The goal of chromatographic data handling is to identify and quantify analyte compounds. To be successful, chromatographers must choose gas chromatography (GC) instrument and data-handling parameters correctly. When fast chromatography is involved, parametric choices that are suitable for routine chromatography might deliver undesirable outcomes. When performing quantitative analysis, chromatographers should pay close attention to parameters related to peak detection and integration — the expression of detector response as a numeric quantity. Other choices that affect computations after integration, such as the type of standardization and calibration, are not influenced strongly by the speed of the chromatography but nonetheless require equally careful evaluation.

Reliable chromatographic quantification depends first upon the accurate and reproducible detection of peak starts, apexes, valleys, end points and baselines. In the past two “GC Connections” columns, I discussed some common data-acquisition and integration parameters for finding peak retention times and identifying peaks when their retention times shift slightly from run to run, especially in a fast chromatography environment.1,2 Knowledge of peak retention times and identities, however, does not imply that a data system can determine their magnitudes accurately or, indeed, even at all. Some refined and additional information is necessary before this can occur. Of course, it is possible to define peaks’ boundaries manually by cutting pieces of chart paper or by drawing shapes on a computer screen, but these solutions defeat the fundamental attraction of chromatography as an automated technique.

**Baselines**

The primary piece of additional information required is the location of the chromatogram’s baseline profile in relation to the peaks that are eluted on top of it. Peak magnitudes can be expressed in terms of either peak height (microvolts) or area counts (microvolt-seconds). In either instance, unless the chromatogram is exceptional, peak magnitudes are measured not from zero on the chart paper or in the data file but in relation to the detector signal level that would remain in the absence of any injected peaks. Chromatographers aren’t looking for the absolute height or area; they want to know the difference that a peak makes relative to its absence.

In some analytical techniques, the background signal — the absence of any analyte — can be determined at a separate time or as a separate signal. From a somewhat simplified point of view, optical techniques such as infrared (IR) spectroscopy commonly use dual light paths, only one of which passes through the sample, to measure the analyte signal and background signal simultaneously. They then subtract the background from the analyte signal to produce an output spectrum. Other instrumental disciplines such as nuclear magnetic resonance (NMR) spectroscopy measure the analytical signal’s departure from a previously determined zero point. This measurement is possible because of the short-term stability of the output signal in the absence of the analyte and the technique’s independence from the nature of the analyte itself.

In the instance of chromatography — especially GC — the situation is different. The background signal varies significantly from run to run — most notably with programmed temperature elution. It also depends upon the current sample and the history of previously injected samples, both of which make simple prerun detector autozero functions nearly useless. Because of their high sensitivity, many GC detectors provide output signals that will drift enough on the same time scale as the chromatographic signal to make the subtraction of separate baseline and analytical measurements too inaccurate and irreproducible in most instances. Many gas chromatographs include a baseline subtraction software routine that records a non-analyte baseline and then corrects subsequent chromatograms by subtracting the recorded background. However, this technique fails to accommodate changes in
the background signal from run to run, most significantly with capillary columns, unless a new background profile is recorded before each run. Most chromatographers will not accept the halving of sample throughput that results from this practice.

**Chromatographers aren’t looking for the absolute height or area; they want to know the difference that a peak makes relative to its absence.**

Simultaneous background and signal determination are possible in GC, and they have been performed widely in the past, primarily with packed-column GC. Chromatographers use dual columns of the same nominal size and the same stationary phase and flow-rate, but only one column receives a sample injection. The separate detector signals are subtracted from each other in real time in the GC instrument by analogue or digital means and the composite signal goes to the data-handling system. This approach is more-or-less successful at the low sensitivities commonly used with packed columns, but it is unsuitable for capillary columns for two reasons. First, it is difficult to find two capillary columns that have identical baseline profiles at high detector sensitivities, and, second, even precisely matched columns won’t remain that way after several real-world samples have been injected onto one of them. And, from an economic point of view, no one wants to buy a pair of each of the distinct columns he or she uses.

Today’s data-handling systems solve these challenges — and create some new ones of their own — by applying what essentially is an implied baseline to analyte peaks. The baseline signal that the data system subtracts from the analyte signal is not measured separately or simultaneously through a non-analyte signal channel. Instead, the same signal that contains the analyte peaks also provides information about the background signal. When a data system subtracts the background under a peak in a chromatogram, it interpolates sections of baseline for each peak according to rules that were determined by software programmers and then selected — knowingly or unknowingly — by a chromatographer who specified the data system’s quantification method parameters. Although the chromatographic applications of these computational techniques are not unique — many other software systems apply similar methods — their use in chromatography data handling represents one of their earliest implementations.

**Integration of Isolated Peaks**

How, then, does a chromatography data-handling system determine the location and shape of the baseline segments? The simplest example is a fully resolved and symmetrical peak on a flat baseline. Figure 1(a) shows a situation in which the baseline segment, in green, extends from peak start to peak end. This figure is an example of fast chromatography in which the peaks are separated by 1.25 s with a resolution (R_s) of 1.04, and the sampling rate is 30 Hz. The peak area is bounded by the baseline segment on the bottom and by the chromatographic signal at the sides and top. In this instance, the operator used the data system to determine appropriate integration parameter settings from an empty section of the flat baseline, as I have recommended previously. Using the default method values or values from another method developed for slower chromatography usually causes the apparent peak start and end points to move to different locations and sometimes fails to detect the peak at all.

For integration, a data system constructs a linear equation between the peak start and end points that describes the amplitude of the baseline segment as a function of the time since the peak start. It then subtracts this value from the signal amplitude at the corresponding retention times. The subtracted values are summed and normalized to units of microvolt-seconds or millivolt-seconds to yield the peak’s total area measurement. Normalization adjusts signals with various data-acquisition rates and analogue-to-digital (A/D) conversion factors to the same scale.

The situation becomes more complex when a peak sits on top of a changing baseline, such as Figure 1(b) shows. This figure shows two baseline segments. The red one extends far down along the baseline profile to the left; it was determined using the area and noise thresholds that the system elected for Figure 1(a) in the absence of baseline drift. The green baseline segment in Figure 1(c) — the one that more closely represents what a chromatographer would choose based on visual appearance — resulted from reassessing the area and noise threshold settings in light of the rising baseline. However, the data system’s straight-line interpolation between peak start and end could still be more of an approximation than the actual baseline profile, and this guess could introduce an error in the apparent peak area.

The example peaks in Figure 1 are identical — the only difference between the chromatograms is the presence or absence of the curved baseline under the peaks. With the threshold settings chosen for Figure 1(c), the data system reported the area under the peak in Figure 1(a) as 626 300 µV-s, and it found an area of 626 500 µV-s in Figure 1(c), which is a very good match. Other situations might not do as well and it is advisable to evaluate integration results throughout a range of sample concentrations. Of course, selecting the wrong peak detection parameters, as shown by the erroneous extended baseline segment in Figure 1(b), will cause serious quantitative errors.

**Overlapping Peaks**

Things become dicier when two peaks overlap. In Figure 2(a), an unresolved pair
of peaks sits on a flat baseline and the valley between the peaks does not return anywhere close to the chromatogram’s baseline. These peaks each have the same nominal area as the single peaks from Figure 1. The data system has detected the valley minimum as a baseline point and has drawn two baseline segments — to and from the valley and the adjacent peak start and end points, as shown by the red line on the plot.

For the purpose of illustration, I forced this obviously incorrect baseline segment allocation by disabling the data-handling system’s peak valley detection, which otherwise is enabled by default. Figure 2(b) shows, in green, the baseline segments the data system created with valley detection disabled. The areas of the two peaks as reported for Figure 2(b) are 625 800 and 625 700 µV-s, which closely match the area of the single peak in Figure 1. Note however, that these results were obtained with the same area and noise threshold settings as used in Figure 1(c). When I set the threshold levels to their defaults, the system actually detected three peaks instead of two because of some noise in the chromatogram, and it did not find the correct start and end points for the first and second peaks, respectively. I obtained the area counts in Figure 2(b) by using the second peak, which has the same area as the peaks in Figures 1 and 2. In Figure 3(a), the data system drew a simple vertical baseline segment at the valley point between the two peaks. From Figure 3(b), the reported area counts are 634 200 and 28 600 µV-s, which raise the area per cent of 3.52% — significantly different from the known level of 5.0%. In Figure 3(b), exponential skimming was disabled, and the data system drew a simple vertical baseline segment as shown by the red line on the plot.

What happens when overlapping peaks straddle a rising baseline? The results in Figure 2(c), which were obtained using the same integration settings as those used in Figure 2(b), show that the data system reported areas of 627 000 and 625 900 µV-s for the two peaks, which match the reported area counts without a rising baseline to within 0.3%. Thus, in these simple situations, a data system with appropriate integration parameters delivers integrated peak areas that correspond quite closely with the expected levels.

Another situation arises when a small peak sits on the trailing end of a larger peak, as Figure 3 illustrates. Most data-handling systems support the allocation of an exponentially shaped baseline segment under the second, smaller peak that approximates the trailing shape of the first peak in a so-called skimming operation. The second peak in Figure 3 has a known area that is 5% of the area of the larger peak, which has the same area as the peaks in Figures 1 and 2. In Figure 3(a), the data system drew an exponential baseline segment to skim off the smaller peak from the larger. The reported area counts, however, are 639 500 and 23 300 µV-s, which give the second peak an area per cent of 3.52% — significantly different from the known level of 5.0%. In Figure 3(b), exponential skimming was disabled, and the data system drew a simple vertical baseline segment at the valley point between the two peaks. From Figure 3(b), the reported area counts are 634 200 and 28 600 µV-s, which raise the area per cent of 4.3% — much closer to the known level. This example reemphasizes the importance of evaluating the effect of data-handling parameters on quantitative results. Even when a data system’s baseline allocations look good, the results can be less satisfactory than those that could be obtained with a different set of parameters. This outcome is especially true with fast chromatography in which the differences in peak widths and sampling rates are large compared with the default chromatograms for which the data system was designed. Fortunately, all the modern data-handling systems I’m familiar with can operate at high speeds, although some go faster than others. Be sure to ask about high-speed operation when buying or upgrading data-handling software and hardware.

A Word about Method Calibration
The above examples show what happens when the relative peak area of a second, unresolved peak drops from the same size to 5% of the first peak. The changes in the integrated peak areas as reported by the data system are non-linear with respect to the known peak sizes in the chromatogram. Looking at the area counts in the examples, when I reduced the size of the second peak by a factor of 20, its reported area counts dropped by a factor of close to 22 — the change is approximately 10% lower than what I expected. To compensate for those effects, which can originate in the chromatography system as well as in the data-handling system, chromatographers should use multilevel non-linear calibration of their data-handling methods. This technique provides a variable-response factor that depends upon the measured area counts and that compensates for non-linear effects, if properly calibrated and executed. I’ll touch on this subject in more detail in an upcoming instalment of “GC Connections.”

Conclusion
Successful peak detection and identification do not necessarily guarantee that subsequent peak integration will yield areas that accurately reflect the analyte amounts present. Applying a data system’s default peak-integration parameters will nearly always guarantee erroneous area count reports, especially with fast chromatography.
chromatography systems. Instead, chromatographers should assess each target analyte peak in relation to its neighbours and to the chromatogram’s baseline profile. Many data systems include software tools that suggest appropriate integration parameter values. When examined throughout a range of amounts or concentrations, integrated peak areas are not necessarily linear with respect to known calibration values. Multilevel calibration routines can provide convenient means to compensate for non-linear effects, both in data-handling and chromatographic instrument systems.

References

“GC Connections” editor John V. Hinshaw is president and principal scientist of ChromSource, Brentwood, Tennessee, USA, and a member of the Editorial Advisory Board of LC•GC Europe.

Direct correspondence about this column to “GC Connections,” LC•GC Europe, Advanstar House, Park West, Sealand Road, Chester CH1 4RN, UK, e-mail: dhills@advanstar.com

For an ongoing discussion of GC issues with John Hinshaw and other chromatographers, visit the Chromatography Forum discussion group at http://www.chromforum.com