Rametrix™ LITE Toolbox v1.1 Tutorial
With additional instructions and frequently asked questions

Last Updated: February 6, 2019

Introduction

The Rametrix™ LITE Toolbox is for use with MATLAB®(version R2016a or later) and can be downloaded from GitHub at https://github.com/SengerLab/RametrixLITEToolbox

The Rametrix™ LITE Toolbox is subject to copyright and the MIT License agreement located on the GitHub site and in the "Terms of Use" tab in the Rametrix™ LITE Toolbox, itself.

Version Updates

Substantial changes have been implemented between the initial version (v1.0) and the current version (v1.1) of the Rametrix™ LITE Toolbox. The following are a summary of changes since the last version:

- "Field" values are no longer stored in the filename of spectral files (separated by underscores). They are now stored in a new "Fields File," which is either a Google Spreadsheet (GSHEET) or comma-separated value (.csv) file.
- Spectral files are now named according to a barcode value, which also appears in the Fields File.
- All spectral files (from multiple experiments) can now be stored in a common spectral database folder. Individual folders for separate experiments are no longer needed.
- After baselining, spectra can now be "trimmed" on either end to enhance the baseline fit, if necessary.
- The spectra averaging function is more efficient, and an "un-averaging" function has been included.
- Several bug-fixes and viewing updates have been included to enhance the overall experience of using the Rametrix™ LITE Toolbox.

Tutorial

This tutorial makes use of a dataset of Raman scans of different concentrations of 2-Nitrophenol.

Obtaining, Installing, and Starting the Rametrix™ LITE Toolbox GUI

1. Download the "Rametrix™ LITE Toolbox v1.1.mltbx" file from GitHub (at this link).

2. Double-click the MATLAB Toolbox .mltbx file to start MATLAB and open the install dialog box. Click "Install."

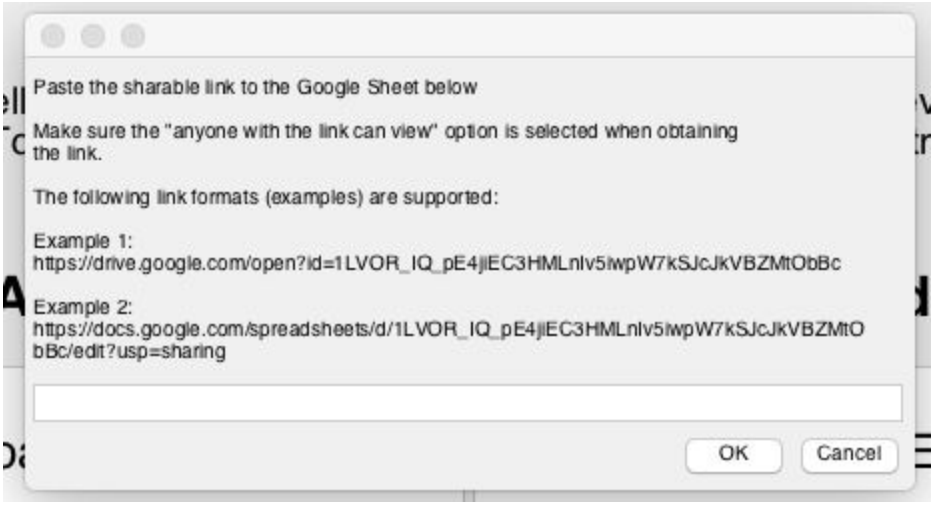
3. Run the Rametrix™ LITE Toolbox by typing "Rametrix" in the MATLAB Command Window. From here, everything runs through the Rametrix™ GUI.

Loading Spectral Files

4. Download the 2-Nitrophenol Calibration Curve Dataset from GitHub (zipped file at this link). Unzip the folder and place in a convenient location.

   Note: Spectral files can be in .SPC, .TXT, or .CSV format. However, files must be named accordingly. In the 2-Nitrophenol Calibration Curve Dataset, the first spectral file included is "NPC1_0001.spc". The portion of the file name "NPC1" is the "barcode," and "_0001" is the scan replicate. The barcode is used to associate other experimental conditions (called "fields" or "factors") with the sample. Often, up to 10 scan replicates are obtained per sample. This produces one spectral file per scan. The Rametrix™ LITE Toolbox has the option to average all scan replicates, so it is important this file naming convention is used. All spectral files must be named with a barcode followed by an underscore and a 4-digit scan replicate value.

5. In the Rametrix™ LITE GUI, Click the “Load Fields (GSHEET)” button in the “Start” tab. This displays the following input box:



6. Copy and paste the following address in this input box and click “OK”: https://docs.google.com/spreadsheets/d/1NRDzVOBL1AeP8cl-pw8ci0R5Xrhkp1lzDWHgWQMIhdg/edit?usp=sharing

   Note: Paste using “control-v” or “command-v” on the keyboard.

7. After seeing the “Success!” notification, Click the “Load Spectra (.SPC)” button. Navigate to the downloaded “2-Nitrophenol calibration curve dataset” folder, and select “Open.” The files will load and another “Success!” dialog box will be displayed. Click “OK.”

Creating and Editing Fields Files

Note: Fields files contain all experimental information about a sample. The example for the 2-Nitrophenol calibration curve dataset can be viewed by copying/pasting the address above (in item #6) into a browser window. This is shown in the figure below.

Note: To use Google Sheets Fields Files in Rametrix™, specific file sharing accesses must be granted. These are discussed in the *Fields Files Sharing* section below.

NPC - 2-Nitrophenol Calibration Curve Log

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

	A	B	C	D	E	F	G	H	I	J	K
1	Index	Barcode	Project	Instrument	Scan Date	Scanned By	Collection Date	Concentration (mM)	Vial	Scan Replicate	
2	1	NPC1	2-Nitrophenol calibration curve	Alpha	20161212	Amanda Fisher	20161212	0.1	1	1	
3	2	NPC1	2-Nitrophenol calibration curve	Alpha	20161212	Amanda Fisher	20161212	0.1	1	2	
4	3	NPC2	2-Nitrophenol calibration curve	Alpha	20161212	Amanda Fisher	20161212	0.1	2	1	
5	4	NPC2	2-Nitrophenol calibration curve	Alpha	20161212	Amanda Fisher	20161212	0.1	2	2	
6	5	NPC3	2-Nitrophenol calibration curve	Alpha	20161212	Amanda Fisher	20161212	0.5	1	1	
7	6	NPC3	2-Nitrophenol calibration curve	Alpha	20161212	Amanda Fisher	20161212	0.5	1	2	
8	7	NPC4	2-Nitrophenol calibration curve	Alpha	20161212	Amanda Fisher	20161212	0.5	2	1	
9	8	NPC4	2-Nitrophenol calibration curve	Alpha	20161212	Amanda Fisher	20161212	0.5	2	2	
10	9	NPC5	2-Nitrophenol calibration curve	Alpha	20161212	Amanda Fisher	20161212	1	1	1	
11	10	NPC5	2-Nitrophenol calibration curve	Alpha	20161212	Amanda Fisher	20161212	1	1	2	
12	11	NPC6	2-Nitrophenol calibration curve	Alpha	20161212	Amanda Fisher	20161212	1	2	1	
13	12	NPC6	2-Nitrophenol calibration curve	Alpha	20161212	Amanda Fisher	20161212	1	2	2	
14	13	NPC7	2-Nitrophenol calibration curve	Alpha	20161212	Amanda Fisher	20161212	5	1	1	
15	14	NPC7	2-Nitrophenol calibration curve	Alpha	20161212	Amanda Fisher	20161212	5	1	2	
16	15	NPC8	2-Nitrophenol calibration curve	Alpha	20161212	Amanda Fisher	20161212	5	2	1	
17	16	NPC8	2-Nitrophenol calibration curve	Alpha	20161212	Amanda Fisher	20161212	5	2	2	
18	17	NPC9	2-Nitrophenol calibration curve	Alpha	20161212	Amanda Fisher	20161212	10	1	1	
19	18	NPC9	2-Nitrophenol calibration curve	Alpha	20161212	Amanda Fisher	20161212	10	1	2	
20	19	NPC10	2-Nitrophenol calibration curve	Alpha	20161212	Amanda Fisher	20161212	10	2	1	
21	20	NPC10	2-Nitrophenol calibration curve	Alpha	20161212	Amanda Fisher	20161212	10	2	2	
22	21	NPC11	2-Nitrophenol calibration curve	Alpha	20161212	Amanda Fisher	20161212	20	1	1	
23	22	NPC11	2-Nitrophenol calibration curve	Alpha	20161212	Amanda Fisher	20161212	20	1	2	

<u>The following columns are contained in the Fields File:</u>

Column A: **Index** - numerical counting so the list can be easily sorted
Column B: **Barcode** - this is the same as the barcode in the spectral file name
Column C: **Project** - the project identifier
Column D: **Instrument** - the name of the Raman instrument used to perform the scan
Column E: **Scan Date** - the date the scan was obtained
Column F: **Scanned By** - the name of the researcher who performed the Raman scan
Column G: **Collection Date** - the date the sample was obtained or prepared (can differ from scan date)
Column H: **Concentration (mM)** - the concentration of 2-nitrophenol in the vial
Column I: **Vial** - the replicate vial number
Column J: **Scan Replicate** - the Raman scan replicate. This is the same as the scan replicate number in the spectral file name.

<u>For new Fields Files, Columns A through G are always required as-is. The next columns are dedicated to experimental parameters of the given experiment (any number can be included). The final column is reserved for the scan replicate. This is also a required column in the Fields File.</u>

This setup allows the user to add, edit, and delete fields in the experiment easily without altering or renaming spectral files.

An added benefit is that rows can easily be added to this Fields File (including barcodes from other experiments). Specific rows can also be copied and added into other Google Sheets. Users may find it useful to sort this file be one or several columns to group rows by the value in a specific column. This often allows for easy copying and pasting into another sheet. For example, for the 2-Nitrophenol Calibration Curve Log (shown above), highlight cell I2. Then, select the "Data" drop-down menu in the toolbar above. Then select "Sort sheet by column I, A→Z." This allows the user to select/copy all entries in either Vial 1 or Vial 2 easily. This can be undone by highlighting cell A2 and repeating the sorting operation.

How To Choose a Barcode

Barcodes must consist of a 3-letter capital-letter prefix, followed by a number. The prefix often describes the project (e.g., NPC for 2-nitrophenol calibration). The numbering always starts at "1" and increments for each sample.

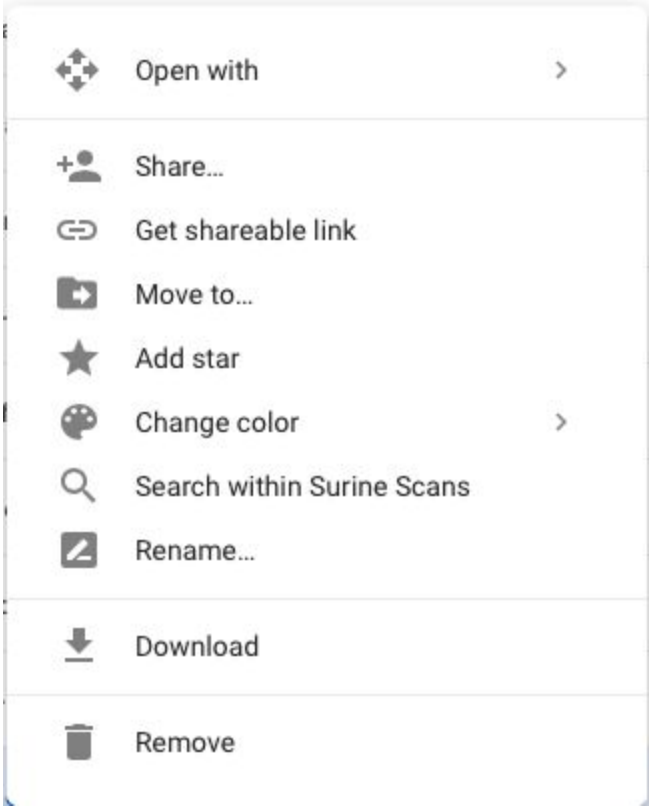
Types of Fields Files

It has been found beneficial to create Fields Files as Google Sheets. However, they can be prepared using any spreadsheet program, including Microsoft Excel. In these cases, the spreadsheet must be exported in comma separated value format (.CSV). This is usually an option under "Save As." Then, the Fields File is read into Rametrix™ using the "Load Fields (.CSV)" button.

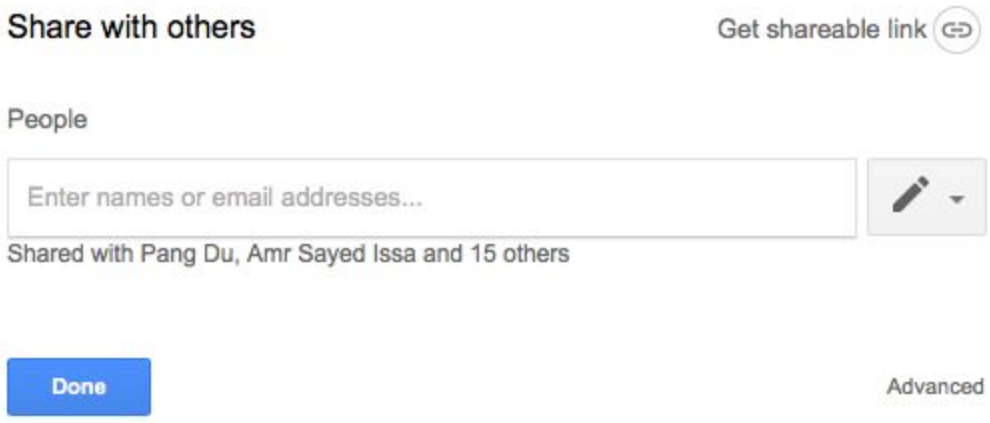
Fields Files Sharing (Required to Use Google Sheets with Rametrix™)

Fields Files must be shared so that "anyone who has the link" can access the file. This is done with the following steps. These instructions apply for an individual Google Drive. Google Team Drives are a bit different but follow the same concept. The goal is to allow access to the file to "anyone who has the link."

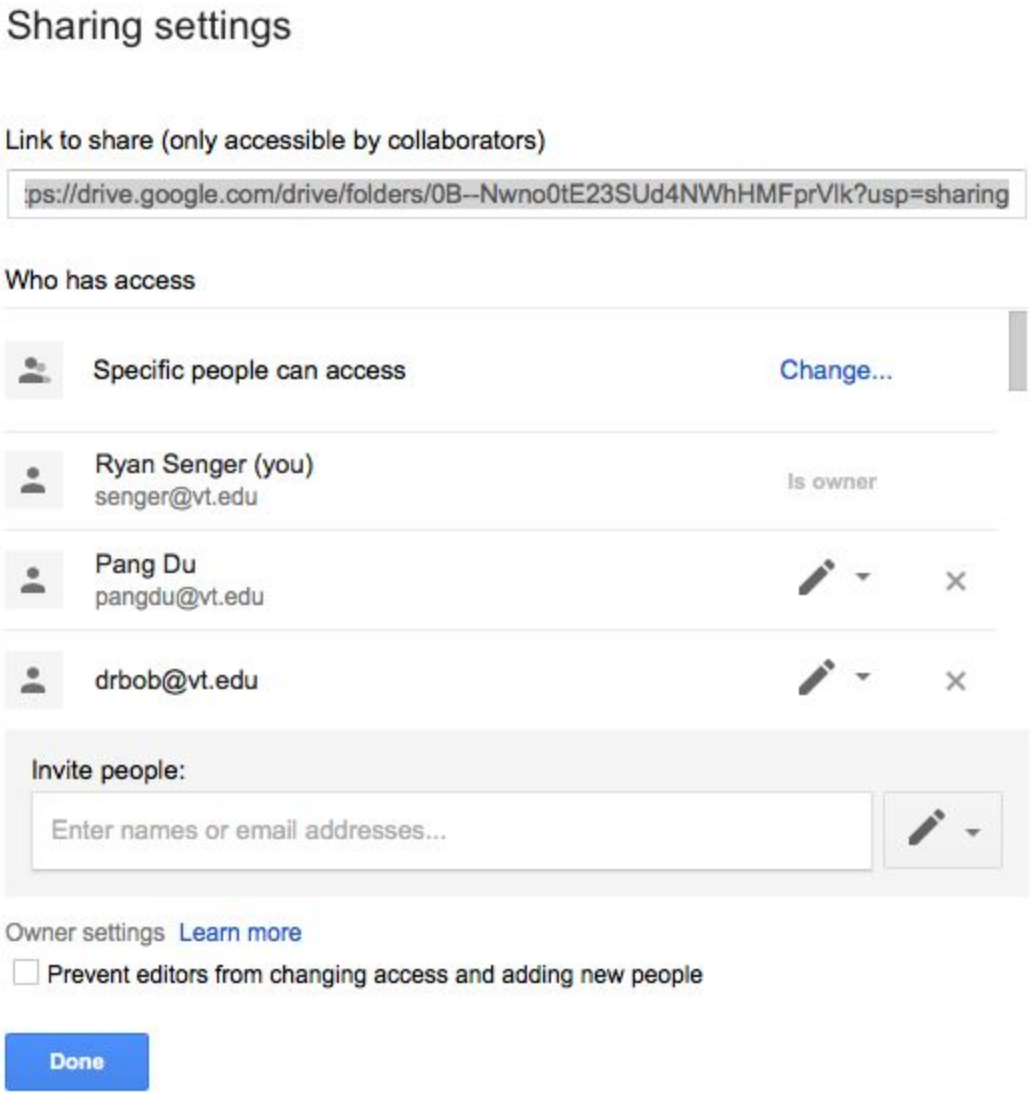
A. Right click on the folder containing the Fields File and select "Share…" as shown in the menu below.



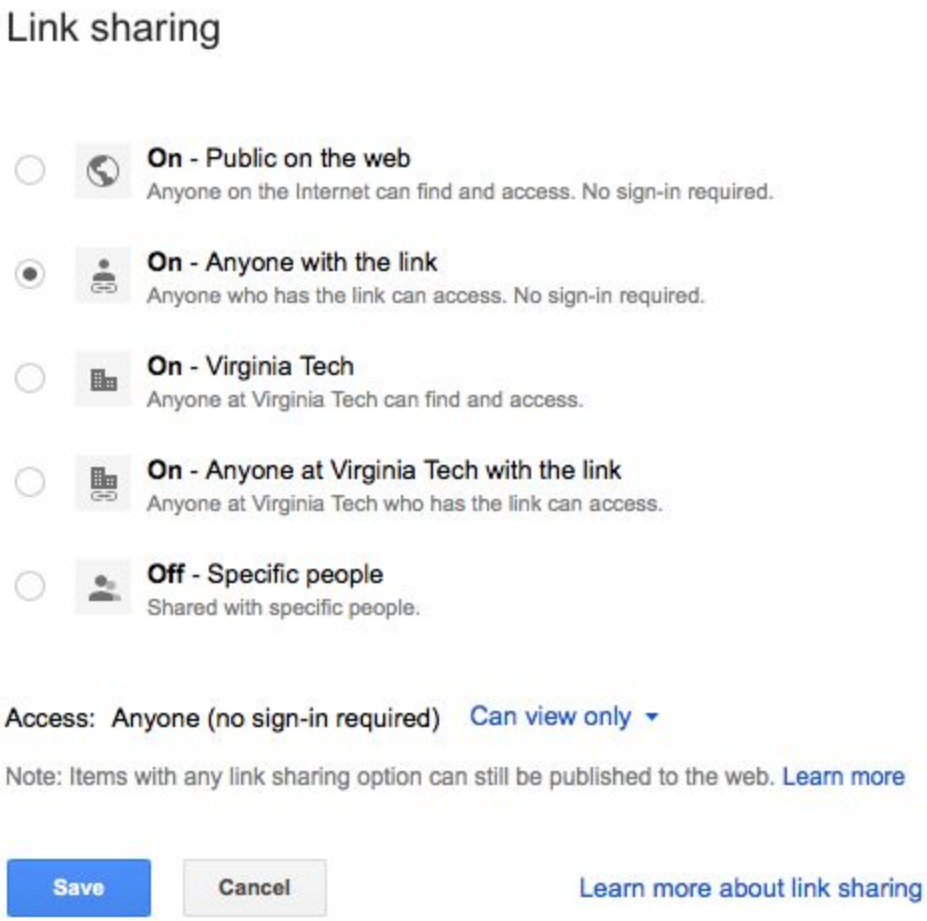
B. Next, click the “Advanced” link in the menu shown below.



C. Click the “Change…” link in the Sharing Settings menu, as shown below.



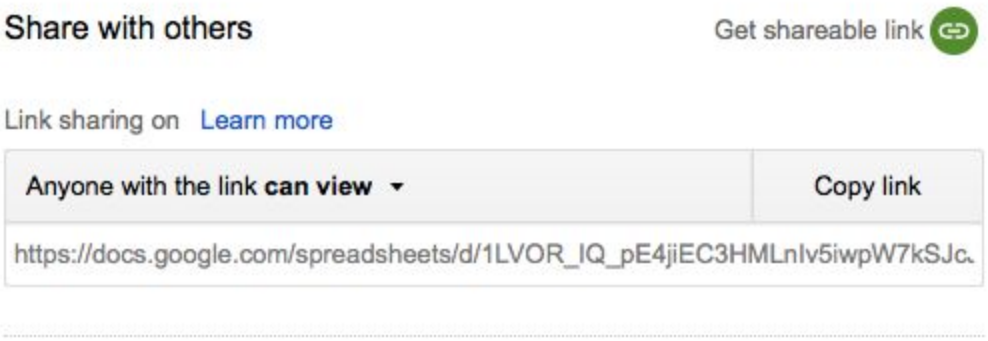
D. Select the “Anyone with the Link” option, and click “Save.”



With these steps, all Fields Files in the shared folder will be accessible to anyone with the link, which is required for Rametrix™ to read the file.

E. Next, open the Fields File to be read into Rametrix™, and select the “SHARE” button on the top right, as shown below.

F. The select the “Copy Link” button. This link will be pasted into Rametrix™ (as shown above in the Tutorial item #6).



Note: Be sure the “Anyone with the link can view” option is selected. If it is not, this can be done from the dropdown menu.

Other Options in the Start Tab of the Rametrix™ GUI

As stated, the Rametrix™ GUI can accept Fields Files in either GSHEET or .CSV format. This is specified in Step 1.

Rametrix™ can accept three different types of spectral files. The first is .SPC formatted files. The other two are .CSV and .TXT formatted files. These files must consist of two columns. The first column contains all wavenumbers, and the second column contains all Raman intensity data.

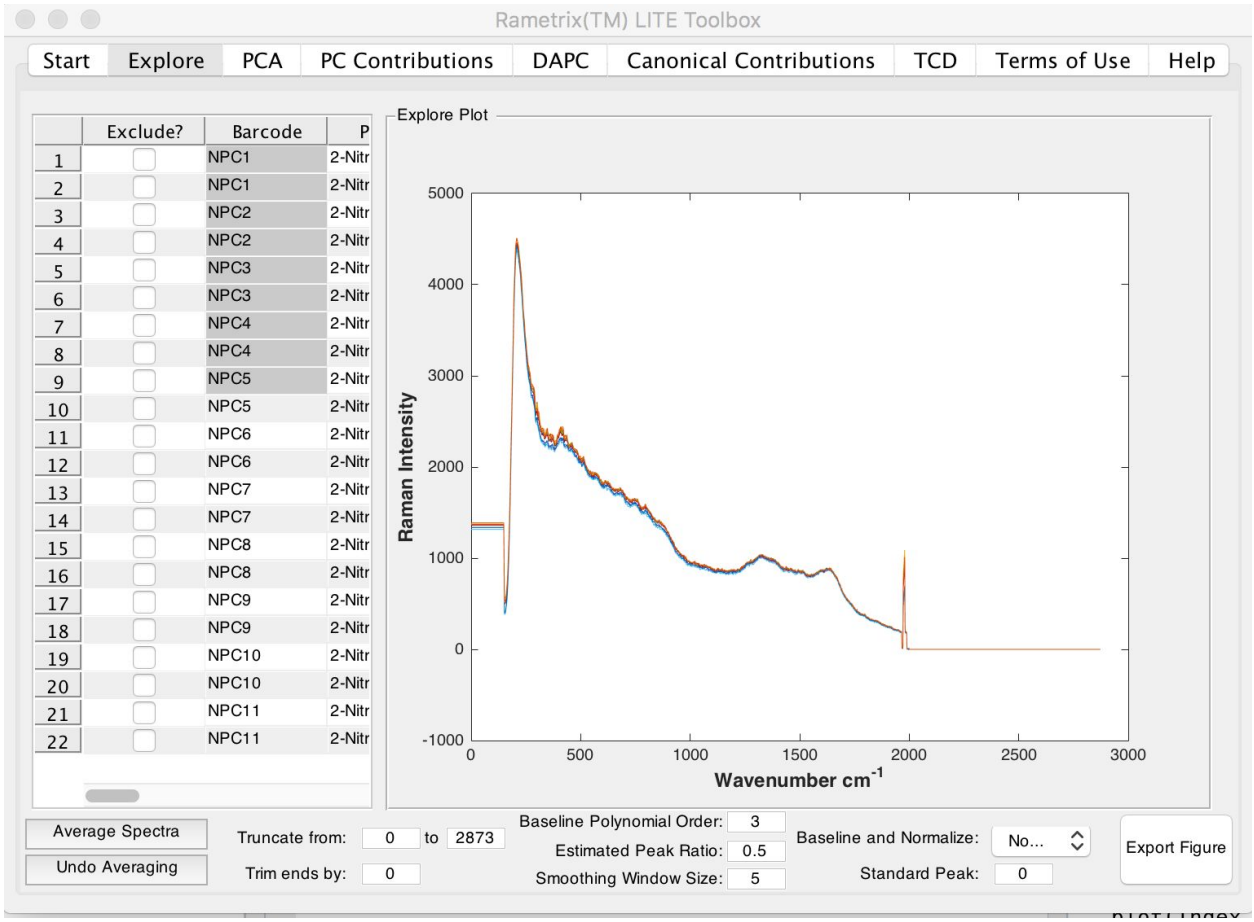
The final row of options in the Rametrix™ GUI Start Tab contains saving and loading options. The entire Rametrix™ analysis can be saved in .RDA format. This saving option preserves all spectral baselining, normalization, and statistical analyses. .RDA format files can be saved and loaded from the Start tab. In addition, data can be exported in .CSV and .XLSX format from the Rametrix™ Start tab.

The Explore Tab of the Rametrix™ GUI

The following steps resume with the 2-Nitrophenol Calibration Curve tutorial.

8. After loading the 2-Nitrophenol Calibration Curve Dataset into the Rametrix™ GUI, select the “Explore” tab at the top.

   Note: All spectra files, with all Fields, are displayed in the window on the left. Selecting a spectra file, displays the spectra data in the main window. Selecting multiple spectra files overlays the spectra. This is shown in the figure below.



9. Click the “Average Spectra” button. This will average all replicate spectra with the same “barcode” value. Averaging can be undone by selecting the “Undo Averaging” button.

10. In the “Truncate from:” box, enter the value “500”, and in the “to” box, enter “1800”. This truncates all spectra immediately. Clicking on another spectra file in the left window will update all calculations.

11. Next, the parameters for the Goldindec baselining algorithm are set. These consist of “Baseline Polynomial Order”, “Estimated Peak Ratio”, and “Smoothing Window Size”. Default values for these parameters are given. These values have been found to be effective in prior research and are changed rarely.

12. Under the “Baseline and Normalize:” drop-down menu, choose the “Baselined and Vector Normalized” option. This will apply the Goldindec algorithm and vector normalization. The Goldindec algorithm fitting can require significant time, depending on the size and complexity of the dataset. Selecting a different spectra file in the left window after calculations complete will update the spectra viewing pane.

Note: The “Baseline and Normalize:” drop-down menu contains several options. The “Show Baseline” option will show how the Goldindec algorithm fits each spectra. Occasionally, the Goldindec algorithm can lead to poor baseline fitting on either end of the spectra. This will impact later statistical calculations. This poor fitting can be fixed by altering the spectra truncation range. It is recommended to increase the truncation (increase the “from” value and decrease the “to” value).

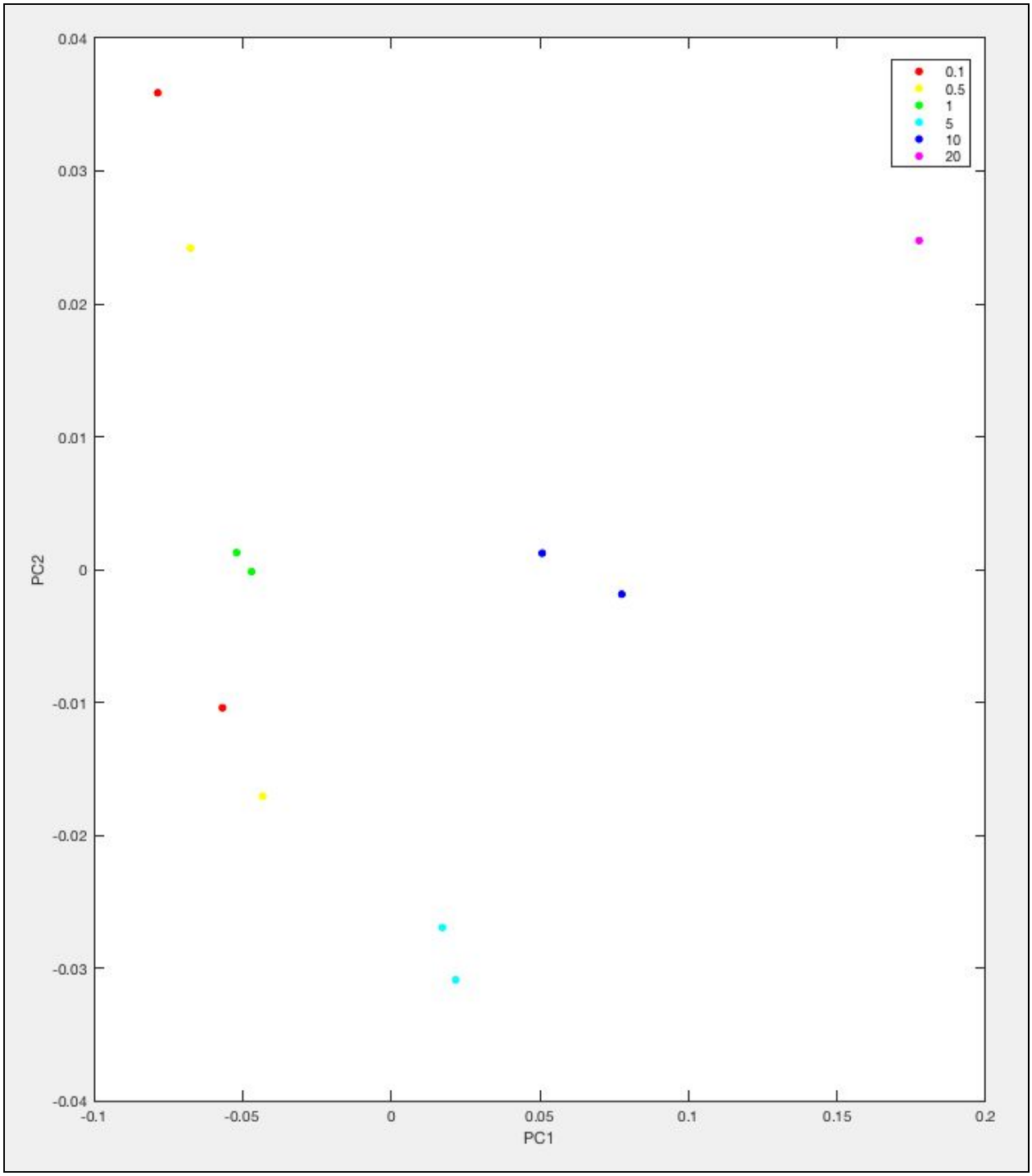
Note: High-quality MATLAB figure(s) of the viewing pane can be exported by clicking the “Export Figure” button. This figure can be edited in MATLAB following exporting.

Note: Should a spectra be found to contain error(s), it can be excluded from further statistical calculations by selecting the “Exclude?” button next to the spectra file in the left window. This can also be useful should a particular field value need to be excluded from further calculations.

The PCA Tab of the Rametrix™ GUI

Principal component analysis (PCA) calculations are performed in the PCA tab of the Rametrix™ GUI. The following steps continue with the 2-Nitrophenol Calibration Curve dataset.

13. In the “Factor to Analyze” “Choose:” drop-down menu, select “Concentration (mM).” Then click the “Run PCA” button. Once results appear, right-click in the viewing pane, and select “Go to X-Y view.” Results should appear as shown below.



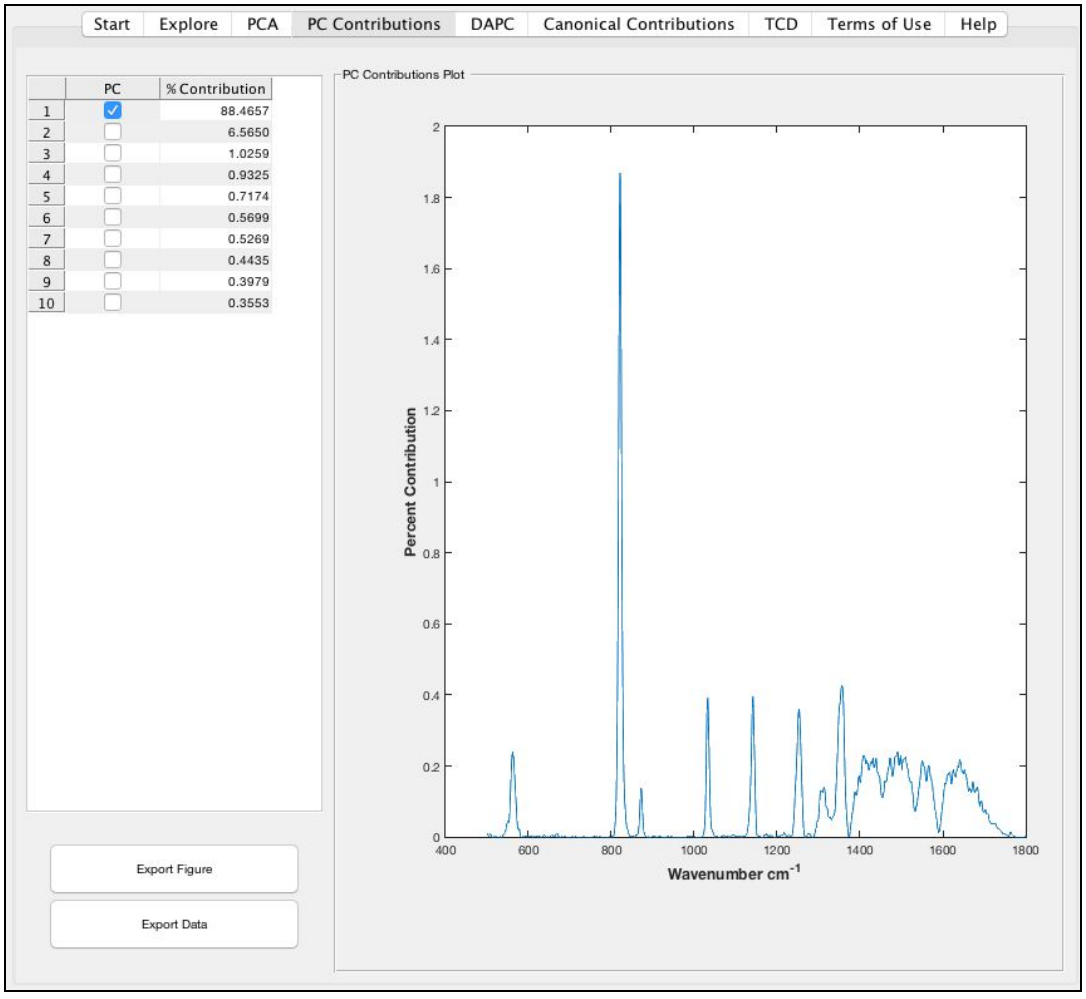
The data points line up along PC1 (left to right) according to concentration. Distinguishing concentrations below 1 mM is significantly harder than for those above 1 mM.

14. The figure can be exported into MATLAB for manual editing by selecting the “Export Figure” button. Additionally, principal component data can be exported in spreadsheet format by selecting the “Export Data” button.

The PC Contributions Tab of the Rametrix™ GUI

Following PCA calculations, the PC Contributions tab is used to obtain information about each principal component. This includes (i) the percentage of dataset variance explained by each principal component and (ii) the Raman shifts most responsible for creating the variance. The following steps continue with the 2-Nitrophenol Calibration Curve dataset.

15. Select the checkbox next to PC1. This principal component captures 88.47% of the dataset variance. This produces the plot below, which shows the contribution of each Raman shift to creating the variance in the dataset.



In the 2-Nitrophenol Calibration Curve dataset, the samples differ only by the amount of 2-Nitrophenol present. Therefore, the major bands appearing in this figure should be those that change the most in the Raman spectra. Suppose PCA were to find differences between Raman spectra of two unknown samples. The bands appearing in this figure would represent the chemicals present in one sample but not the other. Thus, studying PC contributions is a way of discovering the spectral and chemical differences among samples and spectra in the dataset. This has proven extremely useful.

16. This figure can also be exported to MATLAB for further editing using the "Export Figure" button. The data can also be exported in spreadsheet format by selecting the "Export Data" button.

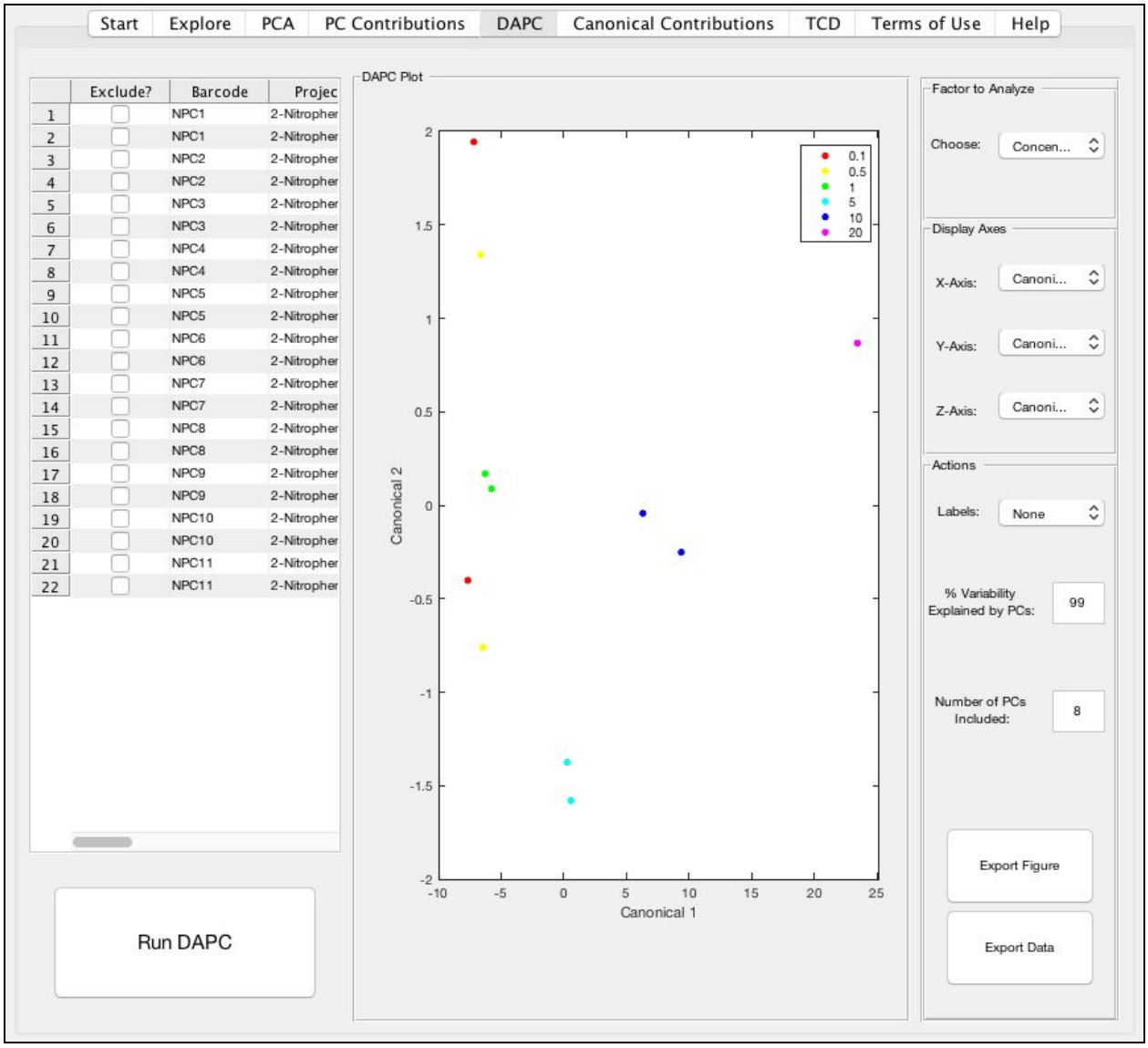
The DAPC Tab of the Rametrix™ GUI

Following PCA calculations, discriminant analysis of principal components (DAPC) models can be built in the DAPC tab. DAPC models work to separate data point clusters further than PCA based on a specified "factor." For the 2-Nitrophenol Calibration Curve dataset, these can be separated further using concentration as the factor.

17. In the "Factor to Analyze" "Choose:" drop-down menu, select "Concentration (mM)."

18. DAPC models use principal components as inputs. The user can specify (i) directly how many principal components are to be used in model construction or (ii) what percent of the dataset variance should be included. If one is selected, the other is calculated automatically. In the box for "% Variability Explained by PCs:" enter "99." Then, select the "Run DAPC" button. You will notice the "Number of PCs Included:" is updated automatically to "8."

19. Once results appear, right-click in the viewing pane, and select "Go to X-Y view." Results are very similar to PCA results for this case and are shown below.



Results are oriented along Canonical 1 by increasing concentration, similar to PCA results. In our experience, PCA and DAPC results are rarely similar. Increasing the number of PCs included or percent variability explained improves the separation of data cluster.

Note: When increasing the number of PCs, care should be taken to avoid dataset overfitting. This is addressed in additional calculations using the Rametrix™ PRO Toolbox for MATLAB®.

The Canonical Contributions Tab of the Rametrix™ GUI

This tab is very similar to the PC Contributions tab in that it shows what Raman shifts are responsible for the DAPC model separating clusters in the dataset. In our experience,

this seems to be less related to chemical composition, when compared to the PC Contributions calculations. It is unclear whether this tab will exist in future version of the Rametrix™ LITE Toolbox.

The TCD Tab of the Rametrix™ GUI

The Total Canonical Distance (TCD) tab uses the DAPC results and calculates the distance across all canonicals between each Raman spectrum and one designated as the "reference." In DAPC, typically only two or three canonical values can be viewed at a time. However, many more can exist. We have found correlation between the distance between a sample and a reference and their degree of chemical similarity.

Note: We have recently found that this distance calculation can be performed using PCA results and even spectral intensity values. These may be more accurate than TCD calculations and will be added to future version of the Rametrix™ LITE Toolbox.

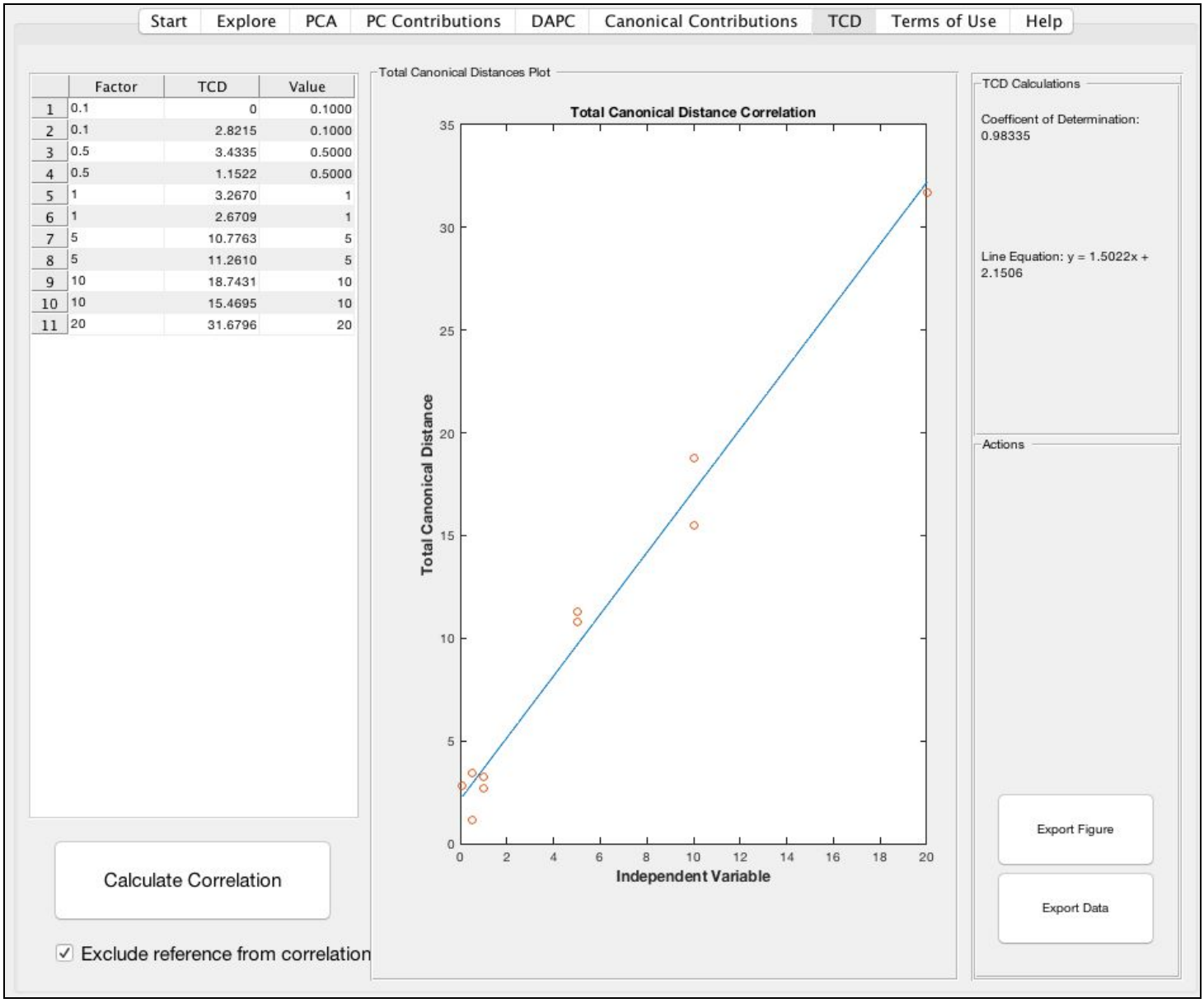
20. In the "Factor" column in the table on the left, put the cursor in the row to be used as the "reference." This usually refers to a spectrum serving as a control or "0" concentration. In the case of the 2-Nitrophenol Calibration Curve dataset, the lowest concentration is 0.1 mM, so one of those spectra was chosen as the reference.

21. When the cursor is placed, the TCD values are calculated automatically. Moving the cursor to a different row will update the TCD values automatically with that row as the reference.

22. Next, numerical values must be filled in manually for the "Value" column. Future versions of the Rametrix™ LITE Toolbox will have an automated solution for this. For the 2-Nitrophenol Calibration Curve dataset, the value in the "Value" column should be the same as the concentration value.

23. Check the box, "Exclude reference from correlation."

24. Click the "Calculate Correlation" button to generate the relationship between TCD (y-axis) and values in the "Value" column (x-axis). In addition to the plot (shown below), the coefficient of determination (R^2) for a linear fit is calculated along with the equation of the fit line. These are given, as shown below.



Results show a good fit and linear relationship among the data with a coefficient of determination (R^2) exceeding 0.98.

25. This figure can also be exported to MATLAB for further editing using the “Export Figure” button. The data can also be exported in spreadsheet format by selecting the “Export Data” button.

Frequently Asked Questions

We are still building this section. Please email your questions to Ryan Senger at senger@vt.edu.