

**КЛАССИФИКАЦИЯ ОБРАЗЦОВ ЦИРКОНА НА ОСНОВЕ ИНТЕНСИВНОСТИ
ХАРАКТЕРИСТИЧЕСКИХ ПОЛОС Al-P, Dy³⁺, Sm³⁺, Fe³⁺ - ЦЕНТРОВ
ЛЮМИНЕСЦЕНЦИИ**

Рассулов В.А. (rassulov@mail.ru)
Московское отделение. ВИМС

**CLASSIFICATION OF SAMPLES OF ZIRCON ON THE BASIS OF INTENSITY OF
CHARACTERISTIC BANDS Al-P, Dy³⁺, Sm³⁺, Fe³⁺ - THE LUMINESCENCE CENTERS**

Rassulov V.A.
Moscow branch. VIMS

Одной из основных задач поисковой геологии и прикладной геохимии является идентификация природных объектов. Принадлежность редкометалльных, полиметаллических и золоторудных объектов и связанных с ними геохимических ореолов к тому или иному рудно-формационному типу влияет на их экономическую оценку. Решение задачи идентификации формационного типа геологического объекта возможно на основе использования циркона, являющегося природным индикатором генезиса и формационной принадлежности геологических образований. В составе элементов-примесей, уровнях их содержаний, морфологии зерен минерала и их количественных соотношениях заключена информация о процессах рудообразования.

Полученные в последнее время данные о типоморфных особенностях минералов стимулируют поиск новых подходов к решению задач идентификации формационных типов с использованием более простых, менее затратных методов.

Одной из важнейших типоморфных характеристик минерала, отражающей состав элементов-примесей люминогенов, является его спектр люминесценции. Интенсивность полос в спектре, связанных с индивидуальными люминогенами, может быть выражена количественно, а значит, является информацией приемлемой для математической обработки. И возможно создание минералогическо-геохимической экспертной системы на основе нормированных интенсивностей полос оптически активных центров (ОАЦ) в спектре люминесценции с использованием вероятностно-статистических методов.

Теоретическим основанием для решения задач распознавания является теорема компактности (Дорохов, 1989). Процедура «обучения» при распознавании складывается из трех последовательных операций: измерение; отбор признаков; классификация. В соответствии методическими рекомендациями (Рассулов, 2005) измеряются нормированные интенсивности характеристических полос ОАЦ - Al-P, Dy³⁺, Sm³⁺, Fe³⁺ в спектре люминесценции цирконов с задержкой регистрации на 180 мкс при возбуждении излучением лазера на молекулярном азоте ($\lambda_{\text{изл}} = 337,1 \text{ нм}$) (Int (Al-P, Dy³⁺, Sm³⁺, Fe³⁺)), характеризующие исследуемые объекты.

Под операцией классификации понимается процедура отнесения объекта к тому или иному классу согласно их расположению в пространстве признаков. Если признаки содержат достаточное количество информации об исследуемом процессе, то можно предположить, что и объекты с близкими свойствами будут группироваться в одной и той же ограниченной области многомерного пространства.

Для определения критериев и разработки процедуры принятия решения (Фор, 1989) о принадлежности образцов циркона по классификационным функциям к одной из априорных групп (Фролов, 2005) на основе Int (Al-P, Dy³⁺, Sm³⁺, Fe³⁺) была исследована представительная коллекция: алмазоносные кимберлиты (группа G1) - 7 трубок, 47 образцов; слабоалмазоносные кимберлиты (G2) - 17 трубок, 104 образца; неалмазоносные кимберлиты (G3) - 10 трубок, 103 образца; карбонатиты линейных зон (G4) - 2 объекта, 54 образца; карбонатиты кольцевых структур (G5) - 2 объекта, 88 образцов. Фигуративные точки образцов циркона из различных по алмазоносности кимберлитов и карбонатитов линейных и кольцевых структур в координатах Int (Al-P, Dy³⁺, Sm³⁺, Fe³⁺) располагаются компактно, но области точек которые они образуют, нельзя охарактеризовать как полностью не пересекающиеся. Следовательно, Int (Al-P, Dy³⁺, Sm³⁺, Fe³⁺) в отдельности не являются четкими критериями принадлежности объекта к одной из групп.

Задача классификации может быть сформулирована следующим образом: найти к какой группе из априорного множества групп принадлежит новый образец, заданный совокупностью значений предварительно установленных Int (Al-P, Dy³⁺, Sm³⁺, Fe³⁺) в спектре люминесценции. Построение функции принадлежности основана на статистической обработке исходной матрицы данных:

$X = \|X_{ij}\|$, где $i = 1, M$ ($M=5$; 1- номер группы, 2-5 значения Int (Al-P, Dy³⁺, Sm³⁺, Fe³⁺)); $j = 1, N$ ($N=408$, число образцов циркона в обучающей выборке).

Критерий принадлежности образца циркона к одной из групп определяют решающим правилом. Решающее правило построено с использованием дискриминантного анализа (Ким, 1989).

Значения F-статистики (Фишера), которые характеризуют различия между группами, $F(16; 1222; 0,01) = 311,5 > F(\text{табличное})$, равное 1,89, поэтому можно отвергнуть нулевую статистическую гипотезу о неразличии между группами на уровне значимости $\alpha = 0,01$. Теперь необходимо выяснить вклад дискриминантных переменных в различие между группами: $F(4; 400; 0,01) > F(\text{табличное})$ - поэтому можно утверждать, что четыре исходные переменные – Int (Al-P, Dy³⁺, Sm³⁺, Fe³⁺) в цирконе значимы при разделении объектов на пять групп при уровне значимости $\alpha = 0,01$ в следующей последовательности: Int (Al-P) < Int (Dy³⁺) < Int (Sm³⁺) < Int (Fe³⁺).

Величина собственного значения ДФ связана с дискриминирующей способностью этой функции: чем больше собственное значение, тем лучше различение. Расчет показал, что кумулятивная доля 1, 2 и 3 ДФ составляет 0,99, а доля четвертой ДФ незначительна. Основная доля дискриминирующей способности приходится на 1 и 2 ДФ, которая составляет 93% при значимом доминировании первой ДФ с 71%.

Другой способ оценки меры значимости ДФ через значения коэффициентов канонической корреляции, которые являются мерой связи с группами. Большая величина коэффициента канонической корреляции для 1, 2 и 3 ДФ указывает на сильную зависимость между группами и соответствующими ДФ. С другой стороны, коэффициент канонической корреляции для четвертой ДФ имеет очень малую величину, которая говорит о слабой связи.

Таблица 1

Остаточная дискриминантная способность и проверка значимости

Номер шага (k)	U-статистика Уилкса	χ^2	Число степеней свободы	Уровень значимости
0	0,007	1997,5	16	0,000
1	0,093	956,3	9	0,000
2	0,444	326,5	4	0,000
3	0,999	0,50	1	0,478

Теоретически можно получить (g-1), где g- число групп объектов (g =5), четыре дискриминантные функции. Необходимо определить число статистически значимых ДФ.

Статистическую значимость ДФ проверили по остаточной дискриминантной способности (способность переменных различать группы, если исключить информацию полученную с помощью ранее вычисленных функций) по так называемой U-статистике Уилкса.

В табл. 1 представлены значения остаточной дискриминантной способности до определения очередной дискриминантной функции по шагам. На «нулевом шаге» ее значение до определения 1 ДФ, на 1 шаге после извлечения 1 ДФ и остаточная дискриминантная способность до извлечения 2 ДФ и так далее. Значение U-статистики Уилкса, которая является «обратной» мерой, на нулевом шаге (k=0) имеет очень низкое значение, равное 0,007. Величина U-статистики Уилкса, близкое к нулю, говорит о четком различии, т.е. центры групп хорошо разделены и сильно различаются друг от друга по отношению к степени разброса внутри групп. Поэтому возможно определить дискриминантные функции.

После определения первой и самой значимой дискриминантной функции остается достаточно остаточной дискриминантной способности для определения второй ДФ (U-статистика = 0,093). Вычисление второй функции (k=1) уменьшает остаточную дискриминантную способность величины U-статистики Уилкса до 0,444, но эта величина еще значима для определения 3 ДФ. После расчета третьей ДФ значительно уменьшается остаточная дискриминантная способность и величина U-статистики Уилкса, равное 0,998, перестает быть значимой.

Это значение близкое к 1 говорит о том, что оставшаяся четвертая дискриминантная функция статистически недостоверна и не обладает информацией о различении центров групп.

На основе U-статистики Уилкса можно получить тест значимости, аппроксимируя функцией распределения χ^2 (хи-квадрат) с $(p-k)$ ($g-k-1$) степенями свободы, где соответственно: p – число дискриминантных переменных ($p=4$); g – число групп ($g=5$); k – номер шага.

В табл. 1 приведены значения χ^2 – статистики для исходных данных о различии центроидов групп. Между центроидами групп есть значимые различия еще до вычисления ДФ ($k=0$), т.к. $\chi^2_{\text{эксп.}} \gg \chi^2_{\text{табл.}}$ на уровне значимости $b \gg 0,001$. После определения значение χ^2 – статистики ($k=1$ и $k=2$, соответственно) стало меньше, но уровень значимости достаточно высок ($\chi^2=956$; $\chi^2=326$), а после вычисления 3 ДФ значение χ^2 – статистики стало меньше ($\chi^2=0,5$), а уровень значимости стал равным $b=0,478$ ($k=3$), то есть вся значимая информация о различии групп уже извлечена. Таким образом, можно утверждать, что первая, вторая и третья ДФ являются статистически значимы в целом как система для разбиения образцов на пять групп.

Таблица 2

Стандартизованные дискриминантные коэффициенты

Наименование переменной	Номер дискриминантной функции			
	1	2	3	4
Int (Al-P)	-0,43	-0,32	0,03	-0,85
Int (Dy ³⁺)	0,32	0,14	-0,91	-0,21
Int (Sm ³⁺)	0,73	0,29	0,43	-0,45
Int (Fe ³⁺)	-0,50	0,86	-0,01	-0,10

В табл. 2 представлены значение стандартизованных дискриминантных коэффициентов, которые применяют для выявления тех переменных, которые вносят наибольший вклад в значение дискриминантной функции: чем они больше по абсолютной величине, тем больше вклад этой переменной. Это является одним из критериев значимости переменных.

Анализ полученных результатов показывает, что во первых, для первой ДФ переменные по знаку подразделяются следующим образом: с положительным знаком – интенсивности Dy³⁺-, Sm³⁺- центров, с отрицательным Al-P-, Fe³⁺- центров, во вторых по абсолютному значению доминируют переменные – интенсивности Sm³⁺- и Fe³⁺- центров. Для второй ДФ характерно различия по знаку переменных – интенсивности Al-P- и Fe³⁺- центров, имеющих большие по абсолютной величине значения. В таблице 3 представлены значения группового «центроида» для четырех дискриминантных функций.

Таблица 3

Значения групповых «центроидов» для четырех дискриминантных функций

Группа	Номер дискриминантной функции		
	1	2	3
G 1	-4,438	-1,905	2,459
G 2	-3,025	-0,036	-1,061
G 3	-1,081	2,222	-0,135
G 4	2,622	-3,818	-1,000
G 5	5,106	0,741	0,626

При сопоставлении расположения фигуративных точек стандартизованных дискриминантных коэффициентов Int (Al-P, Dy³⁺, Sm³⁺, Fe³⁺) и групповых «центроидов» в координатах первой и второй ДФ можно отметить, что, во первых, переменные Int (Al-P, Fe³⁺) расположены в левой стороне плоскости 1 и 2 ДФ и в той же области расположены фигуративные точки цирконов из кимберлитов (рис. 1), а фигуративные точки цирконов из карбонатитов расположены в правой стороне плоскости. При этом фигуративные точки карбонатитов линейных зон расположены в нижней стороне плоскости, а кольцевых структур в верхней.

На рис. 1 представлено расположение фигуративных точек цирконов в координатах первой и второй ДФ. Видно, что образцы G1, G2, G3 и G4 групп расположены более компактно, около групповых центроидов в отличие от образцов из карбонатитов линейных зон (G5). Заметен тренд увеличения значения первой дискриминантной функции в ряду от алмазоносных кимберлитов к карбонатитам кольцевых структур в координатах первой и третьей дискриминантной функции.

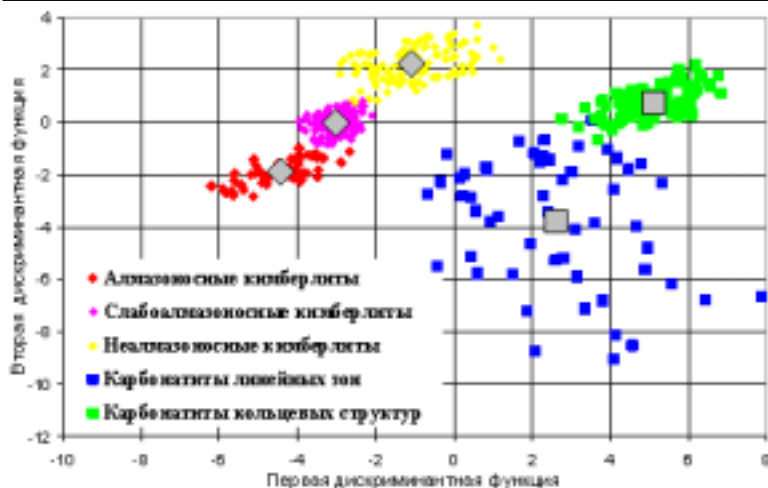


Рис. 1. Расположение фигуративных точек цирконов и групповых «центроидов» в координатах первой и второй дискриминантной функции.

Классификация – это процесс, который помогает принять решение: «принадлежит к» или «очень похож на» указанный образец к данной группе. Такое решение принимается на основе информации которая содержится в классификационных функциях представленных в таблице 4.

Таблица 4

Коэффициенты классификационных функций

Исходные переменные	Группа				
	G1	G2	G3	G4	G5
Int(Al-P)	76,28	72,88	68,68	69,50	63,13
Int(Dy ³⁺)	3,34	7,51	7,60	8,82	8,79
Int(Sm ³⁺)	-1,05	-0,99	0,92	1,36	4,37
Int(Fe ³⁺)	31,25	32,85	34,46	22,65	27,15
Константа	-587,96	-593,98	-575,25	-478,44	-475,95

Классификационные функции имеют следующий вид: $h_k = b_{k0} + b_{k1}X_1 + \dots + b_{kp}X_p$

где h_k – значение функции для класса K; b_{kn} – коэффициенты классификационных функций (табл. 4); X_n – значения нормированных интенсивностей характеристических полос в спектре люминесценции циркона.

Классификационные функции можно использовать для определения принадлежности нового образца к одной из групп применяя эти пять классификационных функций к новым первичным данным и получают пять значений h_k . Образец относится к группе с максимальным значением h_k .

Таким образом, построена статистическая модель по нормированным интенсивностям характеристических полос Al-P, Dy³⁺, Sm³⁺, Fe³⁺ - центров люминесценции в цирконе методом дискриминантного анализа и определены критерии отнесения образца по этим данным.

Алгоритм построения статистической модели состоит из следующих этапов: создание базы данных нормированных интенсивностей характеристических полос; выдвижение гипотезы о различении зерен циркона по Int (Al-P, Dy³⁺, Sm³⁺, Fe³⁺); проведение проверки статистической гипотезы о различении индивидов циркона по люминесцентным характеристикам методом дискриминантного анализа; определение критериев и разработка процедуры принятия решения о принадлежности образца циркона по классификационным функциям к одной из априорных групп на основе нормированных интенсивностей ОАЦ.

Список литературы

Дорохов И.Н., Кафаров В.В. Системный анализ процессов химической технологии: Экспертные системы для совершенствования промышленных процессов гетерогенного катализа. М.: Наука, 1989. 376 с.

Ким Дж.-О., Мюллер Ч.У., Клекка У.Р. и др. Факторный, дискриминантный и кластерный анализ. М.: Финансы и статистика. 1989. 215 с.

Рассулов В.А. Локальная лазерная с учетом кинетики затухания люминесцентная спектроскопия минералов (на примере циркона). Методические рекомендации № 156. М.: ВИМС. 2005. 16 с.

Фор А. Восприятие и распознавание образов. М.; Машиностроение, 1989. 272 с.

Фролов А.А., Лалин А.В., Толстов А.В., Зинчук Н.Н., Белов С.В., Бурмистров А.А. Карбонатиты и кимберлиты (взаимоотношения, минерагения, прогноз). М.: НИА-Природа, 2005. 540 с.