

Инструкция к веб-приложению ArDI 1.4

Библиографическая ссылка, которую необходимо сделать при публикации результатов, полученных с применением ArDI: Шендрик Р.Ю., Плечов П.Ю., Смирнов С.З. [ArDI – система обработки и анализа колебательных спектров минералов](#), Новые данные о минералах, 58 (2024) стр. 26-35, doi: 10.25993/FM.2024.58.2024.008. Текст статьи доступен по ссылке https://fmm.ru/images/2/28/Shendrik_1_2024-2.pdf.

Оглавление

Инструкция к веб-приложению ArDI 1.4	1
1. Обзор вкладок веб-приложения ArDi.....	2
1.1 Вкладка Search.....	2
1.1.1 Примерный алгоритм поиска спектра в базе данных.....	3
1.1.2 Возможные сложности при поиске спектров	3
1.1.3 Загрузка файлов.....	4
1.1.4 Работа с базами данных	4
1.1.5 Поиск похожих спектров в базах данных	7
1.1.6 Рамановский фазовый анализ (RPA).....	8
1.1.7 Изменение загруженного спектра перед поиском	9
1.2 Вкладка Deconvolution	14
1.2.1 Примерный алгоритм разложения спектра на компоненты (деконволюция)	15
1.2.2 Обработка спектра	16
1.2.3 Автоматический поиск пиков	19
1.2.4 Параметры базовой линии	23
1.2.5 Работа с графиками	26
1.2.6 Обрезка спектра	26
1.2.7 Разложение спектра на пики (деконволюция)	27
1.2.8 Добавление спектров из базы данных на график	31
1.2.9 Вычитание спектра минерала хозяина.....	31
1.3 Добавление спектров в базу данных.....	33

1. Обзор вкладок веб-приложения ArDi

1.1 Вкладка Search

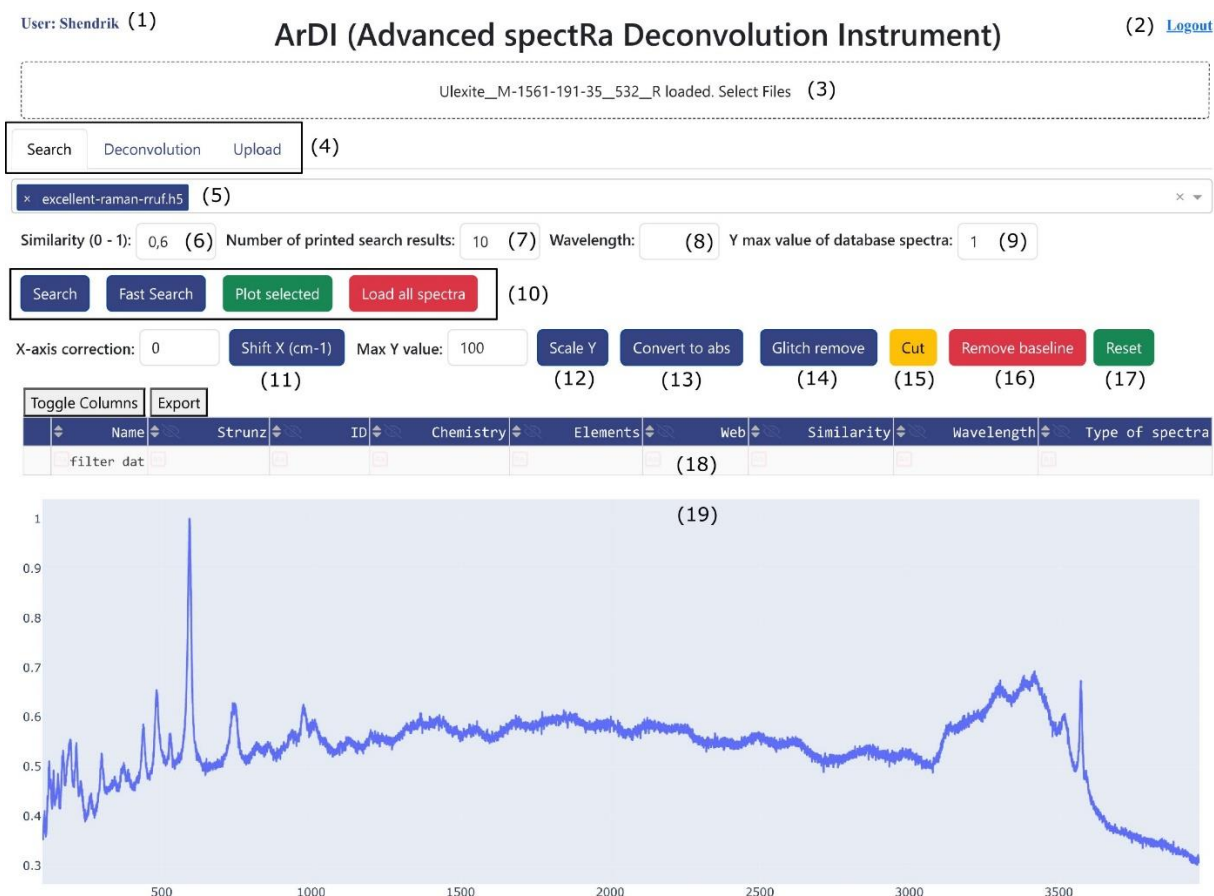


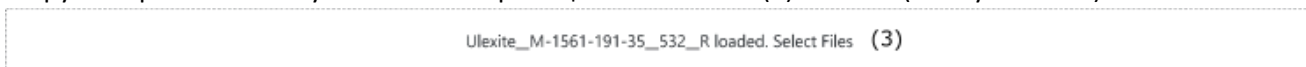
Рис. 1.1 Вкладка Search. Вкладка состоит из: (1) имя пользователя, (2) выход из профиля пользователя, (3) область для загрузки файла со спектром, (4) область основных вкладок: Search, Deconvolution, Upload, (5) название базы данных, в которой ведется поиск схожих спектров при нажатии кнопки Search, (6) значение похожести, при нажатии кнопки Search будут выводиться в таблицу результатов поиска (18) только спектры, похожест которых выше этого значения, (7) количество результатов поиска, которое выводится в таблице (18), (8) – значение длины волны, поиск в базе будет проводится только по Рамановским спектрам, измеренным с длиной волны возбуждения, указанной в этом поле, по умолчанию значение не задано, поиск ведется по всем спектрам, (9) – максимальное значение, к которому приводится значение по оси ординат выводимых по нажатию кнопки Plot Selected найденных в базе данных спектров, по умолчанию 1, (10) – область кнопок для поиска (Search), поиска с использованием машинного обучения (Fast Search), графического отображения спектров в таблице (18) (Plot Selected) и выгрузки всех записей из указанных в поле (5) баз данных (Load All Spectra), (11) – сдвиг загруженного спектра (3) по оси X, (12) – приведение максимального значения оси ординат загруженного спектра (3) к значению указанному в поле, (13) – преобразование загруженного спектра из координат пропускания в координаты поглощения, (14) – удаление из выделенной на загруженном спектре (3) области глитчей или 'cosmic rays', (15) – обрезать загруженный спектр до выделенной на графике области по оси x, (16) – автоматическое удаление базовой линии из спектра, (17) – отмена всех ранее примененных операций (11)-(17), (18) – таблица с результатами поиска или выгрузки из баз данных, (18) – поле графика, где отображается загруженный спектр (3) и выбранные спектры из баз данных при нажатии кнопки Plot selected (10).

При открытии веб-приложения ArDi пользователь после ввода логина и пароля попадает на вкладку Search. Эта вкладка доступна всем пользователям, как привилегированным, которые имеют персональный логин и пароль, так и гостям. Для входа с гостевой учетной записью достаточно в поле логин ввести **guest**, пароль не требуется.

Вкладка Search предназначена для поиска и просмотра спектров в представленных в различных базах данных, выбор которых осуществляется в поле (5). Существует возможность выбора сразу нескольких баз данных. Краткий обзор вкладки Search приведен на Рис. 1.1. Ниже будет приведено подробное описание возможных на этой вкладке действий.

1.1.1 Примерный алгоритм поиска спектра в базе данных

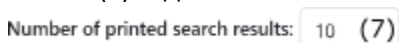
- 1) Загрузите файл в систему нажав или перетащив его на поле (3) Рис. 1.1 (см. пункт 1.1.3).



- 2) Выберите необходимую базу данных в поле (5)



- 3) В поле (7) задайте количество выводимых спектров, можно указать значение побольше



- 4) Попробуйте поискать спектр в выбранной базе нажатием кнопки Search в поле (10) (см. пункт 1.1.5)



- 5) Результаты будут в таблице (18)

- 6) Если поиск выдает неподходящие минералы, тогда попробуйте вычесть базовую линию



спектра нажатием кнопки (16) Remove baseline (16) (см. пункт 1.1.7.6)

- 7) Проведите поиск снова

- 8) Если результат не устраивает вас, попробуйте обрезать спектр (см. пункт 1.1.7.5). Для этого выделите нужную область спектра на графике (19). Рекомендуем выделять область от 100 до



1600 см⁻¹. Далее нажмите кнопку (15) Cut (15).

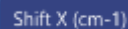
- 9) Постройте найденные в базе спектры для визуального сравнения с загруженным. Для этого оставьте нужные спектры в таблице (18) с помощью фильтров (см. пункт 1.1.4) и нажмите



кнопку Plot selected в поле (10)

1.1.2 Возможные сложности при поиске спектров

Спектры комбинационного рассеяния зачастую содержат достаточно узкие линии, поэтому сдвиг спектра минерала в базе данных или загруженного спектра существенно влияет на результаты. Сдвиг спектра может быть вызван ошибкой в базе данных или отсутствием коррекции при измерении спектров. Особенно чувствительным к сдвигам может быть поиск фаз, в случае, когда линии у минерала в области поиска недостаточно интенсивны. В таком случае рекомендуется на несколько обратных сантиметров сдвигать спектр перед поиском путем нажатия кнопки Shift X (см-1).



Вторым фактором, который может существенно ухудшить результаты поиска, является наличие у спектра фона или базовой линии. Ее перед поиском лучше вычитать. В некоторых случаях базовая линия не влияет на результаты, но зачастую она может оказать существенное влияние.

Также алгоритмы поиска чувствительны к наличию интенсивных линий. Поэтому если интенсивная линия находится в области низких частот, а остальные линии в спектре по тем или иным причинам не

столь интенсивны, то будут находится в первую очередь спектры с интенсивной низкочастотной линией. Часто это соли тяжелых металлов. В этом случае рекомендуем ограничить область поиска, выделив ее на графике и нажав кнопку Cut .

1.1.3 Загрузка файлов

Веб-приложение ArDi умеет искать схожие загруженному спектры в базах данных. Для загрузки необходимо проделать следующие операции. Файл со спектром загружается его перемещением с помощью мышки на область Drag and Drop or Select Files или кликом по ней мышкой левой или правой кнопкой (Рис. 1.1). Во втором случае открывается диалог открытия файла.

В настоящее время поддерживаются следующие форматы:

1. .txt – двухколоночный текстовый файл с разделителями между колонками пробел и символ табуляции (tab)
2. .esp – текстовый формат EnSpectr
3. .csv – csv формат

После успешной загрузки в области загрузки появится имя файла, а также спектр на графиках вкладок Search (19) (Рис. 1.1) и Deconvolution. Файл будет находится в памяти браузера до тех пор, пока не будет обновлена страница, или загружен новый файл, старый при этом очищается из памяти.

1.1.4 Работа с базами данных

На Рис. 1.1 в поле (5) выбираются базы данных для работы. Могут быть выбраны сразу несколько баз данных из списка. С базами данных доступно несколько операций, таких как выгрузка всех записей базы данных в таблицу (18) (Рис. 1.1), отображение не более чем 20 первых записей из таблицы (18) в поле графика (19) (Рис. 1.1). Рассмотрим эти операции более подробно.

1.1.4.1 Просмотр всех записей в базе данных и построение спектров

(1) (2) (5) (4)

Name	Strunz	ID	Chemistry	Elements	Web	Wavelength	Type of spectra
ata...							
070007	10.CA.20	R070007	NiC31H32N4	Ni C H N	RRUF Mindat	532	Proceed
070007	10.CA.20	R070007	NiC31H32N4	Ni C H N	RRUF Mindat	532	RAW
060227	3.DA.30	R060227	Sn2+2106(OH)14Cl16	Sn O H Cl	RRUF Mindat	532	Proceed
060227	3.DA.30	R060227	Sn2+2106(OH)14Cl16	Sn O H Cl	RRUF Mindat	532	RAW
060227	3.DA.30	R060227	Sn2+2106(OH)14Cl16	Sn O H Cl	RRUF Mindat	780	Proceed
060227	3.DA.30	R060227	Sn2+2106(OH)14Cl16	Sn O H Cl	RRUF Mindat	780	RAW
Actinolite_R040063-4	9.DE.10	R040063-4	Ca2(Mg4.5-2.5Fe2+0.5-2.5)Si8O22(OH)2	Ca Mg Fe Si O H	RRUF Mindat	514	RAW
Actinolite_R040063-4	9.DE.10	R040063-4	Ca2(Mg4.5-2.5Fe2+0.5-2.5)Si8O22(OH)2	Ca Mg Fe Si O H	RRUF Mindat	514	Proceed
Actinolite_R040063-4	9.DE.10	R040063-4	Ca2(Mg4.5-2.5Fe2+0.5-2.5)Si8O22(OH)2	Ca Mg Fe Si O H	RRUF Mindat	514	RAW
Actinolite_R040063-4	9.DE.10	R040063-4	Ca2(Mg4.5-2.5Fe2+0.5-2.5)Si8O22(OH)2	Ca Mg Fe Si O H	RRUF Mindat	514	Proceed

(6) (7) << < 1 / 1873 > >>

Рис. 1.2 Таблица, содержащая все записи из выбранной базы данных. (1) – Кнопка выбора скрытых столбцов. Для того, чтобы показать скрытый столбец необходимо поставить напротив его названия галочку. (2) – экспорт таблицы в формате csv, (3) – поле для применения фильтра к столбцу. (4) – сортировка значений в столбце. (5) – скрыть столбец, (6) – удалить запись из таблицы, (7) – перелистывание на следующую страницу таблицы.

По нажатию кнопки Load all spectra (10) данные из базы загружаются в таблицу (18) (Рис. 1.1). Результат такой загрузки показан на рис. Рис. 1.2. Далее с этой таблицей можно проводить следующие операции: скрывание и показ скрытых столбцов, сортировка по отдельным столбцам, применение различных фильтров к столбцам, удаление записей из таблицы и т.д.

Таблица состоит из колонок с названием образца (Name), согласно его имени в базе данных, индекса образца по Штрунцу (Strunz), если минерал с таким названием находится в списке MMA, идентификационный номер в базе данных (ID), кристаллохимическую формулу (Chemistry), последовательность химических элементов, которые содержит образец (Elements). В столбце Web содержатся ссылки на карточку образца на сайте базы данных (например rruf.info, <https://solsa.crystallography.net/rod/>, fmm.ru), а также ссылку на страницу минерала на сайте <https://mindat.org>, если такой минерал там имеется. Также доступны столбцы Wavelength (длина волны возбуждения при съемке Рамановских спектров), Type of spectra – тип спектра (исходный спектр без обработки (RAW) и обработанный спектр, в котором например была вычтена базовая линия, проведено сглаживание и т.п. (Proceed)). В столбце Comments могут содержаться некоторые комментарии к записи в базе данных.

Для того, чтобы скрыть столбец необходимо нажать на кнопку с изображением глаза (5). Для того, чтобы показать скрытый столбец необходимо нажать на кнопку Toggle Columns (1) и поставить галочку у интересующего скрытого столбца.

Для удаления записи из таблицы необходимо нажать крестик в нужной строке таблицы (6). Для перелистывания таблицы используется указатель (7). Для удобства на одну страницу таблицы выводится 10 записей из базы данных.

Имеется возможность применения фильтров ко всем столбцам. Более подробно о доступных возможностях фильтрации столбцов описано по ссылке: <https://dash.plotly.com/datatable/filtering>

Например удобен фильтр по классификации Штрунца, который позволяет показать лишь минералы определенной группы по классификации Никеля-Штрунца на основе классификации Mindat (<https://www.mindat.org/strunz.php>).

Click on a group to view further divisions.	
1	ELEMENTS (Metals and intermetallic alloys; metalloids and nonmetals; carbides, silicides, nitrides, phosphides)
2	SULFIDES and SULFOSALTS (sulfides, selenides, tellurides; arsenides, antimonides, bismuthides; sulfarsenites, sulfantimonites, sulfbismuthites, etc.)
3	HALIDES
4	OXIDES (Hydroxides, V[5,6] vanadates, arsenites, antimonites, bismuthites, sulfites, selenites, tellurites, iodates)
5	CARBONATES (NITRATES)
6	BORATES
7	SULFATES (selenates, tellurates, chromates, molybdates, wolframates)
8	PHOSPHATES, ARSENATES, VANADATES
9	SILICATES (Germanates)
10	ORGANIC COMPOUNDS

Для использования фильтра в по колонке Strunz в строке фильтра в данной колонке (3) (Рис. 1.2). можно вводить желаемые опции:

- 1) Чтобы оставить только фосфаты, в поле вводится 8. И нажимается Enter
- 2) Также можно использовать символы >, <, = и .т.п., чтобы, например вывести минералы с группой по номеру больше 4. (> '5').

Еще один из возможных вариантов применения фильтров – столбцы Chemistry и Elements. Возможно указать структурные комплексы, например (Si₂O₇), так и допустимые химические элементы.

Возможно применять фильтры одновременно к нескольким колонкам.

Первые 20 записей в таблице можно отобразить на графике (19) Рис. 1.1. Для этого следует нажать кнопку Plot selected spectra (10) Рис. 1.1. Максимальное значение по оси у у отображаемых спектров указано в поле Y max value of database spectra (9) Рис. 1.1.

При загрузке спектров из базы данных до загрузки спектра из файла (3) Рис. 1.1. Последний может сразу не отображаться на графике (19). Для его отображения необходимо нажать сначала кнопку Remove baseline (16), а потом, чтобы вернуть спектр к первоначальному виду Reset (17).

1.1.5 Поиск похожих спектров в базах данных

ArDi умеет искать похожие спектры в выбранных базах данных селектор (5) Рис. 1.1. В версии 1.1 реализовано две возможности поиска.

Первый вариант – поиск последовательным сравнением загруженного спектра и спектра из базы данных. Поиск может осуществляться как по спектру целиком, так и по какой-то их его частей. В качестве меры схожести спектров используется евклидова метрика. Для этого происходит приведение спектров к одной размерности, далее вычисляется косинус угла между многомерными векторами, которые представляют собой значения по оси ординат у спектров. Похожесть спектров определяется как косинус угла между их значением ординат:

$$Similarity \equiv \cos \alpha = \frac{\sum_{i=1}^N y_i^{exp} y_i^{base}}{\sqrt{\sum_{i=1}^N (y_i^{exp})^2} \sqrt{\sum_{i=1}^N (y_i^{base})^2}} \quad (1),$$

где y_{exp} - спектр, который ищется, y_{base} - спектр из базы данных, N - число точек в спектре. Метрика схожести спектров выводится в результатах поиска, она лежит в интервале от 0 до 1, где 1 означает полностью идентичные спектры, а 0 совершенно отличные друг от друга. В целом эта метрика схожа с коэффициентом корреляции Пирсона, однако использование Евклидовой метрики более корректно с математической точки зрения [Tap, 2005]. Результаты поиска отображаются в таблице (18). Методы работы с результатами поиска описаны в предыдущем разделе 1.1.4.1. Результаты поиска могут быть ограничены значением Similarity, которое задается в поле (6) Рис. 1.1. В этом случае выводится в таблицу будут лишь те спектры из базы данных, значение Similarity которых, рассчитанное по формуле (1), выше чем указанное в поле (6).

Вторым способом ограничения поиска является количество выводимых результатов. Поиск будет все равно проводится сравнением загруженного спектра со всеми спектрами в указанных базах данных, но выводиться в таблицу будут только количество значений, указанное в поле (7) Number of printed search results.

Третьим способом ограничения поиска в случае поиска по Рамановским спектрам является указание длины волны возбуждения при съемке Рамановского спектра. Значение длины волны задается в поле Wavelength (8) Рис. 1.1. Отметим, что если какое-либо значение в этом поле указано, то процедура сравнения спектров будет проводится только в том случае, если длины волны возбуждения спектра в базе данных в точности равна длине волны указанной в поле (8). Это позволяет ускорить поиск, но если по какой-то причине в базе данных не указана длина волны возбуждения, то сравнение спектров проводится не будет.

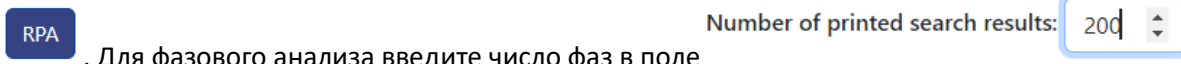
Второй вариант поиска использует методы машинного обучения, основанные на библиотеке FAISS (<https://faiss.ai/index.html>). Поиск осуществляется по нажатию кнопки Fast Search (10) Рис. 1.1. Результаты поиска также выводятся в таблицу (18). Результаты в таблице будут отсортированы от наиболее близких к наиболее далеким. Количество выводимых записей задается в поле (7) Number of printed search results. Имеет смысл задать достаточно большие значения, например 100 или 200. Поиск проводится мгновенно.

Поиск осуществляется следующим образом. Заранее был составлен индекс по спектрам из баз данных Рамановских спектров rruf. Далее проведено обучение модели по этому индексу. Полученная обученная модель используется для поиска. В настоящее время идет подбор параметров индекса и обучения для большей точности в выдаваемых результатах.

При нажатии кнопки Fast Search параметры длины волны и похожести, а также база данных не учитывается.

1.1.6 Рамановский фазовый анализ (RPA)

В некоторых случаях измеренный спектр содержит линии от нескольких минералов. Такое часто происходит при измерении мелкодисперсных образцов, минералов во включениях и т.п. Таким образом, методы поиска, приведенные выше, могут не дать верного результата. Для авторизованных пользователей доступна опция поиска пиков от нескольких минералов в спектре по нажатию кнопки



. Для фазового анализа введите число фаз в поле

Для начала введите небольшое значение в этом поле, например 10. Если спектр основного компонента в базе данных имеет много записей, тогда при поиске смесей будет выводиться только основной компонент. В этом случае для можно попробовать увеличить значение в поле количество выводимых результатов вплоть до нескольких тысяч или десятков тысяч. Обычно устойчивые результаты получаются при значении этого параметра 200-1000. Результаты фазового анализа будут приведены в таблице (18) Рис. 1.1. В поле Similarity будет отображаться вклад данного минерала в спектр. Спектры минералов с наибольшим вкладом показаны на спектре (19) Рис. 1.1.

Этот метод может быть полезен и при поиске минерала по монофазному спектру в том случае, если вышеприведенные алгоритмы сработали неверно. Например на рис Рис. 1.3 показан спектр кимрита, который опознается при поиске как азурит или шпинель, из-за интенсивной полосы в области 400 см-1 и смазанных низкоинтенсивных остальных полос. Однако RPA метод сумел обнаружить наибольший вклад кимрита в спектр.

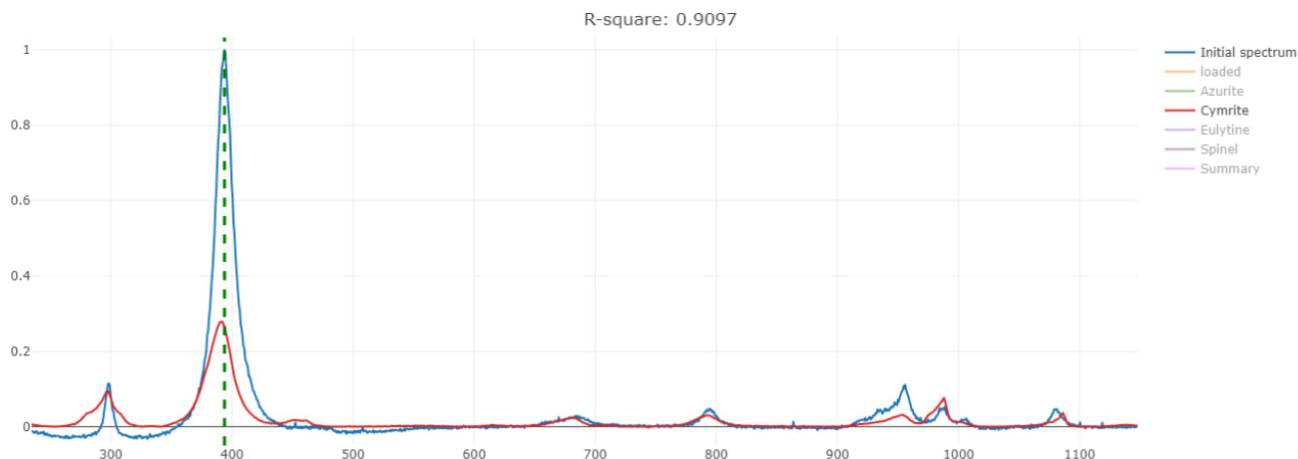


Рис. 1.3 Спектр кимрита, который в поиске может опознаваться как шпинель или азурит из-за единственной интенсивной полосы в области 400 см⁻¹

Алгоритм RPA устроен следующим образом. Сначала производится поиск всех схожих спектров с использованием индекса FAISS. В качестве количества обнаруженных компонент как раз и вводится

Number of printed search results:

значение из поля. Поэтому рекомендуем указывать большие значения в этом поле, чтобы обеспечить наилучший результат.

Найденные алгоритмом Faiss спектры объединяются в матрицу для того, чтобы методом наименьших квадратов подобрать коэффициенты в следующем уравнении линейной регрессии:

$$f_{w,b}(x) = w_0 x_0 + w_1 x_1 + \dots + w_n x_n,$$

Где x_i – i -спектр из найденных FAISS, w_i – его вес. Далее производится подбор параметров w_i методом наименьших квадратов так, чтобы $f_{w,b}(x)$ была наиболее близка к загруженному спектру. Иными

словами решается уравнение $w = (X^T X)^{-1} X^T y$, где X – матрица, составленная из найденных FAISS спектров, y – загруженный спектр, w – веса загруженных спектров.

Отрицательные веса не имеют большого смысла, поэтому должны быть отброшены. Рекурсивно эта процедура повторяется с уже перестроенной матрицей X до тех пор, пока решением уравнения не будут являться только положительные веса w .

1.1.7 Изменение загруженного спектра перед поиском

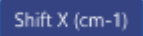
Зачастую загруженный спектр имеет ряд недостатков, например ненулевая базовая линия, наличие на спектре выбросов, сдвиг спектра по оси x при дрефте детектора во время съемки. Это может негативно сказаться на результатах поиска.

На вкладке Search доступны простейшие операции по редактированию спектра перед поиском. Более развернутые инструменты редактирования спектра доступны на вкладке Deconvolution.

1.1.7.1 Сдвиг спектра по оси абсцисс

При съемке спектра положение нуля на оси абсцисс X может не соответствовать нулевому Рамановскому сдвигу. Это происходит из-за того, что не проведена калибровка спектрометра перед началом измерения или из-за дрейфа детектора из-за нагрева при длительных измерениях. Сдвиг может достигать до 10 см⁻¹. Такой сдвиг может исказить результаты поиска. Поэтому в веб-приложении реализована возможность сдвига спектра по оси X.

Для этого необходимо загрузить спектр (3), ввести необходимое значение сдвига в поле X-axis

correction и нажать кнопку  Shift X (cm⁻¹) (11) (Рис. 1.1). Отрицательное значение сдвига соответствует сдвигу в область меньших волновых чисел, положительное в область больших.

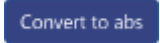
1.1.7.2 Нормировка спектра по оси ординат

Загруженный спектр можно отнормировать к значению, указанному в поле Max Y. Для этого необходимо загрузить спектр в поле (3), далее указать требуемое значение в поле Max Y и нажать

клавишу  Scale Y (12) (Рис. 1.1). После нажатия кнопки Scale Y максимальное значение по оси Y на спектре будет равно указанному в поле Max Y.

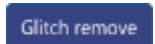
1.1.7.3 Преобразование спектра из координат пропускания в поглощение

Часто требуется преобразовать спектр пропускания, например в случае измерения ИК-спектров, в

координаты поглощения. Это преобразование проводится нажатием кнопки  Convert to abs (13) (Рис. 1.1).

1.1.7.4 Удаление выбросов на спектре

Часто на спектре присутствуют выбросы, которые иногда называют “cosmic rays”. Они могут исказить результаты поиска. Данные выбросы можно удалить в выделенной области нажатием кнопки Glitch remove (14) (Рис. 1.1).

Для этого необходимо выделить нужную область на графике (19) и нажать кнопку  Glitch remove. Чтобы выделить область необходимо удерживая левую кнопку мыши провести выделение области Рис. 1.4. Далее необходимо нажать кнопку Glitch remove. Чтобы увидеть весь спектр необходимо нажать кнопку с пиктограммой домик 🏠.

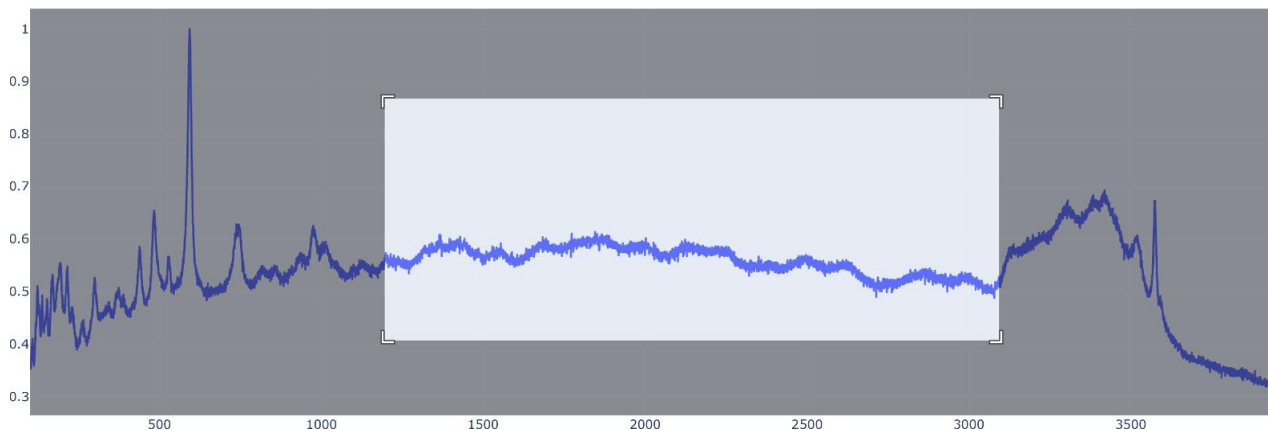


Рис. 1.4 Выделение области на спектре.

После нажатия кнопки Glitch remove происходит наложение на выделенную область фильтра Savitsky-Golay с окном 400 точек. В результате острые выбросы и шумы сглаживаются. На Рис. 1.5 показан результат работы этой функции. Следует заметить, что функцию следует осторожно применять к областям, где имеются пики полезного сигнала, так как возможно искажение спектра. Для удаления отдельных линий в области полезного сигнала следует использовать функцию удаления пиков на вкладке Deconvolution.

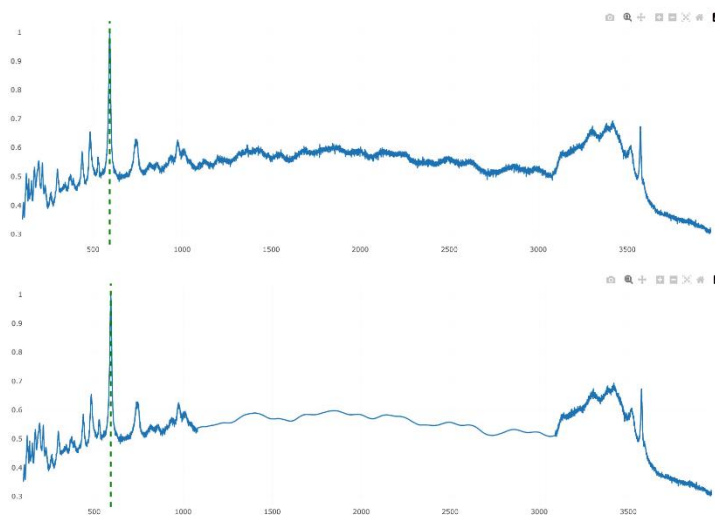


Рис. 1.5 На верхнем рисунке показан исходный спектр, на нижнем результат применения функции Glitch remove к выделенной области.

1.1.7.5 Обрезка спектра

Результаты поиска можно улучшить, используя поиск не по всей области спектра, а только по выделенной. Например в базах данных RRUF часто не приводится область больших волновых чисел, где находятся полосы, связанные с валентными колебаниями О-Н групп. Поэтому при поиске в выдаче могут оказаться спектры, которые содержат область больших волновых чисел. Поэтому зачастую эффективнее бывает выделить область от 150-1300 см⁻¹, где находятся колебания каркаса. Для этого проводим выделение области на спектре, также как показано на Рис. 1.4. Далее нажимаем на кнопку

Cut

Cut (15) Рис. 1.1. В результате спектр будет откадрирован (проведена его обрезка) по оси X в выделенную область.

1.1.7.6 Автоматическое удаление базовой линии

При измерении спектров зачастую нулевая интенсивность не соответствует нулю на оси абсцисс. Присутствует некоторый фон, который описывается базовой линией. Не всегда она имеет вид простой функции. Если уровень фона значителен, то поиск похожих спектров, которые часто содержатся в базе с уже откорректированным фоном, производится неверно.

В этой связи на вкладке Search реализована функция автоматического вычитания базовой линии при нажатии кнопки

Remove baseline

Remove baseline (16) Рис. 1.1.

Вычитание базовой линии проводится по следующему алгоритму. Сначала на спектре находится самый интенсивный пик, далее составляется подгоночная функция, состоящая из суперпозиции функции Псеудо-Фойта, которой описывается наиболее интенсивный пик, и функции базовой линии. В

результате подгонки вычисляются параметры базовой линии и производится ее автоматическое вычитание. Результат вычитания отобразится в поле график (19) Рис. 1.1.

Функция базовой линии вычисляется методом Asymmetric least square smoothing (ALS), реализованный по алгоритму, предложенному в работе [Eilers, 2005] с небольшими модификациями. В исходном алгоритме ALS форма линии задается всего двумя параметрами: p для асимметрии и λ для гладкости. В большинстве случаев асимметрия лежит в диапазоне $0,001 \leq p \leq 0,1$ для сигнала с положительными пиками, тогда как значения гладкости лежат в интервале $10^3 \leq \lambda \leq 10^{10}$, но могут встречаться и исключения. Модификация заключается в том, что при подгонке используется два набора параметров асимметрии и гладкости для двух диапазонов в спектре. В области до 2700 см⁻¹ коэффициент асимметрии лежит в диапазоне $0,001 \leq p_1 \leq 0,1$, а гладкость $10^3 \leq \lambda_1 \leq 10^8$. Для диапазона выше 2700 см⁻¹ используется коэффициент асимметрии в диапазоне $0,001 \leq p_2 \leq 0,01$, а гладкость $10^8 \leq \lambda_2 \leq 10^{10}$. Это требуется для корректного восстановления базовой линии как в области колебаний кристаллической решетки, так и в области валентных колебаний О-Н, которые лежат в области больших волновых чисел. В этой области часто пики имеют большую полуширину, и двухпараметрическая ALS восстанавливает форму некорректно. Пример приведен на Рис. 1.6. При использовании двухпараметрической функции базовой линии видна перекомпенсация с области больших волновых чисел, что может быть полезно при удалении фонового сигнала, связанного с люминесценцией, но в случае, показанном на рисунке, двухпараметрическая линия восстанавливается некорректно (2). Поэтому по умолчанию на вкладке Search используется четырехпараметрическая функция ALS (3).

Значения подогнанных параметров p_1 , p_2 , λ_1 , λ_2 приведены в таблице на вкладке Deconvolution. Там же реализовано автоматическое вычитание базовой линии с двухпараметрической функцией ALS.

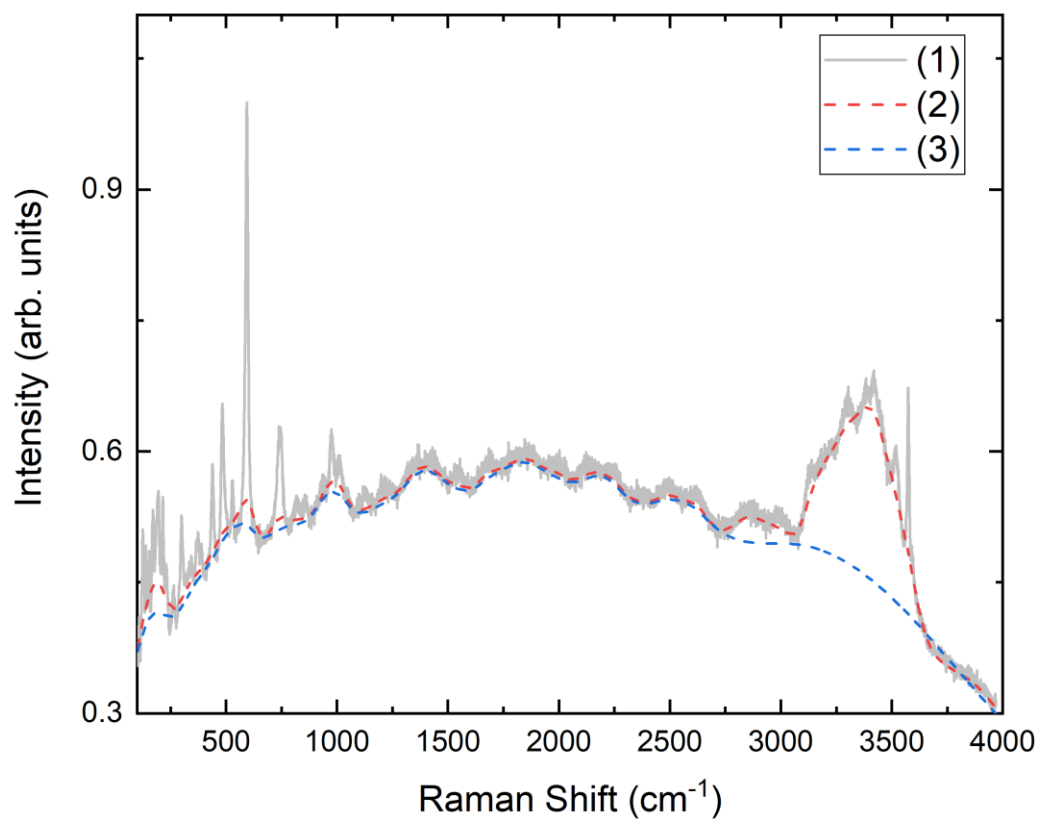


Рис. 1.6 Сравнение результата работы автоматического восстановления базовой линии спектра (1) для двухпараметрической (2) и четырехпараметрической (3) базовой линий.

1.1.7.7 Отмена изменений спектра

Изменения, описанные в пунктах 1.1.7.1-1.1.7.5, можно отменить нажатием кнопки (17)

Reset  Рис. 1.1.

1.2 Вкладка Deconvolution

User: Shendrik (1) ArDI (Advanced spectRa Deconvolution Instrument) (2) Logout

Ulexite_M-1561-191-35_532_R loaded. Select Files (3)

Search Deconvolution Upload (4)

Moving average smoothing ALS smoothing Find peaks and Fit (5)

Look ahead: 1 Delta: 0,5 Find peaks (6)

Tolerance: 1e-15 Max. iterations: 1000 Fit Cancel fit (7)

X-axis correction: 0 Shift X (cm-1) Max Y value 100 Scale Y Convert to abs Cut Remove baseline Reset (8) (9)

125,139,154,171,195,216,230,265,300,344,374,438,483,530,594,746,817,856,937,975,1013,1130,1199,1365,1429,1542,1695,1846,2123,2494,2869,3005,3137,3234,3305,3419,3520,3 (10)

excellent-raman-muf.h5 Ulexite_R050062 6.EA.25: NaCaB₂O₆(OH)₆·5H₂O (10)

Ulexite_M-1561-., R-Square: 0.9869 (12)

Intensity Wavenumber, cm-1 (13)

Residual (13)

Download spectrum Download Subtracted Baseline Data Download Fitted Data Download Fitted Data without baseline Download Peaks Download Fitting Parameters (15)

Fitting parameters

Toggle Columns	Export	Center (C)	Method	A_min scaler	Amplitude (A)	A_max scaler	S_min scaler	Sigma (S)	S_max scaler
x		125	PseudoVoigt	0	1	1000	0.2	4	20
x		139	PseudoVoigt	0	1	1000	0.2	4	20

Add Row (16)

ALS baseline parameters

Second interval	Vary p	ALS p_min	ALS p-parameter (p)	ALS p_max	Vary lam	ALS lam_min	ALS Lambda parameter (lam)	ALS lam_max
	False	0.00250	0.00500	0.0100	False	5.0e+6	1.0e+7	2.0e+7
2700	False	0.000489	0.000978	0.00196	False	5.0e+9	1.0e+10	20000000009.60447

(17)

Рис. 1.7 Описание вкладки Deconvolution. . Вкладка состоит из: (1) имя пользователя, (2) выход из профиля пользователя, (3) область для загрузки файла со спектром, (4) область основных вкладок: Search, Deconvolution, Upload, (5) область вкладок

сглаживания спектра и запуска деконволюции, (6) автоматический поиск пиков, (7) запуск процедуры подгонки пиков и ее отмены, (8) команды редактирования загруженного спектра, (9) поле ввода положения максимумов пиков на спектре, (10) область выбора базы данных и загрузки спектра для сравнения из нее, (11) область легенды и работы с графиком, (12) график, отображающий загруженный спектр и спектры из базы данных, (13) график, на котором построены результаты деконволюции, (14) область легенды графика деконволюции, позволяет убирать лишние пики, (15) ссылки, по которым можно загрузить исходный, обработанный спектры, результаты деконволюции, (16) таблица, содержащая параметры для деконволюции, (17) таблица, содержащая параметры для подгонки базовой линии

Вкладка Deconvolution предназначена для того, чтобы обрабатывать загруженный спектр, искать пики на спектре, раскладывать его на компоненты (деконволюция), удалять выбранные пики из разложения, сохранять параметры разложения и т.п. Операции на этой вкладке доступны только авторизованным пользователям.

Деконволюция является необязательной процедурой, однако зачастую она требуется для представления данных в статью и более глубокой работы со спектрами. Рассмотрим основные применения процедур на этой вкладке.

Во-первых, автоматическое определение пиков с выгрузкой их в виде строки и таблицы. Эта процедура существенно облегчает подготовку статей, содержащих спектры, где часто требуют подписывать положение всех максимумов на графике. Если спектр содержит более 20 полос, то автоматизация, пусть даже и частичная, экономит время.

Во-вторых, собственно процедура деконволюции. Она требуется для оценки не только положения максимумов, но и уширения пиков, площади пиков и других параметров, что может потребоваться, например, для определения плотности включений по спектрам КРС и т.п. В программе доступен широкий набор распределений для подгонки пиков.

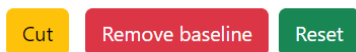
В-третьих, после определения пиков существует возможность убрать лишние пики из спектра. Таким образом возможно отфильтровать интенсивные пики минерала-хозяина и более точно определить минеральный состав включений.

В-четвертых, результат разложения спектра будет загружен в базу данных и в дальнейшем будет использоваться для более точной идентификации минералов.

Ниже приводится описание алгоритма разложения на пики и описание всех функций работы со спектром, приведенные на этой вкладке.

1.2.1 Примерный алгоритм разложения спектра на компоненты (деконволюция)

- 1) Загрузите спектр переместив его в область (3) или кликнув по области (3) и выберите соответствующий файл. Процедура подробно описана в 1.1.3.
- 2) В случае, если в области 2900-4000 см-1 есть широкая полоса, которая не связана с люминесценцией, или которую хотелось бы сохранить при разложении:
 - 2а) На вкладке Search в области (4) нажимаем кнопку Remove baseline (см. раздел 1.1.7.6)



- 2б) Переходим на вкладку Deconvolution и нажимаем кнопку Reset в области 8 на рисунке



. Если провести такие операции, то будет проведен автоматический подбор параметров

базовой линии, чтобы ускорить процесс деконволюции. Автоматически подобранные параметры отобразятся в таблице (17). В дальнейшем ручная подстройка также возможна в таблице (17).

- 3) Произведите автоматический поиск пиков поле нажатия кнопки Find peaks (6). Чтобы увеличить количество обнаруженных пиков следует уменьшать значение в поле Delta (6). Если пиков находится слишком много, то следует увеличить значение Delta.
- 4) Недостающие пики можно указать вручную, добавив положение их максимума в строку добавления пиков (9) и нажав клавишу Enter на клавиатуре. В этом случае на рис. 12 отобразится новая штриховая зеленая линия на месте добавленного максимума.
- 5) Удалить лишние пики можно удаление соответствующего максимума из строки (9) с последующим нажатием клавиши Enter.
- 6) При необходимости можно изменять параметры пиков (их ширину, амплитуду, тип распределения), а также удалить и добавлять пики, используя таблицу (16). Но в этом случае вертикальные зеленые штриховые линии меняться не будут.
- 7) После того, как все пики указаны и их параметры исправлены, следует нажать на кнопку Fit поля (7). После этого начнется деконволюция спектра.
- 8) Результат деконволюции отобразится на рис. (13). Параметры подогнанных пиков будут отображены в таблице (16), параметры базовой линии в таблице (17). Также результат деконволюции и таблицы с параметрами можно скачать по ссылкам (15).
- 9) Если деконволюция не получилась, например, не получилось подогнать ширины пиков, то следует увеличить значение в поле Max iterations или поменять параметры пиков или базовой линии в таблица (16, 17).
- 10) Если в спектре присутствуют «лишние» пики, то они могут быть удалены из спектра на рис. (13) нажатием на соответствующий пик в поле (14). В этом случае спектр на рис. (13) из спектра на рис (13) будет вычтена соответствующая полоса и он будет пересчитан. Результат может быть загружен по ссылкам Download fitted data и Download fitted data without baseline в поле (15). Вернуть пик можно повторным нажатием на соответствующую линию в поле (14).

1.2.2 Обработка спектра

1.2.2.1 Сглаживание спектров

В случае, если спектр зашумленный, то автоматический поиск пиков, обработка и разложение спектра будут проходить достаточно долго. С большой вероятностью процедура деконволюции будет давать неправильные результаты. Чтобы избежать этого необходимо наложить на спектр фильтр низких частот или сгладить его.

В программе предусмотрены три вида сглаживания, два из них находятся на вкладках Moving average smoothing и ALS smoothing  в поле (5) Рис. 1.7.

1.2.2.1.1 Moving average smoothing

Сглаживание сигнала методом скользящего среднего. Для фильтрации задается окно Smooth window, в котором указывается количество точек, по которому будет проводится усреднение. Для сглаживания необходимо нажать на кнопку Smooth. В любой момент можно вернуть первоначальный спектр нажатием кнопки Reset. Ниже приведен пример работы этой функции.

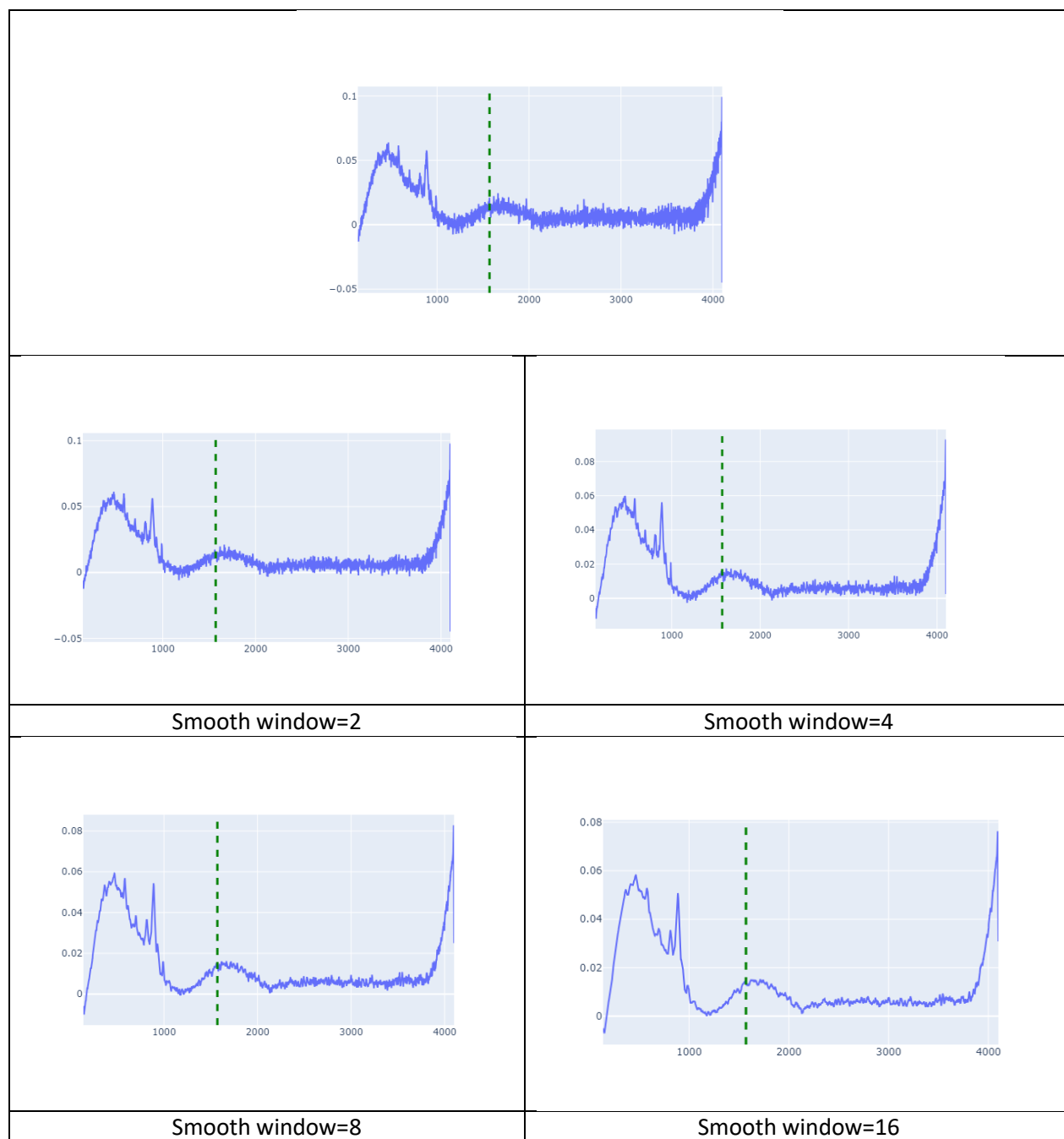


Рис. 1.8 Пример работы Moving average. Вверху исходный спектр

1.2.2.1.2 ALS Smoothing

Вторым методом фильтрации шумов на кривой является ALS Smoothing. Для сглаживания используется два параметра – smooth (гладкость) и p (асимметрия). Их влияние показано на Рис. 1.9 и Рис. 1.10. Чем ниже значение гладкости, тем ближе кривая к исходной. Действие этого фильтра также можно отменить кнопкой Reset. Значения выше 1000 для гладкости используются для восстановления формы базовой линии (см. ниже).

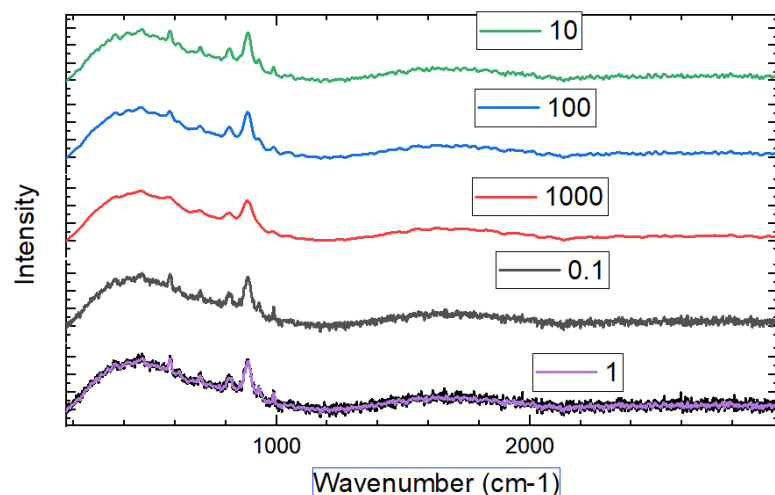


Рис. 1.9 Влияние параметра smooth на вид кривой. p задано равным 0.35, smooth меняется от 0.1 до 1000

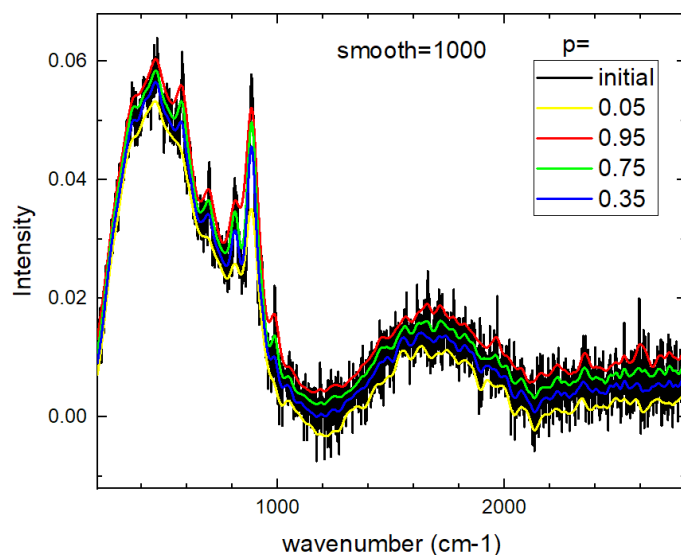


Рис. 1.10 Влияние параметра p на вид кривой.

1.2.2.1.3 Удаление “cosmic rays” или сглаживание Savitzky–Golay

Программа позволяет удалять выбросы на спектрах, называемые cosmic rays. Фактически реализуется сглаживание спектра в выделенной области с помощью фильтра Savitzky–Golay. Для того, чтобы убрать cosmic rays, выделяется необходимая на спектре область. Желательно, чтобы область захватывала только область предполагаемого выброса, далее устанавливаем в поле Smooth вкладки moving average значение порядка 150 и нажимаем кнопку Smooth selected. Далее подобную процедуру можно провести для остальных выбросов. Если выделить широкую область на спектре, тогда фактически будет проводиться сглаживание методом Savitzky–Golay с окном заданным в поле Smooth вкладки Moving average (см. Рис. 1.5).

1.2.2.2 Сдвиг спектра по оси X

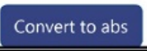
Также как во вкладке поиске на вкладке деконволюции реализована коррекция спектра по оси X путем нажатия кнопки Shift X в поле (9) (см. пункт 1.1.7.1)

1.2.2.3 Нормировка спектра

Загруженный спектр можно отнормировать к значению, указанному в поле Max Y value. Для этого необходимо загрузить спектр в поле (3), далее указать требуемое значение в поле Max Y value и нажать клавишу  Scale Y (9) (Рис. 1.7). После нажатия кнопки Scale Y максимальное значение по оси Y на спектре будет равно указанному в поле Max Y value.

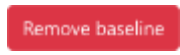
1.2.2.4 Преобразование спектра в координаты поглощения

Если был загружен спектр пропускания, то он может быть преобразован в спектр поглощения

нажатием кнопки Convert to abs  в поле (9) Рис. 1.7. В этом случае происходит преобразование спектра по формуле $A = -\log T$, где T – исходный спектр в координатах пропускания, а A спектр в координатах поглощения или оптической плотности.

1.2.2.5 Автоматическое удаление базовой линии

Автоматическое удаление базовой линии работает по такому же принципу, как и на вкладке Search

(см. 1.1.7.6) и выполняется нажатием кнопки Remove baseline  в поле (9) Рис. 1.7. Отличие от Remove baseline на вкладке поиск состоит в том, что по умолчанию параметры базовой линии подгоняются для одного диапазона. На результат автоматического удаления базовой линии влияют параметры, указанные в таблице (17) Рис. 1.7.

1.2.2.6 Отмена всех изменений

Изменения, описанные выше, можно отменить нажатием кнопки Reset .

1.2.3 Автоматический поиск пиков

В инструменте реализован автоматический поиск пиков на графике. Для поиска пиков необходимо зайти во вкладку Deconvolution, ввести значения Look ahead и Delta в поле (6) и нажать кнопку Find peaks. Положение по оси x найденных пиков будет отображаться в строке ввода (9) ниже, а также в таблице (16) внизу вкладки. Также в таблице (16) в колонке Amplitude будут отображаться относительные интенсивности пиков. Также они будут отмечены на графике (12) в виде вертикальных штриховых линий зеленого цвета (Рис. 1.7).

Можно в ручном режиме добавить пики. Для этого положения их максимума по оси x в виде целого числа необходимо ввести в строку ввода через запятую в поле (6) и нажать клавишу Enter. В этом случае положение введенного пика появится на рисунке слева в виде вертикальной штриховой черты. В случае если воспользоваться таким способом добавления пиков, пропадает информация об их амплитудах в таблице, она станет равной по умолчанию единице. Зачастую это не критично для разложения на пики. Однако если необходимо добавить пик и одновременно сохранить информацию о пиках, то можно воспользоваться второй опцией добавления пиков путем нажатия кнопки Add Row в левом нижнем углу таблицы (16). При нажатии этой кнопки к таблице добавляется пустая строка, куда необходимо вручную ввести положение максимума

нового пика по оси x , а также подкорректировать его ширину на полувысоте и амплитуду при необходимости в таблице (16). В этом случае добавленный пик не будет отображен графике (12), но будет учитываться при разложении спектра на пики.

Для корректной работы автоматического поиска пиков необходимо задать два параметра Look ahead и Delta. Ниже мы рассмотрим, каким образом они выбираются.

Величина Look ahead задает минимальное расстояние в точках, на котором могут располагаться пики. Look ahead = 1, эта величина представляет собой целое число, минимальное значение которого 1, соответствует случаю, когда два максимума могут находиться в соседних точках. На спектрах с хорошим соотношением сигнал/шум этот параметр не влияет на нахождение пиков и его можно задать минимальным, то есть равным 1. На зашумленных спектрах его увеличение может снизить количество «ложных» пиков.

Величина Delta задает минимальную амплитуду пика. Чем ниже значение delta, тем менее интенсивные пики возможно обнаружить. Однако, чем ниже значение delta, тем выше вероятность нахождения «ложных» пиков.

В качестве примера можно привести достаточно зашумленный спектр на Рис. 1.11. Сравним как влияет зашумленность спектра на количество определяемых пиков при одном и том же значении delta.

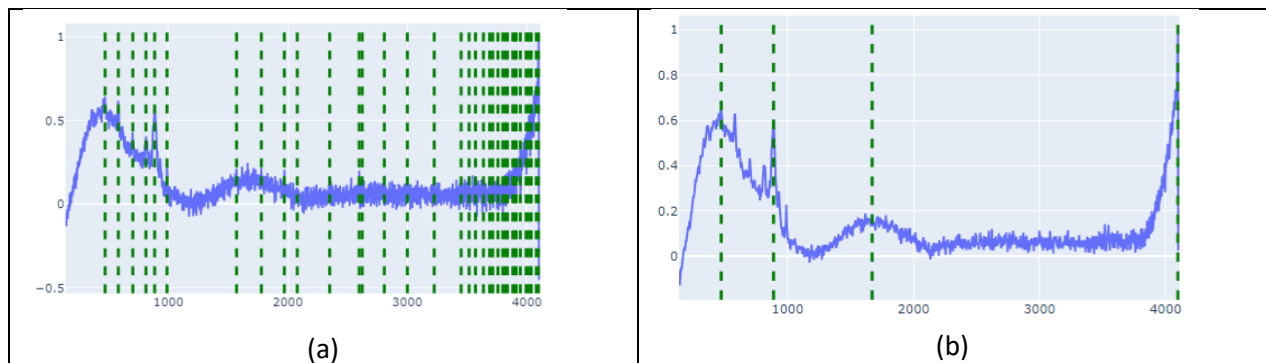
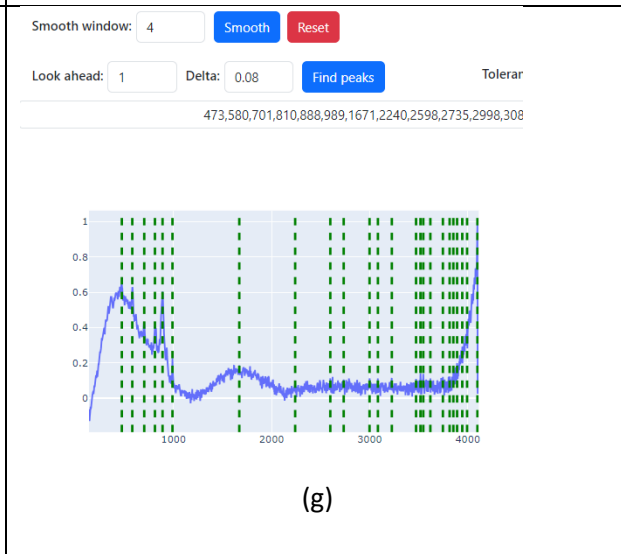
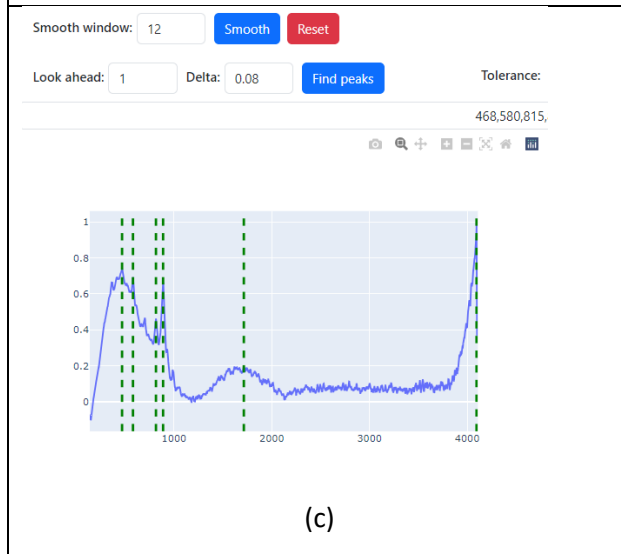
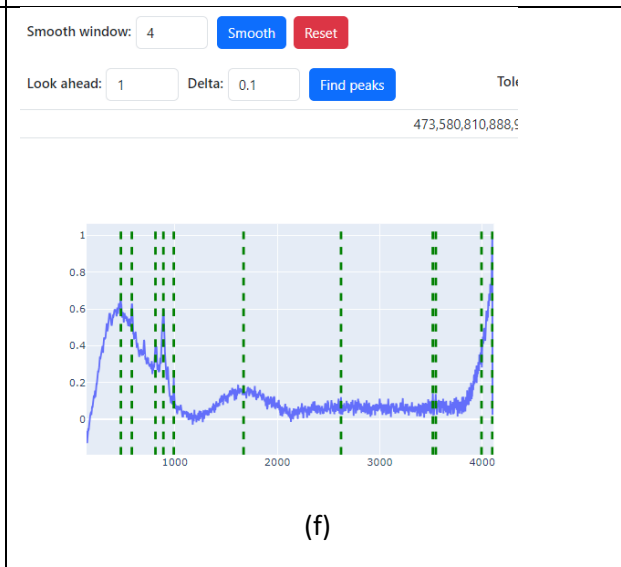
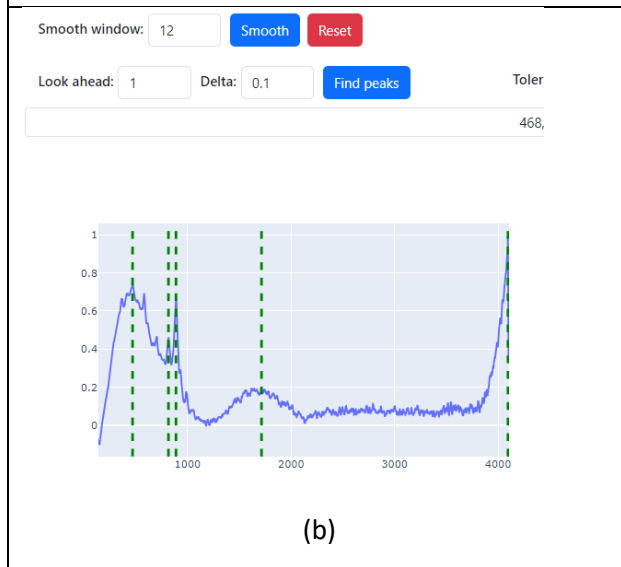
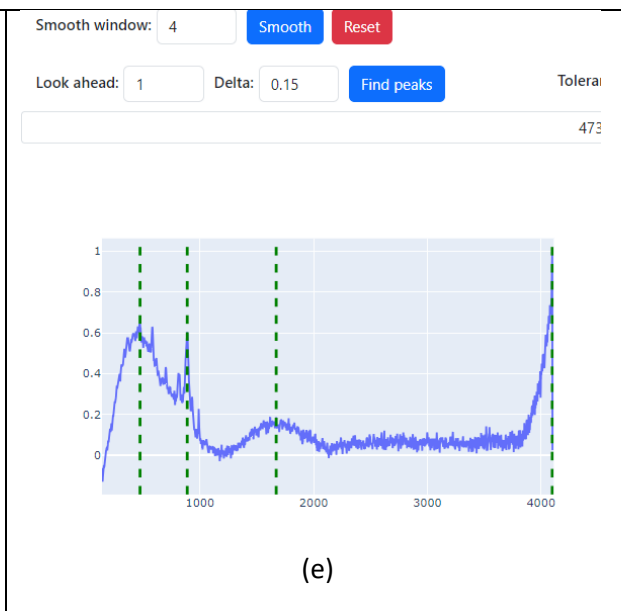
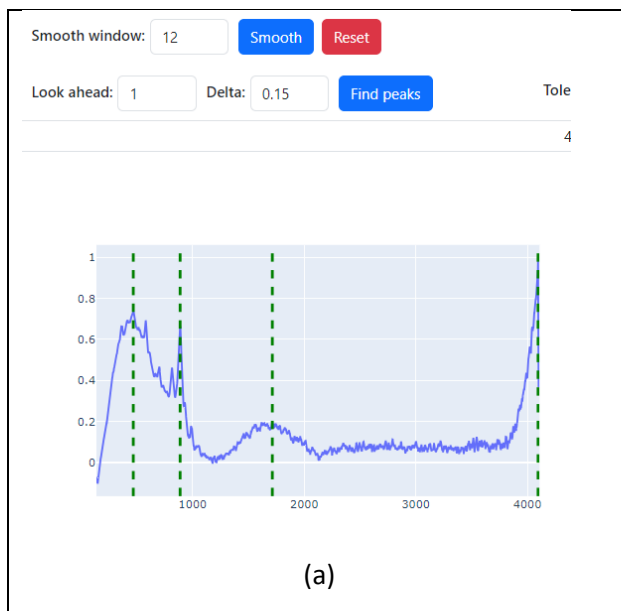


Рис. 1.11 Сравнение автоматического поиска пиков на "зашумленном" (а) и сглаженном спектре (b). Значение параметра Delta одинаково на обоих спектрах и равно 0.15

Фильтрация (сглаживание) спектра позволяет понизить значение дельта без появления «ложных» пиков.



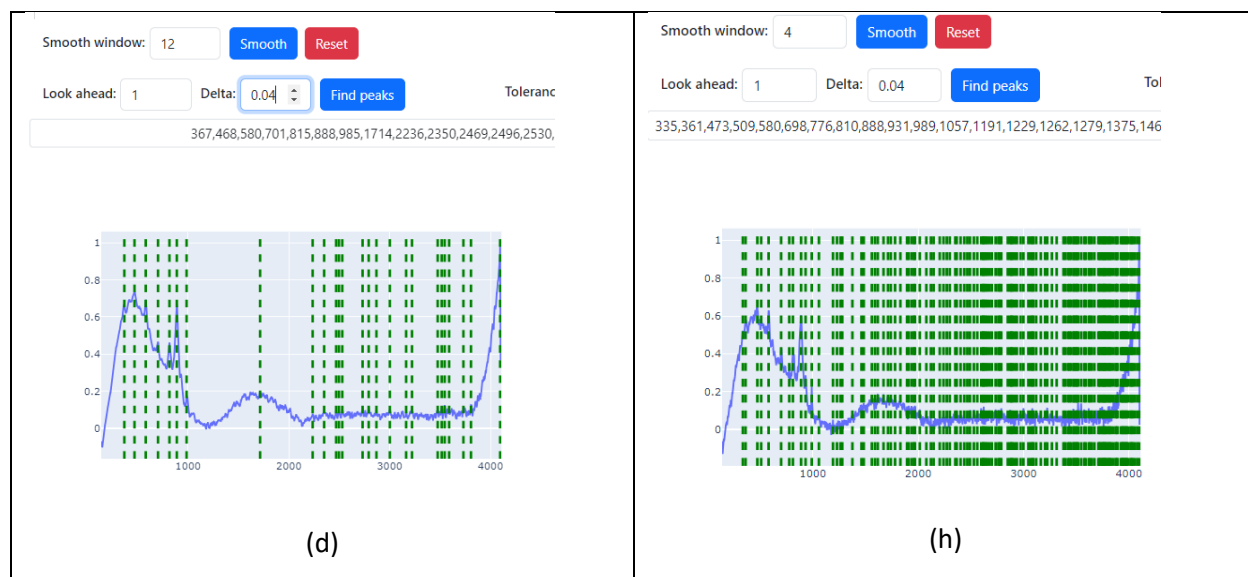


Рис. 1.12 Влияние величины окна при сглаживании спектра и значения Delta на количество найденных автоматически пиков.

На Рис. 1.12 показано, что сглаживание спектров с окном 12 точек приводит к меньшему числу «ложных» пиков и позволяет установить значение delta ниже, чем для более зашумленного и менее сглаженного спектра. Таким образом, наиболее правильным подходом является измерение сразу же спектра с высоким соотношением сигнал/шум, тогда без значительных затруднений удастся найти практически все, даже слабые пики (Рис. 1.13).

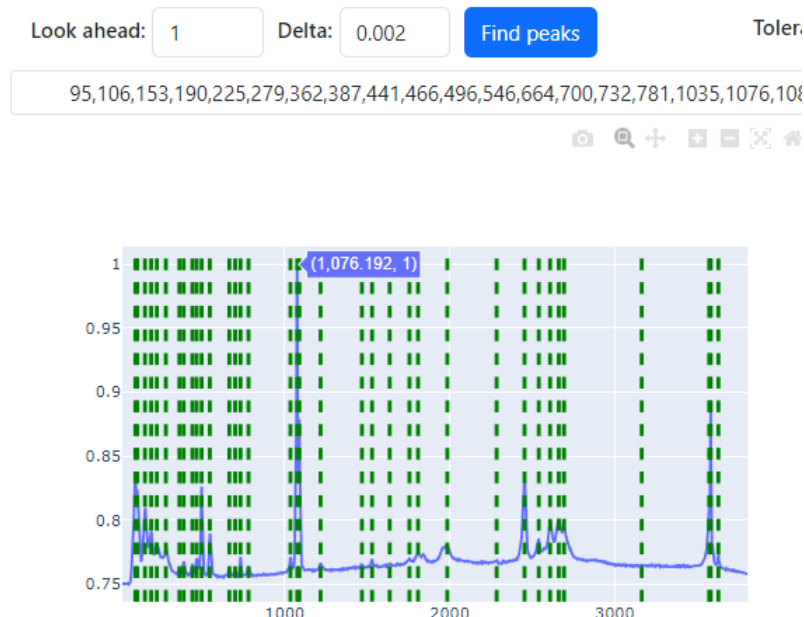


Рис. 1.13 Автопоиск пиков отлично находит практически все пики на незашумленном спектре.

Если же получить хороший спектр не удастся, то перед разложением необходимо его отфильтровать. Однако чрезмерная фильтрация спектра приводит к его безвозвратному искажению и потере важной информации. В этом случае нужно подобрать наименьшую дельту,

чтобы количество «ложных» пиков было не так велико, но в то же время определилось наибольшее число корректных пиков (например, как на Рис. 1.12д). Отметим, что большое количество ложных пиков существенно увеличивает нагрузку на систему и память, поэтому их поиск может завершиться аварийной остановкой программы.

После того как был получен рисунок с достаточным количеством корректных пиков, «ложные пики удаляются путем их стирания из строки ввода пиков и нажатия клавиши Enter.

После удаления «ложных» пиков имеет смысл вручную установить корректные пики, которые не нашел автоматический поиск, путем добавления их через запятую в строке ввода пиков.

Краткие выводы:

- 1) Автоматическое обнаружение и разложение на пики лучше работает на хорошо измеренных спектрах, то есть спектрах с низким уровнем зашумленности
- 2) Если спектр не получается хорошо измерить, то допустимо его отфильтровать, при этом выбирая такие параметры сглаживания, чтобы потерять как можно меньше полезной информации.

1.2.4 Параметры базовой линии

Как было отмечено выше в программе реализован алгоритм восстановления формы базовой линии по модифицированному алгоритму ALS. Параметры базовой линии задаются в таблице (17). Доступно два набора параметров базовой линии для двух участков кривой. Рассмотрим подробнее для чего предназначены столбцы в таблице (17) Рис. 1.14

Second interval	Vary p	ALS p_min	ALS p-parameter (p)	ALS p_max	Vary lam	ALS lam_min	ALS Lambda parameter (lam)	ALS lam_max
	False	0.00250	0.00500	0.0100	False	5.0e+6	1.0e+7	2.0e+7
2700	False	0.000489	0.000978	0.00196	False	5.0e+9	1.0e+10	20000000009.60447

Рис. 1.14 Таблица с параметрами базовой линии

Первая колонка Second interval задает значение по оси x, при значениях больше которого базовая линия будет описываться вторым набором параметров. Если в этом поле не задано значение или оно выходит за пределы измеренного спектра, тогда базовая линия будет описана только одним набором параметров. На Рис. 1.14 показано, что второй интервал начинается с величины 2700 см⁻¹.

Вторая колонка Vary p может принимать два значения True и False. По умолчанию задано значение True. Если в этой колонке указано значение False, тогда при автоматической подгонке параметров спектра (процедура деконволюции) параметр p в соответствующей строке таблицы варьироваться не будет и будет принят за константу.

Колонка ALS p_min означает минимальное значение при варьировании параметра p. Колонка ALS p_max означает максимальное значение при варьировании параметра p.

В колонке ALS p-parameter (p) задается значение параметра p. Далее, если в поле Vary p соответствующей строки указано значение True, тогда варьирование параметра p производится в интервале $p_{\min} \leq p \leq p_{\max}$.

Колонка Vary lam указывает, нужно ли варьировать параметр lam (λ). Если в этом поле указано значение True, тогда параметр lam варьируется в диапазоне $\text{lam_min} \leq \text{lam} \leq \text{lam_max}$.

В большинстве случаев асимметрия лежит в диапазоне $0,001 \leq p \leq 0,1$ для сигнала с положительными пиками, тогда как значения гладкости лежат в интервале $10^3 \leq \lambda \leq 10^{10}$, но могут встречаться и исключения. В любом случае для достижения эффекта значения λ изменяются в логарифмическом масштабе.

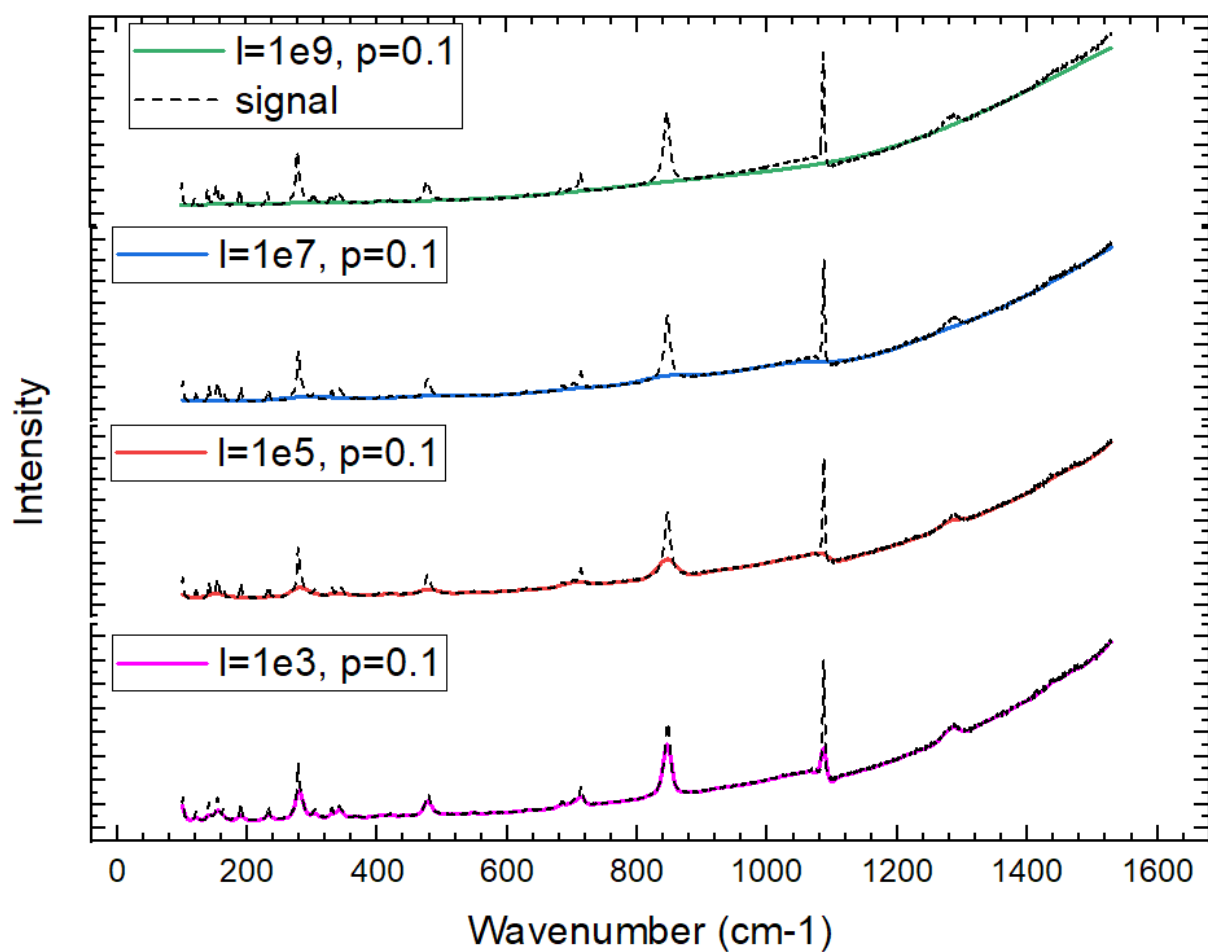


Рис. 1.15 Форма базовой линии (сплошная кривая) и исходный спектр (штриховая кривая) в зависимости от параметра гладкости (на графике в легенде l).

На Рис. 1.15 и Рис. 1.16 показано, как зависит форма базовой линии от параметров гладкости и асимметрии.

Из Рис. 1.15 видно, что малые значения параметра гладкости приводят к тому, что базовая линия частично повторяет форму пиков, если теперь ее вычесть из спектра, то произойдет перекомпенсация, и форма линий в обработанном спектре исказится. Большое же значение параметра гладкости может привести к недостаточно точному повторению базовой линии, что при вычитании приведет к возникновению на спектре широких полос вблизи областей, где базовая линия недостаточно хорошо повторяла форму спектра.

На Рис. 1.16 происходит варьирование параметра асимметрии. Уменьшение этого параметра позволяет точнее воспроизвести область под пиком.

В случае наличия широких полос на спектре (например люминесценция) имеет смысл задать малое значение $1e3 - 1e5$ для гладкости и относительно большое $0.05 - 0.2$ значение асимметрии.

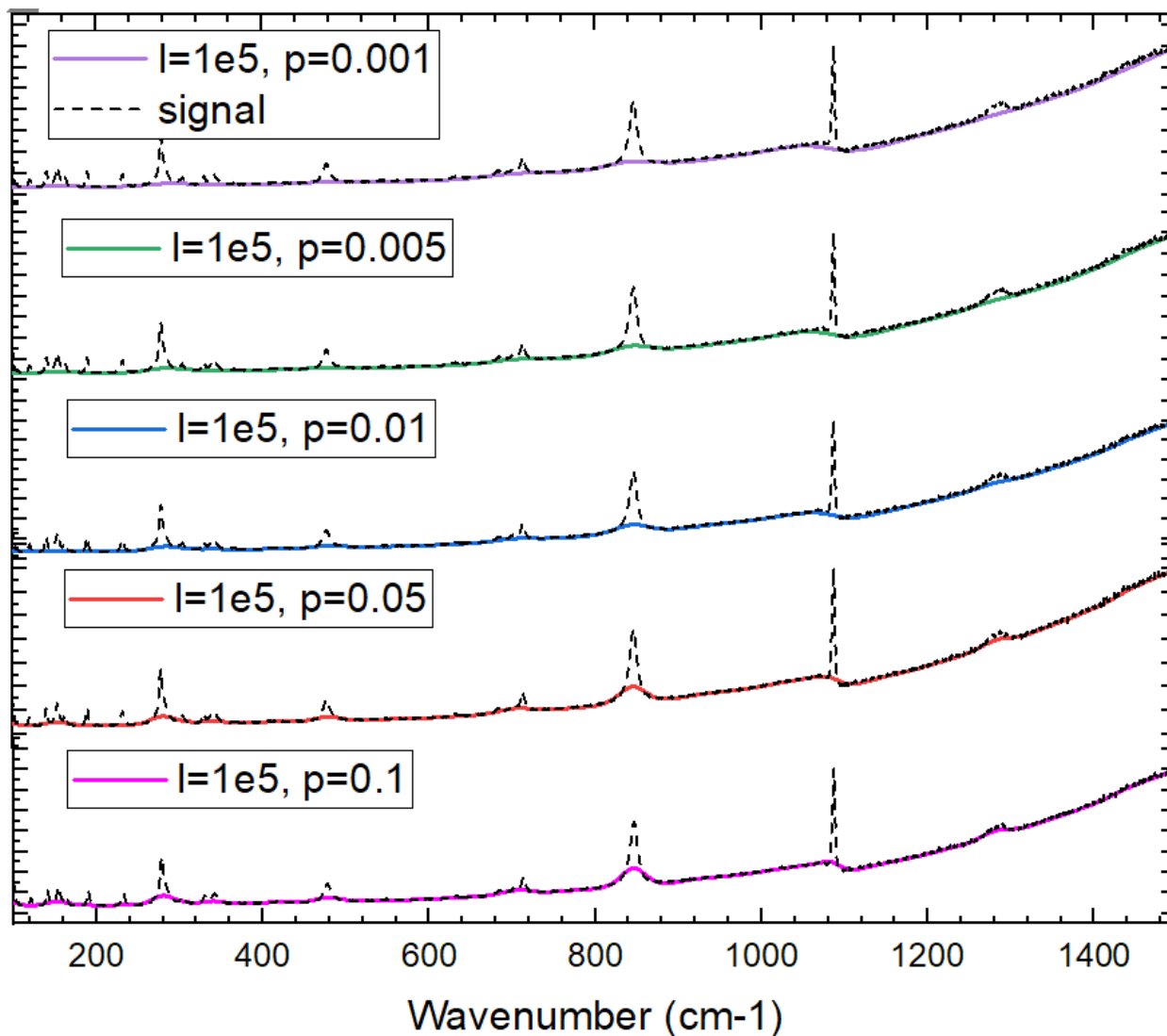
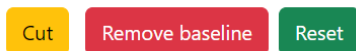


Рис. 1.16 Форма базовой линии (сплошная кривая) и исходный спектр (штриховая кривая) в зависимости от параметра асимметрии (на графике в легенде p).

Второй диапазон (вторая строчка в таблице на Рис. 1.14) имеет смысл указывать, если в области $2700-4000 \text{ cm}^{-1}$ находится комбинация широкого и узких пиков, которые связаны с валентными колебаниями О-Н. Если в этой области нет пиков или их ширина сравнима с остальными пиками на спектре, то второй диапазон можно не указывать. Это упростит процедуру деконволюции.

Отметим также, что если возникли сложности с подбором параметров базовой линии, то можно воспользоваться следующим алгоритмом: указать в поле (6) значение Delta такое, чтобы остался определённым только самый интенсивный пик или вообще не определилось пиков.

Далее на вкладке Search в области (4) нажимаем кнопку Remove baseline (см. раздел 1.1.7.6)



2б) Переходим на вкладку Deconvolution и нажимаем кнопку Reset в области 8 на рисунке



. Если провести такие операции, то будет проведен автоматический подбор параметров базовой линии, чтобы ускорить процесс деконволюции. Автоматически подобранные параметры отобразятся в таблице (17). В дальнейшем ручная подстройка также возможна в таблице (17).

Ручная подстройка параметров базовой линии осуществляется, если все значения в полях Vary p и Vary lam равны False. Параметры, значение которых равно True, будут варьироваться.

1.2.5 Работа с графиками

Графики в инструменте являются интерактивными. Перемещаясь по графику с помощью крестика, можно узнать текущие координаты точек. Зажав левую кнопку мыши, с помощью крестика можно выделять различные области на графиках. Одно нажатие на кривую на легенде позволяет скрывать график, а повторное нажатие вновь показать его. Также есть возможность увеличить области на спектре с помощью инструментов в правом верхнем углу графика (лупа). Чтобы вернуть первоначальный вид графика необходимо нажать на домик (рис. 12), также есть возможность автоматического масштабирования и сохранения рисунка в формате png, создание пометок на графиках. График с пометками также можно сохранить в формате png.

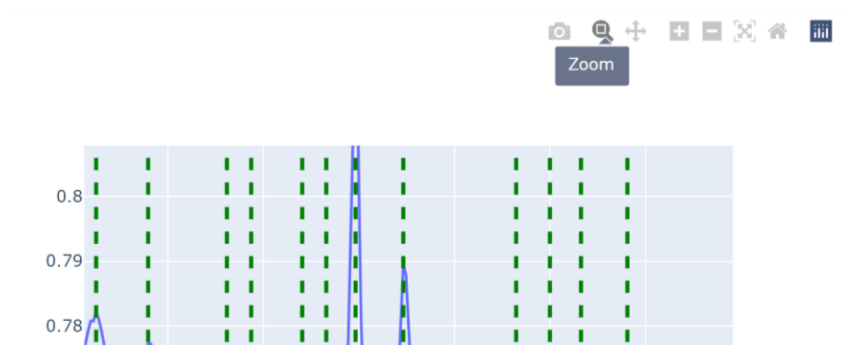


Рис. 1.17 Панель управления графиком в правом верхнем углу. При наведении мышкой появляются подсказки.

1.2.6 Обрезка спектра

Иногда необходимо обрезать спектр, например, в случае подъема на одном из краев, или в случае, когда весь спектр лежит в области до 1000 см⁻¹, и не имеет смысл работать с остальной пустой областью.

Чтобы обрезать спектр, сначала необходимо выделить нужную область на левом графике. Для этого нужно выбрать в правом верхнем углу лупу (zoom) и с зажатой левой клавишей мышки растянуть прямоугольную область до необходимого нам размера (Рис. 1.18). После чего нажимаем кнопку Cut (желтая кнопка в поле (9) Рис. 1.7). В результате произойдет обрезка графика по оси x в заданном интервале. Ось y не обрезается. Дальнейший автоматический поиск пиков, поиск по базам данных, разложение на пики будет происходить только в этой области. Чтобы вернуть первоначальный необрезанный график достаточно нажать кнопку Reset на любом из графиков.

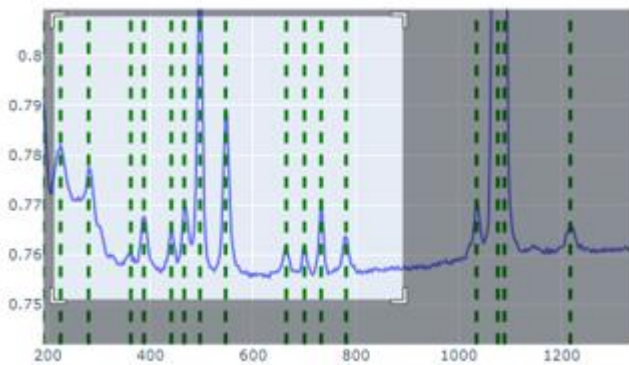


Рис. 1.18 Выделение области на спектре для ее кадрирования

1.2.7 Разложение спектра на пики (деконволюция)

Разложение спектра на пики (деконволюция) спектра проводится по методу наименьших квадратов. Входными параметрами для процедуры подгонки являются следующие для каждого из пиков:

- 1) Предполагаемые координаты центров пиков (C),
- 2) Предполагаемая интегральная амплитуда пика (A),
- 3) Предполагаемая сигма для пика (S)
- 4) Форма пика (выбирается из выпадающего списка).

Также задаются ограничения на область значений при подгонке пиков:

- 1) C_min value, C_max values. Значения координаты центра пика будут находится в соответствующем интервале $[C - C_min; C + C_max]$
- 2) A_min scaler, A_max scaler. Значения амплитуды каждого из пиков будут находится в соответствующем интервале $[A * A_min; A * A_max]$
- 3) S_min scaler, S_max scale. Значения сигмы каждого из пиков будут находится в соответствующем интервале $[S_min * S; S_max * S]$.

Все эти 10 параметров задаются для каждого из пиков в таблице Fitting parameters. По умолчанию значения каждого из параметров одинаковое для всех пиков и приведены в таблице. Однако для каждого из пика их можно изменять прямо в таблице. После изменения нажимается клавиша ввод и тогда значение измененного параметра будет учитываться при подгонке. Также можно изменить значение параметра и нажать одну из стрелок на клавиатуре, значения также сохранятся.

С помощью клавиши Shift можно выделять и копировать данные из таблицы, а также вставлять данные в нее. В поле метод указывается желаемая форма пика. Доступные формы выбираются из выпадающего меню, возможно начать вводит в поле название формы пика, и будут предложены подходящие методы.

В таблице ALS baseline parameters задаются параметры для подгонки базовой линии. Как было отмечено выше это p – асимметрия и lam – гладкость, а также границы изменения этих параметров: $ALS\ p_min < ALS\ p_parameter\ (p) < ALS\ p_max$; $ALS\ lam_min < ALS\ Lambda\ parameter\ (lam) < ALS\ lam_max$.

Процедура деконволюции начинается после нажатия на зеленую кнопку Fit. Также можно задать число итераций до сходимости в поле Max_iter и допустимое отклонение Tolerance (Рис. 1.19).

Tolerance:

1e-15

Max_iter:

1000

Fit

Cut

Remove baseline

Рис. 1.19 Параметры подгонки спектров

Обычно 1000 итераций достаточно для сходимости, но иногда следует увеличить это значение.

Результат деконволюции будет отображен на графике в правой части экрана. Если значение R2 для разложение равно inf, то оно больше, чем 0.9999. Результат разложения появляется на графике (15) Рис. 1.7.

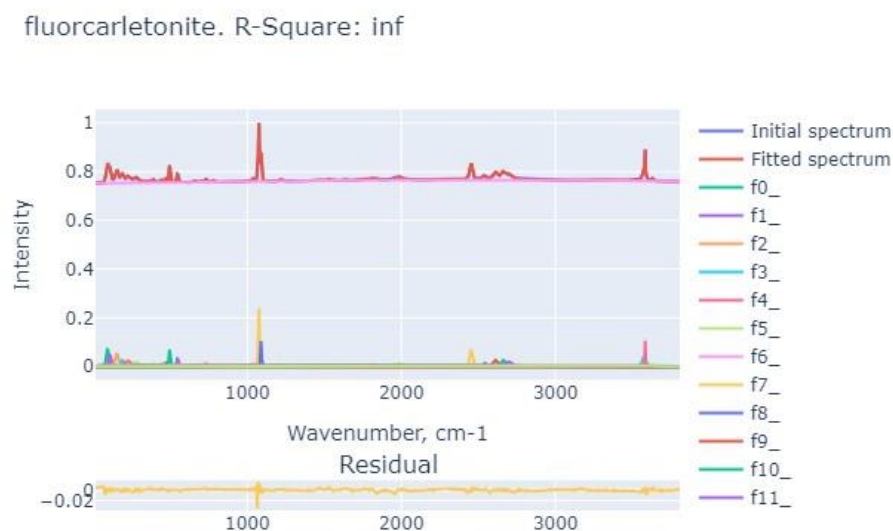


Рис. 1.20 Результат деконволюции в поле (17)

В результате деконволюции становится известны параметры всех пиков и базовой линии. Поэтому возможно скачать в виде csv файлов следующие графики, кликая по ссылкам под графиками в поле (15) Рис. 1.7. Также параметры деконволюции доступны в таблицах (16) и (17). В таблице (16) обновятся значения положения пиков, амплитуды и ширины пика, а в таблице (17) параметры базовой линии, при условии, что они варьировались (см. Параметры базовой линии 1.2.4).

По ссылке Download Subtracted Baseline Data происходит скачивание исходного спектра (который изображен на графике в левой части экрана) с вычитенной базовой линией, по ссылке Download Fitted Data происходит скачивание результата разложения, который является суперпозицией всех пиков и базовой линии, по ссылке Download Fitted Data without baseline скачивается файл содержащий результат разложения исходного спектра, но без базовой линии, по ссылке Download Peaks происходит

скачивание графиков для всех пиков и базовой линии, полученных в результате деконволюции, по ссылке Download Fitting Parameters скачиваются параметры всех пиков.

1.2.7.1 Дополнительные параметры пиков

При использовании для разложения некоторых функций помимо координат центров пиков (C), интегральной амплитуды пика (A), сигмы для пика (S) имеется дополнительный параметр (https://lmfit.github.io/lmfit-py/builtin_models.html). Обычно он задается по умолчанию или вычисляется в процессе подгонки. Однако его значение и интервал в котором будет изменение при подгонке можно задать вручную. Для этого в таблице Fitting parameters необходимо нажать кнопку Toggle Button и выбрать скрытый параметр Additional, а также при необходимости интервал вариации этого параметра при подгонке, который задается как [Additional*Additional_min; Additional*Additional_max]. После чего справа в таблице появятся дополнительные колонки, в которых вводятся необходимые значения. Эти параметры необязательно задавать для всех пиков (Рис. 1.21).

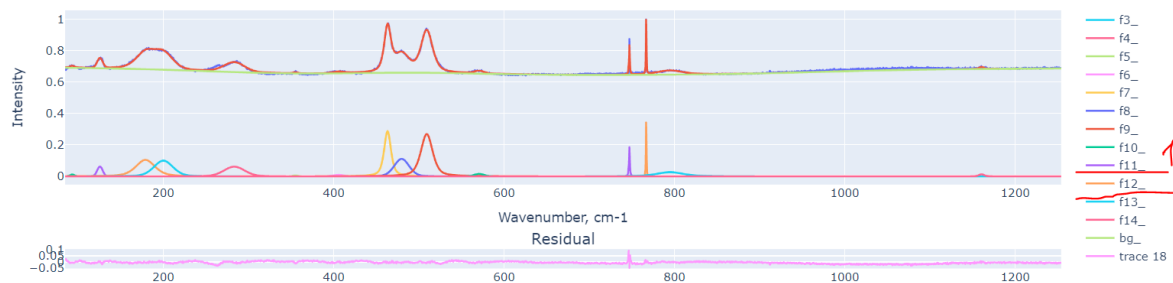
Toggle Columns	Export		Method	A_min scaler	Amplitude (A)	A_max scaler	Sigma (S)	Additional_min	Additional	Addit:
☒ Amplitude (A)			'pseudoVoigt	0	0.4309892924116011	1000	4			
☒ A_max scaler			'pseudoVoigt	0	0.4028605482717521	1000	4			
☐ S_min scaler			'pseudoVoigt	0	0.3800556120778705	1000	4			
☒ Sigma (S)			'pseudoVoigt	0	0.40842272546682556	1000	4			
☐ S_max scaler			'pseudoVoigt	0	0.3779896799364323	1000	4			
☒ Additional_min			'pseudoVoigt	0	0.4154946559396106	1000	4			
☒ Additional			'pseudoVoigt	0	0.6131108267779102	1000	4			
☒ Additional_max			'pseudoVoigt	0	0.5997616114421931	1000	4			
☐ x 1430.080778			PseudoVoigt	0	0.5605880015891935	1000	4			

Рис. 1.21 Выбор и задание дополнительных параметров для деконволюции

1.2.7.2 Удаление лишних пиков

После того, когда процедура разложения экспериментального спектра на пики была проведена, появляется возможность удаления «лишних» пиков из спектра. Удаление обратимо.

Для удаления необходимо на графике с результатом разложения на компоненты нажать на «лишние» пики на легенде к графику (Рис. 1.22). Они исчезают с графика, а результирующая кривая разложения будет пересчитана и перестроена, ее можно загрузить по ссылке Download Fitted Data. Вернуть удалённые пики можно повторным кликом по соответствующему пику в легенде. Результирующая кривая также пересчитывается при каждом выделении кривых на легенде. Результат может быть загружен по ссылкам Download fitted data и Download fitted data without baseline.



[Download spectrum](#) [Download Subtracted Baseline Data](#) [Download Fitted Data](#) [Download Fitted Data without baseline](#) [Download Peaks](#) [Download Fitting Parameters](#)

Рис. 1.22 Удаление «лишних» пиков. Выделение «лишних пиков».

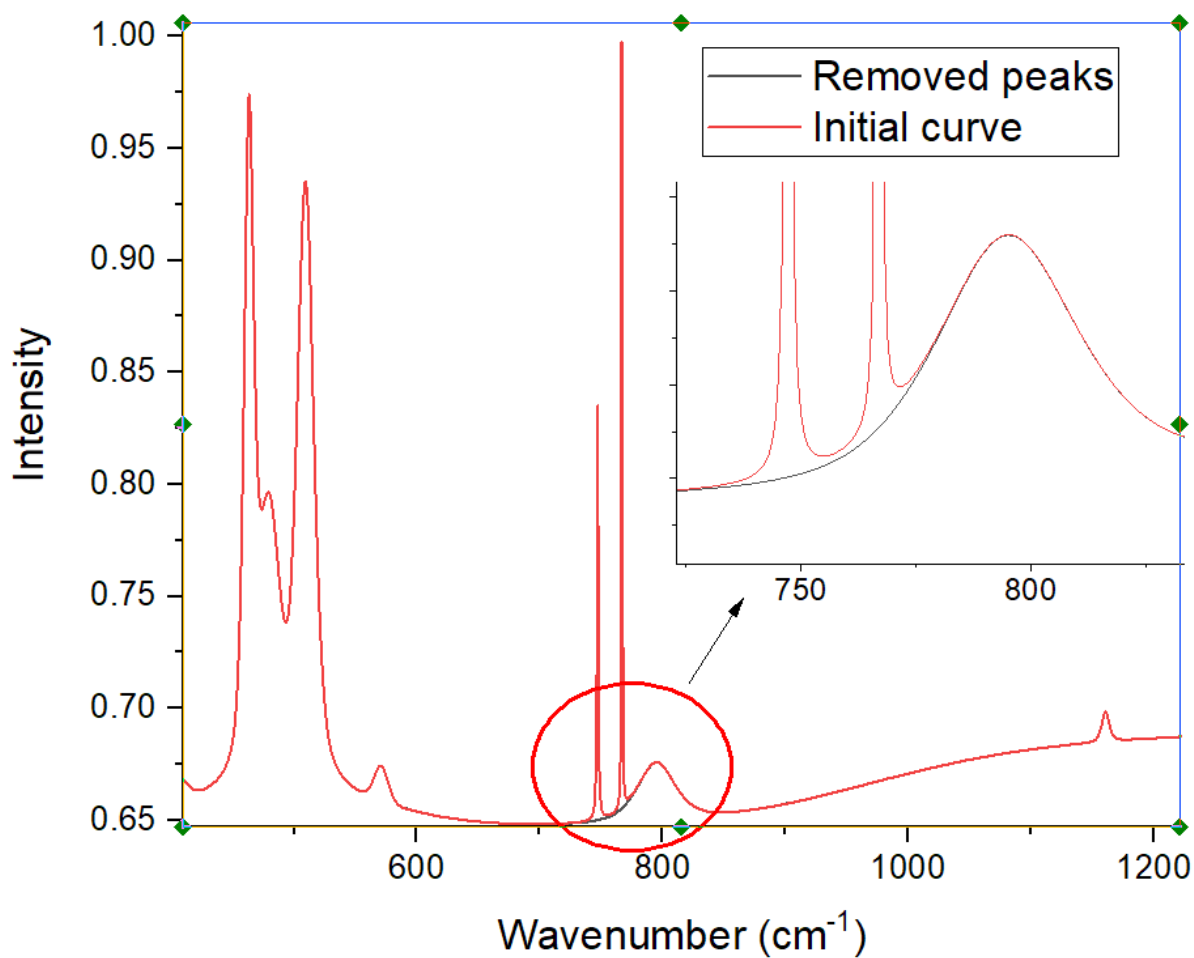


Рис. 1.23 Сравнение кривой с удаленными «лишними» пиками и исходного спектра.

Удаление пиков может быть полезным при удалении “cosmic rays”, а также пиков минерала-хозяина при исследовании включений, а также полос, связанных с люминесценцией (Рис. 1.23).

1.2.8 Добавление спектров из базы данных на график

Для удобства анализа спектров возможно добавить имеющийся в базе данных спектр на загруженный в поле графика (12) Рис. 1.7. Для этого в поле (11) Select DataBase выбирается база данных из списка. Происходит загрузка спектров в этой базе данных. Далее в поле с названием выбирается спектр, который добавляется на график. Можно добавлять несколько спектров (Рис. 1.24).

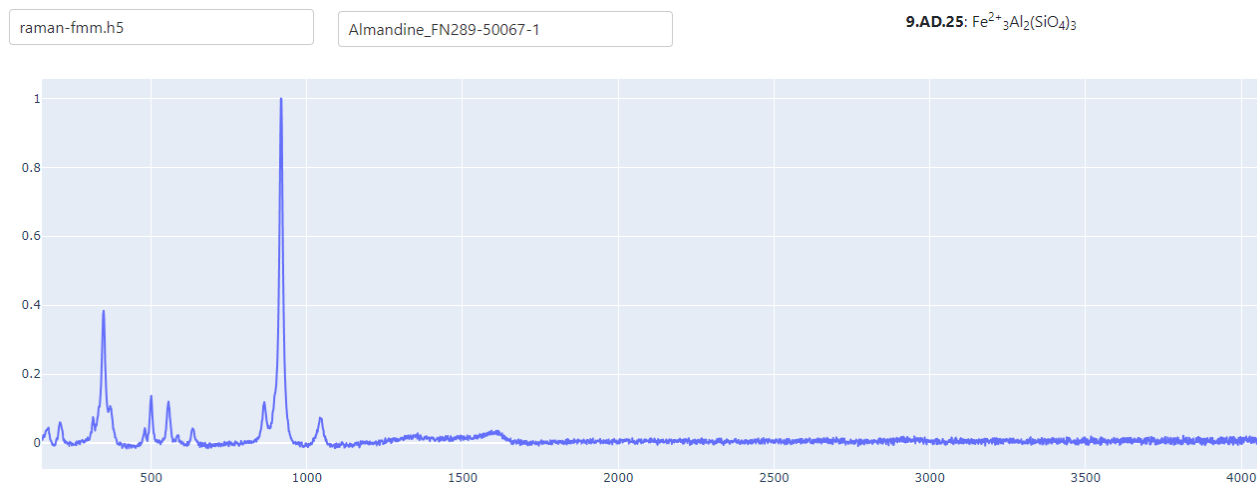


Рис. 1.24 Добавление спектров из базы данных

1.2.9 Вычитание спектра минерала хозяина

Часто спектры КРС содержат полосы от нескольких минералов. Особенно часто такое происходит при съемке включений малого размера. В этом случае спектр КРС содержит как, зачастую интенсивные, пики от минерала хозяина, так и пики от минерала во включении.

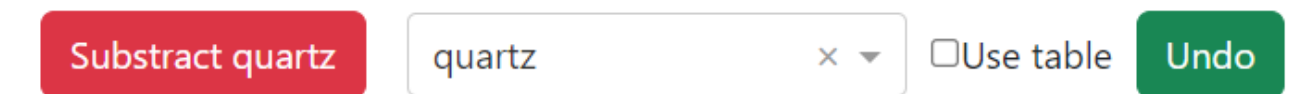


Рис. 1.25 Область вычитания спектра минерала-хозяина

В ArDI реализована функция удаления пиков наиболее распространенных минералов-хозяев. Для того, чтобы удалить полосы от минерала-хозяина необходимо выбрать в поле рядом с кнопкой Subtract <название минерала>, которая расположена в верхнем левом углу над графиком, соответствующий минерал-хозяин и нажать на кнопку Subtract <название минерала> (Рис. 1.25)

В поле графика отобразится помимо исходного спектра вычитенный спектр, с которым можно продолжать дальнейшую обработку и поиск по базе данных, а также пики минерала-хозяина, которые были вычтены из исходного спектра (Рис. 1.26).

Вычитание происходит следующим образом. Положение пиков минерала-хозяина заранее известно. Происходит разложение исходного спектра на полосы, связанные с минералом хозяином с учетом базовой линии. Начальные параметры базовой линии берутся из таблицы 17 Рис. 1.7, также как при обычной деконволюции их можно изменять, количество итераций также как и для обычной

деконволюции задается в поле 7 Рис. 1.7. В таблице 16 Рис. 1.7 отобразятся параметры разложения (Рис. 1.27). Для более точной подстройки их можно изменить в этой же таблице. Чтобы выполнить вычитание с измененными параметрами необходимо нажать кнопку Undo, она вернет график до вычитания, далее поставить галочку в поле Use table, в этом случае будут использоваться измененные данные из таблицы 16, и снова нажать на кнопку Subtract <название минерала>.

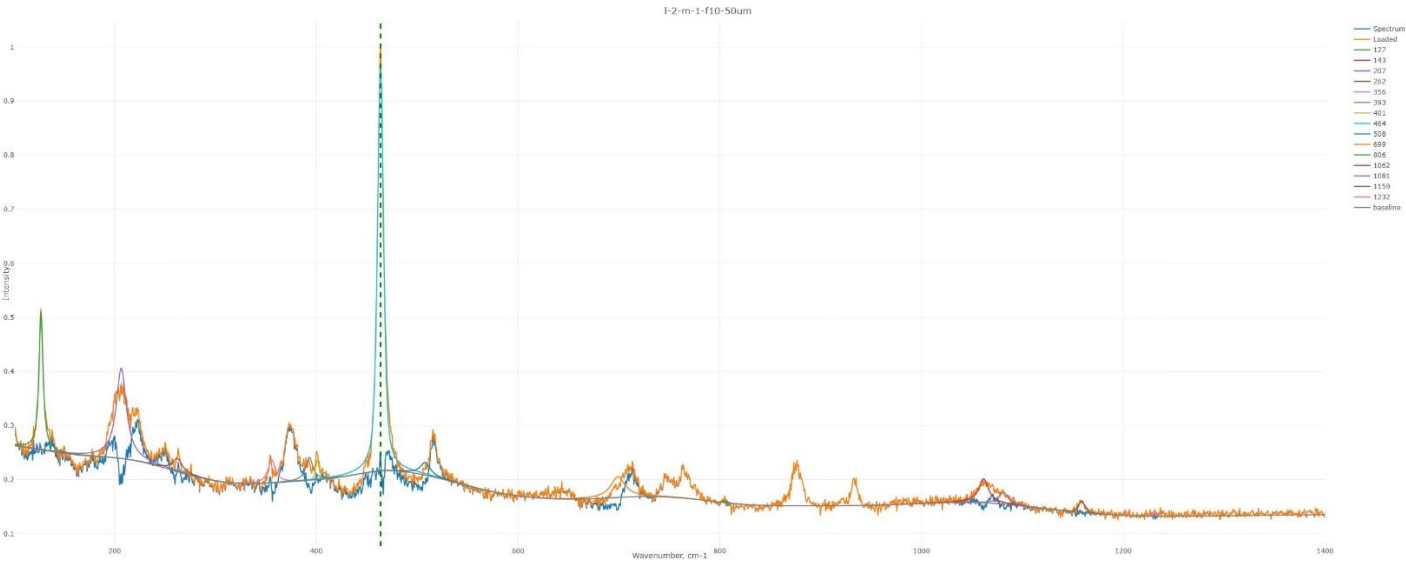


Рис. 1.26 Результат вычитания (синяя кривая) спектра минерала хозяина из исходного (оранжевая кривая) спектра. Пики – положение полос минерала-хозяина (Рис. 1.27)

Toggle Columns		Export							
	Center (C)	Method	A_min scaler	Amplitude (A)	A_max scaler	S_min scaler	Sigma (S)	S_max scaler	
×	127.0	PseudoVoigt	0.0	1.9	1.0e+3	0.20	2.4	1.5	
×	143.1	PseudoVoigt	0.0	0.0077	1.0e+3	0.20	6.0	1.5	
×	206.6	PseudoVoigt	0.0	3.1	1.0e+3	0.20	6.0	1.5	
×	262.2	PseudoVoigt	0.0	0.44	1.0e+3	0.20	5.7	1.5	
×	355.8	PseudoVoigt	0.0	0.60	1.0e+3	0.20	4.5	1.5	
×	393.1	PseudoVoigt	0.0	0.56	1.0e+3	0.20	4.2	1.5	
×	401.1	PseudoVoigt	0.0	0.19	1.0e+3	0.20	2.0	1.5	
×	463.7	PseudoVoigt	0.0	7.4	1.0e+3	0.20	3.6	1.5	
×	508.0	PseudoVoigt	0.0	0.44	1.0e+3	0.20	5.9	1.5	
×	699.0	PseudoVoigt	0.0	1.0	1.0e+3	0.20	8.5	1.5	
×	806.2	PseudoVoigt	0.0	0.022	1.0e+3	0.20	3.0	1.5	
×	1062	PseudoVoigt	0.0	0.82	1.0e+3	0.20	5.9	1.5	
×	1081	PseudoVoigt	0.0	0.37	1.0e+3	0.20	5.5	1.5	
×	1159	PseudoVoigt	0.0	0.22	1.0e+3	0.20	4.0	1.5	
×	1232	PseudoVoigt	0.0	0.041	1.0e+3	0.20	2.1	1.5	

Рис. 1.27 Параметры полос минерала хозяина, использованные при вычитании спектра.

Если нужного минерала нет в выпадающем списке, то можно вручную занести положение его полос для спектра КРС в таблицу 16. При этом точные относительные интенсивности задавать необязательно,

достаточно указать для всех полос одинаковые значения, в процессе деконволюции значения будут подобраны автоматически. Положение полос можно взять, например в книге Chukanov N.V., Vlasina M.F. *Vibrational (Infrared and Raman) Spectra of Minerals and Related Compounds*. 2020. Springer, 1376 p.

1.3 Добавление спектров в базу данных

Если результат разложения спектра получился хорошим, тогда есть возможность сохранить исходный спектр, а также обработанные спектры и результат деконволюции в базу данных. Для этого используется вкладка Download to database На Рис. 1.28 представлено общее описание этой вкладки.

Обязательными для заполнения являются поля Mineral name (3) – название минерала, поле 4 – Sample ID – уникальный номер образца, в базе не может быть двух образцов с одинаковыми номерами. В поле 3 работает автодополнение, достаточно лишь начать вводить название минерала на английском и будут предлагаться варианты для выбора подходящего названия одного из утвержденных IMA минералов.

В случае добавления в базу Рамановского спектра в поле Wavelength (8) нужно выбрать длину волны лазера, с которой проводилась съемка.

В поле (9) вводится метод, которым регистрировался спектр: Raman – спектр комбинационного рассеяния, FTIR – ИК спектроскопия образцов в таблетках KBr, ATR – ИК спектроскопия нарушенного полного внутреннего отражения, Reflectance - ИК спектры отражения, Luminescence - спектры люминесценции, XRD – порошковая дифракция.

При наличии можно добавить файлы с составом и изображением образца.

В поле 1 можно перетащить или по щелчку правой или левой клавиш мышки добавить файл с изображением образца или файл со статьей, где опубликованы данные по образцу.

В поле 2 можно перетащить или по щелчку правой или левой кнопки мышки добавить файл с химическим составом образца, в любом формате, но лучше в .csv или .xls.

В поле (6) есть возможность указать владельца образца или имя того, кто загружает спектр, в поле (7) указывается географическое указание на место, откуда был получен образец, в поле (10) можно оставить комментарий с условиями съемки и какими-то особыми деталями.

После того, как все поля будут заполнены, если перед этим успешно прошла процедура деконволюции, то по нажатию кнопки 11 Submit to server, данные будут отправлены в папку в облаке, и после модерации добавлены в базу данных. В случае, если все прошло без ошибок, в поле 12 появится соответствующая информация, она будет синего цвета, в случае ошибки в этой области появится информация об ошибке, красным цветом.

User: Shendrik

ArDI (Advanced spectRa Deconvolution Instrument)

[Logout](#)

Drag and Drop or Select Files

Search

Deconvolution

Upload

Load image or article Select Files

(1)

Load sample chemistry Select Files

(2)

Mineral name:

Mineral name

(3)

or

Sample name

(4)

Sample ID:

Sample ID

(5)

Sample owner

(6)

Sample locality

(7)

Wavelength:

532

(8)

Method:

Raman

(9)

Comments

(10)

Submit to server

(11)

Load Status:

(12)

Рис. 1.28 Общее описание вкладки загрузки спектров в базу данных: (1) поле для загрузки статьи или фотографии образца, спектр которого загружается в базу, (2) область для загрузки файла, содержащего химический состав образца (csv, xls, xlsx), (3) название минерала или название образца (4), если название не найдено в выпадающем списке, (5) уникальный номер образца в коллекции того, кто загружает спектр, (6) владелец образца, (7) географическая привязка, откуда был взят образец, (8) – длина волны возбуждения, в случае спектров комбинационного рассеяния, (9) Метод измерений, которым был получен спектр (Раман, ИК, РФС и т.п.), (10) – комментарий (условия съемки и т.п.), (11) – кнопка отправки спектра, (12) – область уведомления, после успешной загрузки в ней появится синяя надпись, в случае ошибок – красная с указанием ошибки.

