Why use Signal-To-Noise as a Measure of MS Performance When it is Often Meaningless?

Technical Overview

Authors

Greg Wells, Harry Prest, and
Charles William Russ IV,
Agilent Technologies, Inc.
2850 Centerville Road
Wilmington, DE 19809-1610
USA

Abstract

The signal-to-noise of a chromatographic peak determined from a single measurement has served as a convenient figure of merit used to compare the performance of two different MS systems. The evolution in the design of mass spectrometry instrumentation has resulted in very low noise systems that have made the comparison of performance based upon signal-to-noise increasingly difficult, and in some modes of operation impossible. This is especially true when using ultra-low noise modes such as high resolution mass spectrometry or tandem MS; where there are often no ions in the background and the noise is essentially zero. This occurs when analyzing clean standards used to establish the instrument specifications. Statistical methodology commonly used to establish method detection limits for trace analysis in complex matrices is a means of characterizing instrument performance that is rigorously valid for both high and low background noise conditions. Instrument manufacturers should start to provide customers an alternative performance metric in the form of instrument detection limits based on relative the standard deviation of replicate injections to allow analysts a practical means of evaluating an MS system.

Agilent Technologies

Introduction

Signal-to-noise ratio (SNR) has been a primary standard for comparing the performance of chromatography systems including GC/MS and LC/MS. Specific methods of calculating the SNR have been codified in the U. S., European and Japanese Pharmacopeia [1-3] to produce a uniform means of estimating instrument detection limits (IDL) and method detection limits (MDL). The use of SNR as a measure of IDL and MDL has been useful for optical based detectors for LC, and flame based detectors for GC. As mass spectrometry instrument design has evolved, the ability to accurately compare performance based upon SNR has become increasingly difficult. This is especially true for trace analysis by mass spectrometry using ultra-low noise modes such as high resolution mass spectrometry (HRMS) or tandem MS (MS/MS). SNR is still a useful parameter, particularly for full scan (EI) MS; but the comparison of high performance MS analyzers should be based upon a metric which is applicable to all types of MS instruments and all operating modes. Statistical means have long been used to establish IDL and MDL [4-7] and are suitable to all operating modes for mass spectrometers.

Evolution of Instrumentation

Many sources of noise have been reduced by changes to the MS design such as low noise electronics, faster electronics allowing longer ion sampling (signal averaging), modified ion paths to reduce metastable helium (neutral noise) and signal processing (digital filtering). HRMS and MS/MS are also effective means to reduce chemical background noise, particularly in a complex sample matrix. For the signal, a longer list of improvements to the source, analyzer and detector components have resulted in more ions produced per amount of sample. In combination, the improvements in SNR and the increase in sensitivity have resulted in significant and real lowering of the IDL and MDL.

Lack of Guidelines for SNR Measurements

The measurement of the signal is generally accepted to be the height of the maximum of the chromatographic signal above the baseline (Figure 1). However, some of the changes in GC/MS SNR specifications have been artificial. The industry standard for GC/(EI)MS has changed from methyl stearate, which fragments extensively to many lower intensity ions, to compounds that generate fewer, more intense ions such as hexachlorobenzene (HCB) and octafluoronaphthalene (OFN). The change to OFN held a secondary benefit for noise: the m/z 272 molecular ion is less susceptible to baseline noise from column bleed ions (an isotope of the monoisotopic peak

of polysiloxane at m/z 281 increased baseline noise for HCB at m/z 282).

Improvement in instrument design and changes to test compounds were accompanied by a number of different approaches to measuring the noise. In the days of strip chart recorders and rulers, a common standard for noise was to measure the peak-to-peak (minimum to maximum) value of baseline noise, away from the peak tails, for 60 seconds before the peak (Figure 1) or 30 seconds before and after the peak. As integrator and data systems replaced rulers, the baseline segments for estimation of noise were auto-selected and noise was calculated as the standard deviation (STD) or root-mean-square (RMS) of the baseline over the selected time window.
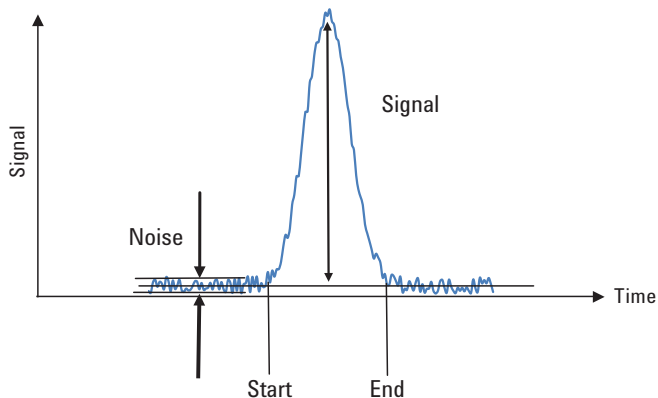

Figure 1.   Analyte signal as a function of time for a chromatographic peak, demonstrating the time dependence in the amount of analyte present.

Automation of the calculation holds many conveniences for a busy lab, but the noise measurement criteria have become less controlled. In some cases, vendors have estimated noise based upon a very narrow window (as short as 5 seconds); and the location can be many peak widths away from the peak used to calculate signal. These variable hand-picked noise windows enable a vendor to claim a higher SNR by judiciously selecting the region of baseline where the noise is lowest. Generally, the location in the baseline where the noise is calculated is now automatically selected to be where the noise is a minimum. Figure 2 shows three different values of RMS noise calculated at different regions of the baseline. The value of the noise at positions a, b, and c is 54, 6, and 120; and the factor of 20 variation in SNR is due exclusively to where in the baseline the noise is measured. However, these SNRs will not correlate to the practical IDL for automated analyses that the lab performs. Therefore, the use of signal-to-noise as an estimate of the detection limit will clearly fail to produce usable values when there is low and highly variable ion noise and the choice of where to measure it is subjective.
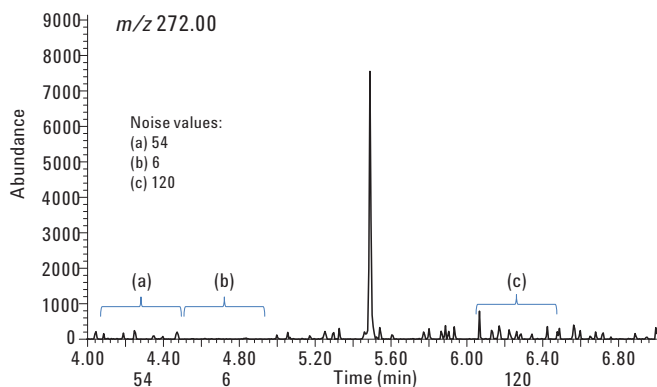
Figure 2. EI full scan extracted ion chromatograms of m/z = 272 from 1 pg OFN.

The situation becomes even more indeterminate when the background noise is zero as shown in the MS/MS chromatogram in Figure 3. In this case, the noise is zero and the signal-to-noise then becomes infinite. The only noise observed in Figure 3 is due to the electronic noise, which is several orders of magnitude lower than noise due to the presence of ions in the background. This situation can be made more severe by increasing the threshold for ion detection. Under these circumstances, it is possible to increase the ion detector gain, and hence the signal level, without increasing the background noise. The signal of the analyte increases, but the noise does not. This is obviously misleading since the signal increased, but there was no increase in the number of ions detected, and therefore no change to the real detection limit. This allows the SNR to be adjusted to any arbitrary value without changing the real detection limit.
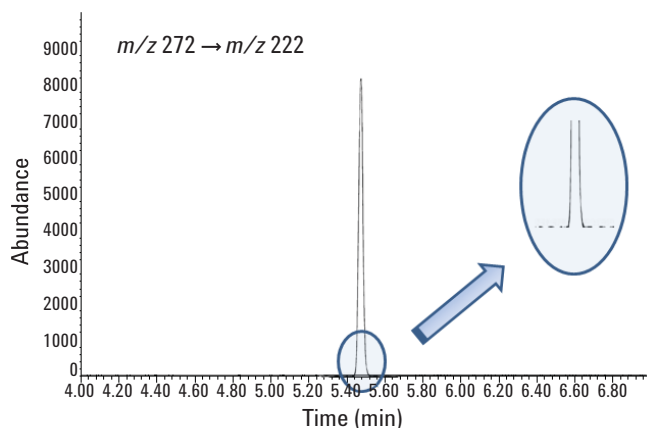

Figure 3. EI MS/MS extracted ion chromatogram of m/z 222.00 from 1 pg OFN exhibiting no chemical ion noise.

Two decades ago, instrument specifications often described the detailed analytical conditions that affected the signal or noise value such as chromatographic peak width, data rate and time constant. Today, those parameters are missing or hard to find; even though a novice chromatographer realizes a narrow peak will have a larger peak height than a broad peak. In many cases, the GC conditions are selected to make the chromatographic peak extremely narrow and high, with the result that the peak is under-sampled (only one or two data points across a chromatographic peak). This may increase the calculated SNR, but the under-sampling degrades the precision and would not be acceptable for most quantitative methods. Once again, a false perception of performance has been provided. Accurate, meaningful comparison of instrument performance for practical analytical use based upon published specifications is increasingly difficult – if not impossible. Alternative means are required to determine the instrument performance and detection limit, which is generally applicable to all modes of MS operation.

What is the alternative for SNR?

SNR is still useful and represents a good first estimate to guide other statistically based estimates of performance. Every analytical lab should understand and routinely use statistics to evaluate and validate their results. The analytical literature has numerous articles that utilize a more statistical approach for estimating an IDL or MDL. The US EPA has dictated a statistical approach to MDL; and can be found in the recommended EPA Guidelines Establishing Test Procedures for the Analysis of Pollutants.[6] The European Union has also supported this position. A commonly used standard in Europe is found in The Official Journal of the European Communities; Commission Decision of 12 August 2002; Implementing Council Directive 96/23/EC concerning the performance of analytical methods and the interpretation of results. [8]

Both of these methods are generally similar and require injecting multiple replicate standards to assess the uncertainty in the measuring system. A small number of identical samples having a concentration near the expected limit of detection (within 5-10) are measured, along with a comparable number of blanks. Because of the specificity of mass spectrometry detection, the contribution from the blank is negligible and is often excluded once the significance of the contribution has been confirmed. The mean value $\overline{X}$ and standard deviation (STD) of the set of measured analyte signals, X_n (that is, integrated areas of the baseline subtracted chromatographic peaks), is then determined. Even for a sensitivity test sample, the sampling process is actually a complex series of steps involving drawing an aliquot of analyte solution into a syringe, injecting

into a GC and detecting by MS. Each step can introduce variations in the final measured value of the chromatographic peak area resulting in a sample-to-sample variation or sampling noise. These variances generally limit the practical IDL and MDL that can be achieved.

The variance in the measured peak areas includes the analyte signal variations, the background noise, and the variance from injection to injection. The statistical significance of the mean value of the set of measured analyte signals being distinguishable from the combined system and sampling noise can then be established with a known confidence level. The IDL is the smallest signal or amount of analyte that is statistically greater than zero within a specified probability of being correct. The IDL (or MDL) is related [9, 10] to the standard deviation STD of the measured area responses of the replicate injections and a statistical confidence factor t_a by:

IDL = (t_a)(STD) where the STD and IDL are expressed in area counts.

Alternately, many data systems report relative standard deviation (RSD=STD/Mean Value). In this case, the IDL can be determined in units of the amount of the standard (ng, pg, or fg) injected by:

IDL = (t_a)(RSD)(amount standard)/100%

When the number of measurements n is small (n < 30), the one-sided Student t-distribution [11] is used to determine the confidence factor t_a. The value of t_a comes from a table of the Student t- distribution using n-1 (number of measurements minus one) as the degrees of freedom; and 1_a is the probability that the measurement is greater than zero. The larger the number of measurements n, the smaller is the value of t_a and less the uncertainty in the estimate of the IDL or MDL. Unlike the statistical method of determining IDL and MDL, the use of signal-to-noise from a single sample measurement does not capture the sampling noise that causes multiple measurements of the same analyte to be somewhat different.

As an example, for the eight replicate injections in Figure 4 (n = 7 degrees of freedom) and a 99% (1_a = 0.99) confidence level, the value of the test statistic from the t-table is 1_a = 2.998. For the eight 200 fg samples, the mean value of the area is 810 counts, the standard deviation is 41.31 counts, and the value of the IDL is: IDL = (2.998)(41.31) = 123.85 counts. Since the calibration standard was 200 fg and had a measured mean value of 810 counts, the IDL in fg is: (123.85 counts) (200 fg)/(810 counts) = 30.6 fg. Alternately, in terms of RSD, the IDL = (2.998)(5.1%)(200 fg)/100 = 30.6 fg. Thus, an amount of analyte greater or equal to 30.6 fg is detectable and distinguishable from the background with a 99% probability.
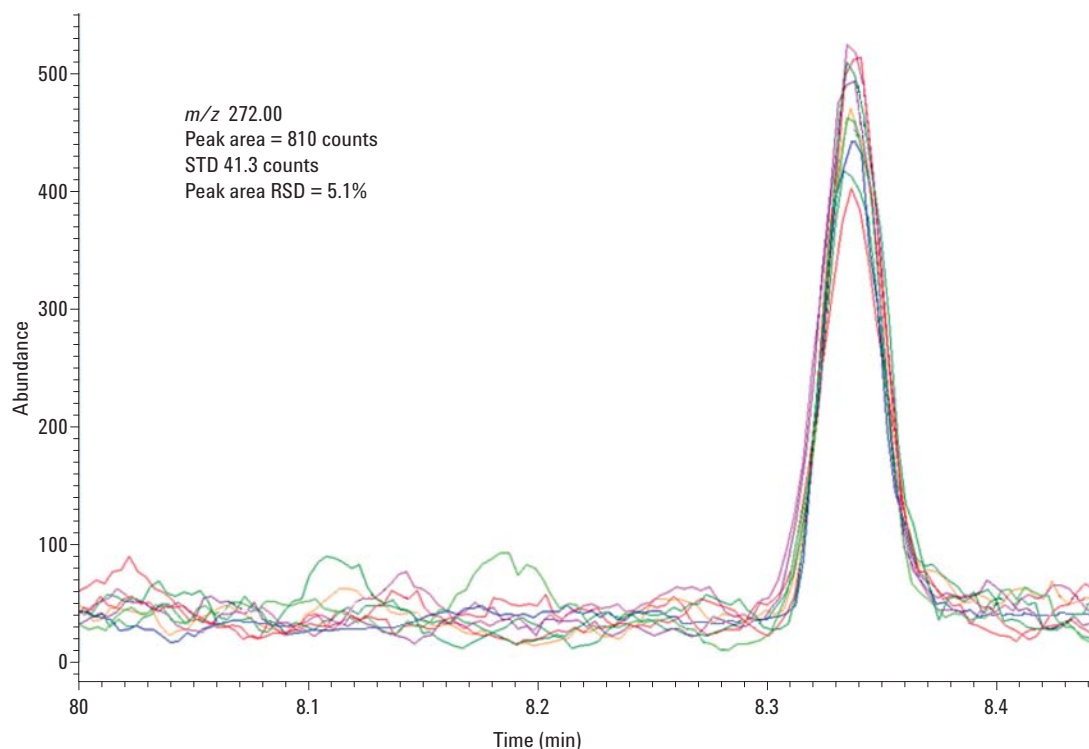

m/z 272.00
Peak area = 810 counts
STD 41.3 counts
Peak area RSD = 5.1%

Figure 4.   *EI full scan extracted ion chromatograms of m/z = 272 from 200 fg OFN; eight replicate injections; 3.3 Hz data rate.*
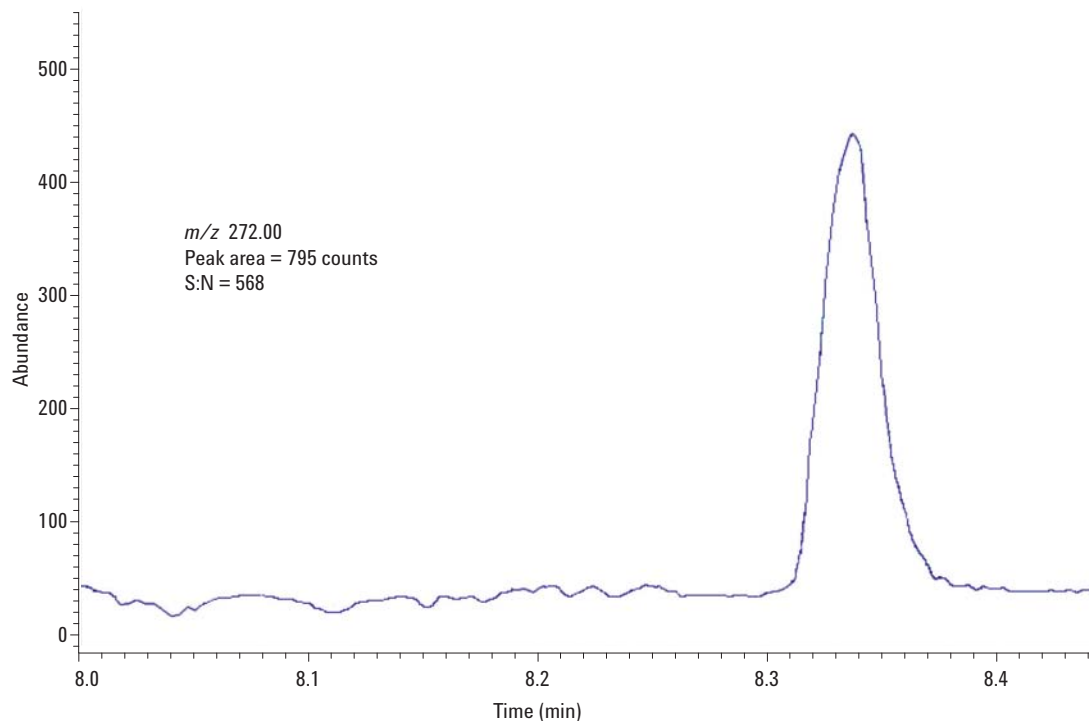
4

Figure 5.   EI full scan extracted ion chromatograms of m/z = 272 from 200 fg OFN; first entry in Table 1; 3.3 Hz data rate.

In contrast, the SNR measured on one single chromatogram (Figure 5) would give an IDL of 1.1 fg, assuming that IDL = 3 X rms noise (first entry in Table 1). The high SNR is the result of a commonly used algorithm that calculates the noise using the lowest noise sections of the baseline adjacent to the peak where the noise is unusually low. The individual SNRs and IDLs measured for each injection in Figure 4 are tabulated in Table 1. The IDL values range from 1.1 fg to 20.5 fg; and even for the largest value is significantly smaller than the IDL determined by statistical means; which is more realistic.

Table 1.   Comparison of S/N for Eight Injections

Injection number	Signal	RMS noise	S/N	IDL=3*RMS (fg)
1	795	1.4	568	1.1
2	821	2.5	328	1.8
3	835	10.9	77	7.8
4	854	2.8	305	2.0
5	818	2.9	282	2.1
6	776	26.6	29	20.6
7	853	2.6	328	1.8
8	735	2.7	272	2.2

Estimating Relative Sensitivity

A secondary benefit to specifying detection limits in statistical terms is to use the RSD as an indirect measure of the relative number of ions ion a chromatographic peak. It is known [12] that if a constant flux of ions impinge on a detector, and the mean number of ions detected in a particular time interval averaged over many replicate measurements is N, then the RSD in the number ions detected in is $1/\sqrt{N}$. Hence decreasing the number of ions in a chromatographic peak will increase the area RSD. This effect can be seen by comparing the RSD in Figure 6 and Figure 4 where the analyte amount is 1pg and 200 fg respectively. The increased RSD at the lower sample amount is due in large part to ion statistics. Five times lower number of ions means $\sqrt{5}$ = 2.24 times increase in RSD. It follows that the 2.1% RSD at 1 pg will become 4.7% at 200 fg just from ion statistics. It is not unexpected that at lower sample amounts there will also be some additional variance from the chromatography. Thus RSD measurements of the same amount of analyte on two different instruments can be used to indicate the relative difference in sensitivity near the detection limit; assuming that other contributions to the total variance of the signals are small. The more sensitive instrument will have the smaller RSD if all other factors are the same – that is, peak width, and data rate. This avoids the uncertainty of inferring sensitivity from the measurement of peak areas where there is no baseline noise.
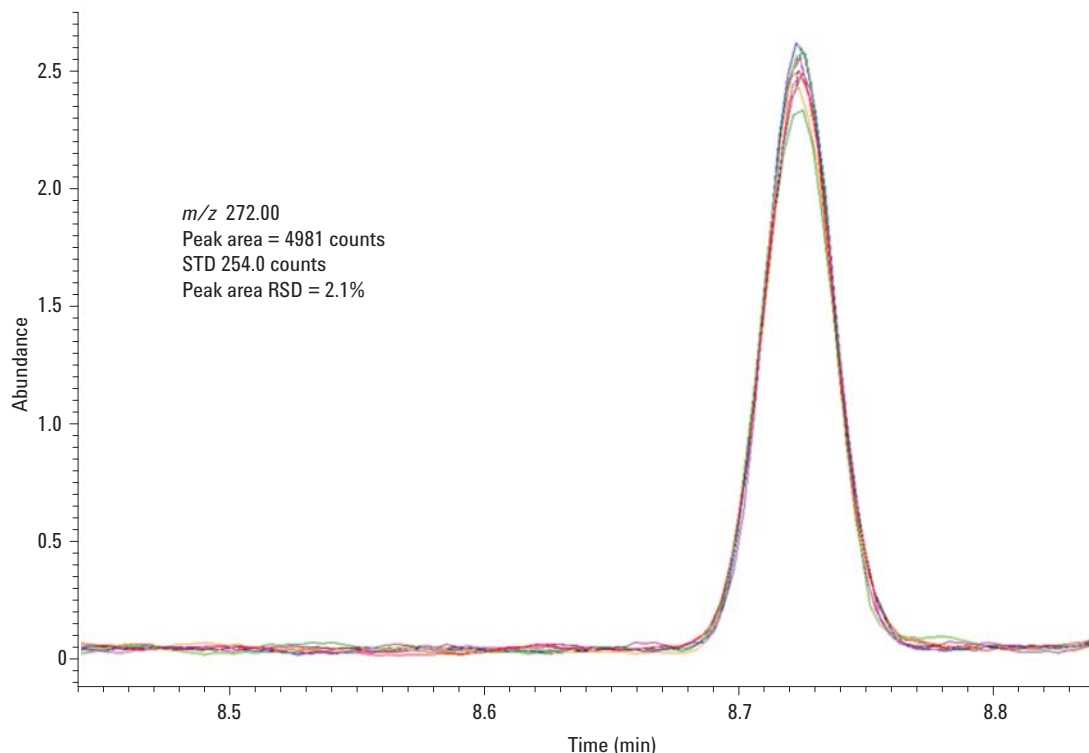
m/z 272.00
Peak area = 4981 counts
STD 254.0 counts
Peak area RSD = 2.1%

Figure 6.   EI full scan extracted ion chromatograms of m/z = 272 from 1 pg OFN; eight replicate injections; 3.3 Hz data rate.

Impact of changing from S/N to RSD

For MS vendors, there is complexity and cost in providing customers this valuable information. If the GC is configured with an autosampler and a splitless inlet, it should be fairly simple to add seven to 10 replicate injections. This might add a little time to the installation, but the use of an autosampler should keep this at a minimum. But how do you test a system that does not have an autosampler or the appropriate inlet? Manual injections will add operator dependent imprecision to measuring the peak area. Reproducible injection using a manually operated injector device like the Merlin MicroShot Injector (Merlin Instrument Co.), are necessary to reduce the sampling noise. If the method uses headspace or gas analysis, it could be costly to test as a liquid injection and then re-configure to for the final analysis. These costs must be managed appropriately; but factors like cost and complexity for some configurations should not prevent a transition to a better, statistically based standard such as %RSD for the majority of the systems which do not have these inlet limitations.

From the customers perspective however, there is a significant benefit to having a system level test of performance that offers a realistic estimate of the IDL, and the system precision near

the limits of detection where it is the most critical. The purchase of a mass spectrometer is a significant capital expenditure, and often purchasing decisions are based on a customer's specific application need. Instrument manufacturers do not have the resources to demonstrate their product's performance for every different application. It would be valuable to have a simple, but realistic means of evaluating a particular MS instrument's performance in the form of the IDL and the system precision. These will form lower bounds to the MDL and method precision for specific applications.

Summary

In the past, the SNR of a chromatographic peak determined from a single measurement has served as a convenient figure of merit used to compare the performance of two different MS systems. However, as we have seen, this parameter can no longer be universally applied and often fails to provide meaningful estimates of the IDL. A more practical means of comparing instrument performance is to use the multi-injection statistical methodology that is commonly used to establish MDLs for trace analysis in complex matrices. Using the mean value and RSD of replicate injections provides a way to estimate the statistical significance of differences between low level analyte

6

responses and the combined uncertainties of the analyte and background measurement, and the uncertainties in the analyte sampling process. This is especially true for modern mass spectrometers for which the background noise is often nearly zero. The RSD method of characterizing instrument performance is rigorously and statistically valid for both high and low background noise conditions. Instrument manufacturers should start to provide customers an alternative performance metric in the form of RSD based instrument detection limits to allow them a practical means of evaluating an MS system for their intend application.

References

1. European Pharmacopoeia vol. 1; 7th Edition

2. United States Pharmacopeia, XX Revision (USP Rockville, Maryland, 1988)

3. Japanese Pharmacopoeia, 14th Edition

4. ASTM: Section E 682 – 93, Annual Book of ASTM Standards, Vol.14.01

5. *Handbook of Residue Analytical Methods for Agrochemicals,* Vol. 1; Edited by: Lee, Philip W.; (2003) John Wiley & Sons; Chapter 4, "Best Practices in Establishing Detection and Quantification Limits for Pesticide Residues in Foods" by Johannes Corley

6. U.S. EPA - Title 40: Protection of Environment; Part 136 – Guidelines Establishing Test Procedures for the Analysis of Pollutants; Appendix B to Part 136 – Definition and Procedure for the Determination of the Method Detection Limit – Revision 1.11.

7. "Uncertainty Estimation and Figures of Merit for Multivariate Calibration"; IUPAC Technical Report, *Pure Appl. Chem.;* Vol. 78, No. 3, pp 633-661, 2006.

8. Official Journal of the European Communities; Commission Decision of 12 August 2002; Implementing Council Directive 96/23/EC concerning the performance of analytical methods and the interpretation of results.

9. "Guidelines for Data Acquisition and Data Quality Evaluation in Environmental Chemistry"; ACS COMMITTEE ON ENVIRONMENTAL IMPROVEMENT; *Anal. Chem.* 1980, 52, 2242-2249.

10. "Signal, Noise, and Detection Limits in Mass Spectrometry," Agilent Technologies Technical Note, publication 5990-7651EN

11. "Statistics;" D. R. Anderson, D. J. Sweeney, T. A. Williams; West Publishing, New York, (1996)

12. "Data Reduction and Error Analysis for the Physical Sciences;" Philip R. Bevington and D. Keith Robinson; WCB McGraw-Hill, Boston, 2nd Edition, (1992)

For More Information

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

Agilent Technologies