Sensitivity, Background, Noise, and Calibration in Atomic Spectroscopy: Effects on Accuracy and Detection Limits

Introduction

Proper calibration in atomic spectroscopy and an understanding of uncertainty is fundamental to achieving accurate results. This paper provides a practical discussion of the effects of noise, error and concentration range of calibration curves on the ability to determine the concentration of a given element with reasonable accuracy. The determination of lower limits of quantitation and lower limits of detection will also be discussed.

Results accuracy is highly dependent on blank contamination, linearity of calibration standards, curve-fitting choices and the range of concentrations chosen for calibration. Additional factors include the use of internal standards (and proper selection of internal standards) and instrumental settings.

This paper is not intended to be a rigorous treatment of statistics in calibration; many references are available for this, such as "Statistics in Analytical Chemistry[1]".

The techniques of atomic spectroscopy have been extensively developed and widely accepted in laboratories engaged in a broad range of elemental analyses, from ultra-trace quantitation at the sub-ppt level, to trace metal determination in the ppb to ppm range, to major component analysis at percent level composition.

A fundamental part of analysis is establishing a calibration curve for quantitative analysis. A series of known solutions is analyzed, including a "blank" prepared to contain no measurable amounts of the elements of interest. This solution is designated as "zero" concentration and, together with one or more known standards, comprises the calibration curve. Samples are then analyzed and compared to the mathematic calculation of signal vs. concentration established by the calibration standards. Unfortunately, preparation of contamination-free blanks and diluents (especially when analyzing for many elements), perfectly accurate standards, and perfect laboratory measurements are all impossible.

The three most common atomic spectroscopy techniques are atomic absorption (AA) spectroscopy, ICP optical emission spectroscopy (ICP-OES) and ICP mass spectrometry (ICP-MS). Of these, ICP-OES and ICP-MS are very linear; that is, a plot of concentration vs. intensity forms a straight line over a wide range of concentrations (Figure 1). AA is linear over a much smaller range and begins to curve downward at higher concentrations (Figure 2). Linear ranges are well understood, and, for AA, a rule of thumb can be applied to estimate the maximum working range using a non-linear algorithm.

This paper will discuss the contributions of sensitivity, background, noise, calibration range, calibration mathematics and contamination on the ability to achieve best accuracy and lowest detection limits.
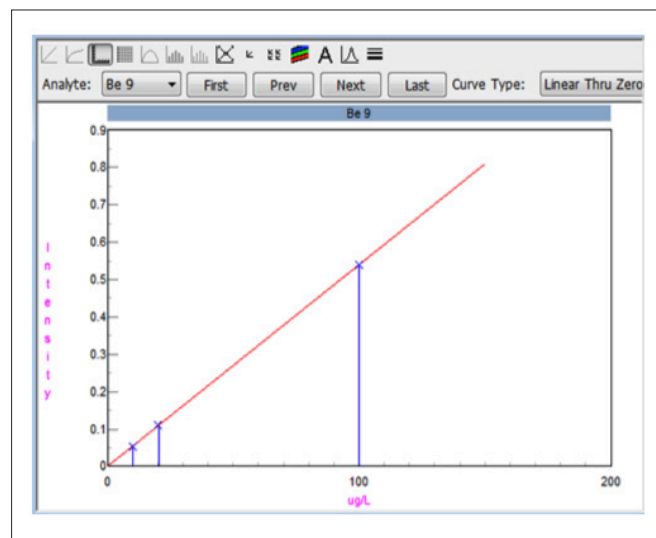

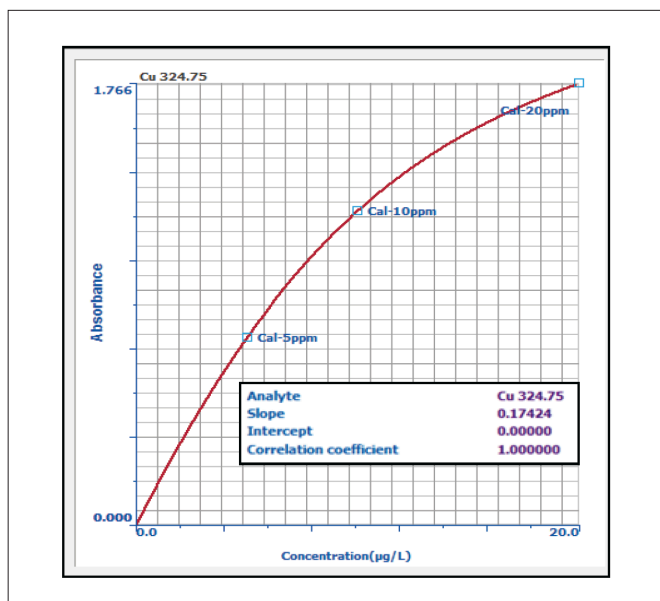*Figure 1.* Example of a linear calibration curve in ICP-MS.

PerkinElmer®

For the Better

Figure 2. Example of a non-linear calibration curve in AA.

Detection Limits – "How Low Can We Go?"

Under ideal conditions, the detection limits of the various techniques range from sub part per trillion (ppt) to sub part per million (ppm), as shown in Figure 3. As seen in this figure, the best technique for an analysis will be largely dependent on the levels that need to measured, among other considerations.
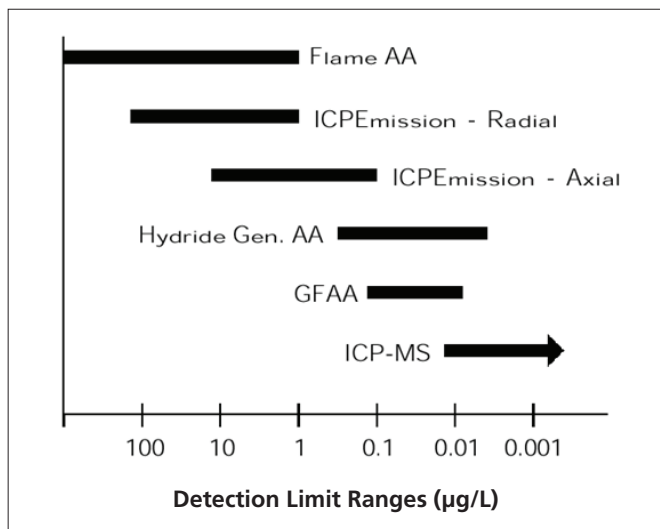

Figure 3. Detection limit ranges of the various atomic spectroscopy techniques.

It is important to realize that detection limits are not the same as reporting limits: just because a concentration can be detected does not mean it can be reported with confidence. There are many factors associated with both detection limits and reporting limits which distinguish each, as will be seen in the following discussion.

A discussion of detection limits can be lengthy and is subject to many interpretations and assumptions – there are even several definitions of "detection limits". In atomic spectroscopy, there are some commonly accepted approaches. Under some analytical protocols, the determination of detection limits is explicitly defined in a procedure, as in U.S. EPA methods 200.7 and 6010 for ICP, 200.8 and 6020 for ICP-MS, and 200.9 for

graphite furnace AA, based on a more thorough and widely-ranging definition[2]. However, these procedures are specific to water, wastewater and other environmental samples.

There are few widely accepted approaches for other matrices such as food, alloys, geological and plant materials, etc. It is left to the lab doing the analysis to establish an approach to answer the question "How low a concentration of a given element can we detect in a particular sample matrix?" Because there exists a long history in most labs and many different techniques are employed (e.g., GC, LC, UV/Vis, FT-IR, AA, ICP, and many others) there are many opinions and approaches to this subject.

How, Then, Do We Establish "How Low We Can Go"?
The simplest definition of a detection limit is the lowest concentration that can be differentiated from a blank. The blank should be free of analyte, at least to the limit of the technique used to measure it. Assuming the blank is "clean", what is the lowest concentration that can be detected above the average blank signal?

One way to determine this is to first calibrate the instrument with a blank and some standards, calculate a calibration curve, and then attempt to measure known solutions as samples at lower and lower concentrations until we reach the point that the reported concentration is indistinguishable from the blank. Unfortunately, no measurement at any concentration is perfect – that is, there is always an uncertainty associated with every measurement in a lab. This is often called "noise", and this uncertainty has several components. (A detailed discussion of sources of uncertainty is also a lengthy discussion[3] and beyond the scope of this document.) So, to minimize uncertainty, it is common to perform replicate measurements and average them.

The challenge, then, is to find the lowest concentration that can be distinguished from the uncertainty of the blank. This can be estimated by using a simple statistical approach. For example, U.S. EPA methods for water use such an approach. After calibration, a blank is run as a sample 10 times, the 10 reported concentrations are averaged and the standard deviation is calculated. A test for statistical significance is applied (the Student's t-test) to calculate what the concentration would be that could be successfully differentiated from the blank with a high degree of confidence. In the U.S. EPA protocol, a 99% confidence is required. This equates to three times the standard deviation of 10 replicate readings. This is also known as a 3σ (sigma) detection limit and is designated as the Instrument Detection Limit (IDL) for U.S. EPA methods.

It is important to note that the statistically calculated detection limit is the lowest concentration that could even be detected in a simple, clean matrix such as 1% HNO_3 – it is not repeatable or reliable as a reported value in a real sample.

To be more repeatable, the signal (with its associated uncertainty) must be significantly higher than the uncertainty of the blank, perhaps 5-10 times the standard deviation of the blank. This is a judgment by the lab as to how confident the reported value should be. This concentration level might be called the lowest quantitation limit, sometimes known as PQL (practical quantitation limit), LOQ (limit of quantitation) or RL (reporting limit). There are no universally accepted rules for determining this limit.

Again, the EPA has guidelines in some methods for water samples that establish a lower reporting limit as a concentration that can be measured with no worse than +/- 30% accuracy in a prepared standard. This rule is only applicable to the specific EPA method and in water samples.

Many labs apply the EPA detection limit methodologies simply because there are few commonly accepted and carefully defined approaches for other sample types. Some industries follow ASTM, AOAC or other industry guidelines, and some of these procedures include lower-limit discussions.

When analyzing a solid sample, the sample must first be brought into solution, which necessarily involves dilution. The estimation of detection or quantitation limits now needs to account for the dilution factor and matrix effects. For example, if 1 g of sample is dissolved and brought to a final volume of 100 mL, the detection limit in the solution must be multiplied by the dilution factor (100x in this example) to know what level of analyte in the original solid sample could have been measured if it could have been analyzed directly.

To use the statistical estimate technique (3σ detection limit), a "clean" matrix sample must be available, but this is not always possible. The very product or incoming material being evaluated may be the best example available, but it may be contaminated with the element being determined – indeed, this is the purpose of the analysis. Finding a true "blank" is difficult.

In many cases, then, a more empirical or practical approach is taken. After preparing the sample, it is spiked with a known amount of analyte and measured. If the spike recovery is accurate within some acceptable limits (this is also subject to many opinions and there is no "rule" about what is acceptable accuracy, although the EPA has a rule for spike recovery when analyzing environmental samples), then a lower concentration spike is attempted. After a series of lower and lower concentration spikes, there will be a point at which "confidence" is lost. The analyst would then set a limit for reporting that is at or above a level that gives reasonable accuracy. This is a gray area and is up to the lab to decide. Again, a multiplier for dilution factor must be incorporated into this reporting limit if the sample has been diluted for analysis.

What factors are important for achieving the best possible detection limits? They are:

1. Signal "strength" or sensitivity

2. Background

3. Noise

4. Stability

Let's examine each of these.

Sensitivity Plays a Role in Detection Limits
Signal strength (intensity) must be sufficient to differentiate the presence of an element above the background AND noise; this is known as "sensitivity" and is an important characteristic. However, sensitivity is not, by itself, sufficient to predict detection

limits. For example, if contamination is present, a 10-fold increase in signal will also increase the background 10-fold. In an ideal situation (i.e. the absence of high background or excessive noise) detection limits theoretically improve by the square root of the increase in signal intensity (i.e. a 9x increase in intensity will improve detection limits 3x). However, as we will see, background level and noise have as much of a contribution to detection limits as intensity.

In ICP-OES and ICP-MS, a common way to express intensity is "counts per second", or cps. We will use this unit in the following discussion.

Background Signal Plays an Important Role in Detection Limits
Consider an example with a signal that gives an easily measurable intensity of 1000 cps. However, if the background is 10,000 cps the signal is small relative to the background. It is common to express the relationship between signal and background as the "signal-to-background ratio" or S/B. (This is often referred to as "signal-to-noise" or S/N, but background and noise are two distinct characteristics). The above example would have an S/B of 0.1. However, if the background were only 1 cps, the same 1000 cps signal would have a S/B of 1000, or 10,000 times better than the first example. These two examples are illustrated in Figure 4, which shows that a smaller signal, say 100 cps, is easier to distinguish from a background of 1 cps than a background of 10,000 cps.

All signals are measured in the presence of some degree of background, which can originate from a variety of sources, including detector and electronic characteristics, emitted light from the excitation source (prevalent in ICP-OES), interfering ions formed in the source or from the matrix (prevalent in ICP-MS), or contamination. A quantitative measure of background level is called the "Background Equivalent Concentration", or BEC, which is defined as the concentration of a given element that exhibits the same intensity as the background, measured at a given wavelength (ICP-OES) or mass (ICP-MS). BEC is calculated with the following formula:

$$BEC = \frac{I_{blank}}{I_{standard} - I_{blank}} * C_{standard}$$

I_{blank} = Intensity of the blank

$I_{standard}$ = Intensity of the standard

$C_{standard}$ = Concentration of the standard

The units of the BEC are the same as the units of the standard.

To illustrate, consider an example of two elements, A and D. Suppose Element A has a sensitivity of 1000 cps/ppb, and element D has a sensitivity of 10,000 cps/ppb. We would say that D has 10 times more sensitivity than A. However, if the mass of D is the same as a common background species produced in the argon plasma (such as ArO+), the background signal would be high (100,000 cps for ArO+ in this example). If Element A has no interfering background species, the background could be 1 cps, due only to electronic effects.

Even though element D is 10x more sensitive, the ability to detect D is worse than A, as shown in Figure 4. The BEC for A is 0.001 ppb while the BEC for D is 10 ppb. Note that BEC is NOT a detection limit, but an indicator of relative size of the signal from the element and background. The lower the BEC, the lower the detection limit.
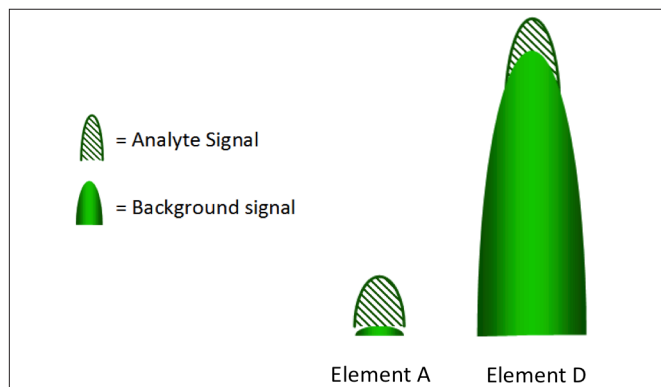

Figure 4. Graphical representation demonstrating that a constant analyte signal (stripes) is more easily seen with a lower background signal (green).

Another case is an element that normally has no background contribution above the instrument background of 1 cps – lead (Pb). In the example in Figure 5, the Pb background is 100 cps, due to Pb contamination. As a result, the Pb BEC and detection limits both increase, compared to an analysis with no Pb contamination.
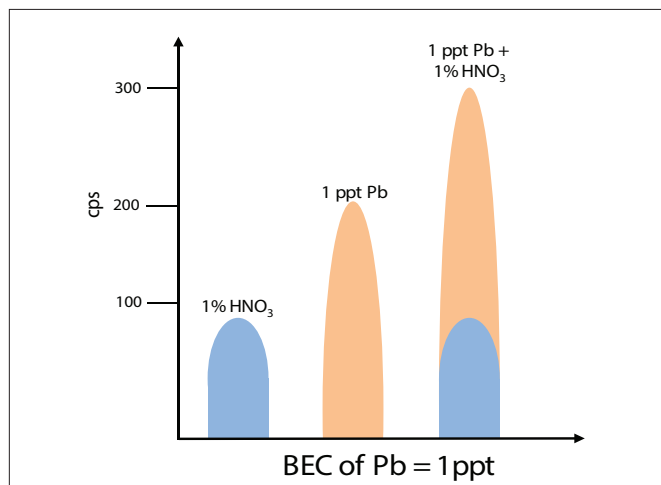

Figure 5. Relationship of BEC and DL on lead (Pb). Signal from Pb is in orange; the signal from the background is blue.

Now consider the effect of increased sensitivity on BEC and detection limit for Pb and holmium (Ho). While Pb is a very common element, Ho is not. As a result, Pb contamination is common, while Ho contamination is rare. If there is no interference or contamination on an analyte mass, a 10x increase in sensitivity improves the DL by a factor of 3 (the square root of the increase). In Table 1, Ho has two different sensitivities.

Table 1. Holmium at Medium and High Sensitivities.

Sample	Medium Sensitivity	High Sensitivity
Blank (cps)	0.2	0.2
Standard (cps)	121000	1270000
BEC (ppt)	0.002	0.0002
DL (ppt)	0.002	0.0007

At medium sensitivity, a Ho blank gives 0.2 cps, and a 1 ppb standard gives 121,000 cps, which yields a BEC and detection limit of 0.002 ppt. With a 10x higher signal (1,270,000 cps) but no increase in blank, the BEC and detection limit improve to 0.0002 ppt and 0.0007 ppt, respectively.

Now consider the same situation for Pb, as shown in Table 2.

Table 2. Lead at Medium and High Sensitivities.

Sample	Medium Sensitivity	High Sensitivity
Blank (cps)	78	701
Standard (cps)	81000	809000
BEC (ppt)	1.0	0.9
DL (ppt)	0.38	0.42

In this example, there is Pb contamination in blank, so when sensitivity increases 10x, the signals in both the blank and standard increase 10x, meaning the S/B remains constant. As a result, there is no improvement in BEC or DL, despite the increase in sensitivity. Since most elements measured by ICP-MS have some interference or contamination on their masses, *higher sensitivity does not always improve BEC or DLs.*

Noise Plays an Important Role in Detection Limits
In the example above, the inherent assumption is that the signal and the background are perfect measurements; that is, 1000 cps has no variability, and the background has no variability. However, this is never the case in a real laboratory measurement: all measurements have uncertainty, which is referred to as "noise". For example, if a signal of 1000 cps was measured five times, it would vary - 995, 1028, 992, 1036, 987 cps. This is why replicate readings are measured: because they are never perfect, we want to measure multiple times and take the average. In the above example, an average of the five readings (1008 cps) would be reported. How much variability there is among the replicates would be indicated by using the Standard Deviation (for absolute variation) and/or the Relative Standard Deviation (for a percentage). So, the reported value in cps would be:

1008 +/- 20 cps (using the standard deviation)

or

1008 +/- 2% (using the relative standard deviation)

Now superimpose "noise" or "uncertainty" on an example, assuming a signal of 1000 cps and a background of 10000 cps:

- If the signal of 1000 cps had an uncertainty of +/- 2%, then the signal could be 980 – 1020 cps.

- If the background also had an uncertainty of 2%, then its range could be 9800-10200 cps.

Figure 6 shows an illustration of a small signal on top of background + noise.
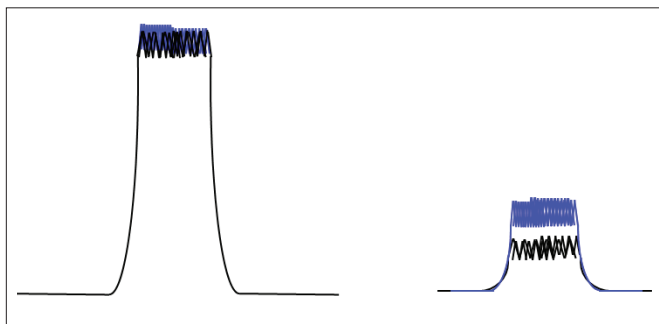

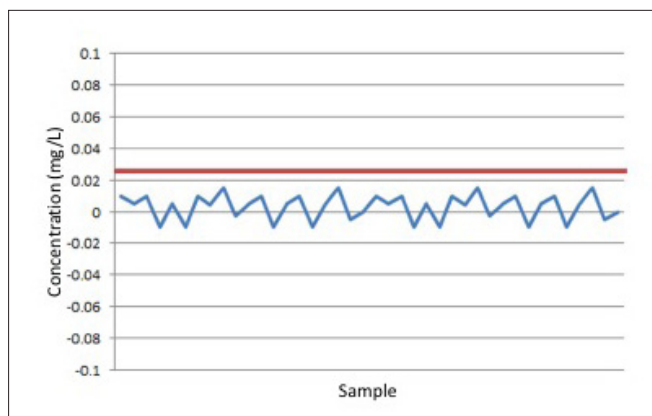*Figure 6.* Effect of background and noise on a small signal.

Figure 7. Concentration measurements of a blank (blue line) and associated detection limit (red line), determined as three times the standard deviation of the blank.
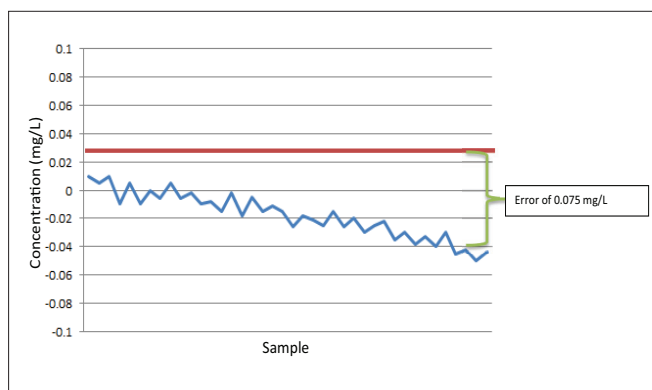

Figure 8. The reported concentration of a blank sample (blue line) decreases over time due to instrument drift. During the final reading, the sample reads 0.075 mg/L below the detection limit (red line).

On the left, a small analyte signal (blue) is "buried" in the natural uncertainty of the high background and is difficult to see. In addition, the analyte signal has its own uncertainty. If the background is high relative to the analyte signal, the ability to quantitate (or even detect) is compromised. On the right, the same analyte signal is easily differentiated from a low background. At very low concentrations near detection limits, longer integration times are necessary to average the noise of both signals.

Stability is Important For Reliable Detection Limits

Stability plays a large role in detection limits in that small signals must be very steady to give valid noise averaging. The detection limit is the smallest signal that, on average, can be distinguished from the average noise of the blank, as shown in Figure 6. It is important to note that the noise and stability are as important as the absolute signal size in being able to differentiate a "real" signal from the noise of the blank.

Furthermore, if an analytical sequence lasts for many minutes or hours, the baseline must be stable for small signals to be measured as accurately as statistical variation allows. Figure 7 shows a blank analyzed repeatedly (after calibration, results in concentration units of mg/L). The variation would be called "noise". Following the EPA protocol for estimating detection limit, the first 10 readings were averaged and the Standard Deviation calculated, giving a value of 0.0086 mg/L; three times the standard deviation (99% confidence level, Student-t test) equals 0.025 mg/L. The red line in Figure 7 shows this 3σ detection limit for this analysis.

However, if the analytical conditions are not stable over the time frame of the sample run (i.e. batch of samples), the reported results can be well below the calculated detection limit. Figure 8 shows the effect of drift: towards the end of the run, reported concentrations for a blank sample are as low as -0.05 mg/L. Therefore, a sample with as much as 0.075 mg/L would report as "below detection limit" of 0.025 mg/L.

Calibration – Effects on Accuracy and Detection Limits

All quantitative answers reported for a sample are calculated from the calibration curve. Proper calibration is crucial for accurate results AND ESPECIALLY for low-level results near detection limits.

A characteristic of atomic spectroscopy instrumentation is linearity: ICP-OES has a linear range of 5-6 orders of magnitude, while the linear range of ICP-MS is 10-11 orders of magnitude. Statistically, these are accurate statements. However, consider a series of calibration standards from 1 ppt to 1000 ppm (a range of 10^9 or nine orders of magnitude) analyzed by ICP-MS. The resulting calibration curve would be linear, giving a correlation coefficient (R^2) of 0.9999 or better (a perfect fit would have a correlation of 1.000000). A common misunderstanding of this statistic is that any concentration from 1 ppt to 1000 ppm could then be read accurately because the curve is "linear".

Figure 9 shows an excellent linear relationship over six orders of magnitude (R^2 = 0.999905). Later we will see there is a hidden problem with this apparently excellent linear calibration, but first some understanding of how real data operates is in order.

The only way perfect accuracy across such wide ranges can be achieved is if every measurement is perfect – which is not possible. Every standard on the curve has an associated error. In fact, the calculation of "best fit" of the calibration data points is a process (Linear Least-Squares fitting) of minimizing the sum of the squares of all the ABSOLUTE errors of the series of data points on the curve, as shown in the example in Table 3.

To find the "best fit" straight line, the higher standards become far more important than the lower standards, which causes linear curves (even with R^2 = 0.99999) to be inaccurate at the low end of a wide range of calibration concentrations, as shown in Figure 10.

Notice how the "Best Fit" line in the graph in Figure 10 is almost at the center of the error bar of the highest standard, while the lower standards are increasingly "off the line" toward the low end. This is the effect of the highest absolute error contributing the most to the overall fit.
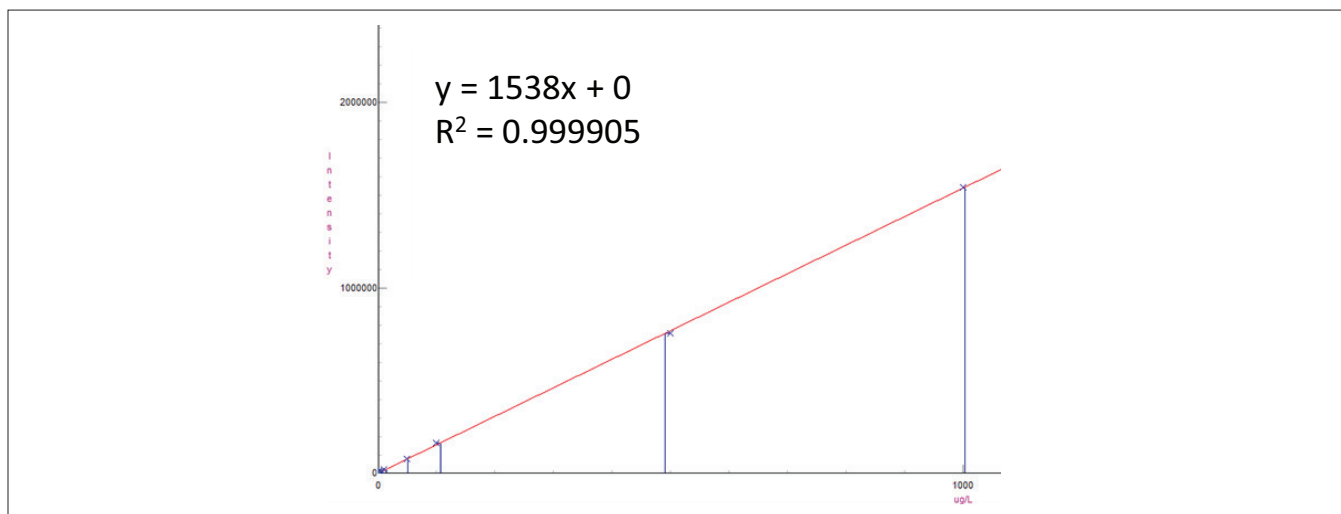

$$y = 1538x + 0$$
$$R^2 = 0.999905$$

Figure 9. Calibration curve for 11 standards from 10 ppt to 1,000 ppb (six orders of magnitude) with an R^2 = 0.999905.

Table 3. Hypothetical calibration with some preparation errors.

Concentration (ppb)	Mean Intensity (cps)	Error of 1% (cps)	Mean Intensity with 1% Error (cps)
0	0	0	0
0.1	100	+/- 1	99 - 101
1	800	+/- 8	792 – 808
10	11000	+/- 110	10890 - 11100
50	46000	+/- 460	45540 - 46460
100	95000	+/- 950	94050 -95950

This statistical reality affects BOTH accuracy at low levels and detection limits. Therefore, if accuracy at concentrations near the lower limits of detection is the most important criteria, a calibration curve that does not include very high standards is preferable. For example, if Se is to be determined and most samples will be below 10 ppb with a need to report down to 0.1 ppb, the instrument should be calibrated with standards in this range. A blank (0) and three standards at 0.5, 2.0 and 10.0 ppb will give far better accuracy at the 0.1 ppb level than a calibration curve with standards of 0.1, 10.0 and 100 ppb. If even a higher standard (say 500 ppb) were included, the ability to read 0.1 ppb accurately would be nearly impossible.
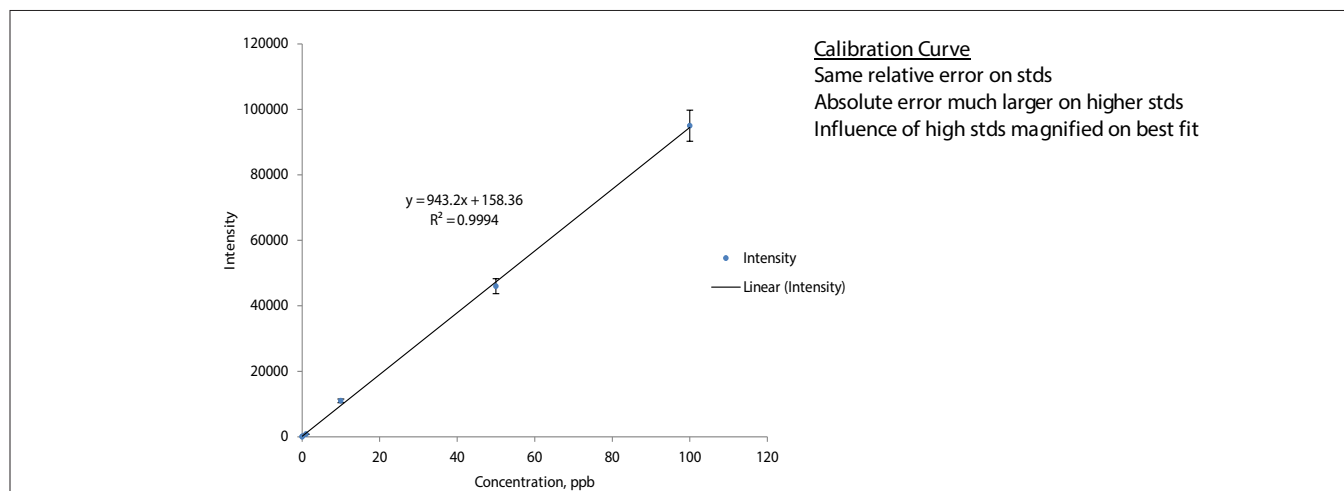

Figure 10. Effect of error on calibration curve with standards covering a wide concentration range.

Using a real example to further illustrate, Figure 11 shows a calibration curve for Zn by ICP-MS (this is the same curve as in Figure 9). With standards at 0.01, 0.05, 0.1, 0.5, 1, 5, 10, 50, 100, 500 and 1000 ppb, the correlation coefficient (R^2) of 0.999905 indicates an excellent calibration. However, when a 0.100 ppb prepared standard is analyzed as a sample against this curve, it reads 4.002 ppb. How can this be?

An expanded view of the low end of the curve reveals the problem: contamination on the lowest seven standards, a common problem with zinc. However, this issue is not apparent from the excellent linear statistics of the complete curve. The lower standards contribute almost nothing statistically to the Least-Squares fit compared to the four highest standards.

The previous example shows the effect of contamination on low-level standards, but blank contamination is another common problem that will compromise accuracy. The calibration blank is assumed to be zero concentration. The measured signal of the blank is subtracted from all subsequent measurements (standards and samples). If the blank signal intensity is higher than that of a sample, then the net blank-subtracted sample signal will calculate to a negative concentration value, as illustrated in the following ICP-OES example.

An analysis of pure aluminum is performed, looking for silicon as an analyte. After calibration, the sample (1000 ppm Al) shows a silicon concentration of -0.210 mg/L. There could be several reasons for this: improper background correction, an interference from a nearby peak, internal standard response. However, as shown in Figure 12, the problem is a poor calibration curve, as shown in Figure 12a. The cause: a contaminated blank, which can clearly be seen in Figure 12b. In this figure, the blank spectrum for Si (yellow) has a peak that is higher than the sample (blue spectrum). As a result, the sample will be reported as negative. Additionally, since the blank is also subtracted from the calibration standards, the calibration plot is not linear. Clearly Si is present in the sample (blue spectrum), but calculating the concentration with a blank that is contaminated with Si is not possible.
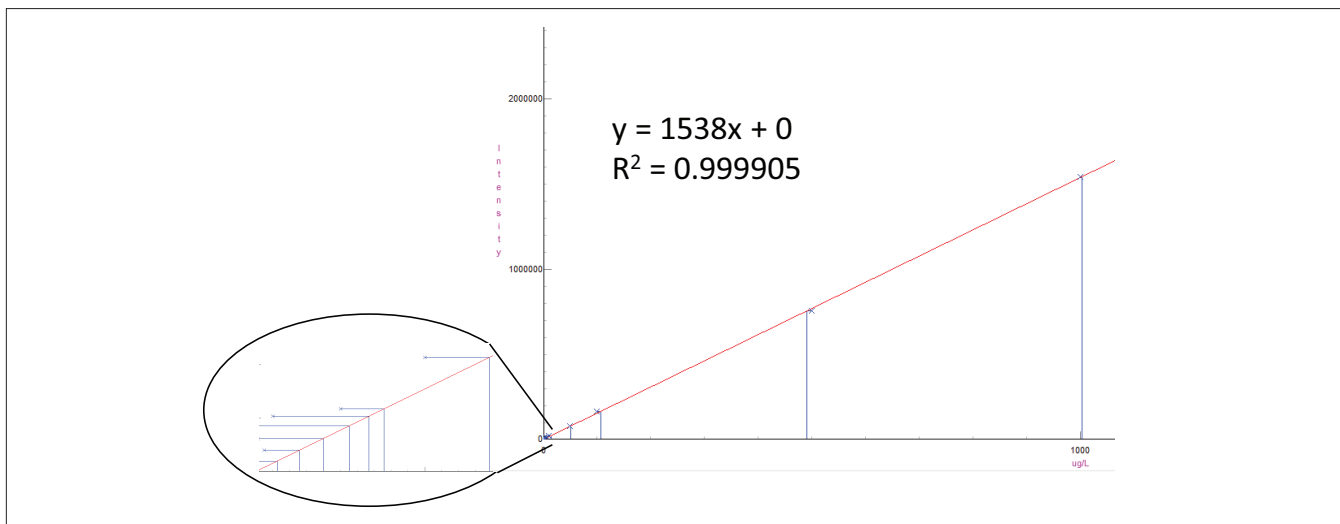

$$y = 1538x + 0$$
$$R^2 = 0.999905$$

Figure 11. Calibration curve for Zn showing excellent linear statistics, but poor low-level performance.


```
Mean Data: 1000 ppm Al
                    Mean Corrected          Calib.
Analyte             Intensity         Conc. Units     Std.Dev.
Si 251.611†          -16018.7         -0.210 mg/L      0.0369
```
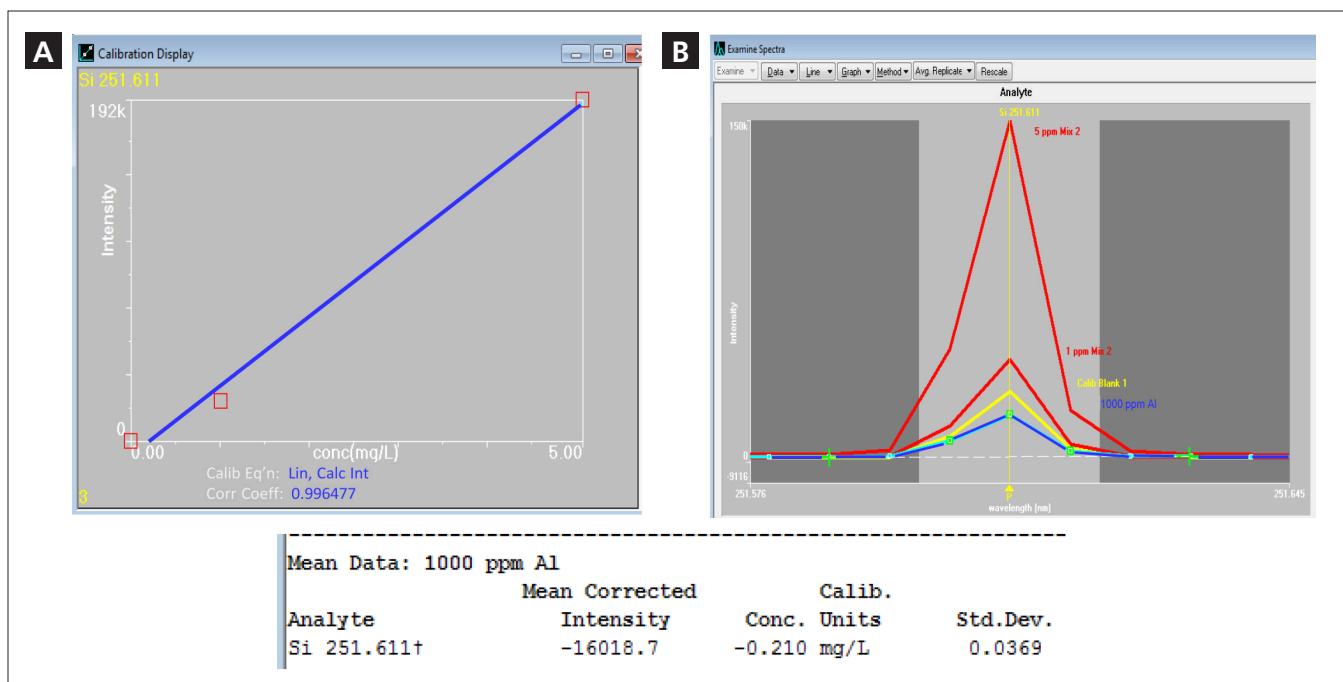
Figure 12. Effect of blank contamination. (A) Si calibration curve with a poor correlation coefficient due to the low standard not falling on the curve. (B) Si spectra of the blank (yellow) and calibration standards (red), and sample (blue). The presence of Si in blank affects the calibration curve and causes the sample to read negative.

To make matters worse in this example, the pure solid aluminum was dissolved and diluted a factor of 1000 to give a final solution of 1000 mg/L. The reported answer of -0.210 would be multiplied by the dilution factor (1000x) to give a final Si "concentration" of -210 mg/L!

Clean blanks and diluents are essential for good accuracy AND detection limits.

Different Calibration Mathematics – Effect on Accuracy

Linear Through Zero

The previous discussions and examples all used "linear through zero" calibrations (i.e. the regression curve is forced through the origin) and have shown that the highest standards carry the greatest influence. In Figure 13, the two lowest standards do not fit the linear-through zero best fit line very well, although the R^2 is a quite acceptable 0.9997. The analytical problem is that a 5 ppb sample would calculate as about 7 ppb (a 40% error), and a 10 ppb sample would calculate as about 11.7 ppb (17% error).

Weighted Linear

Another approach would be to make the lowest standards carry the greatest weight in the linear statistical fit. This is known as "weighted linear" or "inversely weighted linear". Instead of calculating the least squared sum of errors using the absolute errors of the standards, this approach calculates the linear fit from the least squared sum of 1/error of each standard ($1/x^2$). The highest standards now contribute the least to the overall fit.

The advantage to this approach is that the low end of the curve "fits" better than the high end, improving accuracy at the low end. However, accuracy at the high end can suffer.

Using the example from Figure 13, but applying a weighted linear fit to the data, the R^2 improved slightly to 0.9999 and low-end accuracy is much improved, as shown in Figure 14. A 5 ppb sample calculates as 5.01 ppb and a 10 ppb sample reads as 9.97 ppb. At the high end, a 100 ppb reads 101.1 ppb.
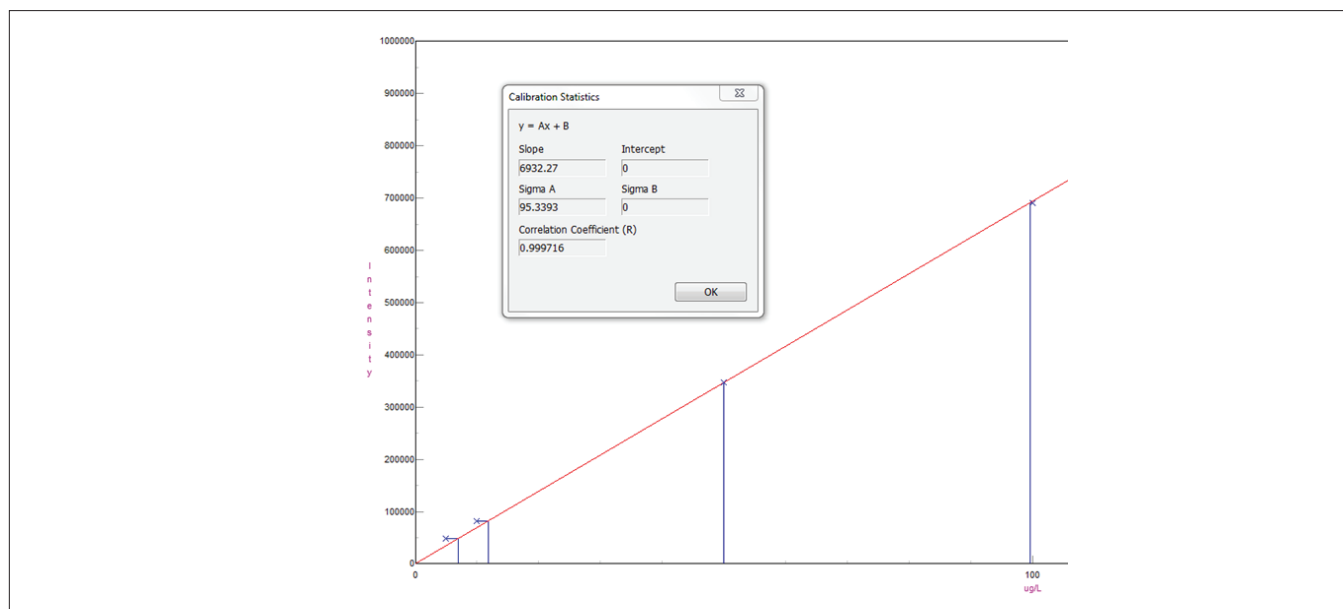

Figure 13. Linear through zero calibration curve showing the influence of high concentration standards on low concentrations.
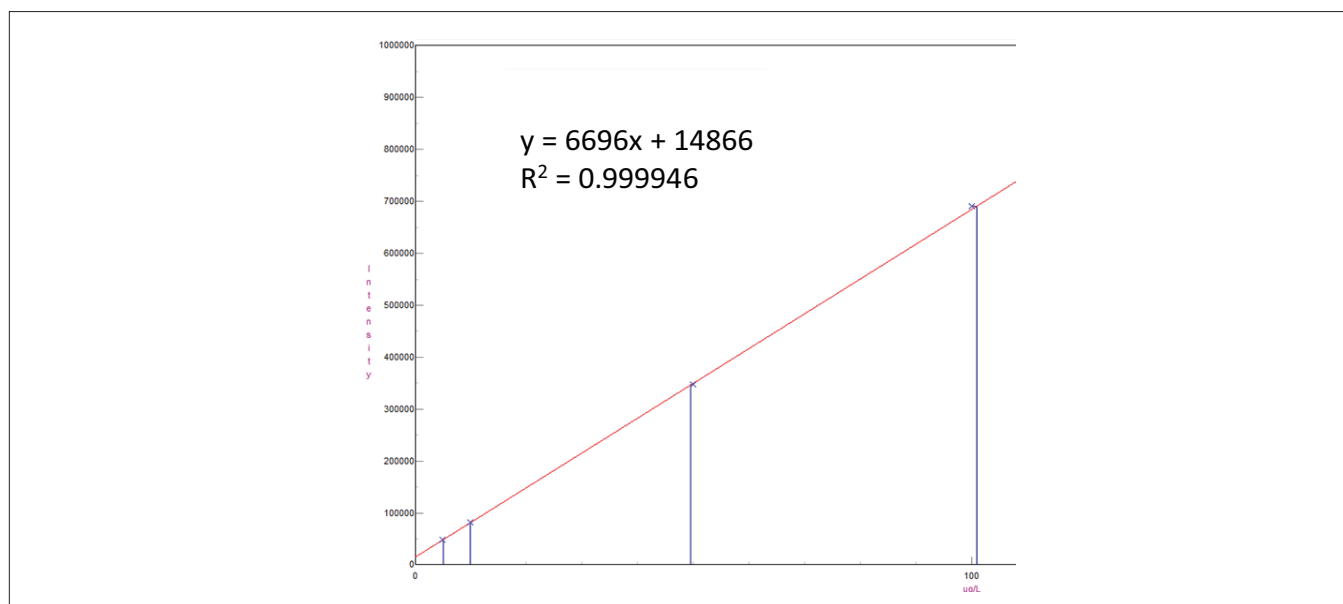

Figure 14. The same calibration curve as in Figure 13, but using a weighted linear fit, which places more emphasis on the low-level standards.

However, if there is contamination in the low standards, the effect/error is magnified and can render a weighted linear curve unfit for any practical purpose.

Simple Linear (Linear With Calculated Intercept)

A popular approach to curve fitting is "simple linear" or "linear with calculated intercept". In this approach, the blank is just another point on the curve and is included in the overall fit – the curve does not have to pass through the origin as with "linear through zero".

Again using the same data as the previous examples, Figure 15 shows a simple linear fit gives R^2 of 0.9999. A 5 ppb sample would calculate as 5.22 ppb, a 10 ppb sample would calculate as 10.13 ppb, and a 100 ppb sample would calculate as 100.3 ppb.

Calibration response in AA is linear only over two to three orders of magnitude, and many labs need to calibrate over a wider range to encompass the samples of interest. This is where non-linear calibration can be useful. The curvature of AA is well understood, and an accurate, non-linear algorithm was published in 1984[4] that allows calibration up to 6X of the linear response (Figure 16).

Method of Standard Additions

The Method of Standard Additions (MSA) is treated separately from the other algorithms as it differs in a significant aspect: MSA creates the calibration curve in the sample itself by adding known concentrations of analyte. All other calibrations in the previous discussions were created from a series of known solutions in a clean, simple matrix known as "external standards".

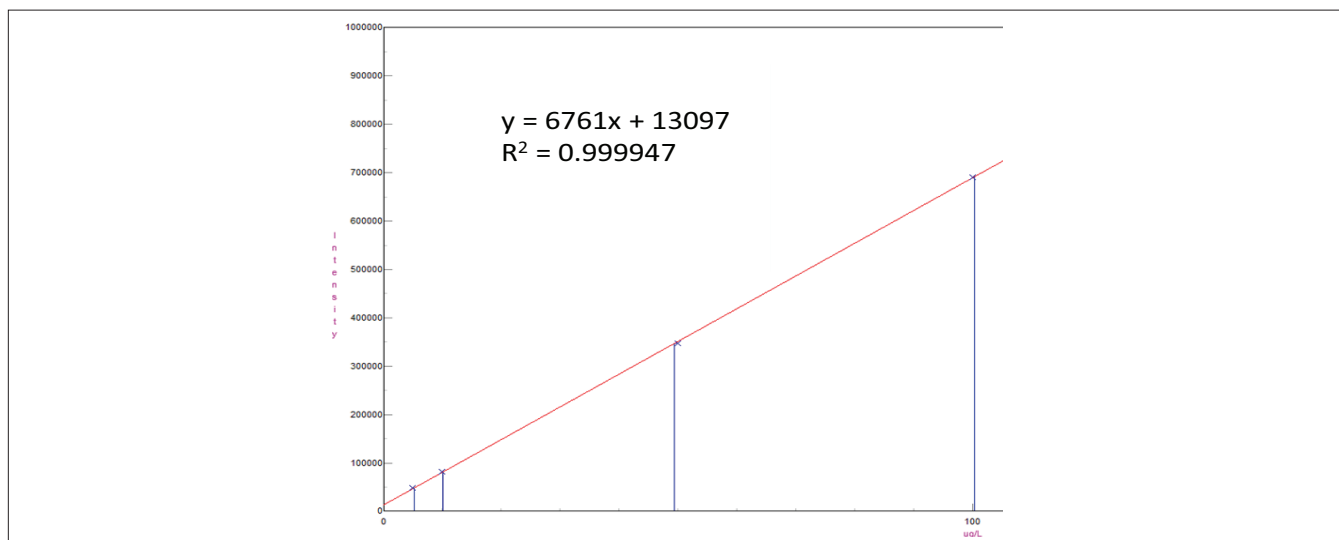

$$y = 6761x + 13097$$
$$R^2 = 0.999947$$

Figure 15. The same calibration curve as in Figure 13, but using a simple linear fit, where the origin is included as a data point in the regression calculation, but the curve does not have to pass through the origin.

There are no hard rules about which algorithm to use unless one is required in a particular prescribed method. A practical approach is to use the curve fitting algorithm that gives best accuracy on known check samples, especially near the low end of the calibration range.

Non-linear Through Zero

Calibration response in ICP-MS should be linear over a wide range of concentrations. If non-linearity is observed, an incorrect setup should be suspected. Variables include:

a. Improper standards preparation

b. Standards that extend above the bottom range of the detector, and detector calibration factors that are not accurate

c. An unrecognized interference that is not properly subtracted or eliminated

Calibration response in ICP-OES is also very linear, but can become non-linear due to a phenomenon known as self-absorption at high concentrations. If the concentration range is too broad to be linear, alternate wavelengths should be considered that will give a linear response over the range of anticipated concentrations. Using a modern ICP optical emission spectrometer with a solid-state detector (i.e. having many alternative wavelengths) should eliminate the need for non-linear calibrations.
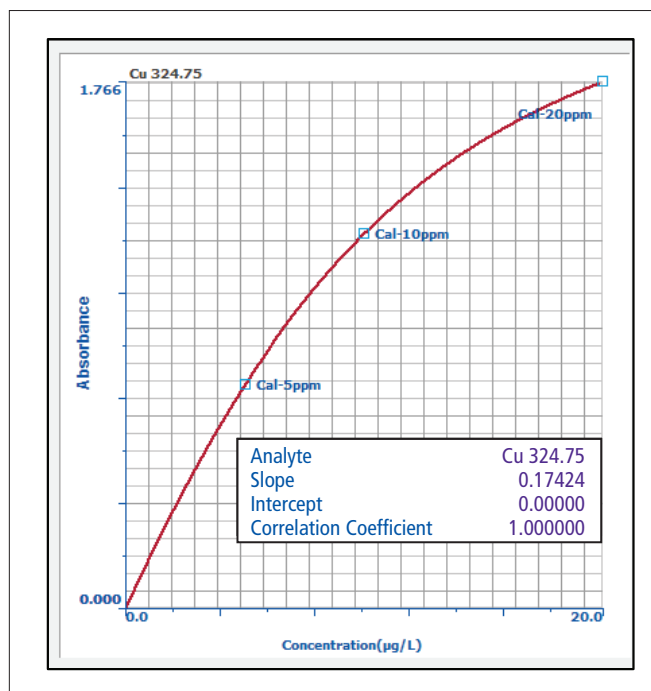

Figure 16. Copper calibration curve by flame AA up to 20 mg/L using a non-linear algorithm, 4x higher than the linear range.

In some sample types, the sample "behaves" differently from the external standards due to a variety of factors, including viscosity, acid content and/or dissolved solids, among others. The outcome: biased results due to a different "sample transport efficiency" between the calibration standards and the samples. The technique of adding internal standards to the blank, standards and samples and using the change in response of the internal standard to correct for these physical effects is a popular way to adjust reported concentrations in samples. Additionally, it is desirable to "matrix match" the blank and standards to the samples to minimize this effect. If samples are digested in a 2% HNO_3 / 1% HCl mixture, preparing a calibration blank and standards in the same acid mix will minimize sample transport differences. An internal standard is still recommended to monitor and correct for any residual effects.

However, in some cases the matrix effects are severe and achieving an accurate result is difficult. Using MSA allows a "perfect match" of the standard response to that of the sample because the calibration slope is calculated in the sample matrix itself.

As an example, consider a maple syrup sample with an unknown amount of aluminum contamination. To analyze for aluminum, external standards prepared in the typical 1% nitric acid diluent will not give an accurate comparison because maple syrup is much more viscous that even a dilution will need an internal standard to correct for the sample transport difference.

A more accurate way to quantitate aluminum in this case is to add known amounts of aluminum to the sample ("spikes" of increasing concentration) and analyze the unspiked sample with the spiked aliquots. A moderate dilution would also be appropriate to reduce the severe viscosity, but to preserve sensitivity and detection limits, perhaps a 1:10 dilution. A calibration curve is created from these solutions using Graphite Furnace AA, as shown in Figure 17.

In this calibration technique, "zero" concentration is the unspiked sample with Al present (a positive signal), but its concentration is unknown. It could be thought of as "Zero ADDED Al". The spiked sample containing the unknown amount plus an added 1 ppb is "1" on the calibration scale, and the unknown amount plus an added 2 ppb is "2" on the calibration scale.
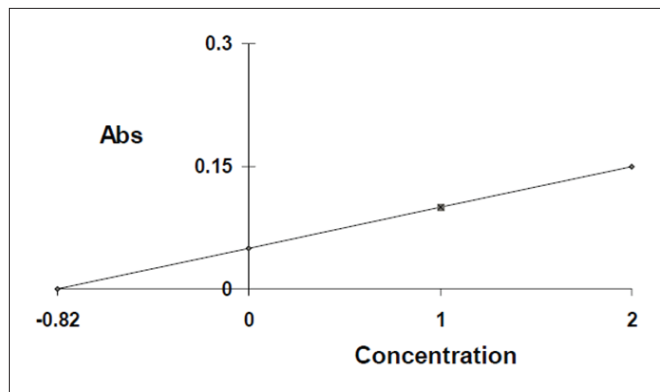

Figure 17. Method of Standard Additions (MSA) calibration curve for Al in maple syrup using AA.

A linear regression is calculated and extrapolated back to the concentration axis. The absolute value of the numeric intercept (Figure 17, -0.82 ppb) is the value of the unknown (0.82 ppb).

The advantage of MSA is that any matrix effects affect both the original sample and the spiked additions in the same way, and good accuracy in difficult matrices is more achievable.

The disadvantage of MSA is the need to prepare multiple aliquots of the same sample and make three or four measurements to get one answer.

A modification of the MSA technique can be used if several or many samples of the exact same matrix are to be analyzed, such as the analysis of urine. One sample can be spiked with the additions, the calibration curve created and the remaining samples just analyzed as unknowns against the calibration. This is often known as "Method of Additions Calibration". The overhead of multiple preparations is therefore limited to one sample.

Effects of Calibration Approaches on Detection Limits

In the Detection Limit section above, the effects of sensitivity, noise, background and stability were discussed. Another important parameter is the effect of calibration on detection limits.

If a calibration curve is poorly fit at the low end, nonsensical values for detection limit calculations can result. In the example from Figure 11, a low level sample of 0.1 ppb calculates to 4.002 ppb. Even though the calculated detection limit based on 3σ of the noise of the blank might be 0.05 ppb, the ability to actually read anything near that level is impossible if contamination or poor curve fitting are present. Sometimes a very low standard will even calculate to a negative value, which also makes the estimated detection limit worthless.

It can be stated that detection limits are dependent on valid calibration, in addition to sensitivity, noise, background and stability.

Summary

Achieving accurate results and getting the lowest possible detection limits are dependent on careful attention to many details. Proper preparation of solutions, avoidance of contamination, selection of optimum calibration ranges for the application and proper selection of calibration scheme are all critical.

Not addressed in this discussion, but of crucial importance, is proper maintenance of sample introduction components and optimization of the instrument. These aspects are the foundation of accurate, precise measurements. Sample collection and preparation are, of course, the gateway to the actual measurements in the lab. If contamination or analyte loss are not controlled at the point of sampling and preparation, all subsequent activities are compromised.

It is said that "the devil is in the details"; this can be a good guiding principle for this discussion. Every variable is important and every detail has an impact on the total quality of data.

References

1. Coleman, D., Vanatta, L. American Laboratory, parts 1-46 (starting in 2002)

2. U.S. EPA Method 40CFR136, Appendix B, https://nepis.epa.gov/Exe/ZyNET.exe/600009W7.TXT?ZyActionD=ZyDocument&Client=EPA&Index=2006+Thru+2010&Docs=&Query=&Time=&EndTime=&SearchMethod=1&TocRestrict=n&Toc=&TocEntry=&QField=&QFieldYear=&QFieldMonth=&QFieldDay=&IntQFieldOp=0&ExtQFieldOp=0&XmlQuery=&File=D%3A%5Czyfiles%5CIndex%20Data%5C06thru10%5CTxt%5C00000000%5C600009W7.txt&User=ANONYMOUS&Password=anonymous&SortMethod=h%7C-&MaximumDocuments=1&FuzzyDegree=0&ImageQuality=r75g8/r75g8/x150y150g16/i425&Display=hpfr&DefSeekPage=x&SearchBack=ZyActionL&Back=ZyActionS&BackDesc=Results%20page&MaximumPages=1&ZyEntry=1&SeekPage=x&ZyPURL

3. ISO Guide to the Expression of Uncertainty Measurement, 2008

4. United States Patent 4,238,830, 1980

For a complete listing of our global offices, visit www.perkinelmer.com/ContactUs