

IMPORTANT FACTORS THAT INFLUENCE THE DETERMINATION OF DETECTION LIMITS

By
Heinz Biermann

July, 1989

Environmental Hazards Assessment Program



STATE OF CALIFORNIA
Department of Food and Agriculture
Division of Pest Management, Environmental Protection and Worker Safety
Environmental Monitoring and Pest Management Branch
1220 N Street, Sacramento, California 95814

EH 89-8

**IMPORTANT FACTORS THAT INFLUENCE THE
DETERMINATION OF DETECTION LIMITS**

BY

HEINZ BIERMANN

JULY, 1989

ENVIRONMENTAL HAZARDS ASSESSMENT PROGRAM

TABLE OF CONTENTS

	PAGE
Table of Contents.....	i
List of Figures.....	ii
List of Tables.....	ii
Introduction.....	1
Specificity of Detection Methods.....	2
Definition of Detection Limit.....	4
Background and Background Noise.....	8
Peak Height.....	10
Calibration Curve.....	15
Linear Regression.....	18
Detection Limit.....	20
Limit of Quantitation.....	22
Sensitivity and Detection Limit.....	24
Signal-to-noise Ratio.....	26
Conclusion.....	28
Glossary.....	30
References.....	31

LIST OF TABLES

	PAGE
Table 1. Limits of detection for various confidence limits and baseline correction methods given as multiples of the standard deviation for the noise in the instrument background.....	32
Table 2. Limits of quantitation for various confidence limits, relative errors and baseline correction methods given as multiples of the standard deviation for the noise in the instrument background.....	33

LIST OF FIGURES

Figure 1. Demonstration of signal-to-noise ratios.....	34
Figure 2. Simulated dissipation curve.....	35
Figure 3. Normalized distributions for background measurements (r_b) with a detection limit at $1.65 \sigma_b$ (95%) level...	36
Figure 4. Dependence of noise and relative error on signal strength.....	37
Figure 5. Practical examples for the theoretical dependence of the relative error on the signal strength shown in Figure 4.....	38
Figure 6. Effect of increased method sensitivity on the detection limit.....	39
Figure 7. Effect of smoothing on the peak shape.....	40

IMPORTANT FACTORS THAT INFLUENCE THE DETERMINATION OF DETECTION LIMITS

INTRODUCTION

A goal in selecting a detection method is to reliably measure as small a quantity as necessary for the intended analysis. Reliability in this context refers to accuracy and precision of the method. Both of these have to be high enough based on the objective for the measurement. A quick screening test, for example, needs less stringent criteria than a precise quantitative analysis. Reliability also refers to the absence of any false signals. It is of no use that a method can detect minute amounts of a compound if there are interferences from other chemicals that cannot be distinguished from the compound of interest. Thus the initial statement contains another, quite different goal of equal importance: specificity.

Detection limit and specificity should be treated separately. A detection limit is an exact, quantitative measure, whereas specificity is a more qualitative expression describing nonrandom events. Let me insert here that there are a number of definitions for detection limits, but each of them is well defined as a quantitative measure. A statistical analysis for the detection limit is possible because only random fluctuations are considered in the calculation. As a consequence, a detection limit becomes meaningless if there are nonrandom sources of errors.

One of the most frustrating sources of nonrandom distortions are interferences by other compounds. If an interference by another chemical is present, there is no way to tell if some or all of a signal is due to the

interference. A second analysis using a different method has to be done in order to determine the amount of the interfering compound and make the necessary correction in the first analysis.

The specificity of a method has to be ensured independently from the detection limit. Because it is an important aspect of a chemical analysis method, I will begin with some comments about the specificity of detection methods.

Specificity of Detection Methods

A detection method is based on a physical process that is quantified by a law of nature. One example is infrared absorption which is based on the interaction of light with the vibration of atoms in a molecule. Another example is the interaction of a charged molecule with a conducting medium.

Infrared absorption is a good example for a very specific method because each molecule that has a unique geometrical arrangement of atoms has a unique infrared spectrum. The second example, a charge transfer, is on the opposite end of the specificity scale. This method cannot distinguish between an H-atom of mass 1 or a biochemical molecule of mass 100,000. All it can do is count charges. Why would anybody choose a completely nonspecific detection method? One condition could be that the method is so much more efficient that a much smaller quantity can be detected. But as long as you know that you are measuring just the compound of interest, it doesn't matter if the method is specific.

The important difference is that if for some reason you're using the wrong compound for the measurement, the specific method will tell you that you've used the wrong compound, whereas the nonspecific method will just yield a result that does not indicate that an error occurred. The best you can hope for is that the value of the result is so unexpected that it gives a clue that something is not correct.

The next level of complication comes in when mixtures of compounds are considered. Whereas methods that are highly specific can sometimes detect compounds unambiguously even in complicated mixtures, nonspecific methods rely completely on pre-separation of the components before analysis. One of the more popular separation techniques is gas-chromatography (GC).

So GC, like all other chromatographic processes, is not a detection method, it just physically separates the components in a mixture. And because a whole variety of compounds will elude from a GC in sequence, a very nonspecific detector is often needed to register them all.

When GC is used with a nonspecific detector (like flame ionization or nitrogen-phosphorus detectors), an identification of the compound is based solely on the retention time. There is a problem associated with the use of retention times, however. For a given set of operating conditions, every compound has a specific retention time. But this retention time is not unique to that compound: there are many compounds that would elude at the same time if they were in the mixture. So GC can prove the absence of a compound but not its presence. If a peak is detected at the retention time characteristic for a certain compound, and there is independent reason to

believe that this compound is present in the sample, then the probability that the peak originated from that compound is high, but it is not a proven fact.

How does this problem of lack of specificity tie in with the detection limit? The ability to distinguish a signal from noise is not affected directly by interfering compounds. The detection limit only indicates that level at which an instrument response can be assumed to be real; it does not guarantee or imply that what has been measured is actually the intended information. Thus a lack in specificity cannot be compensated by a modification of the detection limit. So let's define what a detection limit is and, equally important, what it is not.

Definition of Detection Limit

The detection limit is defined as a threshold that is used to distinguish between random noise and wanted information. If a signal is measured that exceeds this threshold, it is assumed that it did not originate from noise.

There are a number of 'detection limits' using different formulas and parameters. Therefore, one should always explicitly state the basis for the calculation. The selection of these parameters is somewhat arbitrary, but once they are chosen, they have to be calculated from the available data. Estimating the values of these parameters prohibits the use of the detection limit as a quality control measure.

My approach is that the detection limit should cover the same steps that are considered in the calculation of the concentration. It is my understanding

that the standard operating procedure in this Department is to refer analysis results to the final extract used in the analysis; extraction efficiencies and uncertainties are not considered. It follows that the detection limit should not include these factors either.

This is in contrast to, for example, the recommendation of the analytical methods committee of the Royal Society of Chemistry (Analytical Methods Committee, 1987). Their basic definition seems to be quite standard:

"It is recommended that the detection limit of an analytical system be defined as the concentration (c_L) or amount (q_L) corresponding to a measurement level $3\sigma_B$ units above the value for zero analyte. The quantity σ_B is the standard deviation of responses of the field blanks.

It is further recommended that other usages, definitions and named limits be discontinued."

But I disagree with their use of the term analytical system:

"Analysts often refer to the performance characteristics (e.g., the precision) of an analytical method. This usage is misleading because 'the method' per se contributes only part of the total variation that is observed during its use. If the method is used at a number of separate locations and on different occasions, other factors contribute to (and usually dominate) the total variation. Such additional factors

include the analyst, the environment, the brand of instrument used, the quality of the reagents, the nature of the samples, the protocols used for calibration and reagent blank correction, etc. It is these additional factors that, with the method, constitute the 'analytical system.'

Using this definition of an analytical system in the definition of the detection limit will degrade this limit to a least common denominator. The worst possible combination of factors will set the detection limit for all other cases. I fail to see why, for example, a laboratory that can measure a compound very reliably should quote a higher detection limit just because another lab has higher variability in its procedure. The above definition also asks for averaging over brands of instruments; I advocate to use separate limits if different brands of GC, for example, perform differently.

I prefer to calculate a detection limit just from the noise level in the instrument output because ultimately a positive detection depends on being able to distinguish a signal from noise. For this it is of no consequence what errors occurred before this measurement. It doesn't matter, for example, if the extraction efficiency drops from 100% to 50%; the amount of chemical needed for the instrument to produce a measurable signal stays the same. However, if one wants to relate this (constant) minimum detectable signal to a concentration in the original sample, one has to take into account all possible errors and uncertainties in the intermediate steps. As this Department does not take extraction efficiencies into account, relating the reported concentrations to the final extract, the noise fluctuations in

the instrument response are all that is needed to establish a detection limit.

Let me repeat here that measuring an instrument response that is above the detection limit only implies that this response most likely was not caused by random noise. This 'real' signal could still originate from a noise spike or an interfering compound, and one has to use other criteria than the detection limit to make sure that this signal is due to the compound of interest. Unfortunately, this distinction is not made clear in textbooks or articles about the detection limit, and so I am rephrasing this important point one more time.

People often refer to a detection limit as: "The detection limit for compound X is a concentration of y units." This is convenient and short, but contains an unspoken assumption. The proper way to express the detection limit would be: "With 99% probability the instrument can distinguish a signal of z Volts from random noise. If this signal is caused by compound X, this equates to a concentration of y units." It is easy to understand why everybody just uses the short statement. But it omits a very important 'if'. The detection limit only measures the ability of an instrument to separate a signal from noise. It has absolutely nothing to do with the assignment of the signal to a certain chemical. The latter is a question of specificity and has to be answered by different means.

The detection limit is given in concentration terms only because the final analysis result is expressed as a concentration. Again, the fluctuation in the instrument background is the only quantity needed to establish a

detection limit (if one wants to express this limit in concentration terms, one has to establish the calibration curve, too). So let's talk about background and background noise next.

Background and Background Noise

Let me first define the terms signal, noise, background and blank. The part of an instrument response that contains any useful information is called signal. Any unwanted fluctuations in the instrument response are noise. When the average noise level in absence of an input concentration is nonzero, it is called a background. These definitions are based on what is considered to be the useful (or wanted) information.

For instance, if one wants to determine the mean and standard deviation of the random fluctuations, then what was considered to be noise before becomes wanted information and thus the signal. And any change in the response due to the detection of a chemical is now noise because it causes unwanted changes in the mean of the random fluctuations. So what is signal and what is noise is always relative to what one wants to measure.

A blank is defined quite differently. It is the result of an analysis when the input concentration is zero. In practice one should distinguish between laboratory blanks and field blanks. The former tend to have a smaller variability which may underestimate the standard deviation of the noise level.

For direct reading instruments, background and blank are identical, but for methods like GC they are very different. Because in GC analysis the signal

is derived as the difference between the instrument response at a given time and the interpolated value of the background at this time, the blank should be about zero no matter how large the background.

It is important to distinguish between background and blank because the formulas for the calculation of the detection limit are different. To distinguish between the two, the subscript b will be used to refer to background measurements, the subscript B to blank measurements (see the glossary at the end of this text).

The first step in establishing a detection limit is to determine what the smallest response (r_L) is that can be distinguished from a blank response (r_B) with reasonable certainty:

$$r_L = \bar{r}_B + k\sigma_B \quad (1)$$

where \bar{r}_B is the mean, σ_B the standard deviation of the blank measurements and k is a constant.

Often this standard deviation is taken from the distribution of the field blanks (as in the recommendation of the Royal Society of Chemistry), but sometimes this is not adviseable. For example, one cannot determine the standard deviation of the blanks from the output of an integrator. Integrators used in chromatographic analysis try to detect a peak by monitoring the slope of the instrument response. When the slope rises above a certain threshold, it starts integrating. This threshold is set high enough to respond only to 'real' peaks in order to avoid that the integrator

reacts to any random fluctuation. This means that an integrator will always return a value of exactly zero for the blanks, unless there is a contamination or interference problem. In other words, if zero is not within the confidence limit for the intercept of the calibration curve, it indicates that there is a problem with the chemical analysis.

In effect, the integrator has its own built-in detection limit and returns a value of zero for any value below this limit. It is thus impossible to use data processed by an integrator for statistical analysis when the data are close to or below the integrator threshold. Because of this inability, some people determine a detection limit from spiked blanks.

If the variability of the field blanks cannot be determined properly, the noise in the instrument background can be used as a measure. In any case, I prefer the latter method as it takes the special circumstances of each measurement into effect. When the detection limit is based on the variability in the field blanks determined at some point in time, the detection limit would give an erroneous impression of the quality of the data if the instrument noise changes afterwards. Let's now look at how this measure can be used to distinguish between signal and noise.

Peak Height

Often readings are made from an analog recording of a continuous output from an instrument. Because these results are usually stored as a trace on chart paper, I am using some simulated signal traces as examples. Figure 1 shows three such noisy, flat baselines with two peaks each: a narrow peak is centered at $x = 75$ and a wide one at $x = 200$. The plot on the right contains

the same data as the one on the left, just the y-axis is magnified ten times. The peak heights are equivalent to 2σ , 3σ and 6σ values of the background noise. It does not matter what the actual values for the background or peak height are, all I want to convey is a feeling for the signal-to-noise ratios. For instance, if the detection limit is set to 3σ above the background noise, the middle trace shows what you're asking the analyst to measure.

As mentioned before, the noise level in the background does not have to be determined from a separate blank; it can come from an interference-free part of an instrument response containing peaks at other locations. For example, in Figure 1b the noise level could be determined from the flat part of the traces around $x = 100 - 150$. The traces in Figure 1a are useless for this purpose because of the low magnification. Determining the noise level in the same output as the signal is generally preferable because it can take matrix effects into account, and it optimizes the detection limit for each measurement.

The following arguments apply to the determination of both peak heights and areas, but the formulas for peak heights are simpler. There are four different cases to consider:

The peak height h can be equal to the instrument response. This is the case of a direct reading instrument, where

$$h = r_c \quad (2)$$

and r_c denotes the instrument response for an input concentration c . For concentrations near the detection limit one can assume that the variance of the concentration measurement is equal to the variance of the background measurement:

$$v_c = v_b \quad \text{for} \quad r_c \approx r_b \quad (3)$$

In this case, the variance and standard deviation for the peak height are:

$$v_h = v_b \quad \text{and} \quad \sigma_h = q_b \quad (4)$$

If the peak height is determined by subtracting a flat background, the corresponding formula is:

$$h = r_c - r_b \quad (5)$$

with a variance and standard deviation of:

$$v_h = 2 * v_b \quad \text{and} \quad \sigma_h = 1.41 * q_b \quad (6)$$

When the background has a nonzero slope, these formulas have to be modified to:

$$h = r_c - r_{b1} + 0.5 * (r_{b2} - r_{b1}) \quad (7)$$

and

$$v_h = 3.5 * v_b \quad \text{and} \quad \sigma_h = 1.87 * q_b \quad (8)$$

This assumes that the measurement for the peak maximum (r_c) is centered between the two background measurements (r_{b1} and r_{b2}).

Thus the standard deviation of the background noise has to be multiplied by a factor of 1.0, 1.41 and 1.87 when the signal is calculated from one, two or three determinations of the instrument response, respectively.

The last case for the nonzero slope still implies a linear baseline. Often, however, the baseline is curved. When no sophisticated data analysis procedures are available, the nonlinear background is generally approximated by a linear interpolation at the base of the peak. This introduces a bias in the peak measurement which can lead to a significant over- or underestimation. When the curved background is approximated by a nonlinear fitting function (polynomial or exponential, for example), a calculation of the error propagation can become quite complicated and would lead to an even higher multiplication factor for the background noise.

If one wants to get statistical information about the background noise level in an instrument response, it is necessary to use discrete sampling instead of an analog recording. Because a peak is measured as a difference in the instrument response before, after and while a compound eludes, it is necessary to determine the baseline fluctuations that occur in the same timeframe as the detection of this compound (i.e., a few seconds at most). In order to get accurate measurements on this timescale, it is best to use a computer for the data acquisition. If one does not use a computer but evaluates the data from a trace on a stripchart recorder, one cannot determine σ_b accurately. Then the peak-to-peak value or the root mean square of the noise has to be used to set the detection limit at a specific signal-to-noise (S/N) ratio; commonly, the detection limit is then set to $S/N = 2$ (Willard et al., 1981). The only problem with the use of a visually determined S/N ratio is that it cannot be related to a certain probability level.

Note that this argument and all further calculations are based on the assumption that the sample to be analyzed is homogeneous. If this is not true, as may be the case with soil or even liquid samples, different aliquots would actually come from different distributions with different mean concentrations. This type of variability has nothing to do with the detection limit itself. Measurements like this should never be used to establish calibration curves. The resulting uncertainty in the regression may no longer reflect the ability of the method to determine a given concentration.

Another important point is the measurement of negative peak heights. Knowing that negative numbers have no physical meaning, a negative result is often reported as zero. But this practice causes distortions in the calculations of mean and variance of the blanks. When peak heights are calculated from a difference, as in Equ. 5, the instrument response r_c for $c = 0$ is expected to be equal to r_b . The difference between two measurements of r_b will be negative half of the time assuming that the instrument response is normally distributed around a mean value of \bar{r}_b . So even though a negative peak height has no physical meaning, it is a very real result that has to be reported as such.

This implies that the laboratories would be required to report the raw data. A 'not detected' result is not a raw data item. It is a decision made after comparing the analysis result with the noise level (i.e., the detection

limit). For proper statistical treatment of the data, however, the raw data are essential.

After talking about background noise and peak heights, let's take a look at the second ingredient needed for a detection limit, the calibration curve.

Calibration Curve

Three factors influence the decision about how to establish a calibration curve: the spread of the data, the error of the measurement and the model used to fit the data.

There are three ranges that have to be matched: the concentration range of the actual samples, the measurement range of the instrument and the range where the fitting function is valid. It makes a difference to the design of calibration measurements if it is known independently that the model is applicable. Once it has been shown that the model is appropriate, the number of calibration points can be cut down to a minimum. The minimum number of points is equal to the number of parameters in the fitting function (i.e., two for a straight line $[y = a_1 + a_2 * x]$, three for an exponential $[y = a_1 + a_2 * \exp(a_3 * x)]$). This minimum number is generally used only for calibration checks. The original calibration curve should contain more concentration levels to make sure that the model fits.

But even if the model fits the points very well, it is no proof that this is the 'right' model which describes all possible intermediate values. For example, calibrations that involve complicated reaction schemes may yield an S-shaped curve. In these cases even a three point calibration is not good

enough to distinguish between S-shape and linear relationship. If the three points are not on a straight line, a linear relationship can be ruled out. However, if the points are on a straight line, it does not follow that an S-shaped curve is ruled out. On this S-shaped curve one can find an infinite number of sets of three points that each lay on a straight line. Thus a straight line may fit those calibration points very well, but will yield wrong results for intermediate values.

As a consequence, the number of points needed for a calibration curve depends not only on the selected fitting function but also on the models that have to be disproved. If one wants to distinguish between a linear relationship and an n -th order polynomial, for example, one has to make at least $n+1$ different measurements.

The magnitude and distribution of the errors associated with a measurement determine if it is advantageous to make replicate measurements. When the errors are nearly constant throughout the whole range, more useful information is collected by selecting a larger number of different concentrations instead of a few levels with replications. Given that only a limited number of measurements is allocated towards establishing a calibration curve, I favor a diversity in concentration levels over a large number of replications. For a calibration curve a weighted regression usually does not have much effect on the parameter values; it mainly reduces the confidence limits. A weighted regression is valuable only when the absolute errors associated with the measurements are distinctly different between the ends of the measurement range.

The optimum strategy would be to select first one concentration each near the lower and upper limits and perform a few replicate analyses. If the standard deviations at these two levels differ by less than a factor of two, it is not worth while to make replications for the intermediate values.

One assumption made here is that the calibration data have a small coefficient of variation. If it should turn out that there is a large uncertainty associated with the calibration data, I suggest that the whole analysis procedure be discarded. A calibration curve with high precision doesn't have to be right, but a calibration with low precision is definitely not worth much. The calibration curve is usually very precise because the sample preparation steps are often not included in the calibration procedure and thus no matrix effects occur.

Calibration checks should be performed regularly, and the whole calibration procedure has to be repeated when the check yields a result outside the confidence limit for the original regression line. A calibration check for a straight line requires two points. When it is known that the original calibration curve passes through zero, one calibration point is sometimes replaced by a blank. This combination of a one point calibration check and a blank measurement produces satisfactory results most of the time but cannot catch all errors. A better procedure is to select one point in the lower third and one in the upper third of the calibration line.

A linear regression procedure is used most of the time to calculate the calibration parameters from the original data points. Therefore, I insert

here some remarks about linear regression before talking about the detection limit.

Linear Regression

When a linear regression procedure is used to fit a straight line of the form $y = a_1 + a_2 * x$ to the data points, the main information gained in this procedure is the intercept a_1 , the slope a_2 , a measure for the goodness of the fit and sometimes the errors for intercept and slope. But there are additional, useful parameters derivable from the regression calculation.

One additional piece of information is the confidence limit for the regression line. This information is used best in a graphical representation where it provides an immediate impression of the uncertainty in the position of the calibration line.

From an applied standpoint, it is more useful to give the confidence limits for a concentration calculated with this regression line than to present the confidence limits for the line itself. The formula for the uncertainty in the calculation of a concentration x from a measured signal y allows us to attach a confidence interval to any single concentration result obtained using the calibration curve. This interval reflects only the uncertainty from the limited precision of the calibration procedure, it does not cover variabilities introduced during the sample preparation steps (unless they are included in the calibration) (N.R. Draper and H. Smith, 1981; J.C. Miller and J.N. Miller, 1984).

Another factor is the variation of the signal around the regression line ($\sigma_{y/x}$). A basic assumption of the unweighted least-squares fit is that each point (including the blank) has a normally distributed variation in the y-direction with a standard deviation estimated by $\sigma_{y/x}$. Thus the standard deviation for the background measurement σ_b can be estimated from $\sigma_{y/x}$ when an unweighted regression line has been established. A modified formula can be used to calculate $\sigma_{y/x}$ for a weighted regression.

Often 'unweighted' is understood to mean that each point is of equal importance for the regression curve. But all it implies is that each point is considered equally in respect to any possible errors in the measurement. In other respects, a linear regression by least-squares treats the data points very unequally. Take, for example, a dissipation curve where the important quantity is the slope. Let's assume that for one of the points in a data set the y-value is changed by a certain fraction. It can then be derived that the error in the slope caused by this change is proportional to the distance of the x-value from the mean of the x-values and to the y-value at this location.

This relationship is shown in Figure 2. The top trace is the regression line with unequally spaced x-values yielding an off-center mean at 45. The bottom trace shows the absolute value of the relative error introduced when the y-value at that particular x is changed by a factor of two. The line is curved because the relative error depends also on the y-value itself, which approaches zero for the larger x-values. If the error introduced had been

assumed to be a constant, independent of the y-value, the result would have been two straight V-shaped lines.

As a consequence, smaller errors at the extreme ends of the range have a more pronounced effect on the slope than a very large error near the middle. That is why a calibration curve should not extend too far beyond the range of interest. A small error at a high concentration level can change the slope noticeably in the region of interest.

Detection Limit

Once the standard deviation σ_b of the background measurement m_b has been established, the limit of detection depends on the confidence level chosen and the background subtraction method used (see page 12). Confidence levels have to be assigned to both errors of the first and second kind. An error of the first kind (type I) is the error of accepting a signal as real when it actually is part of the background distribution. An error of the second kind (type II) is the error of rejecting a real value as if it were part of the background.

The relationship between type I and type II errors is illustrated in Figure 3. The left distribution in Figure 3a represents the distribution of the background noise. A 95% confidence level corresponds to a limit of detection at $1.65 \sigma_b$. Superimposed in Figure 3a is the distribution of concentration measurements whose mean is equal to this detection limit. The assumption made here is that the variance v_c for a concentration measurement r_c near the detection limit is equal to the variance v_b of the background

measurement r_b (Equ. 3). Half of the time a measurement r_c will be below the average (the hatched area in Figure 3a) and will thus be attributed to the background noise distribution because $r_c < \bar{r}_b + 1.65 \sigma_b$. In other words, this real signal will not be detected half of the time, making the type II error level 50%.

Figure 3b shows the distribution of the background noise values and the distribution for a concentration of $\bar{r}_c = \bar{r}_b + 3.29 \sigma_b$, where both types of errors are reduced to 5% levels. As indicated by the hatched area, only 5% of the time will the measurement be below the detection limit of $1.65 \sigma_b$.

Table 1 lists limits of detection at various combinations of confidence levels for type I and type II errors. In GC analysis, baselines are almost never flat, so a proper factor for a detection limit at the 95% level is 3.08, if type II errors are of no concern.

Let me try to make the importance of type II errors clearer using an example. Assume that a certain chemical poses a serious health risk with a warning threshold set to 4 ppb. It is considered necessary that an analytical method should yield a positive result for an ambient concentration of 4 ppb with 99% probability. A simple method for monitoring this compound is available that has a zero background and a standard deviation of the noise equivalent to 1 ppb. The detection limit based on a 99% confidence interval is then 2.3 ppb. So one can measure 4 ppb with much

better than 99% probability. But it is wrong to conclude that the method can then be used to measure this chemical at the 4 ppb level.

There is a big difference between saying that a measured level of 4 ppb is a positive result and saying that an ambient level of 4 ppb should give a positive result. The crucial difference is that the measurement of the ambient concentration has an error associated with it, too. So a single measurement of the actual 4 ppb level may yield an analysis result of 3 ppb or 5 ppb. More precisely, a measurement of the 4 ppb level will 99% of the time yield a result above 1.7 ppb. This lower limit is below the detection limit! Thus the method is not able to detect a 'true' concentration of 4 ppb in a single measurement with 99% probability, and the method cannot be used for monitoring this chemical. This fact is reflected in a detection limit of 4.7 ppb for 99% confidence levels for both error types (see Table 1).

Limit of Quantitation

When a signal has been recorded just above the detection limit, it can be said that this signal did not originate from noise, but the error in the magnitude of this signal is rather large. Therefore a second, higher limit is sometimes introduced: the limit of quantitation (LOQ) (J.C. Miller and J.N. Miller, 1984). It specifies at what level the relative uncertainty in the measurement drops to a given value. For example, if a 95% confidence level corresponds to $1.65 \sigma_b$, a measured concentration must be at least $16.5 \sigma_b$ above the background mean in order to get a 10% relative error for this measurement. Table 2 lists the limits of quantitation for three

confidence levels (90, 95 and 99%) and for three error levels (10, 20 and 30%). Based on a 95% confidence level and a 20% error, the limit of quantitation is $15.4 \sigma_b$ above the mean background for a sloped baseline.

Earlier, I have made the assumption that the standard deviation in the determination of a concentration near the detection limit is equal to the standard deviation of the background measurements (Equ. 3). This is generally not true for higher concentrations where it is assumed commonly that the error in a measurement is proportional to the measured value (see Figure 4a):

$$\sigma_c = \alpha * r_c \quad (9)$$

Considering that the peak height is determined from a difference (Equ. 5), the error in the peak height is given by:

$$v_h = v_c + v_b \quad \text{or} \quad \sigma_h = (\alpha^2 * r_c^2 + \sigma_b^2)^{1/2} \quad (10)$$

Figure 4b shows the variation of the relative error (σ_h/h) as a function of the peak height h . The specific values used in this example are $r_b = 5$, $\sigma_b = 0.5$ $\alpha = 0.1$. Marked are a LOD at $2.33 \sigma_b$ (equivalent to a peak height of 1.2) and a LOQ at $11.6 \sigma_b$ (equivalent to a peak height of 5.8). This figure shows that for large peaks the relative error becomes constant because the contribution of σ_b becomes negligible. As the peak height approaches zero, however, the relative error grows towards infinity. At the selected LOD, for instance, it is 66% (for this example).

To show that this is not just a theoretical point, I have included Figure 5, presenting two graphs taken from a paper about quality assurance during an air quality study (Fujita and Collins, 1989). It shows the variability in the analysis of hydrocarbons in ambient air between four laboratories and PAN between two laboratories. Even though these graphs do not represent the scatter of replicate measurements within a laboratory but rather the scatter between laboratories, the same functional relationship holds true: a constant coefficient of variation (CV) is observed at higher concentrations with the CV increasing noticeably for lower concentrations.

Sensitivity and Detection Limit

As mentioned earlier, the detection limit depends on the noise in the background and on the calibration curve. Thus one has two options trying to optimize the detection limit: One way is to reduce the noise level in the instrument output, directly lowering the detection limit. The other option is to increase the sensitivity of the analysis method.

Sensitivity is defined in analytical chemistry as the ratio of the change in the response of the instrument to the corresponding change in the concentration of the analyzed compound. The sensitivity of a method can be taken directly from the slope of the calibration curve, i.e., the slope is a measure of the sensitivity (J.C. Miller and J.N. Miller, 1984).

With the slope of the regression line being the sensitivity S , and the intercept being equal to the mean of the blanks \bar{r}_B , the regression function $y = a_1 + a_2 * x$ can be rewritten as

$$r = \bar{r}_B + S * c \quad \text{or} \quad c = \frac{r - \bar{r}_B}{S} \quad (11)$$

Based on Equ. 1, this relationship yields for the detection limit c_L :

$$c_L = \frac{r_L - \bar{r}_B}{S} = \frac{k\sigma_B}{S} \quad \text{or} \quad c_L = \frac{1}{k} \frac{\sigma_B}{S} \quad (12)$$

Thus the detection limit c_L is proportional to the noise and inversely proportional to the sensitivity of the method.

Figure 6 should clarify how an increase in the slope of the calibration curve decreases the detection limit. When only the slope is changed, the background measurement is not affected keeping the detection limit expressed as an instrument response at a constant level above this background. When this detection limit is converted into the corresponding concentration, however, the detection limit does change in proportion to the slope of the calibration curve.

The major assumption made here is that the change in the procedure that caused the increase in the slope does not affect the noise level. One always has to look at what effect a change in procedure has on both the signal and the noise. No matter how large the gain in signal may be, it is not worth it if the noise level increases at an equal or higher rate. Another way to express this is saying that one has to maximize the signal-to-noise ratio and not the signal.

Signal-to-noise Ratio

The signal-to-noise (S/N) ratio is defined as the average signal amplitude divided by the average noise amplitude. The higher the value of the S/N ratio the better the measurement. After acquiring the data the S/N ratio can only be improved by decreasing the value of the noise. Simple magnification of the signal has no effect on the S/N ratio.

Once the analysis procedure is set, the sensitivity of the method (and thus the signal strength) is fixed. There is, however, a way left to modify the noise level even after the data have been acquired and stored on magnetic media. This process is called smoothing. Be aware that this is quite a dangerous manipulation because it also changes the signal shape (P.R. Bevington, 1960).

Figure 7 shows the effect of smoothing on some stored data. The top trace represents the original data with approximately 14 samples taken across the full width at half maximum (FWHM) of the peak. The following traces were derived from the original by low-pass filtering using time constants equivalent to 2, 4, 8, 16 and 32 channels. I have chosen this technique as an example because it shows the same effect as a time constant selector on a strip chart recorder or integrator. So these samples represent also the case of increasing the time constant of the data acquisition system. The difference is that if one selects the wrong time constant for the stripchart recorder, the analysis has to be redone, while an improper manipulation of the stored data can always be undone (provided that the original data are kept unaltered).

Looking at Figure 7a, it is obvious that smoothing can dramatically reduce the background noise. However, it is also apparent that the peak shape becomes very distorted. A characteristic of those distorted peaks in the traces is that the maximum is shifted towards the right and that the falling slope is wider than the rising one.

Figure 7b provides a graphical presentation of the effect that smoothing has on both the signal and the noise, making it easier to visualize the change in the S/N ratio. The graph shows that in this example the optimum operating conditions are achieved for time constants between 4 and 8. Under these conditions the ratio of peak height to detection limit is a maximum. What is true for the data acquisition, is also important in all other steps of the analysis procedure: the signal-to-noise ratio is more important for the quality of the data than the absolute magnitude of the signal.

For any critical measurements near the detection limit, I consider it to be most advantageous to store the original digitized instrument output with high temporal resolution on diskettes. Then the optimum data treatment can be selected and applied to the data afterwards. When only peak height or area are stored no optimization, correction or statistical analysis of the associated noise can be made any more.

The increased flexibility afforded by the off-line analysis is important because there is no standard procedure that yields optimum performance in all cases. There are a number of methods to determine baselines, peak positions, peak heights and areas, as well as methods to minimize interference from noise spikes or overlapping peaks. Being able to look at

the raw data and then to select the optimum combination of data analysis procedures, can make all the difference between a reasonable result and just plain garbage.

The disadvantage is that this type of data analysis takes more time and requires quite a bit of operator skill. But peaks with a signal-to-noise ratio of less than 10 are not suitable for routine analysis anyhow, as an analysis procedure that yields good results for large peaks may produce biased data for small ones. Storing the raw data on disk, one can still evaluate them with a 'black-box' algorithm analogous to an integrator for routine analysis. But there is always the option to go back and reevaluate selected traces for optimum performance.

CONCLUSION

I have tried to provide some information about what criteria influence the determination of a detection limit. I did not provide a standard recipe for the calculation of a detection limit because I don't think there is one. Almost everywhere in the literature one will find detection limits set to three standard deviations above the noise. The problem is to find a proper measure for 'noise' and the distribution from which the standard deviation is taken. How to determine these parameters will depend largely on the chemical analysis method (i.e., what detection principle is used), the data acquisition method (single readout, stripchart or digitized data, for example) and the capabilities of the data analysis procedure.

The important point is that for each analytical problem there is a rational way to calculate a detection limit. This limit allows the person that

interprets the final results to see how likely the reported data could have arisen from random noise, and how the uncertainty in the measurement changes over time. But always keep in mind that the detection limit is not designed to cover systematic errors like interferences from other compounds.

GLOSSARY

FWHM	full width at half maximum
LOD	limit of detection
LOQ	limit of quantitation
S/N	signal-to-noise (ratio)
c	concentration
h	calculated peak height
r	instrument response
S	sensitivity of analysis method
v	variance
σ	standard deviation

Subscripts:

B	blank measurement
b	instrument background
c	concentration
h	peak height
L	detection limit

REFERENCES

- Analytical Methods Committee, "Recommendations for the definition, estimation and use of the detection limit," *Analyst*, 112, 199-204 (1987).
- Bevington, Philip R., "Data reduction and error analysis for the physical sciences," McGraw-Hill Book Company, New York, N.Y. (1969).
- Draper, Norman R., and Harry Smith, "Applied Regression Analysis," Second Edition," Wiley Series in Probability and Mathematical Statistics, John Wilery & Sons, New York, N.Y. (1981).
- Fujita, Eric M. and John F. Collins, "Quality assurance for the Southern California Air Quality Study," *82nd Annual Meeting of the Air & Waste Management Association*, Anaheim, California, June 1989.
- Miller, J.C. and J.N. Miller, "Statistics for analytical chemistry", John Wiley & Sons, New York, N.Y. (1984).
- Willard, Hobart H., Lynne L. Merritt, Jr., John A. Dean and Frank A. Settle, Jr., "Instrumental methods of analysis," Wadsworth Publishing Company, Belmont, California (1981).

Table 1: Limits of detection for various confidence limits and baseline correction methods given as multiples of the standard deviation for the noise in the instrument background.

Confidence Limit		LOD		
Type I ^{a)}	Type II ^{b)}	Baseline Correction Method		
		None	Flat	Sloped
90%	50%	1.28	1.81	2.40
90%	90%	2.56	3.63	4.80
95%	50%	1.65	2.33	3.08
95%	95%	3.29	4.65	6.16
99%	50%	2.33	3.29	4.35
99%	99%	4.65	6.58	8.70

^{a)}Type I error: accepting a result as real when it actually arises from the background noise.

^{b)}Type II error: rejecting a real value as if it came from the background noise.

Table 2: Limits of quantitation for various confidence limits, relative errors and baseline correction methods given as multiples of the standard deviation for the noise in the instrument background.

Confidence Limit	Relative Error	LOQ Baseline Correction Method		
		None	Flat	Sloped
90%	30%	4.3	6.0	8.0
	20%	6.4	9.1	12.0
	10%	12.8	18.1	24.0
95%	30%	5.5	7.8	10.3
	20%	8.2	11.6	15.4
	10%	16.5	23.3	30.8
99%	30%	7.8	11.0	14.5
	20%	11.6	16.5	21.8
	10%	23.3	32.9	43.5

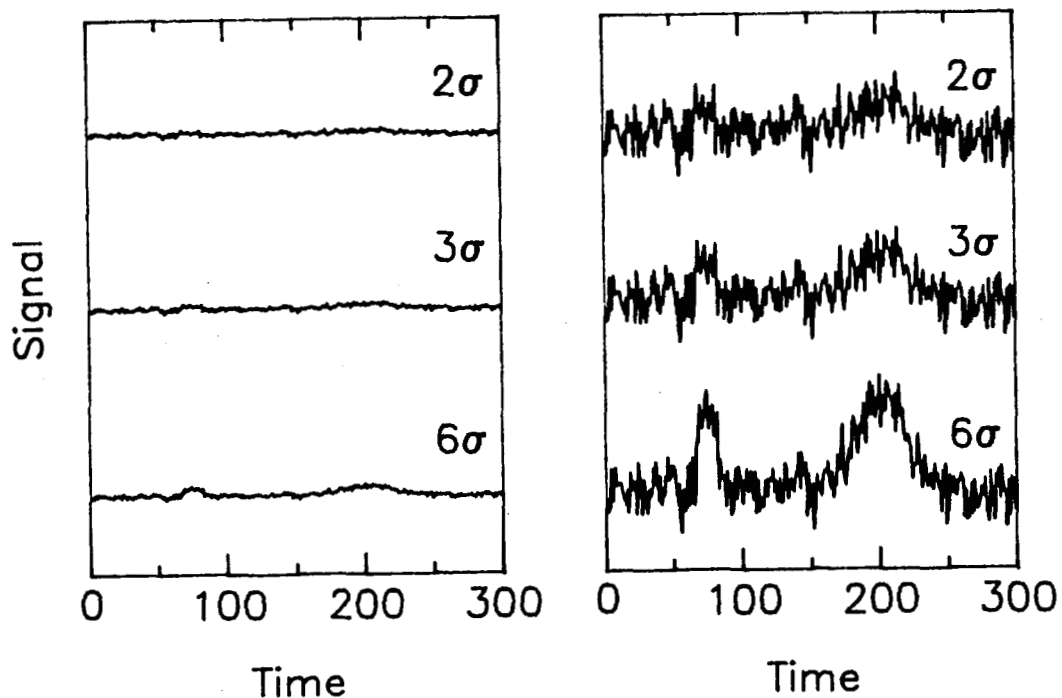


Figure 1: Demonstration of signal-to-noise ratios. Peaks with heights equivalent to 2σ , 3σ and 6σ of the background noise are superimposed on a flat, noisy baseline. A narrow peak is centered at $x=75$, a wide one at $x=200$. Both plots show the same data; the traces on the right are magnified 10 times.

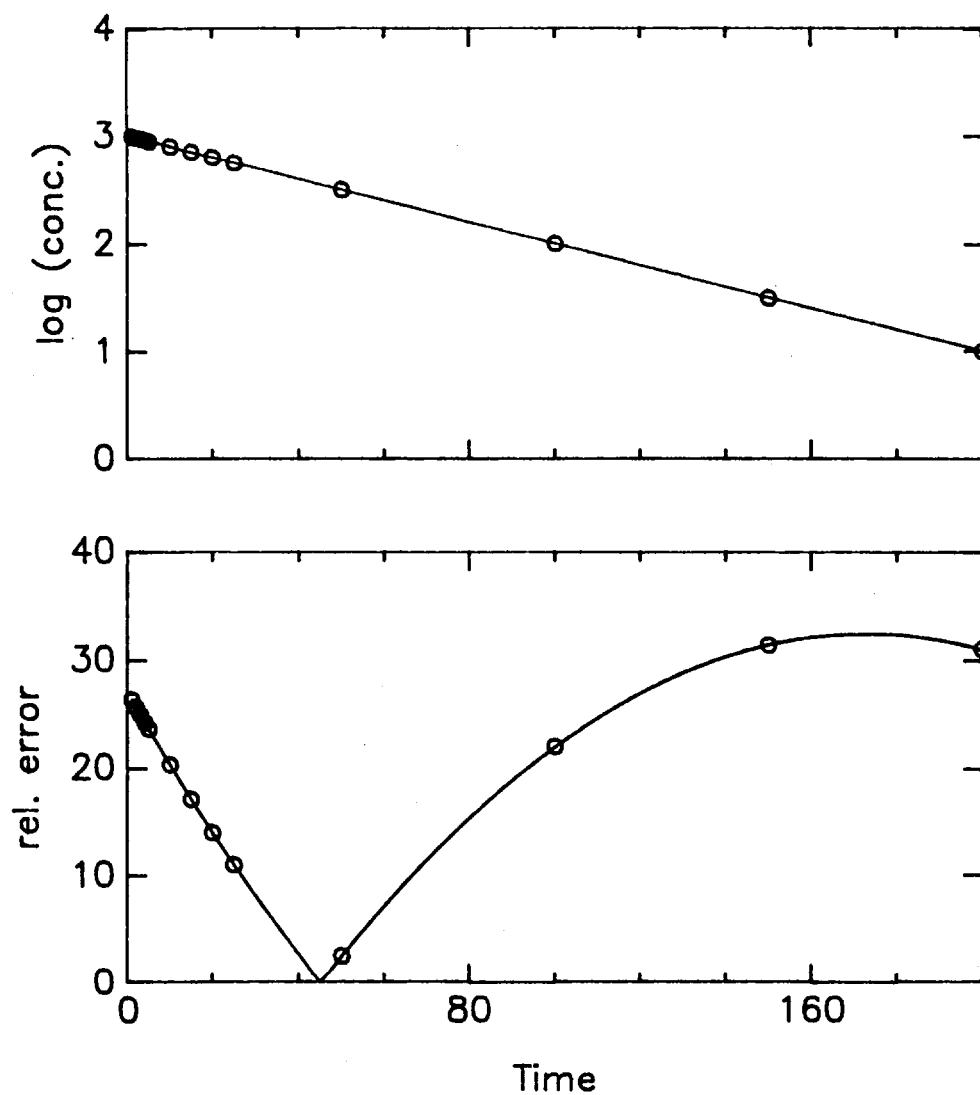


Figure 2: a) [top] Simulated dissipation curve.
 b) [bottom] Error introduced into the slope of the regression line shown in a) by changing the y-value at a single x-value by a factor of two, while keeping all other data points constant.

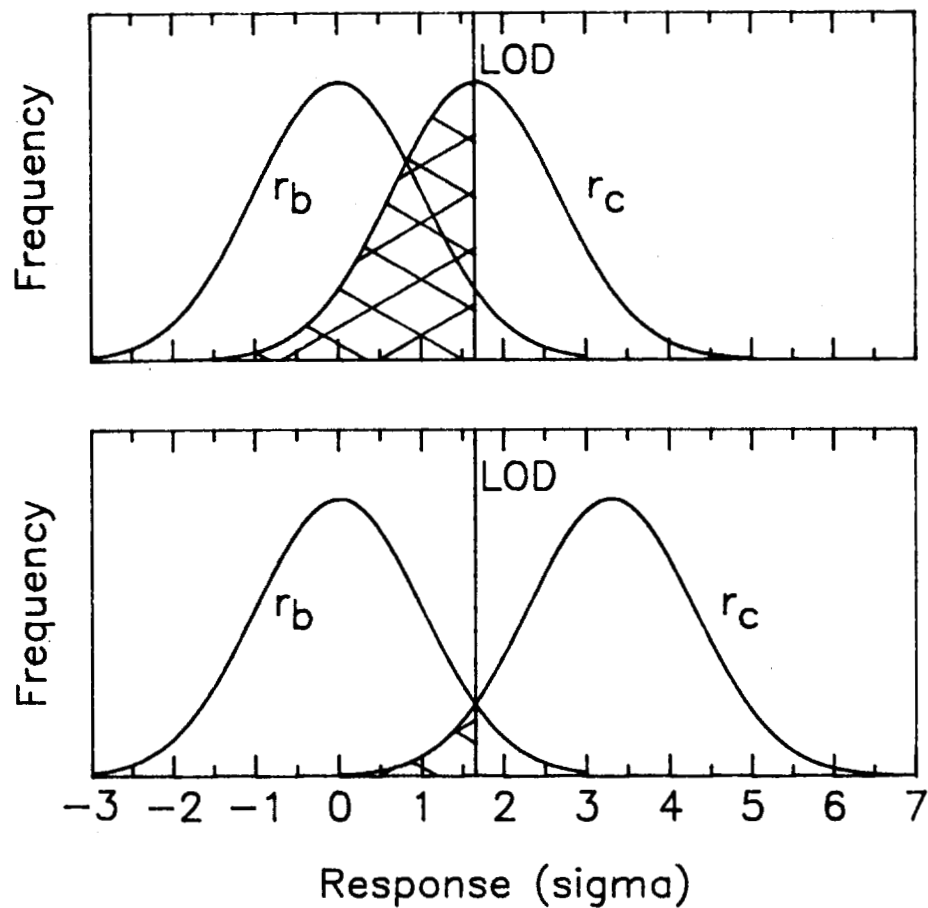


Figure 3: Normalized distributions for background measurements (r_b) with a detection limit at $1.65 \sigma_b$ (95% level). Superimposed are distributions for concentration measurements (r_c) with mean values equivalent to $\bar{r}_b + 1.65 \sigma_b$ [top] and $\bar{r}_b + 3.3 \sigma_b$ [bottom]. The hatched areas show the fraction of concentration measurements below the detection limit.

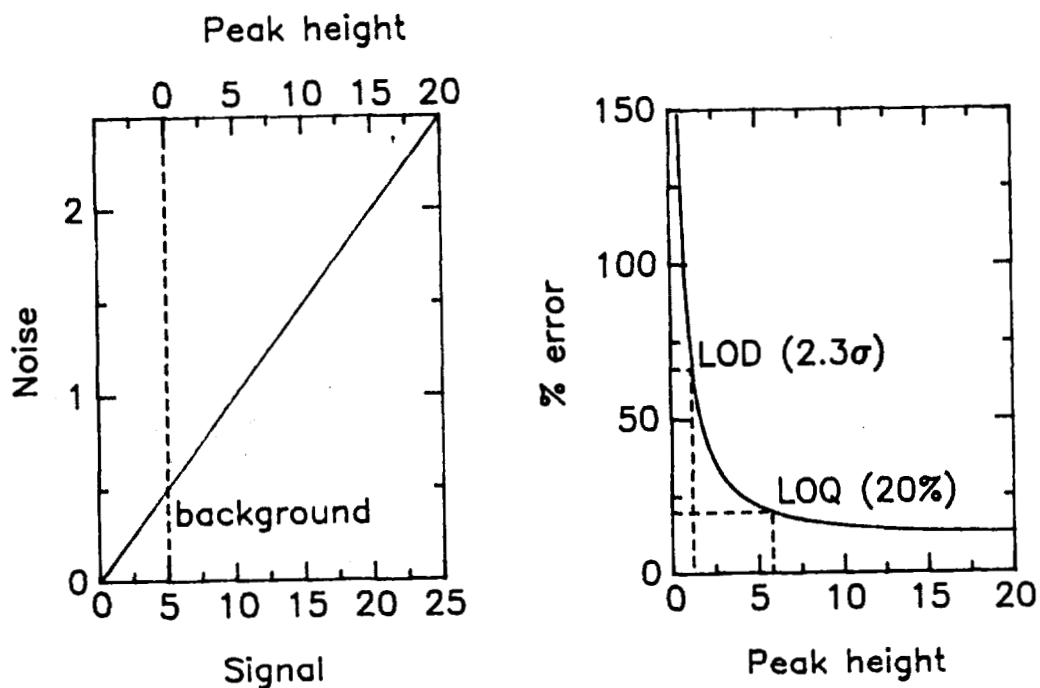


Figure 4: Dependence of noise and relative error on signal strength.

- a) [left] This diagram represents the case where the noise is proportional to the signal. The interesting quantity for the analysis is the peak height, i.e., the difference of the signal minus the background.
- b) [right] This graph shows the variation of the relative error ($100 \times \text{noise} / \text{peak height}$) with the peak height based on the data shown in a).

SCAQs Hydrocarbon Comparison Study Coef. of Variation vs Mean Concentration

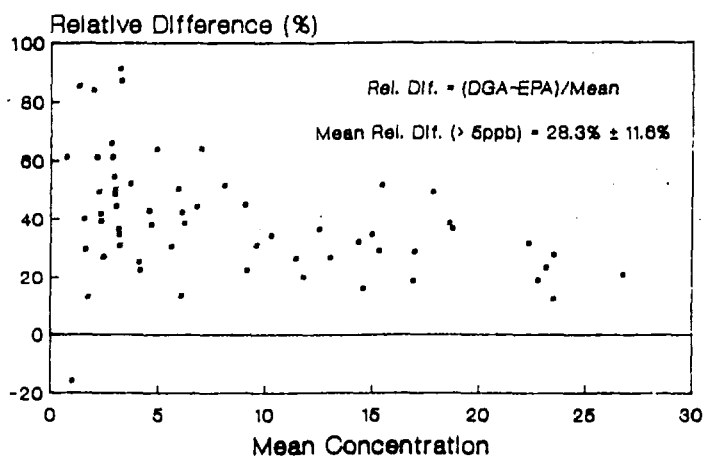
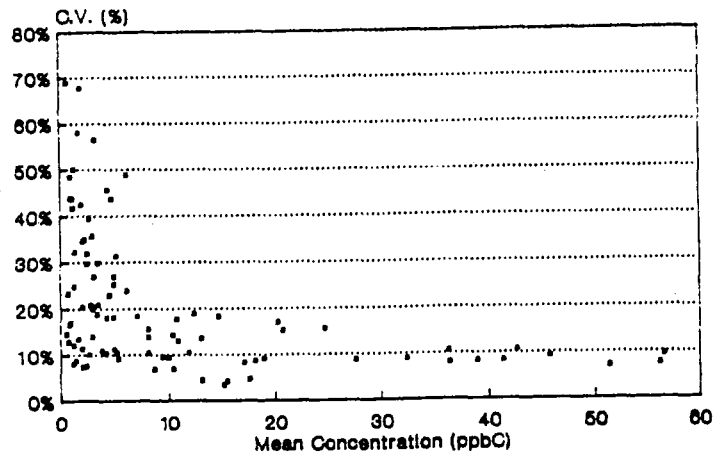


Figure 5: Practical examples for the theoretical dependence of the relative error on the signal strength shown in Figure 4. The top diagram shows the coefficient of variation for the analysis of air samples done by four different laboratories as a function of the mean concentration. The bottom diagram shows the relative difference between two methods for the measurement of PAN as a function of the mean concentration. [Examples are from Fujita and Collins, 1989]

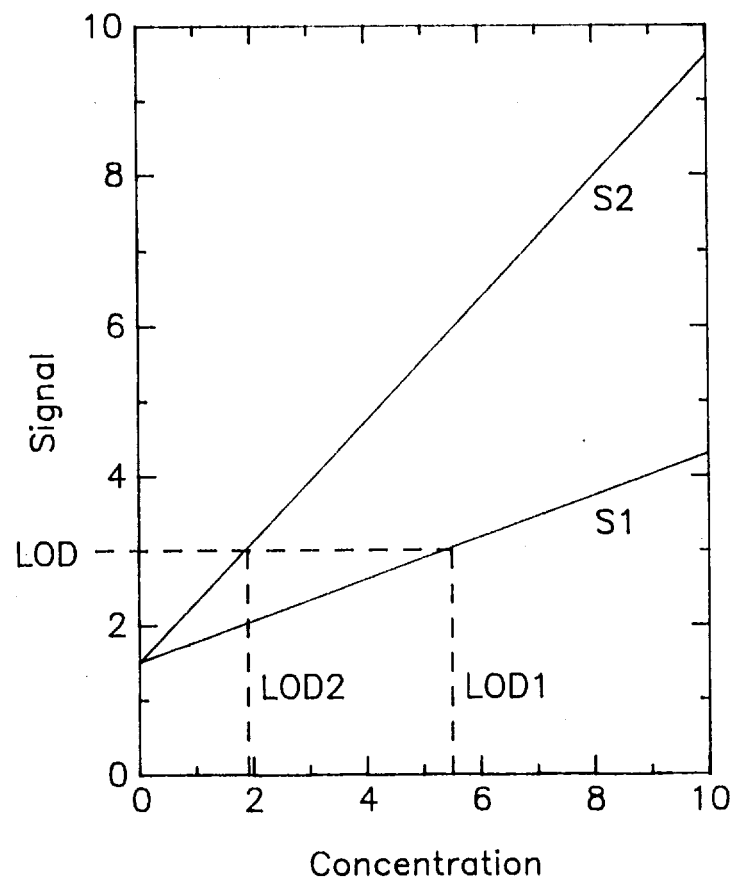


Figure 6: Effect of increased method sensitivity on the detection limit. When the sensitivity is raised from S1 to S2, the background value is not effected, and thus the limit of detection is at the same location along the y-axis. Because of the change in slope, this constant level translates to two different detection limits when expressed as concentrations.

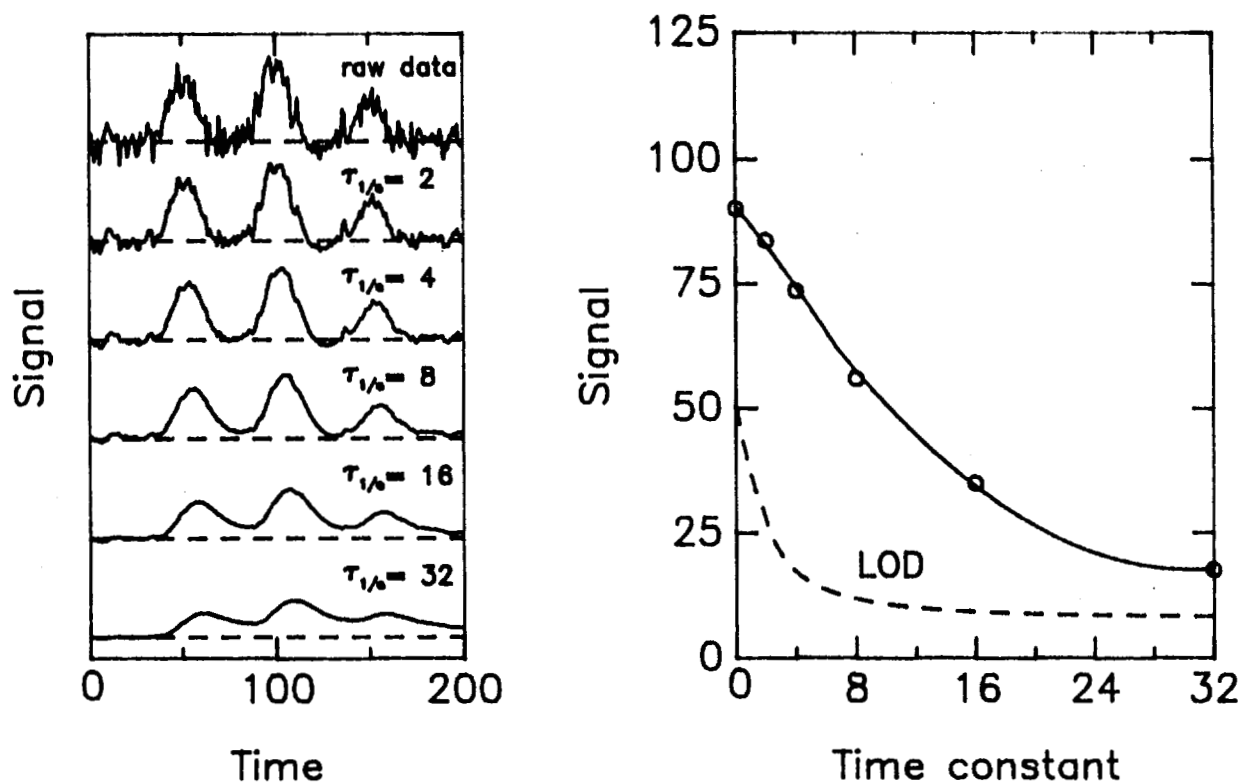


Figure 7: Effect of smoothing on the peak shape.

- a) [left] When the raw data shown in the top trace are smoothed by low-pass filtering with the indicated time constants the noise level is reduced, but the peak shapes become distorted.
- b) [right] The peak height in the data shown in a) is plotted as a function of the time constant of the low-pass filter. Also shown is the change in the detection limit based on the reduced noise level. In this example, optimum results are achieved with time constants between 4 and 8.