменения в исследуемых данных.

Графики счетов [5] применяются для идентификации групп, обнаружения тенденций, нахождения общих свойств и т.д.

После получения системы главных компонент исследователи работают разными способами в зависимости от своих целей:

1) Строят линейную регрессию интересующего технико-экономического показателя в новых ортогональных главных компонент переменных;

2) Рассматривают первые главные компоненты – обычно они несут в себе смысловую нагрузку части исходных параметров, т.к. объединяют в группы скоррелированные исходные параметры;

3) Рассматривают радиусы исходных параметров и радиусы точек измерений в системе главных компонент для разделения параметров и точек измерения на кластеры;

4) Рассматривают расположение исходных параметров x_1, \dots, x_p в новой координатной системе $ГК_1, \dots, ГК_p$ в соответствии с системой.

Для анализа выборка данных содержит 31 наблюдение и 18 переменных (11 основных (4 - технологические параметры, 7 – входы процесса) и 7 вспомогательных (выходы процесса)). Влияние на выход оказывали следующие переменные: температура (Т), объем воздуха ($V_{воз}$), объем раствора ($V_{р-ра}$), рН.

После анализа исходной информации с помощью метода главных компонент с использованием пакета STATISTICA, получены вклады переменных в каждую главную компоненту (таблица 1).

Таблица 1 - Вклады переменных в главные компоненты

	ГК1	ГК2	ГК3	ГК4
Fe общее	0.220410	0.011507	0.067480	0.000702
Cu	0.063182	0.015703	0.237833	0.000540
Ni	0.205437	0.063172	0.008330	0.000347

Pb	0.205634	0.004740	0.053537	0.002427
Zn	0.001759	0.424221	0.050816	0.027835
Ам	0.135648	0.070164	0.016079	0.094107
Температура	0.014018	0.050255	0.013273	0.619733
Vвоз	0.000793	0.044792	0.141217	0.180677
Ph	0.003470	0.046829	0.305001	0.008988
Vp-ра	0.005792	0.260004	0.005324	0.044656
% общей дисперсии	33.38689	14.16325	13.77898	9.961773

Первая главная компонента описывает наибольшее изменение в массиве данных, поэтому рассматривали ее. В данном случае, наибольший вклад в ГК1 внесли три переменные, а именно концентрации Fe, Ni, Pb на входе. Из рассматриваемых переменных для данного процесса именно эти три переменные сыграли решающую роль в изменении выходных параметров в исследуемый период. В таблице 2 представлен порог чувствительности и кумулятивный процент объяснения информации главными компонентами.

Таблица 2 – Порог чувствительности и кумулятивный процент объясненной информации

№ ГК	% общей дисперсии	Кумулятивный %
1	33.38689	33.3869
2	14.16325	47.5501
3	13.77898	61.3291
4	9.96177	71.2909
5	8.49316	79.7841
6	6.47124	86.2553
7	5.23728	91.4926
8	4.56525	96.0578
9	2.11362	98.1714
10	1.48349	99.6549
11	0.34507	100

Как видно из таблицы 2, первые четыре главных компоненты объясняют 71 % всех изменений исходных данных. При этом понималось, что 29% информации остается вне анализа. В методе главных компонент считается, что это маленькие жертвы, по сравнению с тем, что может быть получено при анализе графиков.

Далее было выявлено, какие факторы коррелируют с главными компонентами и проведена интерпретация главных компонент. Первая главная компонента наиболее тесно связана с концентрациями на входе: Fe ($R=-0,89$), Ni ($R=-0,87$), Pb ($R=-0,87$). В

этой связи, первая главная компонента, интерпретирована как химическая, нет влияния ни одного технологического параметра.

Вторая главная компонента, тесно связана с концентрацией Zn ($R=-0,81$) и Vp-ра ($R=-0,64$), была интерпретирована как химико-параметрическая

Третья главная компонента, тесно связана с концентрацией Cu ($R=-0,6$) и pH ($R=-0,68$), была интерпретирована как параметрическо-химическая.

Четвертая главная компонента, тесно связана T ($R=0,82$), была интерпретирована как температурная.

Кроме анализа на основе вклада переменных в главные компоненты, МГК позволяет определять характер зависимости между переменными. Для этого были построены графики счетов и нагрузок, в координатной плоскости первых двух ГК. График нагрузок представлен на рисунке 1.

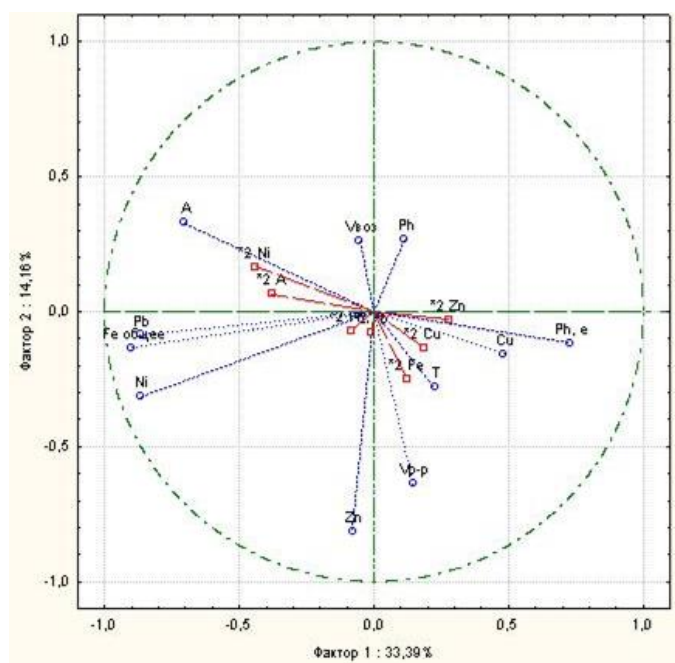


Рисунок 1 – График нагрузок

Данный график показывает, что переменные, расположенные в одном квадранте, имеют прямо пропорциональную зависимость, а в противоположных квадрантах — обратную пропорциональную зависимость. Так выход Ni, Ам имеют обратную пропорциональную зависимость с T, Vp-ра, pH. Выход Fe, Cu, Zn имеет прямо пропорциональную зависимость с T, Vp-ра, pH и обратную пропорциональную с Vвоз. Выход Pb имеет обратную пропорциональную зависимость с pH, и прямо пропорциональную с концентрацией на входе металлов: Pb, Ni, Zn, Fe.

С помощью графиков нагрузок была получена качественная картинка распределения параметров выхода по группам, и выявлена зависимость между ними, технологическими параметрами и параметрами входа.

Таким образом, недостатком семи методов, является низкая эффективность при проведении анализа сложных процессов. Использование метода главных компонент для анализа данных, выявления связей между параметрами процесса, его входами и выходами показало его несомненное преимущество по сравнению с «простыми» статистическими методами. Метод позволил показать влияние некоторых параметров

на процесс даже в условиях ограниченного объема информации (4 ГК объяснили 71% имеющийся информации).

Библиографический список:

1. Михалевич И.М., Примица С.П. Применение математических методов при анализе геологической информации (с использованием компьютерных технологий): учеб. пособ. / Ч.Ш. – Иркутск: Иркут. гос. ун-т, 2006. 115 с.
2. David H. DeYoung/ Analysis of a potroom performance drift, from a multivariate point of view // Light Metals. – 2008. – P. 319–324.
3. Большаков А.А., Каримов Р.Н. Методы обработки многомерных данных и временных рядов: учеб. пособ. для вузов. – М.: Горячая линия-Телеком, 2007. 522 с.: ил.
4. Лоули Д., Максвелл А.. Факторный анализ как статистический метод/ Пер. с англ. Ю.Н. Благовещенского. Издательство «Мир» Москва 1967. 144 с.
5. Радионова О.Е. Интервальный метод обработки результатов многоканальных экспериментов: дис. на соискание ученой степени доктора физико-математических наук; : защищена 29.05.08 : утв. 15.10.08. Москва, 2008. – 273 с.