APPENDIX G
Detection Limits and Quantitation Limits

G-1. Introduction.

G-1.1. Environmental statistical analysis is complicated by a practical constraint on laboratory analysis—the technical impossibility of identifying zero concentrations. This means that it is physically impossible for a laboratory analysis to confirm the complete absence of the chemical or compound of interest. A chemical may be present at some unknown concentration below the low end of the concentration range that the analysis is able to detect. Therefore, for most statistical applications that evaluate site data, there is a need to substitute some number (a "censored" value) that represents the lowest concentration reasonably detected. This threshold or censoring limit is often termed a "detection," "quantitation," or "reporting" limit. However, this Appendix provides separate definitions for the terms "detection" and "quantitation limit" and does not use these terms interchangeably.

G-1.2. To determine which censoring limit should be used for statistical evaluations, it is necessary to understand how environmental laboratories define detection and quantitation limits, as these quantities are used to establish censoring limits. Unfortunately, the subject of detection and quantitation limits is often confused by the highly diverse, and often overlapping, definitions applied to these quantities. Furthermore, no standard approach to establishing censoring limits for environmental data exists. This Appendix describes some of the methods for establishing detection limits and subsequent requirements for substituting values for non-detects in the data set.

G-2. Detection Limits. No instrumental method of chemical analysis is capable of "seeing" a value of zero. All measurement systems are subject to bias and variability. A fundamental contributor to this is the presence of "noise" in the measurement process. Noise can have any number of sources. For example, if one examines the pictorial output from a gas chromatographic analysis (a chromatogram) of a control sample at the normal scale at which it is displayed in a commercial data package, one would observe a Gaussian peak that represents the analyte of interest and what appears to be a straight, smooth line beyond the peak referred to as the "baseline." Figure G-1 depicts a cartoon example. However, that same graph examined at a higher level of magnification would reveal a very different picture of fluctuations across the same line (Figure G-1). Those fluctuations constitute noise and can result from such factors as vibration in the environment around the instrument, fluctuations in electrical current or voltage, the incidental presence of contaminants in the system, or even stray ionizing radiation from universal background.

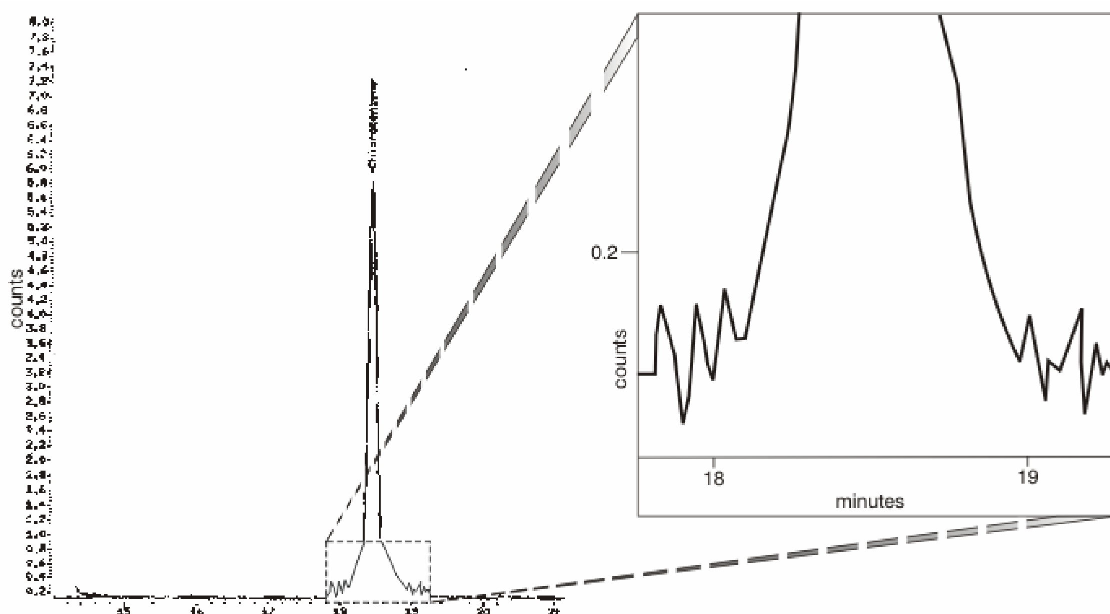


Figure G-1. Noise in GC baseline.

G-2.1. If a very small amount of a target analyte were placed in the measurement system, assuming that the instrument was functioning properly, the analyte would cause a response in the detector that would be translated into a small Gaussian type peak on the chromatogram. However, as the concentration is decreased, the size of the peak decreases until it is “lost” in the noise of the measurement system. Because the amount of noise in the system at any given moment is essentially random, the amount of analyte that can be hidden by the noise is variable but, on average, is always greater than zero.

G-2.2. As the term is typically used in the environmental testing industry, a ”detection limit” (DL) is the concentration that gives rise to an analyte peak or signal that is statistically greater than the surrounding baseline noise at a high level of confidence (typically the 99% level of confidence). The analyte cannot be confidently reported as present when the analyte concentration is less than the DL. Concentrations greater than the DL are reported as “detected.”

G-2.3. However, theoretically, there are two types of “detection limits”: The “Type I DL” that minimizes false positives (Type I error) and the “Type II DL” that minimizes false negatives (Type II error). A false positive occurs when an analyte is absent, or the true concentration is less than the baseline noise but is erroneously reported as present. A false negative occurs when an analyte is erroneously reported as less than or equal to some concentration when it is actually present at a greater concentration. The two types of detection limits are illustrated in Figure G-2.

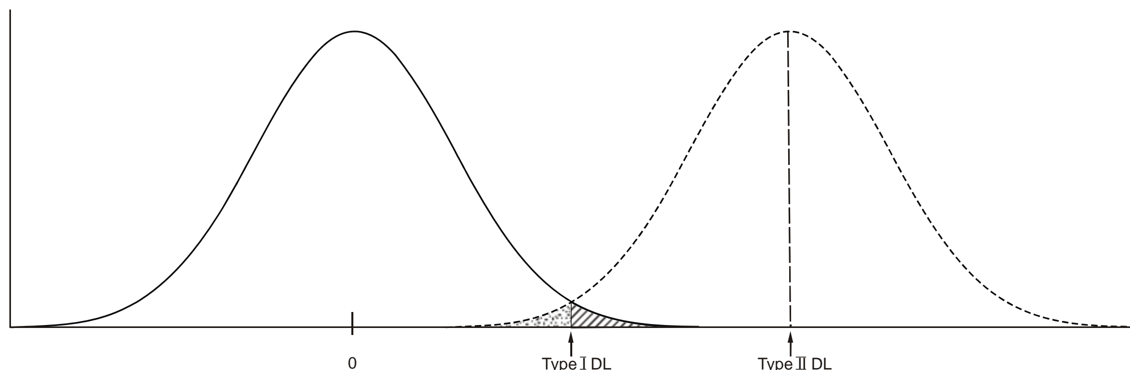


Figure G-2. "Type I DL" (L_C) and "Type II DL" (L_D).

G-2.4. The International Union of Pure and Applied Chemistry (IUPAC), an international, non-governmental organization that supports the advancement of chemical science, refers to the "Type I DL" as the "critical value" and the "Type II DL" simply as the "detection limit." Therefore, for simplicity and to conform with international nomenclature, the IUPAC terminology is predominately used in this document. The critical value is the threshold of analyte or instrument signal attributable to the presence of analyte that is statistically different from zero or baseline noise at a high level of confidence. The 99% level of confidence is used for chemical analyses. When an analyte is reported at a concentration greater than the critical value the conclusion is as follows: The analyte is present at some concentration greater than zero at the 99% level of confidence. The "detection" of the analyte is reported. However, if the analyte concentration reported from a measurement is less than the critical value, the analyte may or may not be present (the true analyte concentration may or may not be greater than zero). Under these circumstances, no conclusion regarding the presence or absence of the analyte is possible. The IUPAC detection limit is established to addresses "non-detections" of the analyte.

G-2.5. When a measurement is taken and the analyte is less than the critical value, the conclusion is that the analyte, if present, is present at some concentration less than the detection limit; the non-detection is reported as "less than the detection limit."

G-2.6. Currie's (1968) approach readily illustrates the nature of the critical value and detection limit on a conceptual level. Currie defines the critical level, L_C, as the concentration at which the binary decision of detection can be made with a specified level of confidence. The shaded area to the right of L_C in Figure G-2 represents the Type I error (i.e., the probability of concluding the analyte is present when the true concentration is zero). Currie defines the limit of detection, $L_{D,}$ to provide an acceptable Type II error rate. The shaded area to the left of L_C represents the Type II error (e.g., the probability of failing to detect the analyte when the true concentration is L_D). In order to calculate quantities L_C and L_D, the following simplifying assumptions are made: The concentrations are normality distributed, the standard deviation is known (or there is negligible uncertainty for the standard deviation), and the standard deviation is not a function

of concentration and the "true" (population mean) concentration is zero. For the 99% level of confidence:

$$L_C = 2.33\sigma$$

$$L_D = L_C + 2.33\sigma = 2\, L_C$$

G-2.7. Unfortunately, it is common practice for environmental chemists to refer to the critical value as the "detection limit." For example, the method detection limit (MDL), defined by 40 Code of Federal Regulations (CFR) Part 136 (Appendix B), is essentially a critical value (as defined by the IUPAC). There is no standard terminology for the IUPAC detection limit for environmental testing. The USACE refers to this as "method reporting limit" or "the reporting limit for non-detections."

G-2.8. The fact that an analyte can be detected, a qualitative statement specifying the presence or absence of the analyte at some level of detection, does not necessarily imply that the analyte concentration can be precisely or accurately specified. The concentration at which quantitative statements can be made reliably is called the "quantitation limit." However, there is no standard terminology for this quantity in the environmental testing industry. It could be referred to as a "report limit" or erroneously referred to as "detection limit." Terms such as "practical quantitation limit" or "contract required quantitation limit" *could* be used. Furthermore, as used by environmental testing laboratories, these terms may, but not would necessarily, refer to the "quantitation limit" as it is defined in this document.

G-2.9. There is a host of terminologies applied to detection and reporting limits depending on the source and the details of the definition. Some more commonly employed terms are listed and cross-referenced in Table G-1. The relationships (cross references) for the various definitions are approximate in nature.

G-3. Alternative Approach to Calculate the Critical Value and Detection Limit. Although the Currie approach is conceptually viable, there is a major practical problem with the approach. Currie did not propose a practical experimental design to calculate L_C, but expressed L_C in terms of the population standard σ (which is usually unknown), rather than the sample standard deviation, s. (In other words, $L_C = 2.33\sigma$ only when the distribution is normal and σ is known.) Similarly, L_D cannot be calculated using σ if this quantity were unknown. However, for a normal distribution, L_C can be defined as an upper tolerance limit for a population mean $\mu = 0$ and can be calculated from s using an equation of the form (Georgian and Osborn, 2003):

$$L_C = K_{p,1-\alpha,n-1}\, s$$

Table G-1.
Common Detection and Quantitation Limit Terminology

Term	EPA Definition	USACE Analogue	IUPAC Analogue
Instrument Detection Limit (IDL)	Concentration that can be distinguished from instrument noise with 99% confidence that the response is not a false positive based on analysis of seven replicates of a standard.	None	Critical value (for determinative portion of method)
Method Detection Limit (MDL)	Concentration that can be distinguished from background with 99% confidence that the response is not a false positive based on analysis of seven replicates of fully processed blank spikes. (Defined in 40 CFR, Part 136, Appendix B.)	MDL	Critical Value (for determinative and preparatory portion of the method)
Sample Quantitation Limit (SQL)	The MDL modified on a sample-specific basis for such factors as dry weight or dilution.		Critical Value (adjusted for sample-specific factors)
Estimated Quantitation Limit (EQL)	The MDL multiplied by a factor between 3 and 5.	Method Quantitation Limit (MQL)—The MQL is at least 3 times the MDL, ½ Action Level, and lies on the initial calibration curve.	Quantitation Limit
Practical Quantitation Limit (PQL)	The EQL “smoothed” to round numbers.	Method Quantitation Limit.	Quantitation Limit
Method Reporting Limit (MRL)	None	Censoring limit for the reporting on non-detects. Determined from project objectives. Commonly set from the Type II DL to the MQL.	Detection Limit

The standard deviation s is calculated from a set of n replicate samples (e.g., a clean matrix such as reagent water spiked with the analyte of interest) that are processed through the entire analytical method. The factor $K_{p,1-\alpha,n-1}$, which depends upon the *coverage probability* (p), level of confidence ($1 - \alpha$) and number of samples (n), can be calculated from Tables B-2 and B-15 using the following equation:

$$K_{p,1-\alpha,n-1} = Z_p \sqrt{[(n-1)/\chi^2_{n-1,\alpha}]} \quad .$$

For example, if $1 - \alpha = 0.95$ (i.e., $\alpha = 0.05$), $p = 0.99$ and $n = 7$, then from Table B-2,

$$\chi^2_{n-1,\alpha} = \chi^2_{6,0.05} = 1.635$$

and, from Table B-15, $Z_p = Z_{0.99} = 2.33$. Therefore,

$$K_{0.99,0.95,6} = 2.33\sqrt{[(7-1)/1.635]} = 4.46 \quad .$$

If a large number of blank samples are analyzed, with 95% confidence, at least 99% of all the measurements will be less than $L_C = 4.46\ s$. The above equation, however, assumes normality and constant variance. A conservative approximation for L_D would consist of initially calculating L_C using the equation above then setting L_D equal to two times L_C.

G-4. EPA Method Detection Limit and Other Detection Limits. There are two major DL estimators: those based on a "single concentration design" and "calibration designs." The major disadvantage of single concentration designs is they assume that variability at a given concentration is constant (i.e., the variability near the DL is similar to that at higher concentrations). Typically, for a single concentration design, a set of replicate samples containing the analyte of interest at a fixed, known concentration are processed to calculate the critical value. Therefore, the critical value is determined at the single concentration for the replicate study and it is assumed that a higher or lower concentration would produce substantively the same value. The MDL is based upon a single concentration design. In calibration designs, the critical value is calculated using multiple concentrations over the range of the critical value. The multiple concentration levels provide a means to model the variance (e.g., or standard deviation) as a function of concentration. In this way, the resulting critical value estimate is not simply a function of sample spike concentration. However, single concentration designs are advantageous relative to multi-concentration designs because they are much simpler and less costly to perform. The critical value can be defined in many different ways; however, only the most commonly accepted method, the EPA MDL procedure, is discussed in detail.

G-4.1. *EPA Method (Single Concentration Design)*. Historically, EPA has used single concentration designs, even though single concentration designs and their associated DL estimators are rarely completely justified. The MDL (defined by 40 CFR) is a single concentration design for the critical value that most environmental testing laboratories use.

G-4.1.1. The EPA defines an "instrument detection limit" (IDL) as an experimentally derived quantity arrived at by repeatedly injecting a small but visible amount of a pure analytical standard into the instrument, measuring the variability in the quantitative results, and calculating the IDL assuming 99% confidence that the observed response is not a false positive. The IDL is generally only performed for inorganic metals analyses. The IDL is typically calculated in the same manner as the MDL, using a Student's *t*-statistic. The two quantities differ predominately in the way the samples are processed. The IDL is determined via the direct instrumental analysis of standards containing the analyte of interest. However, when environmental samples are analyzed, they generally are not directly injected into instruments but are subject to a variety of prior preparatory processes (such as extractions, derivatizaton, solvent exchanges, cleanup, and dilu-

tions). Each step in the processing adds additional noise or uncertainty to the measurement system, which the IDL calculation does not take into account. Therefore, IDLs tend to be smaller in concentration than the corresponding MDLs when samples are subjected to an extensive preparatory process prior to analysis. The minimum quantity of practical importance in environmental analysis is that amount that can be reliably distinguished from the sum of all the various sources of noise involved in the analytical method, the method detection limit (MDL). Thus, environmental laboratories typically use the MDL to characterize detection capability.

G-4.1.2. Although the MDL (as defined in 40 CFR) strictly applies to water matrices, it is applied to a broad range of analytical methods, including those for solid samples. This single concentration design requires a complete, specific, and well-defined analytical method. It is essential for all sample-processing steps of the analytical method to be included in the determination of the method detection limit. MDLs depend upon the sample preparatory procedures and the specific laboratory instrument used.

G-4.1.3. The EPA procedure used to estimate the detection limit is summarized below.

G-4.1.3.1. Prepare a homogeneous matrix that is free of analyte (e.g., reagent water or clean sand).

G-4.1.3.2. Prepare each sample mixture at a concentration of at least equal to or in the same concentration range as the estimated MDL in the matrix of interest.

G-4.1.3.3. Prepare a minimum of seven aliquots of the sample to be used to calculate the MDL and process each replicate through the entire extraction/digestion and analytical method.

G-4.1.3.4. Calculate the variance (s^2) and standard deviation (s) of the replicate measurements.

G-4.1.3.5. Calculate the MDL, using the formula: MDL = $t_{0.99.\nu}\, s$, where $t_{1\text{-}\alpha.\nu}$ is the Student's t value appropriate for the 99% confidence level with $\nu = n - 1$ "degrees of freedom"; and the number of measurements, $n \geq 7$. (The appropriate value of Student's t is typically found in a statistical table, and is equal to about 3.14 for $n = 7$ for the 99% level of confidence).

G-4.1.3.6. Review results to verify the reasonableness of the calculated DL.

G-4.1.4. The use of the MDL for decision-making (e.g., determining environmental impacts) has recently triggered intense scrutiny of the viability of the MDL for measuring detection capability. The following is a partial list of potentially flawed assumptions or problems associated with the MDL as defined in 40 CFR.

G-4.1.4.1. The MDL addresses false positives (i.e., Type I error), but does not address false negatives (Type II error); for example, a non-detection cannot be confidently reported as "< MDL." (However, it should be noted that there is controversy regarding the interpretation of the MDL in terms of the IUPAC definitions of the critical value and detection limit; some individuals have argued that the MDL is actually an IUPAC detection limit.)

G-4.1.4.2. The MDL underestimates method variability as it is typically calculated using a small number of replicates within a short period of time and has been interpreted to be a *prediction limit for the next single future observation,* minimizing false positives at the 99% level of confidence for only one future environmental sample (and not a set of multiple samples) when the analyte is absent (though it should be noted that the interpretation of the MDL as a prediction limit is also controversial).

G-4.1.4.3. The standard deviation is assumed to be constant (i.e., not a function of concentration).

G-4.1.4.4. Normality is assumed.

G-4.1.4.5. No analytical bias is implicitly assumed (e.g., no analyte loss, average analyte "recoveries" of 100%). (The MDL accounts for analytical method variation in the form of random "precision error.")

G-4.1.4.6. The matrix used to perform the MDL study (e.g., reagent water) is assumed to be equivalent (with respect to all physical or chemical properties that would affect detection capability) to the actual environmental matrices that will be tested (e.g., waste water and groundwater).

G-4.1.5. In general, one or more of the assumptions discussed above are routinely violated to some extent for environmental testing. MDLs are statistically derived quantities and are only estimates of the actual detection limit (critical value). For example, based on purely statistical considerations, MDLs are uncertain by a factor of approximately two. Furthermore, because MDLs are typically generated by processing clean material (such as purified water or sand) rather than actual environmental samples, they represent "best case" detection capability. In general, the material analyzed to calculate the MDLs is not representative of the chemical and physical composition of the environmental samples. Detection limits calculated using an actual environmental matrix could be higher than the MDL by an order of magnitude. However, because of these factors, environmental laboratories often report "detection limits" several times greater than MDLs (although there is no uniform standard for how this is done). The detection limits proposed in Paragraph G-3 overcome the first two shortcomings of the MDL discussed above.

G-4.1.6. Lastly, when detection limits such as the MDL are constructed from prediction limits (using either a single concentration or calibration design), in order to minimize false positives at the specified level of confidence, a new detection limit must (in theory) be calculated (from a new study) prior to each new sample being analyzed. However, this is not done in practice. Detection decisions for an enormous number of test samples are calculated based on the results obtained from a single MDL study. This results in a much greater frequency of false positives than 1%. To ensure that false positives are minimized for a large unspecified number of future measurements, detection limits may be constructed from tolerance intervals so that a large proportion of future measurements, p, will be less than the upper tolerance limit (UTL) with a high level of confidence when the "true" concentration is zero. For the critical value, an UTL for p100% coverage (e.g., where $p = 0.99$) at the $(1 - \alpha)$100% (e.g., 99%) level of confidence could be constructed for a "true" concentration of zero (e.g., refer to Paragraph G-3).

G-4.2. *Calibration Designs*. In one type of calibration design, a series of samples are spiked at different known concentrations in the range of the hypothesized critical value, and variability is determined by examining the deviations of the actual response signals from a fitted regression line (instrument response versus concentration). In this design, it is typically assumed that the distribution of the deviations from the fitted regression line is normal with constant variance across the range of concentrations used for the study. The relationship between response signal (Y) and spiking concentration (X) in the region of the critical value is assumed to be a linear function of the form:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

where the (population) "residual" $\varepsilon = Y - (\beta_0 + \beta_1 X)$ is the deviation of the measured value of Y from the "true" regression line $\beta_0 + \beta_1 X$. It is assumed that the distribution of values for ε is normal with mean $\mu = 0$ and some constant variance. A set of n measurements (x_i, y_i) would be used to estimate a line of the form $Y = b_0 + b_1 X$, where the sample slope, b_1, estimates the population parameter β_1 and the sample intercept, b_0, estimates the population parameter β_0. The regression model is used to calculate the critical value and detection limit by constructing either *prediction* or *tolerance* limits for the regression line, $Y = b_0 + b_1 X$. (The specific mathematical formulas used are beyond the scope of this document.)

G-4.2.1. Hubaux and Vos method calibration design is an example of an approach in which statistical prediction limits are used to calculate DLs. The critical, L_C, value is calculated from a 99% prediction interval for the linear regression model. A single future measurement will be less than L_C at the 99% level of confidence when the "true" concentration is zero. The limit of detection, L_D, is then defined as the smallest concentration at which there is 99% confidence a value greater than L_C will be obtained. This method assumes that the variability is constant throughout the range of concentrations used in the calibration design (e.g., if this assumption is violated, a variance stabilizing transformation might be applied and the assumption of constant variance may be reevaluated). The critical value obtained from the Hubaux and Vos design can

be viewed as a multi-design concentration version of the single concentration-designed MDL (e.g., since the MDL is also a prediction limit, minimizing false positives for only one single future observation). Regression models used for multi-concentration designs can also be used to define detection limits based on prediction and tolerance intervals. A tolerance or prediction interval can be constructed for each possible value of the independent variable *X*.

G-4.2.2. As previously stated, the Hubaux and Vos calibration design assumes that the variance is homogeneous (constant) throughout the range of calibration function. This assumption is rarely completely justifiable. In practice, variation in the response signal is often proportional to the concentration. For example, if violations of this assumption are ignored, the variability at low levels can be overestimated and, as a result, detection limits can be overestimated. However, some calibration designs account for non-constant variance. For example, the detection limits for non-constant variance calibration designs can be calculated using a technique called weighted least squares (WLS). The WLS calibration design is similar to the Hubaux and Vos design, but the underlying regression model would assume, for example, that variance is proportional to concentration (Gibbons and Coleman, 2001).

G-5. Quantitation Limits. The ability to distinguish between the presence or absence of an individual analyte, particularly in a complex mixture such as an environmental sample, does not imply the ability to accurately and precisely measure the quantity of analyte present in the mixture. Imagine, for example, a peak partially hidden in the noise of an instrument. If the quantity of analyte is measured as proportional to the height or area of the response, as is the usual case in environmental analysis, from what point is it measured? Where is the baseline? Should it be measured from the lowest point in the noise, the average noise level, or the top of the noise? In other words, because the baseline is constantly shifting, what portion of the observed peak is noise and what portion is response? The magnitude of the response ascribable to the analyte (e.g., peak area) cannot be known with a high degree of certainty (high accuracy and precision); therefore, the measured value must, by definition, be equally suspect. There is a point at which the measured value is so much larger than any possible contribution from measurement noise that the noise becomes negligible relative to the analyte result. That point is the quantitation limit (QL).

G-5.1. In EPA terminology, the QL is, by definition, a value sufficiently removed from the detection limit to ensure that quantitative statements made at that value meet defined degrees of precision and accuracy by most laboratories under most analytical conditions. Because the definition is vague, the QL is also vague. In fact, most practical applications of this concept are altogether arbitrary. For example, in EPA SW-846, the EQL for a given analysis is defined as 5 to 10 times the MDL. However, the multiplication factor is somewhat arbitrary (e.g., various definitions of the QL for various programs have required the MDL to be multiplied by factors ranging from 2 to 10). Some justification for the use of a factor of 5 to 10 is as follows: If the MDL is assumed to be roughly equal to the magnitude of the uncertainty from analytical noise, the relative error should be 20 to 10% at 5 and 10 times the MDL, respectively. However, it should be noted

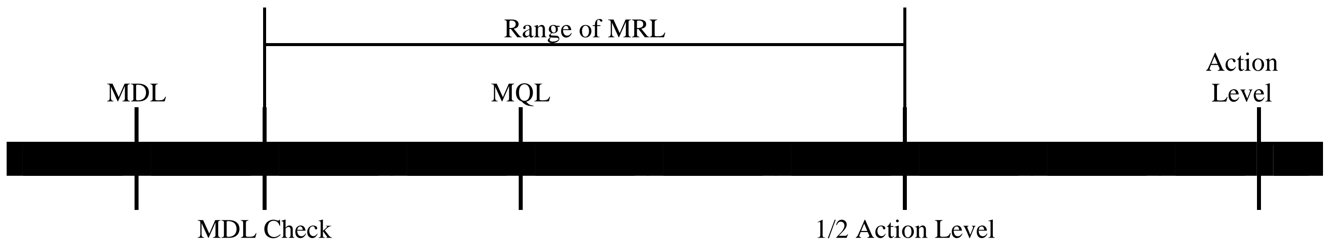
that this assumes that analytical bias is negligible and the standard deviation (used to calculate the MDL) is not a function of concentration and possesses negligible uncertainty. In general, these are not valid assumptions. In particular, the standard deviation is typically an increasing function of concentration and can vary by a factor of about two when it is calculated at a fixed concentration using only seven replicates (as in 40 CFR). Setting the QL at a concentration at least 5 or 10 times the MDL is stated only as guidance (e.g., since the uncertainty at these levels may still be relatively large).

G-5.2. To ensure acceptable precision and accuracy at any arbitrarily defined QL, quality control samples spiked at the QL could be included in the analytical sequence to actually measure the precision and accuracy of the measurement process (e.g., using control charts). Thus, this approach would quantify the uncertainty at the QL for "clean matrices." Unfortunately, environmental testing laboratories do not routinely analyze quality control samples at the QL, but at much higher concentrations (e.g., this would have to be requested when analytical services are contracted). In addition, the quantitation limit should fall within the calibration range of the analytical method. Instrumental response is typically unknown at concentrations less than the lowest initial calibration standard. Environmental testing laboratories usually (but do not necessarily) include the reported QL concentration as the low point of the initial calibration curve.

G-5.3. The "Practical Quantitation Limit" (PQL) is defined as the lowest limit of quantitation achievable by laboratories within specified limits on precision and accuracy during routine laboratory operating conditions. Unfortunately, acceptance limits for precision and accuracy at the PQL are seldom defined. In practice, the PQL is typically established by multiplying the MDL (as derived from 40 CFR Part 136 instructions) by a factor of three to five (from EPA SW-846, Chapter 1). The result obtained is the EQL. The EQL, being a multiple of the statistically derived MDL, will be different for each analyte tested. In the commercial laboratory community, PQLs are frequently set at the low point of the curve and are relatively uniform for methods where multiple analytes are simultaneously determined. The values thus obtained are variously referred to as PQLs, Reporting Limits (RLs), Less Than (< or LT), Non-Detects (NDs), or "U"-values.

G-5.4. The nomenclature that has been historically used by the USACE is defined in EM 200-1-3. The MDL is developed according to the EPA model. The method quantitation limit (MQL) is required to be at least three times greater than the MDL and must fall within the initial calibration range and recommends that the MQL concentration not exceed one half the project-specified action level (decision limit). The MRL, which is established as illustrated in Figure G-3, depends on the end use of the data. The MRL is equal to the MDL for data to be used in support of risk-based decisions. Although this is consistent with current EPA guidance, it should be noted that *false negatives (Type II error) cannot be adequately controlled at the MDL*. The lower reporting limit for non-risk-based data is the concentration of the MDL check sample, which provides a higher level of confidence for non-detections. The MDL check sample is a spike

processed through the entire analytical method that is sufficiently large to consistently results in a detected value.



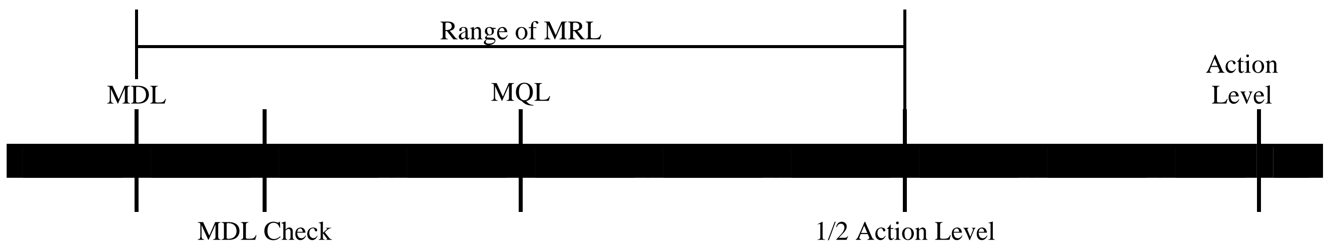


Figure G-3. USACE definition of the method reporting limit.