Classical Least Squares for Essential FTIR

2nd Edition.

Table of Contents

Introduction

This document explains how to use the CLS tool in Essential FTIR. Classical Least Squares (CLS) is a multivariate data reduction technique used quantitative analysis of infrared spectra, usually in the vapor phase. This manual is not meant to explain CLS itself. There are many fine sources for this information, please see the References section.

Installation and Licensing

CLS is installed with Essential FTIR. The latest release of Essential FTIR can be downloaded from http://www.essentialftir.com/download.html. Installation of Essential FTIR is explained in the Essential FTIR manual, which can be downloaded from the same site.

Although included in the Essential FTIR installation, CLS is an add-on package requiring an additional license. Please contact essentialFTIR@tds.net to get a CLS license.

In addition to the CLS tool for Essential FTIR, a stand-alone prediction server is provided, allowing other software to run CLS predictions using methods developed in Essential FTIR. The prediction server is documented in the file 'eftir_cls_predictor.html'.

Overview

A CLS method consists of analytes, which are the compounds that need to be quantified. An analyte needs at least one 'standard'. A standard is a spectrum of the pure analyte acquired at known concentration, temperature, pathlength and pressure. Standards require Analytical Regions to be defined, which are spectral regions in which the analyte has characteristic absorbances. In addition, Interferents can be assigned to a Standard. An Interferent is the spectrum of a pure compound which also has absorbance in the Analytical Regions assigned to the Standard, and which may be present in samples measured in the field. In practice, the Interferents are often other Analytes in the method.

A CLS method can be seen as a hierarchy. A method consists of Analytes. An Analyte consists of Standards. A Standard consists of 1) Analytical Regions and 2) Interferents In Essential FTIR, it is possible to build a library of Analytes, and to easily import them into a Method. Also, Essential FTIR can automatically determine which Analytes included in the method have interfering absorbances with other Analytes in the method, and assign these Interferents automatically to each Analyte in the Method.

The results of a CLS prediction follow the same hierarchy. For a given sample and standard, a separate prediction is performed for each region defined for the standard. These individual region results are combined using a weighted average, a technique

known as ‘Multi-Band Weighting’ (see reference 2). There is a MBW result for each standard for the analyte. The two standards whose MBW results most closely ‘bracket’ the predicted results are used to interpolate a final result for the analyte. So, there is a CLS prediction result for each region, a MBW result for each standard, and an interpolate result for each analyte.

In the case where the predicted result is not ‘bracketed’ by two standards, the closest MBW result is reported as the final analyte result. Extrapolated results are flagged as such, and are not as reliable as interpolated results. The CLS method should be configured so that there are standards with concentrations both lower and higher than any samples that will be analyzed.

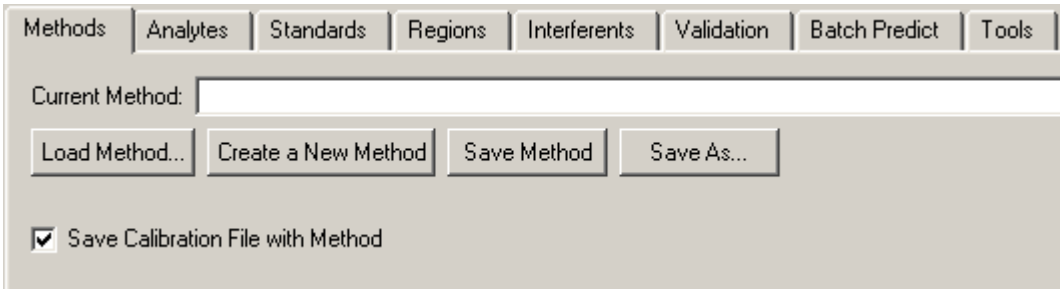
The CLS Tool

The CLS tool is in the Essential FTIR 'Analysis' Tool category.

The CLS tool has these tabs:

Each of these tabs will now be explained.

The Method Tab



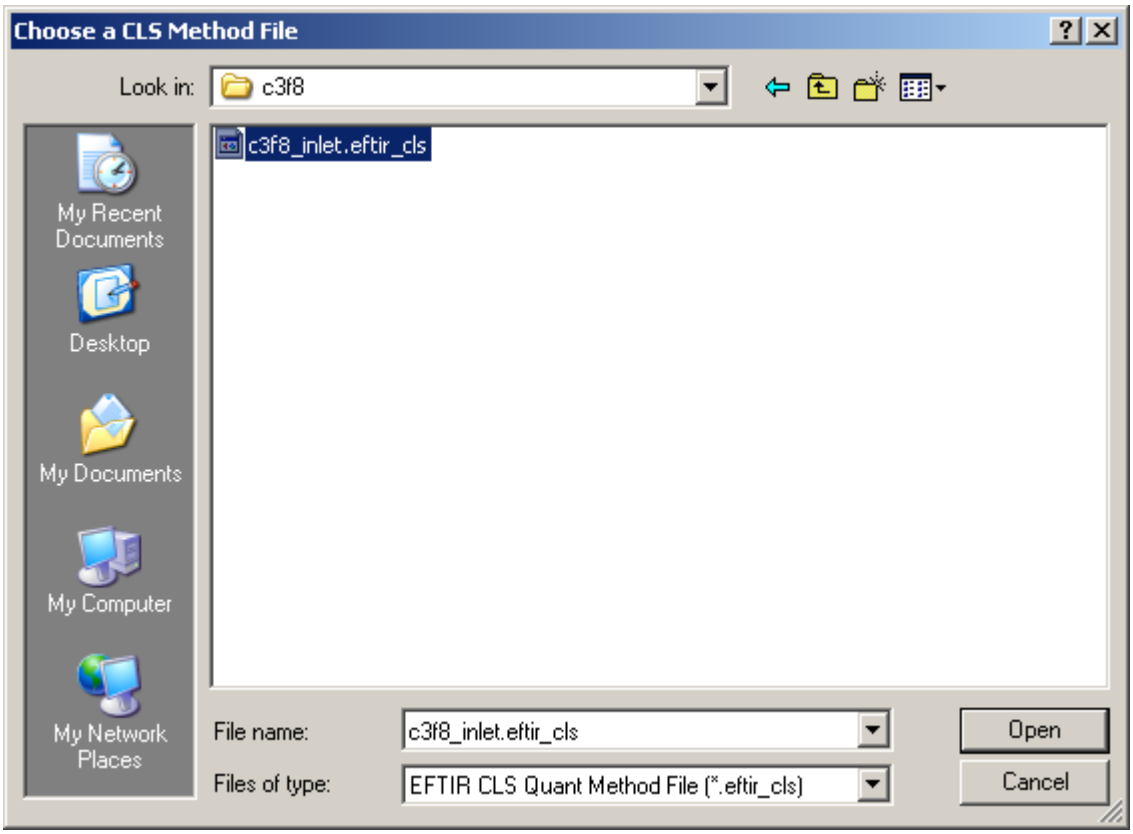
On this tab, methods are loaded, created, and saved.

Load Method...	Load an existing method from disk.
Create a New Method	Create a new, blank, method.
Save Method	Save the current method.
Save As...	Prompt for a filename for saving the current method.
Save Calibration File with Method	Save the calibration matrices alongside the method file.

The Calibration file has the extension '.cal' and is saved in the same directory as the method file. An Essential FTIR CLS method has the file extension '.eftir_cls'.

It's not necessary to save the calibration file, because calibrations will be performed as needed, but it can save time, especially for large methods with lots of Analytes, Standards, and Interferents.

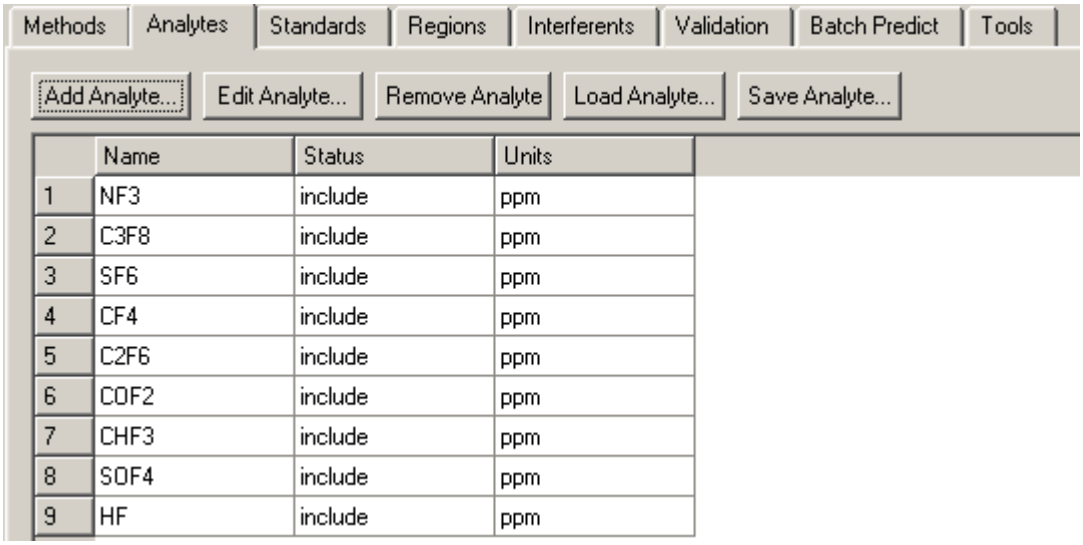
Click on the ‘Load Method’ button. There is a sample method available, which is installed into a subfolder of “C:\Documents and Settings\All Users\Documents\EFTIR\CLS” named c3f8. The method file is named “c3f8_inlet.eftir_cls”. If this method is not on your computer, it can be downloaded from http:\\www.essentialFTIR.com/tools.html.



The Analytes Tab

An analyte is a compound of interest, the concentration of which you want to measure using a CLS method.

This screen shot assumes you have loaded the c3f8 method.



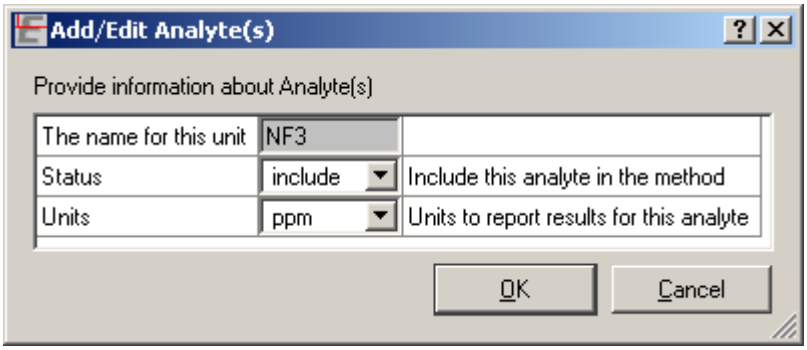
	Name	Status	Units
1	NF3	include	ppm
2	C3F8	include	ppm
3	SF6	include	ppm
4	CF4	include	ppm
5	C2F6	include	ppm
6	COF2	include	ppm
7	CHF3	include	ppm
8	SOF4	include	ppm
9	HF	include	ppm

There is one row in the Analyte table for each analyte. The three columns in the table are labeled ‘Name’, ‘Status’ and ‘Units’.
Status can be ‘include’ or ‘exclude’. This allows quickly including or excluding analytes from the method, without adding or removing them.

On this tab, Analytes are Created, Edited and Removed.

Add Analyte...	Add a new analyte to the method
Edit Analyte...	Edit the selected Analyte or Analytes.
Remove Analyte	Remove the selected Analyte or Analytes.
Load Analyte...	Load an Analyte from disk
Save Analyte...	Save the selected Analyte to disk.

To edit an analyte, highlight its row in the analyte table, and then click ‘Edit Analyte’. If you select multiple rows in the table, you can edit multiple analytes at a time. Here is the Dialog that allows editing analytes:



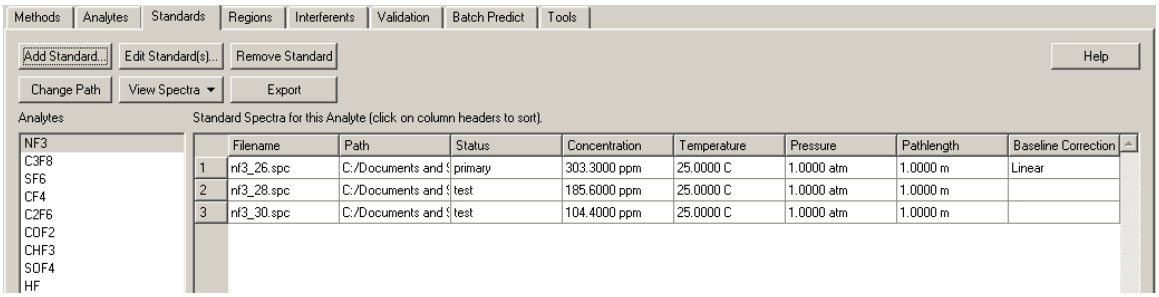
Note that the Analyte Name is grayed out; once an analyte is added to a method, its name cannot be changed. This is because the analyte name serves as a key for organizing Standards and Results.

Analytes can be saved as separate files, along with their Standards, Regions, and Interferents. These Analyte files can be loaded into a method as a package. This saves the tedium of building new methods from scratch; a method can be developed by loading pre-existing Analyte files. An analyte file has the extension '.analyte'. The 'Save Analyte' and 'Load Analyte' allow you to save and recall .analyte files.

The Standards Tab

A Standard is the spectrum of a pure Analyte. It is used to generate calibration information for predicting concentrations of unknowns. These pieces of information must be known about the Standard: its concentration, temperature, pressure and pathlength.

Here is another view of the Standards Tab which includes the Standards Table:



On the left is a list of Analytes. As a different analyte is selected, the Standards table changes to display the Standards for the selected Analyte.
The Standards table has these columns: Filename, Path, Status, Concentration, Temperature, Pressure, and Pathlength, and Baseline Correction.
The standards table can be sorted by clicking on any of the column headers. For instance, to sort the table by Concentration, click on the 'Concentration' column header.

Filename	The filename containing this Standard's spectrum.
Path	The folder the file is in.
Status	This can be include, exclude, primary, or test
Concentration	The known concentration of the Analyte in this Standard
Temperature	The temperature when the spectrum was collected
Pressure	The pressure when the spectrum was collected
Pathlength	The pathlength when spectrum was collected.
Baseline Correction	This can be None, Offset, Linear, or Curve. 'Curve' should be used with care, because baseline corrections are applied individually to analysis regions, not to the entire spectrum. If 'Curve' is used over a narrow peak region, the peak will be flattened out.

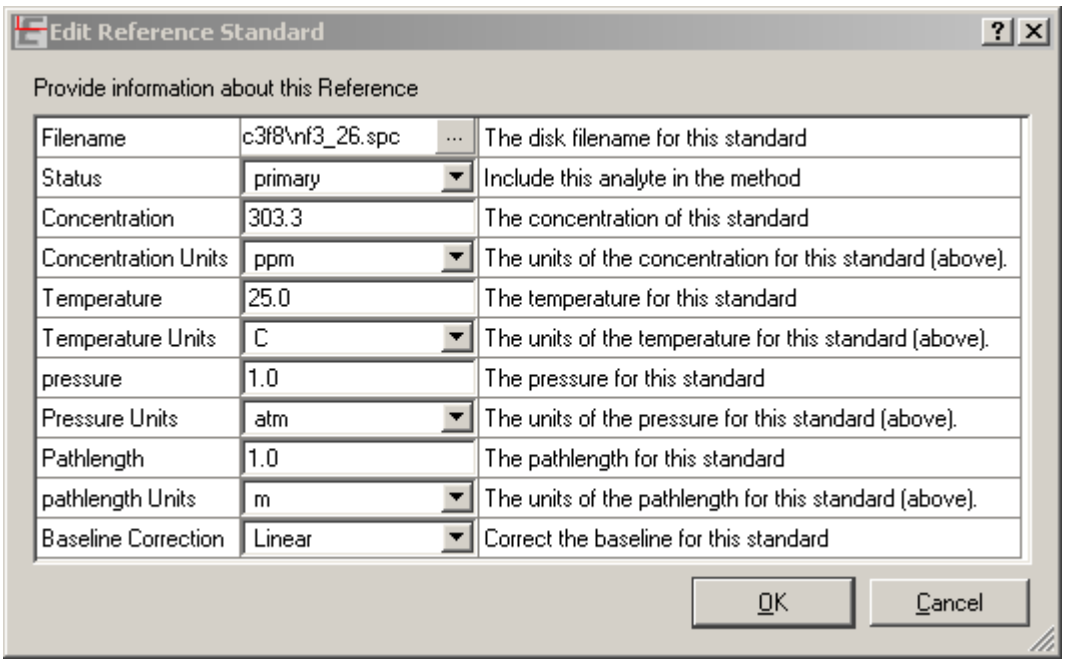
Each standard can have a different baseline correction method assigned to it.

Add Standard...	Add a standard
Edit Standard(s)...	Edit the selected Standard or Standards.
Remove Standard	Remove the selected Standard or Standards
Change Path	Change the folder that the Standards are located in.
View Spectra ▾	View the Standard Spectra.
Export	Export the Standards table to file, clipboard, or Excel.

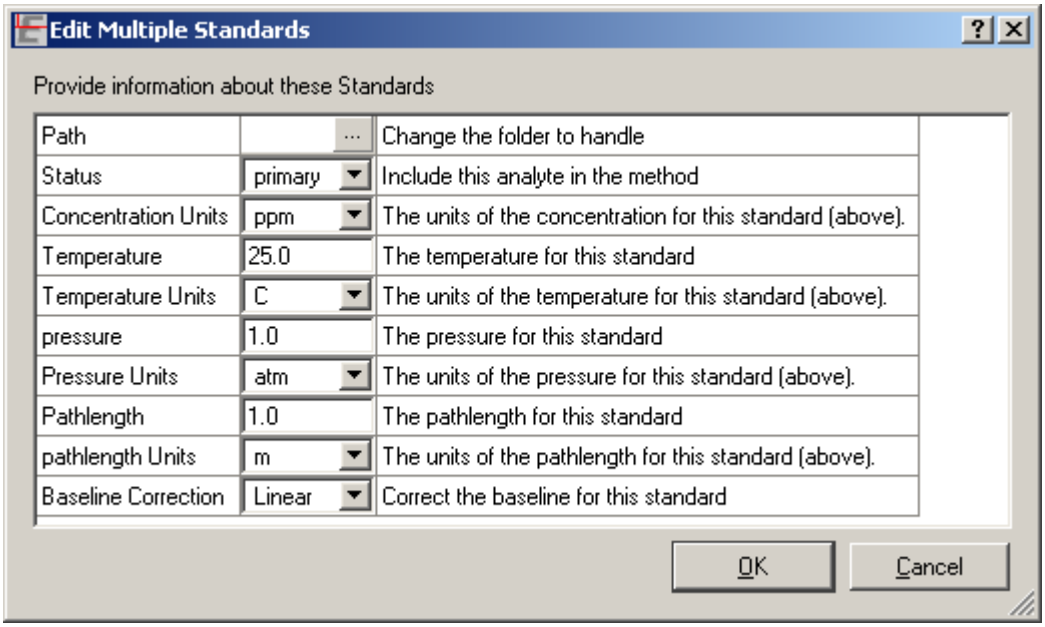
The Path can be changed independently of the Filename; you may have separate folders containing Standards that share the same filename, but may have been pre-processed differently, or have different spectral resolutions, or have been acquired from different instruments. This allows the method developer to experiment during the development phase by switching between sets of Standards.

Standards can be edited by double-clicking on a row in the table, or selected one or more rows in the table and then clicking the 'Edit Standard(s)' button.

If editing a single standard, this dialog appears:



When editing multiple standards, this slightly different dialog appears, which contains a subset of the above fields, showing only those settings that would apply to multiple standards:



Note that this is another way to change the Path of multiple standards.

Note: when a method is loaded, if the standards cannot be found in the Path assigned to them, Essential FTIR automatically looks in the method directory itself. This makes it easier to deploy methods to other computers.

Standards and Primary Status.

A standard can be assigned a status of include, exclude, primary, or test. For a given analyte, there can only be one primary standard. The primary standard has these special properties:

- The analytical regions of the primary standard are used for automatic interferent analysis (more on this in the Regions and Interferents sections, below).
- The primary standard's spectrum is used as the interferent spectrum when automatic interferent analysis detects interference.
- When the given analyte has multiple standards, the primary standard is used to find the bracketing standards.

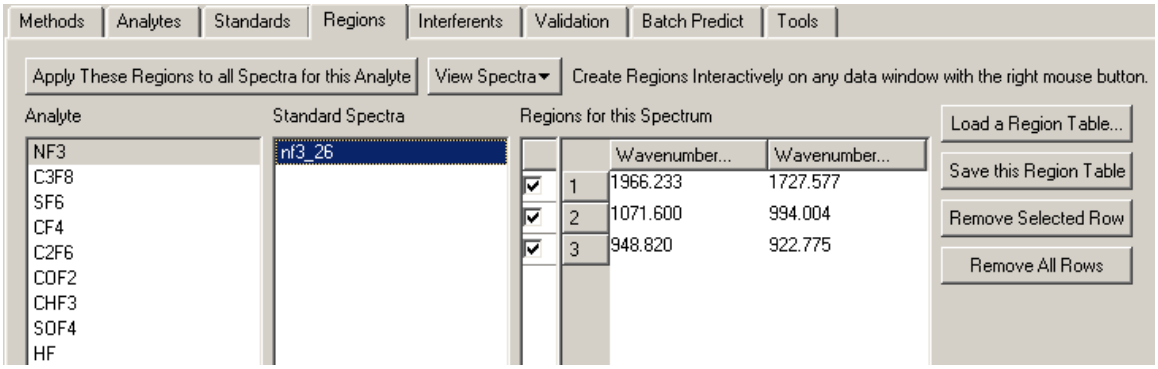
So, in the case where an analyte has multiple standards, care should be taken in selecting the primary standard, and also in selecting the analytical regions for the primary standard.

The Regions Tab

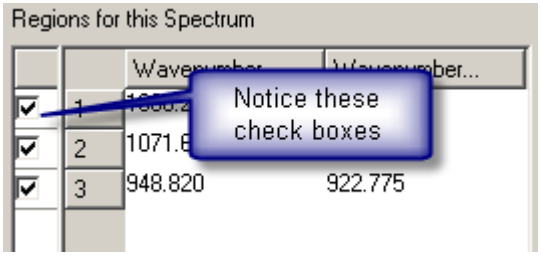
On this tab, analytical regions are assigned to Standards.

Analytes usually have spectral regions that contain information specific to that analyte. Specifying spectral regions will improve the results and make the analysis faster.

Regions are assigned interactively in a spectral display workspace; please see the Essential FTIR manual which has an entire section about creating and changing region selections. Basically, you create regions markers by right-clicking in the spectral display window. For CLS you can create as many regions as you need. Region markers can be moved by dragging with the left mouse button and removed by clicking on it with the right mouse button.



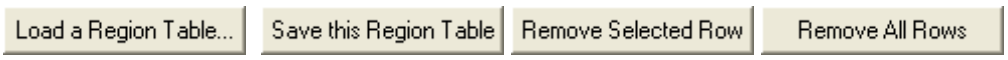
On this tab, there is a list of Analytes included in the method, a list of Standards for the selected Analyte, and a table of analysis regions for the selected Standard.
Note the column of check boxes to the right of the table of analysis regions.



These check boxes allow regions to be defined for the Analyte/Standard, but they are only actually used in calibration and prediction if they are checked. This allows regions to be defined for the Primary Standard that that may interfere with other Analytes in the method (more on this in the 'Interferents Tab' section below). Also, it is common to use a different set of regions for high versus low concentrations of an analyte. You can define all the possible analysis regions for a Standard, and then click the 'Apply these regions to all Spectra for this Analyte' button. Then, clicking on each Standard in turn,

enable the analysis regions, using the check boxes, which are appropriate for that standard.

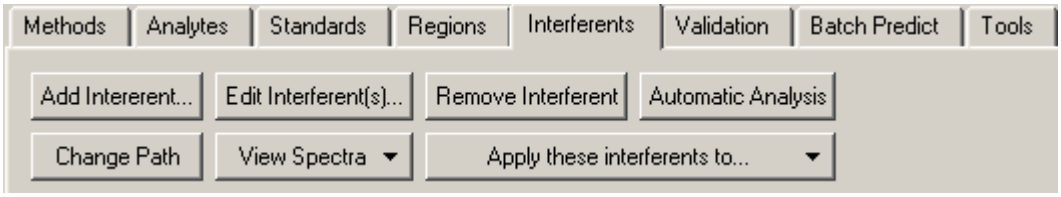
Apply These Regions to all Spectra	Apply the displayed regions to all of an Analyte's Standards
View Spectra	View the spectra associated with Standards.
Load a Region Table...	Load a saved region table.
Save this Region Table	Save this region table to file.
Remove Selected Row	Remove the selected row from the region table
Remove All Rows	Remove all rows from the region table



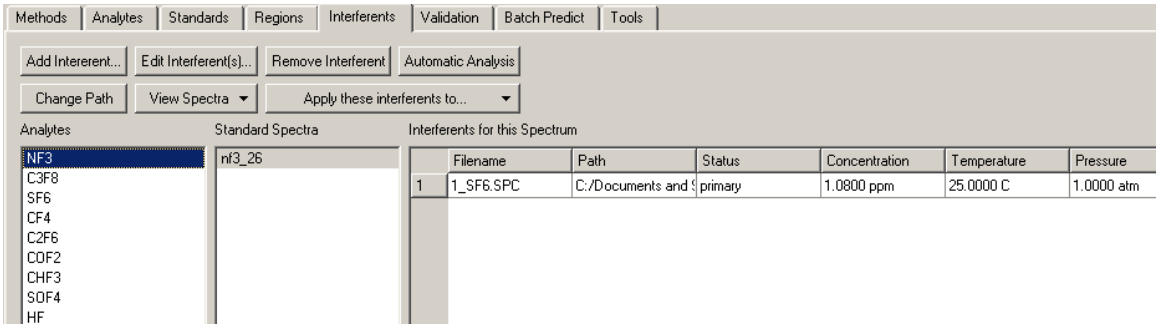
These buttons are covered in detail in the Essential FTIR manual, but their functions are self-explanatory. The same functions are also available by clicking on the 'Wavenumber…' column headers in the wavenumber table.

The Interferents Tab

The interferents tab has these buttons:



Here is another view of the interferents tab that includes the Interferents Table.



For a given Analyte and Standard pair, it is possible that other Analytes have significant absorbance in the given Analyte's analysis regions. These are called 'interferents', and they must be included in the calibration for the given analyte. This tab allows you to define interferents for the given analyte.

An interferent does not have to be an Analyte or Standard in the method; you can include any spectrum. Just like Standards, the concentration, temperature, pressure and pathlength of the interferent spectra have to be known.

This can be a tedious operation, but there are some short cuts built into this software to make the job easier.

First of all, you can define the interferents for one Standard, and then apply them to all other standards for the given analyte, or to all standards for all analytes. Then you can edit the interferents for each Analyte and Standard by removing or excluding those interferents, which are not applicable to the analysis.

Automatic Interferent Analysis

Second, you can use the 'Automatic Analysis' function. This identifies interferences among the analytes and standards that are included in the method. Using the analysis regions defined for the primary standard for each analyte, it looks for overlap with the analysis regions of all the other primary standards. When any overlap is found, that other primary standard is added to the interferents of all the standards for the given analyte.

Add Intererent...	Add an interferent to the selected Standard.
Edit Interferent(s)...	Edit the selected Interferent or Interferents.
Remove Interferent	Remove the selected Interferent or Interferents
Automatic Analysis	Automatically assign interferents to standards
Change Path	Change the folder that the Interferents are located in.
View Spectra ▾	View the spectra associated with the Interferents
Apply these interferents to... ▾	Apply these interferents to other Standards or Analytes

The Validation Tab

Validation means testing the CLS method against known data. You must have assigned some Standards to the 'Test' set in order to validate the selected Analyte.

For this example, the 'C3F8' analyte is selected. The calibration and test samples have been changed from the default method in order to show some features of the table:

		Sample Filename	Known Sample Concentration	Standard Filename	Standard Concentration	Region	Predicted Concentration	Difference (Known-Predicted)	SEC	Fractional Uncertainty	Sigma Squared
1	✚	60_6C3F8.SPC	60.5700 ppm			Interpolated	60.6473 ppm	-0.0773 ppm	2.6325	0.0434	0.0017
2	✚	C3F8_17A.SPC	106.8000 ppm			Interpolated	118.2228 ppm	-11.4228 ppm	4.5885	0.0388	0.0047
3	✚	C3F8_5A.SPC	11.4000 ppm			Extrapolated	12.8574 ppm	-1.4574 ppm	0.0695	0.0054	8.4661e-007

The table contains the results after the analysis. There is a hierarchy of results, and each 'Predicted Concentration' is derived from other results. The plus signs in the first column, ✚, allow viewing of the contributing results. After clicking on the plus sign in row one, the table is expanded to look like this:

		Sample Filename	Known Sample Concentration	Standard Filename	Standard Concentration	Region	Predicted Concentration	Difference (Known-Predicted)	SEC	Fractional Uncertainty	Sigma Squared
1	▬	60_6C3F8.SPC	60.5700 ppm			Interpolated	60.6473 ppm	-0.0773 ppm	2.6325	0.0434	0.0017
2			✚	43_1C3F8.SPC	43.1400 ppm	Multi-Band Weighted	59.6005 ppm	0.9695 ppm	0.0998	0.0017	4.0254e-007
3			✚	29_8C3F8.SPC	29.7900 ppm	Multi-Band Weighted	60.0353 ppm	0.5347 ppm	0.0584	0.0010	2.3923e-006
4			✚	C3F8H13A.SPC	560.5000 ppm	Multi-Band Weighted	92.5022 ppm	-31.9322 ppm	2.7157	0.0294	0.0018
5	✚	C3F8_17A.SPC	106.8000 ppm			Interpolated	118.2228 ppm	-11.4228 ppm	4.5885	0.0388	0.0047
6	✚	C3F8_5A.SPC	11.4000 ppm			Extrapolated	12.8574 ppm	-1.4574 ppm	0.0695	0.0054	8.4661e-007

Now the multi-band weighted results are show for each standard. The two MBW results that were used for interpolating the final result are highlighted with a green background. The final results are highlighted with a yellow background.

Now there are plus signs in column 3 for the rows that contain MBW results. Clicking on these will expand the table so that the individual results for each region that contributes to the MBW result will be shown:

		Sample Filename	Known Sample Concentration	Standard Filename	Standard Concentration	Region	Predicted Concentration	Difference (Known-Predicted)	SEC	Fractional Uncertainty	Sigma Squared
1	▬	60_6C3F8.SPC	60.5700 ppm			Interpolated	60.6473 ppm	-0.0773 ppm	2.6325	0.0434	0.0017
2			▬	43_1C3F8.SPC	43.1400 ppm	Multi-Band Weighted	59.6005 ppm	0.9695 ppm	0.0998	0.0017	4.0254e-007
3						1241.862-1283.275	58.3614 ppm	2.2086 ppm	0.0551	0.0009	8.6228e-007
4						994.085-1020.758	59.6520 ppm	0.9180 ppm	0.0901	0.0015	1.9042e-007
5						1130.509-1169.867	59.6636 ppm	0.9064 ppm	0.1531	0.0026	5.4111e-008
6			✚	29_8C3F8.SPC	29.7900 ppm	Multi-Band Weighted	60.0353 ppm	0.5347 ppm	0.0584	0.0010	2.3923e-006
7			✚	C3F8H13A.SPC	560.5000 ppm	Multi-Band Weighted	92.5022 ppm	-31.9322 ppm	2.7157	0.0294	0.0018
8	✚	C3F8_17A.SPC	106.8000 ppm			Interpolated	118.2228 ppm	-11.4228 ppm	4.5885	0.0388	0.0047
9	✚	C3F8_5A.SPC	11.4000 ppm			Extrapolated	12.8574 ppm	-1.4574 ppm	0.0695	0.0054	8.4661e-007

This allows 'drilling down' to see all the lower-level results that contribute to a final result.

Note the ‘Region’ column for the sample C3F8_5A.SPC (row 9 in the above picture). It says ‘Extrapolated’ and is highlighted in red. This means that the predicted result is not bracketed by two calibration standards. Extrapolated results are not as reliable as interpolated results.

Extrapolated results are also highlighted in red when batch processing files.

Uncertainty

SEC is the ‘Standard Error of Concentration’ and is a statistical measurement derived from the spectral residual. Fractional Uncertainty, also known as UCR, is SEC / (Predicted Concentration). The lower the UCR, the better the result.

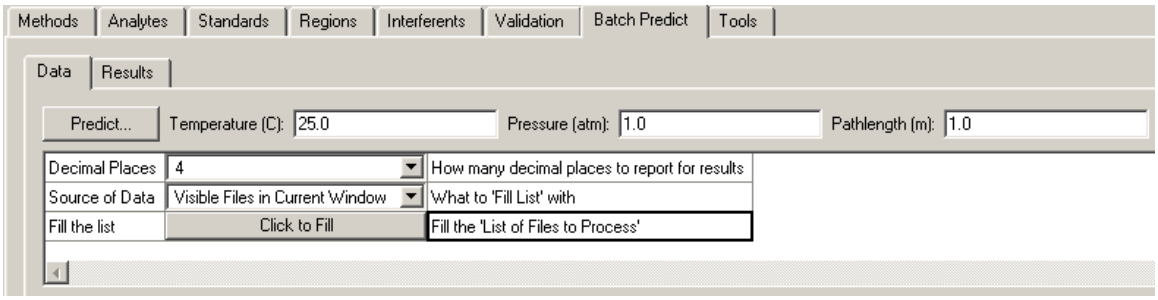
Notice how the predicted concentration for the high concentration test sample, C3F8H13A.SPC, the last one in the table, is way off from the known. This is because of non-linearity in the data: absorbance versus concentration is not a straight line, especially at higher concentrations.

Run Test	Run the validation test.
View Residual	View the spectral residual associate with a result or results.

To view the spectral residual, select a row or multiple rows in the results table, and click ‘View Residual’. If the row contains an interpolated result, two residuals will be displayed, representing the two MBW results that contributed to the interpolation.

The Batch Predict Tab

This tool allows you to predict the concentrations of analytes in unknown samples.



As you can see, this tab has two sub-tabs, named ‘Data’ and ‘Results’

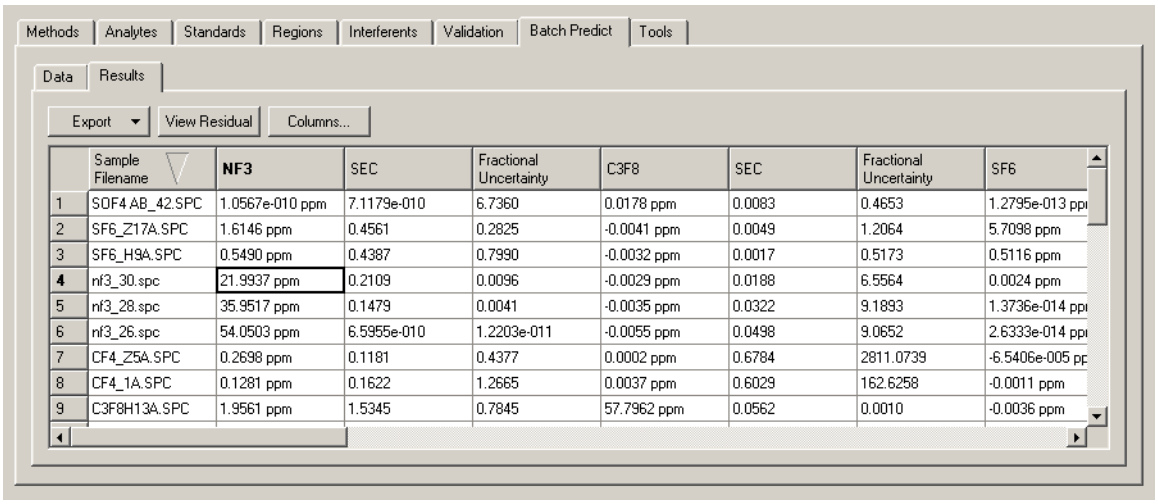
On the ‘Data’ tab are these buttons and fields:

Predict...	Run a prediction over the selected fiels.
Temperature (C):	The temperature, in Celsius, to use in the prediction.
Pressure (atm):	The pressure, in atmospheres, to use in the prediction.
Pathlength (m):	The pathlength, in meters, to use in the prediction.

To analyze samples, the temperature, pressure, and pathlength of the samples must be known. These may be set using the edit fields on this tool. If you get unexpected results, the most likely cause is incorrect settings for these parameters.

To populate the List of Files to process, select the ‘Source of Data’, and then click the ‘Click to Fill’ button. The samples to process can be either data that is already loaded into an Essential FTIR workspace, or files that are on disk.

This screenshot was obtained by clicking ‘View Spectra / All Standards for All Analytes’ on the ‘Standards’ Tab, and then choosing ‘All files in current window’ for the source of data, and then clicking ‘Predict’.



	Sample Filename	NF3	SEC	Fractional Uncertainty	C3F8	SEC	Fractional Uncertainty	SF6
1	SOF4 AB_42.SPC	1.0567e-010 ppm	7.1179e-010	6.7360	0.0178 ppm	0.0083	0.4653	1.2795e-013 pp
2	SF6_Z17A.SPC	1.6146 ppm	0.4561	0.2825	-0.0041 ppm	0.0049	1.2064	5.7098 ppm
3	SF6_H9A.SPC	0.5490 ppm	0.4387	0.7990	-0.0032 ppm	0.0017	0.5173	0.5116 ppm
4	nf3_30.spc	21.9937 ppm	0.2109	0.0096	-0.0029 ppm	0.0188	6.5564	0.0024 ppm
5	nf3_28.spc	35.9517 ppm	0.1479	0.0041	-0.0035 ppm	0.0322	9.1893	1.3736e-014 pp
6	nf3_26.spc	54.0503 ppm	6.5955e-010	1.2203e-011	-0.0055 ppm	0.0498	9.0652	2.6333e-014 pp
7	CF4_Z5A.SPC	0.2698 ppm	0.1181	0.4377	0.0002 ppm	0.6784	2811.0739	-6.5406e-005 pp
8	CF4_1A.SPC	0.1281 ppm	0.1622	1.2665	0.0037 ppm	0.6029	162.6258	-0.0011 ppm
9	C3F8H13A.SPC	1.9561 ppm	1.5345	0.7845	57.7962 ppm	0.0562	0.0010	-0.0036 ppm

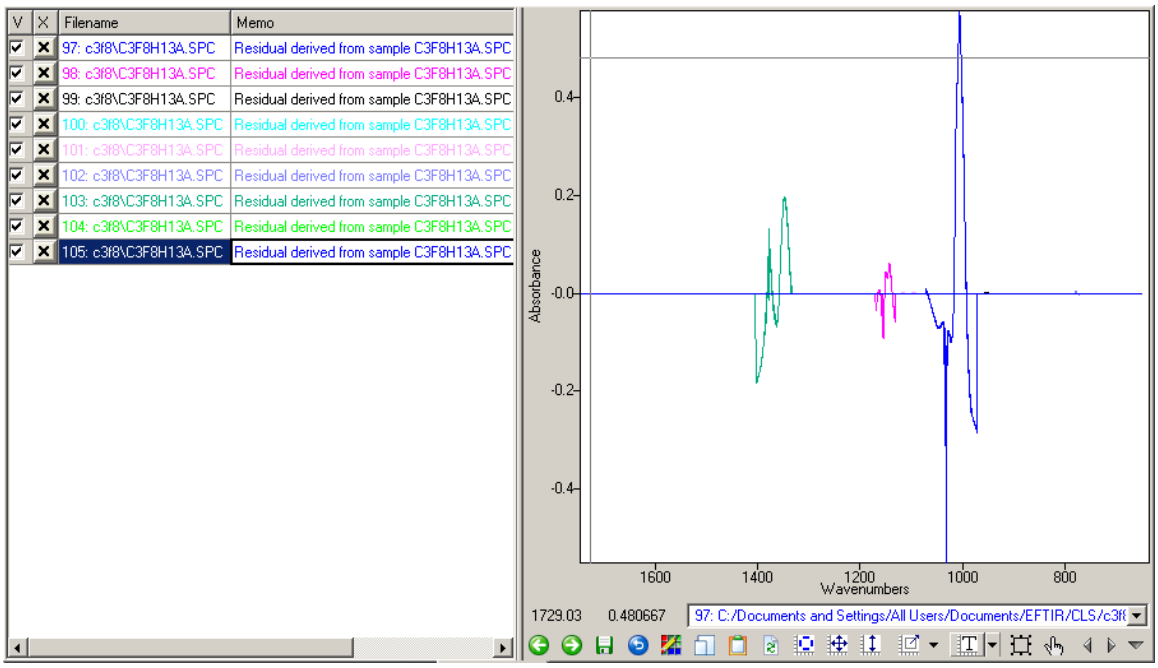
As you can see, the results table has a lot of information. Each sample file is a row in the table, and there are three columns for each analyte: the concentration, SEC and UCR (Fractional Uncertainty). Any extrapolated results will be highlighted in red. The table can be sorted by clicking on any of the column headers.

Export	Export the results table to clipboard, file, or Excel.
View Residual	View the spectral residual associated with a result.
Columns...	Select the information that will be displayed in the results table.

View Residual: For each sample there are one or two residuals for each analyte. If the result is interpolated, there will be two residuals. Select the row(s) you want to view the residuals for, and click 'View Residual'. In the case of this sample method, if you only select one row, you will get as many residuals as there are Analytes included in the method. In this screen shot, the row for the sample file C3F8H13A.SPC was selected. The Memo field of each residual tells which analysis (that is, Analyte) produced that residual.

Columns: Clicking this button will allow setting what data is displayed in the results. The predicted concentration is always displayed, but the error terms are optional.

Show SEC	☑	Show the Standard Error of Concentration
Show UCR	☑	Show the Fractional Uncertainty
Show S2	☐	Show Sigma Squared


V
X
Filename
Memo
97: c3f8\C3F8H13A.SPC
Residual derived from sample C3F8H13A.SPC
98: c3f8\C3F8H13A.SPC
Residual derived from sample C3F8H13A.SPC
99: c3f8\C3F8H13A.SPC
Residual derived from sample C3F8H13A.SPC
100: c3f8\C3F8H13A.SPC
Residual derived from sample C3F8H13A.SPC
101: c3f8\C3F8H13A.SPC
Residual derived from sample C3F8H13A.SPC
102: c3f8\C3F8H13A.SPC
Residual derived from sample C3F8H13A.SPC
103: c3f8\C3F8H13A.SPC
Residual derived from sample C3F8H13A.SPC
104: c3f8\C3F8H13A.SPC
Residual derived from sample C3F8H13A.SPC
105: c3f8\C3F8H13A.SPC
Residual derived from sample C3F8H13A.SPC
0.4
0.2
-0.0
-0.2
-0.4
Absorbance
1600
1400
1200
1000
800
Wavenumbers
1729.03
0.480667
97: C:/Documents and Settings/All Users/Documents/EFTIR/CLS/c3f

References

1. D.M. Haaland and R.G. Easterling, “Improved Sensitivity of Infrared Spectroscopy by the Application of Least Squares Methods,” Appl. Spectrosc. 34(5):539-548 (1980).

2. D.M. Haaland and R.G. Easterling, “Application of New Least-Squares Methods for the Quantitative Infrared Analysis of Multicomponent Samples,” Appl. Spectrosc. 36(6):665-673 (1982).

3. D.M. Haaland, R.G. Easterling and D.A. Vopicka, “Multivariate Least-Squares Methods Applied to the Quantitative Spectral Analysis of Multicomponent Samples,” Appl. Spectrosc. 39(1):73-84 (1985).

4. W.C. Hamilton, Statistics in Physical Science, Ronald Press Co., New York, 1964, Chapter 4.

5. Richard Kramer, Chemometric Techniques for Quantitative Analysis, Marcel Dekker Inc., New York, 1998, Chapter 3.