Using Statistical Peak Identification

Application

Agilent Cerity NDS for Chemical QA/QC

Authors

Lee H. Altmayer, Paul Larson
Agilent Technologies, Inc.
2850 Centerville Road
Wilmington, DE 19808-1610
USA

Abstract

Quantitation demands correct identification for peaks of interest. Traditional peak identification is based upon correlating compounds with their retention times. This retention time correlation is complicated by shifts due to changes in concentration and changes caused by external forces. When peak identification algorithms cannot deal with these complications, they fail to identify peaks correctly. The Statistical Peak Identification feature addresses some of these complications in Agilent's Cerity NDS for Chemical QA/QC. This application shows how Statistical Peak Identification provides a more robust quantitation.

Introduction

There are many ways that retention times (RTs) of peaks can be shifted. Sources can be characterized in terms of the system and of the sample. System issues include effects of ambient temperature and pressure on the ability of the instrument to control column flow rates and zone temperatures. Instruments, such as the Agilent 6890, provide compensation for shifts in ambient temperature and pressure, and also provide precise control of column flow rates and heated zones on the chromatograph. This has reduced variation due to these factors. However, if method conditions are changed, peak identification may no longer be valid. Concentration of components in the sample can also dramatically affect their RTs and, thereby, the peak identification process. Sample concentration also can cause significant shifts in the RT due to peak overload.

In Figure 1 the RT shifts to earlier RTs using gas solid chromatography. Gas liquid chromatography can behave differently. With the overload of an acidic compound, which is not very soluble in the liquid phase, the RT shifts to longer times. The computed symmetries for these peaks are very different. The peaks can move towards other peaks and obscure them.

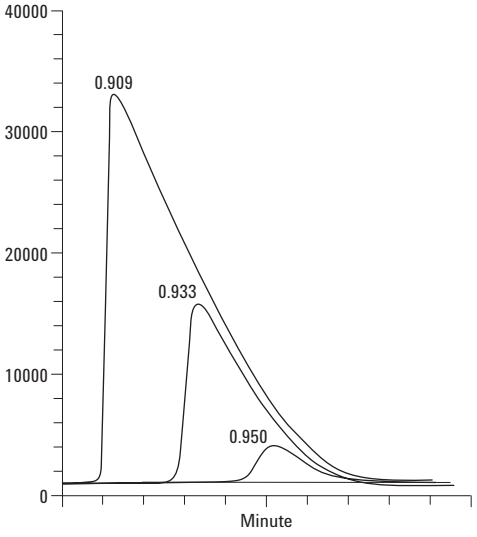


Figure 1. RT as a function of sample concentration on an Alumina column of a refinery gas analyzer (RGA).

The primary means for automated peak identification has been correlation of RTs with the RTs given for compounds listed in the method's calibration table. Focusing only on RT requires that RT windows be defined for identification purposes. When RT windows for adjacent peaks overlap, a decision rule for peak identification is necessary. The rule is normally based on the assumption that the named peak will be the largest peak in the window. For many samples run with gas chromatography (GC), this assumption is sufficient. However, sometimes, even subtle RT shifts can lead to peak misidentification for peaks which are closely eluting due to decision rules used. Figure 2 from the PlotU channel of a Natural Gas Analyzer (NGA) illustrates how a small shift can cause misidentification.

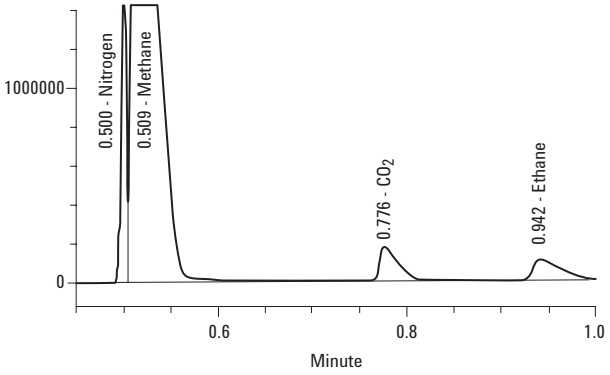


Figure 2. NGA Analysis on the PlotU channel with nitrogen and methane correctly identified.

Note that the RT for nitrogen is 0.500, and for methane is 0.509. A second run shows a 0.003-min shift for the peak which is actually the methane: identification is reversed because of the decision rules used (See Figure 3).

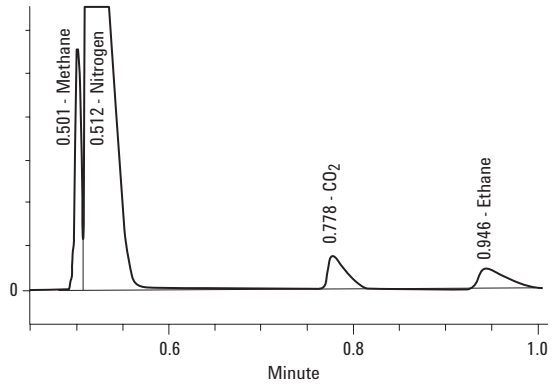


Figure 3. Small RT shift causing misidentification of the nitrogen and methane on the PlotU channel.

Prior to availability of the Statistical Peak Identification (SPID) algorithm, the only way to correct the misidentification in this case was to define one of the other peaks as a reference peak. With release of Cerity 4.03, the user is given a second choice.

Statistical Peak Identification

SPID is a feature added to the Calibration Options user interface (Figure 4) in Agilent's Cerity

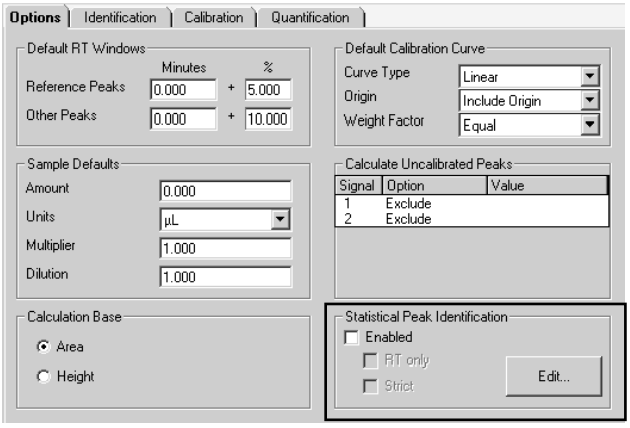


Figure 4. SPID on the calibration options screen.

networked data system (NDS). Using this feature requires the user to enter at least one correctly identified sample in the SPID Data Pool and to check the Enabled checkbox. The Data Pool enables the SPID algorithm to use peak symmetry and area information for each compound in the calibration table. The first set of results is added to the SPID Data Pool prior to enabling SPID in the Options screen for Calibration.

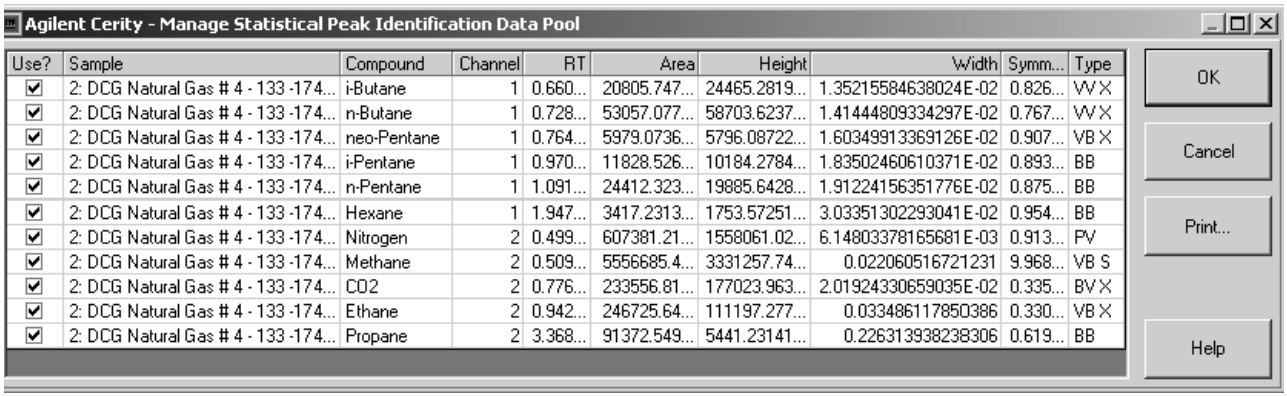
The Data Pool is generated by selecting a sample(s) having correct identification and reprocessing the sample(s) using the Group Reprocessing button at the extreme left of the Reprocess toolbar (Figure 5).

Figure 5. The toolbar for group reprocess on the reprocess view.

Next, results are added to the Data Pool by selecting the Add reprocessed data to SPID data button (Figure 5, adjacent on the right of the Reprocess all selected samples button). The user can then go to the Calibration page under Edit Analysis and click the Edit button. This brings up the following screen, Figure 6, which shows part of the data from the sample in Figure 2. The check box for enabling SPID can now be selected. Peak information is now in the Data Pool and can be used for peak identification (See Figure 7).

current method. The sample can then be reprocessed.

SPID has yet to be used to its fullest advantage. Referencing Figure 1, concentration changes cause large shifts in RT. Developing the SPID Data Pool to include correctly identified components at different concentrations gives a more robust peak identification. Using the Data Pool gives the method developer control of RT changes analogous to Multi-Level Calibration for Amount changes.



Use?	Sample	Compound	Channel	RT	Area	Height	Width	Symm...	Type
☑	2: DCG Natural Gas # 4 - 133 -174...	i-Butane	1	0.660...	20805.747...	24465.2819...	1.35215584638024E-02	0.826...	VV X
☑	2: DCG Natural Gas # 4 - 133 -174...	n-Butane	1	0.728...	53057.077...	58703.6237...	1.41444809334297E-02	0.767...	VV X
☑	2: DCG Natural Gas # 4 - 133 -174...	neo-Pentane	1	0.764...	5979.0736...	5796.08722...	1.60349913369126E-02	0.907...	VB X
☑	2: DCG Natural Gas # 4 - 133 -174...	i-Pentane	1	0.970...	11828.526...	10184.2784...	1.83502460610371E-02	0.893...	BB
☑	2: DCG Natural Gas # 4 - 133 -174...	n-Pentane	1	1.091...	24412.323...	19885.6428...	1.91224156351776E-02	0.875...	BB
☑	2: DCG Natural Gas # 4 - 133 -174...	Hexane	1	1.947...	3417.2313...	1753.57251...	3.03351302293041E-02	0.954...	BB
☑	2: DCG Natural Gas # 4 - 133 -174...	Nitrogen	2	0.499...	607381.21...	1558061.02...	6.14803378165681E-03	0.913...	PV
☑	2: DCG Natural Gas # 4 - 133 -174...	Methane	2	0.509...	5556685.4...	3331257.74...	0.022060516721231	9.968...	VB S
☑	2: DCG Natural Gas # 4 - 133 -174...	CO2	2	0.776...	233556.81...	177023.963...	2.01924330659035E-02	0.335...	BV X
☑	2: DCG Natural Gas # 4 - 133 -174...	Ethane	2	0.942...	246725.64...	111197.277...	0.033486117850386	0.330...	VB X
☑	2: DCG Natural Gas # 4 - 133 -174...	Propane	2	3.368...	91372.549...	5441.23141...	0.226313938238306	0.619...	BB

Figure 6. Manage SPID Data Pool screen.

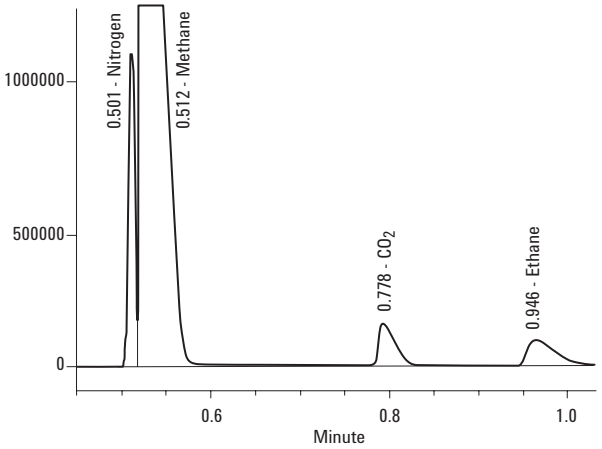


Figure 7. SPID correctly identifying peaks on the previously misidentified chromatogram (Figure 3).

Note that further editing is possible by choosing to use data for an individual sample's results for peak identification. The Data Pool provides means to include additional peak information such as area and symmetry for peak identification. Correcting RT misidentification in this NGA example only required the one sample shown in the Data Pool. From the reprocess screen, the method can be saved either as another version of the original method, or as a new method. However, the user could select a sample with the misidentification without saving, but then must select to retain the

The following example illustrates building a method covering the analytical range of interest. The method is developed using a Refinery Gas Calibration mixture on an Agilent 3000 micro Gas Chromatograph. This multicolumn system rapidly analyzes fixed gases and light hydrocarbons. The Alumina column, used primarily for the unsaturates, will be examined in detail. Table 1 lists the components quantitated on Alumina. Column temperature for the Alumina channel was set at 120 °C. Concentration was varied both by dilution and by injection time (5, 10, and 20 ms for the undiluted sample).

Table 1. Refinery Gas Components on Alumina in Elution Order

Analyte:	**Mole %**
Propane	5.0
Propylene	1.0
i-Butane	10.0
n-Butane	5.0
trans-2-Butene	**5.0**
1-Butene	**10.0**
cis-2-Butene	**5.0**
i-Pentane	2.0
n-Pentane	1.0

Of particular interest are the higher concentration components: 1-Butene and the two 2-Butene isomers. Overloading behavior of these components eventually limits quantitation, thereby establishing the high end of the Analytical range of interest for this analysis. To cover the Analytical range of interest, a series of successive dilutions of the RGA sample was analyzed using extremes as indicated in Figure 8. The range in the area depicted is about 190:1 for the 10% 1-Butene (somewhat different for other components).

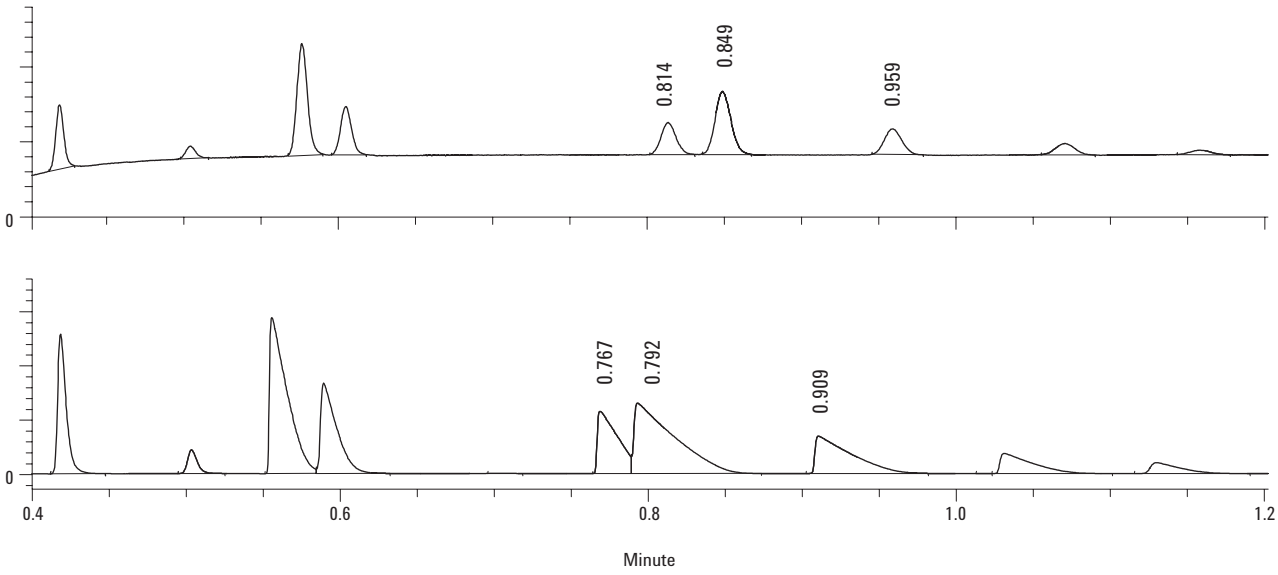
Calibrating from RTs of the dilute analysis, Figure 8 demonstrates results achieved on the overloaded sample. All the ID techniques correctly identify the first and second peaks, Propane and Propylene (not included in the Table). The ID technique in the Table [5% no ref] refers to the Default calibration settings: 5% RT time windows and no time reference peak.

The default technique fails badly. Only one of the seven remaining peaks is correctly identified. Five are missed and n-Butane is misidentified. Increasing the time window to 8% only helps to correctly identify iso-Pentane, but does not handle the 6.7% shift in 1-Butene. Adding Pentane [8% nC_5 ref] as a timed reference peak correctly identifies five of the seven peaks, but still cannot identify two of the Butenes.

Although the chromatogram looks simple, conventional identification methods using the "one peak to a time window" paradigm have significant difficulty. Correct identification of the components over the indicated concentration range is achieved using SPID.

This is a challenging analysis. We will describe how SPID was applied (refer to Table 2). One quick way to construct a SPID Data Pool, which spans a range of conditions, is to start with a center point and to then add the extremes. Once extremes are successfully identified, the bracketed samples should be much easier.

This "Go to Extremes" (GTX) approach has been employed on several other SPID samples with good results. It was not successful here, however, due, as in Table 1, to problems with the Butene components. The GTX approach was attempted for the extremes shown in Figure 8 using a pool from a single run. As the Report summary for the low extreme shows (Figure 9), this approach failed for



	Butanes		Butenes			Pentanes	
ID Technique	i-C_4	n-C_4	trans-2-C_4	1-C_4	cis-2-C_4	i-C_5	n-C_5
5%, No reference	N	I	N	N	N	N	C
8%, No reference	N	I	N	I	N	C	C
8%, n-C_5 reference	C	C	N	I	C	C	C
SPID	C	C	C	C	C	C	C

N = Not identified; I = Incorrectly identified; C = Correctly identified

Figure 8. RGA Chromatograms for the concentration extremes over the analytical range of interest.

Signal 2. The 1-Butene candidate is rejected on the basis of Symmetry and results in one less peak found for a total of eight. At this failure point, an experienced Cerity user could manually intervene via several steps in the Reprocess view (Table 3). We will continue without manual editing of the method.

Report summary:

Signal 1 : Found 5 peaks with fit error 2.854E-04 from 15 of 16 iterations

Signal 2 : Found 4 peaks with fit error 1.3271E-02 from 4 of 5 iterations

1-butene candidate rejected by: Symmetry 0.107/0.859

Signal 3 : Found 8 peaks with fit error 3.3917E-03 from 81 of 91 iterations

Signal 4 : Found 4 peaks with fit error 1.1020E-03 from 5 of 5 iterations

Figure 9. SPID information (Step G, Table 2) in the Report summary following quantitative results for the Low Extreme.

From Figure 10, the symmetry extremes (MIN and MAX values) from the complete 1-Butene results are both outside the single run Data Pool limit in Figure 9.

Only one extreme could be correctly identified from a single midpoint Data Pool. This halted the GTX approach at Step **H** (Table 2) and required a shift to the Maximum Inclusion approach shown to the right in Table 2. As noted in Steps **K-L** (Table 2), when the Data Pool was expanded to include a 9:1 concentration range for 1-Butene, this was adequate to correctly identify all nine components across the 190:1 concentration range. The GTX technique worked for 10 of the 12 samples in the analytical range of interest. The usage of Maximum Inclusion (always the most robust approach) was required for the extremes.

Table 2. Steps to Build a Complete SPID Data Pool

Using extremes and center point	Maximum Inclusion for all correctly identified samples
A. No entries in the Data Pool; SPID Off.	
B. Group Reprocess a mid-range sample.	
C. Verify correct identification.	
D. Add reprocessed data to Data Pool using the Add Reprocessed icon.	
E. A valid Data Pool now exists, so Enable SPID on Calibration|Options screen.	
F. Save the method with a single set of calibration values in the pool. From this point on SPID remains Enabled.	
G. Group Reprocess the extreme samples.	
H. Verify correct identification at extremes, then add the components to the pool.	
If halted by problems identifying extreme components, then use the Maximum Inclusion approach, OR manually edit the Data Pool as specified in Table 3 and then proceed.	
I. Save the method now with a bracketed set of calibration values in the pool.	**I.** Group Reprocess all remaining samples.
J. Group Reprocess, verify and add remaining samples to the pool.	**J.** Verify correctly identified samples and add all of them to the pool.
K. Save the method now with the complete data set of calibration values in the pool.	**Note: All intermediate concentrations, except extremes, were added.**
	K. Save the method with 10 calibration samples over a concentration range ~9:1.
	L. Group Reprocess the extremes.
	M. Verify correct identification at the extremes, then add the components to the pool.
	N. Save the method now with the complete data set of calibration values in the pool.

Table 3. Manual Edit of the Data Pool

1. Select RT only under SPID.
2. Manually edit retentions times.
3. Group Reprocess this manually modified method.
4. Use the Add Reprocessed data to SPID data to add this peak information to the Data Pool.
5. Unselect RT under SPID.
6. Replace the modified RTs with the original values.
7. Save the method. (Return to step I under using extremes and center point in Table 2.)

Results in Figure 10 are generated by selecting the samples and then group reprocessing the entire set following Step N (Table 2). The Cerity Spreadsheet view, requiring Excel to be installed, is selected from the Reprocess tool bar. Monitoring successful identification in these 12 samples is aided by reference to this display: note the extreme values found for 1-Butene. The MIN and MAX RTs are the extremes from Figure 8. Note that statistics for 1-Butene also display the full analytical range of several important quantitative properties: area, width, and symmetry.

Group Reprocess Summary

	C	D	E	F	G	H	I	J
1	Name	Cha	Time	Amount	Area	Height	Width	Symme
107	1-butene	3	0.791506	76.89899	116178.4	64659.11	0.022762	0.092478
112								
113	<NUMBER>		12	12	12	12	12	12
114	<AVERAGE>		0.832253	13.31598	20117.67	16519.32	0.013234	0.386476
115	<STD DEV>		0.014896	20.82839	31467.36	16843.94	0.003336	0.211417
116	<% STDEV>		1.789803	156.4165	156.4165	101.9651	25.20556	54.70366
117	<MIN>		0.791506	0.397567	600.641	843.4225	0.009508	0.092478
118	<MAX>		0.848871	76.89899	116178.4	64659.11	0.022762	0.873529
119								

Figure 10. Group reprocess results for 1-Butene in Cerity using Excel format.

Figure 11 shows the fit error reported for these cases. Fit error and total amount are shown in the log-log plot.

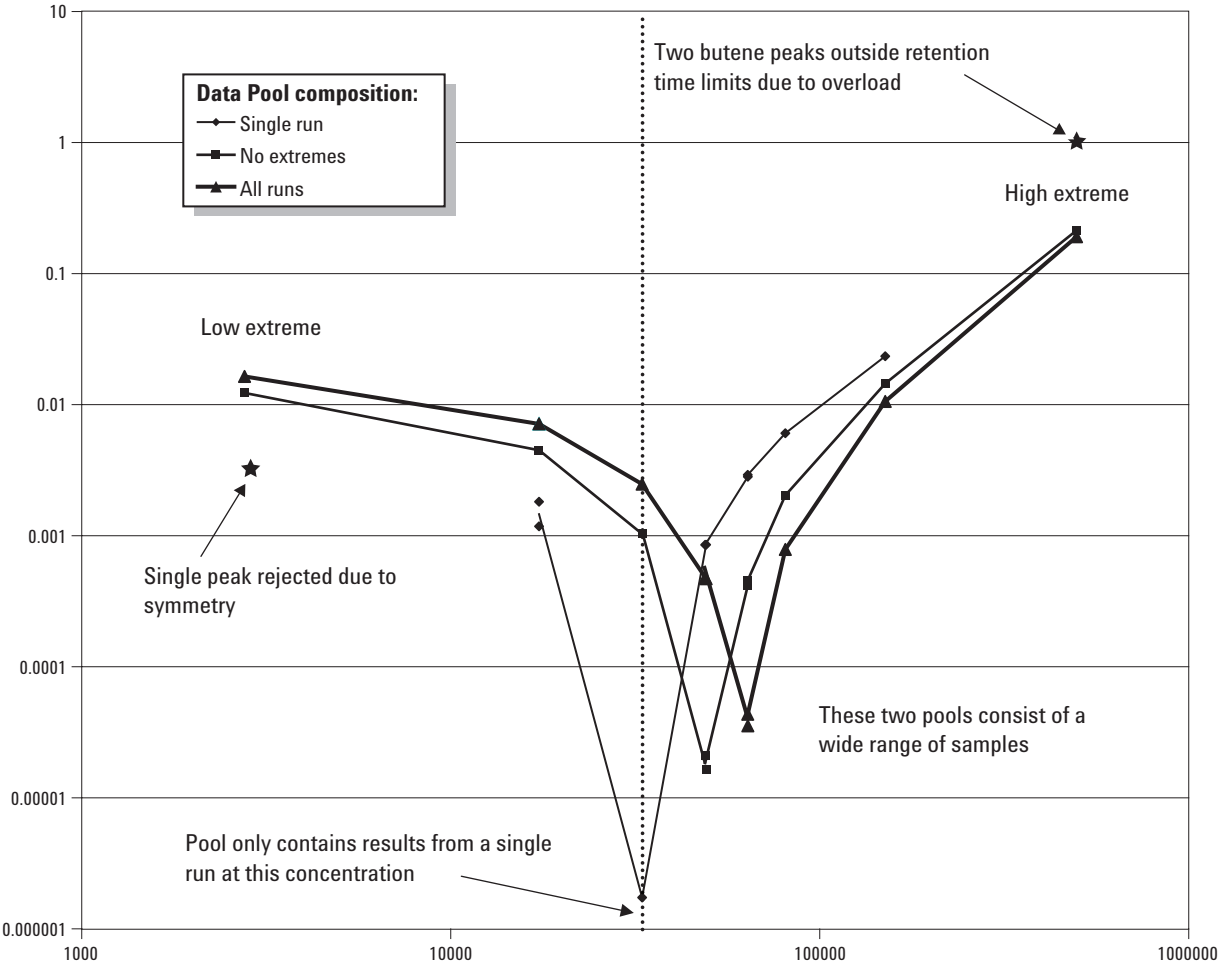


Figure 11. Log-Log plot: fit error vs. concentration of RGA sample for several Data Pool compositions. Increasing fit error correlates with increasing deviation of a given sample's RTs from those at the Data Pool mean.

Fit error is found on the report summary line associated with the Alumina column:

```
Signal 3 : found 9 peaks with fit error 1.7334E-06 from 71 of 71 iterations
```

This is the lowest value of fit error, and therefore the best fit, in the single run Data Pool plot displayed in Figure 11. Note that extreme values are not included because they are not successfully identified.

The fit error at the High Extreme is nearly six orders of magnitude greater than the lowest fit error for this single run Data Pool. This indicates difficulty with fit of the mean pool values for RTs to RTs for the High Extreme case. In this High Extreme case, overloading moves all peaks to lower times by varying degrees. There is no such increase in fit error at the Low Extreme since there is hardly any shift in RT in approaching infinite dilution.

As runs are added producing a “wider” pool with multiple concentrations, extremes are identified successfully.

Limitations

Figure 12 is an example, using temperature programming, demonstrating limitations of various peak identification techniques, and also illustrating the SPID reporting mechanism.

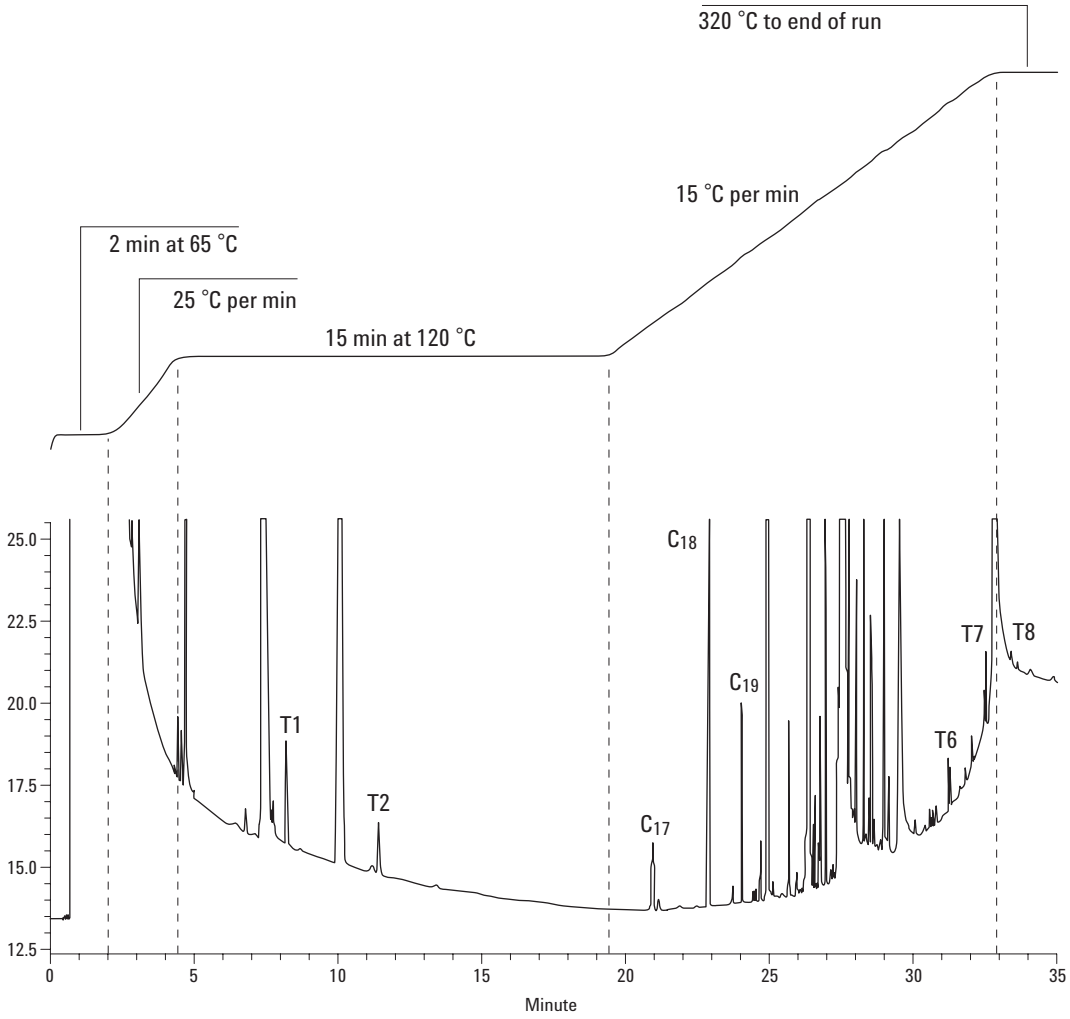


Figure 12. Some peak identities at a 6.0 mL/min column flow rate.

The analysis uses two oven ramps separated by isothermal holds. Flow rate changes are used to simulate effects of column aging, changes in column length, and transfers of analytical methods among instruments and laboratories. To challenge peak identification for this example, a flow rate change of 10% [from 6.1 to 5.5 mL/min] is necessary.

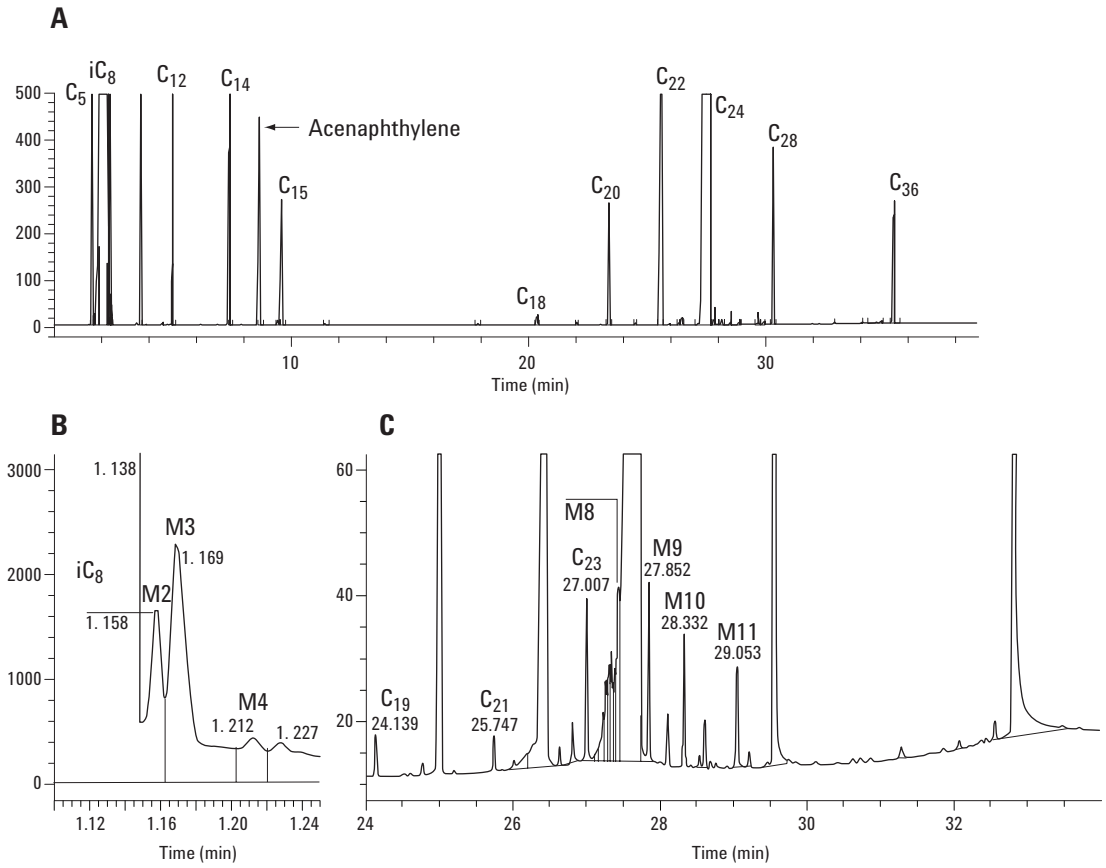


Figure 13. A) Original sample (from [1]); B) Additional peaks near the solvent peak i-C_8; C) Additional trace peaks in the C_{24} region.

The sample, used in an earlier paper [1], ran under different conditions (isothermal region at 140 °C), included 14 hydrocarbon components, including three small peaks, C_{17}, C_{18}, and C_{19}, which approached trace level. To increase SPID computational load, the number of calibrated components was doubled to 28 (close to the maximum of 30 handled per chromatographic signal by SPID) by including most of the trace constituents. These additions are shown in Figures 12 and 13 and include:

Unknown compounds M2, M3, and M4 which elute near the solvent isothermal hold at 65 °C.

Unknown compounds T1 and T2 from the second isothermal hold at 120 °C.

Hydrocarbon impurities C_{21} and C_{23}, and unknowns M8, M9, M10, and M11 from the second oven ramp.

Unknown compounds T6, T7, and T8 from the final hold at 320 °C.

The greatest flow rate induced RT shift occurs in the 15-minute isothermal hold at 120 °C. At just under 5%, the shift is within the default RT window setting. Figure 14 illustrates extremes of flow rate.

Even the most error-prone identification procedure manages to correctly identify C_{15}. If all the chromatographic peaks were this well resolved from possible interferences (resolution of approximately 17 from T1), there would be no need for SPID, or reference peaks, and/or possibly flow control. There is no opportunity to misidentify C_{15} when it is the only peak present in the window.

Less than 3 minutes earlier, at the C_{14}—Acenaphthylene pair, with resolution of 1.3, opportunities to misidentify are plentiful, and a more accurate identification technique can be appreciated.

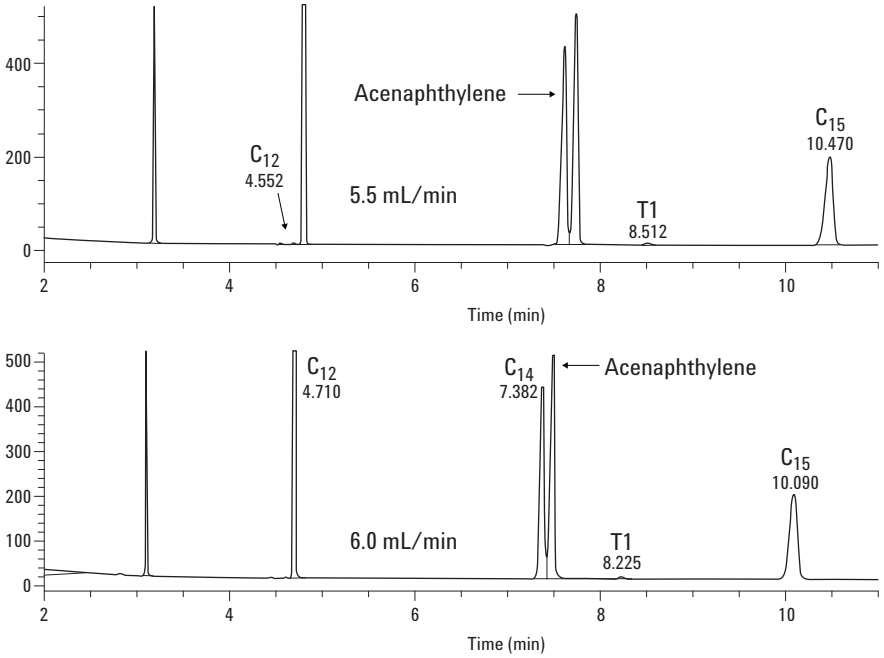


Figure 14. Identifications using 6.0 and 5.5 mL/min flow rate methods.

Table 4 shows procedures of varying robustness in attempting to identify an increasingly distorted chromatogram. Calibration of each identification procedure is done at or near a flow rate value of 6.0 mL/min. Interaction of column flow rate with this multi-ramp temperature programmed analysis produces a non-linear shift in observed RTs. RT shift rises steeply through the isothermal hold at 120 °C, then decreases to a near constant value. Four peak identification procedures are shown:

Least robust technique using fixed time windows and no reference peaks

Adjustment of time windows using a single time reference peak

SPID using a Data Pool composed of 10 runs at a flow rate of 6.0 mL/min (“deep” pool)

SPID using a Data Pool of four runs at flow rates of 5.8, 5.9, 6.0 and 6.1 mL/min (“wide” pool)

Not shown are results from identification using multiple time reference peaks. When multiple reference peaks are carefully chosen to deal with each of the regions in the distorted chromatogram, results improve versus the 5% window using a single reference. Since this is the traditional approach, with only a single iteration performed, there is no testing against statistical limits for RT, area, and symmetry, and no exception reporting for unidentified peaks.

Table 4. Comparative Robustness of Identification Procedures

Procedure	Column flow mL/minute	6.1	6	5.9	5.8	5.5
STD - no ref peaks	Peaks correct	23	28	25	23	14
5% windows	Peaks misidentified	3	0	2	3	7
	Not found	1	0	1	2	7
	Norm total amount	**27.549**	**28**	**4767.117**	**2103.815**	**12065.615**
	Area wrong (m+n)	4630831	0	4656362	4763423	4820000
	Misidentified	IC_8, M2, 3, 4		IC_8, M2, 3	IC_8, M3, 4, Acen, C_{14}	IC_8, M2 ,3 ,4 , $C_{10, 12, 14}$ Acen, T1, T2, C_{23} C_{24}, M8, 9
STD - reference at C_{36}	Peaks correct	23	28	25	23	18
5% windows	Peaks misidentified	3	0	2	3	4
	Not found	1	0	1	2	6
	Norm total amount	**27.549**	**28**	**4767.117**	**2103.972**	**2090.378**
	Area wrong	4630831	0	4656362	4763423	4765346
	Misidentified	IC_8, M2, 3, 4		IC_8, M2, 3	IC_8, M3, 4, C_{14}, Acen	IC_8, M2, 3, 4, $C_{10, 12, 14}$ Acen, T1, T2
SPID - pool: 10 at 6 mL	Peaks correct	26	28	26	23	21
SPID DEEP	Misidentified	0	0	1	4	5
	Not found	1	0	0	1	2
	Norm total amount	**24.806**	**28**	**26.0687**	**24.3037**	**21.9596**
	Fit error	8.9	0.13	3.9	48	284
	Area wrong	0	0	28	82	164
	Misidentified			M11	C_{19}, M10, 11, T6	$C_{18, 19, 21}$, M10, 11, T6
	Key peaks	3	0	3	3	4
SPID - pool : 6.1 - 5.8	Peaks correct	26	27	26	24	21
SPID WIDE	Misidentified	1	1	0	3	6
	Not found	1	0	2	1	1
	Norm total amount	**24.8174**	**28**	**27.51**	**26.041**	**23.2898**
	Fit error	15	1	0.6	27	212
	Area wrong	57	46	13	47	183
	Misidentified	C_{23}	C_{23}	M8	C_{19}, M11, T6	$C_{19, 21, 23}$, M9, 10, 11
	Key peaks	3	3	3	3	4

In this example, “deep” and “wide” pools are contrasted:

The “deep” pool consists of multiple runs under identical analytical conditions- at least eight samples were available for each of 28 calibrated peaks.

The “wide” pool extends across a broad range of flow conditions- computed mean flow rate for the this pool is 5.95 mL/min. Four samples were available for the majority of peaks, but only two samples were available for M3, and three samples were available for M4, C_{23}, M8, M9, T6, and T8.

Response factors provided a normalized contribution of 1.0000 for each of the 28 components at 6.0 mL/min. These contributions sum to 28.0000 when all are identified correctly. Misidentification of the solvent peak has dramatic effects on this summation. SPID's area and symmetry limits prevent this confusion.

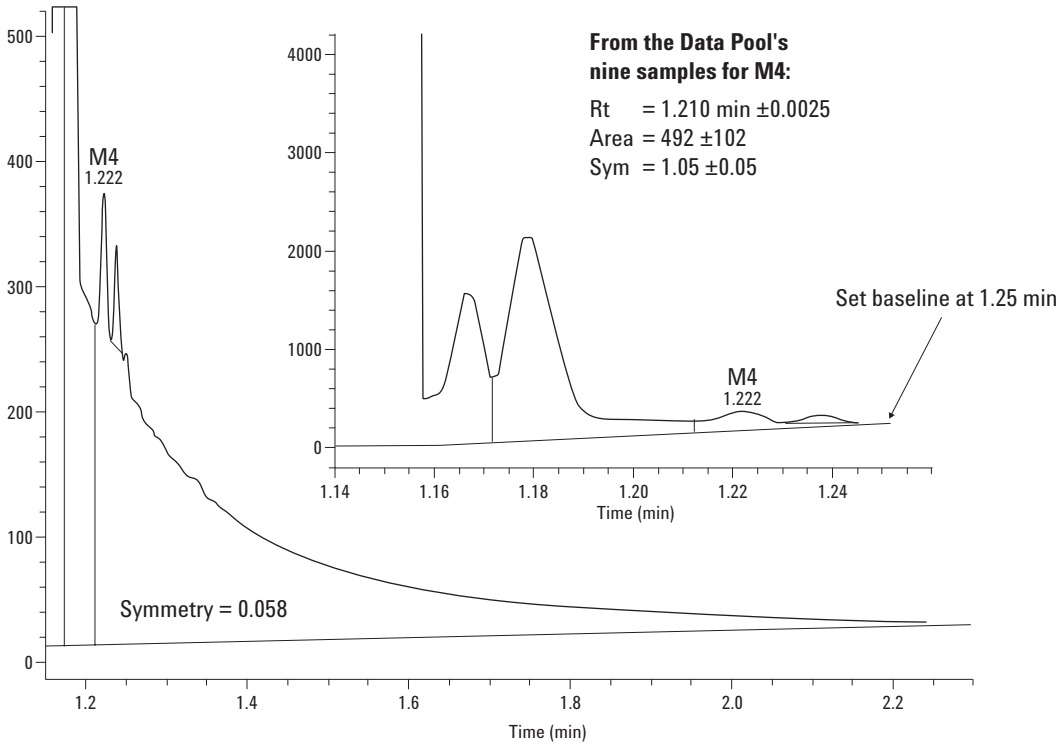


Figure 15. Repairing inconsistent quantitation of M4.

The smallest of the 28 peaks illustrates SPID limitations. The earliest peaks elute immediately following the solvent. Their peak shape is not consistent. No special care was taken to stabilize their detection or baseline treatment, so their respective integration results can vary.

Consider the 5.9 mL/min run (Figure 15) where the peak at 1.222 minutes is a candidate for compound M4. The peak at 1.222 is quantitated as a solvent, grafting onto a huge tail thereby increasing its area by more than an order of magnitude. Peak symmetry, computed as the ratio of frontal area to rear area, drops to 0.058 with addition of the tail. SPID compares this value with M4 symmetry limits, derived from the pool mean and standard deviation, and rejects 1.222 as a match to compound M4. Since there are no other candidates which can be matched to M4, the SPID report summary for this 5.9 mL/min run includes the message:

```
M4 candidate rejected by:Symmetry 0.489/1.695
```

which also lists acceptable limits. The inset (Figure 15) demonstrates how addition of a baseline timed event corrects spurious solvent treatment and permits SPID confirmation of 1.222 as M4 (with symmetry of ~0.84).

Similar, but less dramatic symmetry differences result in other rejections, such as shoulder peak M8. As peaks approach trace level, their quantitation becomes more difficult and results more variable. Examples include trace peaks T6, T7, and T8 which are just visible in the rising baseline around C_{36}. If this variability is not captured in the Data Pool, SPID cannot provide reliable identification.

SPID iterates through multiple choices of peak assignments to minimize fit error and to maximize the number of identified peaks. Fit error measures deviation of identified peak RTs from an “ideal” time, assigned from the mean value of the pool's times. Each SPID iteration begins with a selection of at least three peaks from the pool for use as initial time references. In this case, with half the pool dedicated to the smallest peaks, and with quite a surplus of them, opportunities for misidentification of references exist.

Rather than democratically selecting references at random from the pool, we can instruct SPID in selection of key peaks to serve as initial time references. We should choose at least three easily distinguished peaks covering regions where differences in distortion are likely. The criteria are analogous to that used in selecting conventional time reference peaks. Selecting key peaks was not necessary in the previous examples as their pools did not contain large numbers of trace level peaks.

Another incentive to select key peaks for this sample is to minimize the number of iterations performed. Suppose 10 of the compounds present have two alternatives to evaluate (for example, which peak is Acenaphthylene). This gives 2^{10} combinations (1024) to evaluate per iteration. Fortunately, the number of iterations and combinations is limited since the number of iterations and combinations determines SPID processing time. An estimate of the combination limit is given by:

Combination limit ≅ $2^{(\text{candidates present - 3})}$

For data shown in Table 4, key peaks are chosen in areas of greatest opportunity for mistakes, for example, the large clump of trace peaks between C_{22} and C_{28}. Sometimes, as distortion is increased, it is necessary to anchor identification across the entire chromatogram. For example, in the 5.8 mL/min flow rate case, the “deep” pool uses key peaks at i-C_8, C_{18}, and C_{28}. The “wide" pool substitutes C_{24} as the middle peak. Note that several peaks remain misidentified even using key peaks.

In Figure 16, shape of the fit error plot is similar to that seen previously with RGA results (Figure 11).

There is no value of fit error guaranteeing a good result. Fit error is specific to a particular chromatogram and its associated Data Pool. An error of approximately 0.1 is the best achieved for the 28 peaks in system test sample. For nine peaks on Alumina, errors are consistently below approximately 0.1.

When the analytical range of interest produces only small deviations, it is easiest to acquire replicate runs and to then construct a “deep” pool. More significant changes require capturing the variability and this demands a “wider” pool. This variability was more successfully captured for the RGA case than for the system test sample case.

By the time 5.5 mL/min is attained, all four approaches are seriously deficient. As seen earlier in the RGA case, fit error indicates good chromatogram-to-Data-Pool match. As fit error has increased several orders of magnitude, SPID results have deteriorated markedly, with trace peaks affected the most. Even in this case, however, larger peaks bounded by SPID's area and symmetry limits are correctly identified. This represents a limit to this particular unoptimized analysis.

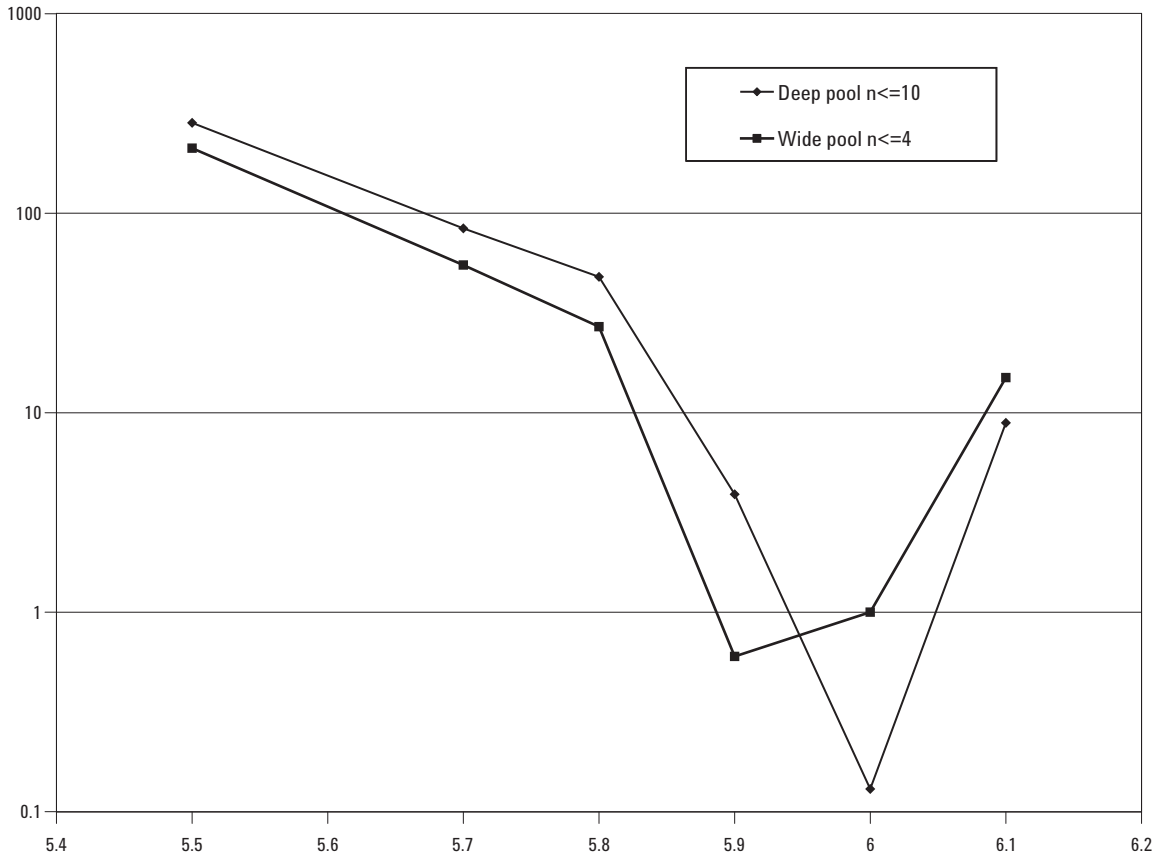


Figure 16. Log fit-error vs. flow rate for the system test sample.

Summary

SPID extends identification over a wider analytical range. Robustness of identification is enhanced for well characterized samples. The Report summary documents why peaks are not included. SPID briefly summarizes its performance on each signal with explicit criteria, peaks identified, fit error, and the number of iterations. Limitations of use were also discussed.

Acknowledgment

We thank Steve Lo for his assistance with the RGA sample discussed here.

References

1. Paul Larson, Robert Henderson, W. Dale Snyder, and Edwin E. Wikfors, “Effect of Environmental Factors on the HP 6890 Series Gas Chromatograph’s Performance Using Manual and Electronic Pneumatics,” Application note 228-321, Agilent Technologies, publication 5963-9807E. www.agilent.com/chem

For More Information

For more information on our products and services, visit our web site at www.agilent.com/chem.

Printed in the USA
January 23, 2003
5988-8533EN

Agilent Technologies