Renishaw plc
Spectroscopy Products Division
New Mills, Wotton-under-Edge,
Gloucestershire GL12 8JR
United Kingdom

Tel +44 (0) 1453 524524
Email raman@renishaw.com

www.renishaw.com

RENISHAW
apply innovation™

TM013 – Multi-file data processing WiRE™ 5

This document aims to show the WiRE™ user how and when to use post processing methods for the removal of cosmic ray features and spectral noise, in multi-file data. The Width of features method can also be applied to single spectra.

These methods can be applied to very large datasets of up to 50 million spectra.

Cosmic ray feature removal

(See Training video TM013 for an interactive example on how to use this option)

Why is cosmic ray feature (CRF) removal by post processing useful?

CRFs can degrade the quality and accuracy of any analysis performed on the dataset, especially when multivariate analysis is used. Intense CRFs can dominate the data analysis and "hide" the true chemical or physical information in the Raman spectra.

WiRE 5 methods for removing CRFs following data collection are desirable for several reasons:

- Removing the unwanted features enables easier analysis of the data, with no constraints on the analysis method (univariate or multivariate).
- The methods are fast and targeted so many CRFs can be removed quickly.
- Very large files can be processed (up to 50 million).
- Multiple methods are available to target specific types of data

What CRF removal methods are available in WiRE 5?

Method 1 – Nearest neighbour (Suitable where the data collection step size is less than the domain size, and performs best when the spectra have a good signal-to-noise ratio)

Benefits of the method:

- CRFs are not zapped, but are instead intelligently replaced by the relevant wavenumber region of the most similar neighbour spectrum, therefore band structure and integrity are maintained.

- CRFs of varying width and shape are removed without having to adjust the parameters.

- The Offset variable changes the 'sensitivity' of the method to CRF intensity.

How it works:

- Correlation coefficients are determined for each spectrum with all its spatial neighbours, in order to select the most similar neighbour (MSN) spectrum.

- A small positive offset is applied to this MSN spectrum and CRFs are identified where the original (test) spectrum has a higher intensity than its offset neighbour spectrum.

- The intensity values contaminated by the CRFs are replaced by scaled intensities from the MSN spectrum.

Method 2 – Width of feature (suitable in all cases, and performs well where the spectra have rapidly changing and intense backgrounds)

Benefits of the method:

- Insensitive to intense or rapidly changing spectral backgrounds.

- Insensitive to the spatial relationship between the mapping step size and the domain size. This method can therefore be applied where the mapping step size > domain size, and where 1D maps have been collected (e.g. time series, temperature series and depth series)

How it works:

- Peaks are identified and compared to width and height parameters, designed to distinguish CRFs from real Raman bands.

These methods are often used together to ensure all significant CRFs have been removed.

When should I use each method?

The nearest neighbour method should be used in two main cases:

1. Data is spatially over-sampled (Step size < domains within sample).

2. Neighbouring spectra **do not** show significant relative spectral background changes

The width of feature method can be used in any case; as the spatial relevance of one spectrum to another plays no part in determining the presence of a CRF.

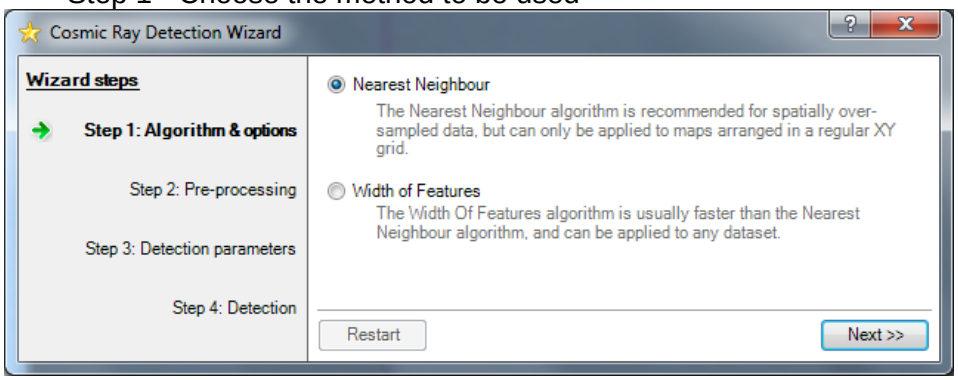
Using CRF post processing removal

Data processing is significantly faster when applying to data stored on a local internal hard drive.

1. Load mapping data into WiRE 5 and select Processing....Cosmic ray removal.

2. The detection Wizard is launched, where CRF candidates are determined.

- Step 1 - Choose the method to be used

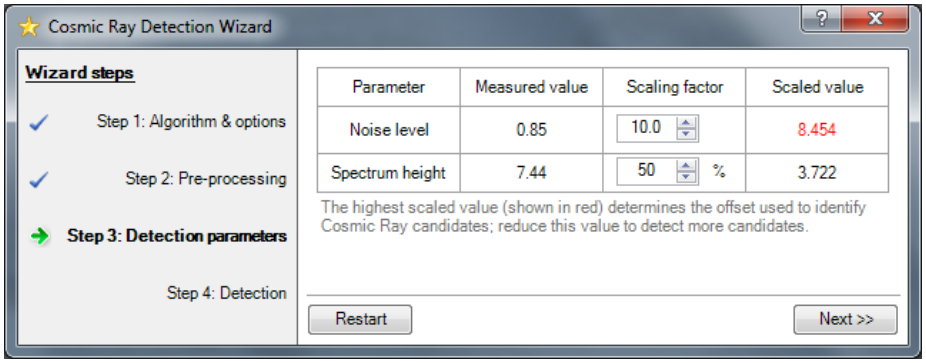


- Select Next to automatically initiate Step 2, pre-processing

- Step 3 enables the user to control the number and severity of the CRF candidates

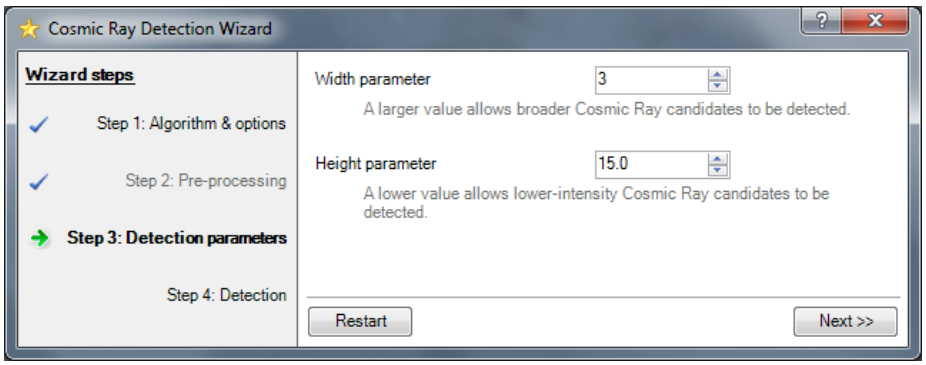
Nearest neighbour – control the scaled value

It is initially recommended that the default value (red) be used, in which case Next can be selected.



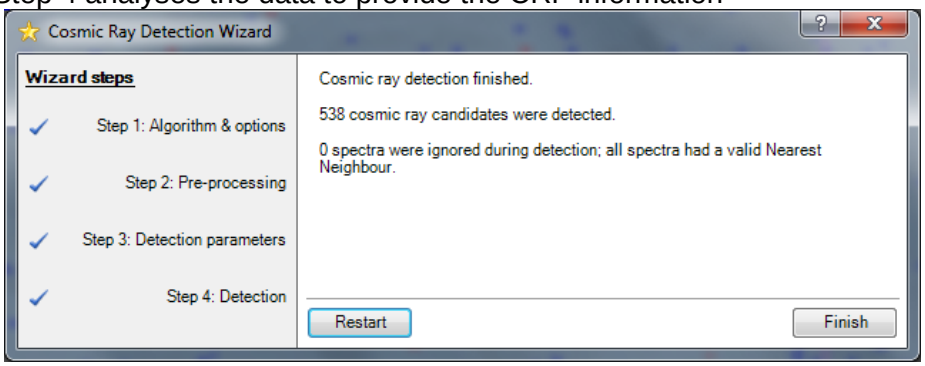
Reducing the maximum of the two values shown (by reducing the scaling factor) will reveal further, less obvious, CRFs.

Width of feature – limit the width and height of CRFs



Increasing the Width parameter allows broader CRFs to be detected, at the possible risk of incorrectly identifying real Raman peaks as CRFs. Decreasing the Height parameter allows weaker CRFs to be detected, again with the same risk.

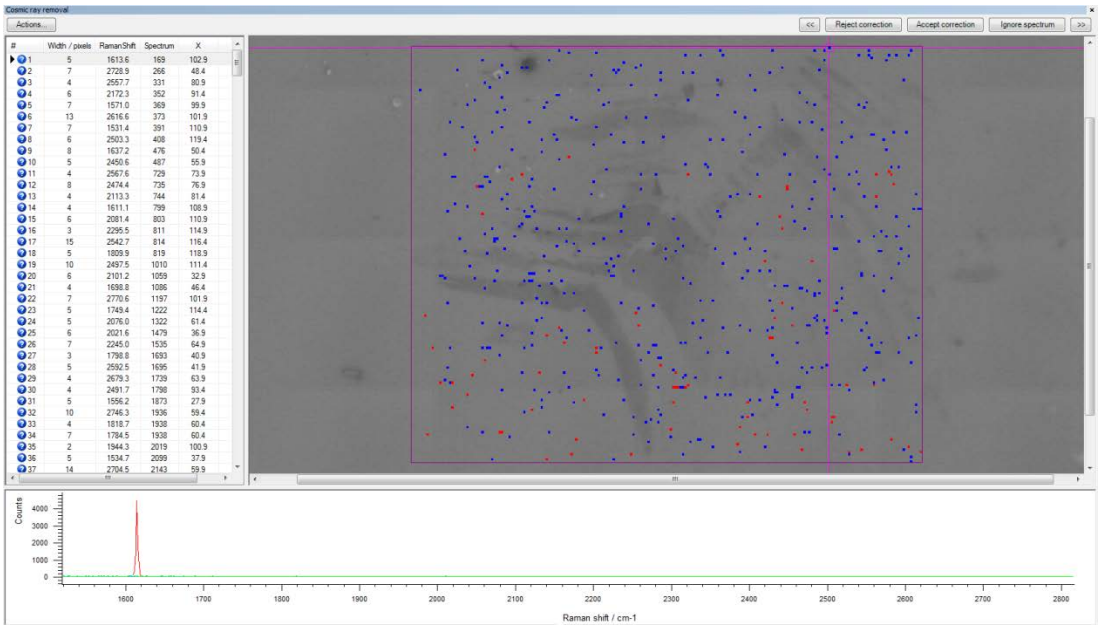
- Step 4 analyses the data to provide the CRF information



The number of CRFs detected is shown.

3. Select finish to then review the provided CRF candidate options.
   This will present a screen containing:

- The white light image of the map region with the location of the CRF candidates overlaid.
- The spectrum at the cross-hair location.
  - Red regions within the spectrum are proposed to be removed (i.e. should represent a CRF).
  - The blue spectrum represents the best nearest neighbour if this method has been selected
  - The green spectrum represents the corrected spectrum (see Spectrum and map colours under the ‘Actions’ button)
- A table containing the CRF candidate information.
- An ‘Actions’ button (providing equivalent information to the right click context menu options).
- Decision control over individual (and grouped) CRFs, top right tool bar.



Most CRF candidates are initially unassigned ( ). The Unassigned state indicates that this CRF candidate has not yet been reviewed. Some CRF candidates are automatically rejected ( ) when the correlation between the spectrum and the best nearest neighbour spectrum becomes worse if the CRF is removed. These auto-rejected CRFs should be checked and can be grouped in the table.

4. There are two main methods by which CRF candidates can be rationalised:

   a. Manual, iterative processing by the user
   b. Targeted grouping using the table

a. Use the options on the top right part of the toolbar to iteratively navigate through single CRF candidates and decide to:

- >> Skip to the next CRF
  This will leave the CRF candidate unchanged, typically unassigned ( ) and available for subsequent detection.
- Ignore spectrum Ignore entire spectrum
  This will disregard the spectrum ( ) from the current or any future detections. Spectra can be un-ignored from the 'Actions' menu (Actions....Ignore spectrum....Un-ignore selected spectra)
- Accept correction Accept proposed CRF candidate removal
  This will remove the red spectral regions ( ) shown in the spectrum and replace these with scaled nearest neighbour information (for the nearest neighbour method). The new spectrum is shown in green.
- Reject correction Reject proposed CRF candidate removal
  This will disregard the proposed removal ( ) from the current detection.
- << Skip to the previous CRF
  This will leave the CRF candidate unchanged and enables previous decisions to be checked or changed.

b. Use the table to order CRF candidates and enable batch assignment

The table contains information on the:

- **assignment status**
- **width of CRF candidate removal (pixels)**
- **position of the CRF candidate centre (cm^{-1} shift)**
- **spectrum number (acquisition number)**
- X spatial point (µm)
- Y spatial point (µm)

The bold table columns can be sorted from low to high by clicking on the column heading. This can be used to target specific feature width and positions, enabling a more targeted approach to CRF candidate removal.

Once a CRF candidate is selected in the table, the Up and Down keys can be used to quickly navigate to other CRF candidates in the table.

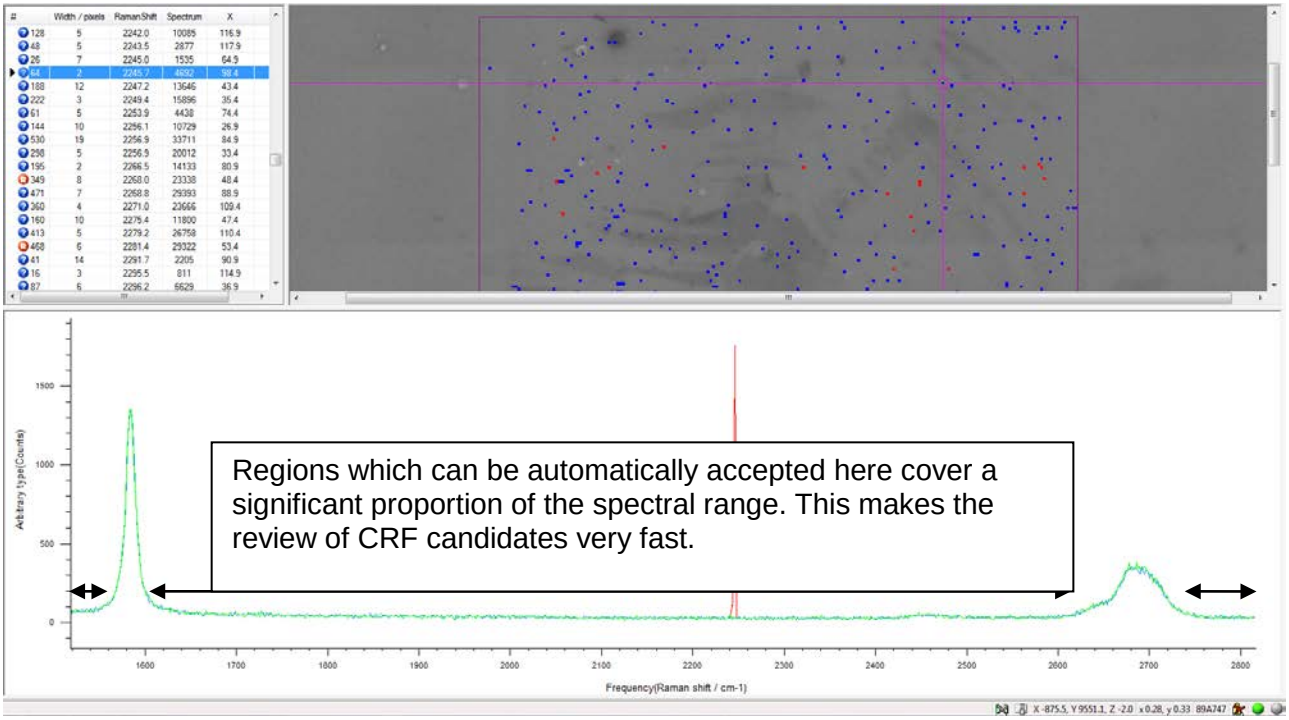
Multiple non-adjacent rows (CRF candidates) of the table can be selected by holding down the CTRL keyboard button. A range of adjacent rows can be selected using the SHIFT button. The tool bar options can now be applied to multiple CRF candidates at the same time.

Sorting by width enables:

- CRFs narrower than typical Raman bands to be quickly and easily accepted.
- Very broad CRFs to be validated
- The threshold between CRFs and real features to be determined, where the sensitivity to less significant CRFs is high, or the Raman band width is very narrow (e.g. gases).

Sorting by position enables:

- CRF candidates to be accepted where Raman bands are known to be not present within the data (e.g. spectral regions between the G band and 2D band of graphene can be automatically accepted – see below)
- CRF candidates do not appear randomly over the spectral range making it faster to decide if they are CRFs or real features.



5. When CRF candidates have been assigned the processing operation can be completed by 'committing' the results. This is done either under 'Actions' or on the Context menu.

6. When this is completed, the CRF removal process can be repeated if desired (through modification of the detection method or parameters in the Wizard).

7. "Committed" Cosmic Ray corrections are only retained whilst the file is open in WiRE. To make the corrections permanent, leave the Cosmic Ray Correction tool and use the "Save file" or "Save file as..." option from the "File" menu in WiRE.

Other useful options

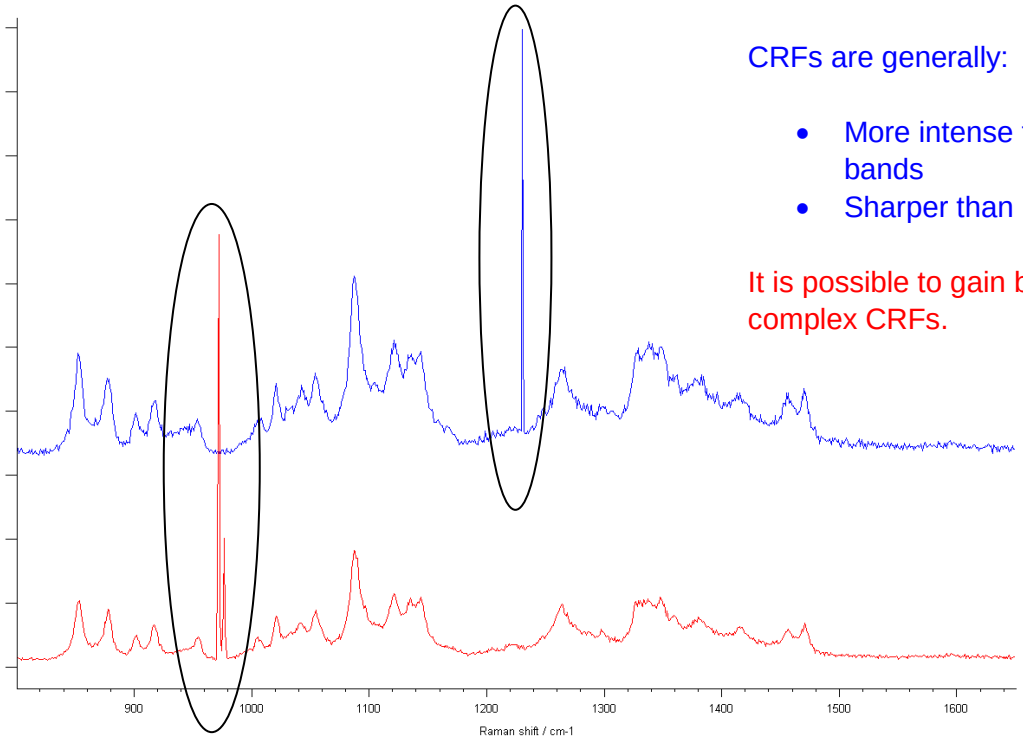
Additionally the 'Actions' button provides the ability to:

- accept or reject all unassigned CRF candidates
- accept, reject or ignore all saturated spectra
- un-ignore spectra
- reset selected CRF candidate to initial state
- exit cosmic ray removal without committing the results
- skip assigned rays when navigating

What are cosmic rays?

Cosmic rays are high energy particles which originate from beyond the Earth. They are effectively completely random but somewhat relate to solar activity. The particle impacts the CCD detector and produces a charge which gets interpreted in the same way as when photons impact the detector. Generally the resulting spectrum contains a cosmic ray feature (CRF) identified by a sharp peak. The angle and energy of the cosmic ray determines the magnitude of the CRF in the spectrum and the number of pixels 'contaminated' in the spectrum. Although unlikely for short acquisitions, using a longer acquisition time will increase the likelihood of a cosmic ray impacting the detector during data collection.

Example spectra show the different types of CRFs which can be observed within Raman spectra.



CRFs are generally:

- More intense than Raman bands
- Sharper than Raman bands

It is possible to gain broad and complex CRFs.

Noise filtering

(See Training video TM013 for an interactive example on how to use this option)

Why is noise filtering by post processing useful?

Noise filtering reduces the level of random noise in a dataset, whilst keeping the important spectral changes we are interested in.

Having less noise in our data has several benefits:

- direct analysis of the raw data produces images with higher signal to noise
- direct analysis may become possible as very weak features may now be visible

- spectra extracted from the dataset are more visually appealing due to the reduced noise level, and can be used as:
    - examples for reports/publication (with suitable acknowledgement)
    - references for component analysis (see TM015)

When using advanced multivariate analysis methods such as principal component analysis (PCA), or Empty modelling™, there is usually no significant benefit in noise filtering.

Which noise filtering method is available in WiRE 5?

Noise filtering is performed most effectively on datasets with relatively large numbers of spectra (>1000) and is therefore ideal for applying to large mapping datasets.

Principal component analysis (PCA, see TM015 Multivariate data analysis) is applied to the dataset and the user is presented with principal components (PCs) in order of decreasing significance. The user can visually inspect each component in turn, and should aim to retain all components that contain real Raman information. All other components are removed (noise components) and the original dataset is recreated.

When should I use noise filtering?

Noise filtering is most effective on data with large and distinct spectral changes throughout the dataset. It is least effective on data which has many small variations (e.g. subtle shifting of Raman band position and width). This is due to the large number of PCs required to appropriately represent the different variants of Raman data.

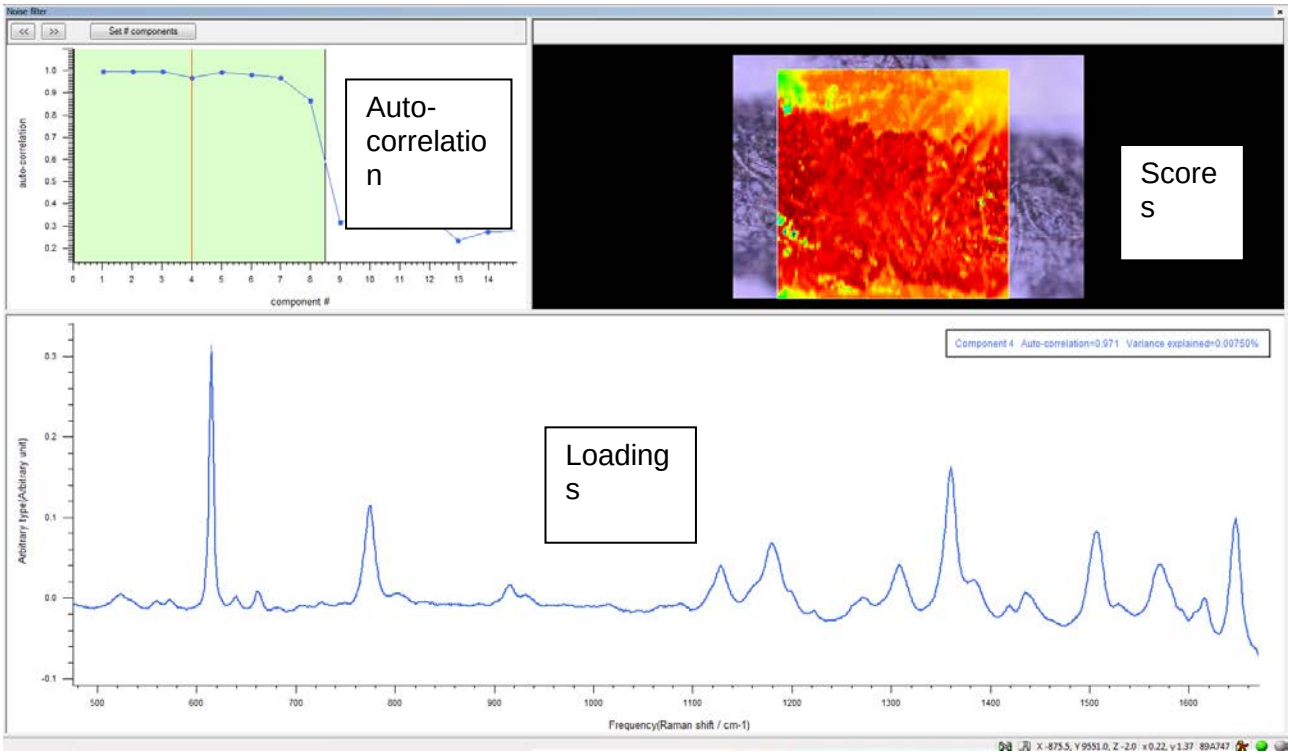
Using noise filtering processing

1. Load mapping data into WiRE 5 and ensure significant cosmic rays have been removed from the dataset.

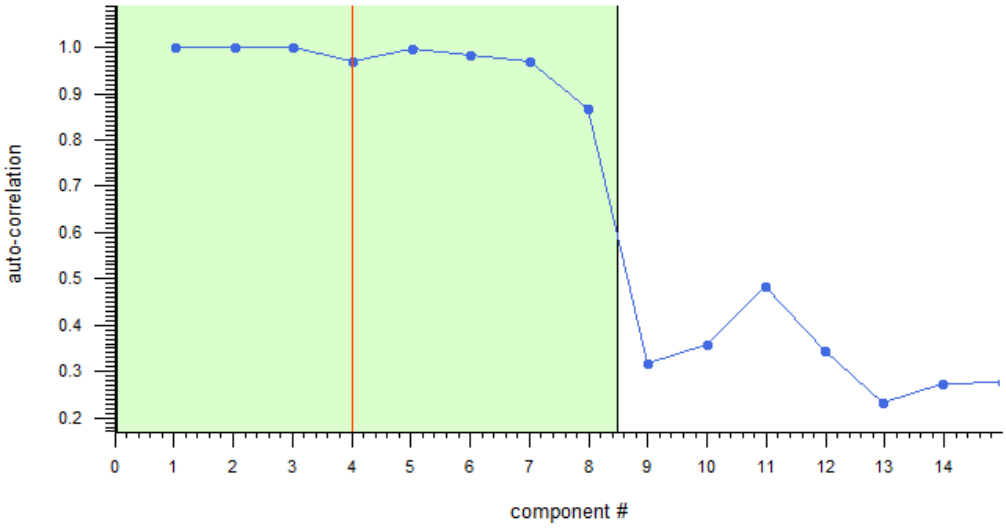
2. Select Processing....Noise filtering.

3. The data are analysed.
   The scores and loadings are available for each PC to allow the user to determine whether the component relates primarily to real Raman signal or noise.
   A plot of the auto-correlation of each component's Loadings vector versus the PC number is also shown.

   Together these displays are intended to aid the user as to where the threshold should be set in removing the noise and keeping the data.



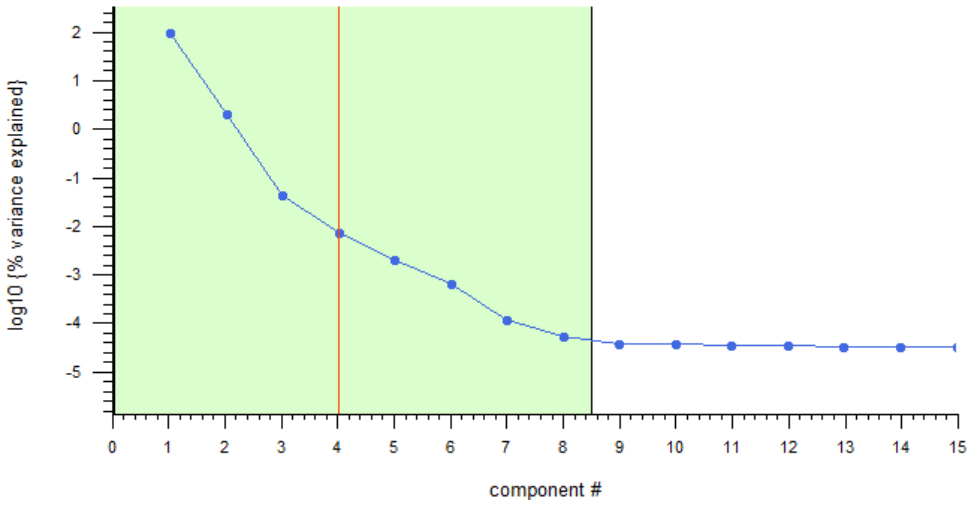
Auto-correlation



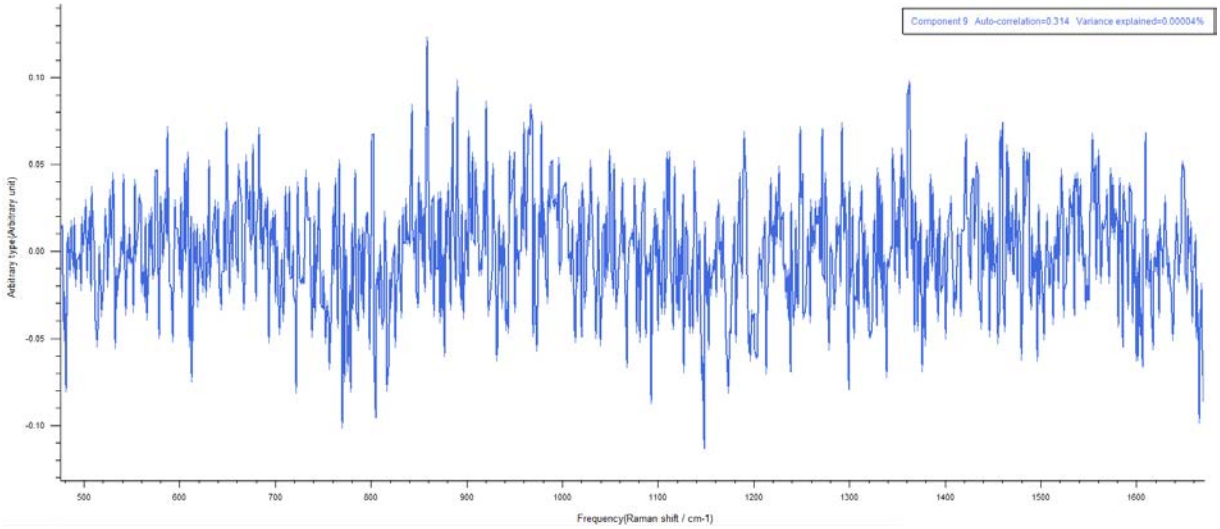
- high loading auto-correlation values are likely to represent Raman data
- low values are somewhat random and typically represents noise

The crossover point between a high and a low auto-correlation value in the principal components may therefore be used as an indicator for which ones to be kept. Typically, a threshold is set, and (starting from the last principal component), the point at which the auto-correlation value rises above that threshold is used to determine how many principal components should be kept. In the above figure, the auto-correlation value for the 9th component is below the chosen threshold (here 0.6), and the 8th is above it, therefore the automatic filtering will keep the first 8 components. In some cases, it will be necessary to review each component in turn, as outliers may have a significant effect on this.

The auto-correlation display can be changed to represent variance explained. Here the point of the curve where it becomes flat is used to judge the threshold point.



Loadings spectrum



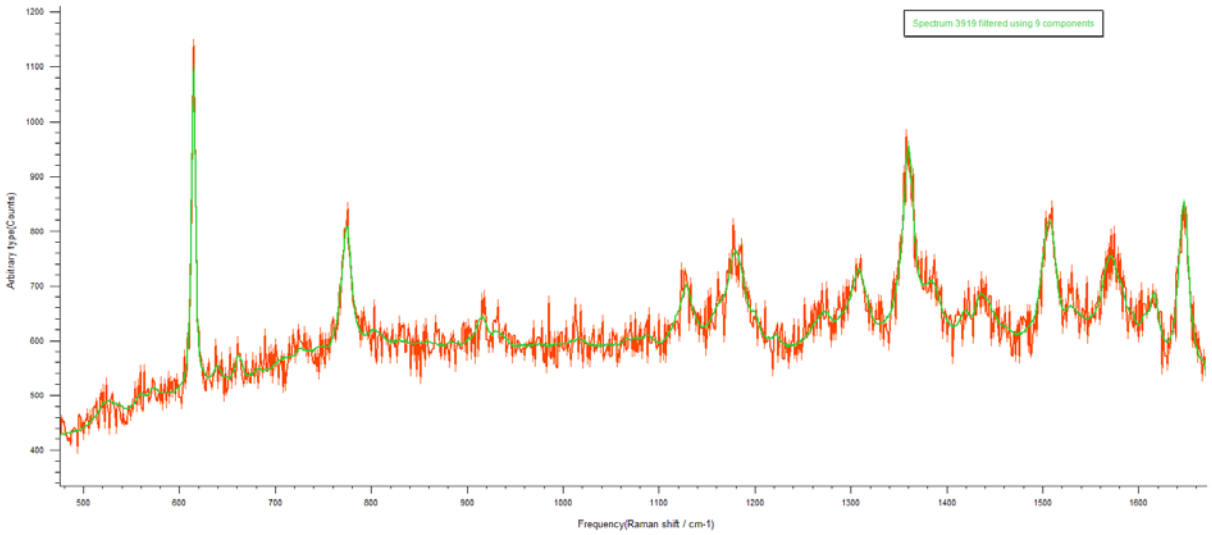
- noisy, featureless loadings indicate the point at which all data has been described through the preceding PCs from a spectral perspective

- a loading is generated for each PC

The loadings display can be changed to preview a noise filtered spectrum against the original. You should check that all the peaks in the original spectrum also appear in the preview of the filtered spectrum, and that their intensities are correct. Missing peaks, or peaks with an incorrect intensity, indicate that the threshold between signal and noise has not yet been reached.

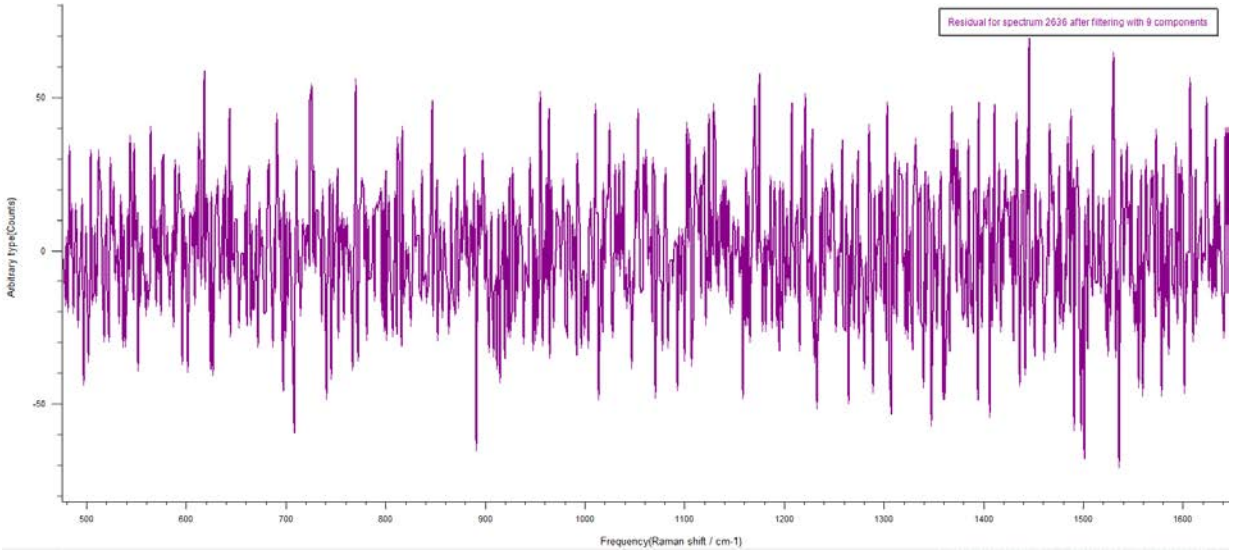
- changing the number of PCs will provide a live update on how this affects the noise filtered spectrum

- click on the scores image to select previews from different sample points

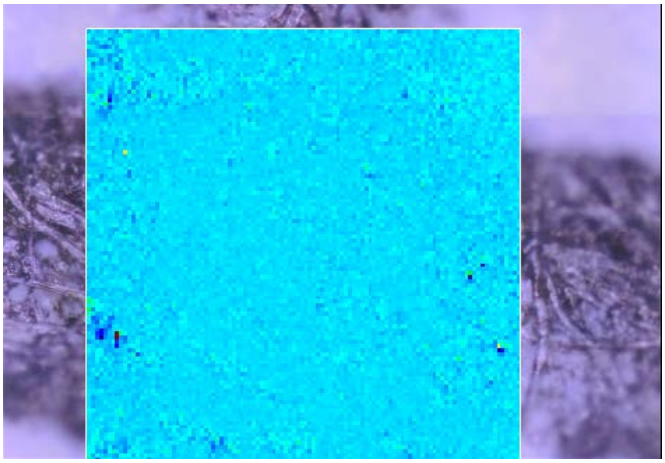


The loadings display can be changed to preview the residual spectrum (difference between original and noise filtered spectrum). If the residual spectrum contains any significant deviations away from random noise, this indicates the threshold between signal and noise has not yet been reached.

- changing the number of PCs will provide a live update on how this affects the residual spectrum
- click on the scores image to select previews from different sample points



Scores image



- noisy, featureless scores images indicate the point at which all data has been described through the preceding PCs from a spatial perspective

- a scores image is generated for each PC

4. Select the PC threshold point by moving the red bar of the auto-correlation (or variance explained) to the last PC to be kept in the data.

5. Select the 'Set # components' button.
   PCs to the left are kept, PCs to the right are removed. The dataset is then recombined.

   The exact number of PCs containing significant Raman information will completely depend on the variance within the data (the number of chemical components resulting in spectral differences) and even intra-component variance (e.g. crystal orientation). Also whether a component is significant or not depends on the aim of the analysis and related requirements within the dataset.

6. Right click to access the context menu from which the processing can be:

   - accepted for all spectra
   - rejected

If accepted the data will be processed. This is not saved to file until a save, or save file as.... operation is performed from WiRE.

Removing single loadings from Raman map data

Principal component analysis analyses map data and reveals spectral patterns presented in order of significance to the total contribution of the data. Most spectral patterns (loadings) represent Raman bands and their changes. As the entire dataset is described by the loadings, any patterns in the data which are not real spectral features (e.g. induced electronic noise) from the sample are also revealed. Within the Noise filtering option it is possible to remove specific loadings and reconstruct the original map data.

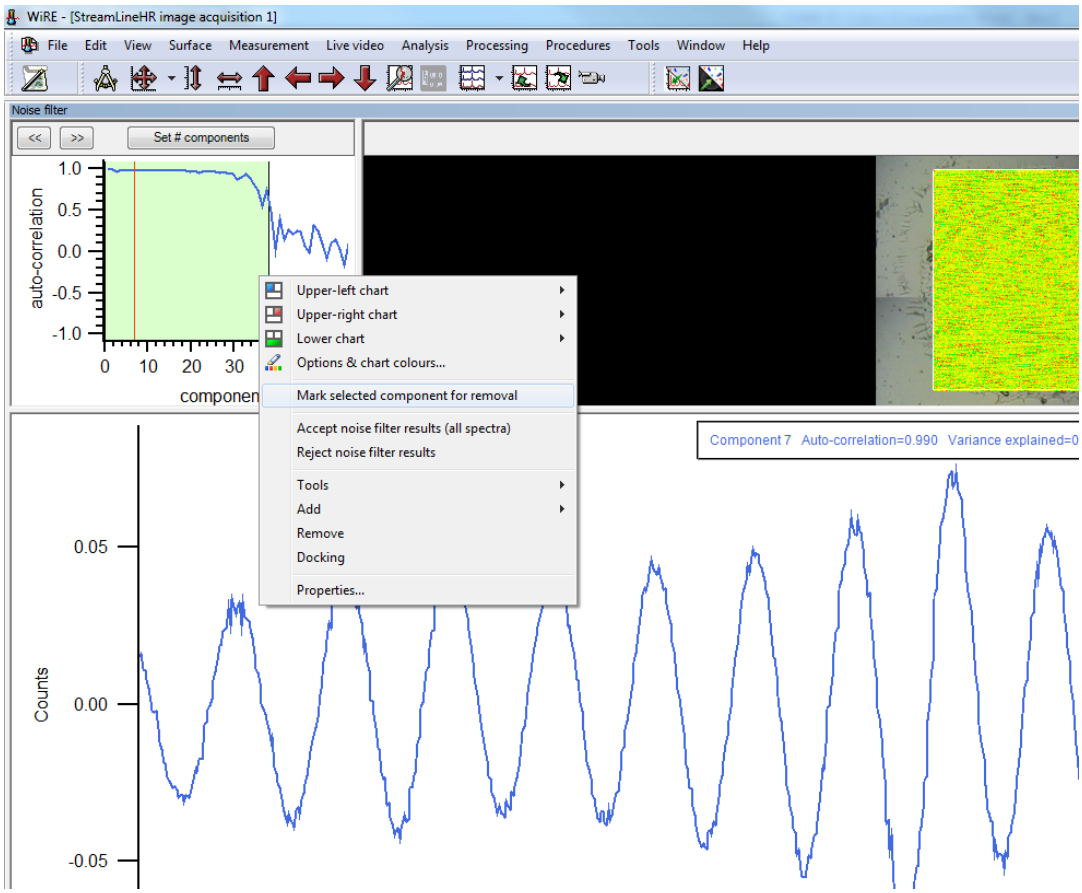
This option should be performed with caution, removing a loadings component which is not noise will change the original data.

1. Load mapping data into WiRE 5 and ensure significant cosmic rays have been removed from the dataset.

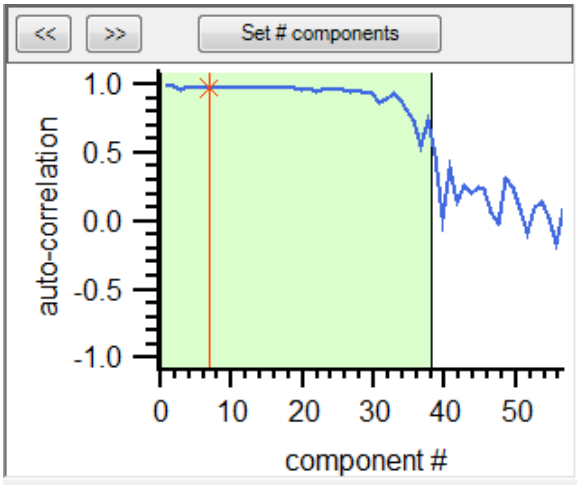
2. Select Processing....Noise filtering.

3. Navigate to the loading to be removed so it is displayed in the bottom viewer

4. Right-click and select *'Mark selected component for removal'*



5. A red cross appears on the auto-correlation plot



Right-click again to either unmark the component for removal or remove the marked component