



WiRE™ software: Cluster analysis and data classification

Gain powerful insight with machine learning in WiRE software

The ability to group highly similar spectral data is a key result for research areas including biology and medicine. It enables qualitative conclusions to be drawn objectively based on statistics, rather than subjectively based on an analyst's experience.

Renishaw's Windows®-based Raman Environment (WiRE) software is available with two options which can be used to group data; these are cluster analysis and data classification.

Cluster analysis is a fully integrated analysis operation performed within WiRE software. Data classification is performed on the separate Data Classifier application enabled by WiRE software. It combines dedicated data processing options with model building, optimisation and testing. Unknowns can be collected in WiRE software and classified using the Data Classifier application in one simple step.

Two methods for maximum flexibility

Whilst the two techniques share a common aim, several key differences exist in the way they work.

Supervision

- Cluster analysis is unsupervised and separates data into clusters based on their spectral differences. With the Data Classifier application you provide data of known classes. Unknown data can then be classified using a supervised approach.

Quantify performance

- Cluster analysis results are limited to the final spatial and spectral output. This means it can be subjective as to how well the data is being analysed.
- Within the Data Classifier application, a range of outputs are provided to evaluate how well the data is separated into groups. You can then assess how reliably the principal component-linear discriminant analysis (PCA-LDA) model can classify unknown spectra.

Knowledge of sample

- Your data can be completely unknown when analysing with cluster analysis.
- With the Data Classifier application, you need spectra from known samples to build the PCA-LDA model. The classification groups are therefore user-defined before data collection.

Extent of differences

- If your spectral components show obvious differences (e.g. different materials in a tablet), then cluster analysis will likely be sufficient to separate the data into groups.
- If component spectra are very similar (e.g. different disease states in biological tissue), then the Data Classifier application can provide greater specificity when grouping the data.

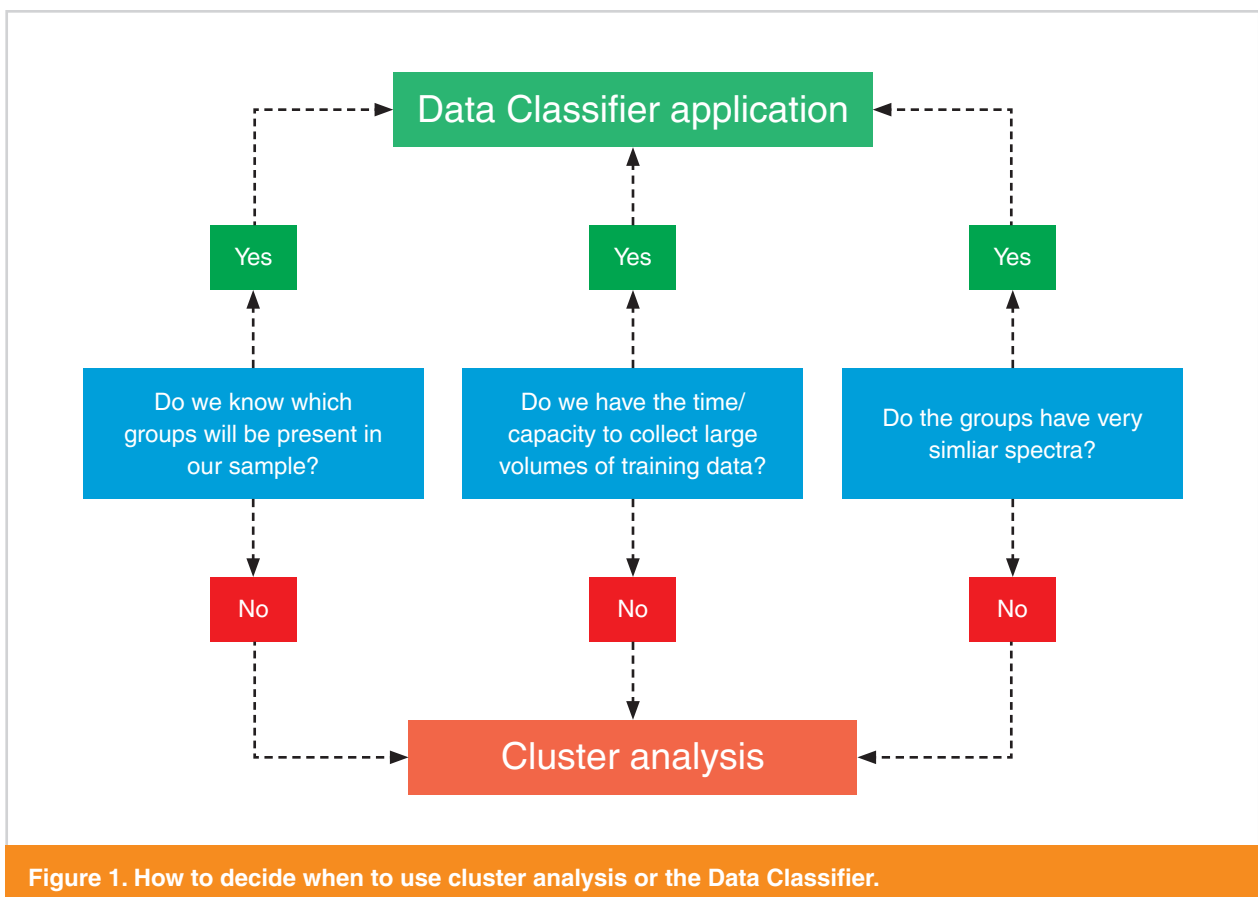


Figure 1. How to decide when to use cluster analysis or the Data Classifier.

Cluster analysis

For speedy, speculative or investigative purposes, cluster analysis is a great tool to use. WiRE software supports K-means and hierarchical options with dynamic dendrogram controls. You can change the number of groups and see how the resulting spectrum from each cluster is affected. Different linkage types and sub-options mean you can apply these methods to any sample type.

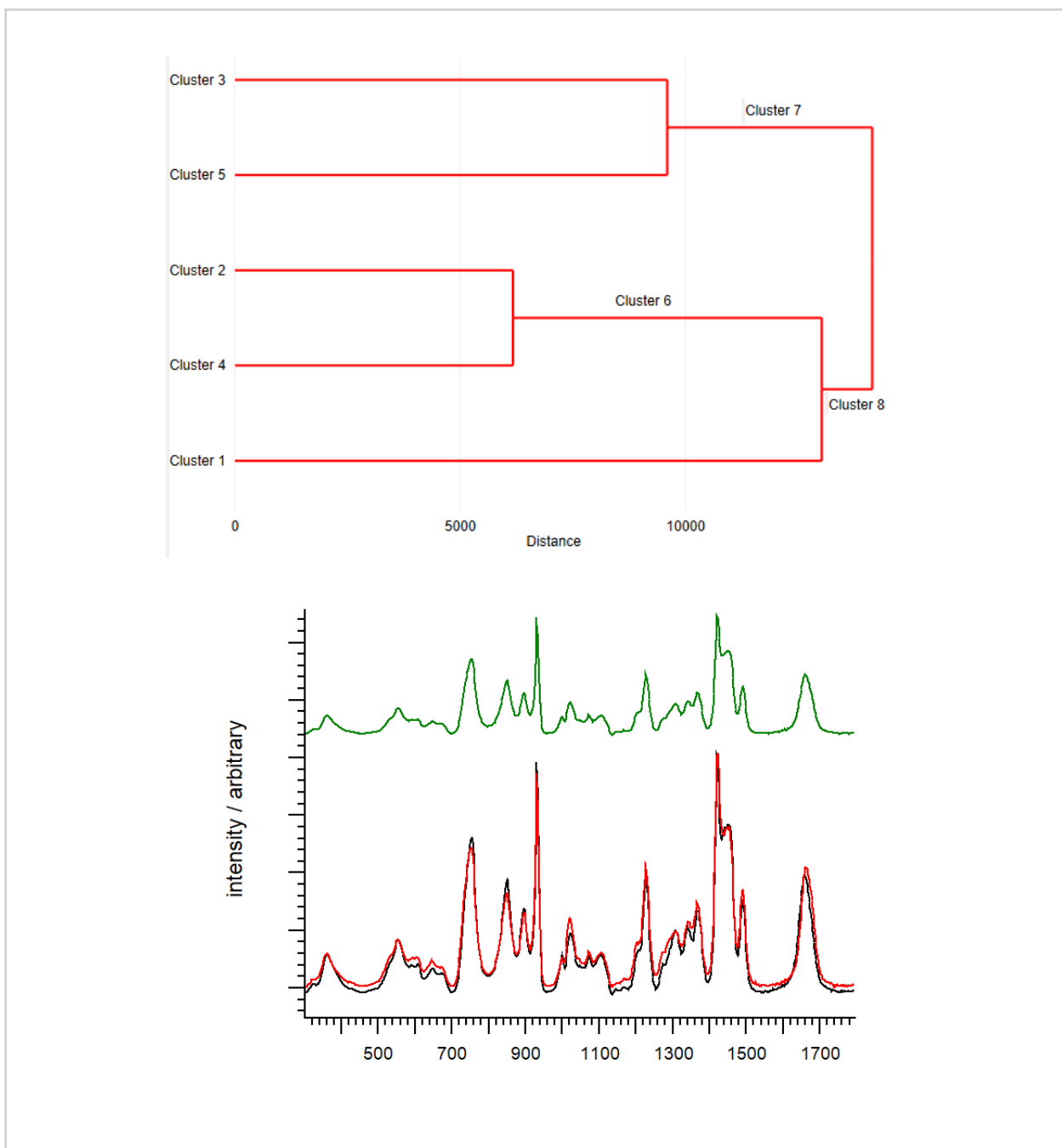


Figure 2. Example of hierarchical cluster analysis to spectrally separate groups of pure pharmaceutical materials. The dendrogram (top) indicates which spectra are clustered together and the relative differences between the clusters. The data first splits into those which are most dissimilar. Clusters can be further split to categorise small but discernible spectral features. For example, cluster 6 (green) splits to cluster 2 (black) and 4 (red) which represent povidone and crosopovidone respectively.

The Data Classifier application

For data analysis where the classes are already known, the Data Classifier application is the tool of choice. Initially, known spectral data are required as a training set for building the classification model. However, each class of spectra is comprehensively quantified with resulting scatter plots and receiver operating characteristic (ROC) curves. This enables optimisation of the classification model; improving sensitivity and specificity of the classified unknown, as shown in Figure 3.

Get clear results from complex data

Grouping your Raman data can be incredibly useful, and sometimes is the main aim of the study. Renishaw's WiRE software clustering analysis and Data Classifier application together support a wide range of data grouping requirements. K-means and hierarchical clustering are ideal options for investigating your data. In contrast, the dedicated Data Classifier application is a complete package. You can build, test and validate your model all-in-one. Results are quantified, so you can confidently apply the model to areas where performance is critical, such as clinical research.

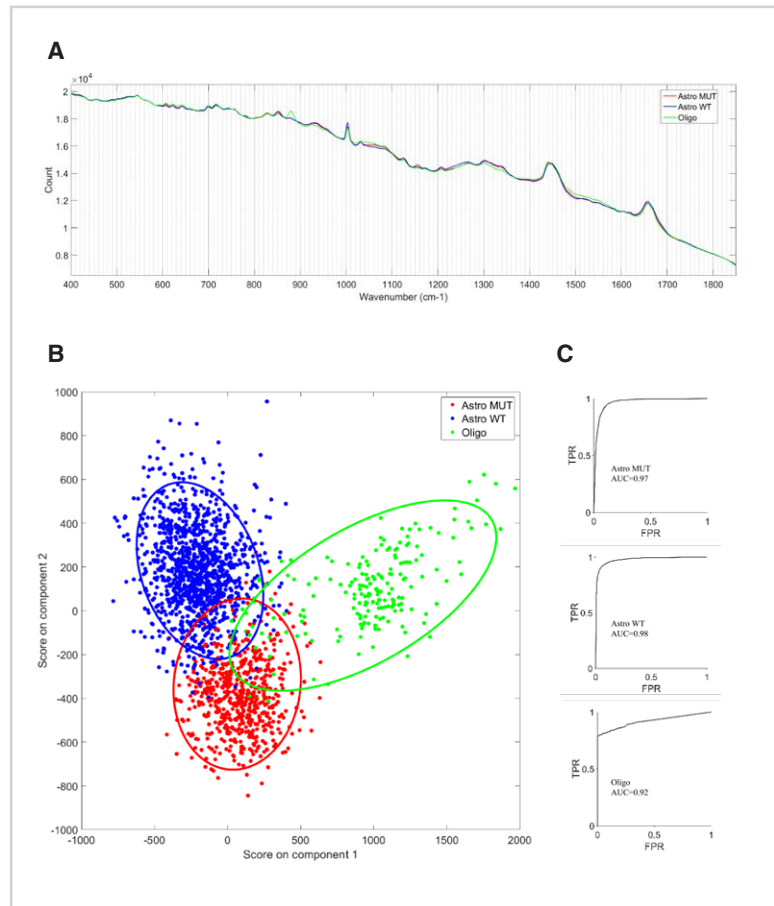


Figure 3. Example of data classification into separate different genetic subtypes of brain tumour tissue. Average raw spectra are shown for three predefined groups (A). Scatter plots show the separation of the three groups (B). ROC curves show how well the data is classified (C).

Want to learn more? Contact our spectroscopy team to discuss your specific requirements. raman@renishaw.com

References

Livermore et al. Rapid intraoperative molecular genetic classification of gliomas using Raman spectroscopy, *Neuro-oncology Advances* 1 (1), 2019.

www.renishaw.com/raman

#renishaw

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

© 2022 Renishaw plc. All rights reserved. RENISHAW® and the probe symbol are registered trade marks of Renishaw plc. Renishaw product names, designations and the mark 'apply innovation' are trade marks of Renishaw plc or its subsidiaries. Other brand, product or company names are trade marks of their respective owners. Renishaw plc. Registered in England and Wales. Company no: 1106260.

Registered office: New Mills, Wotton-under-Edge, Glos, GL12 8JR, UK.

WHILE CONSIDERABLE EFFORT WAS MADE TO VERIFY THE ACCURACY OF THIS DOCUMENT AT PUBLICATION, ALL WARRANTIES, CONDITIONS, REPRESENTATIONS AND LIABILITY, HOWSOEVER ARISING, ARE EXCLUDED TO THE EXTENT PERMITTED BY LAW.

PN252(EN)-01-B