

# Peak Detection with Varian MS Workstation for Varian 220 and 240 GCMS

by:

**Randall Bramston-Cook**

Lotus Consulting

5781 Campo Walk

Long Beach, Ca 90803

310/569-0128

[ebramstoncook@msn.com](mailto:ebramstoncook@msn.com)

February 12, 2010

Copyright 2010 Lotus Flower, Inc.



A primary application of chromatography is measuring concentrations of analytes in unknown matrices. After compounds are separated on a column, they are sent to a detector to measure their responses. Concentrations are a function of the detector signal, usually by an established linear relationship.

A critical factor in maintaining good reproducibility and accuracy of measurements is proper assignment of peak responses above baselines. If peak start and end points are not set appropriately, peak response can be either too large or too small, compared with its true value, and can vary dramatically with replicate runs of the same sample.

Varian MS Workstation<sup>1</sup> for mass spectral data uses a powerful peak detection algorithm originally developed in 1968 with the Varian 200 Data System set up for data collection from multiple gas chromatographs, and has been continually enhanced through the years in Varian CDS 101/111, Varian Vista 401/402, Varian DS601/651/654 and Star Data Systems, and is now in MS Workstation<sup>2</sup>, with many new enhancements over the older programs.

Correct calculation of a peak response involves some judgments to distinguish a real peak from noise, rising backgrounds, change in peak shapes from temperature and flow programming, unexpected disruptions in detector response from instabilities, and flow disruptions (from valve switches, for example). The Workstation, with preset parameter settings, can usually find "peaks" and assign an area or height, but sometimes these initial values do not match criteria that an experienced operator might have given. This reported value may not properly represent the true concentration of an unknown.

Varian MS Workstation provides a number of user selectable parameters to tune peak detection to specific conditions in use. For example, when a high performance capillary column is employed that generates some very sharp peaks, the user can set up parameters readily to fully detect these sharp peaks and ignore baseline drift from temperature programming. At the other extreme, if peaks are broad from use of packed or megabore columns, settings can be employed to expect these broader peaks and ignore detector spikes. Fortunately, preset parameters in this Workstation are usually sufficient to pick out most peaks, albeit not always optimally, and provide a starting point for fine tuning settings to accurately work on subsequent runs of the same ilk without further adjustments.

Varian MS Workstation performs automatic full storage of every chromatogram, with all raw data points, calibration data, and method parameters. Some factors involved in initial data collection - such as data rate, ionization parameters, tune parameters and chromatographic conditions - cannot be changed on this stored data after the original run. Others settings can be altered and their effects on a chromatogram can be visually inspected. Then this modified method can be saved and made available for use with subsequent samples to achieve better results without further intervention.

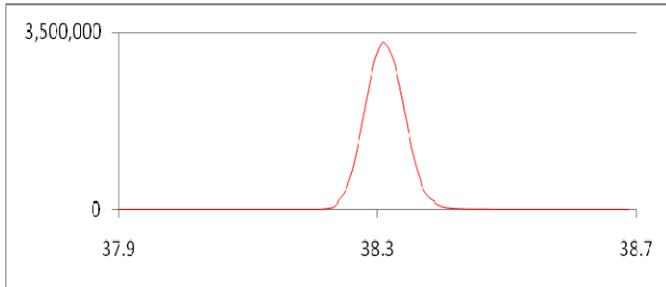
---

<sup>1</sup> Varian MS Workstation has been undergoing improvements over the years, and some significant changes related to peak detection have been implemented in later versions. All discussions here apply to Version 6.9.2.

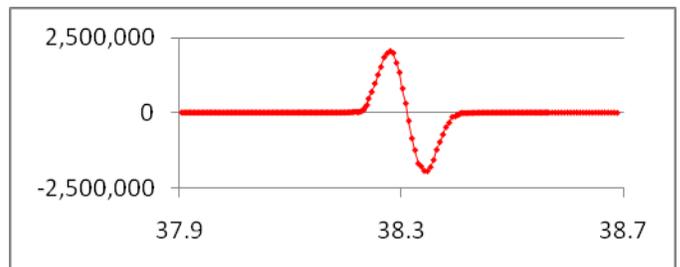
<sup>2</sup> This document presumes that the reader has some familiarity with various screens and operations within Varian MS Workstation. Hints are provided for locating specific operations within the Workstation for the following discussion points, but precise steps are not listed and are assumed to be obvious to the reader.

## Determination of Retention Time

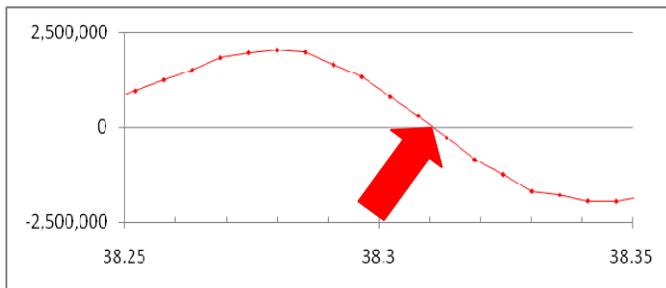
Varian MS Workstation establishes retention time of a peak by computing the zero crossover of the first derivative of the peak. If this value does not correspond to a collected data point, precise timing is determined by linear interpolation of points before and after the zero crossing.



**Retention time is determined by zero-crossing of the first derivative of a chromatographic peak.**



**The derivative is generated by an appropriate digital filter across the peak (see later discussions on Filtering).**

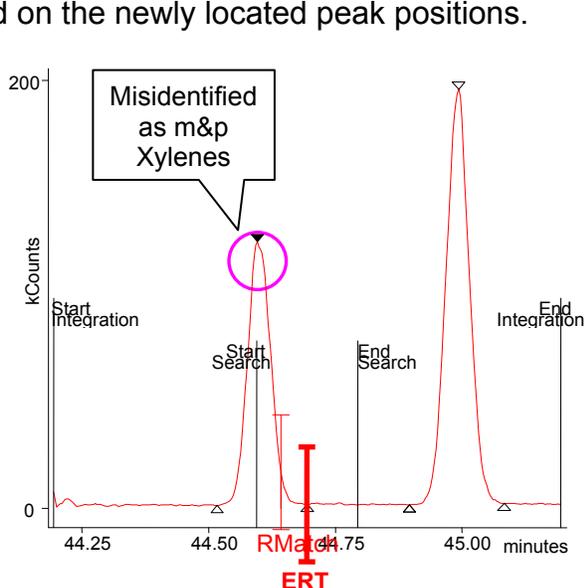


**The precise retention time is set by linear interpolation of points before and after the zero crossing.**

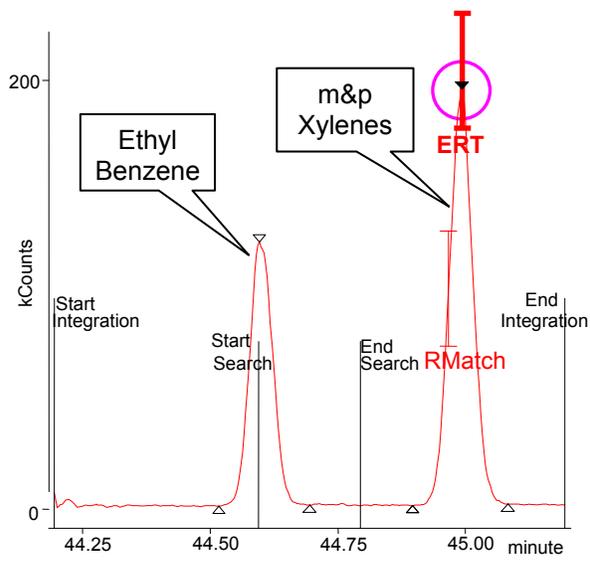
## Retention Time Adjustments with Identification Reference Peaks

Measurements in chromatography cannot be considered exacting, as random and systematic errors can introduce some variability in retention times (along with area measurements). To help correct for systematic deviations in retention times, certain compounds can be assigned as Identification Reference Peaks. Any movement in time of these peaks is detected and a correction based on this change is applied proportionally to all “non-reference” peaks, based on their relative retention to reference peaks. If multiple Identification Reference Peaks are assigned, then the new Expected Retention Times are proportioned based on their proximity to these reference peaks.

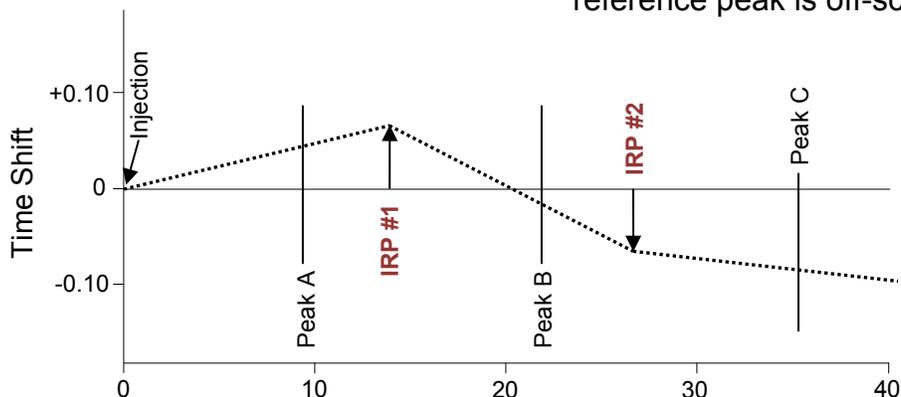
**Identification Reference Peaks** always must be present in **every** sample **and** standard in a series, and must be the **largest** peak within the Search Window; typically internal standards are perfect candidates. **Non-reference peaks** are picked out as the **closest** peak to the Expected Retention Time, after time adjustment from a Reference Peak correction. With a run type of "Analysis", the Compound Table in the Method is not altered as a result of these operations and Identification Reference Peaks are only used to pick out peak movement and properly identify analyte peaks. For a Calibration run type, retention times in the Compound Table are updated based on the newly located peak positions.



Normal variations in chromatographic conditions can shift retention times away from the expected locations. And adjacent peaks can be mislabeled and wrong results are reported.



Proper use of Identification Reference Peaks can sense a shift in retention times and allow corrections to move the "anticipated" Expected Retention Time over to the correct peak. The reference peak is off-scale to the left.

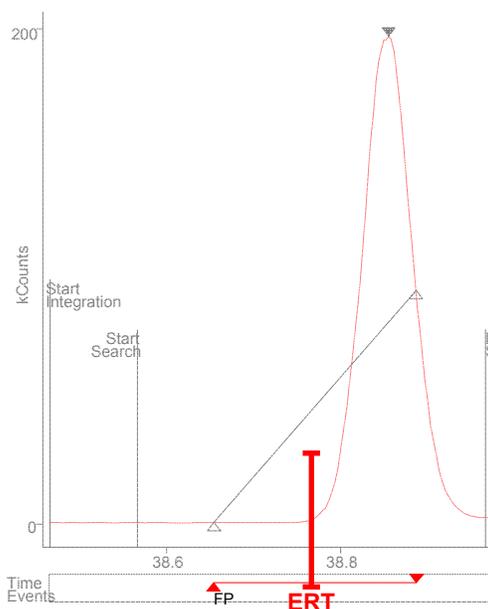


**Multiple Identification Reference Peaks** can provide time corrections over the whole chromatogram. New Expected Retention Times for unknown peaks are adjusted proportionally to their positions relative to adjacent reference peaks. This synthetic example illustrates the corrections involved for two reference peaks.

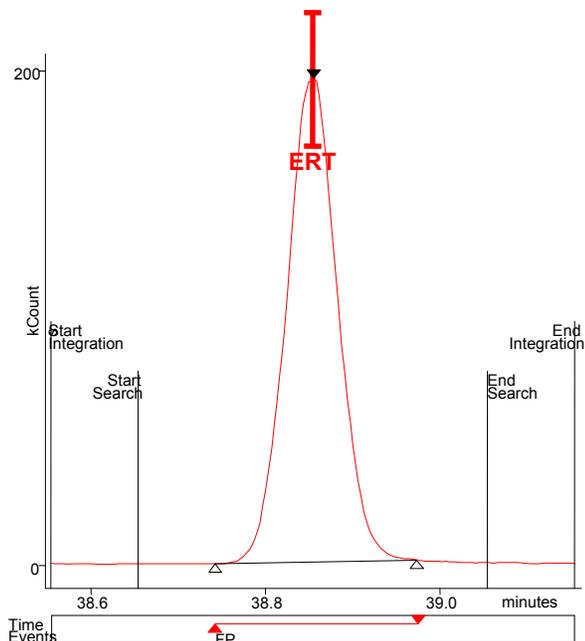
Compound	Expected Retention Time	Measured Retention Time	Measured Shift	Correction	New Expected Retention Time
Peak A	9.374	--		+0.045	9.419
<b>IRP #1</b>	<b>13.841</b>	<b>13.907</b>	<b>+0.066</b>	--	--
Peak B	21.844	--		-0.015	21.829
<b>IRP #2</b>	<b>26.752</b>	<b>26.816</b>	<b>-0.064</b>	--	--
Peak C	35.395	--		-0.085	34.545

## Automatic Retention Time Updates and Timed Event Adjustments

Retention times do not always hold their precise locations in chromatograms and can be expected to move a bit from run to run, especially after the chromatograph sits idle for an extended interval or a new column is installed. If they are not adjusted to the new times, peaks may be labeled with incorrect names, and serious errors in quantitation can occur. In a calibration series, a subtle change in retention time is adjusted automatically by locating the peak within the integration window that is **closest** to the expected retention time and then replacing the listed time with the newly established one in the method Compound Table.



**The Expected Retention Time (ERT) listed in the compound table was originally set to 38.766 minutes, and a Forced Peak (FP) timed event is misaligned with the new peak. Computed area is completely wrong with this invalid baseline.**



**After a Calibration run type is executed, the retention time in the compound table is updated with the true apex value (38.854 minutes), and all associated timed events are also adjusted with the time change. Integration and Search windows are altered as well, now centered on the peak acme.**

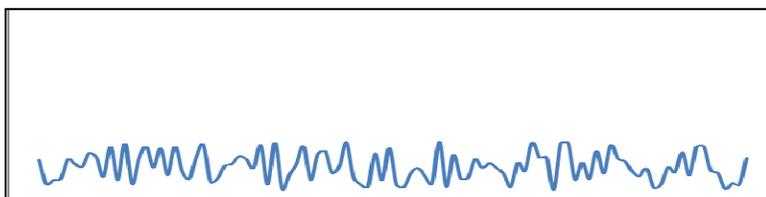
Updates to retention times are **NOT** performed with Analysis run types (unknown samples), as every peak may not always be present here.

When a multipoint calibration is performed, the last run in the series ought to be a midlevel standard so that the final retention time update is accomplished on sizeable peaks, and not on puny peaks where retention time may be skewed by noise, and not on the largest peak where retention times may be distorted from column or detector overload.

## Peak Simulations

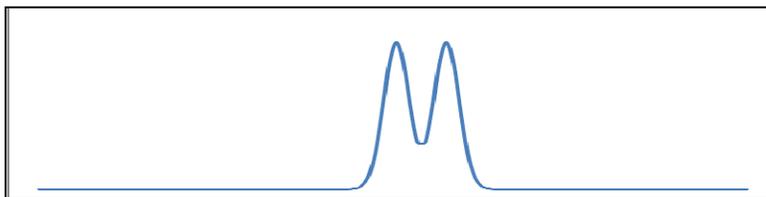
Shape of typical chromatographic peaks (red trace) can be approximated by a Gaussian distribution (shown in blue), as displayed to the right.<sup>3</sup> Using this mathematical model, various data treatments available in Varian MS Workstation, such as data collection rates and smoothing, can be simulated and displayed. In discussions following, 600 points on a Gaussian curve are generated in Microsoft Excel to study various effects on peak processing.

Noise is imitated by generating a series of random numbers within a specified range and then plotted.

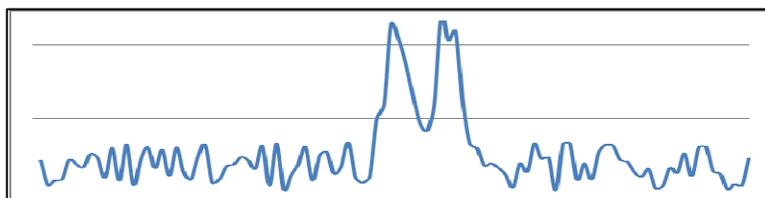


**Noise is simulated by plotting a group of random numbers.**

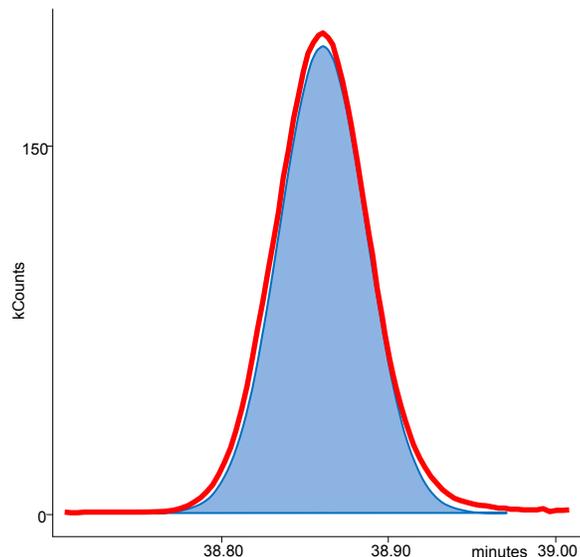
Artificial peaks are created by adjusting Gaussian points to relate to the noise numbers produced. Two closely spaced peaks are used to look at effects of various parameters on resolution of the two, especially when these peaks are combined with random noise numbers.



**Overlapping peaks are created from Gaussian distributions, with a 70% valley penetration.**



**The doublet peak structure is mathematically combined with the random noise to create a data set closely resembling real peaks near their detection limits.**



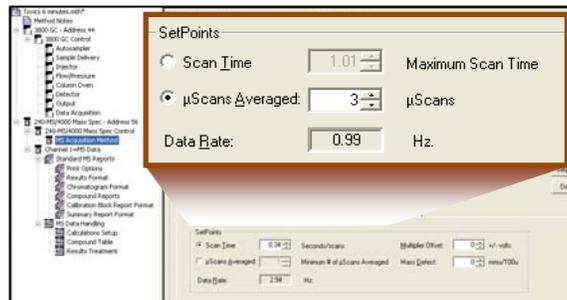
**Normal chromatographic peaks are nearly Gaussian in shape.**

<sup>3</sup> Both actual data and simulated numbers are presented in this monograph to illustrate discussions about peak processing. Synthetic numbers are typically presented in blue graphs and real data are displayed in red.

## Definition of Terms

**$\mu$ Scan Point** – This is a single spectrum scanned from low mass to high and becomes the raw data used for all further data processing.

**Scan Point** –  $\mu$ Scans are averaged together based on Scan Time or  $\mu$ Scans Averaged. If Scan Time is specified, the number of  $\mu$ Scans per Scan increases across the peak to keep the time interval constant. If Minimum # of  $\mu$ Scans is checked, the scan rate increases across the peak. The choice becomes a trade-off between spectral quality (for Scan Time) and chromatographic peak quality (for Minimum # of  $\mu$ Scans).



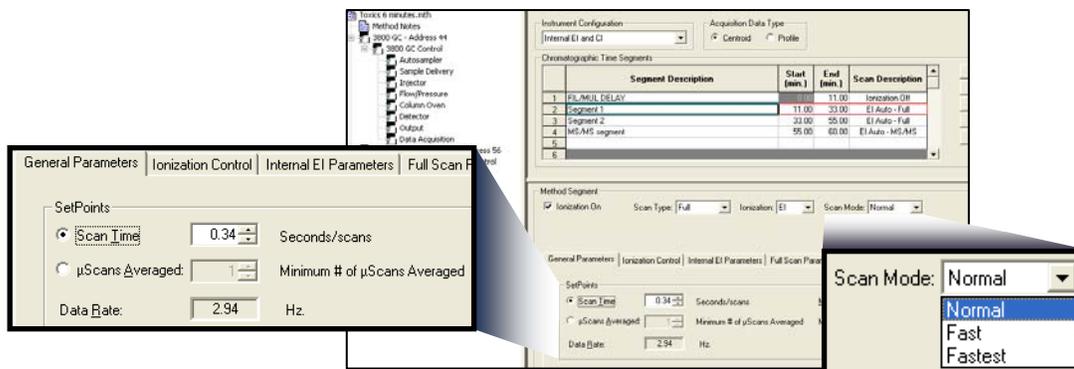
**Bunched Point** – Scans are averaged together to provide ideally 10 to 20 Bunched Points across the top of a peak, based on the Expected Peak Width. Bunched points are used only for peak detection and not for peak integration. They are not displayed.

**Interpolated Point** – Computer generated points for peak detection that are midway between adjacent Bunched Points, used to increase the number of points to find the correct start and end points of a peak when insufficient Bunched Points are collected. These points are only used for peak detection and baseline assignment and not for peak integration. They are not displayed.

## Data Collection Rate

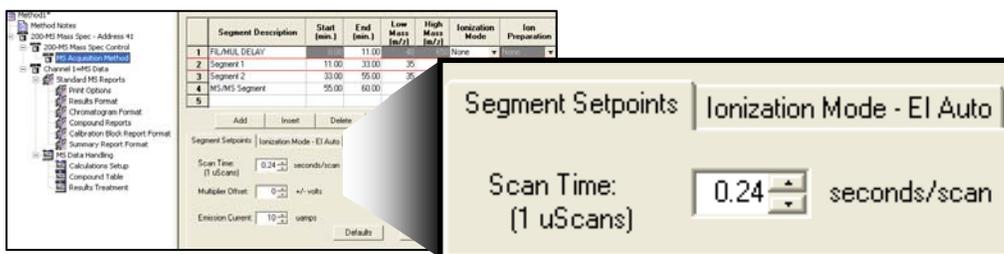
Varian ion traps operate by storing all ions eluting from a chromatographic column within a specific mass range. Then these ions are ejected out of the trap to the detector with a ramping RF field, at a rate 5,000 to 10,000 u/sec over the scan range, to yield mass spectral raw points, labeled as  $\mu$ Scans. Since these  $\mu$ Scans often are generated faster than is required for most peaks, they can be averaged together to help reduce noise and to generate better quality spectra. These averaged data points are then stored in a .SMS or .XMS data file for post-run peak processing and are visible in chromatographic and spectral displays. This data bunching rate applies to the whole segment, and can be altered in other segments if needed.

The rate that these data points are collected is a function of mass scan range, number of scan segments, tune type (where spectra are adjusted to match specific criteria), scan type (Full, MS/MS, SIS and  $\mu$ SIS), and, for the Varian 240/4000, scan mode (Normal, Fast, or Fastest).



**Parameter entries for Varian 240/4000 to set Data Rate.**

With the Varian 220/2200, the entry of scan time sets the  $\mu$ Scans Averaged. Data rate is computed by taking the reciprocal of the Scan Time. For example, for a Scan Time of 0.24 sec/scan, the Data Rate is 4.17 Hz. And the Varian 240/4000 allows entry of **either** Scan Time **or**  $\mu$ Scans Averaged, with the other computed based on all of the criteria listed above and then shown grayed out. The Data Rate displayed is based on the resulting Scan Time.



### Parameter entry for Varian 220/2200 to set Data Rate.

Data rates for Varian Ion Traps are dependent on conditions used for the mass spectrometer. Typical conditions for measuring toxic compounds by EPA Methods 524.3, 624B, 8260 and Method TO-15 are employed for the data rate listings below:

<b>Mass Range</b>	35-300 m/z
<b>Steps in Tune Segments</b>	4

<b>Data Rates<sup>4</sup> for Varian 220/2200 GCMS</b>		
<b>Scan Time (sec/scan)</b>	<b><math>\mu</math>Scans</b>	<b>Data Rate (Hz)</b>
0.24	1	4.17
0.48	2	2.08
0.95	4	1.05
1.89	8	0.53
3.76	16	0.27
5.00	21	0.20

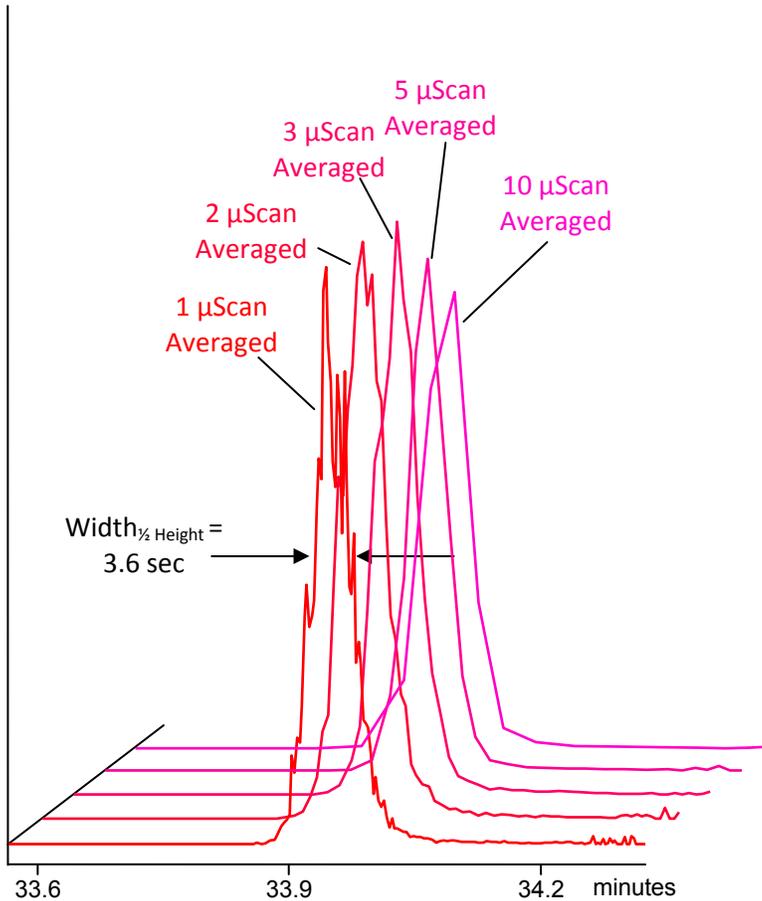
<b>Data Rates<sup>4</sup> for Varian 240/4000 GCMS</b>						
<b>Scan Mode -</b>	<b>Normal</b>		<b>Fast</b>		<b>Fastest</b>	
<b><math>\mu</math>Scans Averaged</b>	<b>Scan Time (sec/scan)</b>	<b>Data Rate (Hz)</b>	<b>Scan Time (sec/scan)</b>	<b>Data Rate (Hz)</b>	<b>Scan Time (sec/scan)</b>	<b>Data Rate (Hz)</b>
1	0.34	2.94	0.31	3.23	0.08	12.50
2	0.67	1.49	0.62	1.61	0.16	6.25
4	1.34	0.75	1.24	0.81	0.32	3.13
8	2.68	0.37	2.47	0.40	0.63	1.59
16	5.36	0.19	4.94	0.20	1.25	0.80
32	10.71	0.09	9.88	0.10	2.49	0.40
64	21.42	0.05	19.75	0.05	4.98	0.20
89	29.79 <sup>5</sup>	0.03	-	-	-	-
97	-	-	29.93 <sup>5</sup>	0.03	-	-
99	-	-	-	-	7.70 <sup>5</sup>	0.13

<sup>4</sup> Displayed data rates are for the specified mass range and segment count. Other settings for these settings will result in differing scan times and data rates.

<sup>5</sup> These  $\mu$ Scans Averaged settings are the maximum allowed values.

## 240 - Normal Scan Mode

uScans Averaged	Points above Half Width	Relative Area	Relative Peak Height
1	20	1.00	1.00
2	13	0.97	0.99
3	7	0.95	0.99
5	3	0.87	0.89
10	2	0.87	0.79

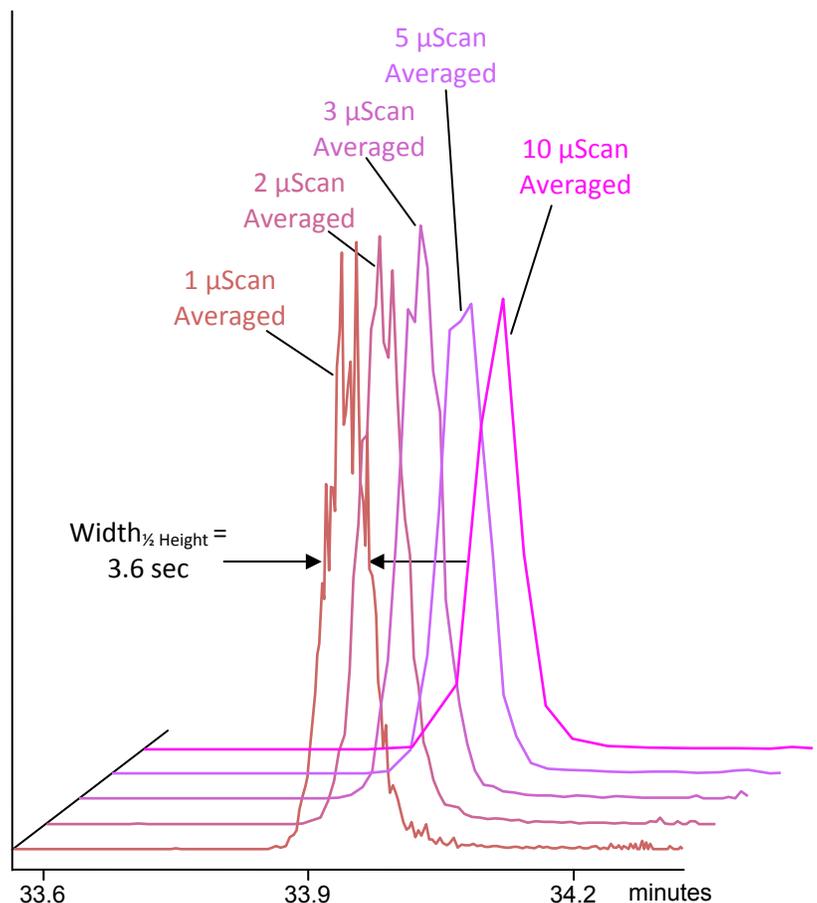


Data rates can have a major impact on peak shapes and displayed noise. The goal for optimum detection is to maintain between 10 and 20 data points across the top of the peak. With a Normal scan mode with the Varian 240, peak areas remain consistent over a narrow range of  $\mu$ Scans Averaged, as well as peak heights. Peaks displayed at left do not any smoothing functions applied. Noise can be suppressed with Mean or Savitsky-Golay smoothing.

## 240 - Fast Scan Mode

uScans Averaged	Points above Half Width	Relative Area	Relative Peak Height
1	22	1.00	1.00
2	10	0.99	0.97
3	7	0.99	0.95
5	5	0.94	0.77
10	2	0.89	0.74

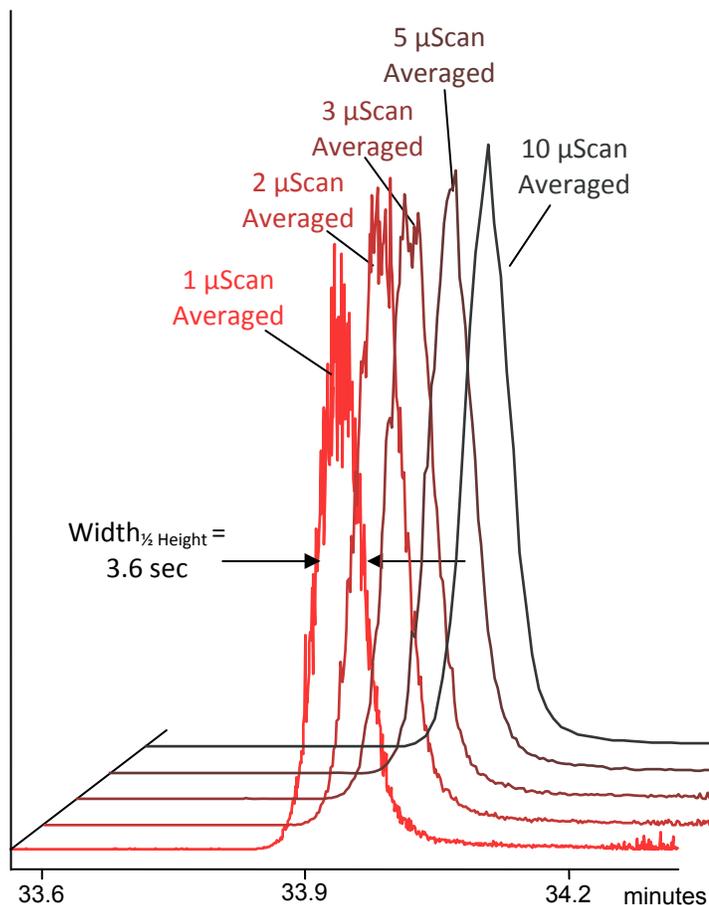
As expected with a limited mass range and similar data rates as with the Normal Scan Mode above, the Fast Scan Mode exhibits nearly the same peak characteristics over a range of  $\mu$ Scans Averaged.



## 240 - Fastest Scan Mode

uScans Averaged	Points above Half Width	Relative Area	Relative Peak Height
1	63	1.00	1.00
2	39	1.00	0.99
3	26	1.00	0.99
5	15	0.98	0.98
10	8	0.97	0.97

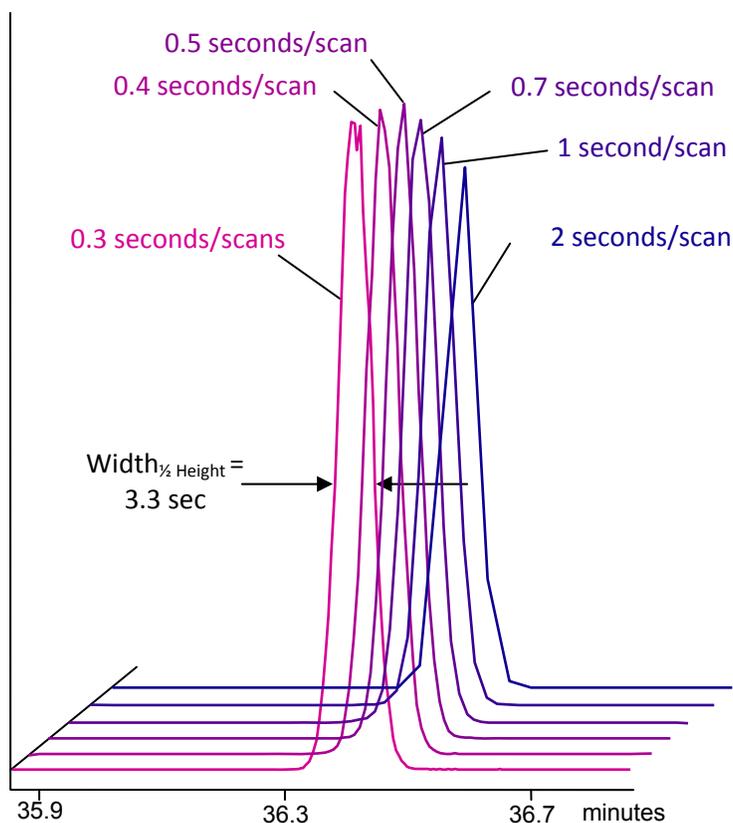
The Fastest Scan Mode is implemented to accurately pick up very sharp chromatographic peaks (< 1 second peak width). However, with "normal" peak widths (typically 3 seconds), the range of  $\mu$ Scans Averaged giving equivalent peak areas is much wider than with the other modes. However, noise is dramatically higher with low settings. Smoothing actions (see below) are needed to curb this degradation of the peak shape and help in making peak detection for the full peak a bit easier.



The Varian 220 parameter entry for data rates differs somewhat from those with the Varian 240. Available choices for the 220 are only in units of Seconds/Scan, and yet demonstrate similar consistencies for a narrow range, with severe degradation if set too coarsely. Performance matches results with 240 - Normal Scan Mode, as anticipated.

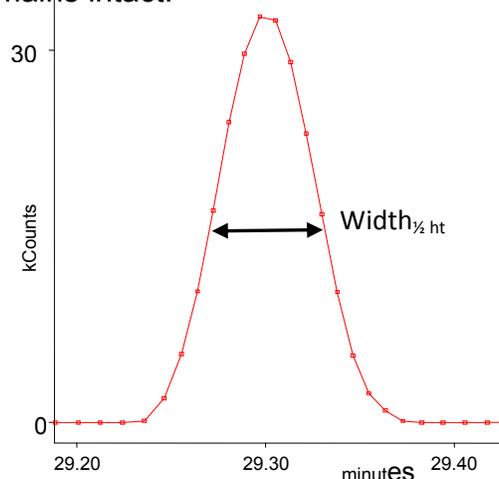
## 220 - Segment Setpoint

Seconds/Scan	Points above Half Width	Relative Area	Relative Peak Height
0.3	12	1.00	1.00
0.4	8	0.98	1.00
0.5	7	0.96	0.99
0.7	5	0.94	0.94
1.0	3	0.92	0.92
2.0	1	0.83	0.80



Data points in a chromatogram are visible in **MS Data Review** screens when the time axis is expanded sufficiently to view them. To achieve full value of the peak height, at least 10 points should span across the top of the peak above the width at half height. Fewer points will degrade the value for the highest point, but remarkably peak area remains intact.<sup>6</sup>

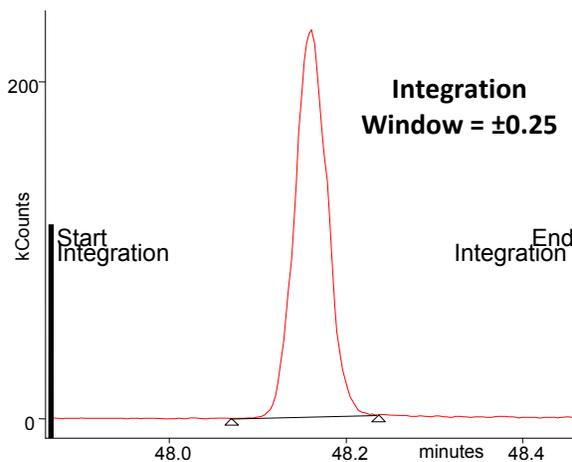
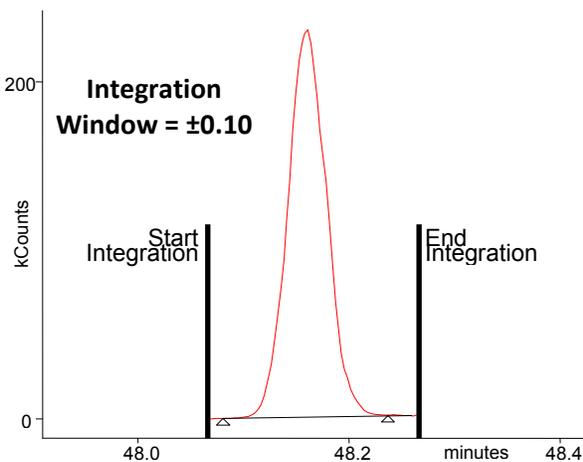
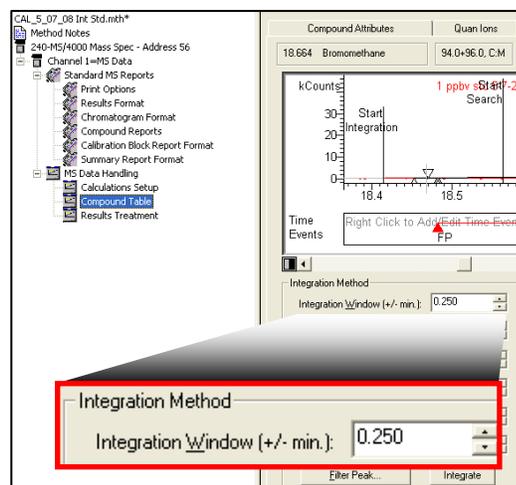
Points across Peak Width $\frac{1}{2}$ ht	Relative Peak Area	Relative Peak Height
$\infty$	1.00000	1.000
17	0.99999	0.998
9	0.99998	0.993
5	0.99997	0.971
3	0.99986	0.950
2	0.945	0.500



Data points defining the peak are displayed by expanding the time axis until these points become visible.

## Integration Window

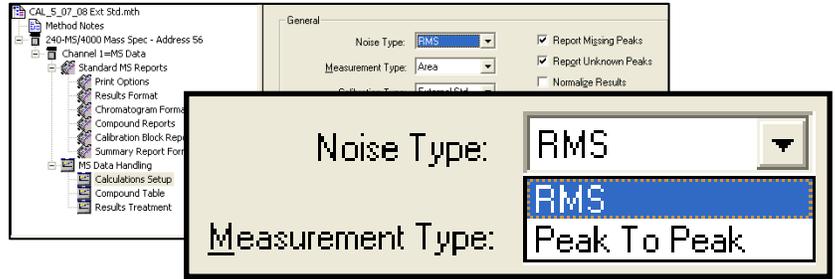
Each identified peak has an assigned window where all integration must take place. Any portion of the chromatogram outside this window is not considered for that peak, with the exception of the noise window. The position of the window is centered on the expected retention time of the peak in the Compound Table. A window for each identified peak is assigned in **Method** → **MS Data Handling** → **Compound Table** → **Integration Method**. The preset value is  $\pm 0.25$  minutes. Size of this window can be readjusted through **MS Data Review** on archived .SMS files, and its absolute location relative to the peak can be changed by correcting the expected retention time for the peak. This parameter positioning should be adjusted to correlate with the chromatography. A window too narrow for the peak can dramatically alter peak integration and noise assessment. A window too wide can effect a switch in peak assignment, especially for Identification Reference Peaks, to an adjacent peak now within the window.



<sup>6</sup> Data presented in this table are generated from synthesized data, without consideration of noise.

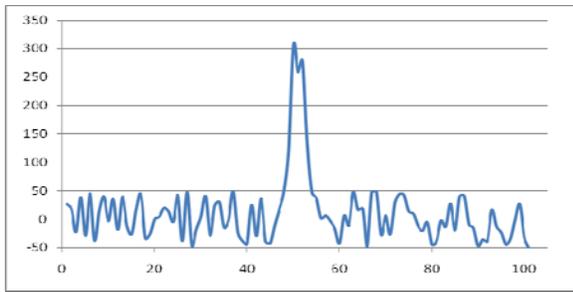
# Computation of Noise for Peak Detection

Peaks are separated from uninformative noise by assessing the level of noise surrounding identified peaks. Then a peak start is sensed when the signal exceeds this established noise. And peak end is perceived when the back of the peak approaches this noise.

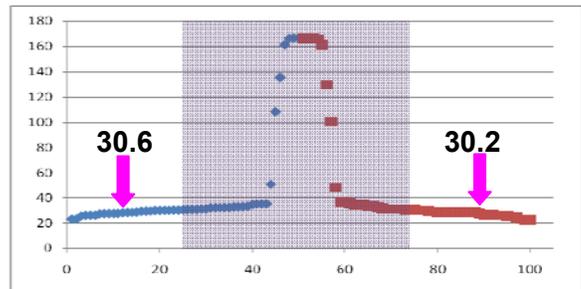


The type of noise is computed and reported based on the user's choice in Calculations Setup in the Method. Peak-to-Peak is simply the difference between high and low data points within a 5-data point window (as shown below). Root-mean-square (RMS) noise is an average of two 10-data points windows - one before the apex and one after. RMS noise is computed by the formula:

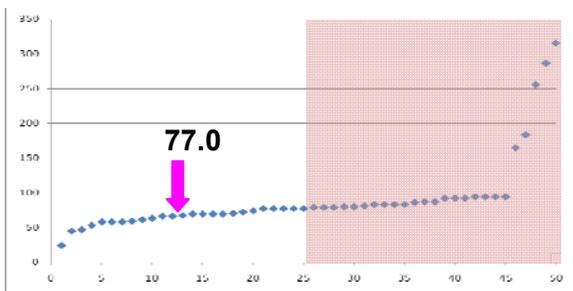
$$y_{rms} = \sqrt{\frac{(y_1^2 + y_2^2 + \dots + y_{10}^2)}{n}}$$



**Synthetic chromatogram provides data points to illustrate the measuring of noise levels.**

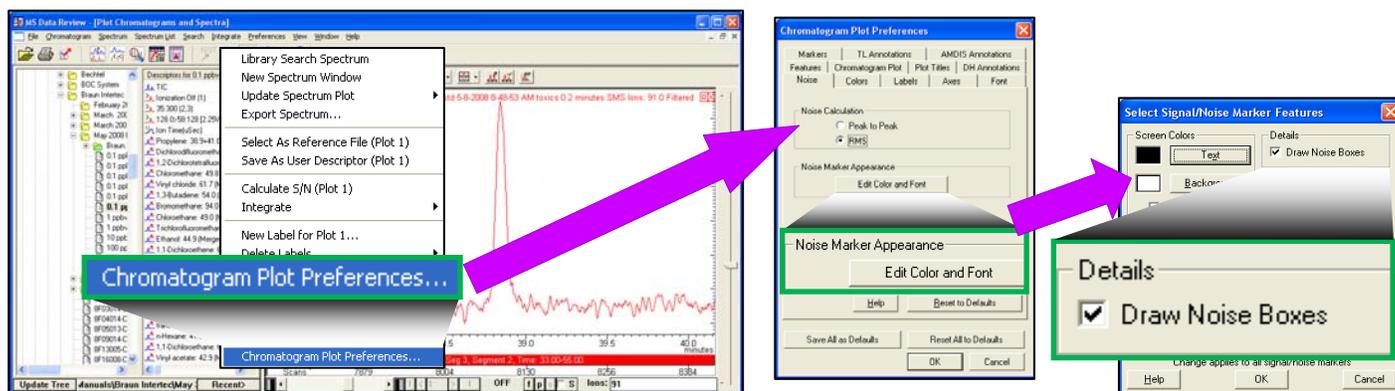


**Root Mean Square (RMS)** is applied to sets of ten points across +/- (50 \* data rate) points from the apex of the synthetic chromatogram. Results then are sorted from low to high to the peak apex, and then high to low to the end. The middle 50% of the data points are excluded, and the numerical average of the two middle points of the remaining for each side of the apex becomes the basis for the RMS noise: 30.4.



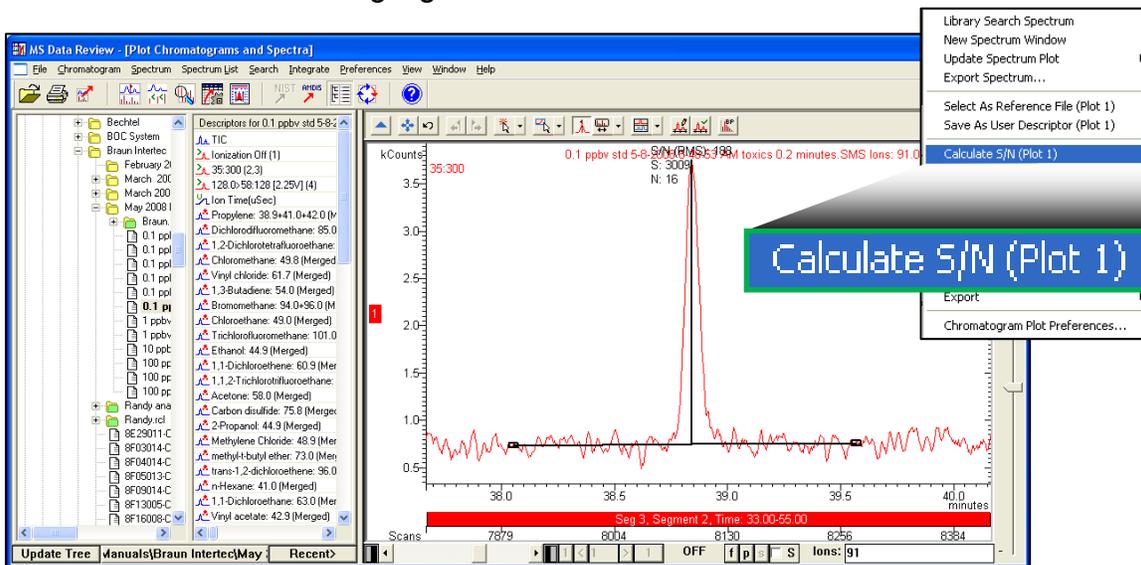
**Peak-to-peak noise (P-P)** is computed by taking the difference between the high and low values within sets of five data points across +/- (50 \* data rate) points of the synthetic chromatogram. The differences are then sorted from low to high. 50% of the high ones are excluded and middle point of the remaining becomes the basis for the P-P noise: 77.

The noise areas used for noise assessment can be visually displayed through **MS Data Review** → **Plot Chromatograms and Spectra**, then right-click mouse anywhere in the chromatogram pane to select Chromatogram Plot Preferences from the menu. Choosing “**Noise**” → **Noise Marker Appearance** - **Edit Color and Font** will pull up a window with a check box to **Draw Noise Boxes**.



The steps to enable display of the chromatographic regions involved in setting the noise value for peak triggering are shown.

Then the peak apex is picked with the mouse pointer and **Calculate S/N (Plot 1)** is chosen, resulting in the areas used to be highlighted.



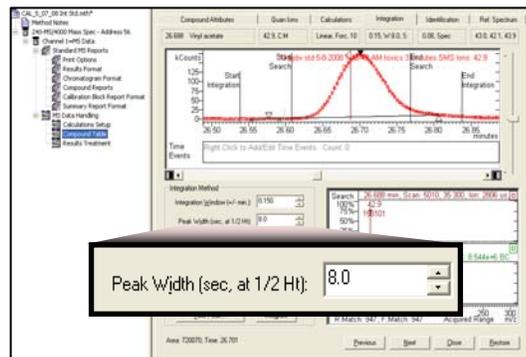
A peak for 3 ppt Benzene (300 ml loading, at 78 m/z, 11-point Saviksky-Golay smoothing) is used to illustrate how regions are shown when computing RMS Noise.

The noise level found is displayed as “N” at the peak apex. This noise value is used for both peak detection and computation of Signal-to-Noise for the peak. The search for the lowest noise interval can extend outside the Integration Window, but will not cross segment changes, such as switching between Full Scan and MSMS. If insufficient space is allotted for this noise evaluation, the window is allocated up to the change.

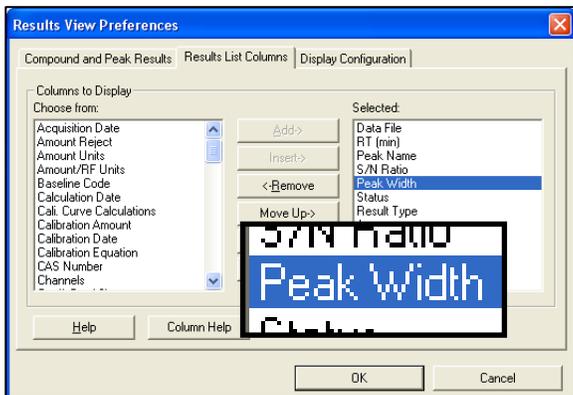
The Noise areas used for peak detection cannot be displayed in the **MS Data Review** → **View Results** → **Chromatogram Pane**, as these are often outside the Integration Window and would not be available for display here.

## Expected Peak Width

The peak detection algorithm is set to be able to sort out unexpected peak shapes. For example, if peaks are anticipated to be sharp from “fast” chromatography, then broad baseline drift or background humps should be ignored. At the other end of the scale, if peaks are predicted to be wide, then sharp “spikes” are to be considered as disruptions, and thus disregarded. The precise setting for this parameter is not critical as Varian MS Workstation readily locates a peak even when expected peak widths do not correlate precisely with the actual widths of detected peaks. The preset value is 4.0 sec. An appropriate entry can be found in Results View when Peak Width is added.



The preset value is 4.0 sec. An appropriate entry can be found in Results View when Peak Width is added.



RT (min)	Peak Name	S.	Peak Width
16.016	Propylene	587	5.9 sec
16.152	Dichlorodifluoromethane	14861	5.0 sec
16.671	1,2-Dichlorotetrafluoroethane	12811	4.9 sec
16.886	Chloromethane	1405	5.7 sec
17.336	Vinyl chloride	4375	5.3 sec
17.492	1,3-Butadiene	2451	5.3 sec
18.674	Bromomethane	2391	4.9 sec
19.113	Chloroethane	3817	5.0 sec
20.051	Trichlorofluoromethane	46704	5.1 sec
21.331	Ethanol	3374	6.7 sec
22.043	1,1-Dichloroethene	8940	4.5 sec
22.066	1,1,2-Trichlorotrifluoroethane	30382	5.2 sec
22.598	Acetone	15	N/A X M
22.660	Carbon disulfide	5335	4.5 sec
23.285	2-Propanol	2357	5.3 sec
24.014	Methylene Chloride	7785	4.1 sec
24.867	methyl-t-butyl ether	5133	4.9 sec

## Detection of Peak Start with Slope Sensitivity

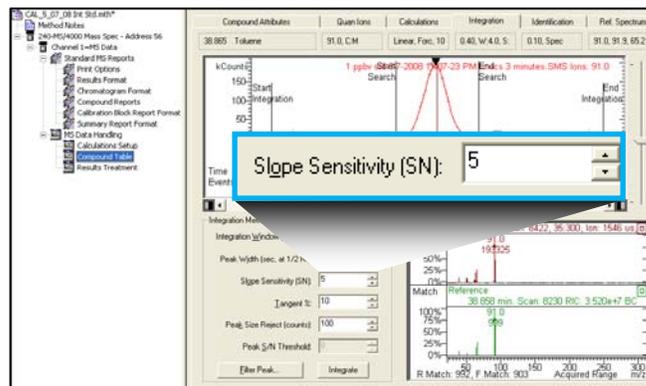
Precise details of the peak detection algorithm in Varian MS Workstation are proprietary. The concepts discussed here are intended to provide a basic understanding of this process to improve the quality of the final results. User-selectable parameters for peak detection can have major impacts of the appropriateness of locating peak start and end points.

After a chromatogram is run, its first derivative is performed<sup>7</sup>, using Cluster points, to determine peak starts. Based on the Noise threshold from the Noise Monitor interval prior to the start of the run and the Initial S/N Ratio, a peak start point is triggered when the derivative value first exceeds this value.

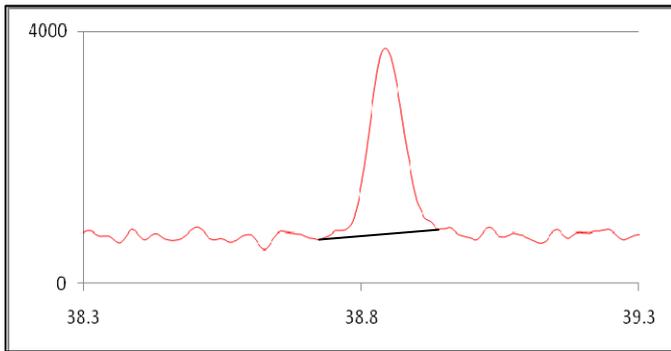
This action point is:

$$\text{Trigger} = \text{Noise} \times (\text{Slope Sensitivity})$$

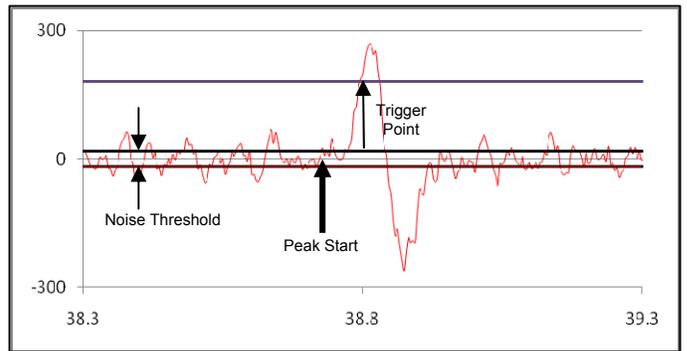
Slope Sensitivity can be adjusted for individual peaks for optimum settings. The available range of values is 1 to 256.



<sup>7</sup> Displayed derivatives of real data are generated through Microsoft Excel after conversion of raw data by ASCII operations.



**A peak for 3 ppt V Toluene is used as an example of the peak detection process.**

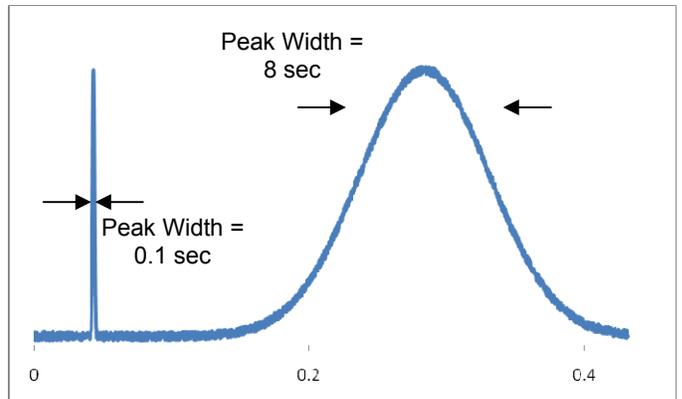


**The region at the front side of the derivative shown above illustrates locations for the trigger point at 5 X Noise and actual Peak Start.**

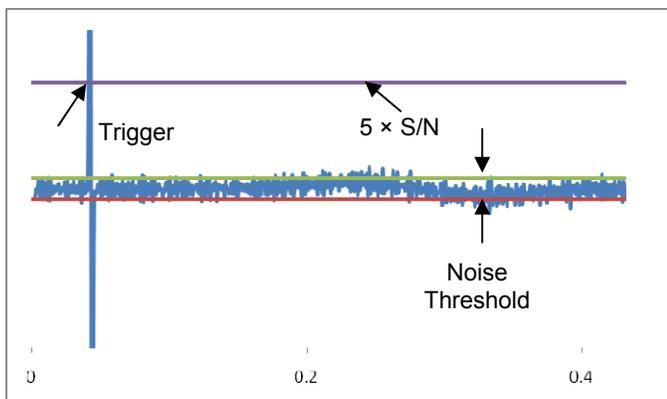
The last 60 Cluster Points prior to this trigger point are examined backwards to find a local minimum, which is then taken as the Peak Start. This step ensures that the true peak start is determined into the preceding noise and the computed area includes an area slice under the whole peak that would have been excluded if the peak start had been set at the trigger point.

To illustrate the process for peak detection, several case studies demonstrate how expected peaks are reported and undesired peaks are ignored.

If very narrow peaks are anticipated, then appropriate parameters, specifically expected Peak Width and Signal-to-Noise, need to be set to pick out the sharp peaks and disregard baseline drifts and unexpected broad peaks. Then when the derivative of the signal is performed, the target peak is properly detected and the broad "baseline drift" is ignored. With judicious choice of Expected Peak Width (0.5 sec) and Signal-to-Noise (5), only the sharp, narrow peak breaks the Signal-to-Noise Threshold; the broad peak does not come close to the Trigger level and is disregarded and not reported.

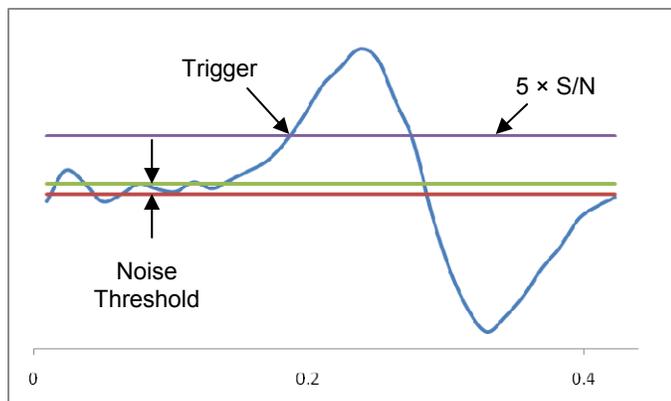


**A very narrow peak is synthesized along with a broad peak. With an expected Peak Width of 0.5 sec, the narrow peak is targeted and the broad peak will be treated as "uninteresting" baseline drift.**

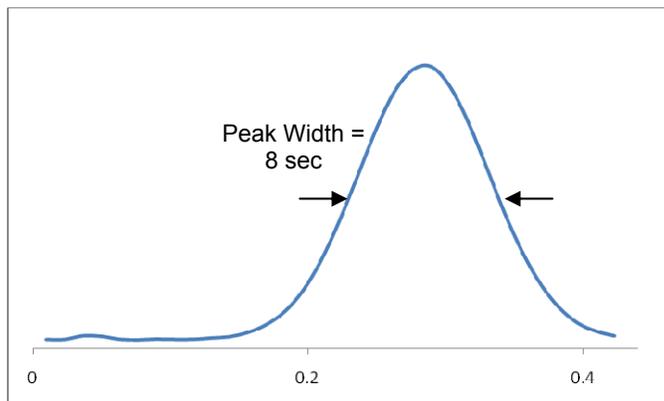


**The derivative of this chromatogram readily picks out the narrow peak as it easily exceeds the trigger point. The broad peak barely breaks the Noise threshold and is not detected.**

Then, when broad peaks are anticipated, expected Peak Width is widened to match the desired peak, and the consequential averaging often completely suppresses narrow peaks, as noise, and boosts detection of the wider target peak.



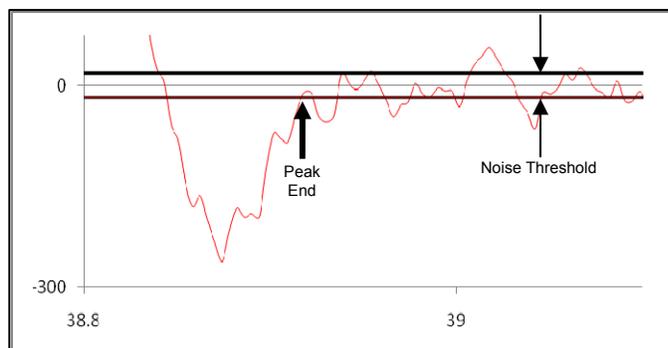
**A derivative of the chromatogram, with a  $WI=8$ , gives a different sense to possible peak starts, as the narrow peak is hardly sensed, and noise on the broad peak is radically suppressed and its derivative readily exceeds the threshold.**



**When broad peaks (width @  $\frac{1}{2}$  height of 8 sec) are expected, the expected Peak Width can be set to match their widths ( $WI=8$ ) and sharp peaks will melt into the noise.**

## Detection of Peak End

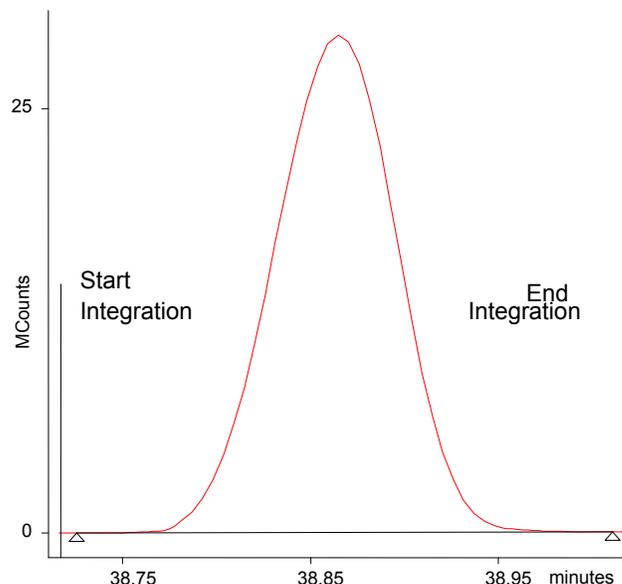
Peak end is detected when the derivative level returns into the Noise Threshold window. If this signal promptly zooms in a positive direction, then a fused peak is detected and a judgment is made with TAN% to determine if the rider peak requires a perpendicular drop to baseline or a tangent skim (see discussions below).



**Peak End point is set when the derivative signal returns into the Noise Threshold window.**

## Interaction of Expected Peak Widths and Slope Sensitivity

Effects of Expected Peak Width and Slope Sensitivity on area counts are studied with a Toluene peak at both high and low concentrations. Each chromatogram is opened in **MS Data Review** and the two parameters are varied over wide ranges. Areas are normalized to the correct area. This series shows how these two parameters have wide latitude in values for major peaks. However, small peaks approaching the noise level have more constraints on adequate values, without some manual interventions discussed below. Data smoothing helps widen the choices for these parameters without severely impacting computed areas.



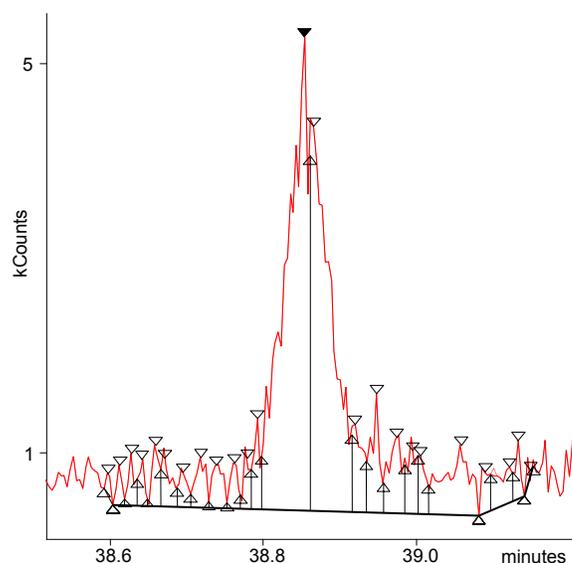
**Major peak for Toluene, 100 ppbV, 300 ml loading, at 91 m/z, raw data (no smoothing), measured peak width - 3.4 sec, Slope Sensitivity - 4, Expected Peak Width - 4 sec.**

		Expected Peak Width (sec.)							
		0.5	1	2	4	8	16	32	64
Slope Sensitivity	1	0.584	0.585	0.612	1.000	1.000	0.998	0.000	0.000
	2	0.584	0.585	0.612	1.000	1.000	0.998	0.000	0.000
	4	0.584	0.585	0.612	1.000	0.998	0.998	0.000	0.000
	8	0.584	0.584	0.612	1.000	0.998	0.998	0.000	0.000
	16	0.584	0.584	0.612	0.999	0.999	0.998	0.000	0.000
	32	0.584	0.584	0.612	0.998	0.999	0.998	0.000	0.000
	64	0.582	0.584	0.612	0.998	0.998	0.998	0.000	0.000
	128	0.581	0.583	0.612	0.998	0.998	0.998	0.000	0.000

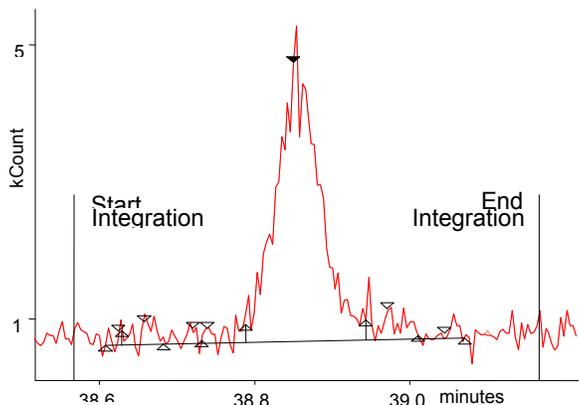
**Expected Peak Width and Slope Sensitivity varied for Toluene peak shown on right. Areas are normalized to the correct area.**

		Expected Peak Width (sec.)							
		0.5	1	2	4	8	16	32	64
Slope Sensitivity	1	0.175	0.174	0.196	0.606	0.605	1.050	0.000	0.000
	2	0.162	0.181	0.196	0.606	0.603	1.000	0.000	0.000
	4	0.051	0.173	0.196	0.606	0.603	1.076	0.000	0.000
	8	0.361	0.173	0.192	0.606	0.647	1.038	0.000	0.000
	16	0.000	0.000	0.174	0.475	0.611	0.869	0.000	0.000
	32	0.000	0.000	0.000	0.164	0.439	0.756	0.000	0.000
	64	0.000	0.000	0.000	0.000	0.000	0.638	0.000	0.000
	128	0.000	0.000	0.000	0.000	0.000	0.889	0.000	0.000

**Expected Peak Width and Slope Sensitivity varied for Toluene peak shown on right and below.**



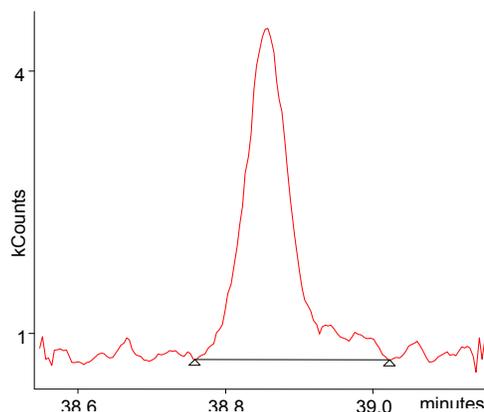
**Trace peak for Toluene, 3 pptV, 300 ml loading, at 91 m/z, raw data (no smoothing), Slope Sensitivity - 4, Expected Peak Width - 4 sec.**



**Trace peak for Toluene, 3 pptV, raw data (no smoothing), measured peak width - 3.4 sec, Slope Sensitivity - 2, Expected Peak Width - 16 sec.**

Judicious setting of Expected Peak Width wider than the actual peak width can help dampen noise through additional bunching of data points, and peak starts and ends are closer to points that are more likely to be selected by the operator. However, the acceptable range for the parameter choices is very narrow and will likely necessitate manual interventions with every chromatogram.

A better procedure might be to add in some smoothing to peaks to significantly reduce noise and allow the peak detection process to have a bit more flexibility to maintain consistent area assignments.



**Trace peak for Toluene, 3 pptV, 11-point Savitsky-Golay smoothing, measured peak width - 3.4 sec Slope Sensitivity - 4, Expected Peak Width - 4 sec.**

**Expected Peak Width (sec.)**

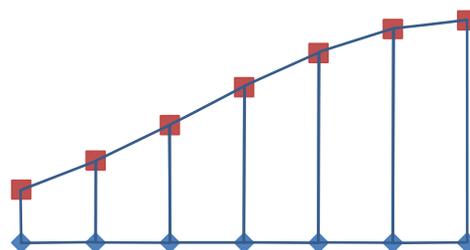
		0.5	1	2	4	8	16	32	64
Slope Sensitivity	1	1.001	0.999	1.026	0.993	1.048	1.009	0.000	0.000
	2	0.838	0.961	1.026	0.989	1.048	1.009	0.000	0.000
	4	0.674	0.858	0.962	1.000	1.053	1.009	0.000	0.000
	8	0.354	0.772	0.863	0.966	1.053	1.007	0.000	0.000
	16	0.000	0.480	0.833	1.049	0.960	1.007	0.000	0.000
	32	0.000	0.000	0.674	1.019	1.047	0.866	0.000	0.000
	64	0.000	0.000	0.000	1.021	1.047	0.866	0.000	0.000
	128	0.000	0.000	0.000	0.000	1.047	0.866	0.000	0.000

**Expected Peak Width and Slope Sensitivity varied for Toluene peak shown on right. Areas are normalized to the correct area.**

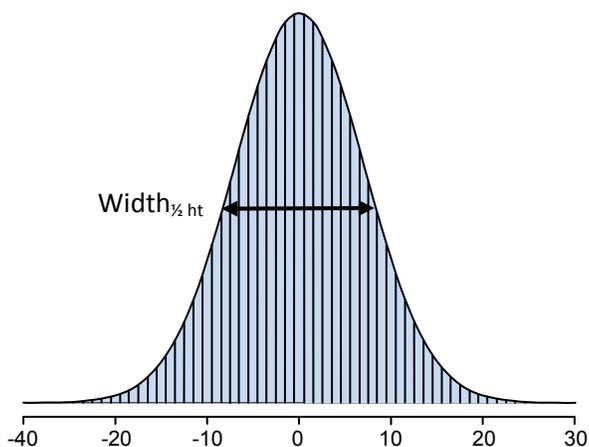
A major impact on area counts can occur when the Expected Peak Width does not closely match the actual peak width, especially with smaller peaks. In some cases the peak is missed completely, or in others the peak is split into multiple peaks with significantly reduced areas for the portion picked out for the analyte. Various manual manipulations are available to force the start and end points for the peak (as discussed below), but tedious review of every chromatogram can be drastically reduced by setting up these parameters to pick out the right areas for a wide dynamic range of concentrations automatically.

**Process for Computing Peak Areas**

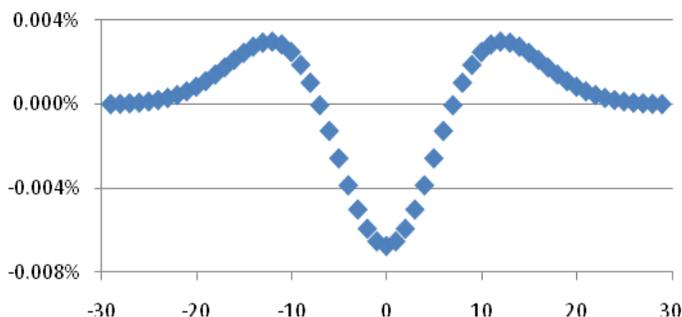
Varian MS Workstation computes areas for detected peaks by a summation of trapezoid areas found between raw data points in the curve and assigned baseline. This simple, but elegant, approach represents accurately the true shape of the peak, particularly if sufficient data points are collected to give at least 10 points across the top of the peak above the width at half height.



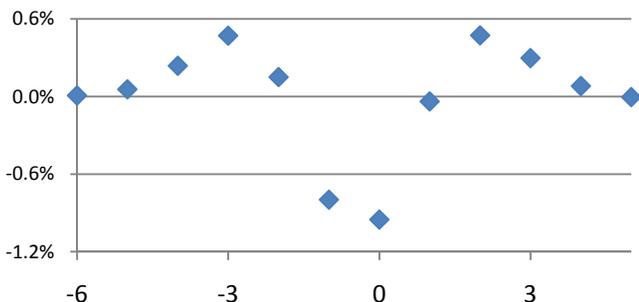
**Area of a Gaussian curve simulated by summation of trapezoids.**



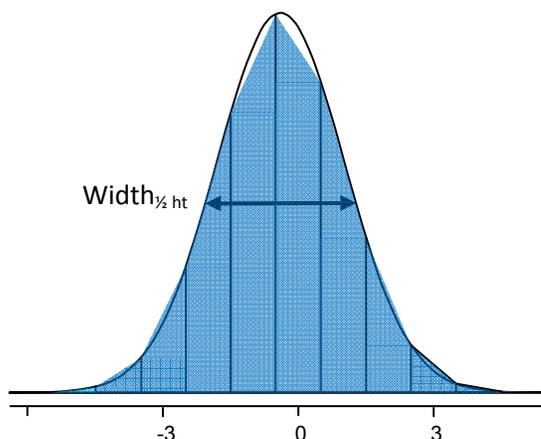
**Summation of Trapezoids with 16 points across top of peak.**



**Percent deviations of area by trapezoids compared with true Gaussian curve, 16 points across top. Total area by trapezoids is 0.99999 of true value**



**Percent deviations of area by trapezoids compared with true Gaussian curve, 3 points across top. Total area by trapezoids is 0.99986 of true value.**



**Summation of Trapezoids with 3 points across top of peak, with top data point not at apex of Gaussian peak.**

Measured Peak widths can be listed in reports and also in **MS Data Review** → **Results View**. The preset display does not have the peak width column enabled and can be added by going to **MS Data Review** → **Preferences** → **Results View** → **Peak Width**. The value listed is the computed peak width at half height, reported in seconds. A similar process is performed to include peak widths in the printed report.

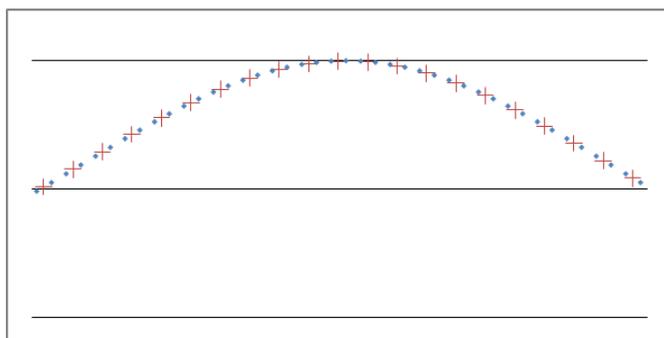
RT (min)	Peak Name	Peak Width
MS 20.060	Trichlorofluoromethane	5.1 sec
MS 21.310	Ethanol	6.7 sec
MS 22.042	1,1-Dichloroethane	4.6 sec
MS 22.059	1,1,2-Trichlorotrifluoroethane	5.2 sec
MS 22.547	Acetone	4.5 sec
MS 22.630	Carbon disulfide	3.6 sec
MS 23.273	2-Propanol	5.6 sec
MS 24.015	Methylene Chloride	3.9 sec
MS 24.912	methyl-t-butyl ether	5.3 sec
MS 24.879	trans-1,2-dichloroethane	4.0 sec
MS 25.694	n-Hexane	6.0 sec
MS 26.466	1,1-Dichloroethane	4.4 sec
MS 26.704	Vinyl acetate	3.9 sec
MS 28.588	cis-1,2-Dichloroethane	3.8 sec
MS 28.685	2-Butanone	5.3 sec
MS 28.880	Ethyl Acetate	3.4 sec
MS 29.406	Bromochloromethane	4.5 sec
MS 29.539	Tetrahydrofuran	4.7 sec

These values can then be used to set the data rate for data collection to assure that sufficient points are created to properly define the shape of that chromatographic peak. For example, if the peak width at ½ height is 4.5 seconds for Acetone, then the data rate should be selected to be near 2 Hz by the formula:

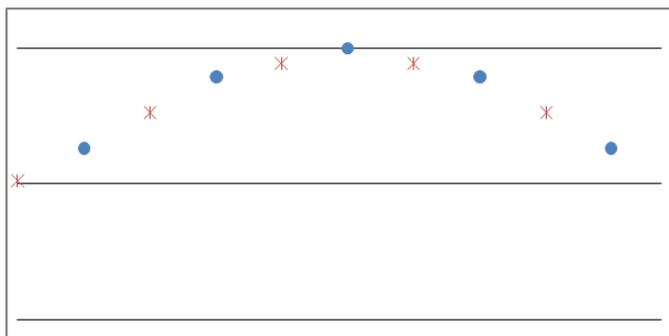
$$\text{data rate} = \frac{10}{\text{peakwidth @ } \frac{1}{2} \text{ height}}$$

Surprisingly areas can be measured quite nicely with dramatically fewer data points than might be anticipated. Typically, at the apex region of a peak, areas by trapezoids are underreported, whereas areas are overvalued at the leading and trailing edge of the peak. These errors are often cancelled out, and the area comes in very close to the true value. One strong advantage of slower data rates is the related reduction in noise, as more raw  $\mu$ Scans are represented in a single data point.<sup>8</sup>

If data points within the Expected Peak Width number more than 20, “bunched” points (+) are created by averaging adjacent data points (♦) to yield between 10 and 20 points across the Expected Peak Width. This process helps to reduce noise by averaging while still maintaining sufficient points to define a peak, and also to better locate peak start and end points.



**Bunching of data points is illustrated with this display of the top half of a Gaussian peak with a peak width at ½ height of 4 seconds. Blue markers are data points at 10 Hz. With 40 data points across the top of the peak, bunched points (red markers) are created by averaging two data points. These bunched points are used for peak processing, but are not displayed.**



**Interpolated points (red markers) are provisionally created when insufficient data points (blue markers) are collected across the top of the peak. The original data rate is 0.8 Hz with peak width of 4 seconds yielding only 5 points across the top. These interpolated points are included in peak processing, but are not revealed.**

At the other extreme, if insufficient data points are available to give 10 to 20 data points across the Expected Peak Width, interpolated points (\*) are temporarily created midway between data points (♦) only for the process of peak detection and integration. Then they are discarded and are not shown.

<sup>8</sup> Signal-to-noise ratios should not be used to judge performance of a mass spectrometer, including sensitivity and detection limits, as the noise value can be manipulated to virtually any number with variations in data rates and mathematical smoothing functions. A better approach is to use statistical methods to determine detection limit performance with repetitive runs of a signal at or near the expected detection limit (see, for example, Electronic Code of Federal Regulations, Title 40, Section 136, Appendix B, ([ecfr.gpoaccess.gov/cgi/t/text/text-idx?c=ecfr&sid=5cb626988c883dd36215bd44abeb4ae4&rgn=div5&view=text&node=40:22.0.1.1.1&idno=40#40:22.0.1.1.0.1.7.2](http://ecfr.gpoaccess.gov/cgi/t/text/text-idx?c=ecfr&sid=5cb626988c883dd36215bd44abeb4ae4&rgn=div5&view=text&node=40:22.0.1.1.1&idno=40#40:22.0.1.1.0.1.7.2))).

## Confirm S/N Threshold

Small peaks can be excluded in final reports based on a minimum area counts (see discussions below), and by an assessment of peak height signal relative to surrounding noise - signal-to-noise. Both of these parameters, if set, are active on chromatograms. Many protocols have a target based in signal-to-noise – S/N – as a judgment to exclude small peaks as below reporting limits. These removals can be reported as “Failed” or “Missing” based on the entry for “Report Outliers As”. The noise value used here can be Peak-to-Peak or Root Mean Square (RMS) based on the choice for Noise Type.

The image displays two screenshots of the MS Workstation software interface. The left screenshot shows a chromatogram for a peak at 17.749 minutes, identified as Dichlorodifluoromethane. The peak is labeled with 'MCounts' and 'Search' values. A callout box highlights the 'Peak S/N Threshold' set to 5. The right screenshot shows the 'General' settings for a method, with 'Confirm S/N Threshold' checked. A callout box shows the 'Report Outliers As' dropdown menu with options: Failed, Failed, and Missing. The bottom right screenshot shows the 'Noise Type' and 'Measurement Type' settings, with 'Noise Type' set to 'Peak To Peak' and 'Measurement Type' set to 'RMS'.

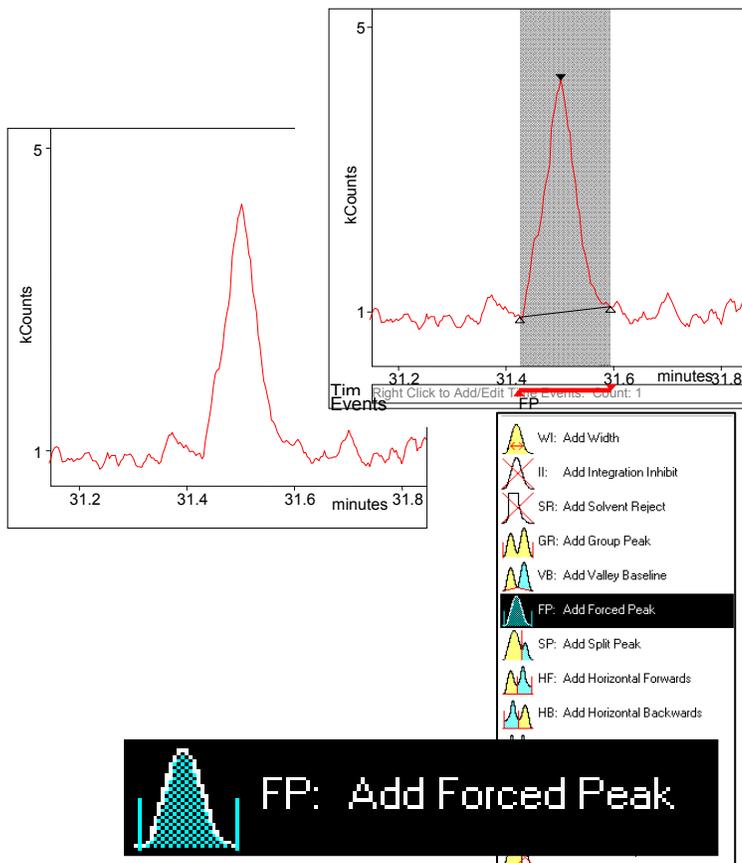
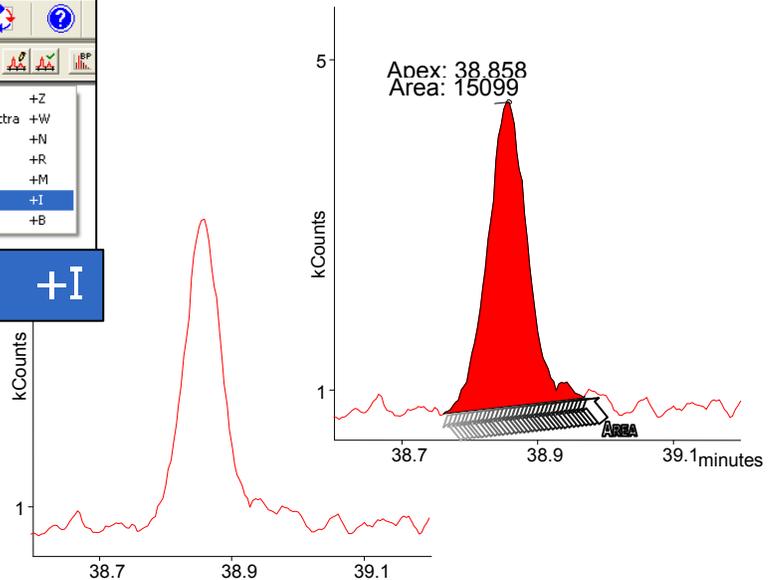
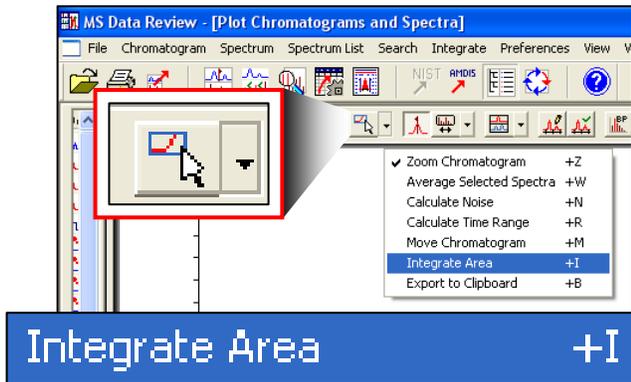
## Automatic Peak Detection

Setting of starting and ending points for a chromatographic peak can have significant impact on computation of its area, and could result in inaccurate and imprecise quantitative results. A myriad of factors enter into this process, including parameters within a MS Workstation method. Fortunately, preset values in a method are often correct for most chromatograms. However, occasionally preset choices turn out to be inappropriate for specific peaks. Adjustments can be performed through **MS Data Review** → **View Results** to visualize the impact of changes, and then the altered method can be saved so that all subsequent runs of a similar type can have the updates applied automatically. A correct choice of parameters will greatly minimize subsequent manual manipulations of chromatograms.

Major peaks in a chromatogram normally do not have difficulty in having their start and end points located, or if these spots are not valid, their wrong positioning likely will have very little impact on area assigned. The problem becomes significantly more important when dealing with small peaks approaching detection limits. Subtle movement in starting and ending sites can yield major changes in the area allocated.

## Manually Integrate Area

Sporadically a peak is not detected, despite attempts to set up proper automatic peak processing prior to data collection, especially small peaks near detection limits. The simplest approach is to activate the process with “Integrate Area” icon in **MS Data Review** and click the mouse at the start of the “peak”, and then drag it to the end of the peak. Retention time and peak area are then displayed. This result is not archived in the data file and not reported with normal results. The value only serves as an indicator for the size of a peak. This operation can occur during collection of a live chromatogram.



## Forced Peak

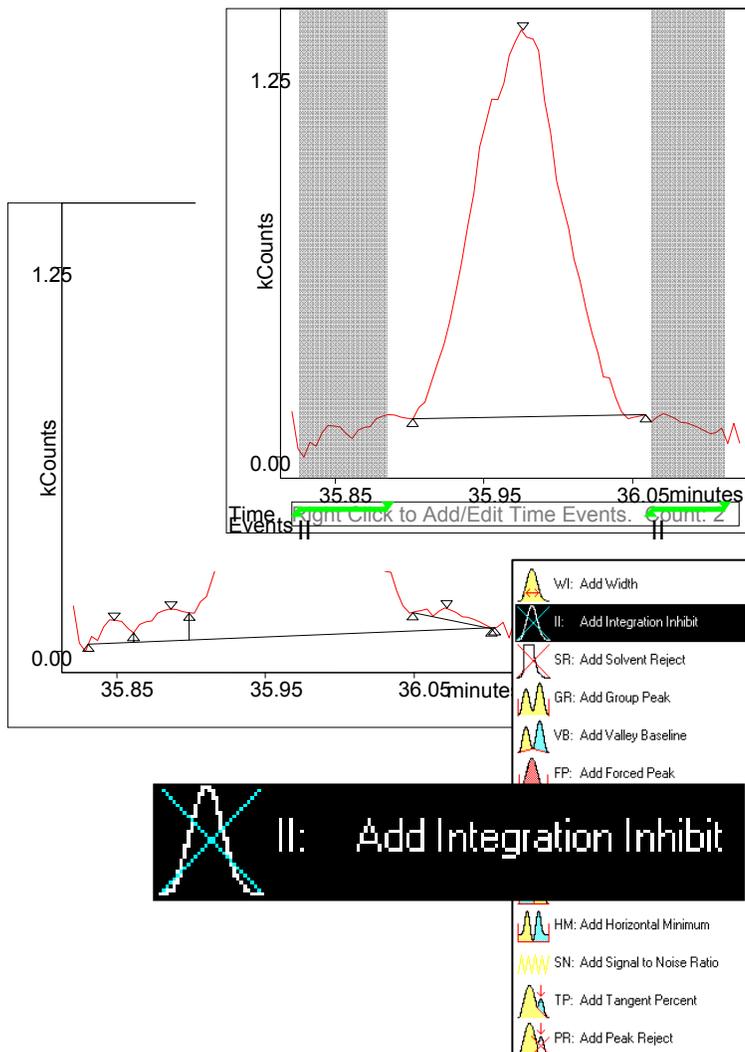
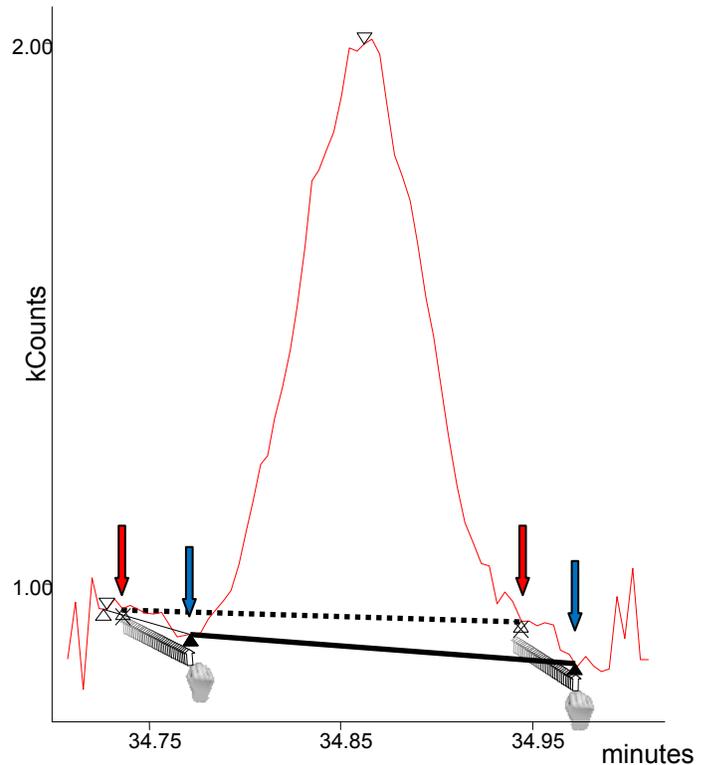
Another choice to create areas for peaks not assigned a baseline, and thus not integrated, is use the time-programmable event “Forced Peak - FP”. This parameter is accessed in **Data Review** → **View Results**. A right mouse click in the Time Events window opens a menu for all events. Selecting “FP” inserts a bar that can be moved around and resized by the mouse to adjust the peak start and end as needed. The peak must be reintegrated for this event to take effect. Results from this action are included with final results.

All time programmable events are method entries and will apply to all chromatograms using this method. They should be used sparingly, as subtle shifts in chromatography can dramatically impact their effects.

## Manually Move Start and End Points

If start and end points are not set appropriately, they can be manually readjusted through **MS Data Review** → **View Results** → **Chromatogram Display** by grabbing the offending triangle with the mouse pointer and dragging it a new desired location. New peak points are indicated by solid black triangles, and original points are crossed out. The new area is generated by “Processing” the run, and results are reported in the final summary. The new peak points can only be moved along the chromatogram tracing and not suspended in “midair”.

This operation will only apply to the active chromatogram and will not be part of any method, and will not impact any other chromatogram.



## Integrate Inhibit

Sometimes the automated process to locate peak start and end points are not quite set as the user would have chosen. One mechanism to force operations is to utilize the time-programmable event “Integrate Inhibit - II”. By placing this event before and after a peak, all data points inside its each span are not considered and the start and end points of the event always become possible baseline points. Established peak detection settings still apply before and after this action

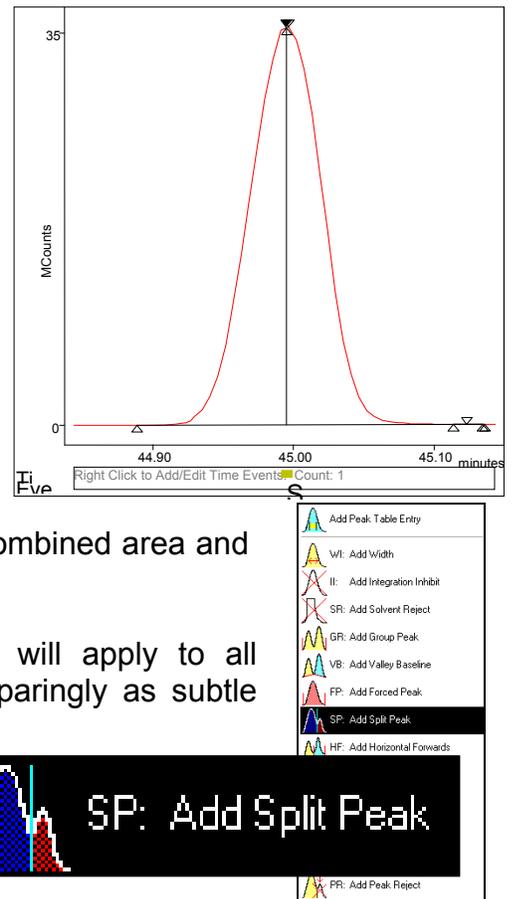
If retention shifts are not dramatic, this event can be set into the method and be applied to all subsequent runs, as peak detection parameters still apply for the span between two of these events.

All time programmable events are method entries and will apply to all chromatograms using this method. They should be used sparingly as subtle shifts in chromatography can dramatically impact their effects.

## Split Peak

Occasionally peaks occur very close together. The chromatography employed, along with smoothing functions, can make their individual areas appear as one peak, maybe with a hint of splitting, but not enough to be picked up as separate peaks by the chosen parameters. The time-programmable event “Split Peak - SP” can be added in to force the peak into two areas. A notable pair in MS lore is meta- and para-Xylenes, with the exact same spectra for both and with nearly the same chromatography. Use of this function can help divvy up areas between the two peaks. This action can be a very subjective call on where to place this event, made a bit easier if the two exhibit a hint of a valley between them. Often these two are reported instead as the combined area and labeled as “m+p-Xylenes”.

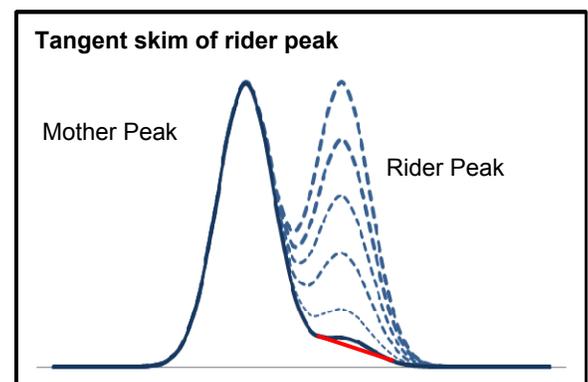
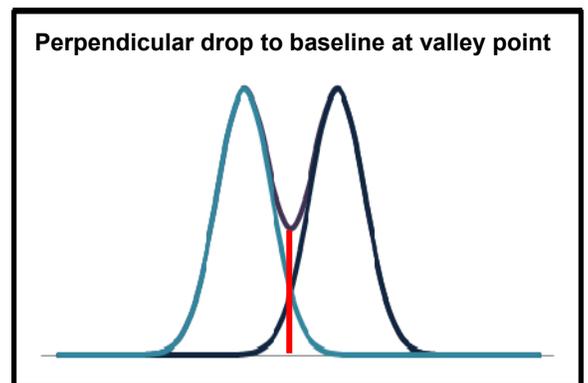
All time programmable events are method entries and will apply to all chromatograms using this method. They should be used sparingly as subtle shifts in chromatography can dramatically impact their effects.



## Tangent Percent

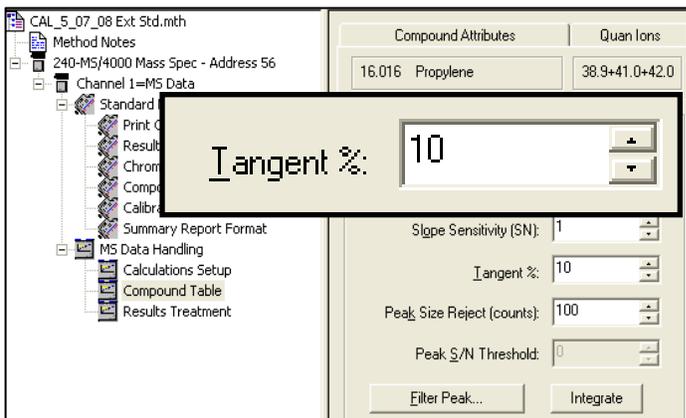
Overlapping peaks can introduce some significant errors during allocation of areas between peaks. If two peaks are equal in size and overlap, a simple drop to baseline at the valley point between the two will allow a reasonable assessment of the areas. However, as a rider peak fades compared with its mother peak, at some point a perpendicular drop does not properly assign areas. The tail of the mother peak adds too much to the area of the rider. Instead, a tangent skim should be applied to the rider peak to provide its “baseline” for computing its area. The triangle under the rider is added to the mother peak’s area. The decision point for switching between perpendicular drop and tangent skim is based on the parameter Tangent Percent - TP. This compares the height of the rider peak to the height of the mother. If the ratio exceeds the Tangent Percent value, a perpendicular drop is executed. If less than this value, a skim is undertaken and marked as “Tangent Skim - TS”.

This allocation of areas is, at best, a crude approximation, but at least small rider peak areas will not be grossly over-exaggerated with inclusion of a portion of the mother peak, and yet when the rider is large, the areas for the two peaks can be reasonably assigned with a perpendicular drop.



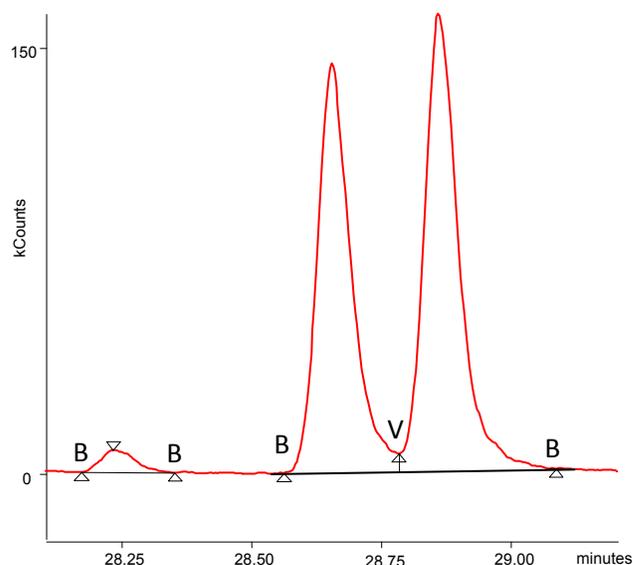
Each analyte in Compound Table can have its own Tangent% parameter. This setting can be time-programmed as needed. The preset value is 10%. A value of 0% sets all judgments to perpendicular drop and 100% treats all rider peaks as tangent.

How the peaks are treated is listed as "Separation Code" in **Results View** and in reports.



	RT (min)	Peak Name	Baseline Code	S	Peak Width	Statu
SMS	16.016	Propylene	VM	587	5.9 sec	
SMS	16.152	Dichlorodifluoromethane	MB	14861	5.0 sec	
SMS	16.671	1,2-Dichlorotetrafluoroethane	BB	12811	4.9 sec	
SMS	16.886	Chloromethane	BB	1405	5.7 sec	
SMS	17.336	Vinyl chloride	MB	4375	5.3 sec	
SMS	17.492	1,3-Butadiene	MB	2451	5.3 sec	
SMS	18.674	Bromomethane	BB	2391	4.9 sec	
SMS	19.113	Chloroethane	MB	3817	5.0 sec	

Baseline Codes	Peak Onset	Peak End
BB	Baseline	Baseline
BV	Baseline	Valley
VV	Valley	Valley
VB	Valley	Baseline
MB	Mended End <sup>9</sup>	Baseline
BM	Baseline	Mended End
MM	Mended End	Mended End
MV	Mended End	Valley
VM	Valley	Mended End
TS	Separated Tangent Peak	
TF	Fused Tangent Peak	
GR	Group Peak	
HF	Horizontal Forward	
HB	Horizontal Backward	
HM	Horizontal Minimum	



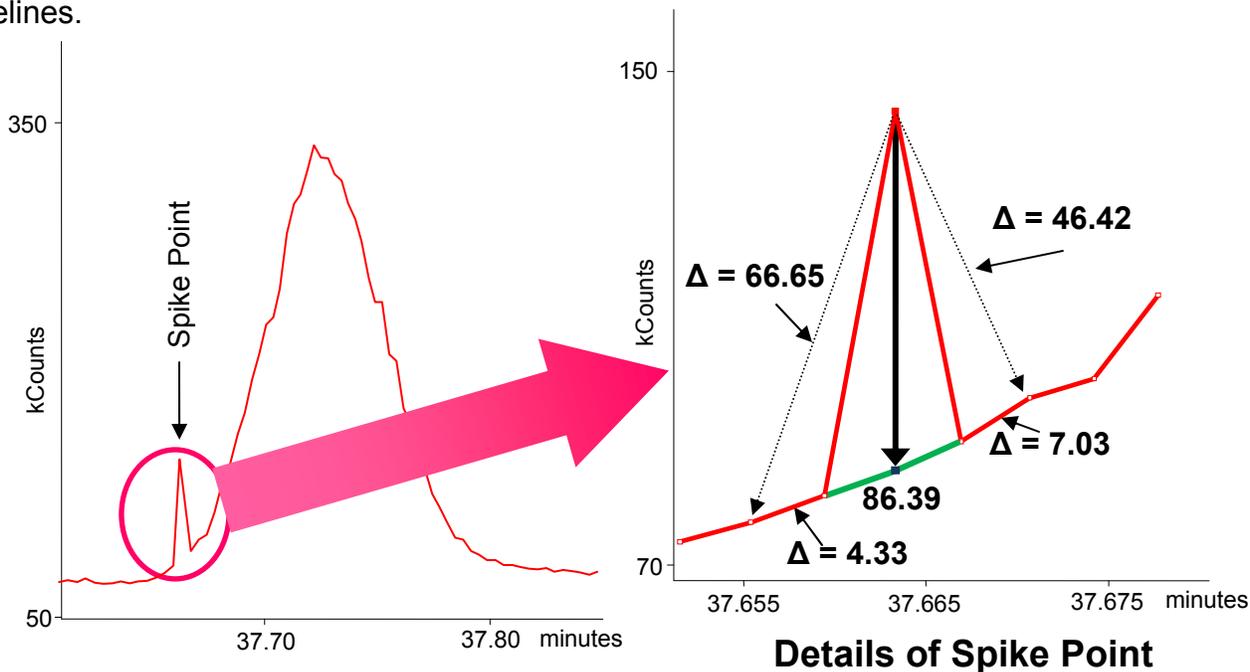
**Representative case where two overlapping peaks have their areas assigned from a perpendicular drop to baseline between them.**

Note: Several time-programmable events are included in the events list for very special functions in gas chromatography and are not generally applicable to mass spectrometer data, including Solvent Reject - SR, Valley Baseline - VB, Horizontal Forward - HF, Horizontal Backward - HB, and Horizontal Minimum - HM. These functions are not discussed here.

<sup>9</sup> Mended End indicates that the baseline has been adjusted for a set of fused peaks generated by a penetration into the straight line baseline for that group.

## Remove Spikes

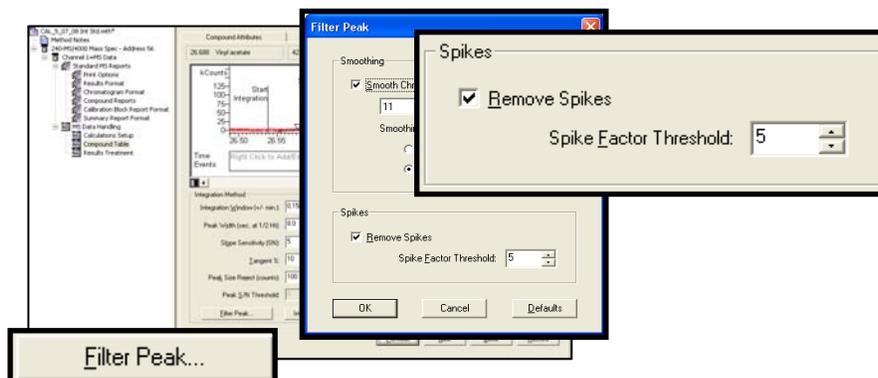
Occasionally mass spectrometers can generate spurious data points from a glitch in the ionization process or a hiccup in the data transfer between spectrometer and workstation. If not treated, this deviant point can artificially split peaks, add area counts, or improperly assign baselines.



These spikes can be removed automatically by setting a Spike Factor Threshold value. This test occurs before quantitation and prior to smoothing actions, if specified. A point is determined to be a spike if difference between its amplitude and that of the spot two points away is greater than the Spike Factor Threshold multiplied with the amplitude difference between the first and second data points away from the candidate point. The threshold must be exceeded on both sides to be a spike. The amplitude of a spike point is replaced by the average amplitude of the four points that surround it. This automatic process helps to distinguish between normal random noise or the onset of a real peak, and a troublesome spike that should not be considered. The available range for this parameter is 2 to 20, with a preset value of 5. The initial setting is with the Remove Spikes “unchecked” and disabled. Spike removal occurs prior to any peak detection processes.

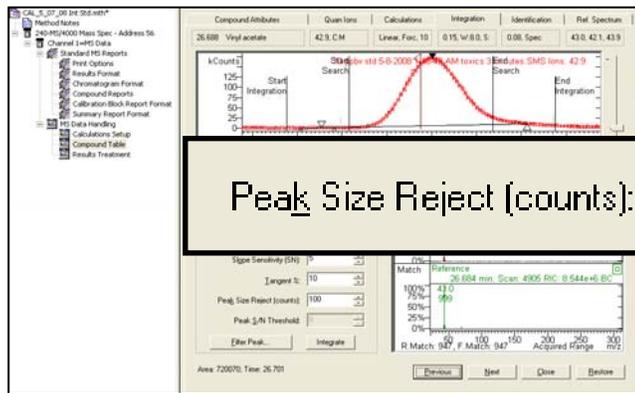
Point	Amplitude	Difference	Threshold $\Delta \times 5$	Spike ?
1	76.95	--	--	
2	81.28	[2-1] 4.33	21.65	
3	143.60	[3-1] 66.65	> 21.65	YES
4	90.15	[3-5] 46.42	> 35.45	YES
5	97.18	[5-3] 7.03	35.45	

<b>Average (1,2,4,5)</b>	86.39
--------------------------	-------



## Peak Size Reject

Peak areas below an established detection limit are not likely to be reportable and can be excluded from consideration in quantitation. As the Varian MS Workstation can eject low areas, this parameter becomes a criterion for eliminating small peaks below detection limits from results. Preset value is 500 and has a range from 0 (no rejection) to 2,000,000,000 [sic]. A valid entry should be slightly lower than the area anticipated for the method detection limit of that compound.



When performing a calibration series, this value can be set higher to ensure exclusion of minor peaks from inappropriate considerations in these computations. Once this process is completed, the reject setting can be lowered to a level to include analytes at their detection limits.

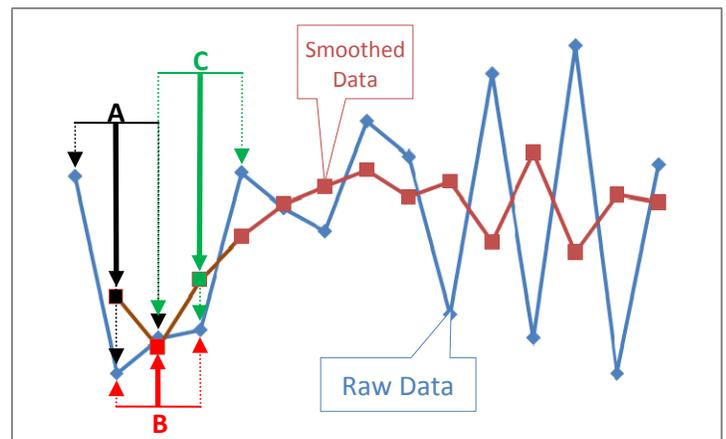
This parameter works in conjunction with Confirm S/N Threshold (see above) to exclude inconsequential peaks. They are making separate and independent judgments on this elimination.

## Smooth Chromatogram

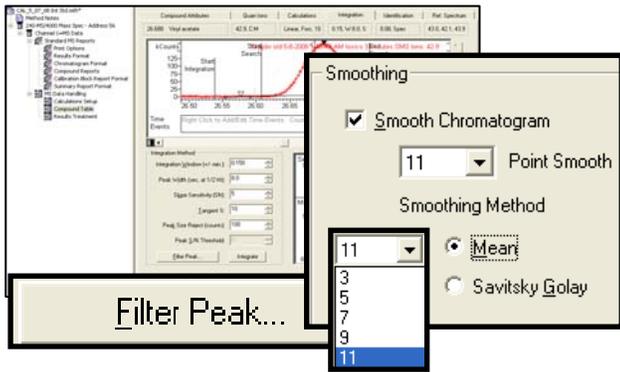
Occasionally chromatograms are tough to integrate, as noise messes up the process to pick out legitimate peaks from baselines. A previously discussed approach to reduce noise is to slow down data collection rates, but still maintain fidelity of the peaks with sufficient data points across the top of the peak. A different approach is to perform mathematical treatments on the peaks to still keep their shapes and reduce the noise level. Varian MS Workstation offers two "smoothing" tactics - Mean and Savitsky-Golay. These smoothing actions are applied to a chromatogram prior to any peak detection process.

Mean smoothing, also called "moving boxcars", takes a specified number of data points before and after a center point, performs a simple average, and replaces that entry into a new data set. Typically the noise is reduced by the square root of the points in the moving average.

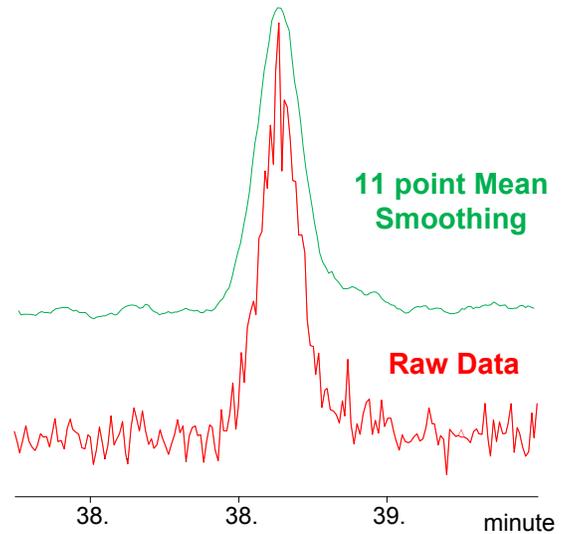
Data Point	Data Point Amplitude	Average for A	Average for B	Average for C
1	63	32.7	19.7	36.7
2	13			
3	22			
4	24			
5	64			



**3-point "mean" smoothing is accomplished by averaging three data points and replacing the center point with average, and then moving to the next point to repeat.**



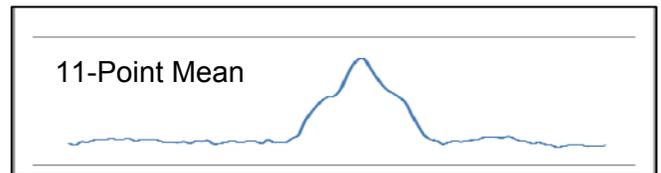
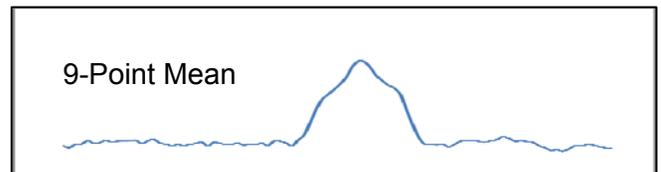
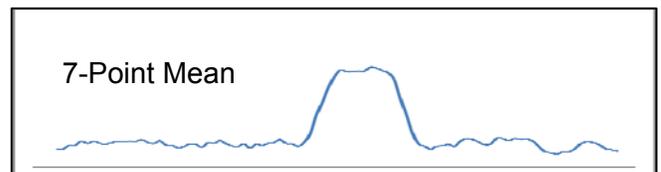
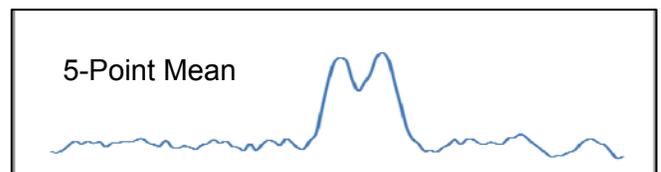
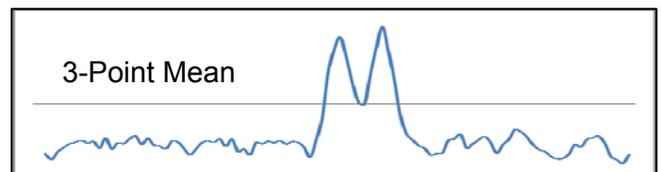
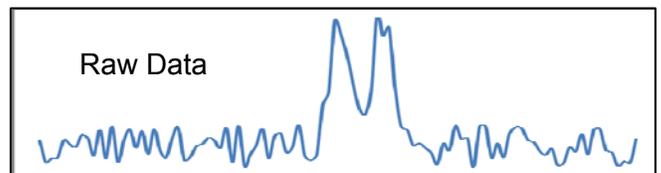
Choices for number of points across the span in the moving Mean are 3, 5, 7, 9 and 11.



Mean smoothing performs as expected in decreasing noise, as shown in a chromatogram at right. Surprisingly peak areas remain almost intact over the full range of smoothing values. However, this process severely degrades peak height at all levels.

Noise improvement is shown for a chromatogram of 3 pptV Toluene, comparing raw data with an 11-point Mean Smoothing.

Number of Points	Signal/Noise - RMS	Peak Area	Change from Raw Data	Height	Change from Raw Data
Raw Data	92	14,873	--	4,439	--
3	136	14,792	99%	3,685	83%
5	215	14,793	99%	3,658	82%
7	267	14,298	96%	3,472	78%
9	335	14,365	97%	3,350	75%
11	404	14,341	96%	3,233	73%



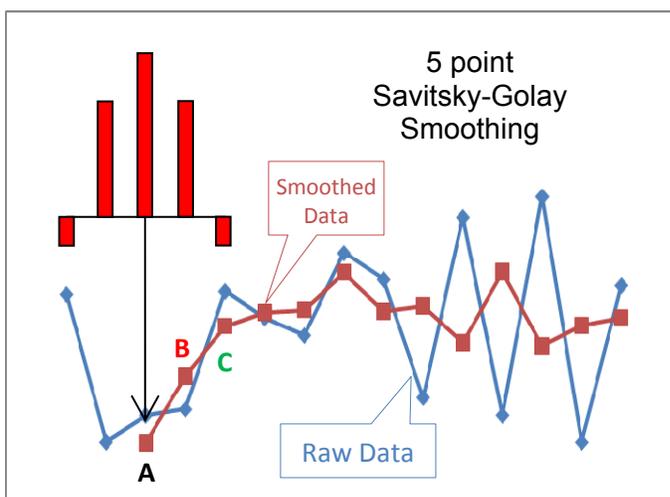
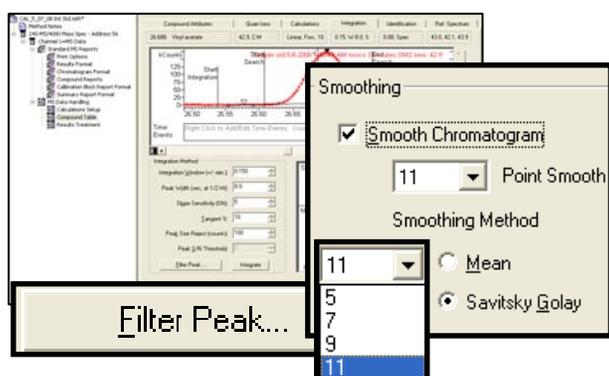
Another distortion of peak shapes by Mean smoothing occurs with two closely eluting peaks. The illustration at right shows the effect of increased averaging. Indeed, noise is diminished with each wider Mean, but any hint of the doublet peak disappears after the 5-point Mean. Wider averaging converts the peak pair into a single peak.

Mean Points	Valley Penetration	Relative RMS Noise
Raw	63%	1.00
3	54%	0.54
5	32%	0.39
7	0	0.31
9	0	0.27
11	0	0.25

An alternative smoothing function is a Savitsky-Golay filter. Like Mean smoothing (discussed above), Savitsky-Golay performs a moving filter through the data set, but instead of applying equal weights to each data point, differing coefficients<sup>10</sup> are applied based on their proximity to the center point. This application smoothes out random noise, but does not distort peak shapes as much as a Mean average would.

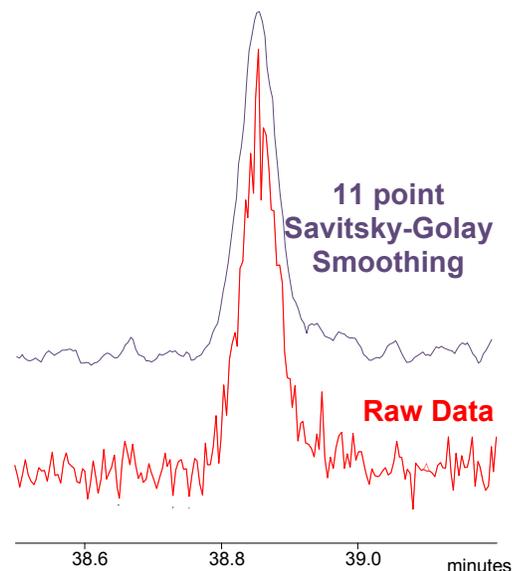
Data Point	Data Point Amplitude	S-G Weight	Value for A	S-G Weight	Value for B	S-G Weight	Value for C
1	63	-0.0857	12.49	-0.0857	35.31	-0.0857	52.09
2	13	0.3429					
3	22	0.4857					
4	24	0.3429					
5	64	-0.0857					
6	55	0.3429					
7	49	-0.0857					

Choices for available number of points in the Savitsky-Golay filter are 5, 7, 9 and 11.



Savitsky-Golay smoothing performs as expected in decreasing noise, as shown in a chromatogram at right. Amazingly and just as with Mean averaging, peak areas remain nearly intact over the full range of smoothing values. However, a distinct advantage with Savitsky-Golay is that the peak shape ends up not as distorted, and peak heights remain closer to the original, even with an 11-point fit. Undoubtedly, much of the height decrease with this smoothing example can be attributed to suppression of noise at the apex.

	Signal/ Noise - RMS	Peak Area	Change from Raw Data	Height	Change from Raw Data
Raw Data	92	14,873	--	4,439	--
5	113	14,697	99%	3,825	86%
7	149	14,686	99%	3,716	84%
9	199	14,422	97%	3,670	83%
11	214	14,424	97%	3,630	82%



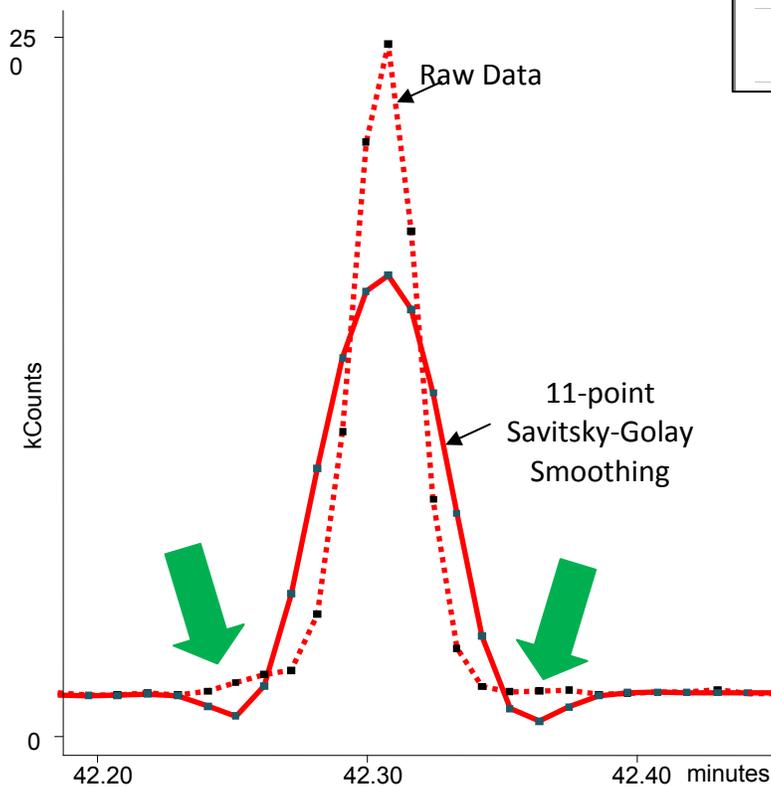
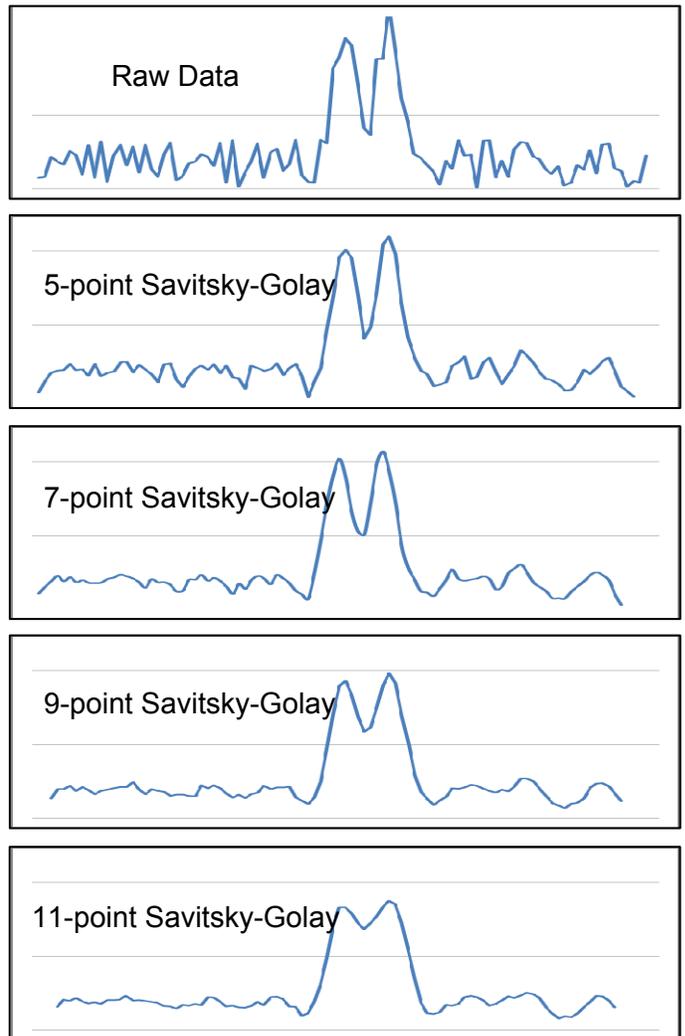
Noise improvement is shown for a chromatogram of 3 pptV Toluene, comparing raw data with an 11-point Savitsky-Golay Smoothing.

<sup>10</sup> Savitsky-Golay coefficients can be found at [www.vias.org/tmdatanaleng/cc\\_savgol\\_coeff.html](http://www.vias.org/tmdatanaleng/cc_savgol_coeff.html).

The same synthetic data used to study effects with Mean averaging are considered with Savitsky-Golay functions. The double peaks are not lost, even with an 11-point treatment. However, noise did not decrease as much as in the other mode.

Mean Points	Valley Penetration	Relative RMS Noise
Raw	63%	1.00
5	62%	0.64
7	52%	0.51
9	39%	0.42
11	20%	0.38

The choice between smoothing functions becomes a user judgment on perceived enhancements of one approach over the other. Approaches can be varied for peaks in the Compound Table, at the user's option. The preset value in the method has Smoothing disabled.



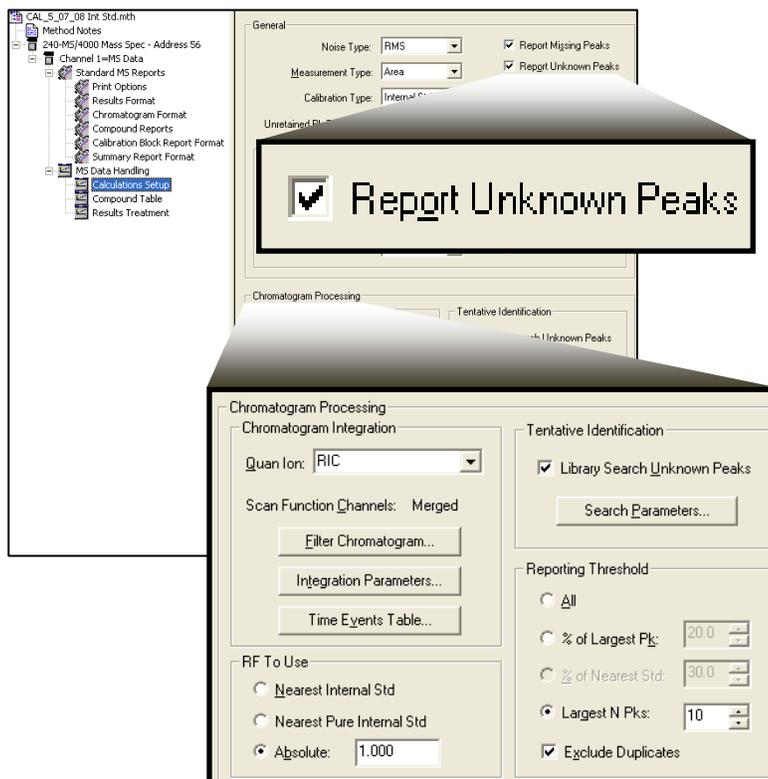
Overuse of Savitsky-Golay smoothing with peaks possessing too few points to define their profile can dramatically distort their shapes and generate inappropriate dips before and after the peak. Peaks areas would be miscalculated since assignment baselines would be in error with these drops. These depressions are generated by negative coefficients on the wings of the smoothing function. This error can be corrected by employing a lower number of points in the filter.

**A peak with only 3 raw data points across its top is severely distorted with a 11-point Savitsky-Golay smoothing function.**

## Treatment of Unknown Peaks

All post-collection operations can be set for each compound individually, or parameters can be set to apply to all “Unknown Peaks”. Filtering, Integration Parameters and Timed Events can be selected to apply globally to all peaks not listed in the Compound Table.

If “Report Unknown Peaks” is unchecked, then this “Chromatogram Processing” portion is grayed out and is not accessible.



## Potential Sources of Data Loss

The amount of data involved in one measurement with a mass spectrometer is huge. The data set is a three-dimensional array of time, mass, and signal amplitude. Data transfer between spectrometer and computer is normally quite significant during the collection process. Every data point possesses a list of masses and intensities, and the transfer can occur at 12 Hz, or even faster with smaller mass ranges. The spectrometer can wait a bit for the computer to a reply to a hardware interrupt for the data transfer, but if the computer fails to react, possibly due to its doing many other tasks, some data points can be lost. When operations involve very fast data rates, then secondary operations with open Windows, especially actions involving accessing the hard drive, should be suspended until after the experiment is completed.

Many anti-virus software programs want to intercept all hard drive accesses, and look at the data stream before data can be written to the hard drive. This process can dramatically slow down data transfer from the spectrometer, and data can be lost. If possible, any anti-virus package on board should be set to ignore the VarianMS directory, or alternatively, to be disabled completely. A major impact of anti-virus programs is with .TMP files created by MS Workstation, and anti-virus programming needs to be set to disregard these as well.

## Guidelines for Setting Optimum Parameters for Peak Detection

1. Collect data for a midrange standard with chromatography fully optimized.
2. Set Expected Peak Widths for each compound to match peak widths at  $\frac{1}{2}$  height from the report for this midrange standard.
3. Set data rate and Scan Mode to enable at least 10 data points across top of the narrowest peak in the chromatogram.
4. If sample peaks could have some low, noisy signals - say near their detection limits, then set the Savitsky-Golay smoothing on with at least a 9-point smoothing function for all peak sizes.
5. Enable Spike Removal with a Spike Factor Threshold of 5.
6. Set Integration Windows wide enough to see "adequate" baseline before and after each peak.
7. Set Identification Reference Peaks for compounds present in every sample, typically Internal Standards and Surrogates.
8. Set Peak Reject Value just below lowest standard area counts.
9. Rerun or recalculate a midlevel standard to reset Expected Retention Time windows.
10. Then **SAVE** method.
11. Run multi-level calibration sequence as required, with the last run being a mid-level standard.
12. Reset Peak Reject Value to just below established detection limit and resave method.
13. Run unknown samples.

**This author thanks Ed Nygren, Dave Michelson and Dave Rhoades of Varian, Inc.,  
for assistance in preparing this monograph.**

Screens are copyrighted by Varian Inc.,  
and are reprinted (reproduced) with the permission of Varian, Inc. All rights reserved.  
Varian and the Varian logo are trademarks or registered trademarks of Varian, Inc.  
in the U.S. and other countries.

---

---

**Lotus Consulting**  
310/569-0128 Fax 714/898-7461  
email ebramstoncook@msn.com



5781 Campo Walk  
Long Beach, California 90803