PARS : Programs for Analysis of Raman Spectra

V 1.2.2

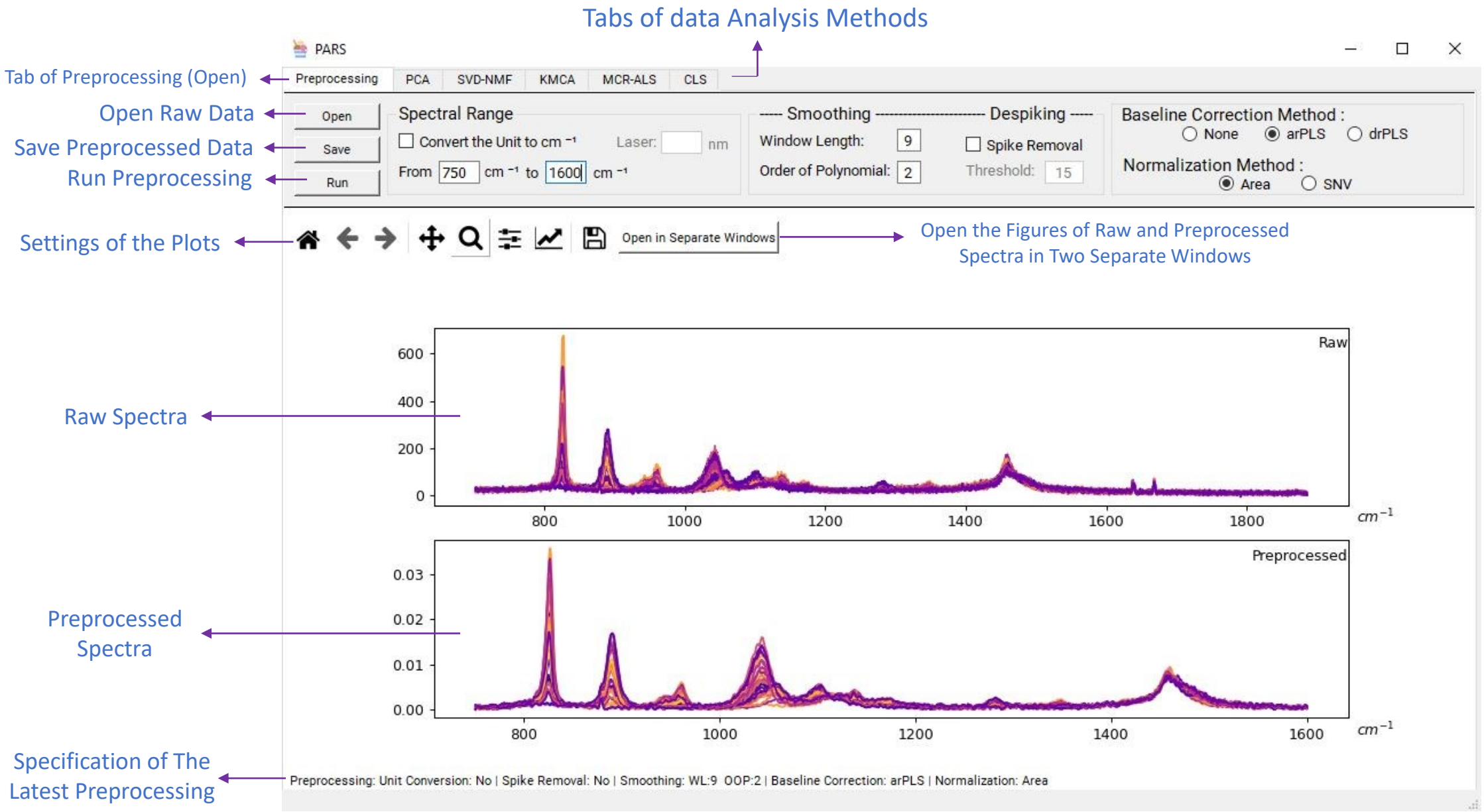
Table of Contents

What does it do?

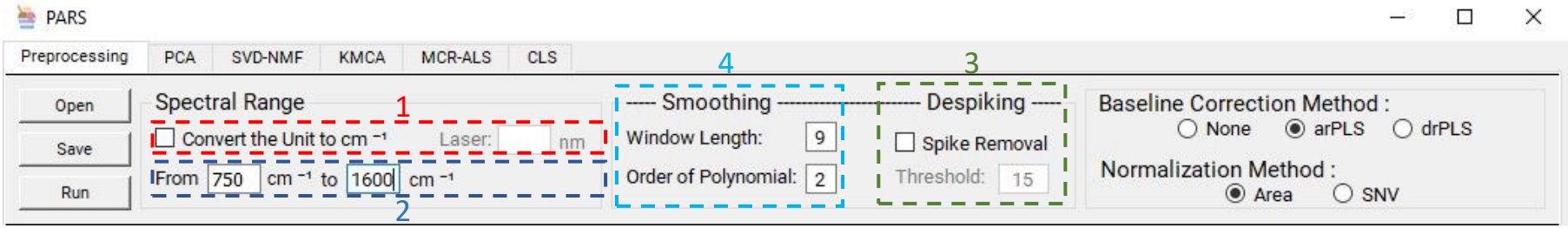
- It is a tool for preprocessing and analysis of Raman spectroscopic data.

- It could be used for
  - Cleaning Raman spectra from noises and artifacts (Preprocessing).
  - Extracting information from Raman spectra. For example, information about compositions of a sample, their pure spectra and their relative concentrations.
  - Reconstructing the distribution maps of the compositions.

Preprocessing (1)



Preprocessing (2)



Steps of Preprocessing:

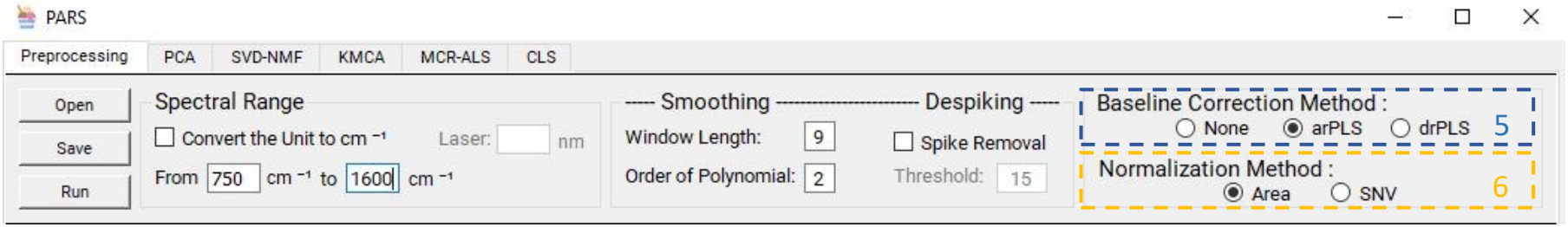
1) Convert the unit to cm^{-1} : It is optional and could be used when the Raman data are saved with wavelength unit. To use this option, the user must enter the excitation wavelength in Nanometer (Wavelength of the Laser).
The transformation formula is : $Raman\ Shift\ [cm^{-1}] = 10^7 \times \left(\frac{1}{\lambda_{Excitation}\ [nm]} - \frac{1}{\lambda_{Scattered}\ [nm]} \right)$

2) Spectral range: truncates the spectrum and only keeps the wavenumbers between the entered numbers. If the first edit box is left empty, the software assumes from the beginning of the spectrum and if the second edit box is left empty, the software assumes until the end of the spectrum.

3) Despiking: This is optional. It removes the spikes in the spectra made from incidence of Muons to the spectrometer's detector. The algorithm searches spikes in the spectra using a moving box of size "Window Length". The lower the Threshold number is, the stronger the algorithm becomes. The number is 15 by default. The best number should be found via trial and error.

4) Smoothing: Reduces the noise of the spectra via the Savitzky-Golay filter. "Window Length" is the number of the points (wavenumbers), and "Order of Polynomial" is the order of the polynomial that is fitted to those points. Window Length must be an odd number.

Preprocessing (3)



5. Baseline Correction: The Raman spectra sometimes have an oblique or curved baseline that are usually made from fluorescence emissions of the samples. This non-straight baseline should be corrected before analysis. The software provides two newest and most efficient methods for this purpose. The methods are:
drPLS : doubly reweighted penalized least squares (DOI: 10.1364/AO.58.003913)
arPLS : Asymmetrically Reweighted Penalized Least Squares Smoothing (DOI: 10.1039/c4an01061b)
The parameters of drPLS are set stronger than arPLS in the software.

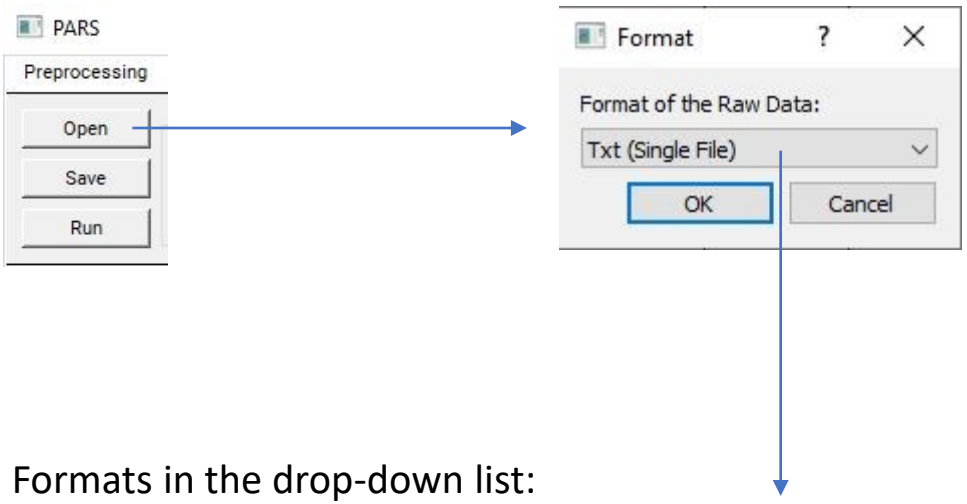
5. Normalization: In order to enable the comparison of spectra with each other for further data analysis and eliminate the systematic differences among measurements, the spectra must be normalized. Normalization is necessary, since the scale of some spectra would change because of parameters such as laser intensity fluctuations or Mie scattering of excitation and scattered light in different parts of the sample. The software provides two famous methods for this purpose:
Area: In some literature it is called vector normalization. This method in the software is coupled with setting the zeros of the spectra to the zero of y-axis, after normalization.

$$norm = \sqrt{S_1^2 + S_2^2 + \ldots + S_N^2} \qquad S_{i\ (Area\ Normalized)} = \frac{S_i}{norm}\ ; i = 1, 2, \ldots, N$$

SNV: Standard Normal Variation. This method suits the best for PCA.

$$Mean = \frac{S_1 + S_2 + \ldots + S_N}{N} \qquad SD = \sqrt{\frac{((S_1 - mean)^2 + \ldots + (S_1 - mean)^2)}{(N-1)}} \qquad S_{i\ (SNV\ Normalized)} = \frac{(S_i - mean)}{(SD)}\ ; i = 1, 2, \ldots, N$$

Preprocessing (4)



Formats in the drop-down list:

1) Txt (Single File) : To choose a single raw Raman spectrum that is saved as Text file. E.g., can be used for seeing and processing of a single Raman spectrum.

2) Txt (Image) : To choose a folder that contains data of a 2d spectral Raman image. The raw Raman spectrum of each pixel must be saved in a separate Text file (With two columns: Wavenumbers and Intensity values). The name of the file must include the location of the pixel in this form: [aa x bb]. E.g.: [00 x 01]

3) HDF5 (Image) : To choose a hdf5 file of raw Raman data that includes two datasets. One dataset must be an array of wavenumbers, and the other dataset must be a 3d matrix of intensity values in this form: hypercube[Intensity values, row, column]

An Example of Naming of The Text Files For The Images

Positions of the Pixels

2Propanol2-Ethanol98_exp27s-acc1 [00 x 01].asc
2Propanol20-Ethanol80_exp27s-acc1 [00 x 02].asc
2Propanol35-Ethanol65_exp27s-acc1 [00 x 03].asc
2Propanol45-Ethanol55_exp27s-acc1 [00 x 04].asc
2Propanol50-Ethanol50_exp27s-acc1 [00 x 05].asc
2Propanol60-Ethanol40_exp27s-acc1 [00 x 06].asc
2Propanol70-Ethanol30_exp27s-acc1 [00 x 07].asc
2Propanol85-Ethanol15_exp27s-acc1 [00 x 08].asc
2Propanol95-Ethanol5_exp27s-acc1 [00 x 09].asc
2Propanol100_exp27s-acc1 [00 x 10].asc
Ethanol100_exp27s-acc1 [00 x 00].asc

Principal Component Analysis (PCA) (1)

PCA is one of the famous unsupervised methods for dimension reduction and feature extraction. It maps the original dataset onto uncorrelated vectors that are called Principal Components (PCs). In other words, It gives a coordinate system based on data to represent the statistical variations in the dataset. Hence, by removing all the redundant data, PCA summarizes data into a few PCs. This simplification enables us to analyze and visualize complicated multidimensional Raman data.

Common ways to display information extracted by PCA are:

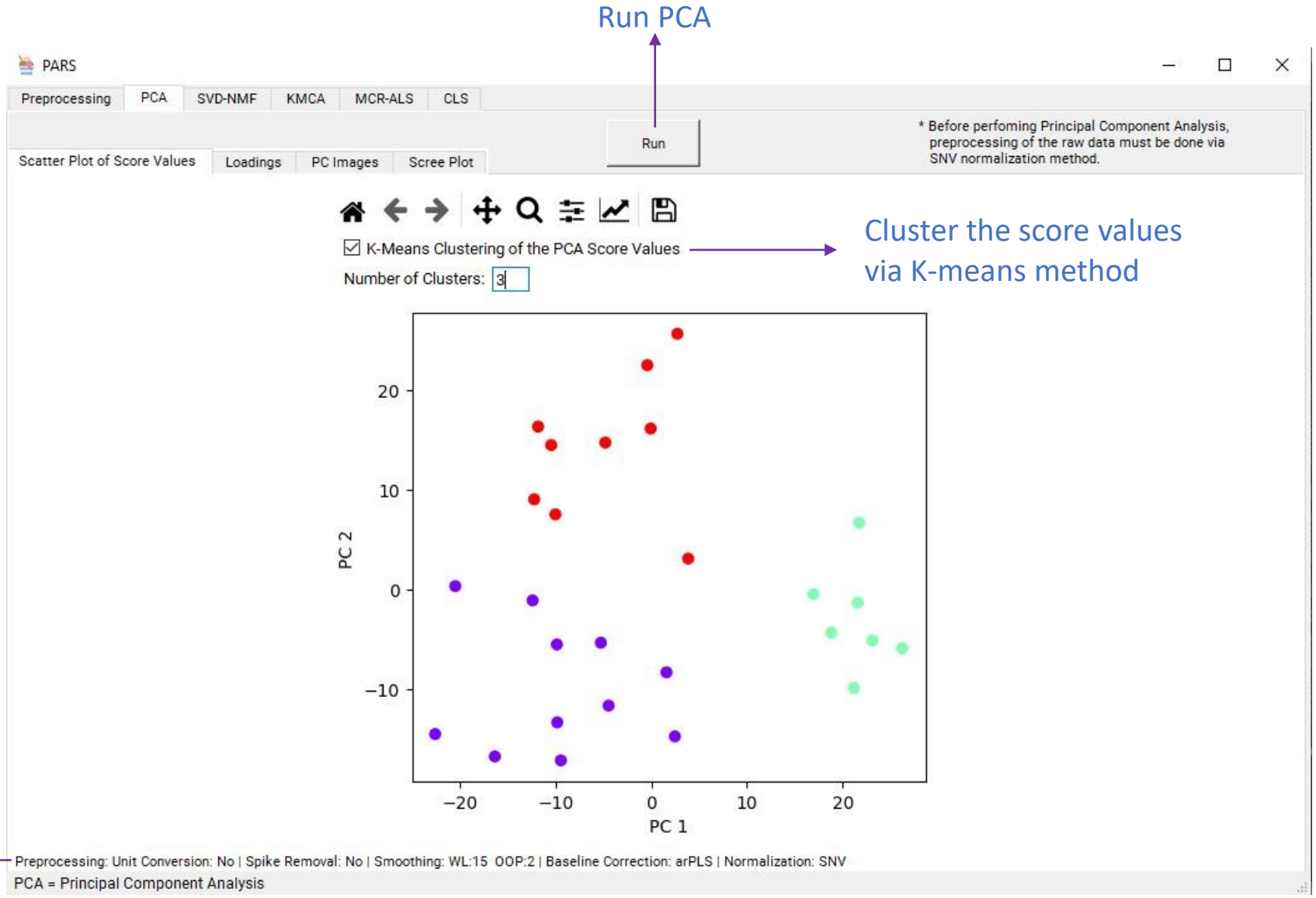
1) Scatter plot of the PC score values of each spectrum.
2) Loadings of the PCs.
3) False-color reconstructed images for each PC based on the score value.

Principal Component Analysis (PCA) (2)

Scatter plot of the PC score values of each spectrum

Please note that before performing PCA on the data, they should be preprocessed. Besides, SNV normalization suits better for PCA.

At the moment, the software shows only the first and the second PCs.



Principal Component Analysis (PCA) (3)

Loadings of the PCs

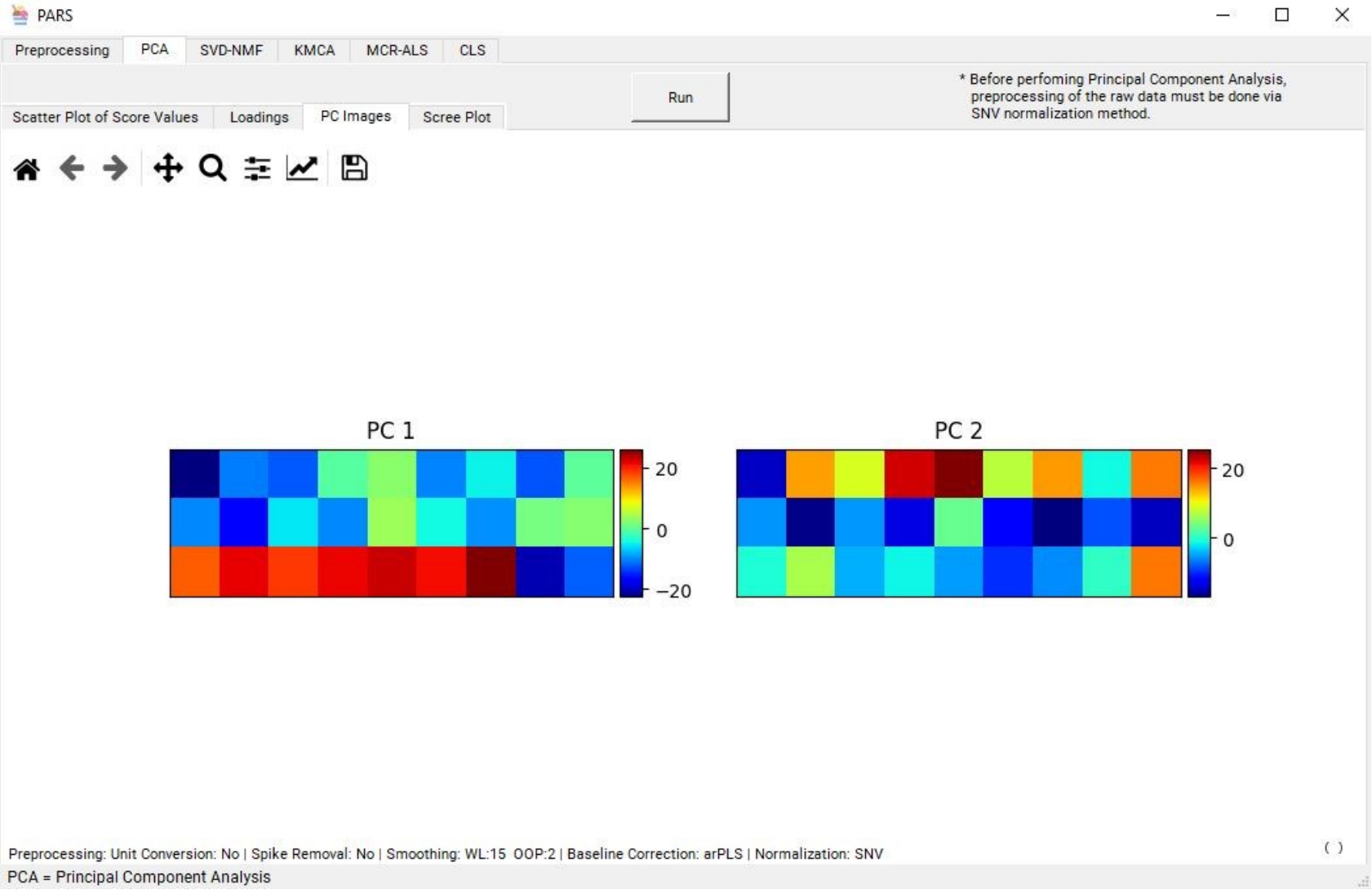
Average of All the Spectra

Loading of PC 1

Loading of PC 2

Principal Component Analysis (PCA) (4)

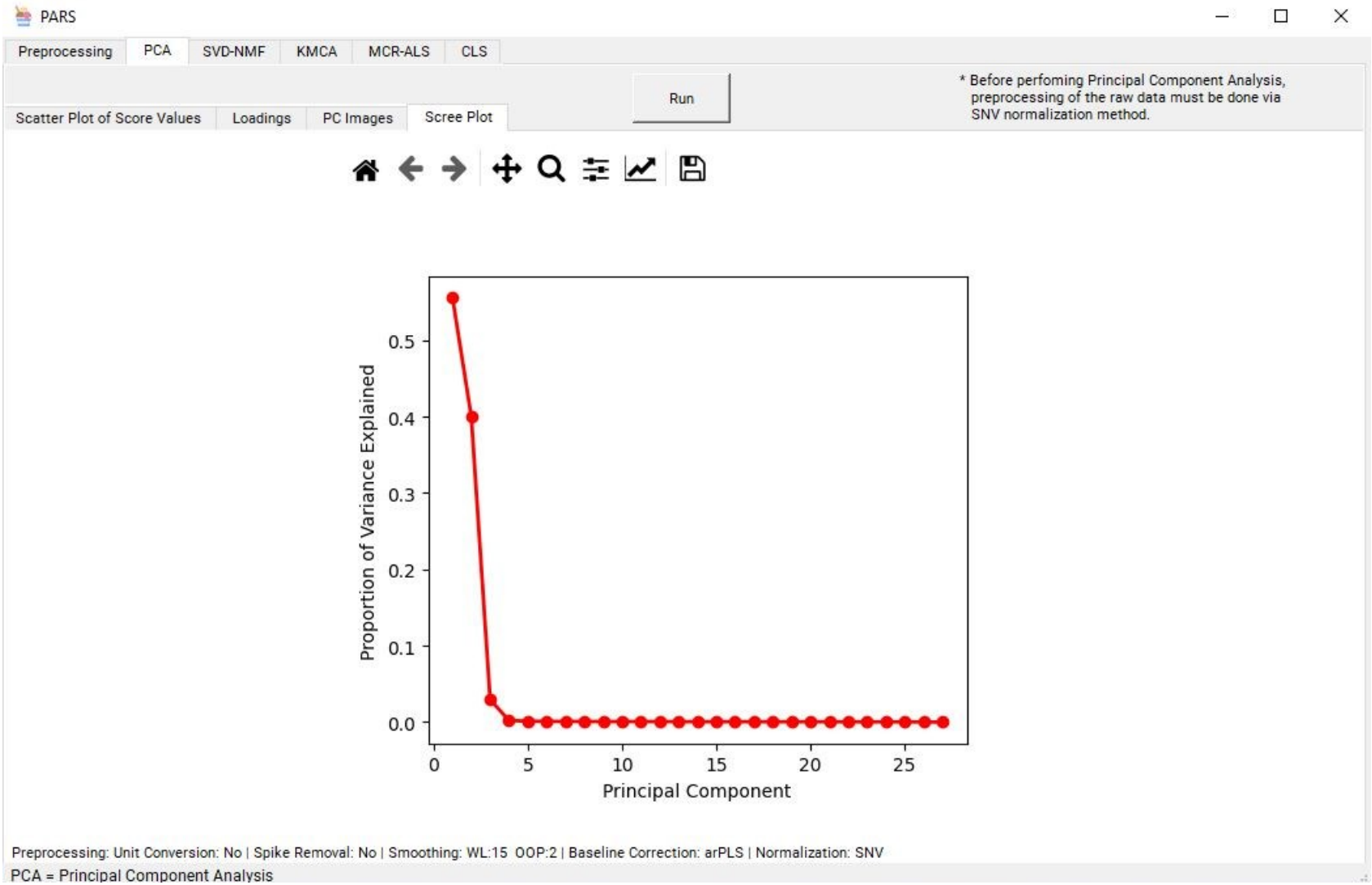
False-color reconstructed images of score values of PC1 and PC2



Principal Component Analysis (PCA) (5)

Scree Plot

The Scree plot of PCA could be used for estimating the number of components

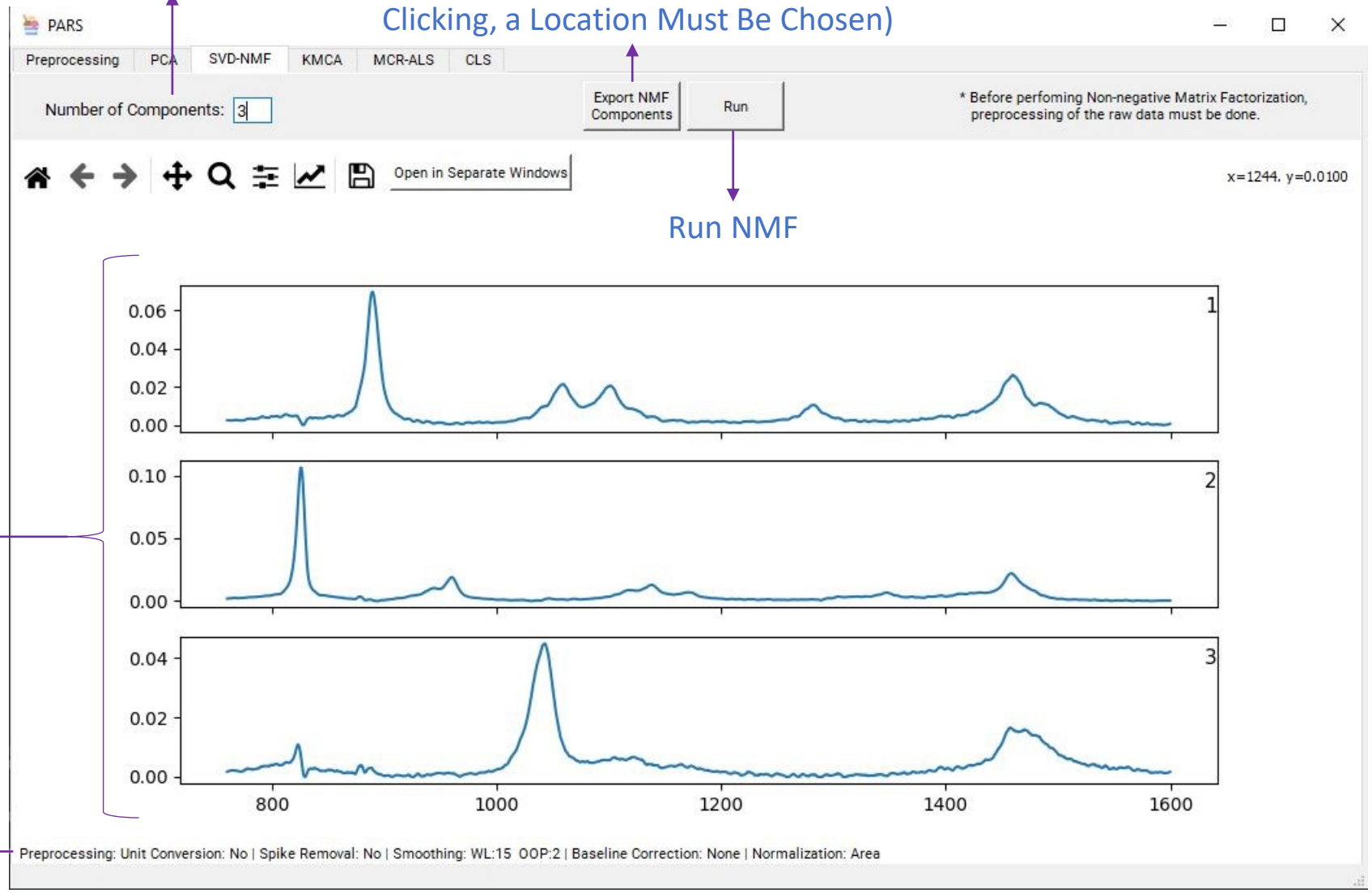


Non-Negative Matrix Factorization (NMF)

Could be used separately for resolving pure spectra (unmixing), as well as providing initial spectra for MCR-ALS.

The Number of Pure Components

Export the Pure Spectra as Text Files (After Clicking, a Location Must Be Chosen)



Run NMF

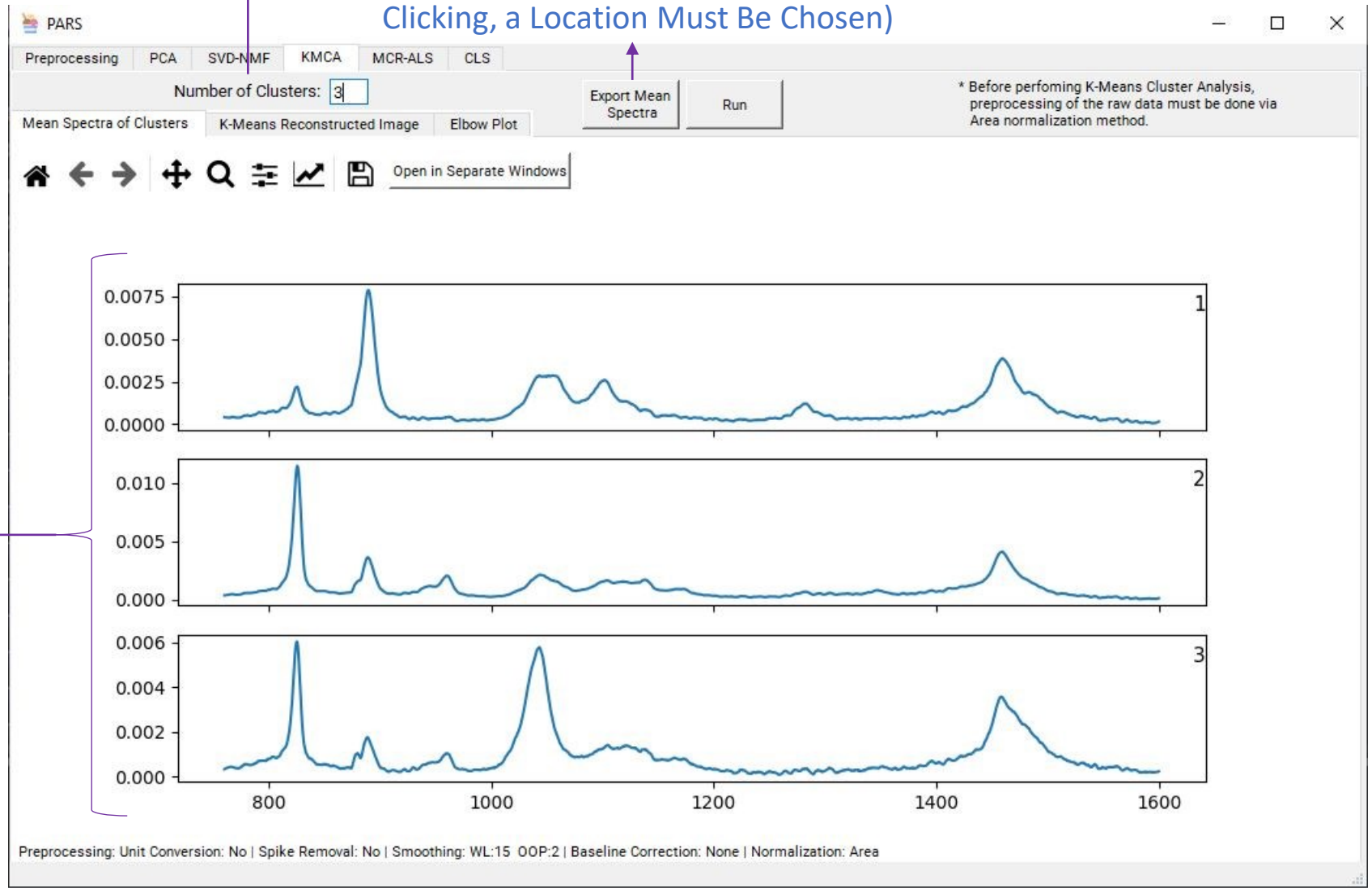
The Unmixed (Pure) Spectra

Specification of The Latest Preprocessing

K-Means Cluster Analysis (KMCA)(1)

The Number of Clusters

Export the Pure Spectra as Text Files (After Clicking, a Location Must Be Chosen)



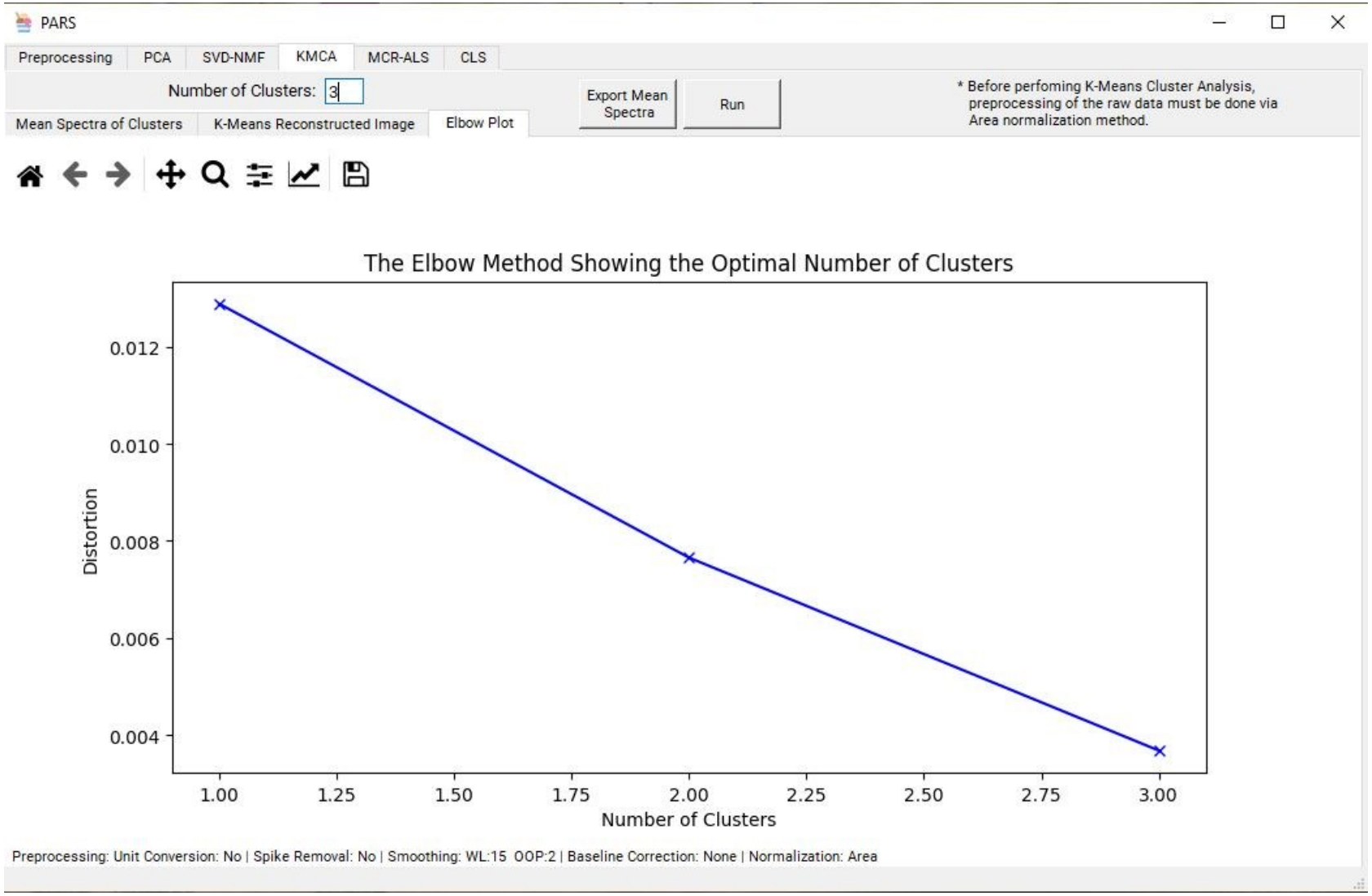
Average Spectrum of the Spectra of the Clusters

K-Means Cluster Analysis (KMCA)(2)

K-Means
Reconstructed Image

K-Means Cluster Analysis (KMCA)(3)

Elbow plot could be used for getting an overview of the efficient number of clusters. The last point before the plot becomes flattened is the optimal number. Therefore, to check the elbow plot, one should first set the number of clusters a large number and checks in what number the plot becomes flattened. The optimal point could be an indication of the number of pure components.



Multivariate Curve Resolution Alternating Least Square (MCR-ALS) (1)

MCR-ALS is a chemometric method that could be used to analyze the Raman spectra collected from mixtures to extract the relative concentrations and the pure spectra of the constituents. The method receives an initial estimate (e.g., initial spectra) and retrieves the components' concentration profiles and pure spectra. In PARS, the initial spectra could be provided via NMF or KMCA. Therefore, before performing MCR-ALS, NMF or KMCA must be done. NMF is recommended for this purpose. Please note that the number of components must be given to NMF and KMCA.

Multivariate Curve Resolution Alternating Least Square (MCR-ALS) (2)

Export the Pure Spectra as Text Files (After clicking, a location must be chosen)

Method For Creating Initial Spectra

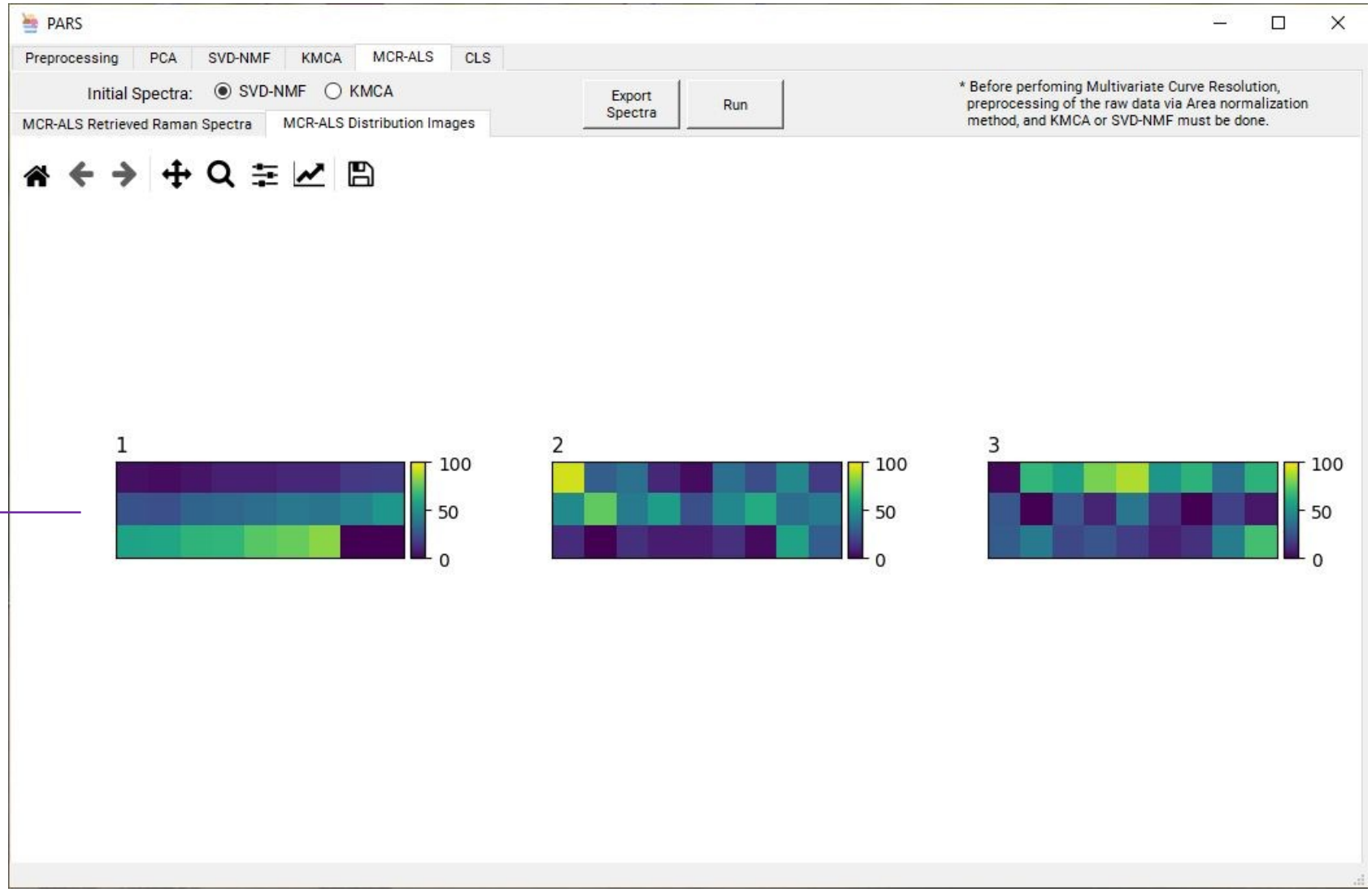
Run MCR-ALS

The Unmixed (Pure) Spectra

Multivariate Curve Resolution Alternating Least Square (MCR-ALS) (3)

The maps of MCR-ALS score values show the relative concentration of the components.

Distribution Maps Reconstructed via MCR-ALS Method



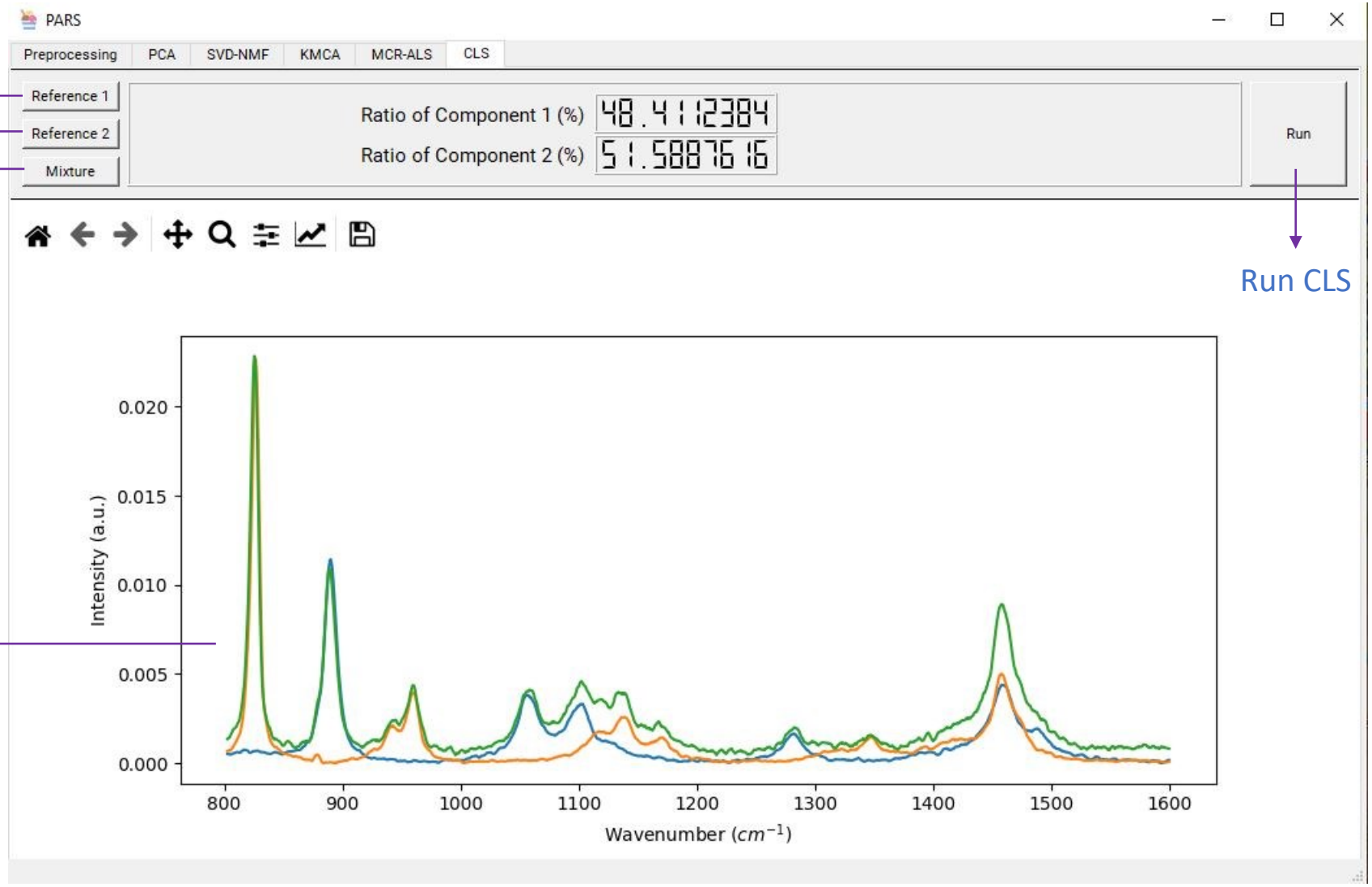
Classical Least Squares (CLS) (1)

Could be used for getting a rough estimation of the ratio of the constituents in a binary mixture.

Pure Spectrum of the First Substance

Pure Spectrum of the Second Substance

Spectrum of the Mixture



Run CLS

Fitting the Spectra of the Pure Substances and the Mixture