Фомин В.Н., к.х.н., руководитель ЛИП «ФХМИ».

**Автоматизированная обработка данных лазерной атомно-эмиссионной спектроскопии.**

Спектр ЛАЭС часто содержит достаточно много аналитических линий элементов (сигналов, линий, пиков), интенсивность которых зависит от концентрации элемента, природы матрицы, условий регистрации спектра. В силу ряда причин более подвержен случайным влияниям, чем спектры, полученные от других источников возбуждения. Наряду с сигналом в спектре всегда есть фоновый шум, в виде невысоких пичков. Как следствие, для получения приемлемого качества результатов анализа требуют математической и статистической обработки.

В литературе достаточно много примеров обработки разного рода спектров, однако все они относятся к решению конкретной задачи (данные одного прибора/типа приборов) и не всегда напрямую применимы в других случаях. В специализированном ПО, например Fityk /1/, PeakFit /2/, реализованы некоторые алгоритмы, однако эти программы не подходят для автоматической обработки большого количества спектров. При обработке спектров по одному PeakFit часто пропускает значимые, пики, и не всегда понятно, почему. В общем, в области специализированного ПО сложилась классическая ситуация: постановщик задачи недопонимает возможностей программирования, программист недопонимает сути того, что он обрабатывает и должен получить на выходе, и оба они недостаточно владеют математикой. Коммерческое ПО, поставляемое с приборами, так же не решает проблем такого рода, поскольку заранее разработанно для решения фиксированного круга задач. В общем, сложилась классическая ситуация: постановщик задачи недопонимает возможностей программирования, программист недопонимает сути того, что он обрабатывает и должен получить на выходе, и оба они недостаточно владеют математикой.

Работа обслуживающей университетской лаборатории подразумевает постоянную смену объектов изучения, и, как следствие постоянную необходимость оптимизации условий регистрации спектров и хроматограмм, а для этого нужно извлекать параметры пиков из многих (до нескольких сотен) файлов одновременно.

С учетом вышеизложенного, представляется уместным обратиться к математикам за помощью в разработке собственного ПО для анализа результатов физико-химических измерений. Для сохранения определенной пластичности и совместимости с уже имеющимися нашими наработками предлагается использовать среду R /3/. Для предварительного моделирования работы алгоритма, если есть такая необходимость, предлагается использовать свободно распространяемую GNU Octave /4/. С учетом тенденции к возрастанию количества одновременно анализируемых спектров можно предусмотреть выполнение наиболее вычислительно ёмкой части на GPU, например, с помощью Julia /5/. В итоге получить связку R+Julia. R отлично справляется с любой статобработкой и хемометрикой, а Julia «из коробки» поддерживает вычисления с применением GPU. Обе среды бесплатны, что немаловажно. Наша группа располагает некоторыми навыками работы упомянутыми средами разработки, поэтому нас устроит решение в виде алгоритма, не привязанного к языку.

Что требуется при работе с одним спектром: 0) Сглаживание без потери ширины и интенсивности сигналов. 1) Обнаружить все максимумы, и, при необходимости, границы пиков. 2) Выработать математически аргументированый критерий отнесения максимумов к сигналу или шуму, применить его. 3) Для всех пиков, отнесенных к сигналам, ориентировочно определить положение максимума (λ), интенсивность (I), ширину на полувысоте (FWHM). 4) Аппроксимировать контурами Войгта (по умолчанию, но можно на выбор добавить Лоренца и Гаусса), найти новые значения λ, I, FWHM. 5) Посчитать площади под каждым пиком.

При работе с пакетом спектров следует сначала определить положения всех имеющихся аналитических линий по усредненному или, лучше, по суммарному спетру, применив к нему все пункты работы с одиноким спектром, и лишь после этого сохранять таблицу, содержащую данные по интенсивности/площади для каждого отдельного спектра (т.е. если в данном спектре какого-то пика, присутствующего в других спектрах группы, нет, значение должно быть 0 или уровень шума, но не NA или NaN).

Основная проблема, с которой мы столкнулись, это предварительная автоматическая разметка границ пиков и их аппроксимация. Остальное решаем более успешно.

Литература:

1. M. Wojdyr, J. Appl. Cryst. 43, 1126-1128 (2010) [reprint]
2. <https://grafiti.com/peakfit/>
3. R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
4. <https://octave.org/>
5. Bezanson J., Edelman A., Karpinski S., Shah V.B. Julia: A fresh approach to numerical computing. SIAM Review, 2017, 59(1): 65–98. https://doi.org/10.1137/141000671