Chemometrics Practice

In this tutorial, you'll practice using a few chemometric tools we have discussed in class.

Exercise 1: Peak Picking

Objective: To investigate peak picking and reflect the results of peak picking using calibration curves.

Using the browser-based software GustieChrom (https://homepages.gac.edu/~dstoll/GustieChrom.html), you will import LC chromatograms and perform peak picking.

1. Navigate to the GustieChrom web page.
2. Watch the tutorial in the tab at the top (8 min).
3. Download the data files below.

Caffeine Standard 02.dx_DAD1A.CSV

Caffeine Standard 20.dx_DAD1A.CSV

Caffeine Standard 100.dx_DAD1A.CSV

Caffeine Standard 200.dx_DAD1A.CSV

Caffeine Unknown.dx_DAD1A.CSV

1. In GustieChrom, delete the demo data that automatically loads, and import the data files.
2. The peak between 0.1 and 0.15 min is a caffeine peak. Use the Peak Picker tool to integrate the caffeine peak in every standard and the unknown.
3. Copy the data to an excel file and build a calibration curve of the standards. Note that the file name includes the concentration of caffeine in the standard in ppm. The sample listed as 02 is 0.2 ppm. Perform a linear regression or use a line of best fit on the data.
4. Using the calibration curve, find the concentration of the unknown standard.

Optional Practice:

1. Delete the data from above and redo your analysis. Perform "poor" peak picking by extending the integrated area far out of bounds. Did this change the results?
2. Instead of treating the unknown sample as an unknown, treat it as a 2 ppm standard. How does this change your calibration curve results?
3. Instead of plotting the 0 ppm standard, correct the other peak areas using the 0 ppm standard. In this case, did your results change?

Exercise 2: Normalization & Unsupervised Processing

Objective: To perform exploratory data analysis on GC-MS data using unsupervised statistics.

Using the browser-based software MetaboAnalyst, you will generate a PCA and HCA analysis for a set of GC-MS chromatograms from essential oils. The data is taken from this application note, which you can view for more background:

Application Note_GCxGC of sandalwood oils.pdf

1. Download the following summary file. It is a comma separated values (.csv) file. Open it in excel and familiarize yourself with the data.

Essential oils data set.csv

1. Navigate in a browser to the MetaboAnalyst (https://www.metaboanalyst.ca/) program. Familiarize with the landing page.
2. Click on "Click to start here".
3. Choose "Statistical Analysis [one factor]".
4. In the next step, you need to upload your data. For our analysis, you should use the "A plain text file" option. Select peak intensities, samples in rows (unpaired), upload the data file, and click on Submit.
5. The next page is a data integrity check. Based on your familiarization with the .csv file initially, you should be able to verify why the file meets each of the criteria. Read through this page and click on proceed.
6. The normalization overview page provides some valuable pre-processing strategies that you can try. See if you can identify some of the concepts we covered in lecture. When you select an option, you can click the "Normalize" button at the bottom of the screen and then "View results". Check how the data changes its distribution based on your different selections above. Try a couple different strategies in each category, and view the results of each. Does the result change when you stack different options together?
7. Once you have selected a normalization strategy, click "Proceed".
8. Based on the data you input, you will have options for paths you can explore. Review the options. Some of them will not be clickable because of the structure of the data set you are using.
9. Attempt a One-way ANOVA on your data. Explore the results.
10. Attempt to perform a PCA and an HCA. If you aren't sure which values to use for a setting, use the default values. You can also try adjusting different options and looking at the results before and after.
11. What have you learned from the data based on your exploration? A good exercise would be trying to write 4-5 bullet points about what you have learned about the essential oil data based on the outputs you have generated.

Optional Practice:

1. Attempt to run other tests and observe the output using the essential oil data. Can you discern the meaning of the output?

2. Attempt to run some of the same tests using the test data provided by Metaboanalyst. Can you pull any conclusions from the data?
3. Fill out the User Survey that is presented at the start of using Metaboanalyst. Providing feedback on using this tool helps open source tools like this continue