

# A manual for Porchlight

Jakub Konkol  
George Tsilomelekis

August 15, 2022



A light to guide you home

# Contents

<b>1</b>	<b>Running Porchlight</b>	<b>3</b>
<b>2</b>	<b>The main program window</b>	<b>3</b>
2.1	The menu bar . . . . .	3
2.2	Left Panel . . . . .	4
2.3	The Right Panel . . . . .	5
<b>3</b>	<b>Importing Data</b>	<b>5</b>
<b>4</b>	<b>Exporting Data</b>	<b>5</b>
<b>5</b>	<b>File Format Support</b>	<b>6</b>
<b>6</b>	<b>Transformations</b>	<b>6</b>
6.1	Trim . . . . .	6
6.1.1	Trim . . . . .	6
6.1.2	Inverse Trim . . . . .	6
6.2	Baseline . . . . .	6
6.2.1	Polyfit . . . . .	7
6.2.2	AsLS . . . . .	7
6.3	Smooth . . . . .	7
6.3.1	Rolling . . . . .	7
6.3.2	SGSmooth . . . . .	7
6.4	Normalization . . . . .	8
6.4.1	SNV . . . . .	8
6.4.2	Area . . . . .	8
6.4.3	Vector . . . . .	8
6.4.4	MinMax . . . . .	8
6.4.5	MSC . . . . .	8
6.4.6	Peak Normalization . . . . .	9
6.5	Centering . . . . .	9
6.5.1	LastPoint . . . . .	9
6.5.2	Mean . . . . .	9
6.6	Derivative . . . . .	9
6.6.1	Savitzky-Golay Derivative . . . . .	9
<b>7</b>	<b>Known Issues</b>	<b>9</b>

# 1 Running Porchlight

1. Install Python. I recommend using [Anaconda](#) to perform the installation. Follow the [installation guide](#). When you get to *step 8*, make sure to select “Add Anaconda 3 to my PATH environment variable.”

Alternatively, you can install Python from the [official website](#).

Finally, you can install Python from the Windows store.

2. Download program folder from <https://www.gtsilomelekis.com/>
3. Run installation.bat from the folder to install the required packages.
4. Run RunMe.bat to run the program.

# 2 The main program window

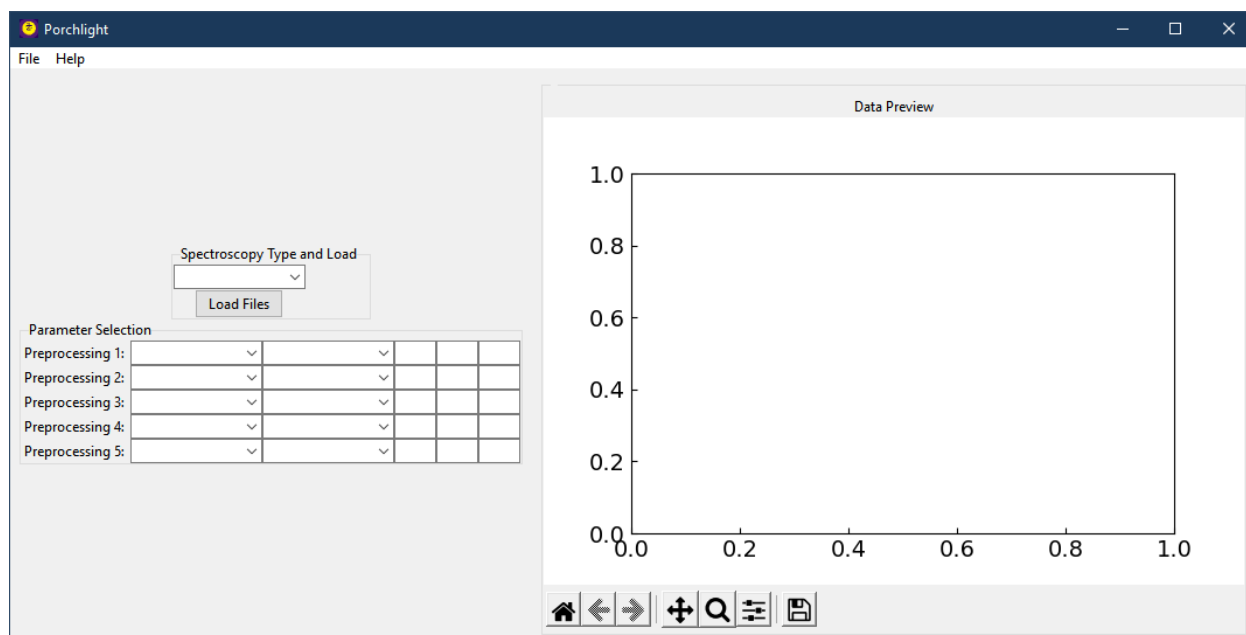


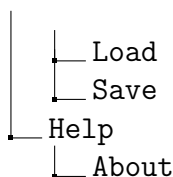
Figure 1: The main Porchlight window.

The main program window for Porchlight is where the user spends most of their time. It is currently broken up into three main parts: the menu bar, the left panel, and the right panel.

## 2.1 The menu bar

The menu bar contains the standard expected options.

Menu Bar  
└─ File



## 2.2 Left Panel

Going from top to bottom:

**Type of spectrum** is a drop down that allows the user to choose the type of data they will load in. Current choices are:

- IR - Transmission
- IR - Absorbance
- Raman
- UV-Vis

The type of spectrum will change the plotting style of the chart on the right, as well as some common motifs seen in that type of spectroscopy, such as the reversed axis in IR spectroscopy.

**Load Data** lets the user choose the files they would like to load. Supported filetypes include:

- CSV
- Text
- Excel
- SPC

Thermo-Galactic uses `spc_spectra` to read the data, so users can install that package using `pip install spc_spectra`.

Porchlight can load multiple files at a time, and those files can contain one spectra or multiple. The main thing to understand is that Porchlight makes the assumption that the longest axis in the table is the direction of the spectra. For example, if a file contains a dataset that is  $3250 \times 24$ , Porchlight will read it as 23 spectra, with the first column containing the energy axis (such as wavenumbers for IR or Raman) and vice versa.

If the energy values don't match up perfectly, Porchlight has some logic that will try to identify and match energies. For example, say there is a section of File A where the energy points go 406.4311, 406.9132, and 407.3953; while File B has the same section as 406.4311, 406.9133, and 407.3953. Porchlight will identify a mismatch in 406.9132 and 406.9133 and combined those into a column at 406.91325.


**Parameter Selection** allows the user to choose a category of transformation to apply, then a specific technique within that category, and then some options. The specific methods and options are detailed in Section 6.

## 2.3 The Right Panel

This section contains the plot of the data. There the data can be interacted by using the widgets at the bottom of the plot.


**Home**  resets the plot to its original position.

**Forward and Back**  and  will redo and undo your changes, respectively.

**Pan/Zoom**  will let you pan and zoom the plot. Use left click to pan the plot. Hold “x” to lock the movement to only the x-axis, and “y” to lock the movement to only the y-axis. Use right click to zoom. You can manipulate the x-axis by pressing “x”, moving to the right to zoom in and to the left to zoom out. Manipulate the y-axis by holding “y” and then move the cursor up to zoom up or down to zoom out. Holding “CTRL” zooms while maintaining the aspect ratio.

**Zoom**  lets you draw a rectangle to zoom to on the plot.

**Subplot Configuration**  lets you manipulate the visible settings of the plot, useful is a label is extending past the visible region but you want to save it.

**Save**  lets you save plot figure to an image file. Supported formats include:

- png
- ps
- eps
- svg
- pdf

## 3 Importing Data

1. Select the type of spectra from the dropdown.
2. Click on the “Load files” button and choose your files. CTRL+left click to select multiple individual files at once, Shift+left click to choose a sequence of files.
3. Press the “Open” button to load the files into Porchlight.

## 4 Exporting Data

To save the data, use the Menu bar by going to File → Save. Porchlight will save to CSV and Excel files.

Please note that the default file format is CSV. Due to a Tkinter bug, if you want to save in a different format such as Excel .xlsx, you must type in the file extension when saving. Meaning, to save the file as a an Excel file name “spectra”, you must type “spectra.xlsx” as the file name.

## 5 File Format Support

Porchlight will natively support loading text files (.txt), comma seperated values files (.csv), and Microsoft Excel files (.xlsx). Thermo-Galactic spectral files (.spc) can be read if the user has the package spc\_spectra installed.

In txt, csv, and xlsx files, Porchlight assumes that the longest dimension of the data is the spectral waveform. To this end, the first column or row will be taken as the spectral axis, and the subsequent data as the spectral intensity values. Files can contain single or multiple spectra as long as a wavelength axis supplied.

## 6 Transformations

This section describes the types of transformations and the options needed.

### 6.1 Trim

Trimming the spectra is recommended unless you are certain the spectra loaded are all the same range. While many methods are one-way methods that are applied to each spectra individually, some methods such as MSC assume the spectra are all the same length.

#### 6.1.1 Trim

Trimming cuts down the spectra to a specified range. Option A is the lower bounds, option B is the upper bounds. The spectra removes anything smaller than option A or greater than option B.

#### 6.1.2 Inverse Trim

The “inverse trim” cuts out data between two points. Option A is the lower bounds, option B is the upper bounds. Useful for cutting out high noise regions in the middle of spectrum, such as the 2300-1900  $\text{cm}^{-1}$  region of FTIR data collected on a diamond ATR.

### 6.2 Baseline

Baseline correction functions attempt to correct for changes that occur in the baseline. Two examples include a drift in the spectrometer calibration or fluorescence in a Raman spectrum.

### 6.2.1 Polyfit

Polyfit takes one parameter, the order of the polynomial in Option A.

The polyfit algorithm works by fitting the data to a polynomial, and then replacing values in the fit by any spectral intensities that are less than the fit. Then, it performs the polynomial fitting again for a set number of iterations or until no changes occur. The idea here is that by always replacing the fit data with lower values, the fit continues to pull down until it follows the baseline.<sup>[1]</sup>

### 6.2.2 AsLS

AsLS is an asymmetric least squares method that uses multiple iterations to attempt to fit a line beneath the spectrum. It takes two parameters, where optA is a penalization term and optB is an asymmetry term. The implementation is based on the one provided Eilers and Boelens.<sup>[2]</sup> They recommend the penalty term to be set between  $10^2 - 10^9$ , while the symmetry term between 0.001 – 0.1. When working in Porchlight, you can use e notation. For example, 10e6 represents  $10^6$ .

The method can be difficult to use because the “optimal” parameters are fairly subjective. Here are some hints based on my experience. The penalty factor (optA) controls the the baseline’s pliability. If optA is small, then the baseline is soft and follows the curves of the spectrum more easily. If the penalty factor is large, then the baseline is stiff and will follow broader trends in the data. Be warned that if the baseline is too soft, then it will start fitting into broad peaks. Too large, and you may not be able to follow the baseline. The author’s describe the asymmetry parameter as being related to noise. The noisier the data, the larger optB should be. It’s challenging to describe it’s behavior, I recommend trying 0.001, 0.1, and 0.1, seeing which gives you a better spectrum, and work from there. If you have a spectrum with sharp peaks and little noise try a small optB value first.

## 6.3 Smooth

### 6.3.1 Rolling

Rolling is a rolling window smoothing, or “boxcar” averaging. OptA is an odd integer for the number of datapoints used in the averaging. The larger the value, the more smoothing you will have but the greater the deformation of the peaks.

### 6.3.2 SGSsmooth

This technique applies a Savitzky-Golay smoothing to the data.<sup>[3]</sup> The Savitzky-Golay smooth is a moving window that fits a polynomial to a subsection of data points, and then sets the center of that window to the fitted polynomial value. Option A is the order of the polynomial and option B is the number of data points. Option A must be an integer greater than 0 that is less than option B. Option B must be a positive, odd integer.

Generally, the following trends are true for smoothing data: a higher order polynomial will result in less smoothing of the data while a larger window will result in more smoothing. As one can imagine, a higher order function will be able to fit the data better, which is why

it appears to have less of an effect on smoothing. However, too much smoothing will reshape your data to the point it may no longer be representative of the underlying spectrum. Peaks will broaden and shrink. Therefore, smooth sparingly.

## 6.4 Normalization

### 6.4.1 SNV

This technique applies the Standard Normal Variate (SNV) transformation.[4] It takes no options.

$$sn_i = \frac{s_i - \bar{s}}{\text{std}(sn)} \quad (1)$$

$sn_i$  is the transformed spectrum,  $s_i$  is a datapoint in that spectrum,  $\bar{s}$  is the spectrum mean,  $\text{std}(sn)$  is the standard deviation of the spectrum.

### 6.4.2 Area

This technique normalizes the spectrum by the integrated area under the curve. It takes no options.

$$sn_i = \frac{s_i}{\int s \, ds} \quad (2)$$

### 6.4.3 Vector

This technique applies a vector normalization.[5] It takes no options.

$$sn_i = \frac{s_i}{\sqrt{\sum s_i^2}} \quad (3)$$

$sn_i$  is the transformed spectrum,  $s_i$  is a datapoint in that spectrum. In literature, often applied after derivatives.

### 6.4.4 MinMax

This technique applies a min-max transformation that normalizes the data between 0 and 1.[5] The largest value in the spectrum becomes 1 and the lowest value in the spectrum becomes 0.

$$sn_i = \frac{s_i - s_{min}}{s_{max} - s_{min}} \quad (4)$$

### 6.4.5 MSC

Applies a Multiplicative Scattering Correction (MSC).[6] The algorithm requires a reference spectrum that is considered the “ideal” spectrum of that dataset. One can be provided by setting option A to a positive integer of the array value in the series, or if set to 0 the reference spectrum is the mean of all spectra.



The algorithm then fits each spectrum to the reference spectrum to a line  $y = mx + b$ . Each datapoint is then manipulated by the equation

$$sn_i = \frac{s_i - b}{m} \quad (5)$$

If a reference spectrum is chosen, it is ignored by any further transformations.

#### 6.4.6 Peak Normalization

Normalized the spectral intensity to be 1 at a given wavelength.[5] OptA is the position of the peak.

### 6.5 Centering

#### 6.5.1 LastPoint

Subtracts each datapoint in the spectrum by the last point in the array.

#### 6.5.2 Mean

Mean centering can refer to subtracting by the mean of the spectrum or by the mean spectrum of all the spectrum.

**Option A = 0** then Porchlight calculates the mean value of the spectrum and then subtracts each point in the spectrum by that value.

**Option A = 1** then Porchlight calculates the mean spectrum of all the spectrum loaded into Porchlight then subtracts each spectra by the mean spectrum.

### 6.6 Derivative

#### 6.6.1 Savitzky-Golay Derivative

This performs a Savitzky-Golay derivative on the data.[3] This method is popular as it simultaneously smooths as it calculates the derivative spectrum. OptA is the window size, OptB is the fitting polynomial order, and OptC is the order of the derivative. Similar caveats to the window size and fitting polynomial apply here as the smoothing. For the derivative order, note that the higher the derivative the lower the signal to noise ratio. Usually, first or second derivatives are enough.

## 7 Known Issues

The loading of files is slow, you may need to wait a few seconds for the data to display.

Certain operations such as baseline fitting are very computationally intensive, if performing baseline fitting be prepared for the program to lock up.

Tkinter on Windows has a bug where you need to explicitly type out the filetype of the saved file, otherwise it defaults to CSV. For example, to save as an Excel file named "data", you need to explicitly type "data.xlsx" into the filename when saving.

## References

- [1] C. A. Lieber and A. Mahadevan-Jansen, “Automated method for subtraction of fluorescence from biological raman spectra,” *Applied Spectroscopy*, vol. 57, pp. 1363–1367, nov 2003.
- [2] P. H. C. Eilers and H. F. M. Boelens, “Baseline correction with asymmetric least squares smoothing,” techreport, Leiden University, Leiden, The Netherlands, Oct. 2005.
- [3] A. Savitzky and M. J. E. Golay, “Smoothing and differentiation of data by simplified least squares procedures,” *Analytical Chemistry*, vol. 36, pp. 1627–1639, July 1964.
- [4] R. J. Barnes, M. S. Dhanoa, and S. J. Lister, “Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra,” *Appl. Spectrosc.*, vol. 43, pp. 772–777, May 1989.
- [5] R. Gautam, S. Vanga, F. Ariese, and S. Umaphy, “Review of multidimensional data processing approaches for raman and infrared spectroscopy,” *EPJ Techniques and Instrumentation*, vol. 2, June 2015.
- [6] P. Geladi, D. MacDougall, and H. Martens, “Linearization and scatter-correction for near-infrared reflectance spectra of meat,” *Applied Spectroscopy*, vol. 39, pp. 491–500, may 1985.