KnowItAll Software Training

Chemometrics Analysis with KnowItAll Trendfinder

Spectrum Analysis

How to Use KnowItAll Trendfinder to Perform Chemometrics Analysis

Purpose

These exercises demonstrate how to use KnowItAll Trendfinder to perform Chemometics analysis on various spectra.

Objectives

These exercises will teach you to apply KnowItAll Trendfinder to

- IR
- LC-MS
- GC-MS
- Raman
- UV-Vis

Background

The KnowItAll Trendfinder application allows one to perform Chemometrics analysis of spectral and chromatographic data.

Training Files Used in This Lesson

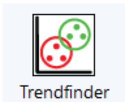
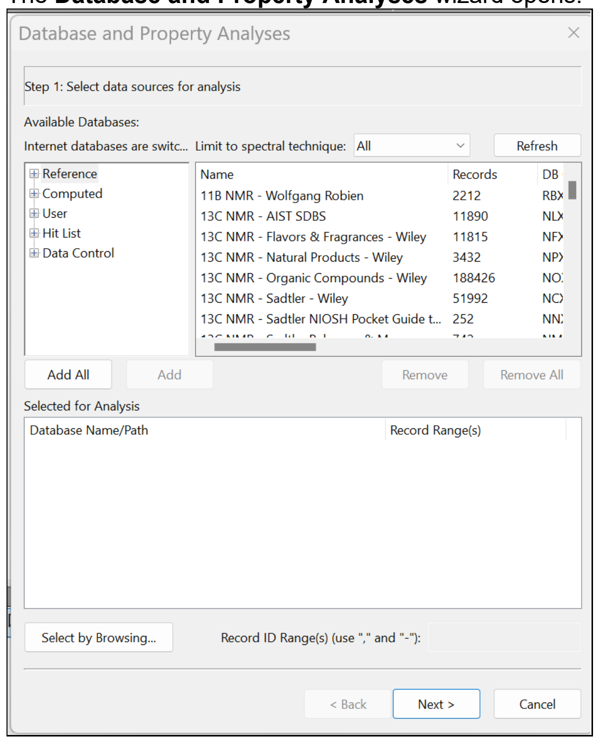
C:\Users\Public\Documents\Wiley\KnowItAll\Samples

KnowItAll Applications Used

- KnowItAll Trendfinder

IR Example

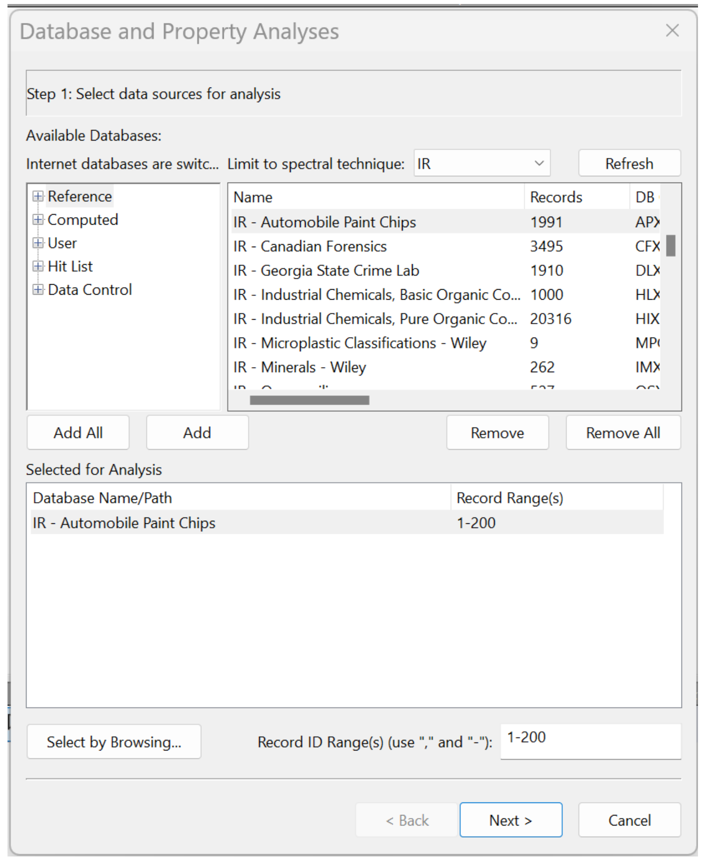
A. Principal Component Analysis (PCA) of Automobile Paint Chips

	Action	Result
1	Open the **Trendfinder** application by clicking its icon in the **Spectral Analysis** toolbox. 	The application opens.
2	Choose **File > New PCA Analysis**.	The **Database and Property Analyses** wizard opens. 

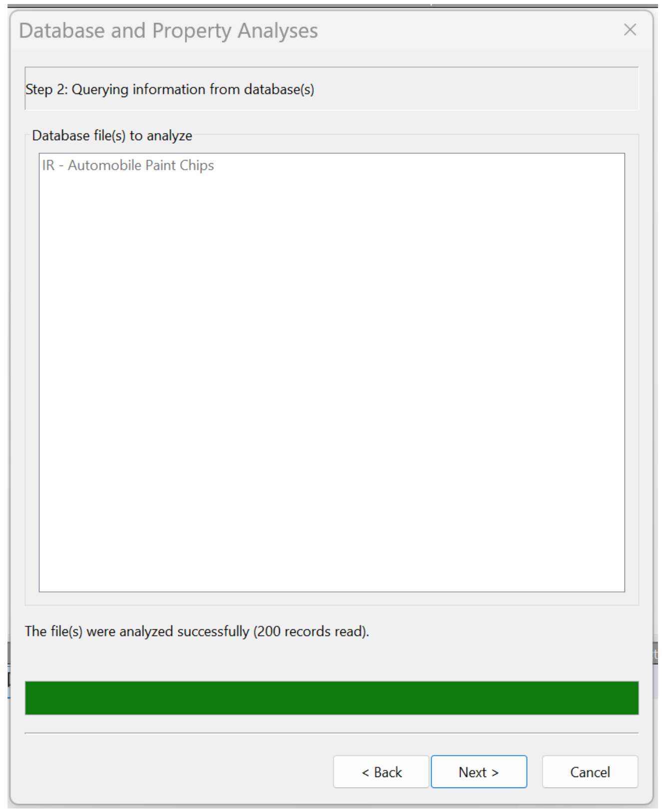
WILEY

<table>
<tr><th></th><th>Action</th><th>Result</th></tr>
<tr><td>3</td><td>

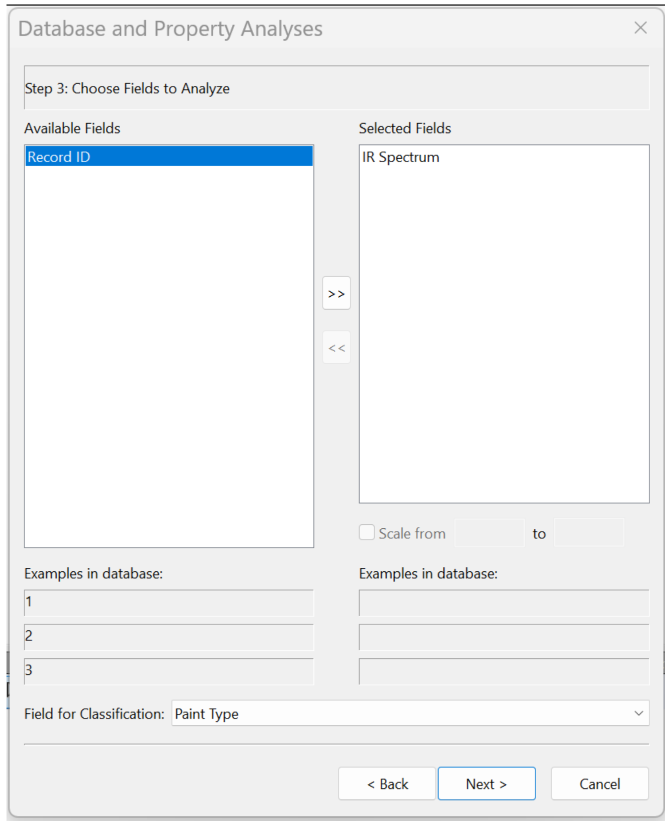
- Select the reference database **IR - Automobile Paint Chips** (APX) and click **Add**.
- Highlight the database in the Selected for Analysis box, type **1-200** in the **Record ID Range** field (to use the first 200 records).
- Click **Next >**.

</td><td>


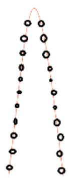
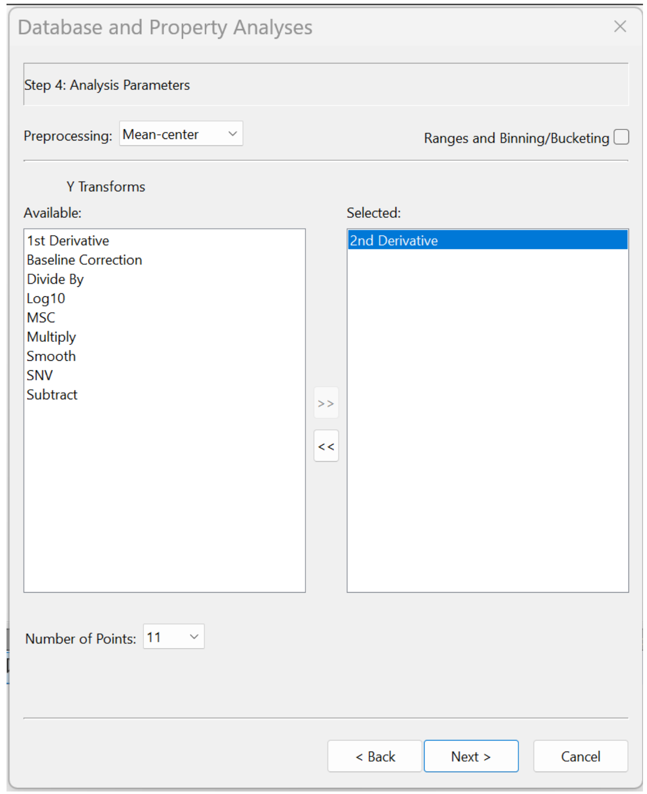
</td></tr>
</table>

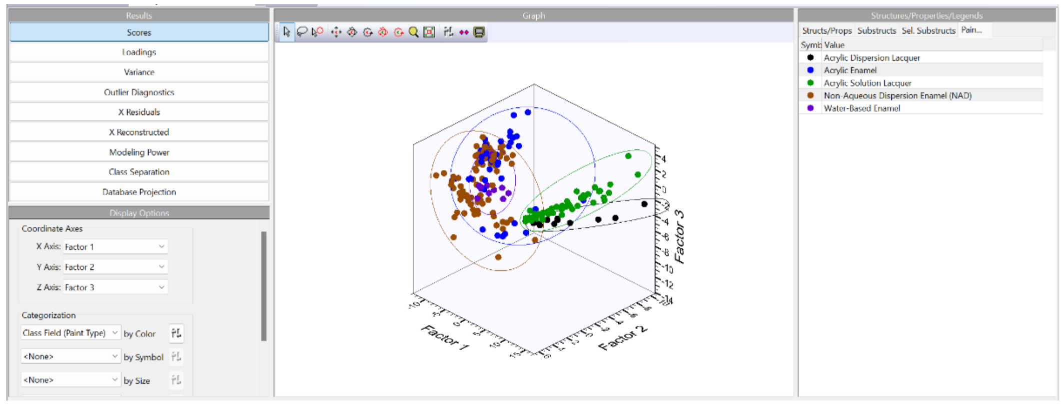
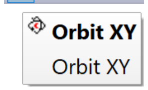
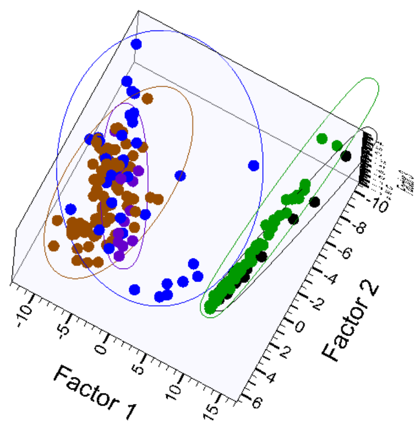
	Action	Result
4	Click **Next** > again.	

WILEY

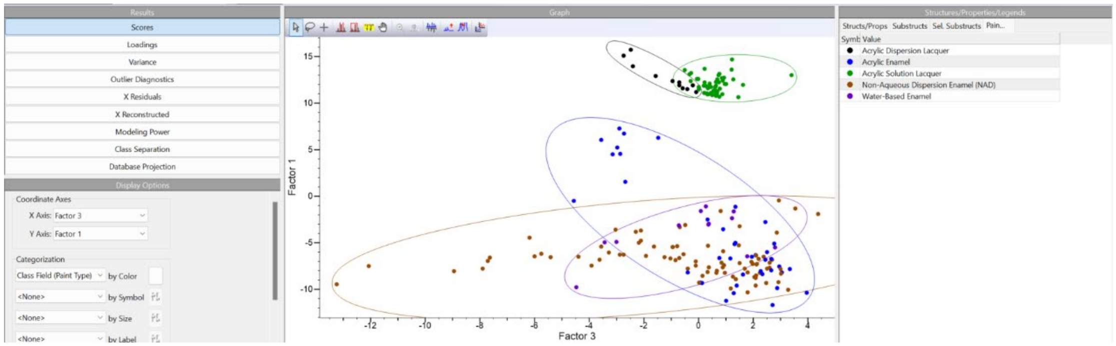
	Action	Result
5	In this dialog: • Move **IR Spectrum** from the left box to the right one by selecting it in the left and using “>>” or double-clicking to move it to the right • Set **Field for Classification** to **Paint Type.** Click **Next >.**	

WILEY

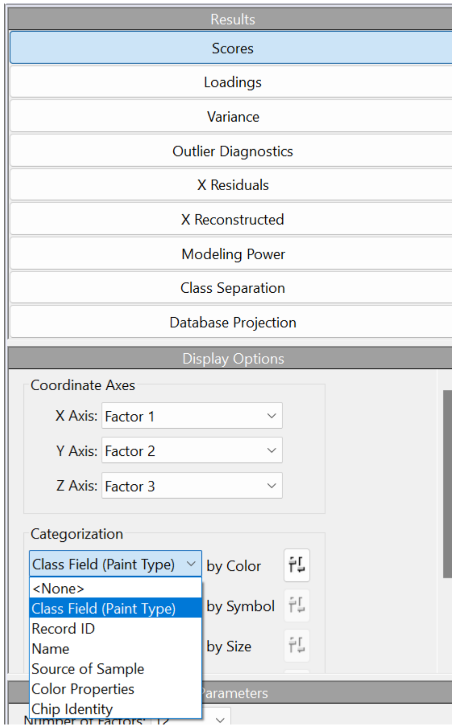
	Action	Result
6	In this dialog: • Move **2nd Derivative** method to the right box to perform **Y Transforms** on the IR spectrum. • Set the **Number of Points** on each side of an IR peak to be 11 (picture below as an example).  • Click **Next.** *Note: 1st derivative pinpoints an apex, 2nd derivative nails down peak shoulders.*	

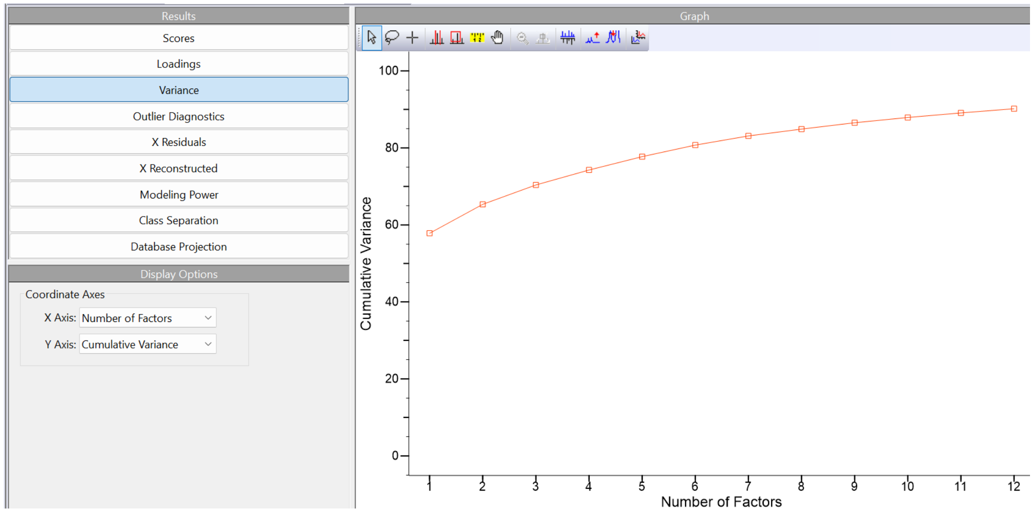
	Action	Result
7	Click **Finish** when the analysis is done. *Note: Factors are mathematical axis' orthogonal to each other.*	 This is a 3 factors view of the analysis, showing the distribution of different type of paint. A circle (**Trajectory**) defines the boundary for a value in the **Field of Classiciation**.
8	 • Click  • Rotate and observe the spectra distribution in 3 factor space.	

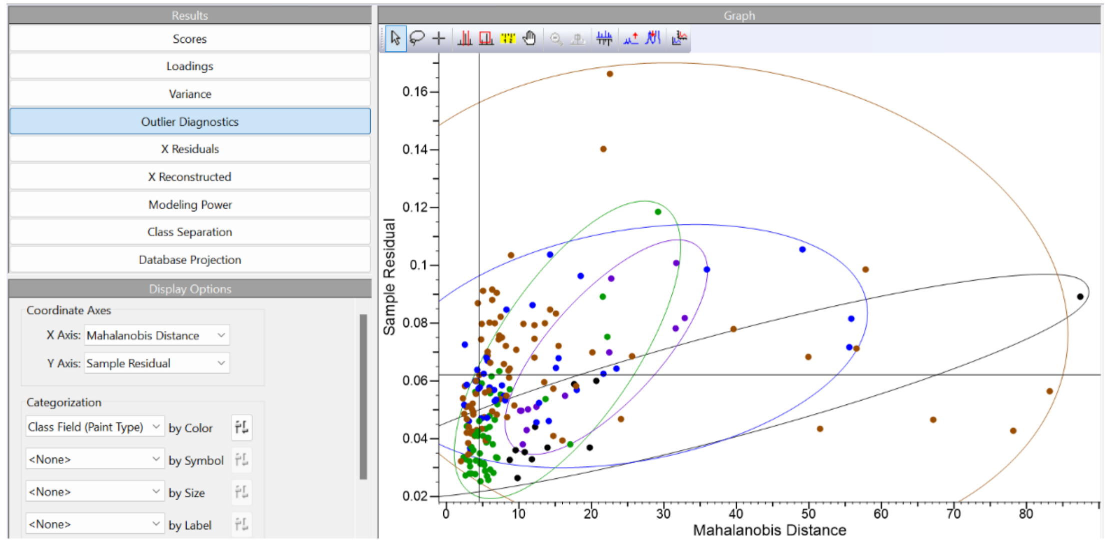
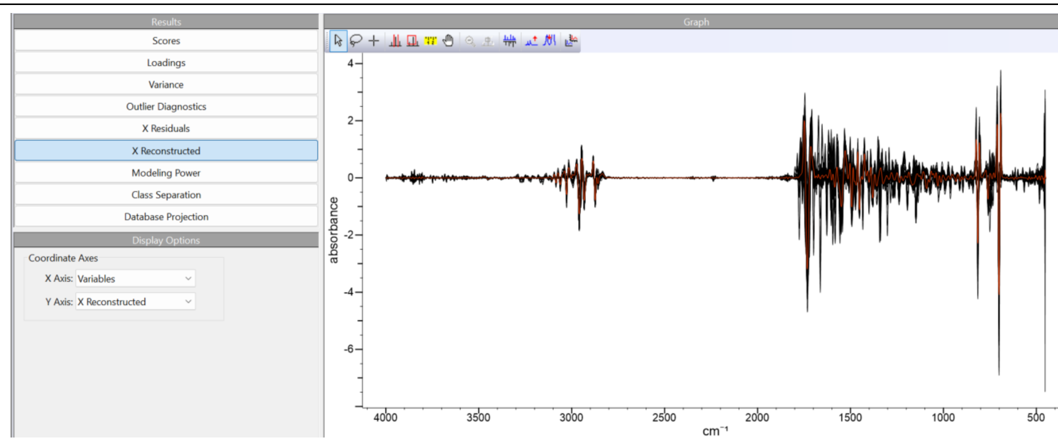
WILEY

	Action	Result
9	• Click the **2D Plot** toolbar button to switch the graph display in 2 dimensions. • Set the **X-Axis** to be **Factor 3** and the **Y-Axis** to be **Factor 1.**	 The lacquer types (upper) are separated from enamels (lower).

WILEY

	Action	Result
10	The **Categorization** dropdown list allows you to see how other properties correlate to the spectral space.	

	Action	Result
11	**Variance** allows you to examine the PCA quality.	 As shown above, 6 factors can explain over 80% of the case.

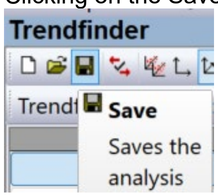
	Action	Result
12	Select **Outlier Diagnostics** to examine potential outliers. ***Manhalanobis Distance*** *is the most common measure to determine if a sample is an outlier.*	 The right most black dot (spectrum sample) is an obvious outlier of the group – it is deviating almost 100% from the center of distribution.
13	**X Reconstructed** should look like the illustration to the right.	 *Note: The above is a "reconstructed" 2nd derivative form of Y.*

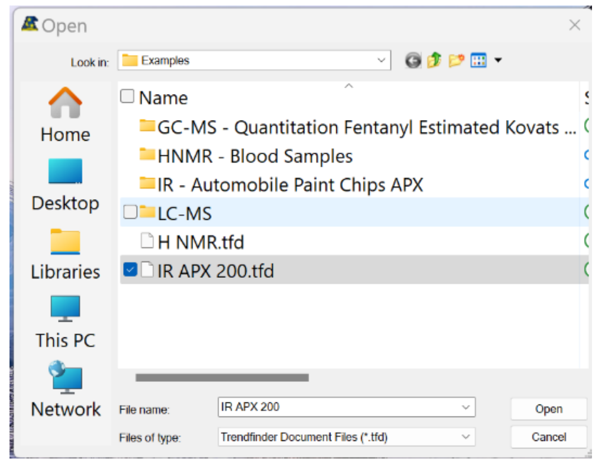
WILEY

B. Save the Analysis and Create a Report

<table>
<tr><th></th><th>Action</th><th>Result</th></tr>
<tr>
<td>1</td>
<td>
Save the analysis by:

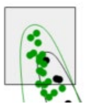
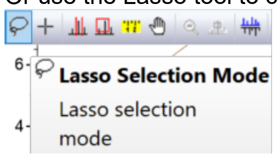
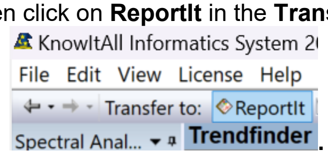
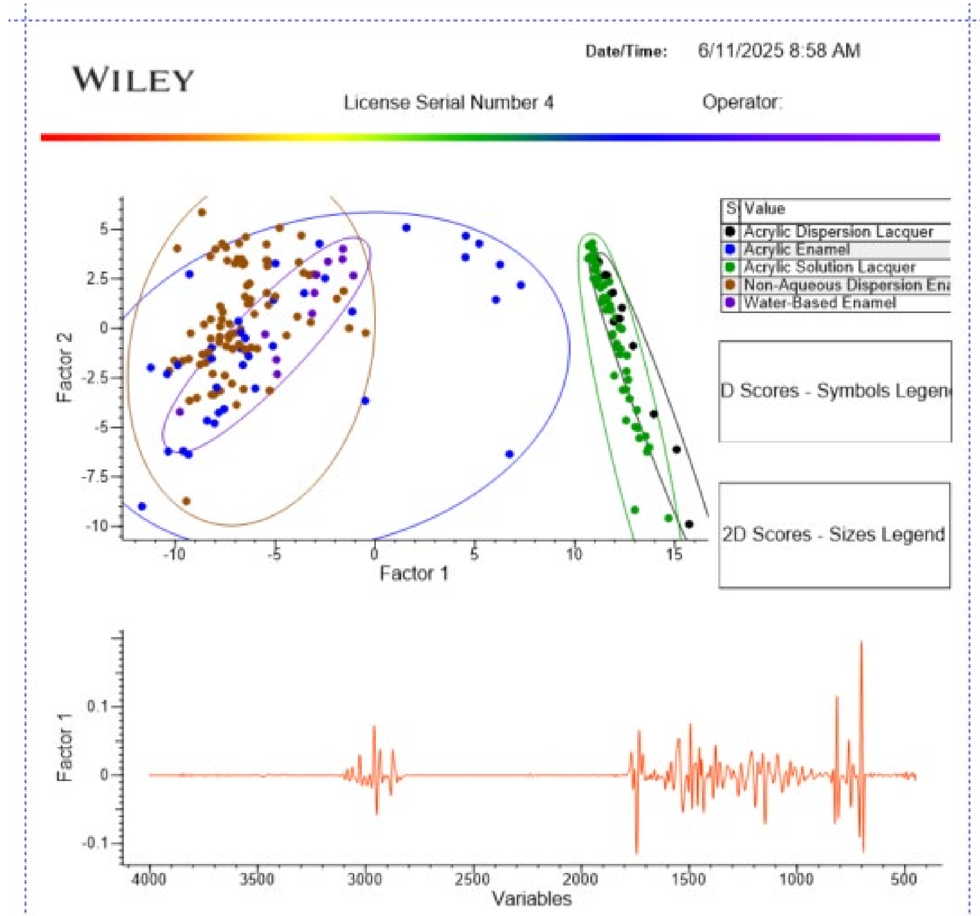
- Clicking on the Save symbol



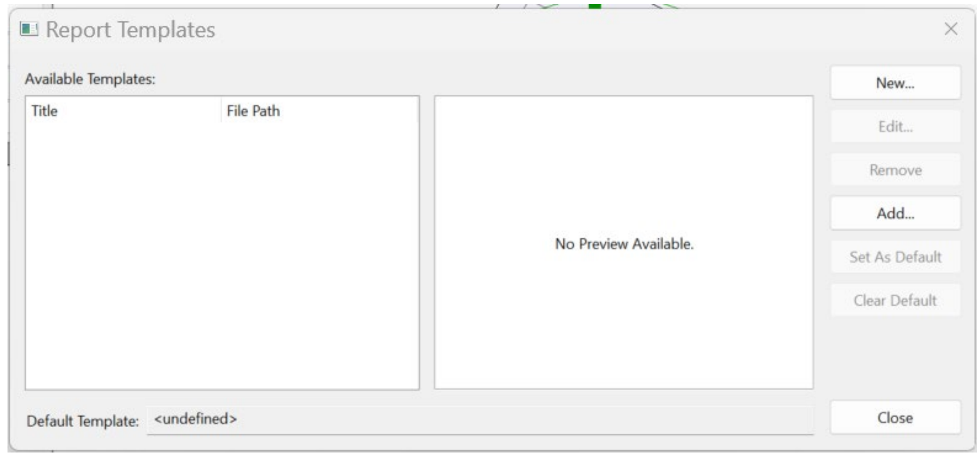
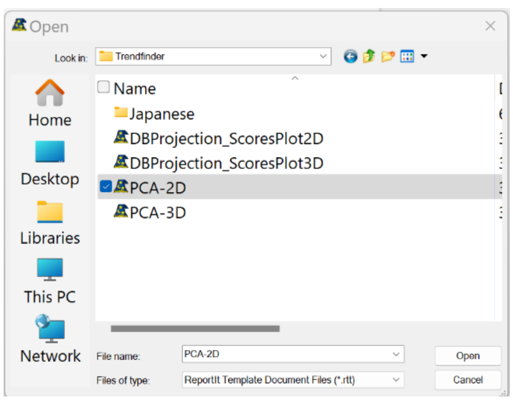
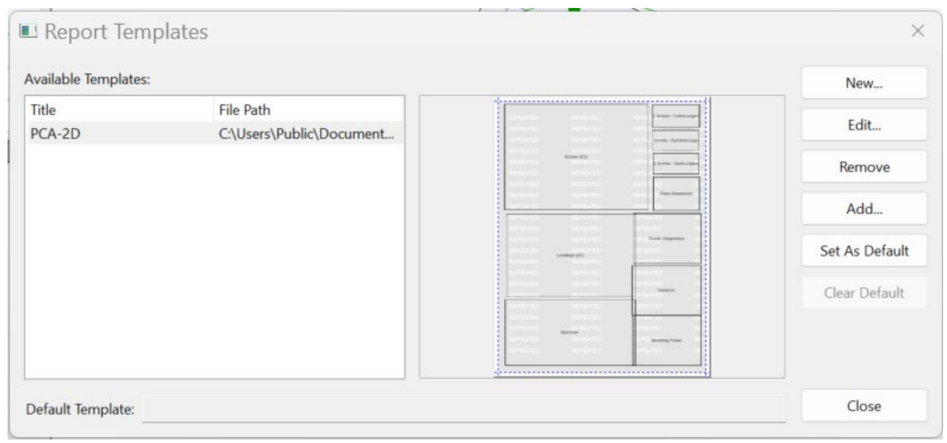
- Or by going to File>Save Analysis.
- Or **Control + S**.
</td>
<td>


</td>
</tr>
</table>

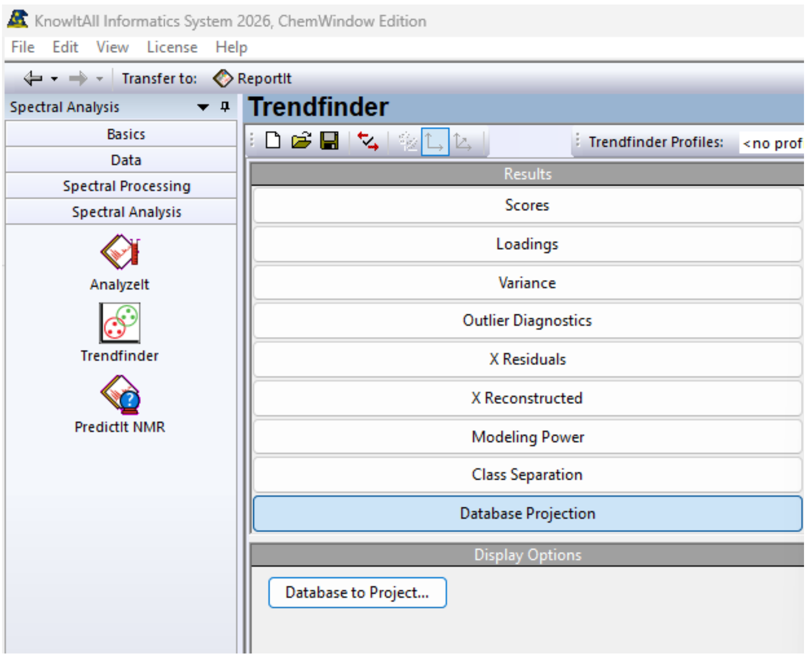
WILEY

	Action	Result
2	Transfer an analysis to **ReportIt by:** • Selecting a group of spectra and: ○ using the mouse to click a position and drag drop to another position.  ○ Or use the Lasso tool to circle points.  ○ Or **Control + Click** points. • Then click on **ReportIt** in the **Transfer to** bar .	

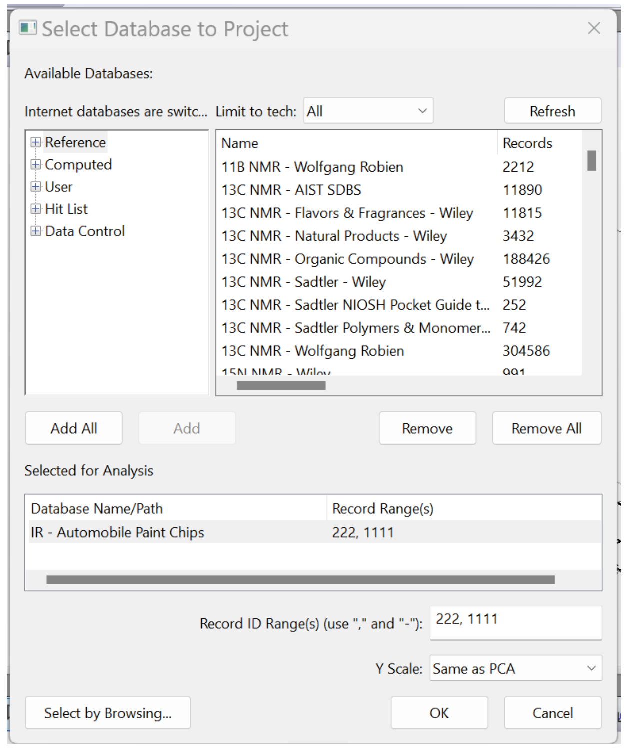
WILEY

	Action	Result
Note:	• If a template is used for the first time, the user has to do the following before transfering data to the **ReportIt** application: **File > Edit Report Templates** • Click **Add button** • Navigate to the template file **C:Users/Public/Public Documents/Wiley/KnowItAll/Report Templates/Trendfinder** to see the options. • Click **Open** to add.	
3	Add as many templates as desired and click **Close** when finished. 	

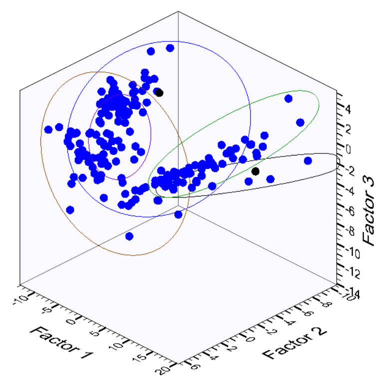
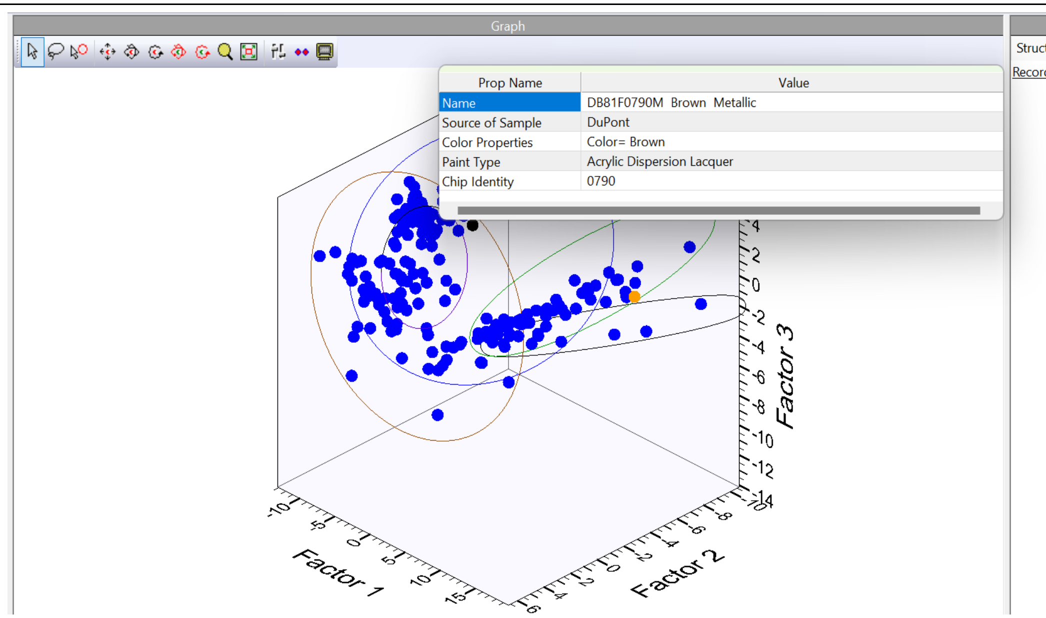
C. Project Unknown Spectra to the PCA "Space"

	Action	Result
1	Using the same data as before, click the **Database Projection** button under the **Results** section.	

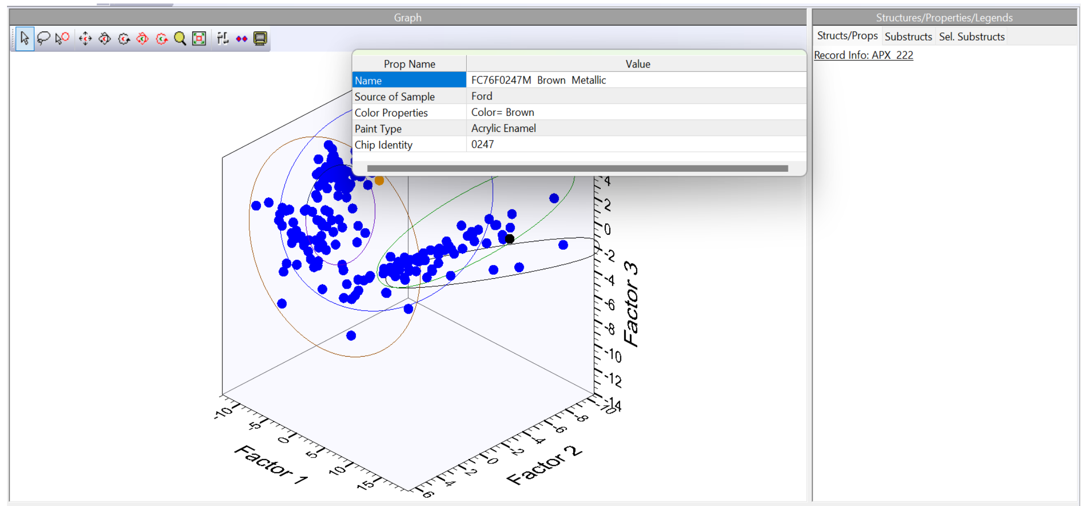
WILEY

	Action	Result
2	Ensure the IR-Automobile Paint Chips database is selected and then: • Type in 222 and 1111 under “**Record Range(s)** (use “,” and “**-**“) to test the classification of these ununsed spectra based on the first 200 spectra. • Set **Y Scale: Same As PCA.** • Click **OK.**	

WILEY

	Action	Result
3	See how two black dots fit into the PCA space (now blue).	
4	• Click one of the dots to see details for that record. (The selected dot will turn orange.) • Hover over the **Record Info** link to see more details.	 This dot is within the Acrylic Dispersion Lacquer circle .

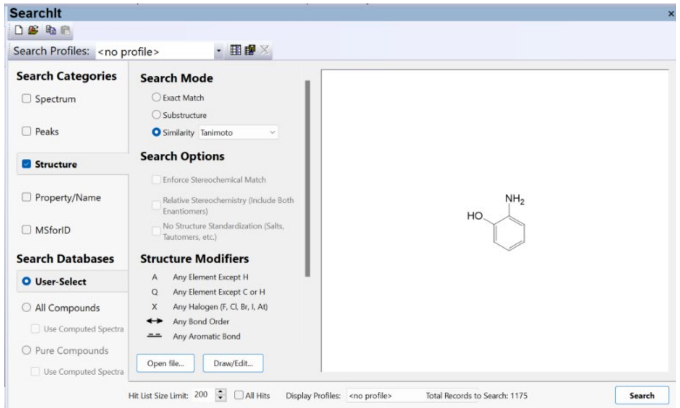
WILEY

	Action	Result
5	Click on the other one dot and hover over the **Record Info** to see more details.	 This dot is within the Acrylic Enamel circle.

WILEY

LC-MS

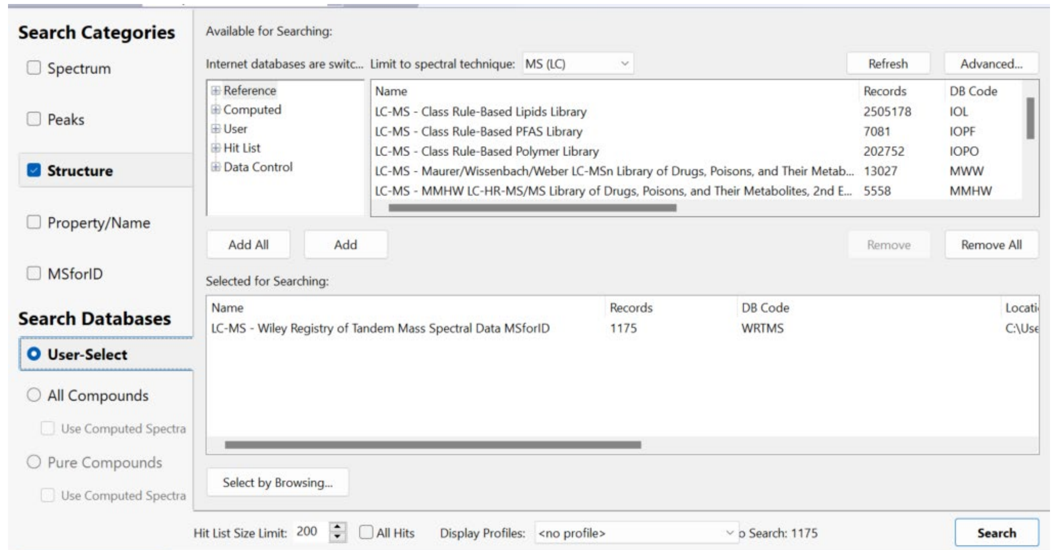
Hit List Analysis of Similar Structures

	Action	Result
1	In **SearchIt,** set a **similarity** structure search as shown in the Result column.	

WILEY

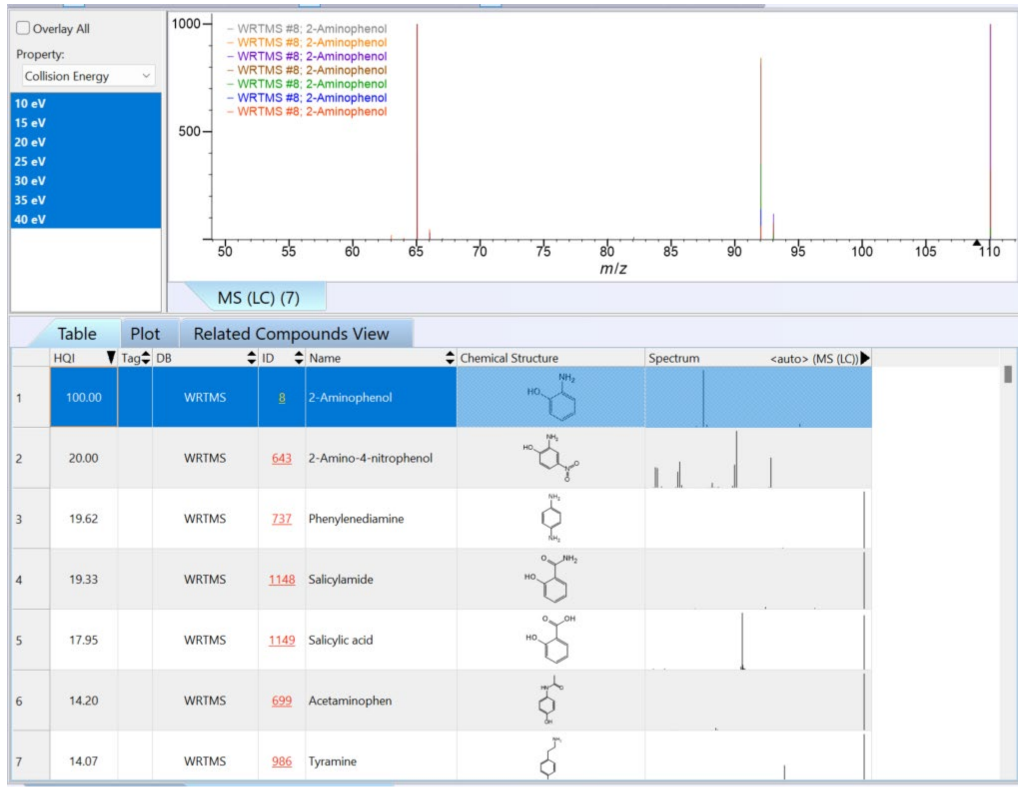
<table>
<tr><th></th><th>Action</th><th>Result</th></tr>
<tr><td>2</td><td>

Go to Search Databases, User-Select.
Set Limit to spectral technique to MS (LC).
Select the database LC-MS – Wiley Registry of Tandem Mass Spectral Data MSforID, click Add.
Set Hit List Size Limit: 200.
Click Search.

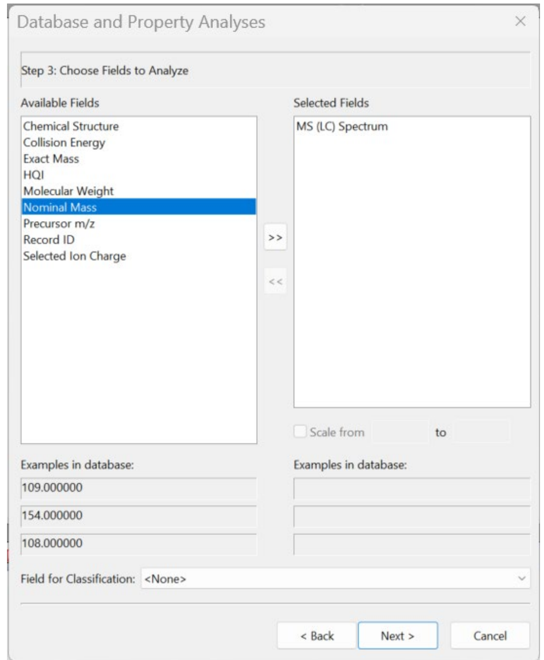
</td><td>


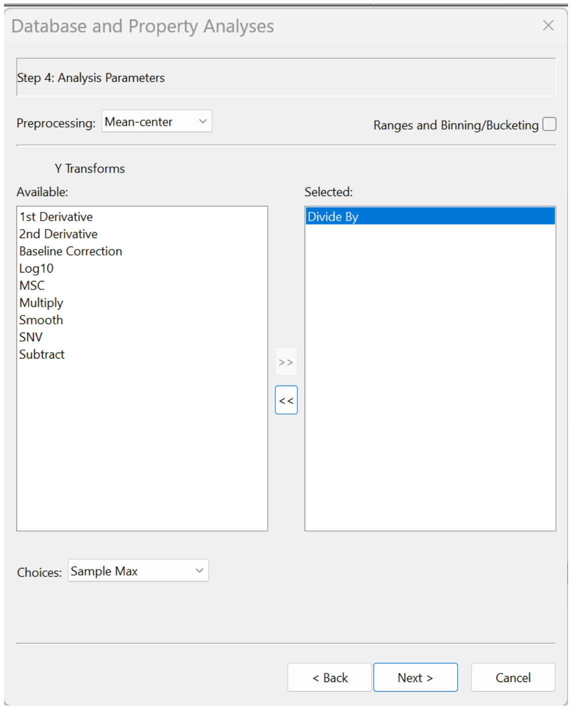
</td></tr>
</table>

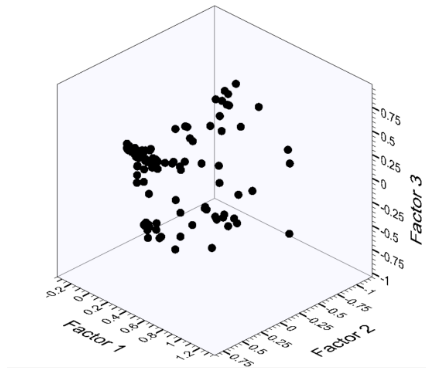
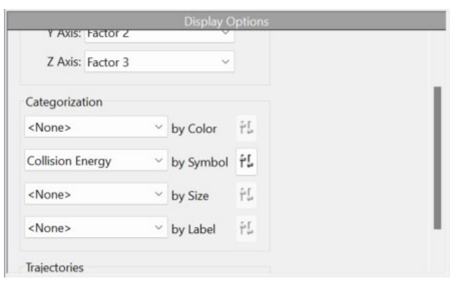
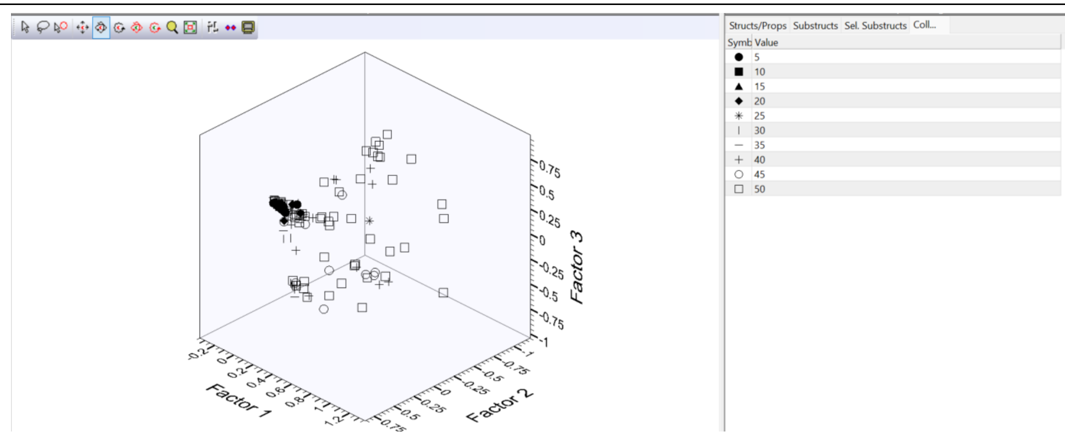
WILEY

	Action	Result
4	Choose **Collision Energy** in the **Property** drop down list and select all.	 This database contains records with MS spectra acquired at different collision energies.
4	• Transfer this hit list to Trendfinder using: Transfer to: Trendfinder ***Note:*** *To start the Trendfinder application without transferring a hit list, open it directly by clicking its icon in the Spectral Analysis toolbox.* • Click **Next >.**	

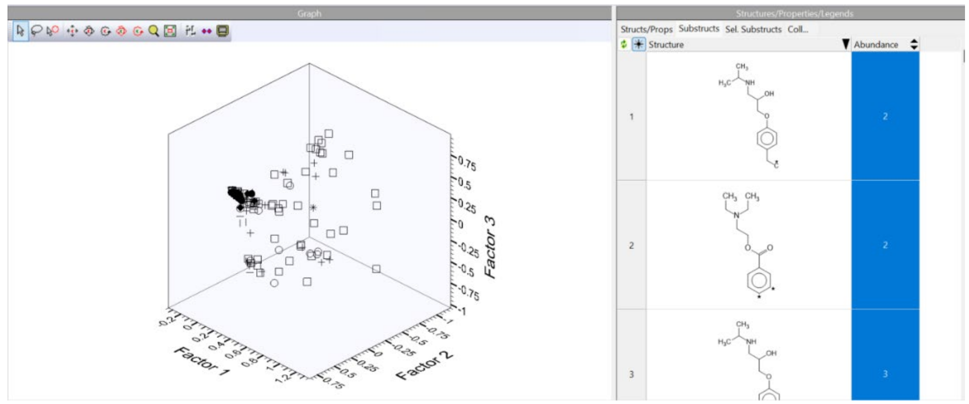
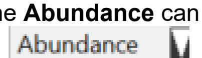
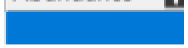
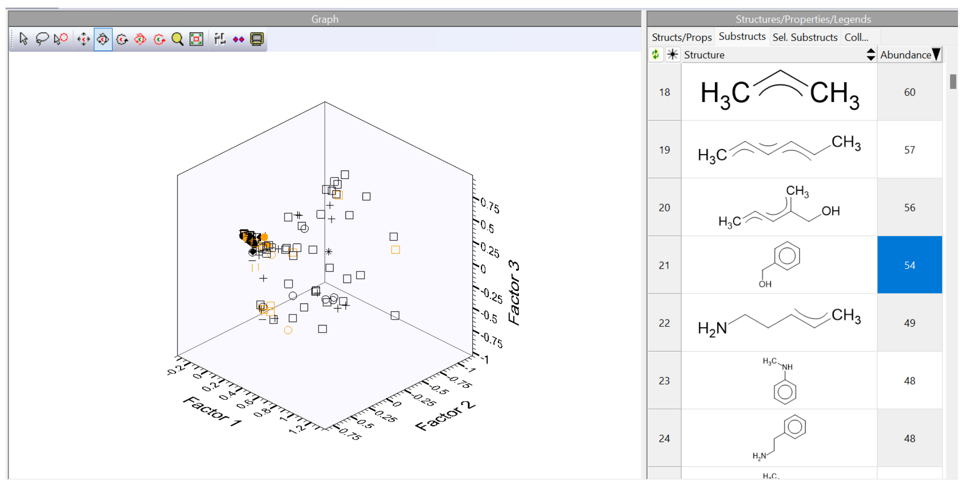
WILEY

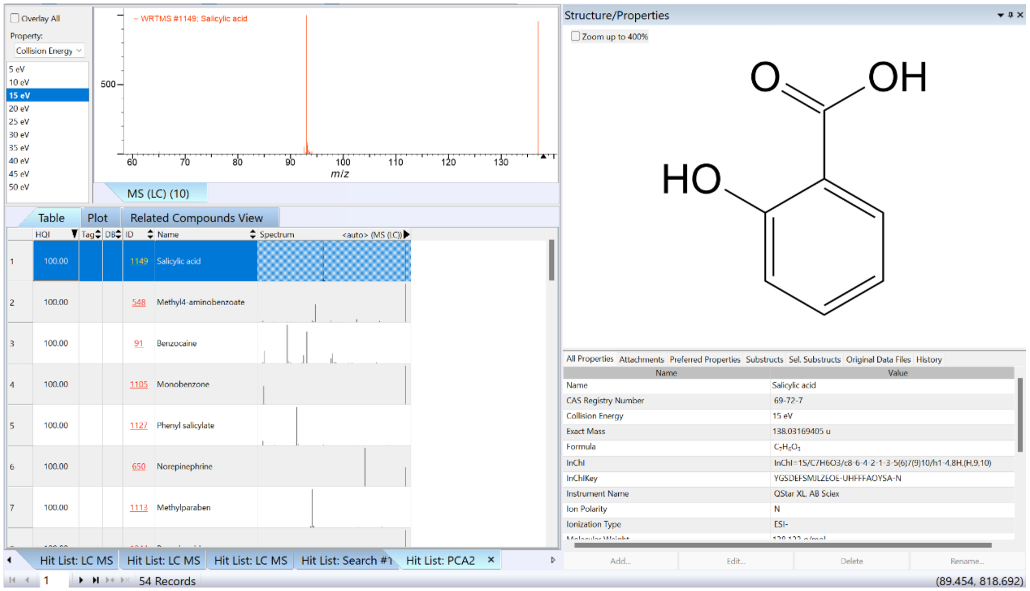
	Action	Result
5	• Select **MS (LC) Spectrum** to analyze. • Click **Next >.**	
	Tip: Based on experience, selecting **Divide By from the Available list** in the **Y Transform and Sample Max** from the Choices drop down is effective for MS spectra.	

	Action	Result
6	Click **Next >** to continue.	

	Action	Result
7	Click **Finish.**	 **Scores** plot opens as above.
8	Under **Categorization**, select to show **Collision Energy by Symbol**: 	 This plot is an interesting way of looking at this set of data. High collision energy breaks molecules more thoroughly resulting in differences in their spectra!
9	Select a particular spot (*)	

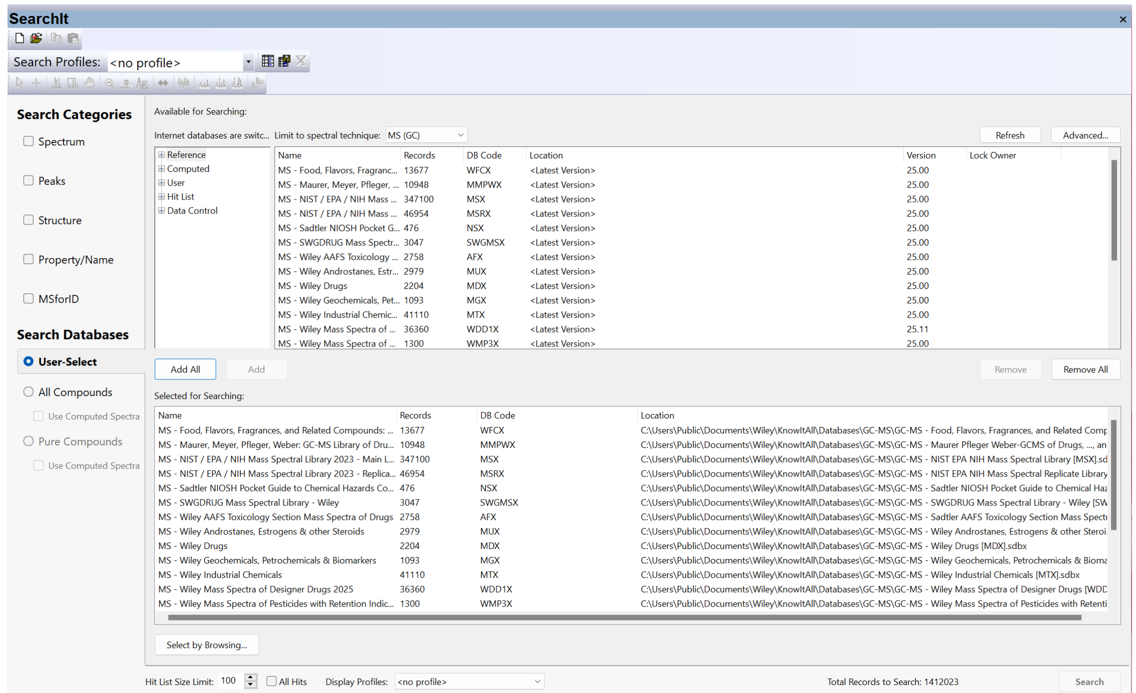
Common Fragments

	Action	Result
1	Continue with the above example. • Open the **Substructs** tab under the **Structures/Properties/Legends** pane. • Click **Start Calculations.**	 The fragments and abundances (frequency to appear) within this entire dataset are shown in the right table.
2	• The **Abundance** can be sorted by clicking  on  • Select a fragment.	 25 Spectra containing this fragment are highlighted in the **Scores** plot. The **Sel. Substructures** tab will allow the user to select a group of dots and see the fragment abundance within that group.

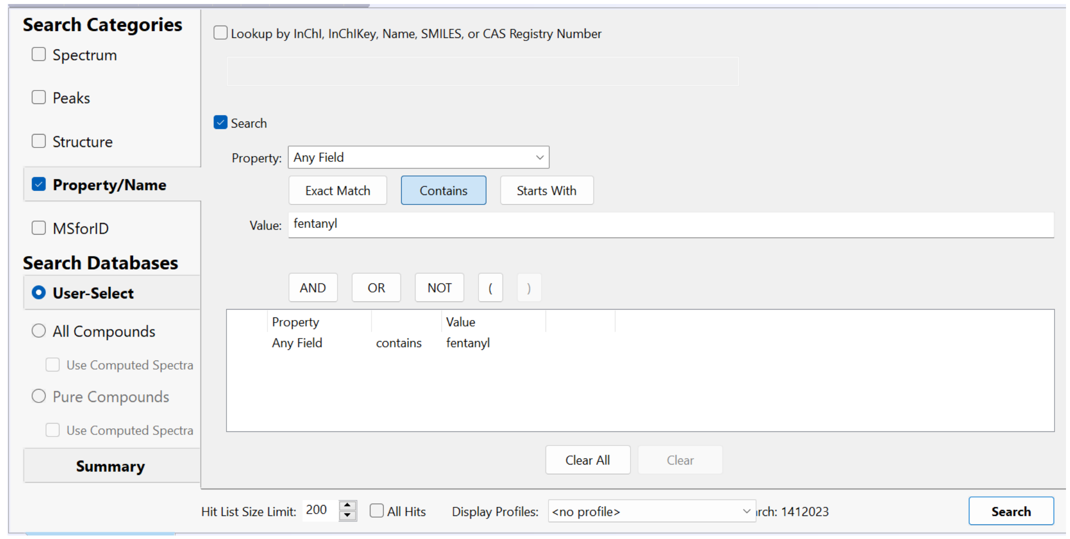
	Action	Result
4	• Transfer the highlighted collection of 54 records **to MineIt** using the **Transfer to** button. • The database records will open in **MineIt.**	

GC-MS

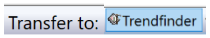
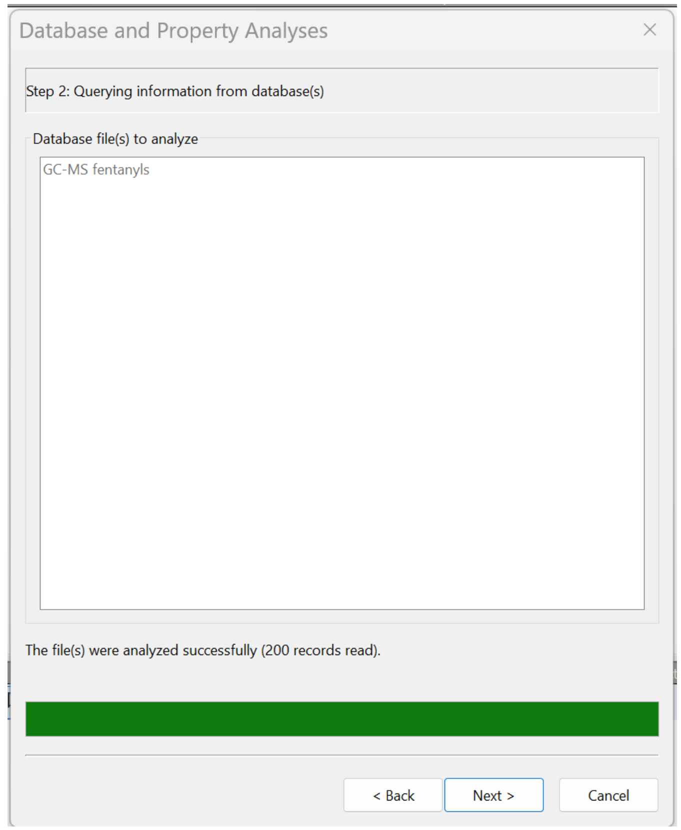
Hit List Analysis Of Fentanyls

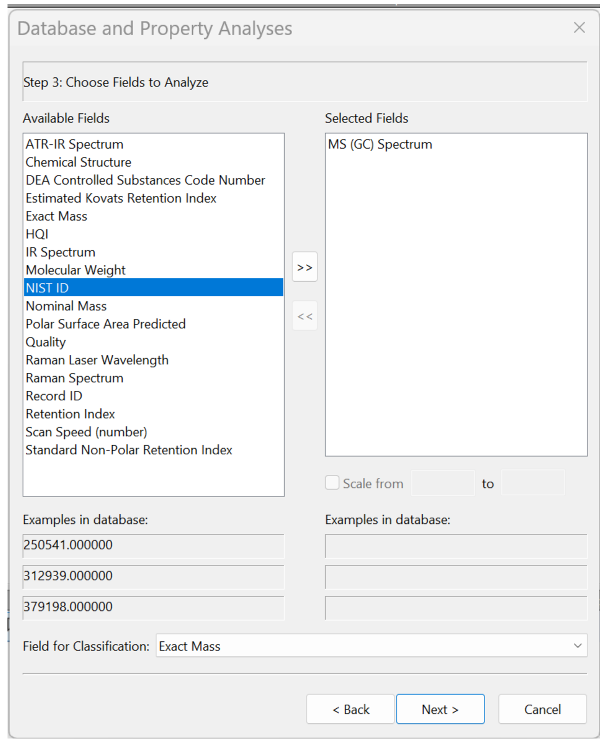
	Action	Result
1	• In SearchIt, go to **User-select.** • **Limit to spectral technique** to **MS(GC)** and • **Add All** GC-MS databases for searching.	

WILEY

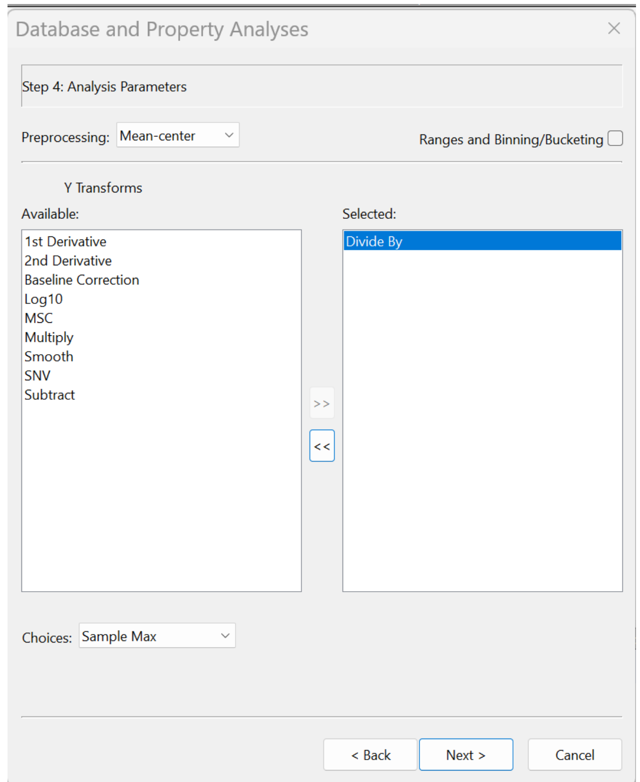
	Action	Result
2	• Set **Property/Name** to search for **Any Field** contains "**fentanyl**". • Set **Hit List Size Limit**: 200. • Click the **Search** button.	

WILEY

	Action	Result
3	**MineIt** provides a hit list of various **fentanyls**. • Transfer this hit list to Trendfinder.  ***Note:*** *To start the Trendfinder application without transferring a hit list, open the Trendfinder application by clicking its icon in the Spectral Analysis toolbox.* • Click **Next >** at the prompt.	

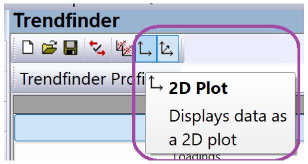
	Action	Result
4	• Move **MS (GC) Spectrum** to **Selected Fields** for analysis. • Select **Exact Mass** as **Field for Classification.** • Click **Next >**.	

WILEY

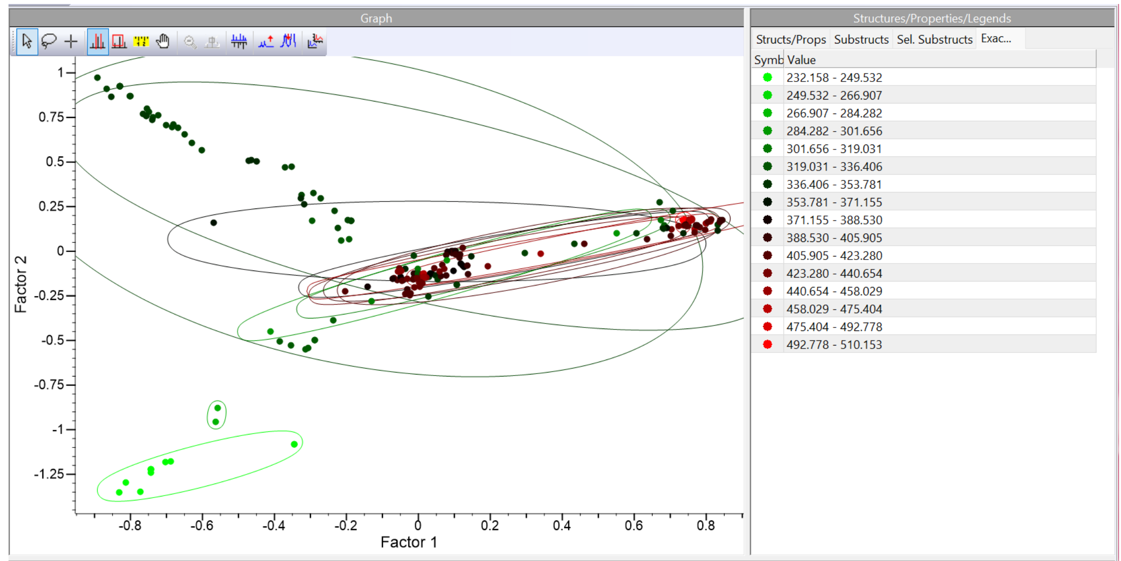
	Action	Result
5	• Choose **Divide By** in **Y Transform** and **Sample Max** in **Choices** to obtain the best results for MS spectra. • Click **Next >.**	

<table>
<tr><th></th><th>Action</th><th>Result</th></tr>
<tr><td>6</td><td>

- Click **Finish.**
- Select **2D Scores** plot view.

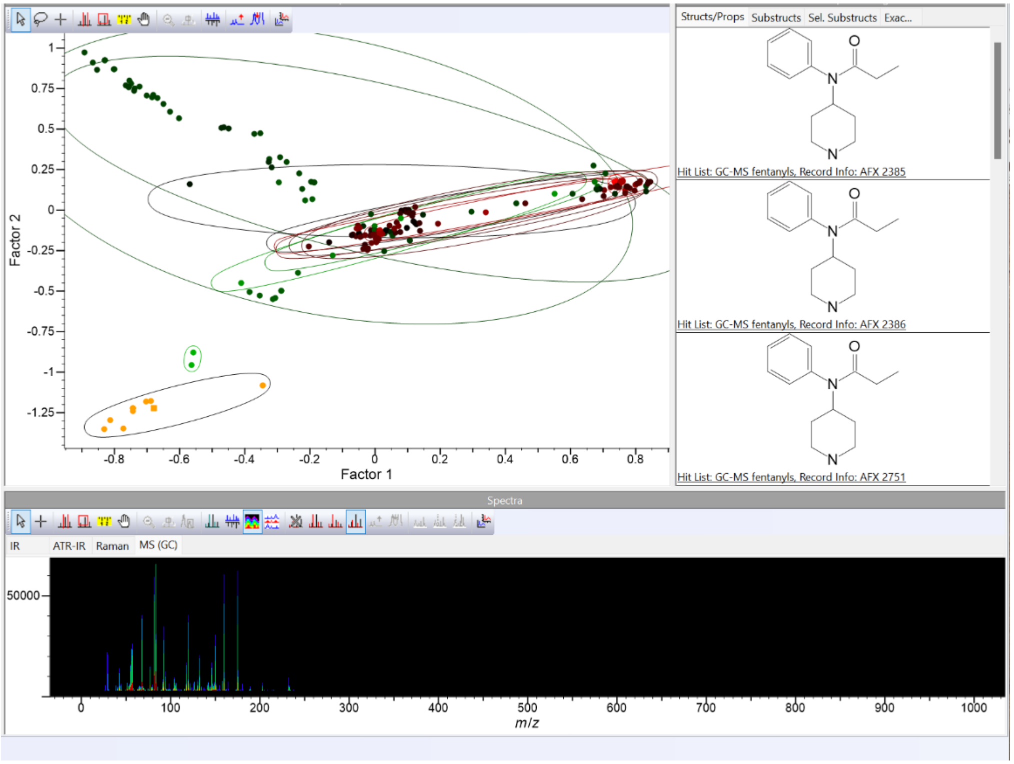


</td><td>

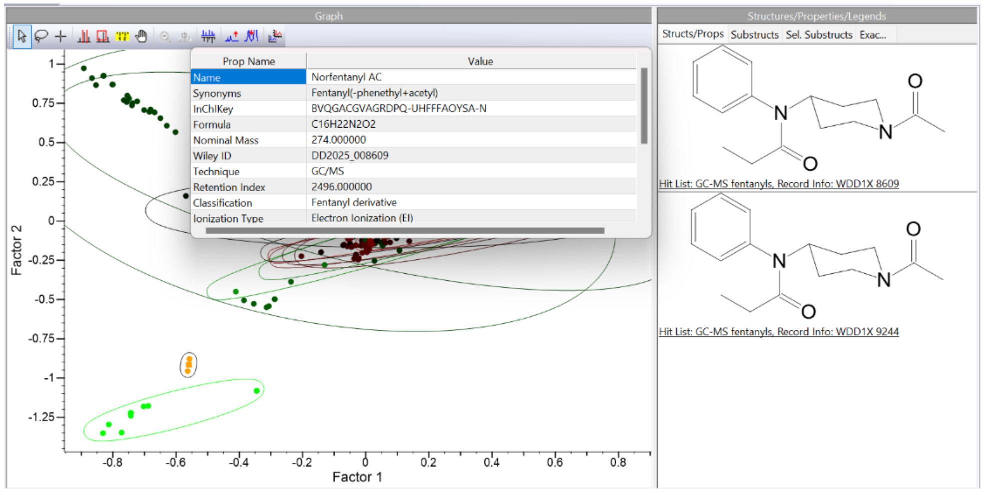
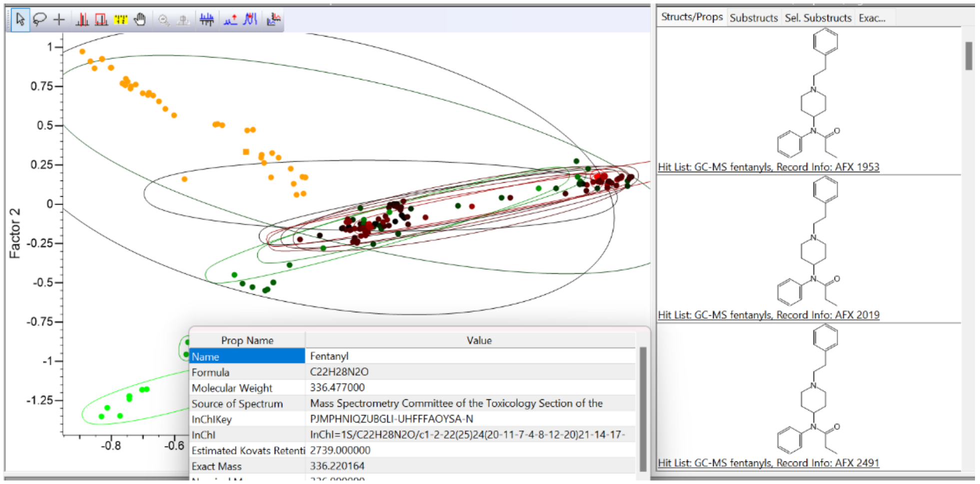


You can see a group of light green (smaller mass) spectra.
A circle (**Trajectory**) defines the boundary for a value in the **Field of Classification.**

</td></tr>
</table>

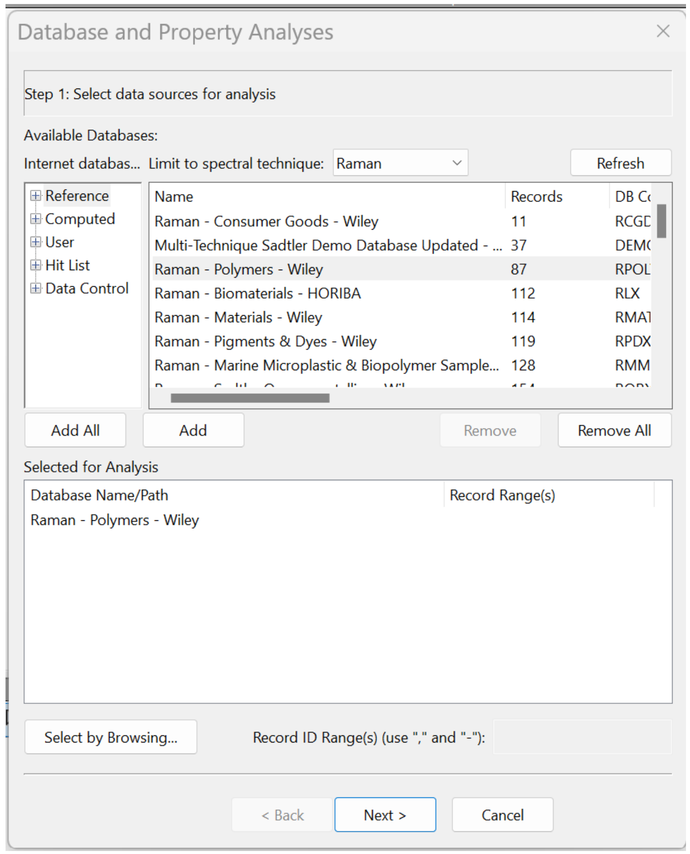
	Action	Result
7	• Use **left mouse** to draw a rectangle so that the green spectra in the lower left corner are selected and turn orange. • Go to the **Structs/Props** tab under the **Structures/Properties/Legends** pane. • Click the **MS (GC)** tab in the bottom pane to open the spectra.	 These are varous **norfentanyl** MS spectra. The bottom pane shows the selected MS spectral **Overlay Heatmap**.

WILEY

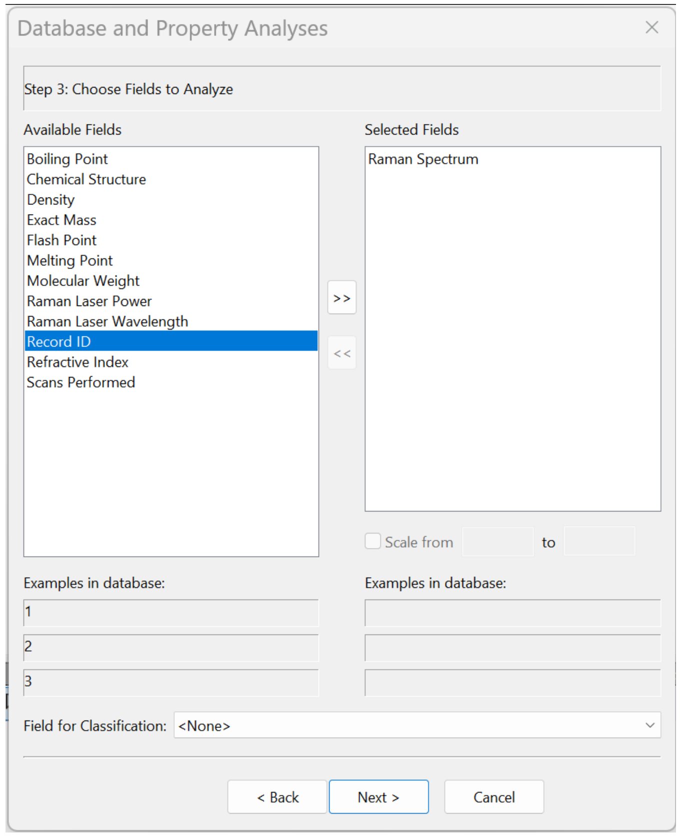
	Action	Result
8	Selecting the two darker group dots groups the norfentanyl AC spectra.	
9	The line of green linear dots from top left to center are fentanyls	

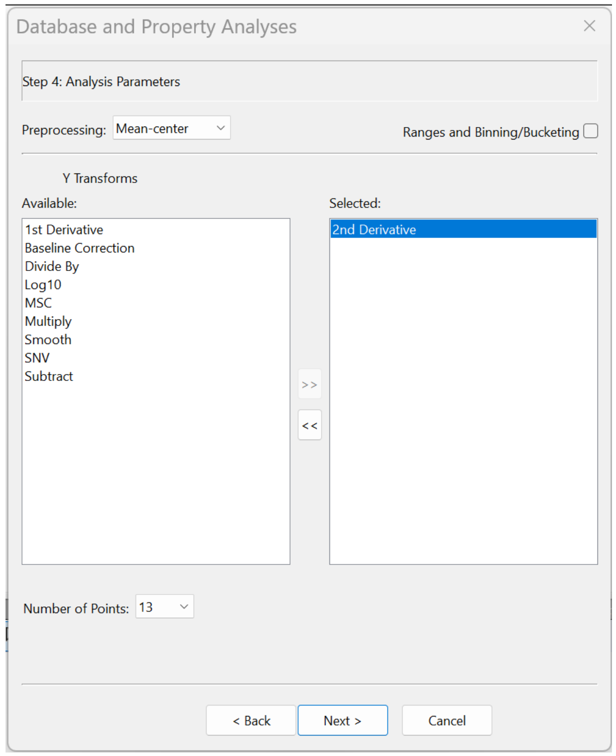
Raman

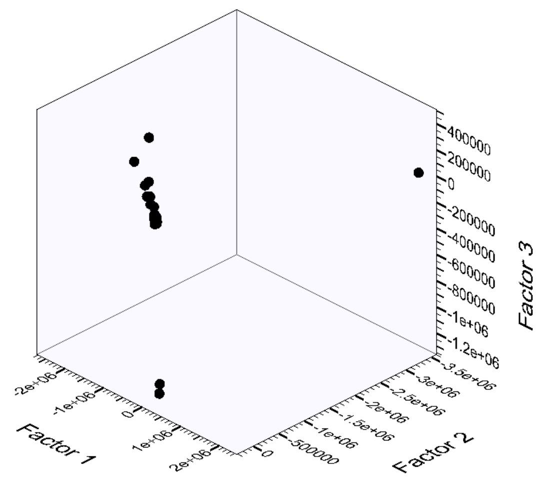
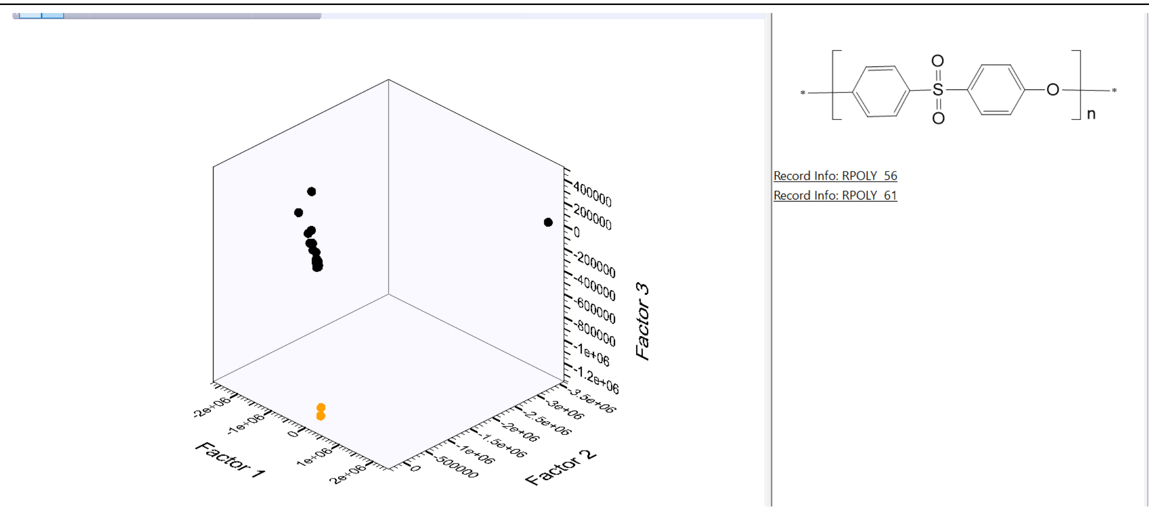
Polymer Analysis

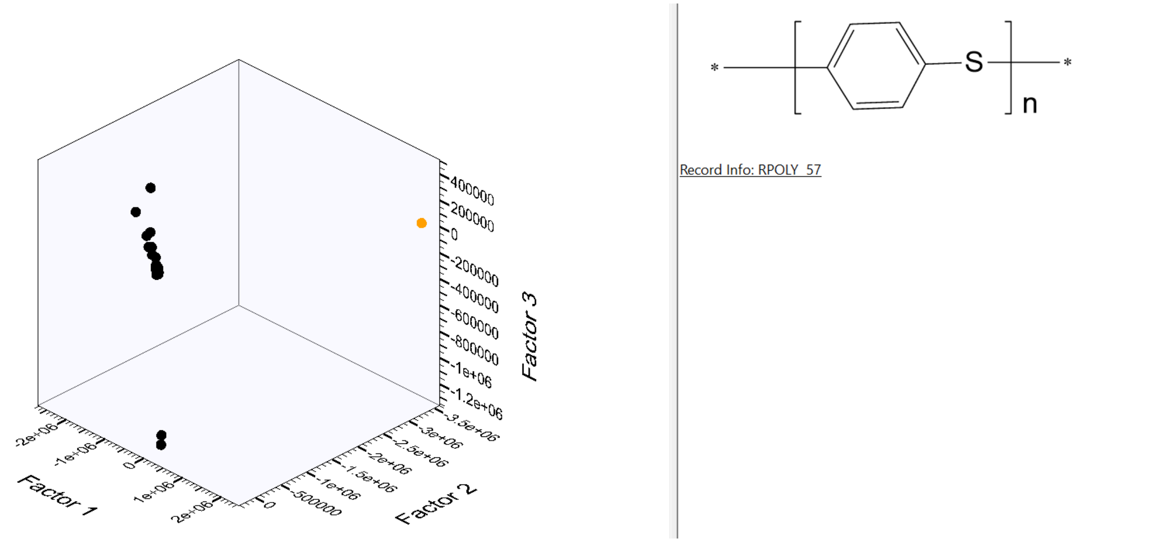
<table>
<tr><th></th><th>Action</th><th>Result</th></tr>
<tr><td>1</td><td>Open the Trendfinder application by clicking its icon in the Spectral Analysis toolbox.</td><td>The application opens.</td></tr>
<tr><td>2</td><td>Limit to spectral technique should be set to Raman.
Select the database Raman – Polymers – Wiley for analysis and click Add.
Click Next >, Next >.</td><td>
</td></tr>
</table>

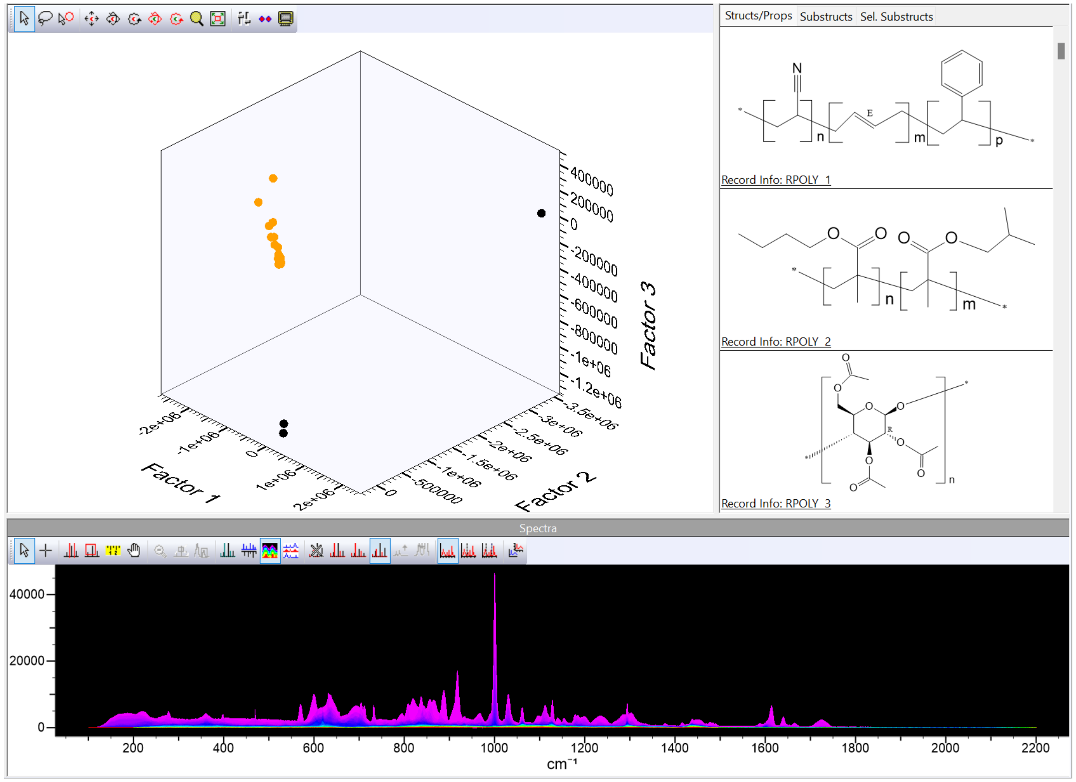
WILEY

	Action	Result
3	• Move **Raman Spectrum** to the right either by: • Clicking **>>**, or • Double clicking on **Raman Spectrum**. • Click **Next >.**	

	Action	Result
4	• Select **2nd Derivative** in the **Y Transform Available** list. • **Set Number of Points** to **13.**	
5	Click **Next >.**	

	Action	Result
6	Click **Finish.**	 3 big groups are clearly identified.
7	Use the **left mouse** to select a group of points.	

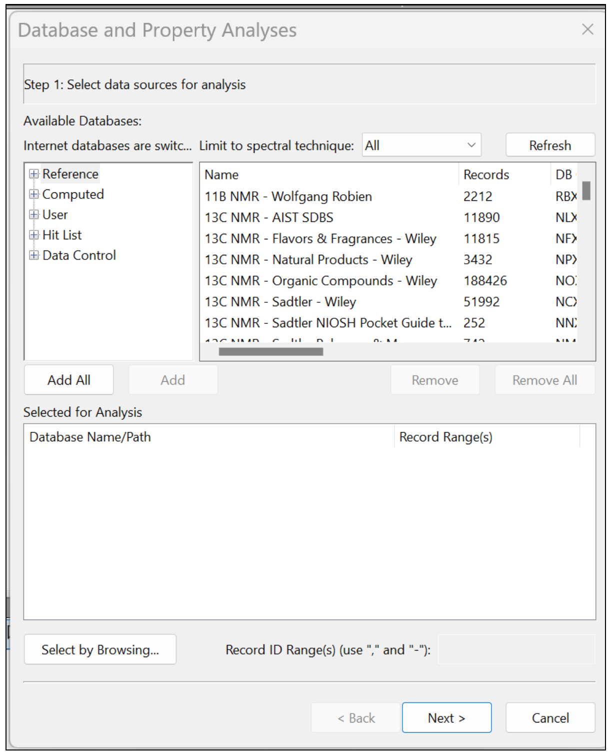
	Action	Result
8	Move to another point to examine a different polymer.	

	Action	Result
9	Select another group of points by clicking and holding the left mouse button examine a different polymer.	 They are quite different polymers. The bottom pane shows the **Overlay Heatmap** of selected spectra.

UV-Vis

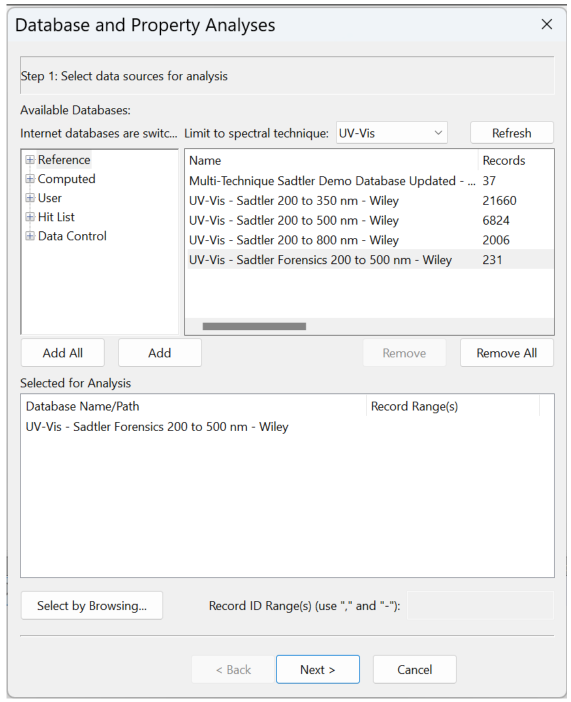
Forensic Material Analysis

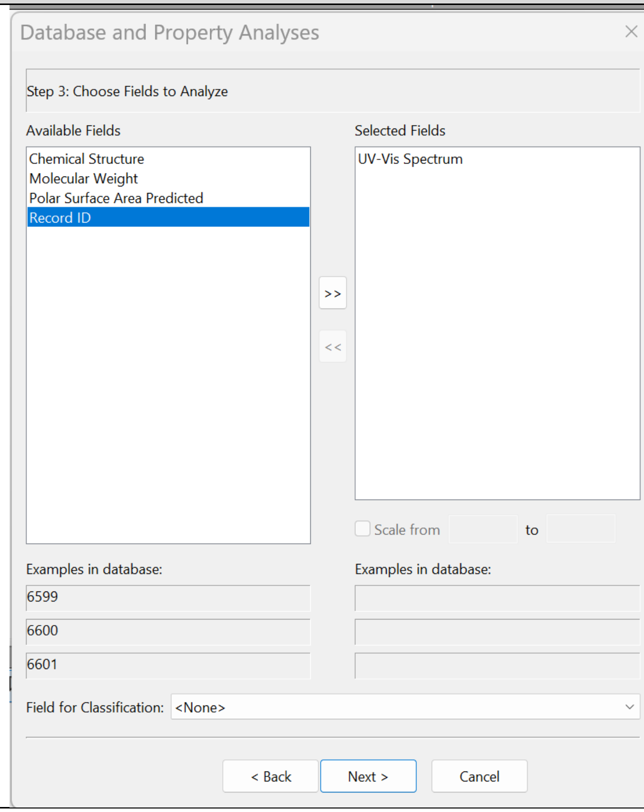
<table>
<tr><th></th><th>Action</th><th>Result</th></tr>
<tr><td>1</td><td>Open the Trendfinder application by clicking its icon in the Spectral Analysis toolbox.</td><td>The application opens.</td></tr>
<tr><td>2</td><td>Choose File > New PCA Analysis.</td><td>The Database and Property Analyses wizard opens



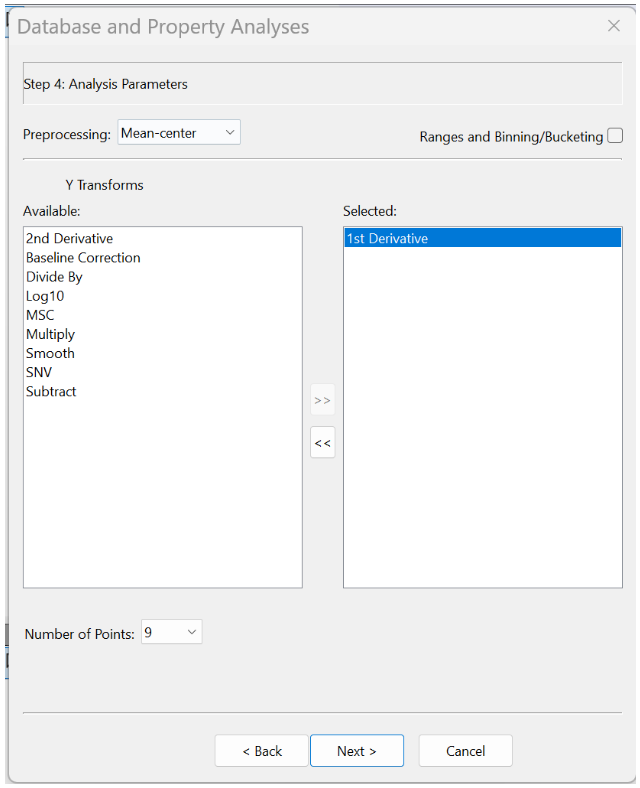
.</td></tr>
</table>

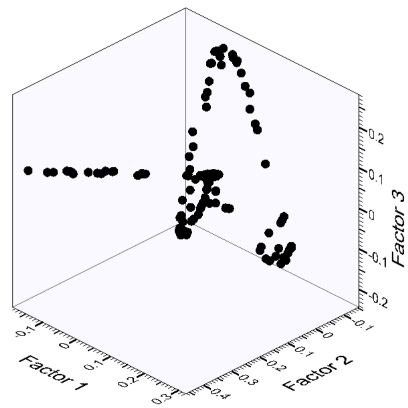
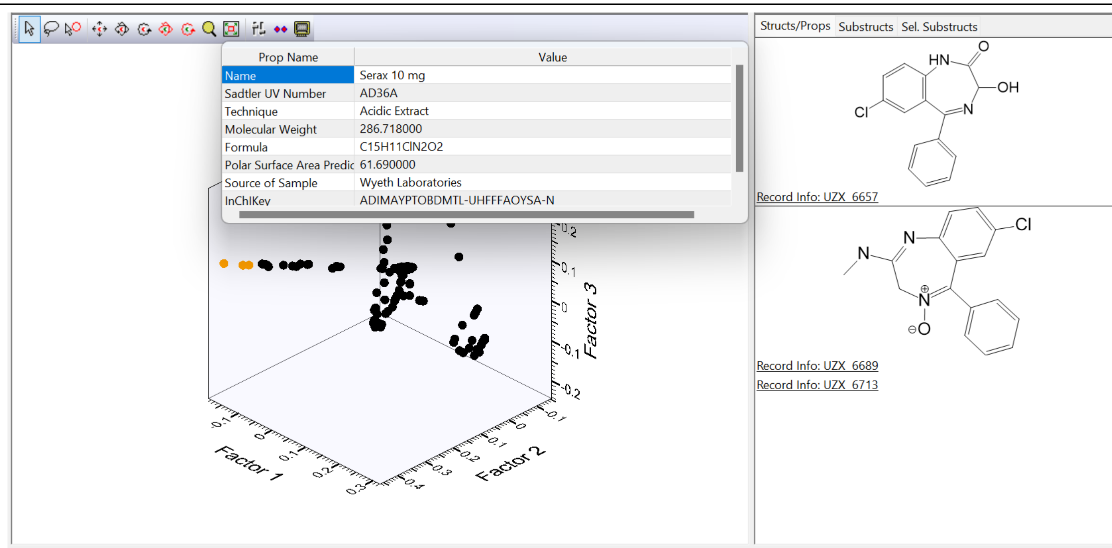
WILEY

	Action	Result
3	• Select the **UV-Vis** database **UV-Vis – Sadtler Forensics 200 to 500 nm – Wiley** for analysis and click **Add.** • Click **Next >, Next >.**	

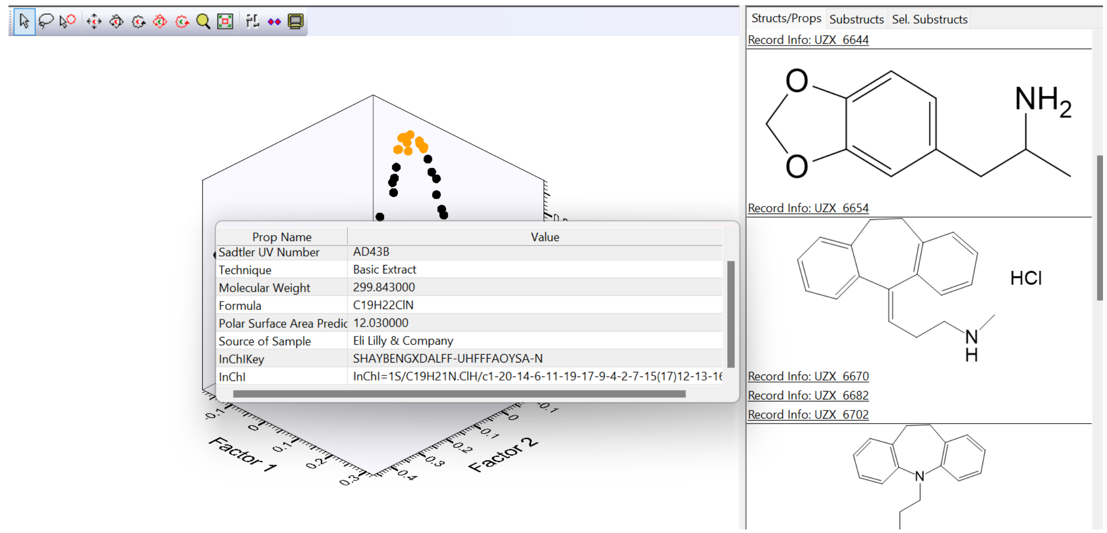
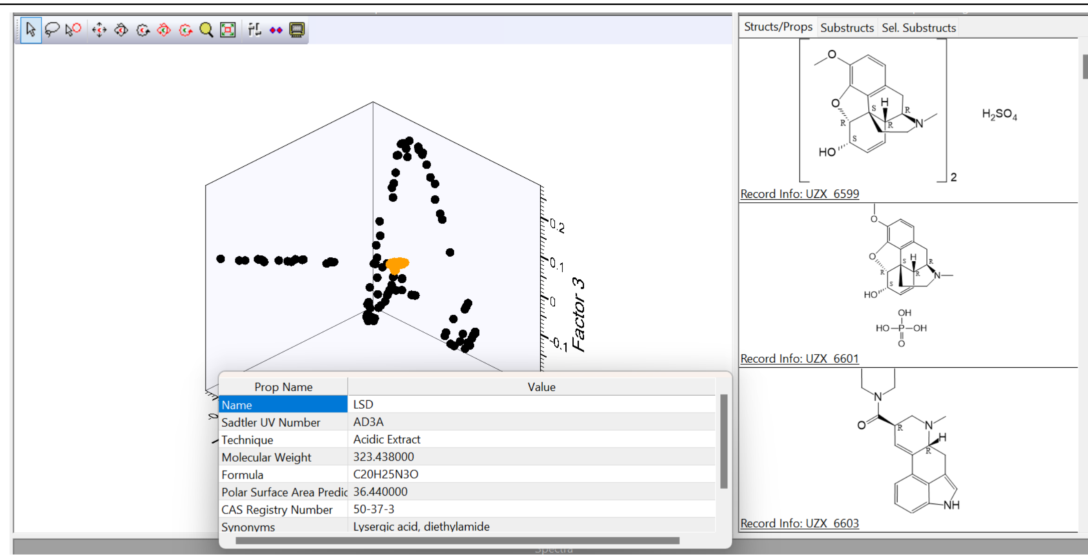
	Action	Result
4	In this dialog: • Move **UV-Vis Spectrum** from the left box to the right one by selecting it in the left and using “>>” or double-clicking to move it to the right. • Click **Next >.**	

WILEY

	Action	Result
5	In this dialog: • Move **1st Derivative** method to the right box to perform **Y Transforms** on the IR spectrum. • Set the **Number of Points** on each side of an IR peak to be 9. • Click **Next.** *Note: For UV-Vis, 1st derivative, which locates the peak top position is sufficient.*	

	Action	Result
6	Click **Finish.**	 The above 3 factors view is returned.
7	Use the **left mouse** to select box each group of points, which turn orange upon selection.	

WILEY

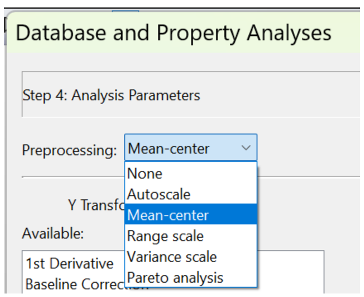
	Action	Result
8		
9		

Principal Component Analysis (PCA) Theory

Principal Component Analysis (PCA) is a process which performs dimensionality reduction. Thus, it transforms a large set of variables into a smaller set of uncorrelated variables (principal components). At the same time, it retains most of the original data's variance. This portion of training will explain how it is applied to spectral data analysis by KnowItAll.

Parameters

Preprocessing



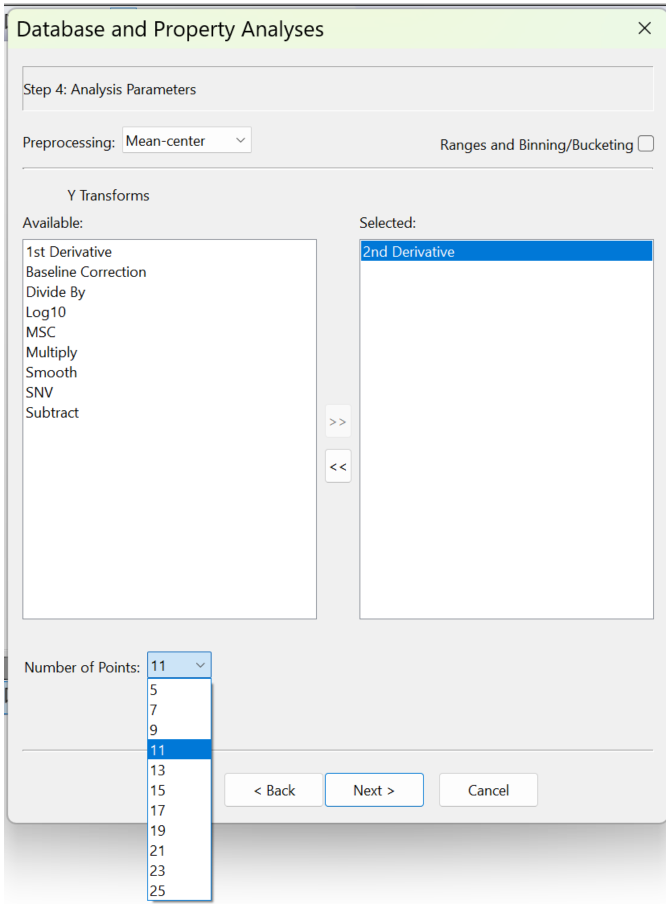
- **None:** No preprocessing occurs.
- **Autoscale**:
  In Principal Component Analysis (PCA), "autoscale" refers to a preprocessing step where each variable in the dataset is standardized. This involves two main actions:
  1. **Centering**: Subtracting the mean of each variable so that the mean of the transformed variable is zero.
  2. **Scaling**: Dividing each variable by its standard deviation so that the variance of the transformed variable is one.

  Autoscaling ensures that all variables contribute equally to the PCA, regardless of their original scales or units. This is crucial because PCA identifies directions of maximum variance, and without autoscaling, variables with larger scales could dominate the analysis.
- **Mean-center:** Centers the data relative to a reference point[1].
- **Range scale:** Min-max scaling which transforms the data to fit within a specific range[2].
- **Variance scale:** First, the variance for each valuable is calculated. Then, each variable is divided by the standard deviation.

- **Pareto analysis:** Pareto analysis is a decision-making tool based on the idea that 80% of a project's benefit can be achieved by doing 20% of the work, or conversely, 80% of problems can be traced to 20% of the causes. In other words, it posits that not all inputs have the same or even proportional impact on a given output[3].

[1] Eigenvector Research Documentation Wiki (2012) *Advanced Preprocessing: Variable Centering,* https://www.wiki.eigenvector.com/index.php?title=Advanced_Preprocessing:_Variable_Centering (accessed 2025-08-19).
[2] Geeks for Geeks (2025) *Normalization and Scaling,* Normalization and Scaling – GeeksforGeeks (accessed 2025-08-19).
[3] Kenton, W. (2025) *Pareto Analysis: Definition, How to Create a Pareto Chart, and Example,* *https://www.investopedia.com/terms/p/pareto-analysis.asp* (accessed 2025-08-19).

Y-Transforms



- **1st Derivative:** parameter “Number of points” is the number of points on each side of a bell curve (peak).
- **2nd Derivative:** parameter “Number of Points” definition like 1st derivative.
- **Baseline Correction:** to correct the baseline.
- **Divide By:** Various data matrix transformation and normalization methods.
    - **Sample 1-norm**: area normalization.

- **Sample 2-norm**: vector length normalization.
- **Sample Max:** sample max value (suitable for MS spectra).
- **Sample Range:** consider various instrument measurement differences.
- **Value at Variable:** normalize to the value at a particular variable n.

- **Log10:** Applies $\log^{10}$ to scale Y values.
- **MSC:** Multiple Scatter Variate[4].
- **SNV:** Standard Normal Variate[5].
- **Multiply:** Multiple Y by an editable numerical value.
- **Smoothing:** Uses our standard Savitzky-Golay algorithm to smooth noisy data.
- **Subtract:** The user can define the value to subtract.

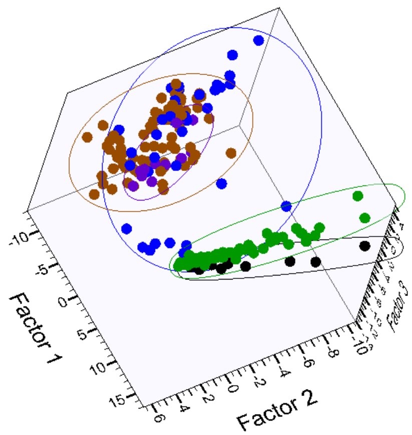
[4] Fearn, T., Riccioli, C., Garrido-Varo, A., and Guerrero-Ginel , J.E. On the geometry of SNV and MSC, *Chemometrics and Intelligent Laboratory Systems,* **2009** *96*(1), 22-26. Doi: https://doi.org/10.1016/j.chemolab.2008.11.006.

[5] Standard Normal Variate (2025) Standard Normal Variate - an overview | ScienceDirect Topics, ScienceDirect (accessed 2025-08-19).

Result

Scores (plot)

A graphical representation that displays the scores of the first 2 or 3 principal components as Factors, mathematical axis' orthogonal to each other, allowing for the visualization of the relationships between observations in a dataset[6]. Selecting the first 200 records of the IR - Automobile Paint Chips (APX) database with the parameter **Mean-center**, results in the plot below:

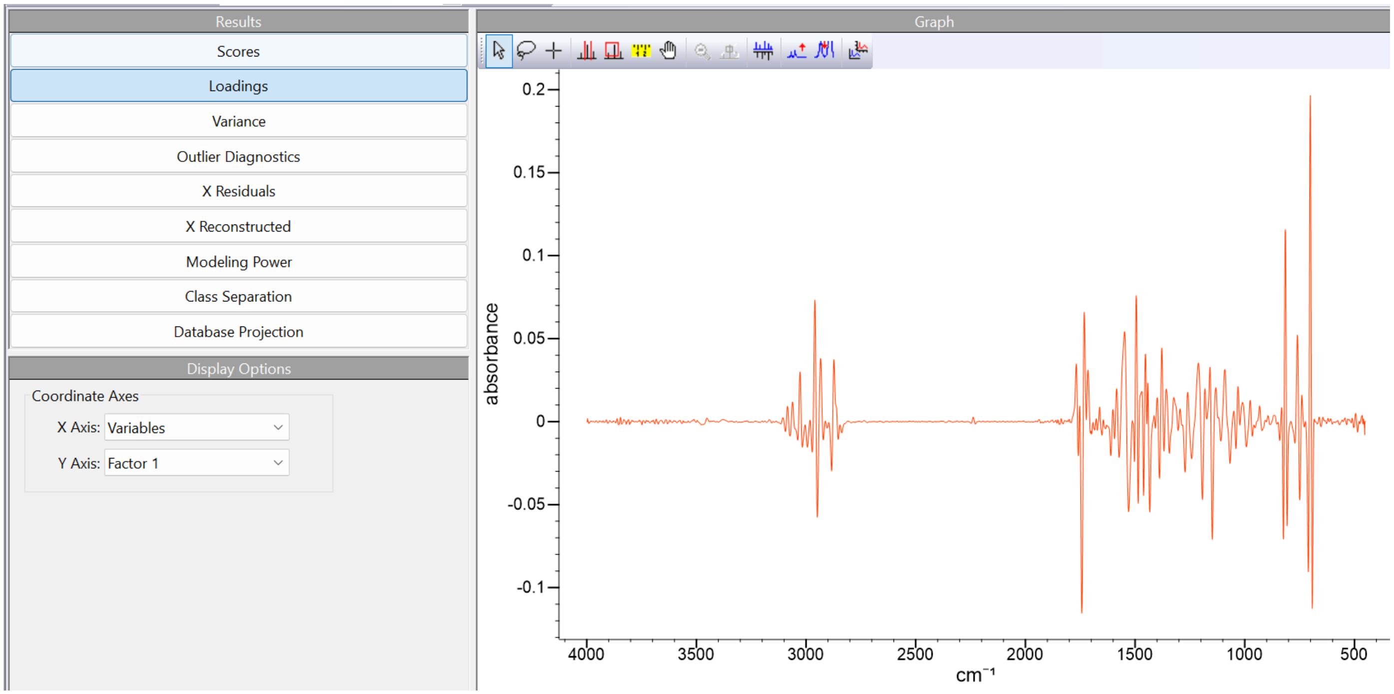


A circle (**Trajectory**) defines the boundary for a value in the **Field of Classification**).

[6] Dunn, K.G. (2025) Latent Variable Modelling in *Process Improvement using Data,* (https://learnche.org/pid/latent-variable-modelling/principal-component-analysis/interpreting-score-plots-and-loading-plots) (accessed 2025-08-19).

Loadings

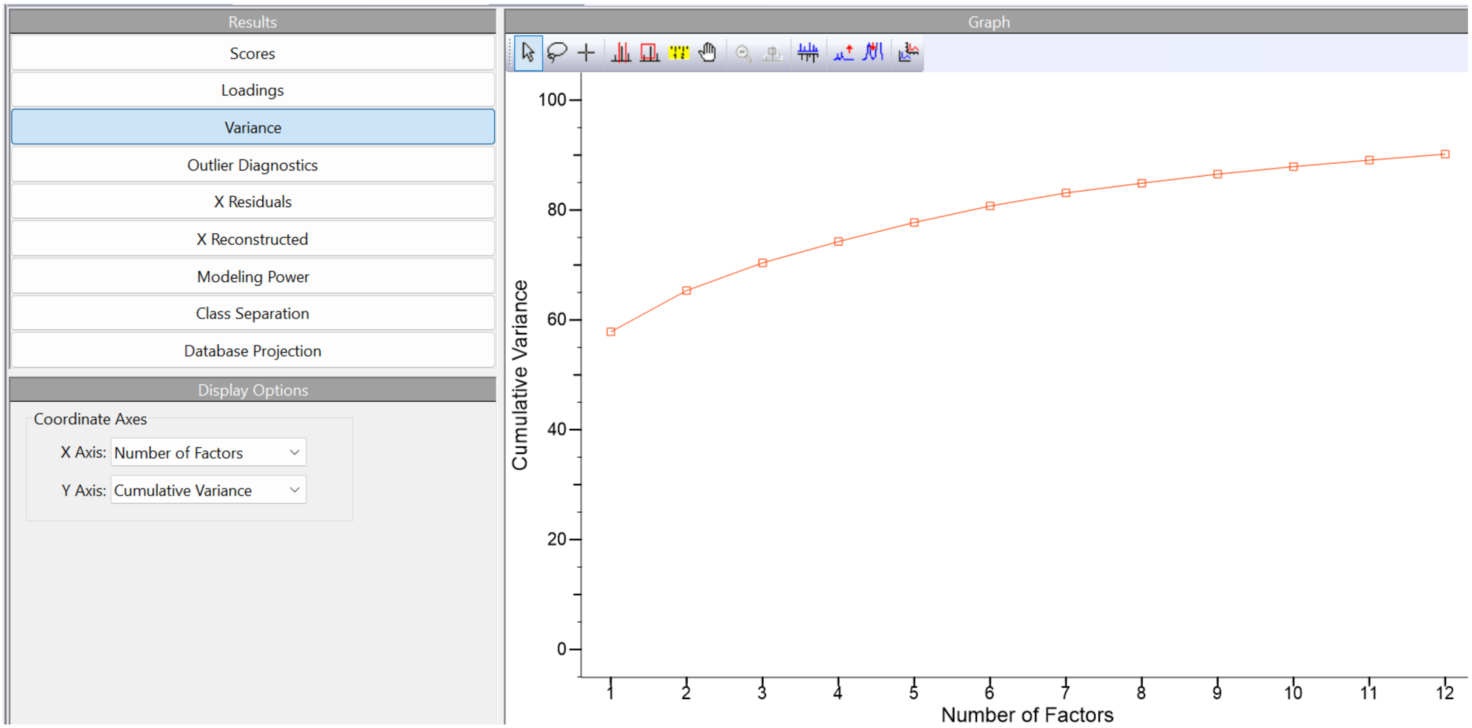
In order to show the relationship between the original variables and the principal components, it helps to understand how much each original variable contributes to the principal components and the nature of these contributions[7]. Using the above example, it would be:



[7] Schork, J. (2025) What are Loadings in PCA? *Statistics Global* (https://statisticsglobe.com/what-are-loadings-pca#loadings-in-pca) (accessed 2025-08-19).

Variance

Variance refers to the proportion of the total variance attributed to each principal component[8]. It helps us understand how much information is retained after dimensionality reduction. The fraction of variance explained by a principal component is the ratio between the variance of that principal component and the total variance[9] When applied to the above example, the result is:



[8] Chouinard, J.C. (2023) What is the Explained Variance in PCA (Python Example), https://www.jcchouinard.com/pca-explained-variance/ (accessed 2025-08-19).
[9] Cheplyaka, R. (2017) Explained variance in PCA, https://ro-che.info/articles/2017-12-11-pca-explained-variance (accessed 2025-08-19).

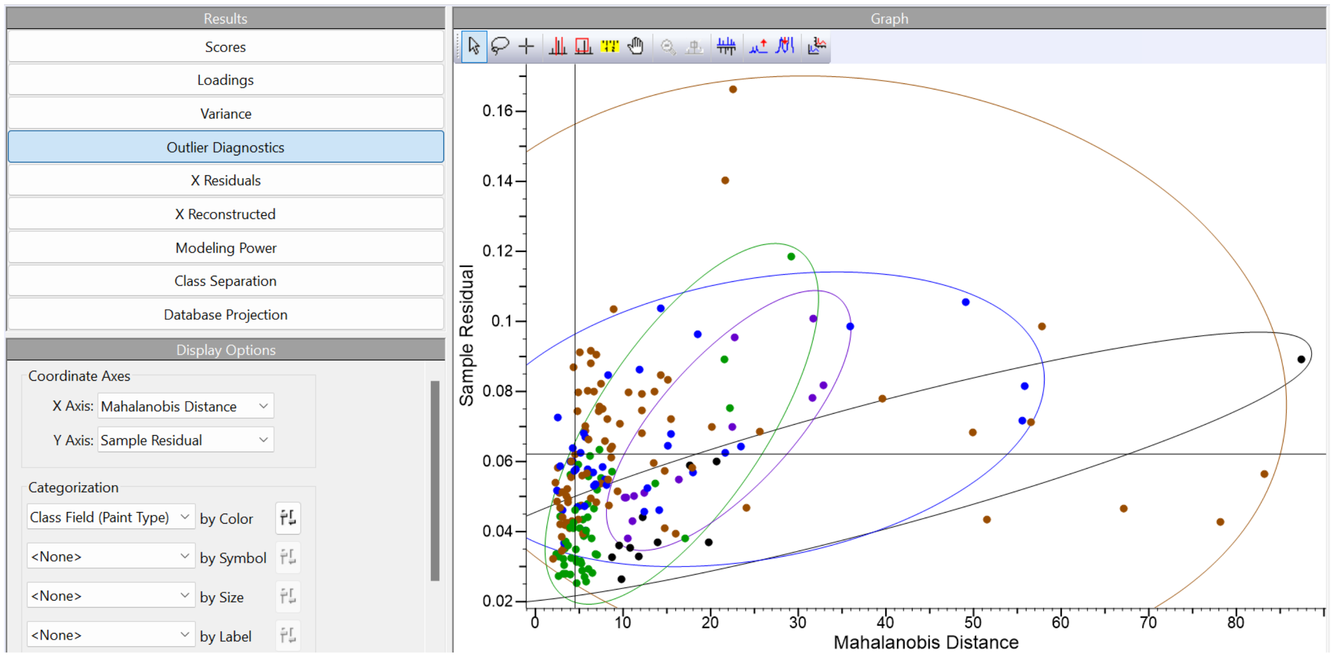
Outlier Diagnosis

It identify outliers by reducing dimensionality and visualizing data on scores plots. Here are the terms used in KnowItAll under the Outlier Diagnostics pane:

- **Mahalanobis Distance**[11]**:** Measures a distance between 2 points. It is a multivariate generalization of the square of the standard score z=(x−μ)/σ: how many standard deviations away P is from the mean of D. This distance is zero for P at the mean of D and grows as P moves away from the mean along each principal component axis. If each of these axes is re-scaled to have unit variance, then the Mahalanobis distance corresponds to standard Euclidean distance in the transformed space. The Mahalanobis distance is thus unitless, scale-invariant, and takes into account the correlations of the data set. In short, ***it measures how far a point is from the center of the data***

distribution, considering the covariance structure. PCA can be used to compute this distance in the reduced feature space. The following steps apply:

- Apply PCA to reduce dimensionality.
- Compute the Mahalanobis distance for each data point in the reduced space.
- Identify points with distances exceeding a threshold (e.g., based on a chi-squared distribution) as outliers. Using the above example, the result is:



- **Samples:** The ID of each sample (i.e., 1 to number of rows/spectra).
- **Sample residual**
- **F Ratio:** This is the statistic for evaluating whether two variances or standard deviations are significantly different. It is calculated by dividing one variance by another variance. If the null hypothesis is true, you expect F to have a value close to 1.0 most of the time. A large F ratio means that the variation among group means is more than you'd expect to see by chance. The F-distribution or F-ratio is a continuous probability distribution that arises frequently as the null distribution of a test statistic, most notably in the analysis of variance (ANOVA) and other F[15].
- **Probability**

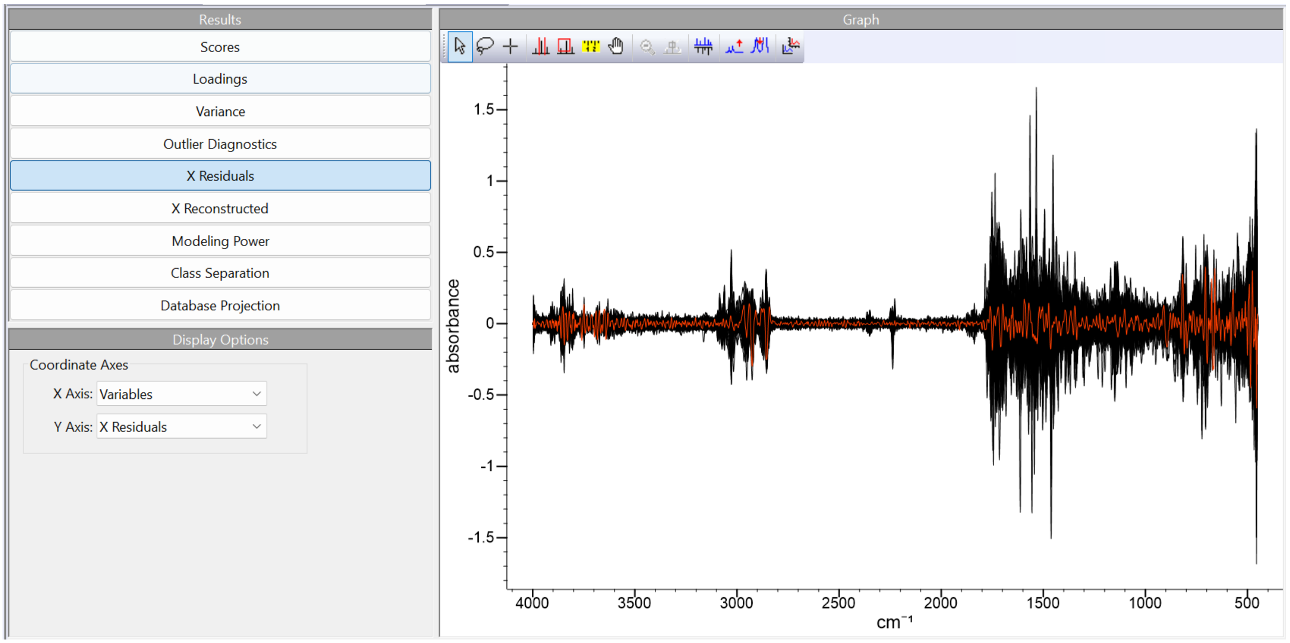
WILEY

- **Record ID**

[10] Datathatmatters (2024) Outlier Detection Simplified: PCA Techniques for Improved Data Analysis https://datathatmatter.com/2024/11/03/outlier-detection-simplified-pca-techniques-for-improved-data-analysis/ (accessed 2025-08-19).
[11] Wikipedia (2025) Mahalanobis distance, https://en.wikipedia.org/wiki/Mahalanobis_distance (accessed 2025-08-19).
[12] GraphPad Software, LLC.(2025) Interpreting results: One-way ANOVA, https://www.graphpad.com/guides/prism/latest/statistics/f_ratio_and_anova_table_(one-way_anova).htm (accessed 2025-08-19).

X Residuals

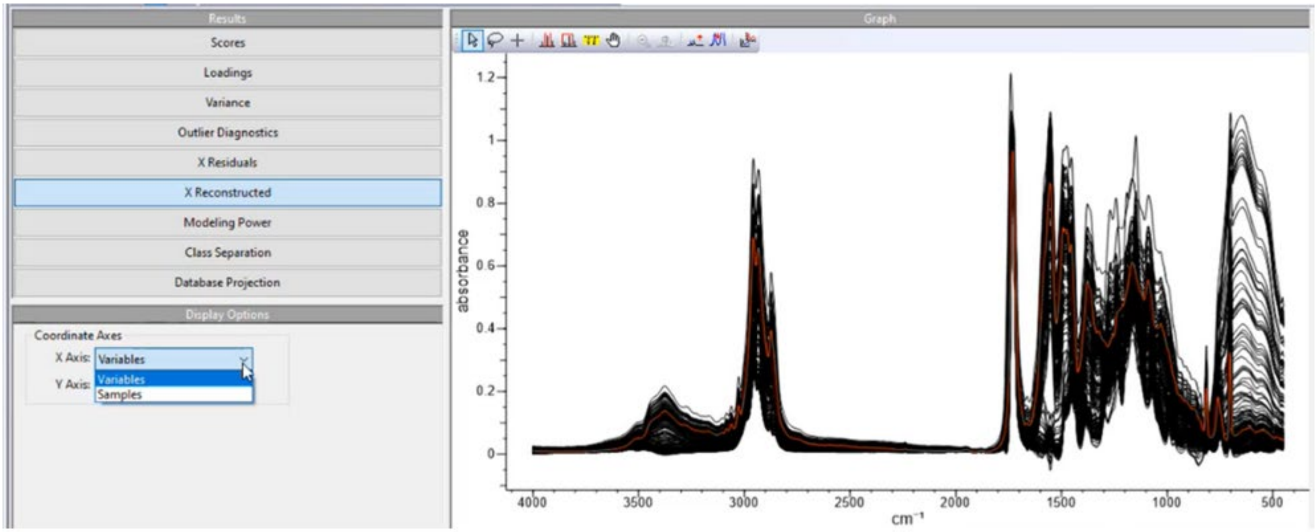
PCA residuals are calculated using the residual matrix E = X − T P ′ = X − X ^, where X is the original matrix and T P ′ is the PCA model. The residuals for each column in the original matrix can be calculated using the R 2 value, which gives an indication of how well the PCA model describes the data from that column. The function pcares(X,ndim) returns the residuals obtained by retaining ndim principal components of the n-by-p matrix X. PCA of a model residuals is based on how well the model translates the effect of variables in Z Z on the data we are analyzing. Using the **residual** matrix E = **X** − T P ′ = **X** − X ^, we can calculate the **residuals** for each column in the original matrix. This is summarized by the R 2 value for each column in **X** and gives an indication of how well the **PCA** model describes the data from that column[13]. Using the above example, we get:



[13] Dunn, K.G. (2025) Latent Variable Modelling in *Process Improvement using Data*, https://learnche.org/pid/latent-variable-modelling/principal-component-analysis/interpreting-the-residuals (accessed 2025-08-19).

X-reconstructed

Reconstruct the original variables from a principal component[14] It can be viewed as "reverse PCA". For the above dataset, we get:

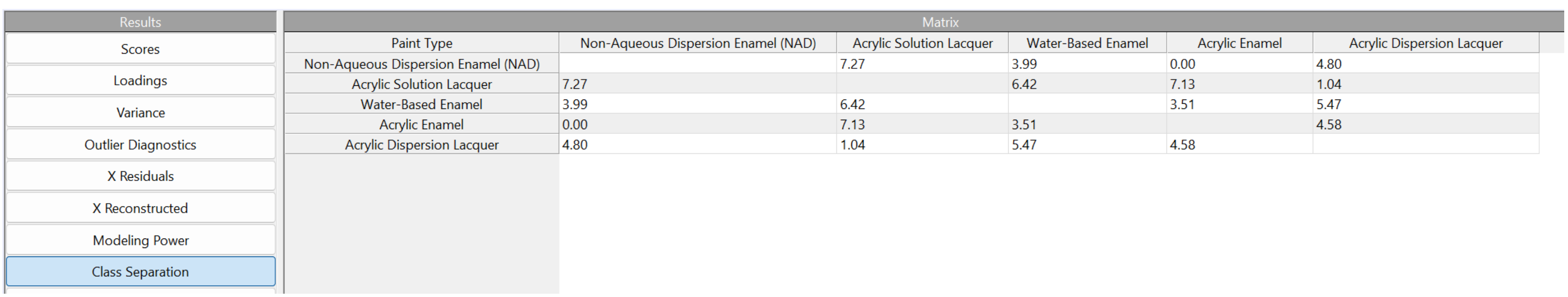


[14] Stack Exchange (2025) How to reverse PCA and reconstruct original variables from several principal components? https://stats.stackexchange.com/questions/229092/how-to-reverse-pca-and-reconstruct-original-variables-from-several-principal-com (accessed 2025-08-19).

Modeling power

Class Separation

It measures how well different class are separated[15]. Our particular example for the 1st 1000 spectra:



Paint Type	Non-Aqueous Dispersion Enamel (NAD)	Acrylic Solution Lacquer	Water-Based Enamel	Acrylic Enamel	Acrylic Dispersion Lacquer
Non-Aqueous Dispersion Enamel (NAD)		7.27	3.99	0.00	4.80
Acrylic Solution Lacquer	7.27		6.42	7.13	1.04
Water-Based Enamel	3.99	6.42		3.51	5.47
Acrylic Enamel	0.00	7.13	3.51		4.58
Acrylic Dispersion Lacquer	4.80	1.04	5.47	4.58	

[15] Stack Exchange (2025) Measures of class separability in classification problems, https://stats.stackexchange.com/questions/46780/measures-of-class-separability-in-classification-problems (accessed 2025-08-19).