MsMetrix

MsXelerator™
User Manual: GCMS QC / GCMS Quant

Introduction:

This document describes the GCMS QC and QUANT module of the MsXelerator software. This module was developed to perform data analysis and data processing in GC/MS Metabolomics projects. A typical GC/MS workflow for Metabolomics applications includes a number of distinct steps; experimental design, sampling, sample preparation, data acquisition, data processing, data analysis and data interpretation. Besides Full Dataset Peak Picking, analysis and deconvolution, the GCMS QC/Quant module includes Quality control procedures and Analysis of Variance (ANOVA) to properly control each step in the workflow. The layout of the QC/Quant module is shown in Figure 1.

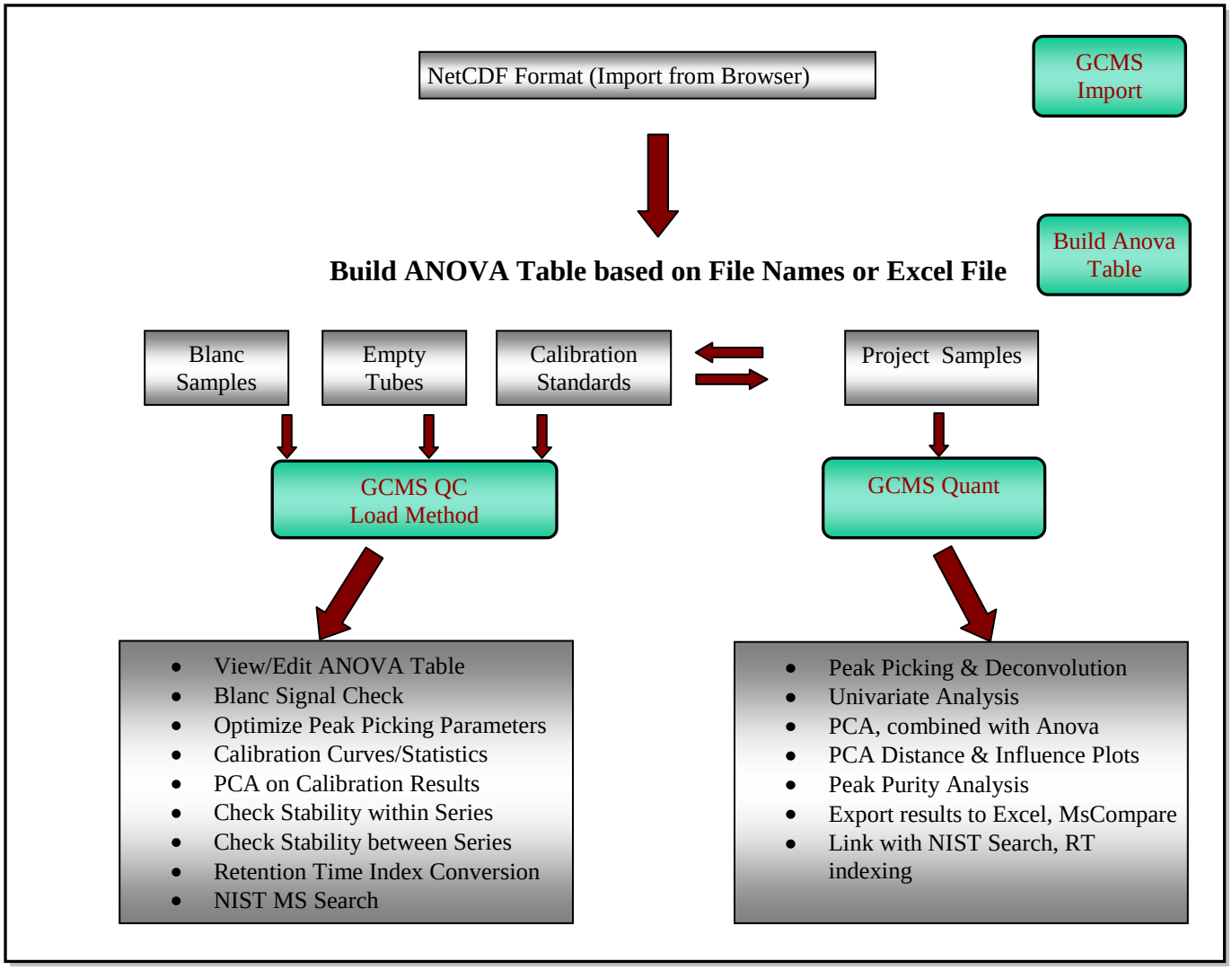


Figure 1: Overview of the GCMS QC Quant Module

GCMS QC: generating ANOVA Tables from Filenames

Based on the NetCDF Filenames the ANOVA Table can be automatically constructed. However, this requires the filenames to have strict names and fields. The ANOVA Table will have the following 12 fields:

ANOVA Table Fields:

1	queue_nr	a sequential number for the project files
2	filename	full filename of each sample
3	year	the year of the project
4	project name	the project name
5	sample id	sample id number or name
6	pan no.	pan number
7	blender no.	blenderkom number
8	tube no.	tube number
9	variety	"variety/ras" name of the sample
10	biorep	biological replicate number e.g. 1 or 2
11	techrep	technical reqplicate number e.g. 1 or 2
12	remark	a field to add comments (not part of the filename)

Sample Naming Rules:

The ANOVA Field values are generated from the sample disk names; below the rules for extraction of the different sample types and field values are explained:

1. Year and Project name: both fields are extracted from the Folder Name, e.g. 120607 Pommonde. There must by a **single space** between the date and project name. The year value is obtained from the first two characters of the date field (***12***0607).

2. QCL, QCS and QCB sample types are directly recognized from their filenames. QCL samples are the empty tubes, QCB samples are blank samples and QCS are the standards (four different samples) containing the calibration components). QCL samples have no tube number.

3. A QCB sample contains water only (blank water). The name could be e.g.: QCB buis14.d. Between QCB and a 'buis-number' there must be a space. There must be no space between "buis" and number.

4. There can be 4 different Calibration Standard samples, QCS1, QCS2, QCS3 and QCS4. The number relates to the standard number (different levels of the calibration components). After the name a space must be present, followed by the 'buis-number', e.g. QCS1 buis12.d.

5. A project sample should have a name from which all the field values can be extracted. An example for the name of a project sample is: **Bite_D24_Frankrijk 98475_M3 B1 T1 pan5 blender1 buis13.d**

   The ANOVA Fields are extracted from the name and should be separated by single spaces according to the following rules: The order of the Fields in Name must be: Ras, Sample ID, Biorep, Techrep, Pannumber, Blendernumber, Buisnumber and the fields must be separated by single space! If needed you can use underscores within the fieldnames Ras and Sample ID.

Ras;	use underscores for better naming
Sample_ID;	use underscores for better naming
Biorep;	name must be B1 or B2
Techrep;	name must be T1 or T2
Pannummer;	name must be pan1 or pan 2 etc.
Blendernr;	name must be blender1, blender2 etc.
Buisnr;	name must be buis12, buis15 etc.

Below the field assignment is given for the above example.

Anova Fields:	**Ras**	**Sample_ID**	**Biorep**	**Techrep**	**Pannr.**	**Blendernr**	**Buisnr.**
Sample name on Disk:	Bite_D24_Frankrijk	98475_M3	B1	T1	pan5	blender1	buis13.d

If automatic conversion of ANOVA fields from the file names fails, the samples type will be set to "Sample" having no further ANOVA details. The ANOVA table is build when importing of GCMS files has been completed and the user loads the list file (list of imported files) into the GCMS QC / Quant module.

GCMS QC: generating ANOVA Tables from Excel Layout File:

The filenames and ANOVA details can also be specified in an Excel File. This Excel file has a strict layout. Below the layout of this file is specified using an example.

Example of Excel Layout file:

GCMS Quant Anova Input Table											
Number of Samples		20									
ANOVA Fields		12									
Path to Files		D:\HZPC\Juni_2012_Project\120607 Pommonde									
Sample Type Definition											
Name	**Description**										
QCB	Blank Sample										
QCL	Empty Tube Sample										
QCS1	Standard1										
QCS2	Standard2										
QCS3	Standard3										
QCS4	Standard4….etc.										
Sample	Normal Project Sample										
Do not change Bold Fields Names !!											
Field1	**Field2**	**Field3**	**Field4**	**Field5**	**Field 6**	**Field7**	**Field8**	**Field9**	**Field10**	**Field11**	**Field12**
Queue No.	**Filename**	**Year**	**Project**	**SampleID**	Pan no.	Blender no.	Tube no.	Ras	Tech. Rep	Bio.Rep	**Sample Type**
1	QCL1	2012	Pommonde	Empty Tube							QCL
2	QCL2	2012	Pommonde	Empty Tube							QCL
3	QCB buis1	2012	Pommonde	Blank			1				QCB
4	QCB buis2	2012	Pommonde	Blank			2				QCB
5	QCS1 buis3	2012	Pommonde	Standard1			3				QCS1
6	QCS2 buis4	2012	Pommonde	Standard2			4				QCS2
7	Light_G23…	2012	Pommonde	98475_M3	7	2	17	Light	1	1	Sample
8	Light_G23…	2012	Pommonde	98475_M4	7	2	18	Light	1	2	Sample
9	Light_G23…	2012	Pommonde	98475_M5	8	2	19	Light	2	1	Sample

10	Light_G23…	2012	Pommonde	98475_M6	8	2	20	Light	2	2	Sample
11	Fine_B23…	2012	Pommonde	78505_M3	5	1	21	Fine	1	1	Sample
12	Fine_B23…	2012	Pommonde	78505_M4	5	1	22	Fine	1	2	Sample
13	Fine_B23…	2012	Pommonde	78505_M5	6	1	23	Fine	2	1	Sample
14	Fine_B23…	2012	Pommonde	78505_M6	6	1	24	Fine	2	2	Sample
15	QCL9	2012	Pommonde	Empty Tube							QCL
16	QCB buis7	2012	Pommonde	Blank			7				QCB
17	QCS1 buis9	2012	Pommonde	Standard1			9				QCS1
18	QCS2 buis10	2012	Pommonde	Standard2			10				QCS2
19	QCS3 buis11	2012	Pommonde	Standard3			11				QCS3
20	QCL11	2012	Pommonde	Empty Tube							QCL

Rules:

1. The first header line of the Excel file should not be changed; otherwise the program cannot recognize it.
2. At the third line you will have to specify the total number of samples to read.
3. The fourth line (second column) holds the number of ANOVA Fields. There is a maximum of 12 Fields and the first 5 fields should not be changed.
4. The 5th line holds the path to the samples. Be aware that all converted NetCDF files must be in the same folder. You cannot use different sub-folders.
5. Following is a section relating the samples to the Sample Type Field. In column number 12 of the ANOVA table you will need to specify the Sample Type for each sample. Check the definitions.
6. At line 19, the Field Headers are displayed (default names)
7. Below the Field Header, the actual Anova Field names can be entered. You can only change field names for columns 6-11. Do not change the other fieldnames.
8. Below the fieldnames, the actual data are entered.

9. The first column holds the queue number, the second column the filename. The name should exactly match the name of the file on disk (no extension). You will get a warning if files cannot be found (typo error?).
10. Columns 3,4 and 5 hold the year, project name and SampleID code. Do not change these Fieldnames.
11. Following are 6 columns for which you are free to set any name you like and the data belonging to it.
12. The table must be ended with the Sample Type info. You need to use the names specified in Sample Types Definition columns.

To create the ANOVA Table use: **Menu > File > Create Project from Excel ANOVA Layout Table**. Browse to the excel file and press Open. If samples cannot be matched with the files on disk, you will need to correct the Excel File.

When successfully converted, you will get a message. Next load the BLF file manually into the MsX-GCMS QC application.

Importing NetCDF GCMS Files:

Start the MsXelerator Program. From the Browser you can import a series of NetCDF files by selecting:
Menu > File > Import MS Vendor Data > NetCDF Batch. The following window will be displayed (Figure 2). Browse to the proper folder where your NetCDF files are located. You can select individual files or a series by making a selection with the mouse. When all files have been selected, **press Add**. Check the Selected Files List on the right. Be sure that you have only selected NetCDF files and no other files or folders. If correct, press **Import**.

If your NetCDF files are located in different subfolders, you might want to search for them by pressing SubDirs and the program will search all subfolders for NetCDF files.

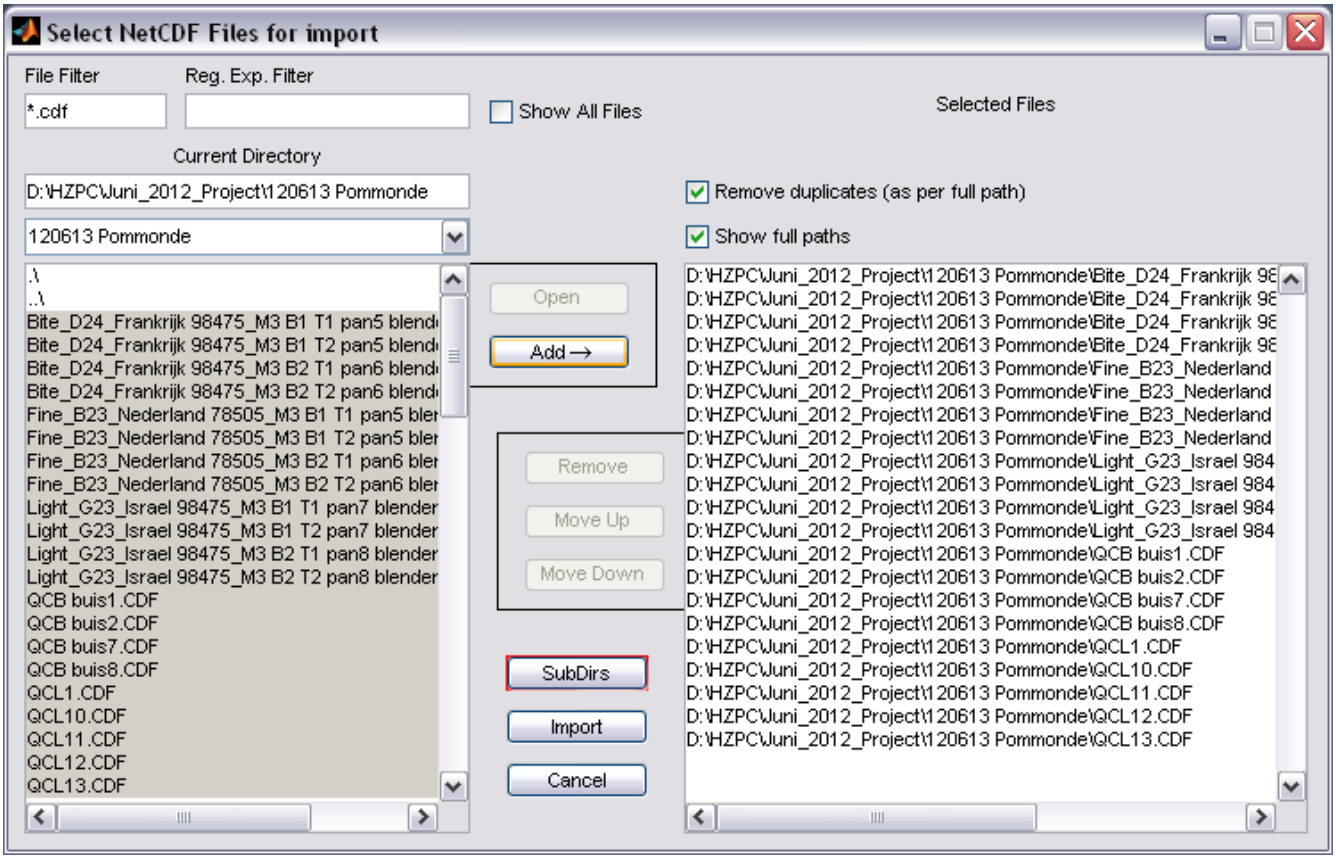


Figure 2: Browser Import GUI for series of NetCDF files

The Browser will start the conversion of the NetCDF files to MsXelerator format. When completed, you will be asked for a name and location for the List/Sequence File. This list file (*.blf) is a file containing the names and locations of the converted NetCDF files. This list file can be used directly by GCMS QC / Quant. Select a proper name and location. After that you will be asked if the list file should be sorted on the date of the NetCDF files. **Select Yes.** Importing is a onetime operation, the next time you start MsXelerator you can just open the MsX data file.

When importing is completed the Browser will display the last loaded sample from the list, like shown in Figure 3. From the Browser you can explore individual GCMS data files, extract and overlay extracted ion current (EIC) for any m/z value, view MS spectra, perform baseline correction, determine typical peak widths, noise levels and much more. For GCMS data you can link the displayed MS spectrum with the NIST Search program (see configuration on page 8), or first perform a proper baseline correction before starting the identification. For a complete overview of the Browser, please read the appropriate chapters from the MsXelerator manual, which can be found in the installation folder.

Browser Deconvolution of individual Peaks:
A new tool especially for GC/MS data processing is peak deconvolution. Using the Browser, it can be used for individually selected peaks to extract the “pure” MS spectrum for identification.

Select a Peak from the TIC window and extract the largest ion-fragment. Zoom in, if not already done. This EIC will be plotted to the middle window. From the menu select: **Task > GCMS Deconvolution.** You will be asked to first set the base ion. You should select the ion with the proper chromatographic shape for this peak. You can plot other ions by clicking on an m/z ion in the MS window. When satisfied, you will be asked for a minimum correlation threshold value for matching peaks. Ion fragments from the same component should co-elute with the selected peak. The co-elution is expressed by the correlation coefficient. A default a value of 0.8 will be used.
A cross-hair cursor will be shown. Click to the left and right of the peak. Then apply the Linear Baseline Correction when asked. Only peaks having the same shape will be part of the MS spectrum in the top window. You can now use this "deconvoluted" spectrum for component identification using a NIST Search (**Menu > Task > GCMS NIST Search**).

The above procedure can be automatically applied to a complete file using the MPeaks Module (Peak Picking / Deconvolution, see the chapter on MPeaks Data Processing/Clustering).

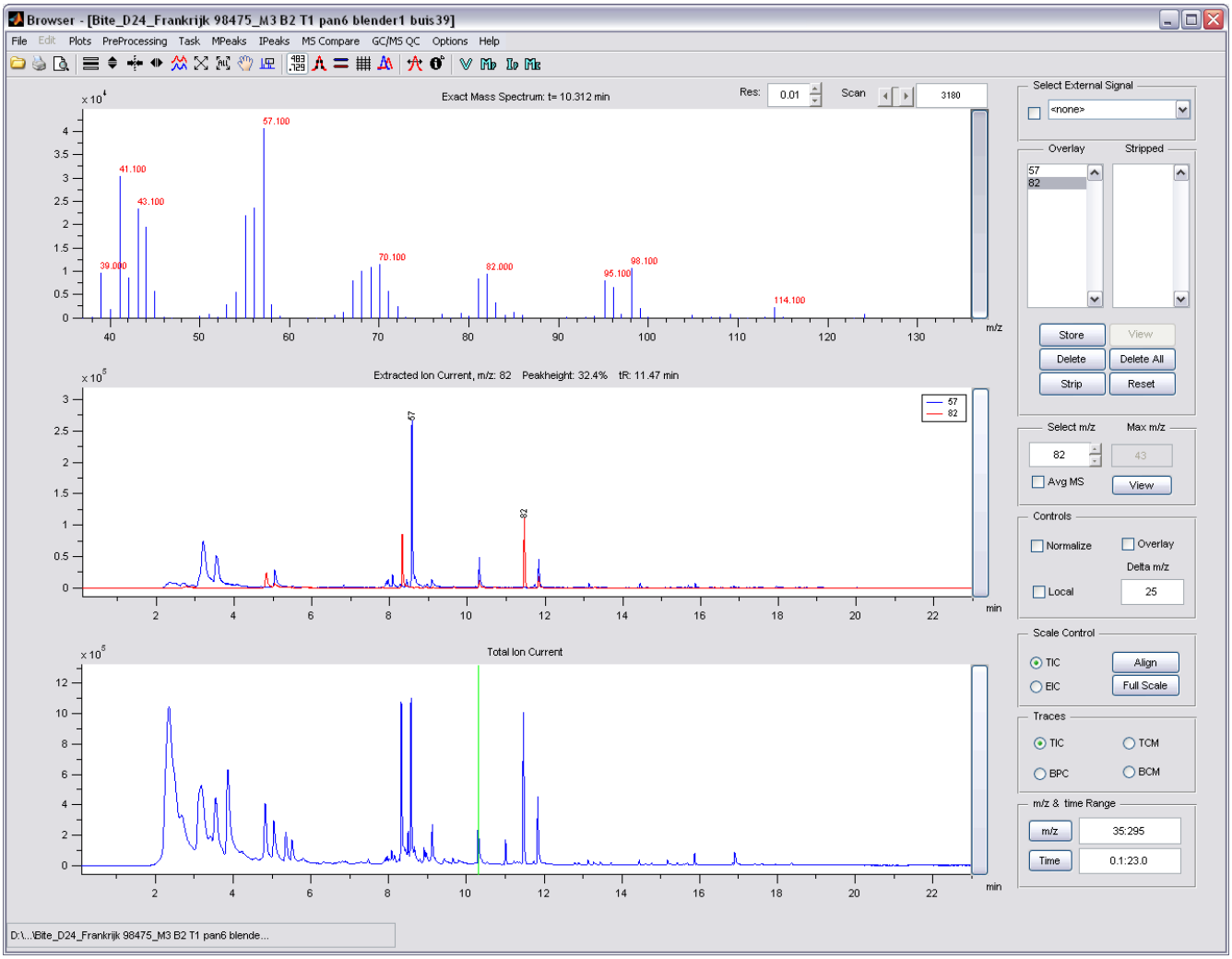


Figure 3: Browser Overview (TIC, EIC, MS spectrum). Overlaid are two Extracted Ion Currents in the middle window (m/z 57 and m/z 82).

Starting GCMS QC:

The GCMS QC module can be started from the Browser by selecting: **Menu > GCMS QC > Start GCMS QC.** The GCMS QC window will be shown, Figure 4. The GCMS QC layout, buttons and menu items are explained below.

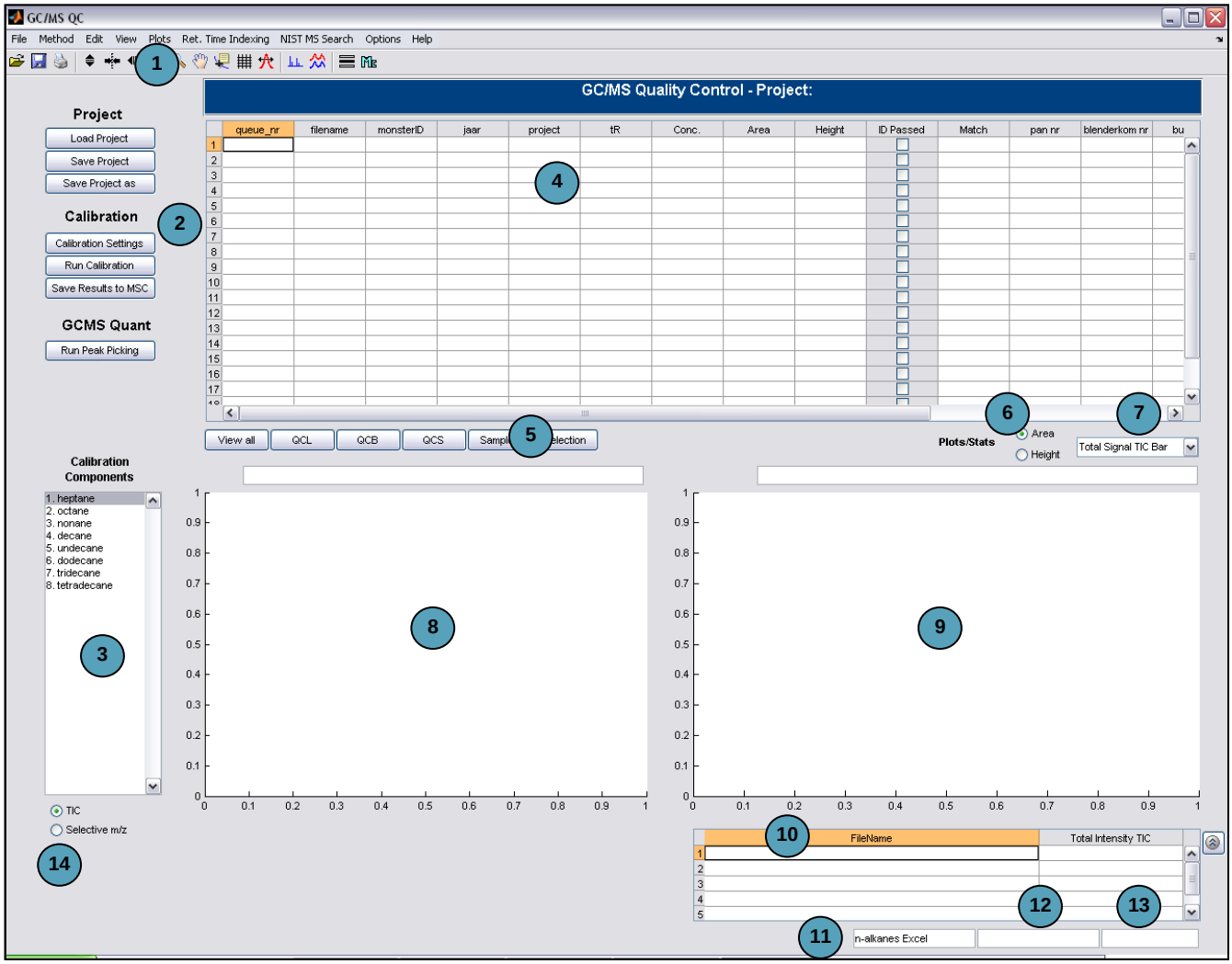


Figure 4: GCMS QC Overview

At the top of the window, below the menu items, a series of toolbar buttons are displayed (**1**). These buttons perform short cut actions. Below the toolbar buttons the Title of the project will be displayed.

Layout and Buttons:

2 In the upper left corner a number of Push Buttons are located. These have the following functionality:

Load Project: Loads a project from disk.
Save Project: Saves the current project using the same name (e.g. after calibration or changing some values).
Save Project as: Saves the current project using a new name.

Calibration Settings: Opens a new GUI for setting and optimizing Peak Picking Parameters for TIC analysis or for peak picking based on selective mass traces.
Run Calibration: Runs the calibration on the QCS samples for all calibration components using the peak picking parameters specified before. The method file defines how the peak picking is done: using TIC or selective mass calibration.

Save Results to MsC: Copies the QCS samples and the results from the calibration to the MsCompare module.

Run Peak Picking: Starts the GCMS Quant module for peak picking based on the TIC's of all or selected samples. **Currently only TIC Peak Picking is supported.**

3 In the lower left corner the list of calibration components is displayed. The list is based on the information found in the Calibration Matrix on disk (see Methods Section). When calibration has been performed, selecting another calibration component will update the ANOVA table with the results for the selected component. Any linked statistics plots will also be updated.

4 The table at the top shows the ANOVA table having 18 columns/fields. Besides the 12 fields from the ANOVA table explained on page 2, there are 6 additional columns: retention time, concentration, area, height, ID Passed and Match. These will be explained below.

5 Below the table, on the left, a series of buttons are displayed. These buttons control the display of the sample types, e.g. clicking on the QCS button will only display the samples having Sample Type "QCS" in the table.

6 On the right hand side of the table, two radio buttons are displayed; Area and Height. Clicking on either one of these will show the results (after calibration has been performed) based on area or peak heights.

7 To the right of these radio buttons a popup list box is displayed. The list box contains 9 different plots from which to Quality of Calibration can be studied. One plot is available based on the full TIC traces.

8 The plot region in the lower left of the screen is used to plot individual or multiple TIC traces. Traces can be zoomed, plotted in overlay or in stacked mode. Sample information regarding the selection is made visible in the edit box on top of this plot.

9 The plot region in the lower right region of the screen is used to plot the output of the calibration results. The plot type can be selected from the list box above the plot region. The comment field found above this plot holds specific information from the analysis. The lower right plot area is also used to display MS spectra of any selected sample at any retention time.

10 The table below this plot region displays any statistics information regarding the selected plot. The arrows next to the table can be used to enlarge the table to see more information if necessary.

11 The method field displays the name of the method (calibration components) that is currently in use.

12 The comment fields found in the lower right part of the screen holds the project name.

13 The Sample Info Field displays the number of samples in the project.

14 As a default the TIC traces will be plotted. However, if selective masses are defined in the method, you can also plot selective mass traces for each component.

Menu Items:

The items found in the Menu bar are explained below:

File:

1. Load/Create Project: loads an existing project from disk. If the ANOVA table does not exist, you will be asked if it should be created.
2. Create Project from Excel ANOVA Layout File. Reads the Filenames and ANOVA entries from Excel File
3. Append Project: to append an existing project on disk to the currently active project, select Append Project. The program will give a warning if calibration was performed using different peak picking parameters compared to the current project.
4. Save Project: saves the current active project including any changes to disk.

5. Save Project as: saves the current active project including any changes to a new project name.
6. Exit: exits the GCMS QC / Quant module
7. History List: you can quickly load a project from the history list.

Method:

1. Load/Apply new Calibration Method. A Method holds the calibration names, retention times, retention indices, formulas, selective masses, names of spectrum files for each component and the calibration levels for each calibration sample. A method is completely specified in a Excel file.
   Use the option to load a new method. Changing a method will delete the current calibration results.
2. Opens Excel with the currently loaded method file.
3. Select a method file on disk and view with Excel.

Edit:

1. Delete all Calibration Results: deletes the calibration results.
2. Delete Selected Samples: deletes the selected samples from the project. This action can only be performed when **View All Samples** has been activated.

View:

Not active yet.

Plots:

Not active, use the items found in the Plot List Box

Retention Time Indexing:

Select Retention Time Indexing after calibration has been performed successfully. The retention times for the calibration components can then be used for retention time index conversion for all samples and results.

NIST MS Search:

1. Run Search: To activate NIST MS Search, first plot a MS spectrum for a given sample at a selected retention time. To do this first click on the Spectrum icon in the Icon bar. Next, click on any peak in the displayed spectrum. The MS Spectrum will be extracted and plotted in the lower right corner. If retention time indexing was activated, the Retention Time index will be displayed in the title and also exported to the NIST Search program.
   Click on the NIST MS Search menu item to search to current displayed MS spectrum. If more than one TIC is displayed the program will select the one closest to the mouse position. The actual sample name will be displayed in the MS Spectrum title.
2. Save Current MS spectra as. The currently displayed MS spectrum will be saved to a location on disk with the choosen filename and extension *. msp. These MS files can be entered when building a new method. The spectra are used to check the purity of the calibration peaks by comparison with the stored spectra.

Options:

1. Set Plot Identifiers. Opens a window from which ANOVA fields can be selected that will be used for identification of selected points in the plots. If the **Identify on** icon is active, a cross will be displayed. Clicking on a single point will bring up the information box. Please try to click at the corner of displayed points.
2. Calibration: Select either Area or Peak Height. Can also be controlled directly from the radio buttons below the table.
3. View Project Calibration Parameters: Present a list of the peak picking parameters for the current project.
4. View Disk Calibration Parameters: the latest used peak picking parameters might be different from the ones used to calibrate the QCS samples in the active project. Select this option to view the list of peak picking parameters most recently used (disk).
5. Set Signal Threshold Empty Tubes. This option allows you to set the threshold for clean samples (Total TIC Signal) for the QCL samples. A horizontal line will be added to the Bar or Line Total Signal Plot. The threshold only applies to the current active session. As a default a value of 3.0E+007 will be used.
6. Apply Narrow Column Widths: used narrow column display for the ANOVA table to view more information on the screen.
7. QC Software Testing: for development only.

Help:

Displays this manual.

Toolbar Icon Buttons:

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

The icons from the Icon Bar perform the following actions:

1. Loads a project from disk
2. Save Project As
3. Printing, not active yet.
4. Scale TIC in Y-direction; when zoomed on re-scales the displayed tics in the y-direction only.
5. Not active yet.
6. Not active yet.
7. Reset to full scale plotting
8. Activate zooming; when zooming is not active click this button
9. Activate Panning:
10. Set data tip cursor to on; using the opion the information on the sample is displayed in TIC plots and some statistics plot.
11. Display a grid on both plots
12. Peak Width Measurement Tool; use this tool to measure the widths of peaks in scans and minutes. To be used for setting and optimizing peak picking parameters.
13. MS spectrum plotting; click this option to activate MS spectrum extraction. Next click on any position in the TIC window.
14. Plot the TIC traces in stacked mode.
15. Export currently active sample (single) to the Browser for more detailed analysis.
16. Export currently displayed samples to MsCompare for detailed analysis. Only the traces are exported, no result table. To export only the QCS (Calibration) samples and results to MsCompare use the push button: Save to MsC.

Setting Configuration for linking with NIST Search Programs:

Some manual actions should be performed to be able to link a MS spectrum from MsXelerator to the NIST Search Program. Perform the step below.

- Current installation folder for NIST is assumed to be C:\NIST08.
- Go to the subfolder MSSEARCH.
- Create a text file named: AUTOIMP.MSD
- Use an editor and put on the first line the following text: C:\NIST08\MSSEARCH\filespec.fil
- Save the file.

Building a Method File: Specification of calibration compounds and levels

A method file holds all the information necessary to perform calibration. This includes:

1. Number of calibration components and the number of calibration samples.
2. Component names.
3. Retention times and Retention indices.
4. Compound formula (optional) .
5. Selective mass for each of the component.
6. A link to the disk file that holds the mass spectrum for each of the components.
7. The concentration level for each component in the calibration samples.

The method file is created in Excel and follows a strict layout. Below an example for a method file is shown (n-alkanes).

GCMS Quant Method File		***03-05-2013***	***Attention: Keep two empty Lines between Section Headers!!***			
Method Name		n-alkanes Excel				
Number of Calibration Compounds:		8				
Number of Calibration Samples:		4				
Path to GCMS Spectrum files:		E:\MATLAB\R2006a\Gmd7\GCMS_Files				
Calibration Sample Definition (do not edit)						
Name	Description					
QCS1	Standard1					
QCS2	Standard2					
QCS3	Standard3					
QCS4	Standard4					
QCS5	Standard5					
QCS6	Standard6					
QCS7	Standard7					
QCS8	Standard8					
QCS9	Standard9					
Calibration Compounds						
No.	**Name**	**Ret. Time (min)**	**Retention Index**	**Formula**	**Selective Mass**	**MS Spectrum File name on disk**
1	Heptane	3.550	700	C7H16	100	heptane.msp
2	Octane	5.009	800	C8H18	114	octane.msp
3	Nonane	6.799	900	C9H20	128	nonane.msp
4	Decane	8.570	1000	C10H22	142	decane.msp
5	Undecane	10.210	1100	C11H24	156	undecane.msp
6	Dodecane	11.724	1200	C12H26	170	dodecane.msp
7	Tridecane	13.129	1300	C13H28	184	tridecane.msp
8	Tetradecane	14.448	1400	C14H30	198	tetradecane.msp
TIC or Selective Mass Calibration						
Use Selective Mass Calibration :		No				

Calibration Levels						
Calibration Units:		pg				
No.	**Cal File**	**heptane**	**octane**	**nonane**	**decane**	**Undecane**
1	QCS1	1250	1250	1250	1250	1250
2	QCS2	2500	2500	2500	2500	2500
3	QCS3	5000	5000	5000	5000	5000
4	QCS4	10000	10000	10000	10000	10000

Description of the Layout:
A method file is divided in several sections. Important: there must be **two empty lines** between the sections.

Line 1 holds a fixed text to recognize the method file: GCMS Quant Method File. Do not change this text.

The next few lines hold general information regarding the number of components and calibration samples.
Line 4 holds the method name in the 3th column.
Line 5 holds the number of calibration components in the 3th column.
Line 6 holds the number of calibration samples in the 3th column.
Line 7 holds the path to the folder in which the MS spectra for each of the components can be found. To create a MS spectrum file see page 20-21.

The next section, **Calibration Sample Definition**, show the predefined names of the calibration samples. You can have up to nine different calibration samples. **Do not change the names.**

The following Section, **Calibration Compounds,** defines the information needed for each of the components.
Column 1 holds the component number.
Column 2 holds the component name.
Column 3 holds the retention times in minutes.
Column 4 holds the retention index for this component.
Column 5 holds the formula.
Column 6 holds the selective mass for the component.
Column 7 holds the spectrum filename for each column. The path was defined in the top of the method file.

In the next section, it is defined whether calibration should be based on the TIC or on individual selective masses. If selective mass calibration is requested for each of the calibration components enter “Yes” in the third column.

The last section holds the concentration levels for all components in each of the calibration samples. In the example 4 calibration samples are specified and 8 components (only 4 displayed). In the line above, the concentration units can be specified e.g. pg or mg/ml etc.

The user can change the method for a project by loading a new method from disk. **Menu > Method > Load/Apply new Calibration method.** Any existing calibration results will be deleted.

Loading and Viewing a Project:

From the Menu, select Load Project (or press the push button Load Project). Locate the proper BLF file that holds all sample names. This file was created during import and should be present at the location of your samples or at the selected location when saving this file.

If no ANOVA table exists, you will be asked if it should be created. Select Yes. Please be aware that the ANOVA table will only be correct if the filenames have correct fields as explained on page 2.

The screen will look like shown in Figure 5. A few samples have been selected and the Total TIC bar plot was activated.

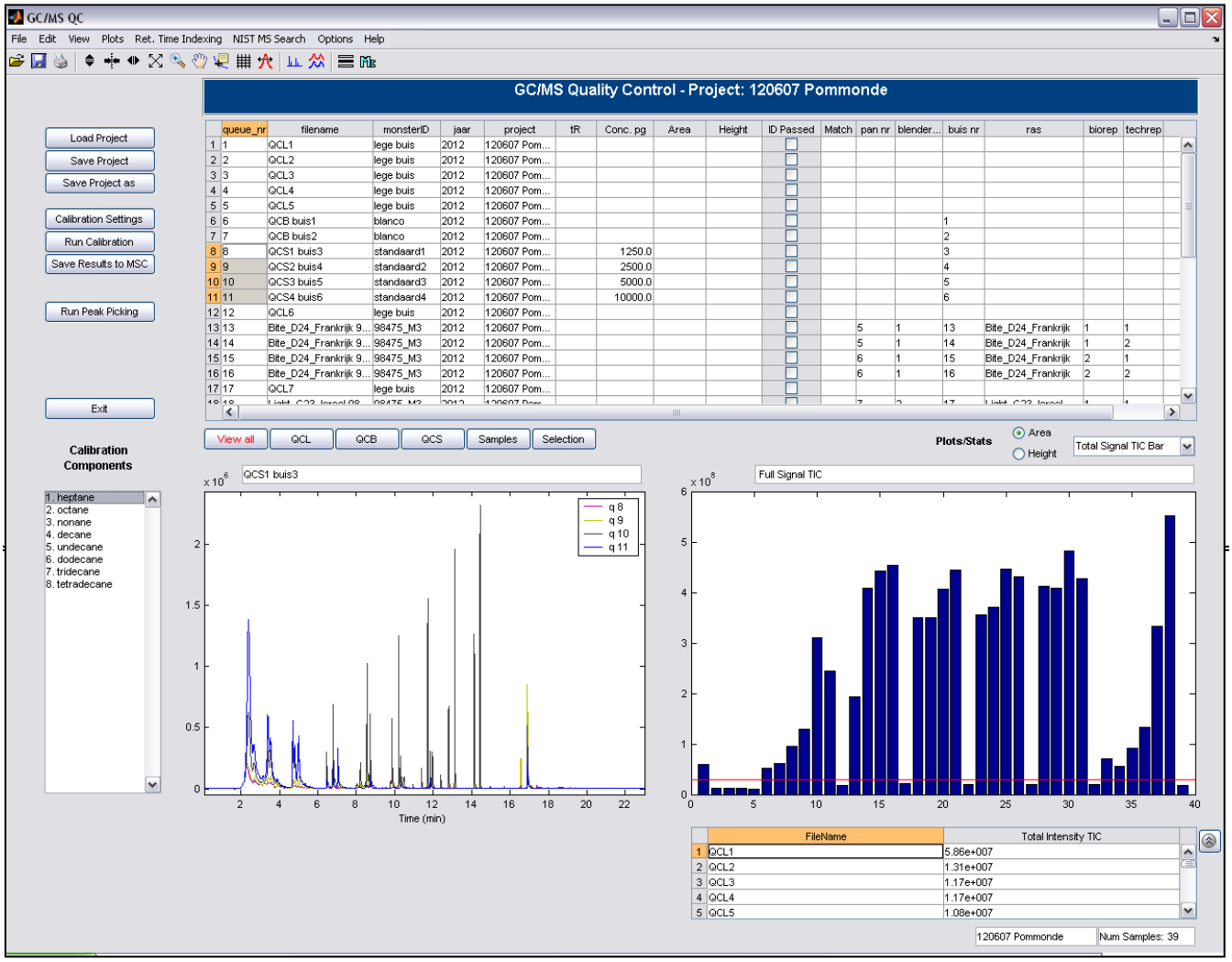


Figure 5: GCMS QC Screen after loading a project

You can zoom in on the TIC traces in combination with data marking to explore the different samples. Alternatively you can just you the mouse to select samples from the Table for plotting. The Plot Legends in the upper right corner of the TIC plot will display the Queue Numbers for the selected samples.

To view specific Sample Types click on the **Sample Type Buttons** below the ANOVA table. Now you can explore these group of samples in more detail.

Active Plots before Calibration:

Total Signal TIC:
To create a bar or line plot of the total TIC intensity for the currently displayed samples, select the Total Signal TIC Bar in the Plot List box. The bar plot will be created and the threshold level for clean samples will be displayed. Optionally create a line plot if preferred. The Intensities for each sample in the table will be displayed in the Statistics Table at the lower right. This plot is especially useful to check the QCL samples (Empty Tubes) for having low signals. Of course, the TIC chromatograms can also be plotted directly and checked for strange observations.

PCA score Plot on selected Samples:
You can create a PCA score plot based on TIC traces for all samples or any selection (QCS, QCL, QCB or Samples). Please be aware that your TIC traces should be well aligned to be used for PCA. Select View Samples and then click the PCA TIC Chromatograms selection from the Plot List box. The PCA result is shown below in Figure 6. You can immediately see that two samples seem to be strong outliers (row numbers 1 and 2 from the table, belonging to the "Ras": Bite_D24_Frankrijk. These samples should be checked in more detail). Before creating the plot you will be asked to define the grouping (colors). You can select any of the ANOVA fields for grouping the PCA results.
When no selection is made you will be asked to select between "SampleID" and "Ras".

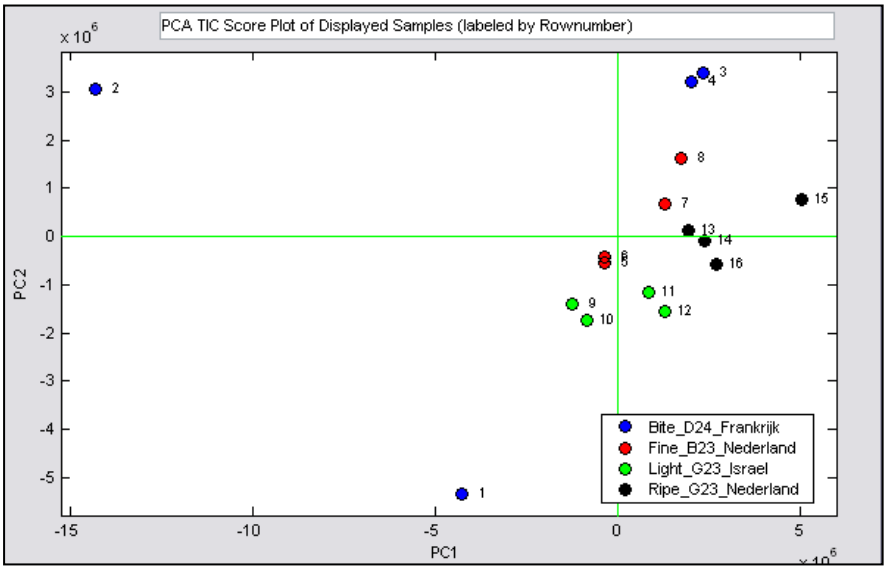


Figure 5: PCA score plot using the Samples only

All other plot types are not active without first running a calibration on the QCS samples.

Running a Calibration on QCS Samples:

Before running a calibration on the QCS samples based on the TIC or using selective masses, the Peak Picking Parameters should be optimized and saved. To enter the Peak Picking GUI, **press Calibration Settings**.

Figure 6 will be shown. The List box on the left displays all samples from the project. For calibration purposes you should select one of the standards, QCS1, QCS2, QCS3 or QCS4. In this case the QCS3 standard was used. The TIC will be displayed in the plot area. In figure 6, also the expected retention times (green lines) for the 8 calibration components are made visible. Furthermore, the **View Peaks button** was pressed, which shows all the detected peaks using the current peak picking parameters. Additionally, the **Check Purity of Calibration Peaks** option was activated. Any calibration peaks found will be marked with a solid green dot. If the peak does not match the purity of the standard on disk, a red dot will be displayed.

When using selective mass calibration, select the m/z radio button and enter the mass. The plot will be updated with the Extracted Ion Current of the selected m/z value. Be aware that the peak parameters for TIC and Selective Mass can be quite different, especially regarding the noise level and the absolute intensity level.
Below the Sample List box, all the Peak Picking Parameters are shown. The last used and saved settings will be active and displayed. The peak picking algorithm parameters are explained below.

Basic Peak Picking Parameters:

Peak Width: the width in scans for separating closely eluting peaks. If you use a very narrow width, noise on broader peaks could be mistakenly seen as real peaks. For Calibration purposes the calibration components are normally separated by quite some distance. This parameter will be more important for peak picking of the actual project samples (GCMS Quant).
Intensity Threshold: specify the absolute intensity threshold. Peak below this level will be deleted.
Noise Level: this is one of the most important peak picking parameters. A peak should have an intensity compared to its baseline larger than the noise level.
S/N Threshold: you can specify the Signal/Noise ratio. Peaks smaller than this value will be deleted.
Min and Max FHWW: specify the minimum and maximum number of scans at halve height for each peak. Peaks narrower or broader than the specifications will be deleted.
Smoothing: optionally you can apply smoothing to the TIC or EIC trace to better estimate noisy peaks. Be careful not to include noise. The Smoothing Width (scans) determines the amount of smoothing applied.

Peak Purity Parameters:

Two parameters control how calibration components are found and checked; the time window and the match limit. From the Calibration Matrix the expected retention times for the calibration peaks (n-alkanes) are known. These positions are fixed and stored on disk. You can visualize the calibration retention times by activating the **View tR of Cal Peaks**. Vertical green lines will be displayed where the calibration peaks should elute.

A calibration peak can only give a positive match if the detected peak is within a certain window around the expected peak position and if the MS spectrum of the calibration compound matches that of the detected peak. The time window controls the amount of shift that a calibration peak may shift compared to the retention times in the Calibration Matrix. The match limit expresses the minimum correlation coefficient for the MS spectrum of the calibration component and detected peak. Only a peak detected in e.g. +/- 0.5 minutes from the expected retention time and having a correlation coefficient larger than the match (0.8) value will be a real match.

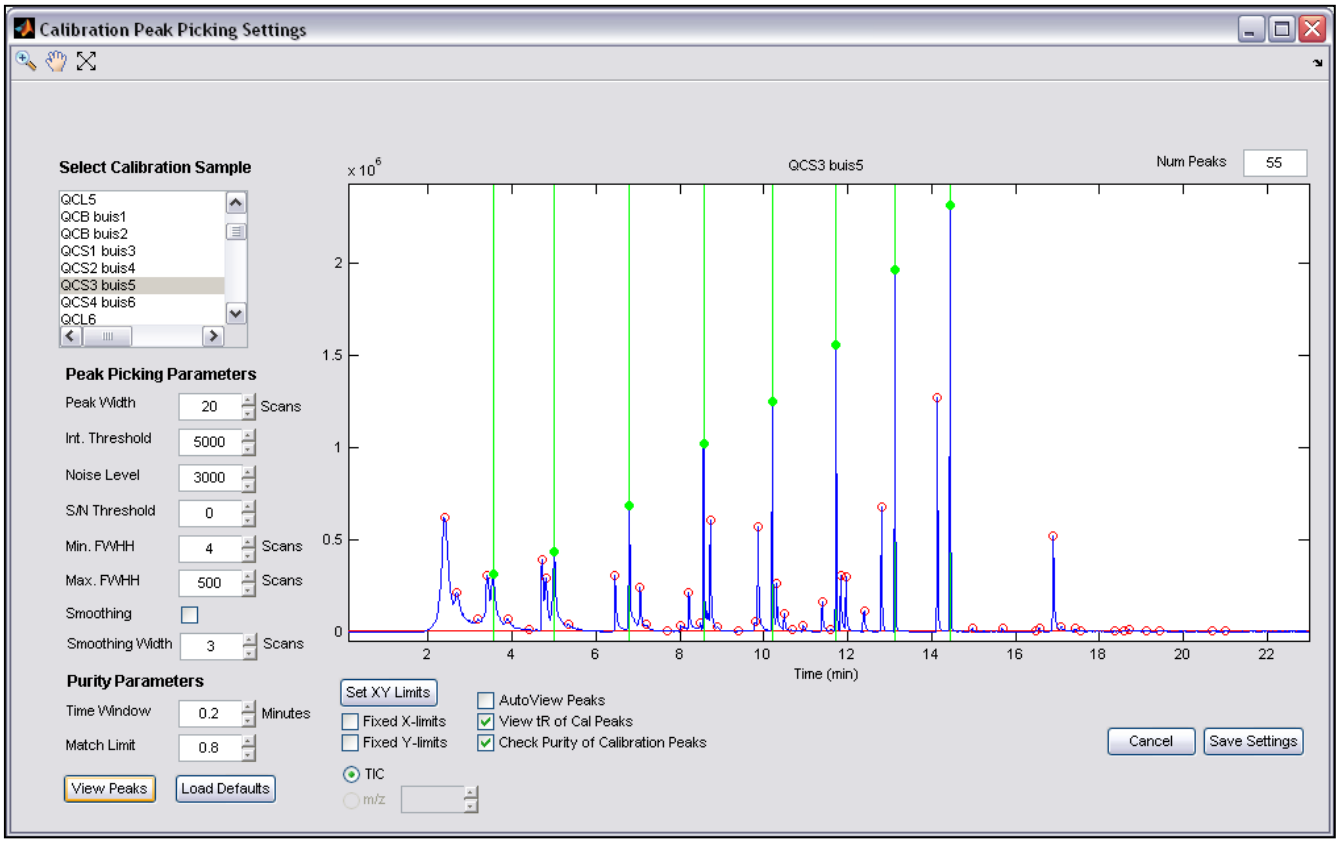


Figure 6: Peak picking overview (TIC). Activated are the expected retention times for the calibration peaks and the check Purity of Calibration Peaks. Matching calibration peaks are marked using solid green dots (8 peaks) and non-matching calibration peaks will have solid red dots. All other detected peaks are marked using open red circles.

From figure 6 is can be seen that using these settings in total 55 peaks have been detected based on TIC calibration. All calibration peaks from the n-alkane series were found.

Options to be used in Peak Picking GUI:
The GUI has three Icon Bar Buttons: one for zooming in, the second one can be used to measure the peak widths and the third icon is used to return to full scale plotting.

View Peaks: performs and displays the peak picking results for the currently selected parameters. The number of detected peak for the current sample will be displayed in the N**um Peaks** box at the top right of the window. In the example 55 peaks have been detected and all calibration peaks conform to the Matching Limits.
Load Defaults: load predefined default setting for GCMS TIC peak picking from disk.

AutoView Peaks: when this option is activated, automatic updating of the peak picking results is performed when changing one of the parameters or when another sample is selected from the list.
View tR of Cal Peaks: show the expected retention times for the calibration components.
Check Purity of Calibration Peaks: show the detected calibration peaks. A green solid dot means a positive match (correlation coefficient larger than the Match Limit). A red solid dot means that a peak has been found in the window, but that the correlation coefficient is smaller than the Match Limit. If more peaks are found within a calibration window, the one having the largest correlation coefficient (MS spectra) with the spectrum on disk will be used.

Set XY Limits: after zooming in you can press this button. The displayed plotting limits will be remembered. When you now change to another sample, the TIC is zoomed using the stored limits. This is easy for checking peak picking results for small peaks in a restricted region.
Fixed X-limits: only fixes the x-axis for scaling.
Fixed Y-limits: only fixes the y-axis for scaling.

TIC/EIC: Select between TIC and selective m/z display. When a selective m/z mass trace is plotted, also set the m/z value.

Cancel: return to the GCMS QC module without making any changes to the peak picking parameters.
Save Settings: return to the GCMS QC module and stores the current peak picking parameters.

Experiment with the peak picking settings and be sure to detect the calibration peaks before returning to the GCMS QC module. Click **Save Settings** and you will be returned to the GCMS QC module. You can now start the calibration procedure for the QCS samples. The Peak Picking GUI will be more important when peak picking has to be performed for the project samples (GCMS Quant).

When returned to the GCMS QC module, **press Run Calibration**. The results will be added to the ANOVA Table. Below an example is shown of the updated table. Displayed are the QCS samples and the results for component Hexane. If you select another calibration component from the Component List, the table will be updated with the results for this component.

	queue_nr	filename	monsterID	jaar	project	tR	Conc. pg	Area	Height	ID Passed	Match	pan nr	blenderkom nr	buis n
1	8	QCS1 buis3	standaard1	2012	120607 Pom...	3.554	1250.0	7244.0	48094.0	☐	0.661			3
2	9	QCS2 buis4	standaard2	2012	120607 Pom...	3.554	2500.0	13774.0	101567.0	☑	0.934			4
3	10	QCS3 buis5	standaard3	2012	120607 Pom...	3.544	5000.0	34470.0	316028.0	☑	0.999			5
4	11	QCS4 buis6	standaard4	2012	120607 Pom...	3.531	10000.0	44654.0	419344.0	☑	0.988			6
5	35	QCS1 buis9	standaard1	2012	120607 Pom...	3.57	1250.0	7517.0	51106.0	☐	0.656			9
6	36	QCS2 buis10	standaard2	2012	120607 Pom...	3.547	2500.0	12923.0	96537.0	☑	0.931			10
7	37	QCS3 buis11	standaard3	2012	120607 Pom...	3.541	5000.0	33355.0	302281.0	☑	1.0			11
8	38	QCS4 buis12	standaard4	2012	120607 Pom...	3.544	10000.0	43890.0	408018.0	☑	0.993			12

Table 1: updated ANOVA table showing calibration results for hexane.

The table contains the detected retention time, concentrations, areas and peak heights for hexane for each of the 8 QCS samples. The ID-Passed column shows if the MS spectrum matches the calibration component (on disk). The actual match factor is displayed in the column Match. It appears that 2 QCS samples (QCS1 buis3 and QCS1 buis9) do not have proper correlation. These samples have the lowest concentration levels, however it appears that for other components at higher concentrations also one of the calibration samples shows strange behavior.

Figure 7 shows the QC overview when component tetradecane was selected. From the table it can be seen that the peak

for tetradecane was not found in the fourth QCS sample. The retention time displays the value NaN (not a number), which means that the peak was not found. In this case the area and height will be set to zero, the ID Passed will be set to fail and the Match Factor will be zero.

When selecting all peaks, the TIC plot will automatically zoom in on the retention time of tetradecane. Indeed, it can be seen that only 7 samples have signal in this range. The 4th sample does not have a peak at this location. Furthermore, from the calibration curve it can be seen that at the highest concentration level (10.000 pg), the QCS sample with queue number 11 is a clear outlier. Below the calibration curve the statistics table for this component is listed. Five entries are displayed: Component number and name, the R^2 value for the calibration line, the root mean squared error (RMSE) , and the two regression parameters for the calibration curve (B0 and B1). From the calibration plot and the R^2 value it is obvious that we have a bad calibration.

	Regression Parameters	Value
1	Component	8. tetradecane
2	Rsquared	0.284
3	RMSE	35843
4	B0 parameter	6763.717
5	B1 parameter	5.836

Something clearly happened to this calibration sample. Many of the Statistics plots will lead to the same conclusions, which is actually the task of the GCMS QC module.

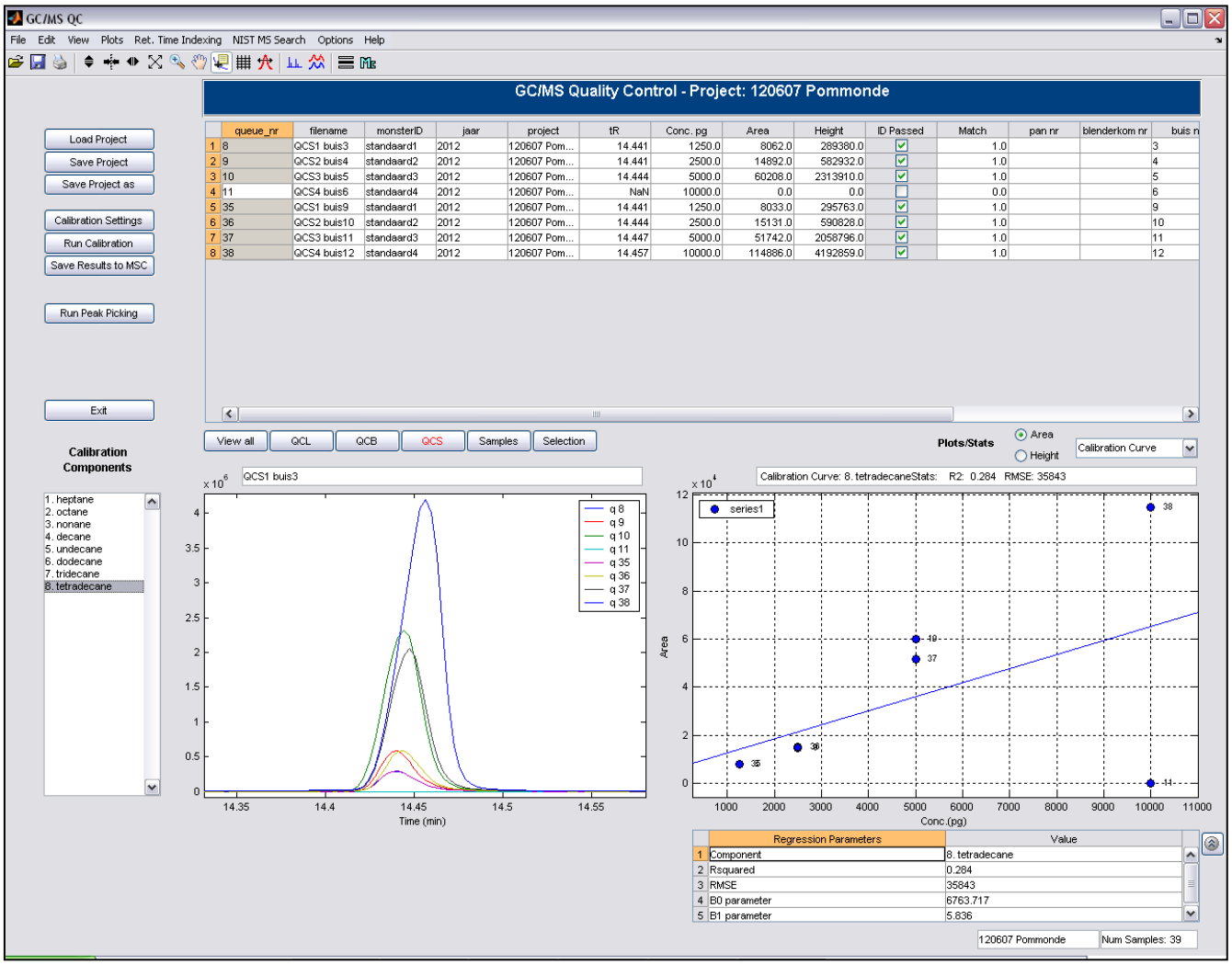


Figure 7: GCMS QC overview for component tetradecane. Displayed are the TICS for the 8 QCS sample, the ANOVA table and the calibration line / statistics table for tetradecane.

GCMS QC Plots available after Calibration:

The other type of plots that can be made after a calibration has been performed are: Total TIC Bar and Line, Calibration Curve and Calibration residuals, individual area/height plot, stability plot (% Coefficient of variation or Averages for each standard), a PCA score plot on the calibration results for all components and a PCA score plot based on the TIC chromatograms for the QCS samples. Below, examples are given for each of these plots:

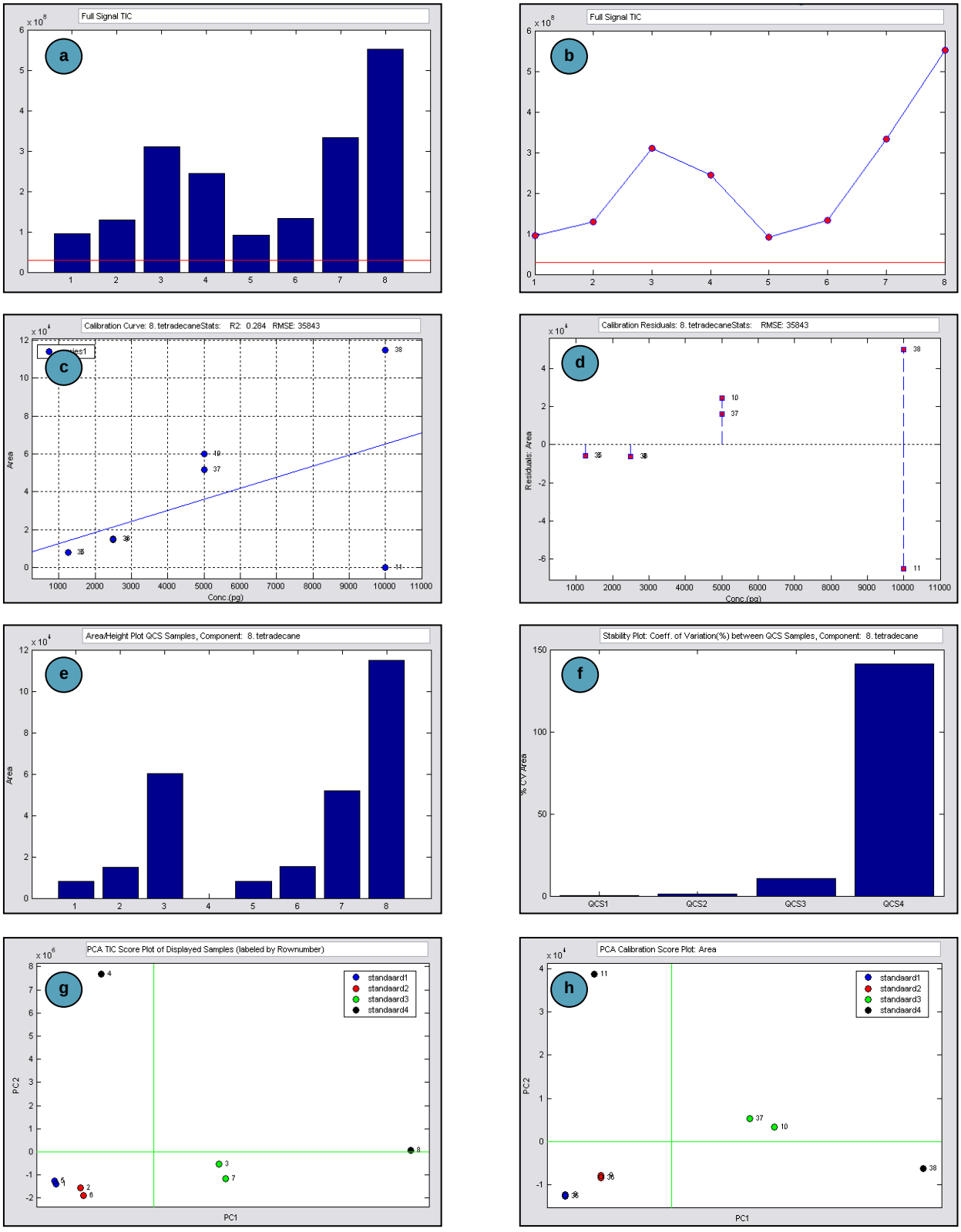


Figure 8: Overview of different QC plot types after calibration has been performed.

Figure 8a: displays the full TIC intensity for each of the 8 samples. From this plot it can be seen that the fourth sample has a low intensity in the series (the series is ordered).

Figure 8b: same as 8a, but presented as a line plot.

Figure 8c: the calibration curve for component tetradecane. At high concentration levels the fit is very bad. It can be seen that no signal was found for sample 11 at the highest concentration level. The title of the plot shows the calibration component, the R^2 value and the Root Mean Square Error (RMSE).

Figure 8d: shows the regression residuals as a function of the concentration level.

Figure 8e: individual peak area or peak height values as a function of the samples.

Figure 8f: stability plot for the current series. You can select to plot the %CV or the mean values. The bars show the % coefficient of variation or means calculated from replicate QCS samples. The x-axis holds the four calibration samples QCS1, QCS2, QCS3 and QCS4. For each pair of samples the %CV and the means are calculated. Figure 8f shows the %CV for each QCS group. It can be easily seen that a problem exist in one of the QCS4 samples. The Statistics table will report the %CV or the means for the four QCS standards.

Figure 8g: PCA score plot of the full calibration table (all QCS samples, all calibration components). From this figure is can be see that sample 11, standard4 is a strong outlier. Attention; the numbers in this plot refer to the row numbers in the full Anova table. For a single project these numbers are the same as queue numbers. However, for combined projects this is no longer the case.

Figure 8h: shows the PCA score plot when performed directly on the TIC traces. The plot basically gives the same information as a PCA run on the calibration matrix.

Setting Anova Fields as Plot Data-Tips:

The ANOVA information for the different samples in figures 8a-8h can be made visible by activating the View Anova Information Icon button (for some plots this feature is already active at start). You can select which information should be displayed when clicking on one of the samples (points/bars) in the plots.
To control the type of ANOVA information displayed, select: **Menu > Options > Set Plot Identifiers.** This opens the window shown in Figure 9. You can select the ANOVA fields to be shown. Press OK when done. Now click on one of the bars or points in the plots. As an example figure 9b shows the Total TIC signal for the samples. When clicked on the first sample a text box with the selected ANOVA info is shown. Right click on the data-tip for options or to delete the data-tip.

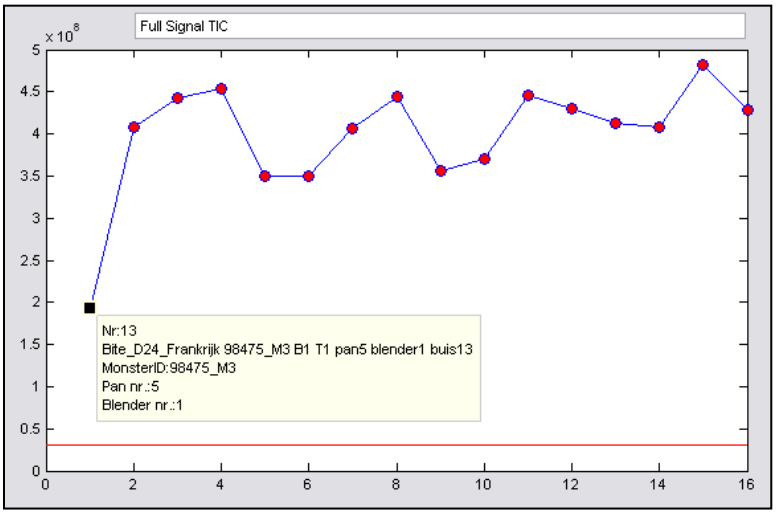


Figure 9a and 9b: Select the ANOVA info to be displayed and a example for one of the plots showing the data-tip using 5 of the ANOVA fields. The data-tip also operates on TIC traces when multiple traces have been plotted in overlay.

Plotting TIC Traces: in overlay or stacked:

You can plot the TIC or EIC traces in overlay or stacked, see Figure 10. For a stacked plot, the plotting order is from top to bottom. The first sample will be plotted at the top, the last selected sample to the bottom. Making a new selection from the table will automatically update the plot. To plot selective mass traces, activate **Selective m/z Traces.**

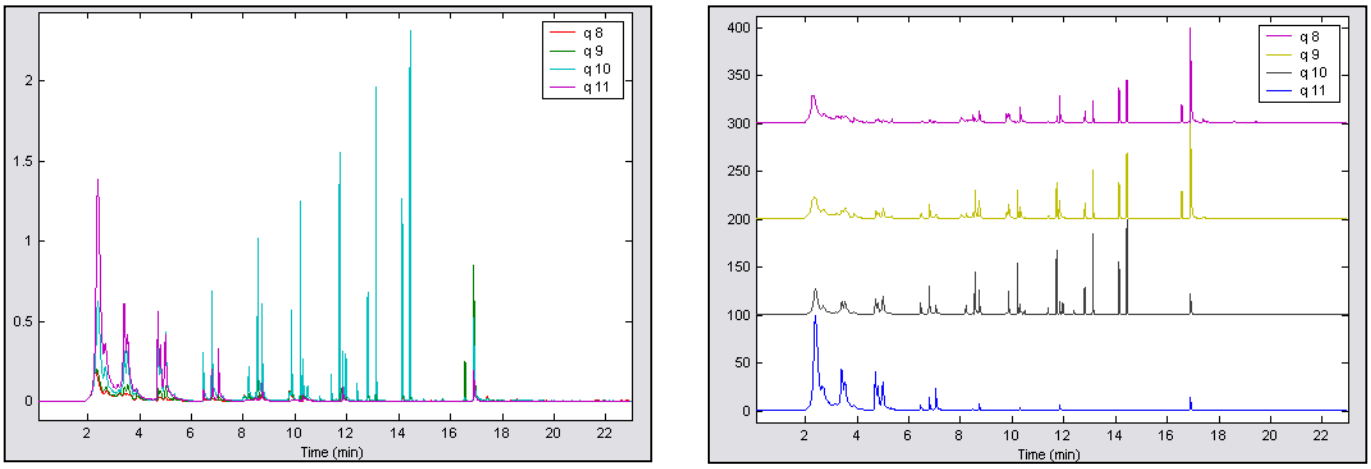


Figure 10, TIC traces in overlay or stacked plotting mode.

Plotting MS spectra / Spectrum Search using NIST:

To plot the MS Spectrum at a certain position for one of the selected samples select the MS plot icon from the toolbar. Then click at a position of a peak in the TIC window. You can perform this action when multiple TICs have been plotted; the sample closest to the click position will be selected. Probably you will get the following warning.

This means that you will first have to accept the calibration results, otherwise the program cannot calculate a RT index values for the current selected retention time. You will get this warning only once. To accept the current calibration results for retention time indexing, select: **Menu > Ret. Time Indexing > Accept Calibration Results for Rt Indexing.**

Now you are ready to plot the MS spectrum including the RI value, see Figure 11. The title of the MS spectrum plot displays the selected time in minutes, the RI value and the Queue Number for the selected sample. If you want to search to current spectrum in the NIST library, select **Menu > NIST MS > Start Search.**
Please be aware that the raw plotted spectrum will be exported to the NIST search. If you need more dedicated baseline correction or deconvolution procedures during spectrum extraction, please link this file to the Browser. The Browser has more specialized tools for deconvoluting MS spectra, see page 4.

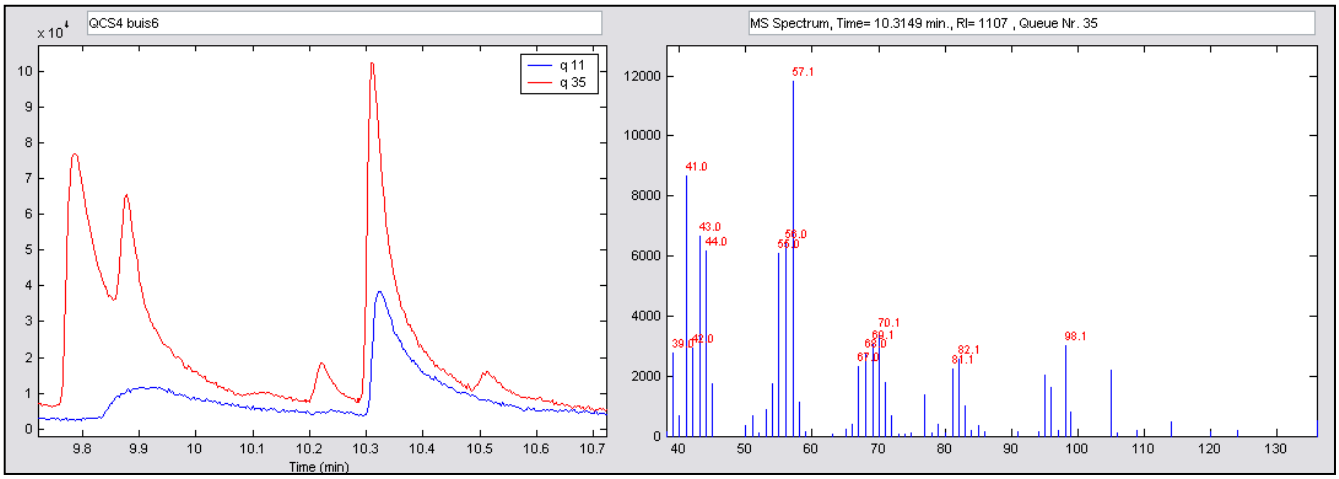


Figure 11: Direct MS spectrum extraction for NIST Search Identification.

Saving the current Spectrum to Disk:

To save the current displayed MS spectrum to disk, select: **Menu > NIST MS Search > Save current Spectrum as.** The extension will be .msp. The MS spectrum files are needed to perform peak purity checking.

Exporting/Linking Files to the Browser or MsCompare:

Any sample from the table can be explored in more detail by loading it into the Browser module. Select one single sample from the table and press the Browser Icon on the Toolbar. The sample will be loaded into the Browser.

To load the displayed samples in the table to the MsCompare module, press the MsCompare Icon on the Toolbar. The MsCompare module will start and the displayed samples are loaded into MsCompare. MsCompare will automatically display the Sample ID names in the Sample List. You can optionally switch to viewing full filenames if preferred.

Merging Projects:

You can append as many projects as you want. After appending projects some special plots become available to see the stability of calibration results between series / over time.

To append a project to the current project select **Menu > File > Append Project.** If both projects contain calibration data for which the settings were different, you will get a warning. In this case you might want to re-calibrate the QCS samples. However, as long as your calibration peaks are estimated correctly, you don't need to recalibrate.

You will get a message if appending was successful. The new number of samples in the project is shown in the lower right corner of the screen and the split in series can be seen from the table by checking the project name field. Attention: if you now save the project, the old project will be overwritten. Please save the combined projects using a new filename.

Appended Projects and QC Plots:

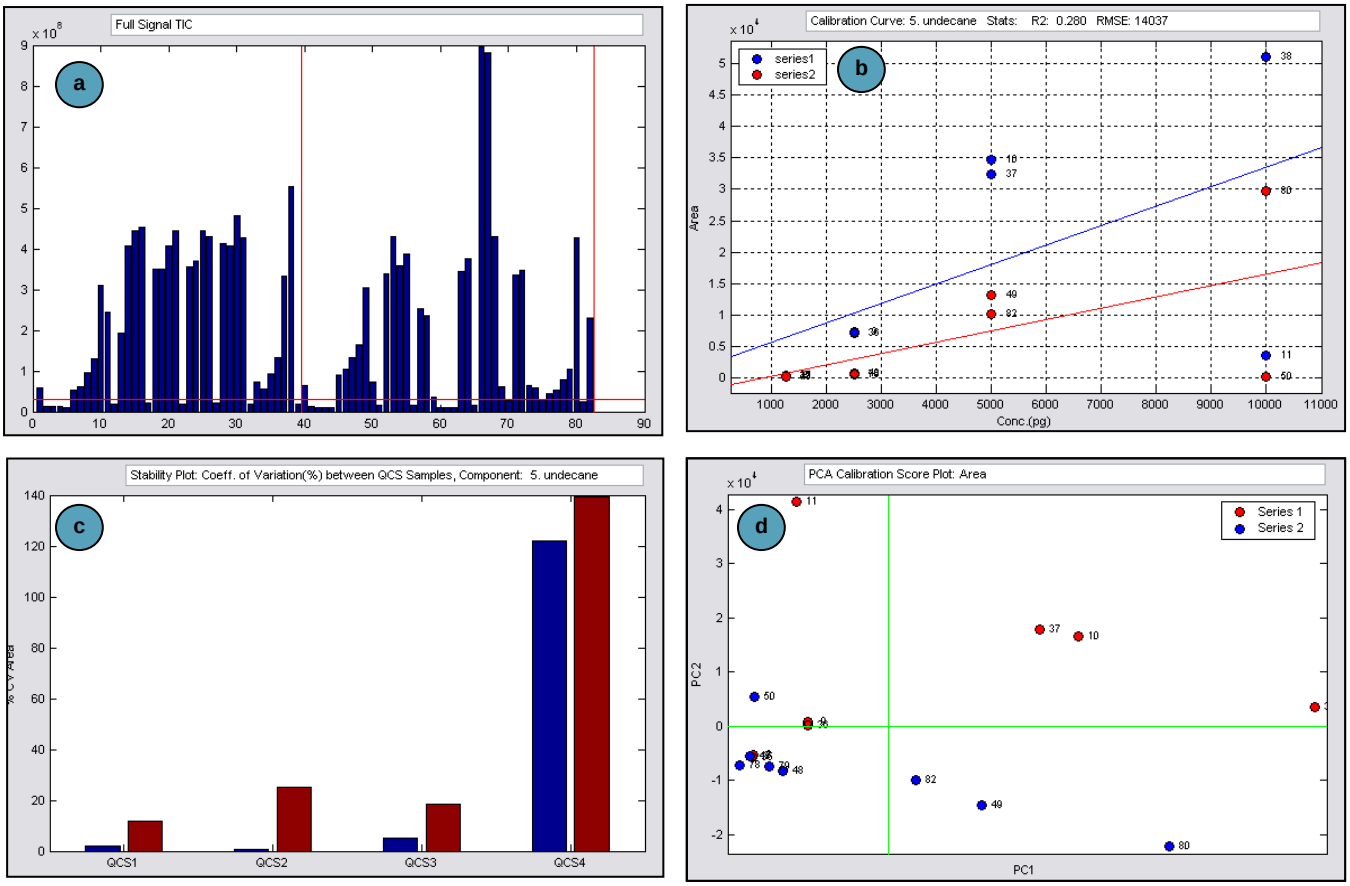
The combined project will keep the original queue numbers. Most of the plots will now show markings which data are from the first or the second project. Some examples will be given below (Figure 12 a-e).

Figure 12a shows the Total TIC Bar Plot for two series. The first series has 39 samples, the second 43 samples. A red vertical line is drawn at the position where a new series (project) starts.

Figure 12b shows the calibration line for component undecane. A least squares line is fitted for both series. However, the statistics table will show the result for the combined data (single regression line results). From this figure it is seen that the two calibration lines are quite different. The second series in general shows much smaller peak areas and peak heights. This can also be seen when overlaying the same QCS sample from the two series.

Figure 12c shows the within series stability plot. The stability for each series is calculated and both results are shown in one single plot. The first series uses blue bars; the second series uses red bars. It is seen that the %CV values for the second series are much larger for the first three QCS samples. This is not true for the last QCS4 group, but it was already shown that within the first series an outlier in QCS4 was present. The statistics table below the plot will hold the %CV values for each series as a separate column, see figure 12e.

Figure 12d displays the PCA score plot of the calibration results. Each series is plotted with different colors. It can be seen that the calibration results (height and area) for both series are quite different.



e

	QC Sample Type		CV-Value (%) Area 2 Series
1	standaard1	2.18	11.81
2	standaard2	0.76	25.22
3	standaard3	4.9	18.23
4	standaard4	122.22	139.28

Figure 12: GCMS QC plot for two combined projects

When projects are combined, a new plot type becomes available: **the Between Series Stability Plot**. You can select three different presentations for plotting the calibration results between series; comparison of the RMSE between series in a Bar plot; comparison of the R^2 as a Bar plot and the individual regression residuals for each standard sample for both series. The regression will be performed for each series separately. Figure 13 shows all three plots.

From the first figure it can be concluded that the calibration fit for the first series is worse compared to the second series. This is probably due to the outlier detected in series one. The R^2 value is somewhat better for the second series; both have

very low values due to the presence of outliers. The third plot shows that individual residuals for the second series are smaller, which is in line with the lower RMSE errors. Statistics for both series can be found below each plot.

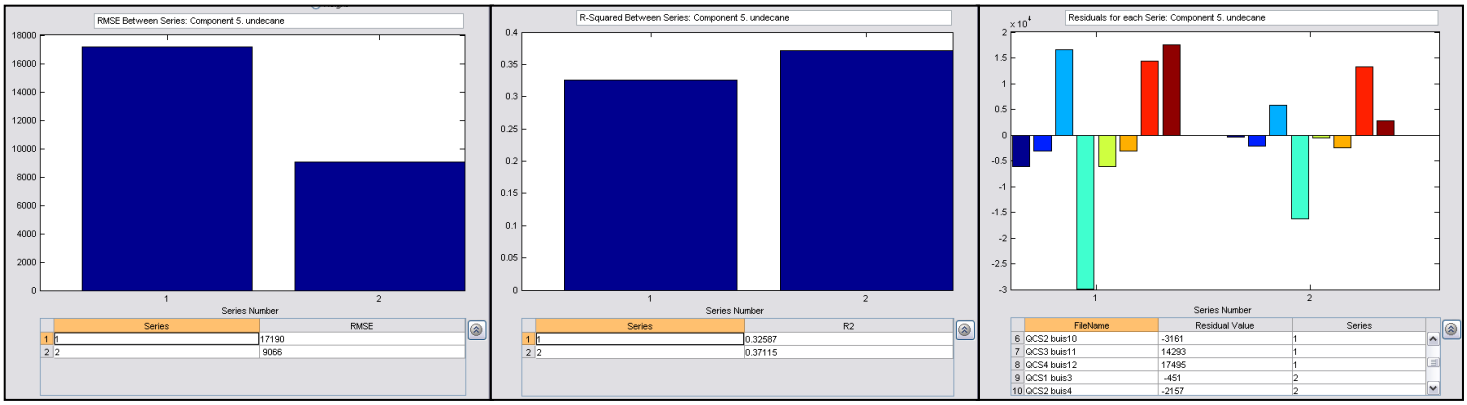


Figure 13: Stability plot between series; comparison of results from two merged projects. RMSE, R^2 and residuals.

You can optionally re-calibrate and study results after deleting the outlying samples from both series. To delete outliers first return to the View All Table, next select: **Menu > Edit > Delete Selected Samples.**

General remarks/tips:

- To activate the TIC plot or the Statistics plot just click in the plot area. You can then use the scaling buttons like full scale plotting or adding/removing a grid.
- Not all plots support the data-tip feature (merged projects).
- Stacked plots are always plotted having a normalized scale. The plot is useful for comparison of relative differences, to compare absolute differences use overlay plotting or export to the data to the MsCompare module. More options for plotting are available here.

GCMS Quant:

In the first part of the document the GCMS QC module was explained. The purpose was to control and check instrument performance (over time) based on calibration and control samples.

The figure below outlines the structure and features of the GCMS Quant module, for processing of Project Samples. In principle the module can be used for peak picking of all samples in a project, but this manual will focus on processing of the project samples only.

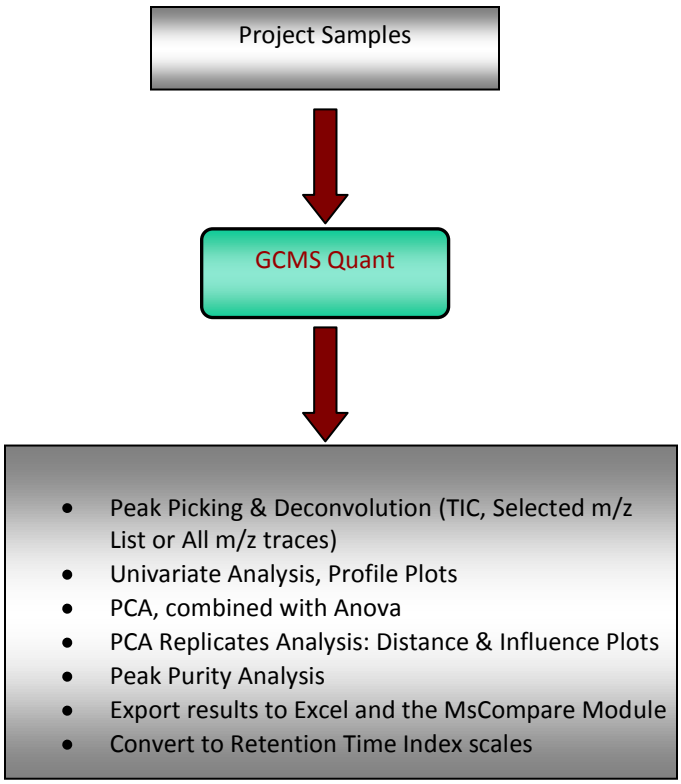
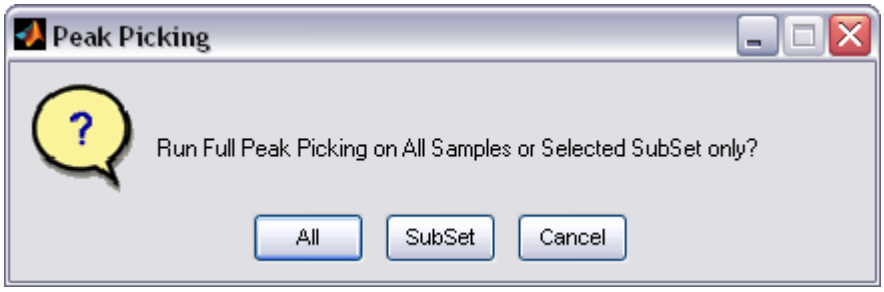


Figure 14: Overview of GCMS Quant Module

From the GCMS QC module select view "Samples" only, by pressing the **Samples Selection** button. Next, start the GCMS Quant module by pressing the **Run Peak Picking** button. You will get the following question below. You can of course use all Samples from the project or only the "Samples" subset. Press **SubSet** as we only want to perform peak picking on the project samples.



The GCMS Quant module is shown in Figure 15.

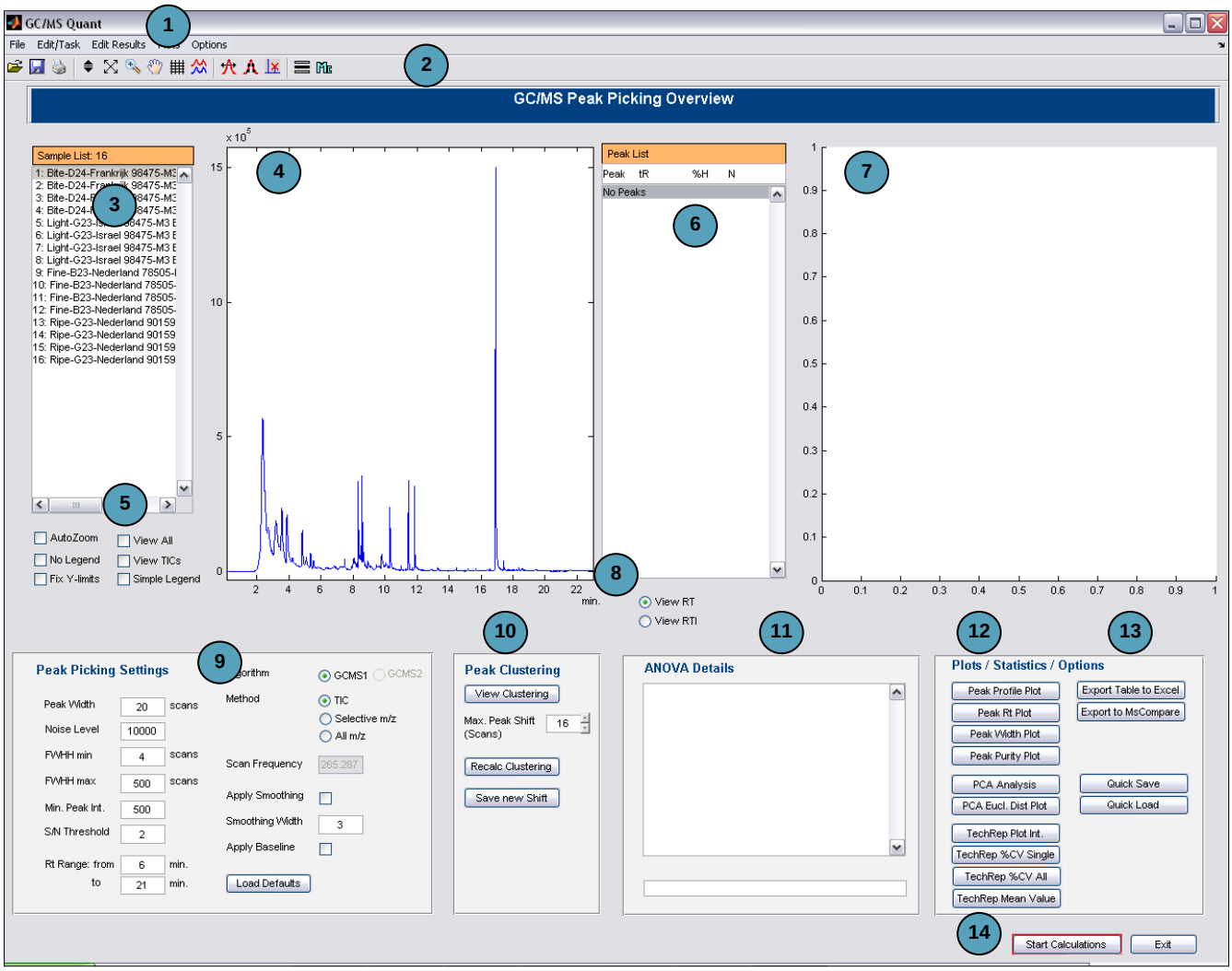


Figure 15: GCMS Quant opening screen

Layout and Buttons:

At the top of the window, the menu items are found **(1)**. Below the menu items, a series of toolbar buttons are displayed **(2)**. These buttons perform short cut actions.

(3) - Sample List Box: contains all the loaded data (full filenames).

(4) - Plot area 1: single TIC traces or multiple TIC traces can be viewed here. If peak picking has been completed, peak positions will be made visible.

(5) - Scaling options for TIC traces: Auto Zooming around detected peak, View all Traces, use simple or no legends.

(6) - Peak List box: contains the list of all detected peaks. The list contains: peak number, retention time, % peak height and the number of samples in which this peak has been detected.

(7) - Peak Profile/Statistics Plot area: this region is used to plot the peak heights for all samples or any of the statistics Plots.

(8) - View RT or RTI: view either retention times or retention index values. Calibration from (GCMS QC) must have been performed and excepted. Excel result table will be exported with RT and RTI values.

(9) - Peak Picking Settings: all peak picking parameters for individual files are presented here.

(10) - Peak Clustering Settings: after peak picking, peaks are clustered between samples. The Maximum allowed Shift between peaks can be specified here. If you need to use other values for peak clustering, you don't need to run a new peak picking, only change the peak clustering value. You can save the new peak clustering parameter after changing it. The **View Clustering Button** is not active yet.

(11) - ANOVA Info Section: selected samples from the sample list box or from the statistics plots will display the

ANOVA information in this area for all selected samples. You can control which ANOVA fields are displayed from the Options Menu (Set Fields for ANOVA details).

(12) - Plots and Statistics area: select from different plots to view peak detection results.

(13) - Export results to Excel or to MsCompare.
Quick Load and Save; using these features you can quickly load results from the last (quick) saved analysis.

(14) - Start the Calculations using the current settings or exit the program.

Menu Items:

The items found in the Menu bar are explained below:

File:

1. Load Results; loads previously saved results (*.gcq files) from disk. Both samples and peak picking results will be loaded.
2. Save Results as; saves the current sample list and peak picking results to a new file name.

Edit/Task:

1. Delete selected Samples: deletes the selected samples from the sample list. This feature can only be used after peak picking has been performed. After deleting the samples you need to re-run the peak clustering only. It is not necessary to re-run peak picking.
2. Keep only Calibration Components: if only QCS samples are loaded you can select to remove all peaks from the results list except the peaks matching the calibration components (as defined in the GCMS QC module).
3. Convert Table to Retention Time Index: the calibration results in GCMS QC should have been accepted. After that you may switch the table format between Retention Times and Retention Time Indices.

Edit/Results new!!**:**

1. Recalculate Zero Peak Values: Due to thresholding, Peak Picking results for very small peaks might not be present for all samples. This option will re-calculate the Peak Heights for all peaks and samples that have zero/empty entries. After this operation, the table will be more complete for analysis.
2. Keep Consistent Peaks Only (Peaks must have values for all Samples). This option will remove all peaks for which some of the values are zero or empty.
3. Keep Consistent Peaks (X % of peaks should have values). This option will delete (smaller) peaks found in less than X % of the samples, e.g. specify that at least 80 % of the samples should have a calculated value.
4. Keep Consistent Peaks (N values non-zero). This option will delete (smaller) peaks in which the user can define N. E.g. N=10 means that a least 10 samples should have a calculated value. If not, the peak will be deleted from the table.
5. Keep Consistent Peaks (X % of peaks should have values). This option will delete (smaller) peaks found in less than X % of the samples, e.g. specify that at least 80 % of the samples should have a calculated value.

6. Delete Peaks with High Relative Standard Deviation in Technical Replicates Group. Specify which group to use and specify the Coefficient of Variation (%). Peaks from replicate samples that have high CV can be deleted. Peaks can be deleted for each Tech. Rep Group separately.
7. Delete Peaks with High Relative Standard Deviation in **all** Technical Replicates Group. Peaks from replicate samples that have high CV can be deleted automatically in all Technical Replicates Groups.

8. Delete Low Intensity Peaks (Relative). Specify the Relative Peakheight as a percentage. Peaks that have a smaller average PeakHeight will be deleted from the table.
9. Delete Low Intensity Peaks (Absolute). Specify the Absolute Peakheight. Peaks that have a smaller average PeakHeight will be deleted from the table.

10. Reload Original Table: not implemented, please use the Quick Save and Quick Load Buttons.

Plots:

1. Change the Peak Marker in TIC trace plots. You can plot a rectangular box or a single vertical line plotted at the average retention time of the detected peak(s)

2. Smooth Chrom. LineStyles; option to plot smooth lines for traces and statistics plots.
3. Pop-Out View (Printing). Use this option to view and print the selected graphical window.

Options:

1. Set Fields for ANOVA details; opens a window in which you can select which fields to use for displaying in the ANOVA Details area for each sample. The selected fields will be saved to disk.
2. Set Fields for PCA legends.
3. Set Technical/Biological Replicates based on Field Number **new!!**. You can specify which field to use that specifies the Technical Replicates. E.g. select from Sample ID or Variety. It is assumed that samples having the same Sample ID or the same Variety are all replicates. Special plots can be made to view results from replicate analysis.
4. Use small GUI fonts; uses small fonts sizes on some monitors.

Toolbar Icon Buttons:

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

The icons from the Icon Bar perform the following actions:

1. Loads peak picking results from disk
2. Saves peak picking results under new name
3. Printing, not active yet.
4. Scale TIC in Y-direction; not active yet.
5. Not active yet.
6. Not active yet.
7. Reset to full scale plotting
8. Activate Zooming
9. Activate Panning
10. View grid style plot
11. Stacked trace view; not implemented yet
12. Export all traces and results to MsCompare
13. Export single selected sample to the Browwer
14. Peak Width measurement tool
15. Plotted dotted line style for TIC traces to see scans.
16. Remove selected peak(s) from result table; not implemented yet.

Running Peak Picking and Clustering:

The Peak Picking Parameters are explained below: Please experiment with the peak picking settings in the GCMS QC module. Select Calibration Settings and then select one of the project samples to optimize peak picking settings.

Peak Width: the width in scans for separating closely eluting peaks. If you use a very narrow width, noise on broader peaks could be mistakenly seen as real peaks. You can select a few TICs for plotting and use the Peak Width Measurement tool from the Icon Toolbar to measure the peak widths in scans. Click on the left and right side of the peak. The number of scans and time-range will be displayed. Use this tool to check how far apart closely eluting peaks are.

Noise Level: this is one of the most important peak picking parameters. A peak should have an intensity compared to its baseline larger than the noise level.

Min. Peak Int.: specify the absolute intensity threshold. Peaks below this level will be deleted.

S/N Threshold: you can specify a Signal to Noise ratio cutoff. Peaks smaller than this value will be deleted.

Rt Range: specify the retention time range if you want to run peak picking in a selected time window only

Apply Smoothing: optionally you can apply smoothing to the TIC traces to better estimate noisy peaks. Be careful not to include noise.

Smoothing Width: The Smoothing Width (scans) determines the amount of smoothing applied.

Apply Baseline: optionally runs a full baseline correction (not in this version)

Load Defaults: load defaults Peak Picking parameters from disk.

Max Peak Shift: select the maximum allowed shift between similar peaks from different samples. Careful; a small value will cause many peaks to be present in single samples.
You can experiment with the Max. Peak Shift parameter without the need to re-run peak pickings. This makes it very fast to experiment with the clustering parameter. Select a new shift value and press **recalc clustering.** If the peak clustering value gives good results, you can save it to disk.

After running Peak picking, the settings will be automatically saved to disk and the last used settings will be reloaded when you start GCMS Quant.

New: The user can run 3 different types of Quantitation/Peak Picking procedures. Peak Picking can be based on Total Ion Currents only; a list of selected m/z values (as defined by the Method) or perform Peak Picking on all m/z traces in a GCMS data File.

Quant Example:

This example includes 16 project samples (4 samples of similar "ras" type, each having 2 technical and 2 biological replicates). To start peak picking using the current settings press: **Start Calculations**. You will get a warning on a missing layout file. Please ignore, this is for a future releases. After a few seconds the results will be displayed in the PeakList box. The title of the box will display the total number of detected peaks, in this case 199. Peaks are sorted on retention time. Select a few samples and activate the Auto Zoom option. You will be asked for a time window for zooming in around the position of the selected peak. If TIC traces are very close together you can identify the traces by just clicking on it. The sample will be highlighted in the Sample List Box.
The peak position and width will be displayed in the TIC plotting area. Click on one of the detected peaks in the Peak List Box; the Peak Profiles for all samples are plotted and the selected samples are marked (x) in the peak profile plot. The last value in the peak table relates to the number of samples in which this peak was found. Single sample peaks, or peaks found in only a few samples are suspicious. You should check the Max. Peak shift parameter and also perform some basic diagnostic analysis. Outlying samples can be the cause for the detection of many peaks that only appear in a single sample.

You can select multiple peaks for profile plotting. These will all be marked in the TIC plotting area.

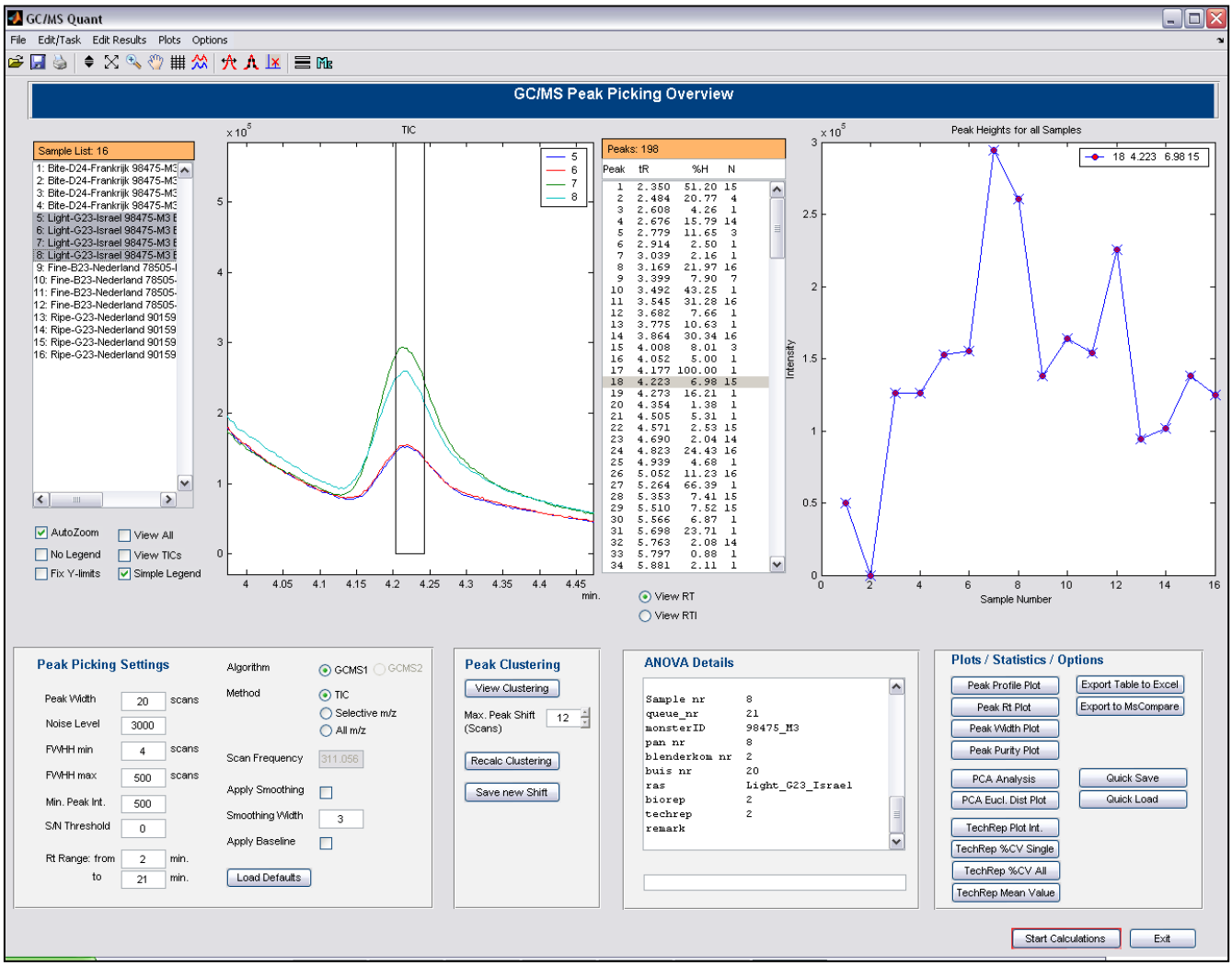


Figure 16: GCMS Quant Overview after Peak Picking has been completed

Figure 16 gives the overview of the GCMS Quant screen after peak picking (TIC) has been performed. Four samples have been selected (two technical and two bio replicates) and a peak was selected from the PeakList box. The detected peak positions are updated in the TIC plotting area. The peak is present is almost all samples except in the second sample. The ANOVA area displays the sample information for the selected samples. A selection was made to view only the "queue number" and "ras". From the results it can be seen that many peaks belong to sample 2 only. When plotting the group of samples to which the second sample belongs, it appears that the second sample is a strong outlier in this group. This very much influences the peak clustering.

GCMS Quant Basic Plots:

After peak picking 4 basic plot types become available, see figure17 a-d:

A: Peak Profile Plot: select a peak entry from the Peak List, the peak height values will be plotted for all samples. You can select more peaks to view multiple profiles. Selected samples from the Sample List box are marked in the Profile plot with the marker X. Clicking on a point in the profile plot will highlight the sample in the Sample List Box. See Figure 17 upper left.

B: Actual Retention Time plot: shows the actual retention times for the selected peak for all samples. A missing peak has a value of zero.

C: Actual Width plot: shows the actual widths for the selected peak for all samples. Missing peaks get the value zero.

D: Peak Purity plot: displays the peak purity values for the selected peak. Samples having missing peaks do not have a purity value. The displayed purity level is set to a value of 0.8. You can see that the selected samples from figure 16 all have high purity values, but other samples are definitely not pure. The purity is calculated by comparison of the MS spectrum in the top of the peak with the spectra at either side of the peak (FHWW). Using a peak picking procedure for TICs only will probably generate many impure peaks when the sample contains many peaks.

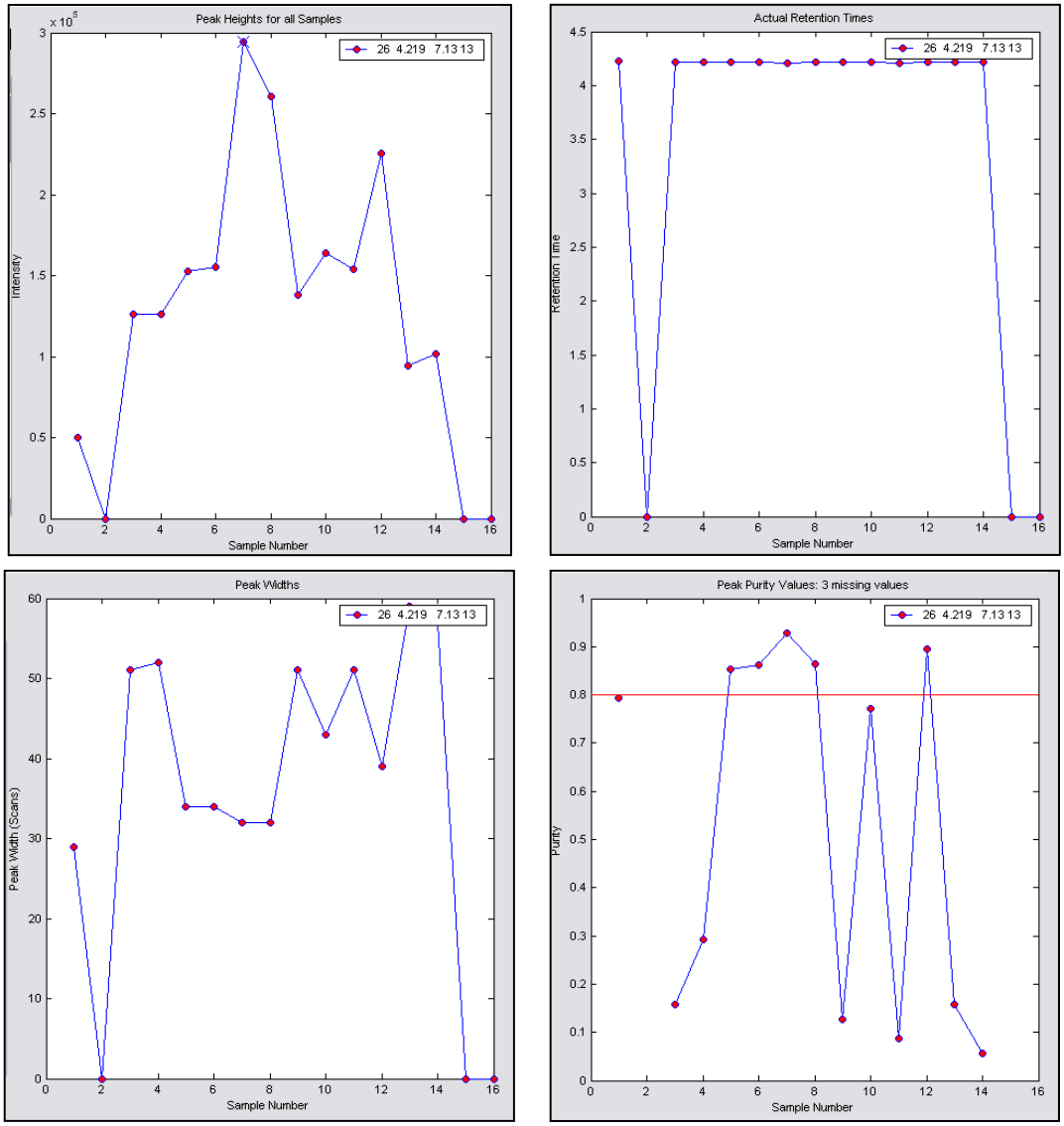


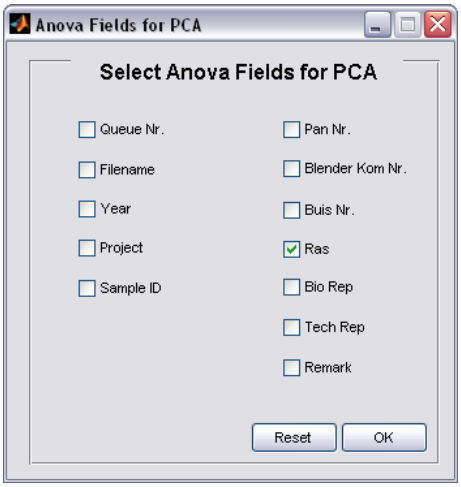
Figure 17 A-D: Basic Quant Plot types: upper left - Peak Profile plot; upper right - actual retention time plot; lower left, actual width plot and lower right displays the peak purity plot for all samples.

GCMS Quant Special Plots:

A number of specialized plot are available for diagnostic purposes:

A: PCA Score Plot: the PCA scores plot gives a good overview of your data results in two dimensions. Outliers and strange behavior can be observed quickly. PCA is performed on the full mean centered result table. Push the PCA button to get the score plot; Figure 18. The samples are marked with their row number from the Sample List Box.

The user may choose from different ANOVA legends for grouping the samples in the score plot. To select the grouping choose: Menu > Options > Set Field for PCA legends. You will then get the following window on the left from which the grouping can be selected.



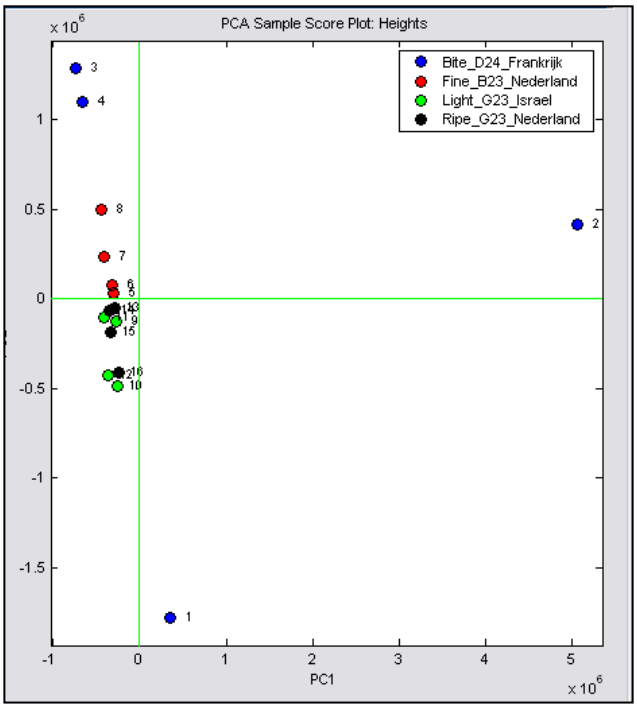


Figure 18, PCA score plot. Grouping is based on "Ras"

The "Ras" field was selected for grouping samples, but you can make any selection. After changing the legend, the plot will be updated directly. From Figure 18 it can be concluded that we have 2 strong outliers, sample 1 and 2 both belonging to the Bite_D24_Frankrijk "Ras". Especially sample 2 is a very strange sample on principal component number 1. As samples from the same group are either technical or biological replicates, we expect these samples to be very close together in the PCA score plot as is the case for the other three groups.

The PCA score plot is interactive. Clicking on a point will highlight the selected sample in the Sample List Box and the ANOVA Details box will show the selected info for this sample. You can drag a rectangular area in the PCA score plot to select more samples. The Sample List Box and ANOVA box will be updated. In this case it can be seen directly that both samples are from the same biological replicate.

B: PCA Euclidian Distance Plot: based on a selected grouping, the Euclidian distance plot calculates the average distance for a group from the group mean. Also the standard deviation is calculated. Figure 19 show the Euclidian Distance plot for grouping "Ras". On the x-axis the average distance from the group mean is plotted, on the y-axis the distance standard deviation for each group is plotted. It can be seen directly that the scores (see also figure 18) for group Bit_D240_Frankrijk are quite scattered compared to the other groups.

For this group, both the average distance and standard deviation are large.

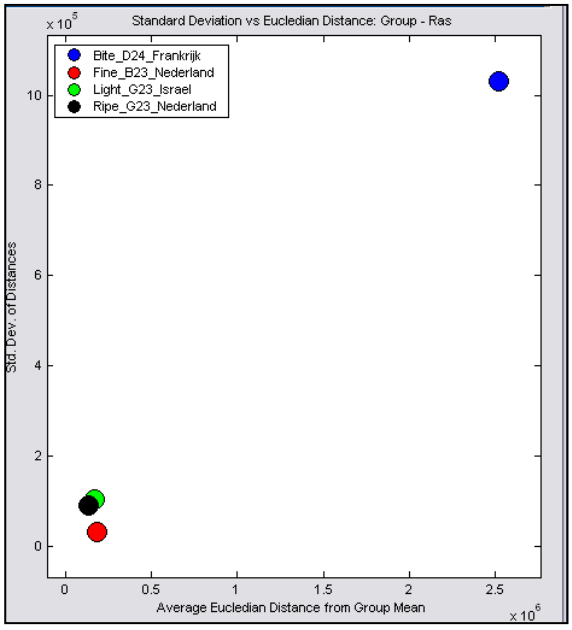


Figure 19: PCA Euclidian Distance Plot

You can base the grouping on a number of ANOVA Fields: Sample ID, Ras, Blender, Pan, Tech Rep and Bio Rep. Before creating the plot you will be asked which ANOVA field to use. The title of the plot will display the selected grouping for the plot.
Clicking on a point in the Euclidian Distance Plot will highlight all samples belonging to that group in the Sample List Box. Also the ANOVA Box will now contain the samples from the selected group only. When selecting e.g. "Blender" as the grouping field, it can be easily seen that samples from the same "Ras" always have the same blender number.

C: Technical Replicate Plot Intensities: in the Metabonomics setup each "Variety" sample has two technical and two biological replicates. The Technical Replicate plot will display the four replicates values for the selected "Variety" in one plot as a function of the average of these four values. At the start the plot will be created for all detected peaks, but the user has the possibility to do sub-panel plotting, in which each time a group of 50 peaks will be displayed to get a more detailed overview. Using the left and right arrow key you can move forward and backward between sub-panel plots.

After the plot has been created, the selected "Variety" samples will be marked in the Sample List Box and in the ANOVA Box. Clicking on one of the points in the plot will highlight the selected peak in the Peak List box and an overlay plot of TICS of the selected "Variety" samples will be displayed. The ANOVA Box will display which sample was selected. Please activate AutoZooming to see a more detailed overview of the peaks. Figure 20 shows the GCMS Quant full screen with all peaks present in the TechRep Plot. One point has been selected, and this sample is displayed in the ANOVA Box. The TIC plot displays the four samples for the selected "Vareity" zoomed around the selected peak, which has been highlighted in the Peak List box. Because it can be very crowded in the plot, the selected peak will be marked with an arrow.

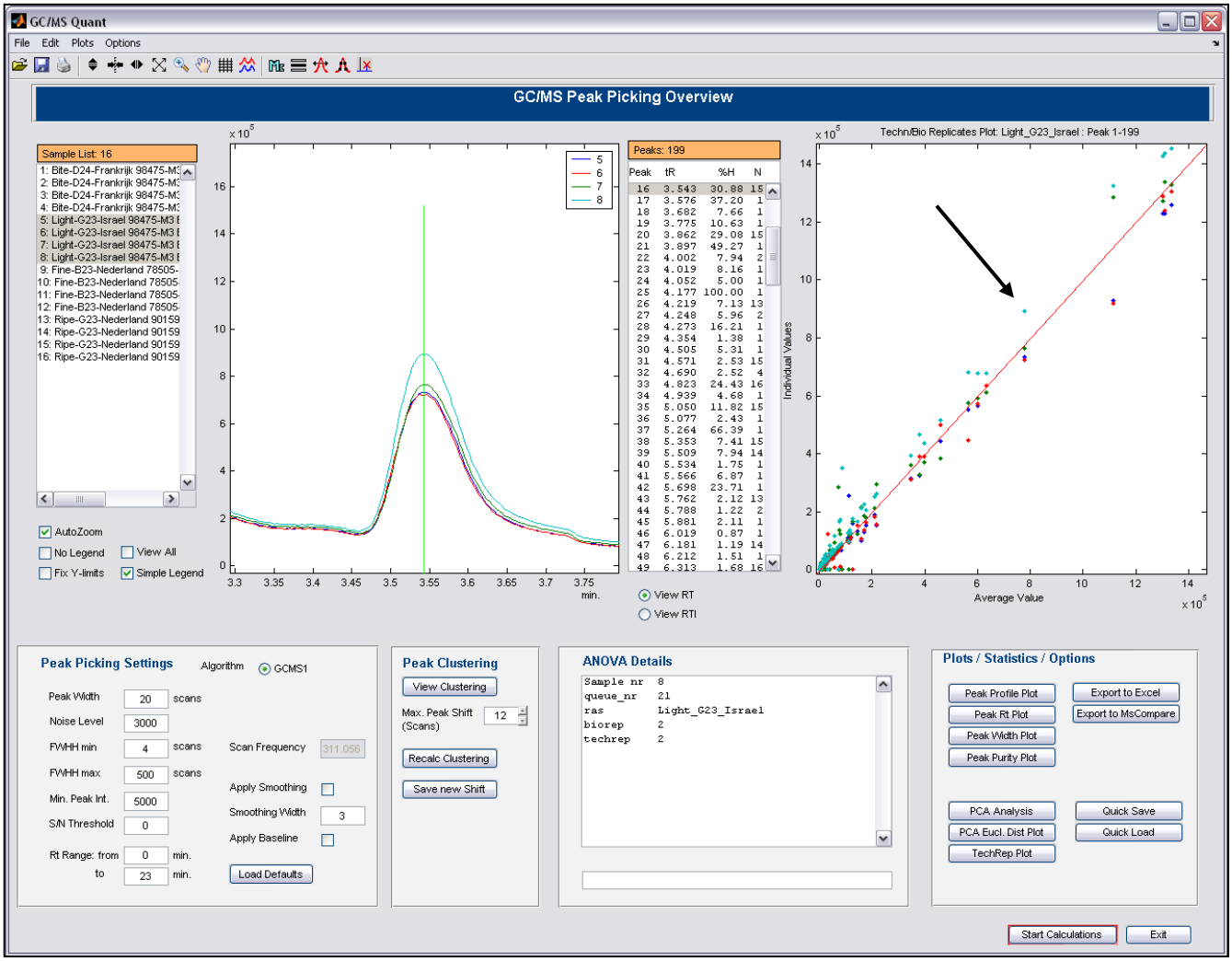


Figure 20: Technical Replicate Plot for group Light_G23_Israel. All peaks are displayed for the 4 replicates (no subpanels). TICs for the group samples are plotted and the ANOVA box shows the selected sample while the Peak List box highlights the selected peak number.

D: Technical Rep. % CV Single (Group) new!! **:** Using this Plot, the % CV will be plotted for all peaks from a Selected Group. You will have to select the Technical/Biological Replicate group first. The x-as will now display the Coefficient of Variation for all replicates. You can optionally remove peaks having high CV from the Edit/Results Menu.

E: Tech Rep. All (Groups) new!!**:** Select one or more peaks from the Table and press this plot. You will get a bar graph of %CV for this peak for all the Groups separately. Optionally select more than one peak.

F: Tech Rep. Mean Value new!!**:** Select this plot to present an overview of the Mean values in each of the Technical Replicate groups. The mean and standard deviation are presented. Optionally select multiple peaks.

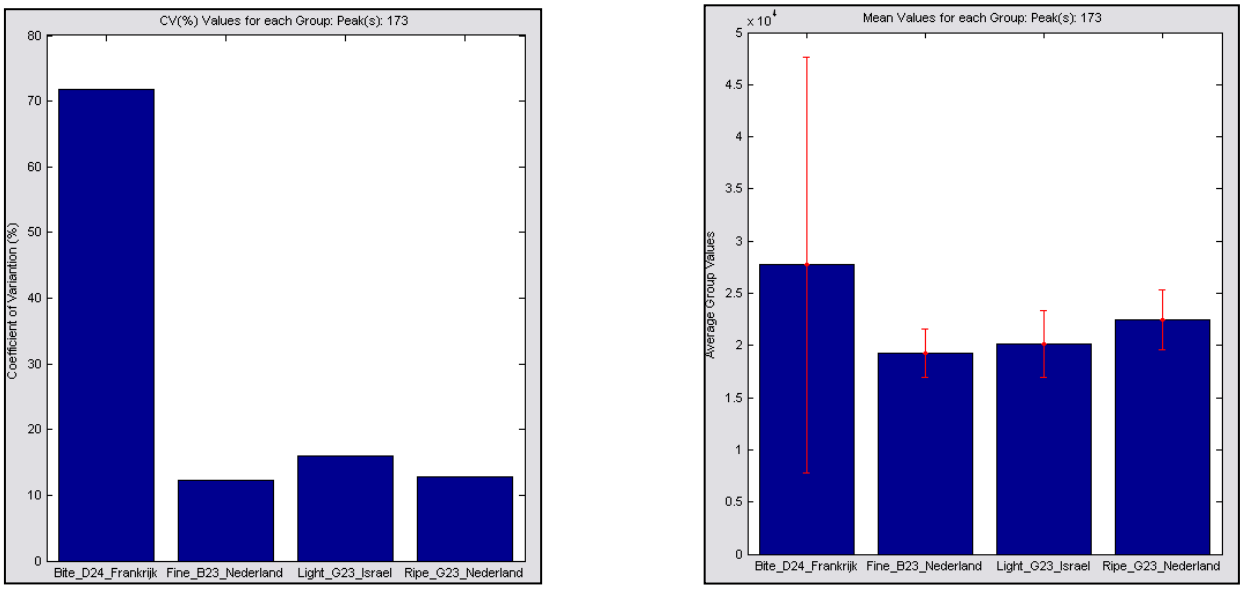


Figure 20A: Technical Replicate Plots: Left; % Coefficient of Variation for a selected Peaks in the 4 Technical Replicate Groups. Right; Mean values + standard deviation for the selected Peak.

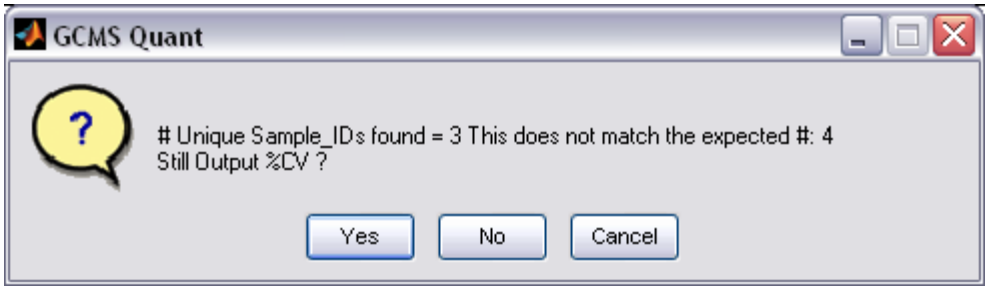
Converting Retention Time to Retention Time Indices:

Below the Peak List Table, two radio buttons can be found to view the table in normal retention time units or in retention time index units. If you select **View RTI,** you will probably get a warning telling you that **NO RTI DATA HAVE BEEN CALCULATED, SELECT FROM MENU.** From the Edit Menu select: **Convert Table to Retention Time Index**. If you get a warning message, you should first accept the calibration results in the **GCMS QC module**. Go to GCMS QC and select from the **menu: Ret. Time Indexing > Accept Calibration Results for Rt Indexing.**
The Quant table will be converted to RTI indices. You can now easily switch between both units. Also exporting to Excel will now export peak results in the original retention times and retention indices.

Exporting Results to Excel:

Select Export to Excel to save 6 different result tables in Excel Format. After calculation of all purity values you will be prompted for a filename and location. Be patient, it can take some time to write the tables to disk. When finished, the Excel file will be opened.

The % Coefficient of Variation calculates the spread for peaks having the same Sample ID. The program will check the number of Sample ID groups. If e.g. 16 samples are present, the program expects 16/4 unique Sample ID's. If the calculated value does not equal the value found, you will get the following question. Select Yes to Output % CV, otherwise % CV output will be skipped.



Each Excel file will have 3 empty sheets, followed by the following GCMS Quant output sheets:

Raw: Starts with a single header line. The header line contains the date, the name of the table and the number of samples and peaks. The next line contains the info headers: Peak Number, Retention Time, Retention Index and m/z value (to be used in future releases). After that the results for individual samples can be found. The sample header contains the full filenames. The results start at the next line.

Norm: Basically the same type of table as Raw, however the individual values are normalized by the total TIC signal for each sample. Furthermore, this table contains all the ANOVA values and also the Total TIC Signals for each sample.

%CV: This sheet contains the % Coefficient of Variation results for all peaks. These values are calculated from samples having the same Sample ID. The Table Header will show each Sample ID name.

N: belongs to % CV and displays the number of samples present in each Sample ID group.

Parameters: Contains all filenames and the parameters used for peak picking.

Purity: Contains all the peak purity values for all samples and all peaks.

Export Data and Results to MsCompare for further Statistical/Multivariate Analysis:

MsCompare is MsXelerator's specialized module for Metabonomics data analysis based on LC/MS or GC/MS data. It contains many univariate and multivariate features for analyzing and studying multiple samples and groupings. See the MsCompare manual and tutorials to learn more about this Module.

You can export the results from GCMS Quant to MsCompare by clicking on the **Export to MsCompare button**. MsCompare can use Classes or Groups based on any of the ANOVA Fields you specify. Click on the MsCompare Icon button in GCMS Quant. You will be asked to define the ANOVA Class Field that will be used to set colors and classes in MsCompare.

MsCompare will open and show all samples. Press the Update Table Icon button on the toolbar to update the exported table. The result is shown in Figure 21. The exported table can be used for PCA/PLS, create sample groupings, sorting, viewing of TICs, EICs and MS spectra or calculate all kind of Group Statistics to find discriminating peaks . If you want to explore results in MsCompare you should save the table, as MsCompare uses its own format.

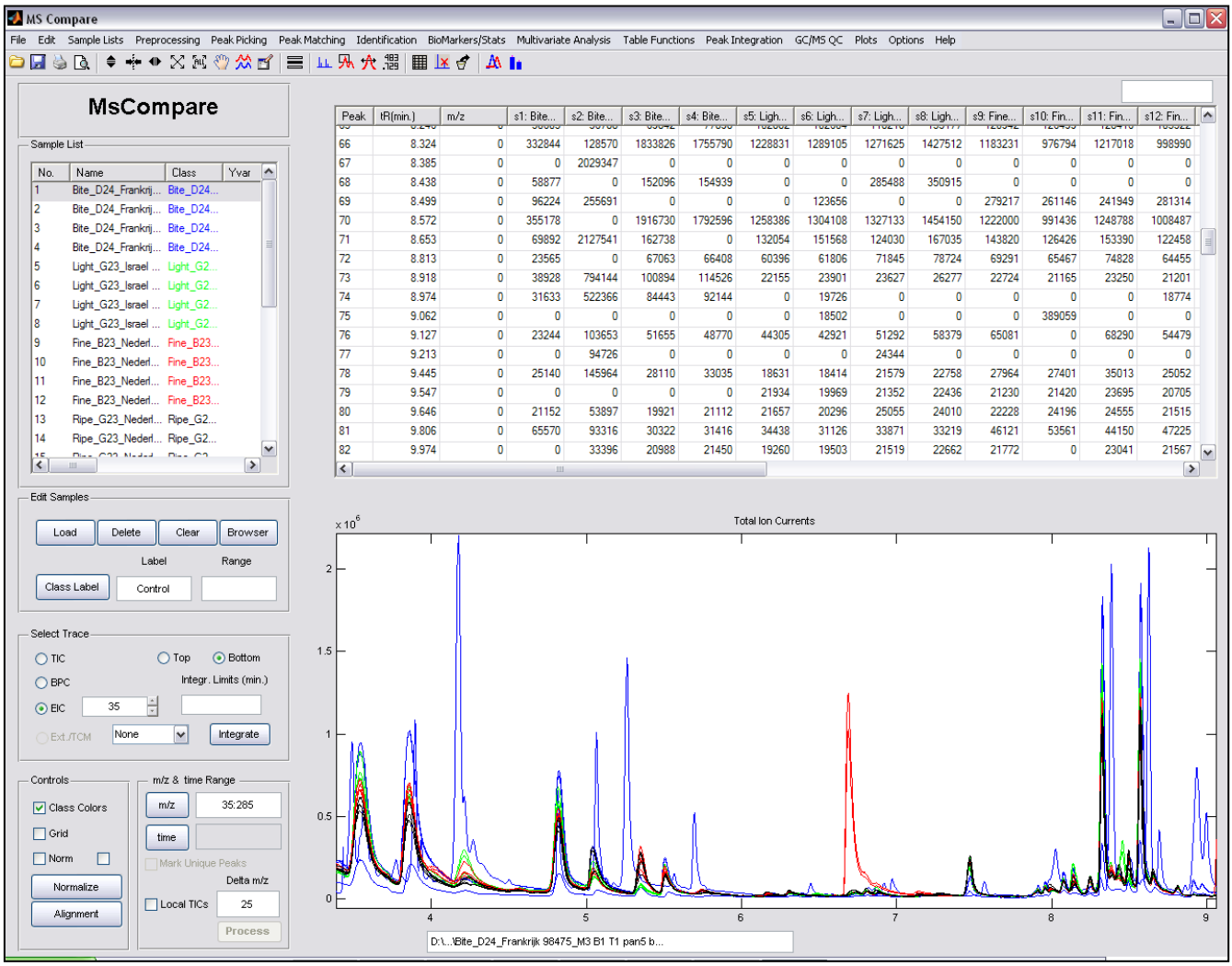


Figure 21: MsCompare Screen showing exported Samples and results from GCMS Quant. Classes were taken from the ANOVA Field: Variety.

Quick Save/ Quick Load:

In GCMS Quant you can use the Quick Save / Quick Load buttons to save the result table and the samples to a temporarily file on disk without specifying a filename for saving. The next time you restart GCMS Quant use Quick Load to load the last results that were saved. The program will tell you how many samples and peaks were stored in the temporarily file. Each Quick Save will overwrite the temporarily file on disk.

Zooming Scaling:

Below the Sample List Box some scaling options can be found:

AutoZoom: activate Autozoom to automatically zoom in on the peak of interest. You can select the window for zooming in.

View All: select View All to plot all TIC traces in overlay. Clicking in the Sample Box will reset the View all option. When all TICs are plotted the plot may get crowded. You can always identify the sample by clicking close to a TIC trace in the plot. This will highlight the sample in the Sample List Box. If **View All i**s active and you click at a sample position in the Profile plot, the TIC trace will be plotted in thick line style, so it can be easily distinguished.

Simple legend: you can use simple legends (numbers) instead of the full filenames.

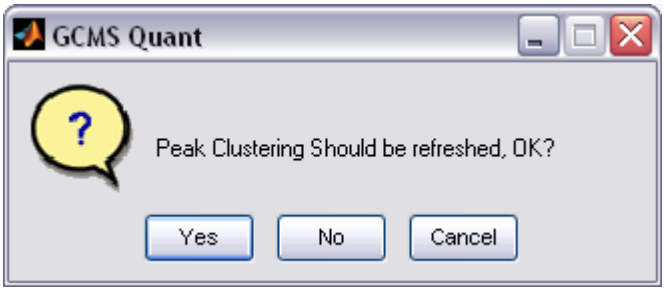
Fix Y-limits: select this button to set the current y-level for further plotting. Only works when AutoZoom is active. This option can be handy to keep zoom in on small peaks.

No Legend: no legend will be added to the TIC trace plot when clicked in the Peak List Box.

Running Diagnostics, Removing Samples and Re-Building a Result Table by updating Clustering:

From the PCA score plot in Figure 18 it is clear that two outliers are present; sample no. 1 and 2. This was also clear from a visual inspection of the TIC traces and from some of the other diagnostics plots. Especially sample 2 is responsible for many unique peaks in the peak picking table. Of course, this can be a true result, but from the replicate analysis we can conclude that these are outlying samples, especially sample 2 in which many peaks seem to be shifted. Sample 1 has a low intensity compared to sample 3 and 4 of the same group (see the Total Signal TIC plot in GCMS QC).

Removing Outliers can be done by selecting the samples to be removed. From the Menu select: Edit > Delete Selected Samples. The Sample List will be updated and you will get the following question.

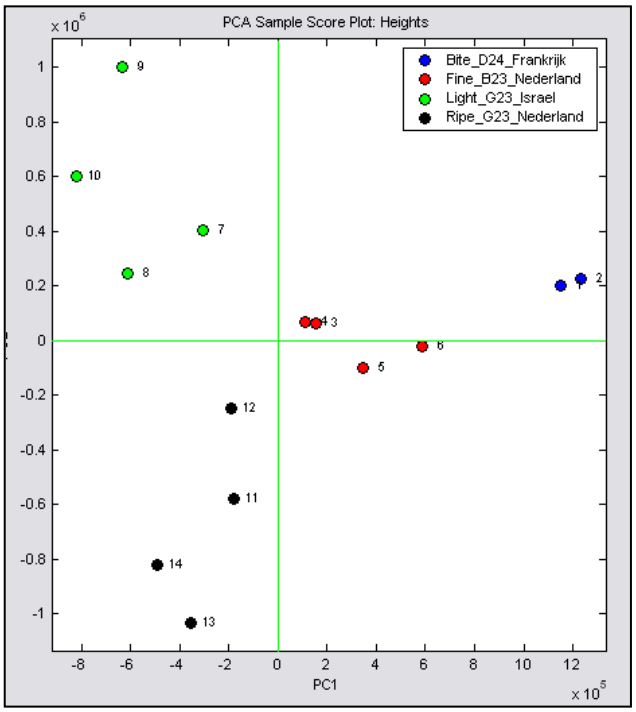


Press “Yes”, there is no need to perform a new peak picking as the peak picking result tables for individual files have not change, only the clustering should be refreshed. The peak picking list will now only have 125 peaks, many single sample peaks were caused by the outliers.

There will still be some single sample peaks in the result table, but these are often low intensity peaks at the limit of peak detection. As a result of the low intensities they are only present in a few peaks. Of course they might also be really unique for only a few samples. To explore these peaks, activate View All. Plot the peak profile and click on the samples in which this peak is present. The sample TIC trace will be plotted having a large line width so that it can be discriminated from the rest easily. Click the sample again to return to normal TIC trace plotting.

The PCA score plot on the smaller sample list is shown in Figure 22a and the Euclidian Distance plot is shown in Figure 22b. The PCA groupings are now quite clear. The Distance Groups in the distance plots are now more in line what can be expected from replicates. The blue group has the lowest distances, but this is now the group which only has two samples.

The TechRep plot is now also easier for interpretation, as many peaks have been removed which previously all had zero values from some of the groups. You could export the result to MsCompare for more detailed analysis tools and graphics.



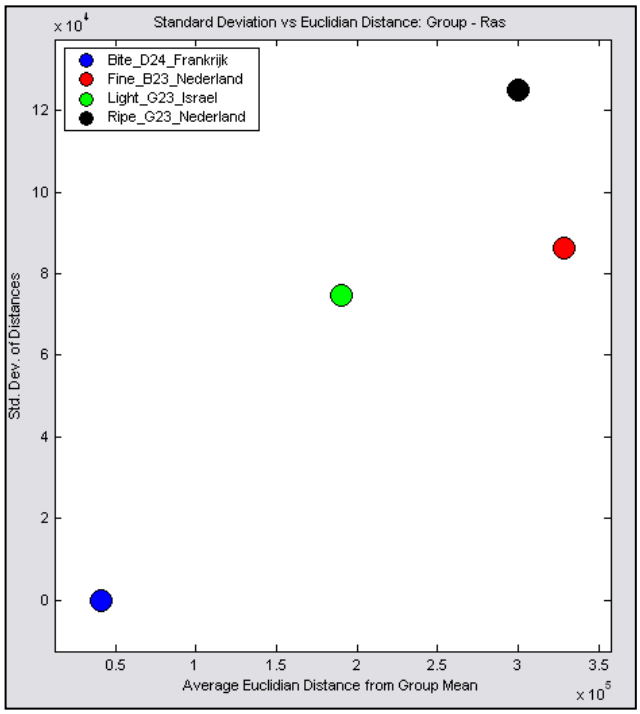


Figure 22a, 22b: PCA Score Plot and Euclidian Distance plot of the reduced data set with no outliers.

View Calibration Peaks only:

From the menu you can select: View Calibration components only. The full table list will be reduced to calibration peaks (n-alkanes) only. Within the GCMS QC module you should have accepted the results for retention time indexing.

End of Manual