DECLARATION OF SHAWN DUFFY

DECLARATION OF SHAWN DUFFY

I, Shawn Duffy, declare:

1. I am employed by CH2MHill as a Senior Chemist. My resume is attached to this Declaration as Exhibit A. Pacific Gas and Electric Company ("PG&E") has engaged CH2MHill to assist with issues surrounding the chromium plume in Hinkley, California. I have been working on chromium chemistry and testing related issues for PG&E since 2004. I was asked to lead a team tasked with assessing the reliability and variability associated with the analysis for hexavalent chromium in a typical commercial laboratory setting with a theoretical Method Detection Limit ("MDL") of 0.02 µg/L as required by the draft Cleanup and Abatement Order No. R6V-2011-0005A1 (the "Draft CAO").

2. I have more than seventeen years of experience in the field of analytical chemistry. On behalf of PG&E, I have been responsible for technical oversight of laboratory analytical activities, quality control and quality assurance, and data validation for the Hinkley remediation project. I perform on-site field audits for the evaluation of sampling protocols and sampling team performance, as well as conduct laboratory audits to ensure quality control and compliance with the project's quality assurance program plan. For the last seven years, I have worked with the subcontracted laboratory's analysts and quality control personnel to ensure that the analytical techniques are consistent and correct according to the applicable methods for hexavalent and total chromium analyses to achieve the current reporting levels of 0.2 ug/L for Cr(VI) and 1.0 ug/L for total chromium.

3. My opinion is that:

(a) A reporting level below 0.10 µg/L for Cr(VI) is beyond the capacity of the typical commercial laboratory to detect and accurately quantitate using EPA Method 218.6.

(b) A reporting level at or below 0.10 µg/L for total chromium is beyond the capacity of the typical commercial laboratory to detect and accurately quantitate using EPA Method 200.8.

4. As part of the process of forming my opinions, I conducted the following study:

(a) Ten water samples were collected from Hinkley wells that had previously been reported as non-detect or just above the detection limit; with a 0.2 µg/L detection limit for Cr(VI) and as non-detect with a 1.0 µg/L detection limit for total chromium. Six groundwater monitoring wells and four water supply wells were sampled. In addition, three of the ten wells were randomly selected and field duplicate samples were collected from the three locations. Each sample was filtered through two, certified clean, inline 0.45 micron filters and collected into a single certified clean container using clean hands/dirty hands methodology, a sampling method that prevents contamination of the samples by limiting the contact of the sample and sample containers with any object that could cause contamination. The Cr(VI) sample aliquot for each well was carefully homogenized – that is, shaken after being filtered into a single large container – and an aliquot split into three certified clean bottles pre-preserved with a buffer solution and submitted to three separate California-approved Environmental Laboratory Accreditation Program("ELAP") laboratories that claim to achieve an MDL of 0.02 µg/L or less for Cr(VI) using CFR Part 136, Appendix B procedures. The total chromium (all forms of chromium) portion for each of the samples listed above was also carefully homogenized and split into a certified clean bottle pre-preserved with nitric acid and submitted to the same three California-approved ELAP laboratories that were asked to analyze the total chromium by EPA Method 200.8 and report the results to the lowest possible MDL using CFR Part 136, Appendix B procedures, which has been labeled as a technically-flawed procedure for assessing level of

detection by the U.S. EPA in the *Revised Assessment of Detection and Quantitation Approaches, EPA-821-B-04-005.* (Exhibit B).

(b) All samples were collected under formal chain of custody and placed into coolers, on ice, immediately after sampling. The samples were then transferred to each lab's courier for delivery to the contracted project laboratories. All of the sample containers used for this study were supplied by one laboratory, and the samples were labeled using generic alpha-numeric identifiers that were similar to the double-blind sample identifiers used for the quality control samples included in this study.

(c) The study included positive, negative and duplicate quality control samples, all of which were submitted as double-blinds to each of the contracted project laboratories.

(d) Six of the quality control samples were performance evaluation ("PE") samples containing known concentrations of Cr(VI) from 0.01 to 0.10 ppb. Environmental Resource Associates ("ERA") was contracted to prepare and certify three sets of double-blind PE samples for total chromium and Cr(VI) at concentrations of 0.01 µg/L, 0.02 µg/L, 0.04 µg/L, 0.06 µg/L, 0.08 µg/L and 0.10 µg/L. ERA prepared the PE samples and transferred them into pre-preserved bottles provided by the same laboratory used for the samples, and using deionized water that was demonstrated to be free of total chromium and Cr(VI) to a MDL of <0.02 µg/L (also provided by the same laboratory). ERA was unable to certify the final concentration of the PE samples due to their ultra-trace level concentrations.

(e) The PE samples were delivered to the project team in the field on the afternoon of the field sampling event to insure that all the study samples were shipped together to each of the three contracted project laboratories. Upon receipt of the PE sample by the project

team, fictitious sample labels were affixed to the PE sample containers and the sample identifiers entered on the chain of custody.

(f) Two field-derived blanks were also included. An aliquot of the same chromium-free deionized water provided to ERA was transferred to identical sample containers, labeled and sent as one of the study samples. A second blank was prepared at one of the ten ground water sample locations, an aliquot of the same chromium-free deionized water was put through the same in-line 0.45 micron filters as the samples, split into preserved containers, labeled and sent as one of the study samples.

(g) The samples were sent to three California accredited ELAP laboratories: Truesdail Laboratory, Inc. ("TLI"), Advanced Technology Laboratory ("ATL") and BCLab, Inc. ("BCL"). The three laboratories are listed on the California Department of Public Health web page, subgroup code 103.310, which is specific to the certification for EPA Method 218.6 for Cr(VI). The laboratories were requested to report the Cr(VI) results to a level of 0.02 µg/L using EPA Method 218.6 with a modified buffer, and as low as possible for total chromium by EPA Method 200.8.

5. The three laboratories reported results for total chromium and Cr(VI) as follows:

Trace-Level Chromium Study – June 2011

Sample ID	Sample Type	ATL Cr(VI) µg/L	ATL Cr (initial analysis) µg/L	ATL Cr (Second analysis) µg/L	BC Lab Cr(VI) µg/L	BC Lab Cr µg/L	TLI Cr(VI) µg/L	TLI Cr µg/L
H-01-Q2	0.10 - PE STD	0.096	ND < 0.02	0.117	0.031	ND < 0.85	0.103	0.146
H-02-Q2	0.08 - PE STD	0.082	ND < 0.02	0.096	ND < 0.026	ND < 0.85	0.085	0.396
H-03-Q2	0.06 - PE STD	0.068	ND < 0.02	0.077	ND < 0.026	ND < 0.85	0.071	0.170
H-04-Q2	0.04 - PE STD	0.048	ND < 0.02	0.061	ND < 0.026	ND < 0.85	0.031	0.096
H-05-Q2	0.02 - PE STD	ND < 0.02	ND < 0.02	0.046	ND < 0.026	ND < 0.85	0.046	0.069
H-06-Q2	0.01 - PE STD	ND < 0.02	ND < 0.02	0.035	ND < 0.026	ND < 0.85	0.035	0.140
H-07-Q2	Supply Well	0.41	0.23	0.546	0.34	ND < 0.85	0.429	0.646
H-08-Q2	Supply Well-Dup	0.4	0.21	0.545	0.33	ND < 0.85	0.421	0.635
H-09-Q2	Supply Well	ND < 0.02	ND < 0.02	0.034	ND < 0.026	ND < 0.85	ND < 0.02	0.099
H-10-Q2	Supply Well	ND < 0.02	ND < 0.02	0.029	ND < 0.026	ND < 0.85	ND < 0.02	0.061
H-11-Q2	Supply Well-Dup	ND < 0.02	ND < 0.02	0.042	ND < 0.026	ND < 0.85	ND < 0.02	0.055
H-12-Q2	Supply Well	ND < 0.02	ND < 0.02	0.023	ND < 0.026	ND < 0.85	ND < 0.02	0.510
H-13-Q2	Monitoring Well	ND < 0.02	ND < 0.02	0.190	0.091	ND < 0.85	0.050	0.190
H-14-Q2	Monitoring Well	ND < 0.02	ND < 0.02	0.118	ND < 0.026	ND < 0.85	ND < 0.02	0.140
H-15-Q2	Monitoring Well	ND < 0.02	ND < 0.02	0.130	ND < 0.026	ND < 0.85	ND < 0.02	0.278
H-16-Q2	Monitoring Well	0.23*	ND < 0.02	0.255	0.82*	ND < 0.85	0.422 *	0.304
H-17-Q2	Monitoring Well	ND < 0.02	ND < 0.02	0.103	ND < 0.026	ND < 0.85	ND < 0.02	0.112
H-18-Q2	Monitoring Well	ND < 0.02	ND < 0.02	0.266	ND < 0.026	ND < 0.85	ND < 0.02	0.303
H-19-Q2	Monitoring Well-Dup	ND < 0.02	ND < 0.02	0.218	ND < 0.026	ND < 0.85	ND < 0.02	0.276
H-20-Q2	DI –Blank	ND < 0.02	ND < 0.02	0.207	ND < 0.026	ND < 0.85	0.079	0.288
H-21-Q2	Field-Blank	ND < 0.02	ND < 0.02	0.197	ND < 0.026	ND < 0.85	0.163	0.354

* Cr(VI) samples received in lab at pH < 2

6. The results from the three certified laboratories show the inherent variability of analytical results for samples at these ultra-trace level concentrations, and demonstrate the inability to report quantitatively accurate results for both EPA Method 218.6 Cr(VI) and EPA Method 200.8 total chromium:

(a) The laboratories for the Hinkley project are required to analyze a Cr(VI) low level laboratory fortified blank spiked at the RL (currently 0.2 µg/L) and have routinely met the standard laboratory fortified blank criteria of 90 to 110% recovery (in other words, the labs test a sample with a known concentration and report a result between 90% and 110% of the known concentration). However, only four of the fifteen PE sample results in this study met the same recovery criteria.

(b) Two of the three laboratories, ATL and TLI, reported EPA Method 218.6 Cr(VI) results for the 0.06 µg/L concentration PE standard with recoveries of 113% and 118%, respectfully. The third laboratory reported the results as non-detect at 0.026 µg/L or 0% recovery.

(c) Two of the three laboratories, ATL and TLI, reported EPA Method 218.6 Cr(VI) results for the 0.04 µg/L concentration PE standard with recoveries of 120% and 77%, respectfully. The third laboratory, BCL, reported the results as non-detect at 0.026 µg/L or 0% recovery.

(d) Two of the three laboratories, ATL and BCL, reported EPA Method 218.6 Cr(VI) results for the 0.02 µg/L concentration PE standard as non-detect at 0.02 and 0.026 µg/L, respectfully, or 0% recovery. The third laboratory, TLI, reported the results at 0.046 µg/L, a 240% recovery.

7. At the required level for the MDL of 0.02 µg/L and the required level for the RL of 0.06 µg/L from the Draft CAO, false positive blank detections are a serious problem that can cause erratic non-reproducible trace-level sample results:

(a) ATL analyzed the samples twice for total chromium. During the first round of analysis, the calibration blank showed a high base line and the PE standards were all reported as non-detect. On second analysis, the calibration blank base line was significantly lower and ATL reported all the PE standards with a high bias, 230% at 0.02 µg/L to 117% at 0.10 µg/L.

(b) The 0.01 µg/L PE blank, which was below the MDL for all labs, was reported by one lab at a Cr(VI) concentration of 0.035 µg/L, a clear false positive.

(c) The deionized water blank and the field blank were also reported by one lab at a Cr(VI) concentration of 0.079 and 0.163 µg/L, more clear false positives.

(d) All three blanks were reported by two labs as detects for total chromium at levels between 0.035 µg/L and 0.354 µg/L, an order of magnitude difference.

8. The three labs could not produce inter-laboratory results that would meet the precision (that is, the measurement of reproducibility) or accuracy (the amount of agreement between a measured value and the "true" value) of the methods:

(a) BCL reported the EPA Method 218.6 Cr(VI) results as non-detect for all the PE standards except one. That one standard (0.10 µg/L) was recovered at 31%.

(b) BCL was also unable to detect total chromium, using EPA Method 200.8, below a MDL of 0.85 µg/L, and all samples were reported as non-detect.

(c) One sample (H-13-Q2) had results reported for Cr(VI) ranging from ND <0.02 µg/L to 0.091 µg/L by the three laboratories. While the total chromium was reported ranging from ND < 0.02 µg/L to 0.190 µg/L.

(d) TLI was biased high on all the PE standards for both EPA Method 218.6 Cr(VI) and EPA Method 200.8 total chromium, the result of what I suspect was lab contamination. The high bias included positive detects of 0.140 µg/L, 0.288 µg/L, and 0.354 µg/L for the three blanks and a minimum recovery of 146% for the total chromium.

(e) ATL analyzed the samples for total chromium twice. During the first round of analysis, the calibration blank showed a high base line and the PE standards were all reported as non-detect. During the second round, the calibration blank base line was significantly lower and ATL reported all the PE standards with a high bias, 230% at 0.02 µg/L to 117% at 0.1 µg/L.

I declare under penalty of perjury under the laws of the State of California that the foregoing is true and correct, and that this Declaration was executed on July 9, 2011, at 6:30 PM, Redding, California.

Shawn Duffy

EXHIBIT A

Shawn P. Duffy

Project Chemist

Education

B.A., Biology, Humboldt State University, 1988

Distinguishing Qualifications

- More than 17 years experience in analytical and environmental chemistry
- More than 25 years experience in the field of science
- Experience in a wide range of sampling techniques and situations
- Biological, physical, and chemistry field, office, and lab experience.

Relevant Experience

Project Chemist/CH2M HILL; Mr. Duffy is a chemist with more than 17 years of experience in the field of analytical chemistry. He provides oversight for an analytical laboratory budget of one - two million dollars annually. He provides senior oversight for a CH2M HILL staff of 8 chemists and data managers, providing chemistry, data management, and data validation for the PG&E Program. Mr. Duffy manages subcontracts with laboratories, provides statements of work, manages purchase orders, approves invoices, provides schedules, oversees laboratory corrective actions, and provides a point of contact between the laboratories and CH2M HILL staff. Performs onsite field audit for evaluation of sampling protocols and sampling team performance and conducts laboratory audits to ensure quality control and compliance with the project's quality assurance program plan.

As the lead CH2M HILL chemist on the PG&E Hinkley and Topock Sites; Mr. Duffy has, For the last seven years, worked with the subcontracted laboratory's analysts and quality control personnel to ensure that the analytical techniques are consistent and correct according to the applicable methods for hexavalent and total chromium analyses to achieve the current reporting levels. And with a primary goal of continuing to meeting the forever increasing regulatory requirements and providing cost effect, accurate, low level analyses.

Representative Projects

Pacific Gas and Electric Projects

PG&E Topock Compressor Station - **March 2004 to present.** Provides chemistry support for all aspects of the Topock Compressor Station Project including the Ground Water Monitoring Program, Compliance Monitoring Program, River Monitoring Program, East Ravine Groundwater Investigation, AOC4 Time Critical Removal Action, RCRA facility investigation/remedial investigation, Background Study, and Interim Measures 1, 2, and 3. Oversight of data validation and senior review of data and data quality reports; provides statements of work and coordinates with CH2M HILL contract administrators to supply purchase orders for lab services, and provide information to project managers and field crews to verify that project work is following the work plan and QAPP. Assist with onsite sample collection and sampling coordination.

IM3- As the lead chemist on the design and startup of a field laboratory facility at the Topock IM3 site Mr. Duffy has researched, implemented and trained field personnel for the characterization of hexavalent and total chromium at Topock to meet discharge permit requirements. Investigated and evaluated equipment necessary for IM3 onsite lab and provide design requirements. Provides onsite and on-call chemistry support and training of facility operators

PG&E Hinkley Compressor Station - **March 2004 to present.** Provide chemistry support for all aspects of the Hinkley project, including the Performance Monitoring Program, Nitrate Monitoring Program, Interim Pumping Program, Injection Monitoring Program, and Background Study. Oversees data validation, senior data review, and data quality reports; provides statements of work and coordinates with CH2M HILL Contract Administrators to supply purchase orders for laboratories services; and provides information to project managers and field crews to verify that project work is following the work plan and QAPP.

Other PG&E clean-up and investigative projects include – Antioch investigation, Colusa site investigation and remediation, Humboldt site characterization, Merced MGP, Red Bluff MGP, Selma site remediation, Shell Pond investigation, Wildcat remediation, Woodland MGP, and the Pipeline Hyrotest.

Federal Projects –

Beale Air Force Base, California; Assisted project chemist with data validation and reviewed: evaluated laboratory data, interacted with the laboratories, and provided data quality information to the project chemist and project manager. Drill rig oversight for sampling and monitoring well installation.
Lennar Mare Island; Mare Island, California; Assisted project chemist with data validation. Reviewed and evaluated laboratory data, interacted with the laboratories, and provided data quality information to the project chemist and project manager.
Hill Air Force Base, Utah; Assisted project chemist with data validation. Reviewed and evaluated laboratory data, interacted with the laboratories, and provided data quality information to the project chemist **and** project manager.
Hickam Air Force Base, Utah; Assisted project chemist with data validation. Reviewed and evaluated laboratory data, interacted with the laboratories, and provided data quality information to the project chemist **and** project manager.
Massachusetts Military Reservation (MMR), Massachusetts; Assisted project chemist with data validation.

Experience Prior to Current CH2M HILL

Biologist/Group Lead; Klamath Wildlife Resources, Redding, California; Lead a Spotted Owl Inventory Survey in Plumas National Forest. Identified, photographed, and marked locations of species by map/GPS. Conducted an Aquatic Amphibian survey in the Lassen National Forest, Almanor Ranger District and lead a Northern Goshawk survey in Lassen National Forest, Hat Creek Ranger District. Using a standard compass, GPS, aerial photographs, and topographic maps to run transects thru suitable habitat establishing survey points

On-call Biologist; ENPLAN; Redding, California; Acted as a biological observer during construction of new wetlands for mitigation purposes. Planted seeds gathered from nearby wetlands. Perform monthly hydro- and photo monitoring.

Technical Support Manager/Customer Support Manager; Michrom BioResources, Auburn, California. Provide technical and customer support for an HPLC (High Performance Liquid Chromatography) instrument manufacturer, LC and LCMS (Liquid Chromatography Mass Spectrometry) supplies, and applications technologies company. Performed column packing, column quality assurance/quality check (QA/QC), instrument QA/QC (including HPLC, Auto-sampler, and other OEM instrumentation), and in-house service of instrumentation parts. Provide technical support to in-house and external clients for HPLC/LCMS, columns, and chemistry related issues. Travel to customer sites for installation, training, service and equipment demonstrations. Generate company-wide marketing materials including catalogs, calendars, brochures and magazine advertisements. Part of a three-member team responsible for worldwide sales and technical documents. Advise clients on new and improved LC/MS methodologies and applications.

Hydrologic Technician; U.S. Bureau of Reclamation; Shasta Lake City, California. Install, maintain, and calibrate continuous monitoring satellite telemetry sites. Member of a sampling teams collecting various sample types (soil, water, fish, biological/climactic); team members from the U.S. Geological Survey, U.S. Bureau of Reclamation, U.S. Fish and Wildlife Service, California Department of Fish and Game, and CH2M HILL. Organized /summarized data for official documents and publications. Project Manager and investigation team member for the NCAO Shasta Lake Limnology Study. Member of the Interagency Technical Team for the Anadromous Fish Screen Program. Converted historic database information into a Microsoft Access database and advised the Mid Pacific Regional Data Center on the construction of a Microsoft Access regional database.

Biological Science (Fisheries) Technician; U.S. Bureau of Reclamation; Shasta Lake City. Monitor water quality in the regional rivers and reservoirs as part of the Water Quality Branch of the Environmental and Natural Resources Division. Collected and analyzed water samples to monitor levels of Cu, Cd, and Zn from the Spring Creek water shed (Iron Mountain Mine - EPA Hazardous Waste Site). Interfaced with contract laboratory personnel to ensure compliance to project specific holding times, instrumentation methodologies, and reporting limits. Member of the Shasta Lake Limnology Study team (NCAO/Denver).

Team Leader; North State Resources; Redding, California. Perform Forest Inventory for Six-Rivers and Shasta-Trinity National Forests. Located field plots using aerial photos and topographic maps. Identified and recorded data on the size and condition of trees, the species and cover of understory vegetation, ground cover and various site attributes, habitat types, and forest health indicators. Conducted a Vegetation Survey contracted from the U.S. Forrest Service (USFS).

Sample Custody Group Leader/Supervisor/ GC Chemist/Extraction Chemist; CH2M HILL Quality Analytical Laboratory; Redding, California. (Sample Custody) Train and supervise sample custodians. Maintain client relationships and ensure the flow of information between the client's project team and the laboratory personnel. Manage over-the-counter projects and filled in as project manager for Client Service Personnel. Provide all project specific information necessary for sample analysis and subsequent reporting of analytical data. Maintain custody of

the samples within the laboratory and assisted clients with information about analytical procedures and results. Order stock and. Subcontract analyses that required work to be performed at another laboratory. Evaluate and revise laboratory Standard Operating Procedures. Interpret and implement DOT regulations for shipping preserved containers. (GC chemist) Analyze Gas, BTEX, Herbicides, and PNAs. Interpret raw data and providing report information. (Extraction chemist) Train existing and entry-level personnel in environmental sample extraction procedures. Inventory, order, and prepare stock standards and solutions. Perform specialized extraction procedures. Routinely used perform different extraction methods to comply with different program requirements including; SW846, CLP, Calif LUFT, Alaska and Wisconsin Diesel Range Organics (DRO) and TCLP extractions for semivolatile and volatile organic preparations. Member of CH2M Hill Laboratory Health & Safety Committee and confined space entry team.

Assistant Entomologist/Field Inspector; Shasta Mosquito Abatement District; Redding, California. Identify adult and larval forms of the mosquito population for state reports. Conducted inspections on the effectiveness of control measures, lethal dose level (LDL) studies, and used maps and aerial photos to identify and locate mosquito sources.

Lab Assistant; Humboldt State University, Life Sciences Department; Arcata, California.

Lab Assistant; Shasta College, Life Sciences Department; Redding, California.

Professional Development

Certifications

Certified Motorboat Operator,

40 Hr. *HAZWOPER* training,

"Working at a Watershed Level" Training

Instrumentation

HACH Spectrometers, Horiba Multiparameter Unit, Various Ground Water Pumps, Hydro lab, D.O. Meter, Sontek Velocity Meters, HPLC, Gel permeation instrument (ABS model 1000), Continuous liquid/liquid extractors, Zero headspace TCLP extractors (Millipore), CP Thermister, Turbidimeter, Graphite furnace, Sparger units (Tekmar), Sonic Disrupters, Analytical balances, CAM extractors, pH meter, GPS, GCs (Varian models 6000, 3400, 3400 STAR, 3600 and 3700), GC Autosamplers/Detectors - FID, ECD, PID, IR, Fluorometer, and UV Instrumentation (models 8000 and 8010), Nitrogen delivery systems, Kuderna Danish Apparatus,

EXHIBIT B

Revised Assessment of Detection and Quantitation Approaches

October 2004

Engineering and Analysis Division
Office of Science and Technology
Office of Water (4303T)
U.S. Environmental Protection Agency
1200 Pennsylvania Avenue, NW
Washington, DC 20460

EPA-821-B-04-005

October 2004

Disclaimer

This document has been reviewed and approved for publication by the Engineering and Analysis Division, Office of Science and Technology. Neither the United States Government nor any of its employees, contractors, or their employees make any warranty, expressed or implied, or assumes any legal liability or responsibility for any third party's use of or the results of such use of any information, apparatus, product, or process discussed in this report, or represents that its use by such party would not infringe on privately owned rights.

Table of Contents

FOREWORD ..... *i*

CHAPTER 1: INTRODUCTION 1-1
- 1.1 Background
- 1.2 Clause 6 Settlement Agreement Requirements
- 1.3 EPA's Approach to Conducting this Assessment
- 1.4 Peer Review of the Agency's Assessment
- 1.5 Proposal and Request for Public Comments
- 1.6 Terminology used in this Document

CHAPTER 2: OVERVIEW AND HISTORY OF DETECTION AND QUANTITATION LIMIT APPROACHES 2-1
- 2.1 Currie's Call for Standardization
- 2.2 Development of the MDL and ML as Practical Embodiments of Currie's Proposal
- 2.3 Other Detection and Quantitation Approaches

CHAPTER 3: ISSUES PERTAINING TO DETECTION AND QUANTITATION 3-1
- 3.1 Analytical Chemistry Approaches to Detection and Quantitation
- 3.2 CWA Regulatory Issues Affecting Detection and Quantitation
- 3.3 Statistical Issues

CHAPTER 4: EVALUATION CRITERIA 4-1
- 4.1 Criterion 1
- 4.2 Criterion 2
- 4.3 Criterion 3
- 4.4 Criterion 4
- 4.5 Criterion 5
- 4.6 Criterion 6
- 4.7 Consensus Principles

CHAPTER 5: ASSESSMENT 5-1
- 5.1 Detection Limit Approaches
- 5.2 Quantitation Limit Approaches

CHAPTER 6 FINDINGS AND NEXT STEPS 6-1

APPENDIX A: LITERATURE SEARCH REGARDING DETECTION AND QUANTITATION LIMIT APPROACHES A-1

APPENDIX B: COMPUTATION OF DETECTION AND QUANTITATION LIMITS B-1

APPENDIX C: EXAMPLE CALCULATIONS C-1

Foreword

EPA has assessed current procedures for determining the sensitivity of test methods and their application to Clean Water Act (CWA) Programs. The assessment was required by a settlement agreement with the Alliance of Automobile Manufacturers, *et al.* We announced the availability of our preliminary assessment for public comment on March 12, 2003. This assessment discussed statistical, chemical, and regulatory issues related to detection and quantitation and different approaches to detection and quantitation. The Agency has revised the preliminary assessment Document to incorporate public comment on that assessment.

In a related action on March 12, 2002, we proposed to revise EPA's method detection limit (MDL) definition and procedure, and codify our minimum level (ML) procedure. The MDL and ML, respectively and in order of increasing magnitude, are the EPA's embodiment of a detection and a quantitation limit.

In this revised assessment, we have:

- Explained why and how we conducted this assessment (Chapter 1),
- Identified relevant concepts to include in the assessment (Chapter 2 of this document),
- Identified issues that may be relevant to the assessment from an analytical chemistry, statistical, or regulatory perspective (Chapter 3),
- Used six criteria to evaluate the ability of each procedure or concept to support activities under the Clean Water Act (Chapter 4),
- Assessed how well each concept meets the evaluation criteria (Chapter 5),
- Summarized our findings and discussed next steps (Chapter 6), and
- With real-world data and several different procedures, calculated and compared detection and quantitation limits, and evaluated the theoretical and practical limitations of each concept (Appendices).

Public comment on the preliminary assessment and the proposed regulatory revisions expressed many divergent views that conflicted with the proposed revisions. Commenters noted that: (1) the MDL does not adequately address analytical variability or systematic error (bias); (2) the MDL does not always achieve a one percent (1%) false positive rate; (3) EPA should provide better guidance on the intended use of the MDL and ML in compliance reporting; and (4) the MDL and ML are not appropriate for all applications in CWA programs. Several commenters expressed support for two alternatives to the MDL and ML that were submitted by a laboratory association and the U.S. Geological Survey, respectively. Although none of the alternative procedures recommended by commenters fully satisfied EPA's needs under the CWA, several procedures contain steps, such as blank correction, that EPA believes warrant further consideration. There was no agreement among commenters as to which of the competing alternatives or revisions to adopt. Commenters suggested that we work together to discuss mutual concerns and possible solutions rather than proceed with the proposed revisions. We agree and recognize that these concerns provide a strong starting point for a continued dialog with stakeholders.

Based on this new information, it is clear that there is a broad interest in improving current procedures and uses, but no consensus for a specific procedure or procedures has emerged among the laboratory, industry, regulatory or regulated communities. In addition, EPA sees merit in alternative procedures suggested by commenters; however, none of these completely satisfy EPA's needs. Thus, we believe that it is appropriate to withdraw the March 2003 proposed revisions, take final action on the 2003 assessment to complete the terms of the settlement agreement, and obtain additional stakeholder

input. In a *Federal Register* notice published on September 15, 2004 [69 FR 55547], we announced that a neutral party is exploring the feasibility of a process by which a broad group of stakeholders would work together to define and address concerns about the way detection and quantitation limits are calculated and used to support CWA programs. This stakeholder process would include stakeholders representing constituencies such as citizens, environmental organizations, permit writers, regulators and regulated industries. We trust that this stakeholder process will address the wide variety of views held by stakeholders and lead to recommendations for possible improvements to current EPA procedures and/or use of alternative procedures.

To facilitate open discussion and consideration of issues, we have made every effort to ensure that this Revised Assessment Document does not prejudge the result of a future stakeholder process. We look forward to further stakeholder participation in this process.

Chapter 1
Introduction

1.1 Background

On June 8, 1999 (64 FR 30417), EPA promulgated (i.e., published in a final rule) Method 1631B: *Mercury in Water by Oxidation, Purge and Trap, and Cold Vapor Atomic Fluorescence Spectrometry* (the "method") for use in EPA's Clean Water Act programs. The method was developed specifically to measure mercury at ambient water quality criteria levels and includes a method detection limit (MDL; see 40 CFR part 136, Appendix B) of 0.2 nanograms per liter (ng/L).

Following promulgation, a lawsuit was filed challenging EPA on the validity of the method. The basis of the challenge included several specific aspects of Method 1631 as well as the general procedures used to establish the MDL and minimum level of quantitation (ML) published in the method. In order to settle the lawsuit, EPA entered into a settlement agreement (the "Settlement Agreement") with the Alliance of Automobile Manufacturers, Inc., the Chemical Manufacturers Association, and the Utility Water Act Group (collectively the "Petitioners") and the American Forest and Paper Association ("Intervenor") on October 19, 2000. Under Clause 6 of the Settlement Agreement, EPA agreed to perform an assessment of detection and quantitation limit concepts. The complete text of Clause 6 is provided in Exhibit 1-1 of this chapter. A summary of Clause 6 is provided in Section 1.2. The summary is followed by a description of EPA's approach to the assessment, including the material and data evaluated (Section 1.3), the use of an independent peer review to evaluate the Agency's assessment (Section 1.4), and EPA's March 2003 publication of and request for comment on the February 2003 assessment, and a related proposal concerning potential changes to detection and quantitation limit procedures approved for use under the Clean Water Act (Section 1.5). A brief discussion of the terminology used in this document is provided in Section 1.6.

1.2 Clause 6 Settlement Agreement Requirements

Clause 6 of the Settlement Agreement is titled *Reassessment of Method Detection Limit and Minimum Level Procedures*. Clause 6 consists of five subclauses, a - b and d - f. (There is no subclause c.)

1.2.1 Clause 6a

Clause 6a broadly defines the scope of the assessment and provides a schedule for completing the initial phase. Specifically, Clause 6a requires EPA to:

- Sign and forward to the Office of Federal Register (OFR) a notice inviting public comment on a reassessment of existing EPA procedures for determining the detection and quantitation limits of contaminants in aqueous samples.
- Forward the notice to the OFR on or before February 28, 2003.
- Provide a period of at least 120 days for public comment on the notice.

- At a minimum, include the MDL procedure published at 40 CFR part 136, Appendix B, and the ML procedure described in Section 17.8 of Method 1631B, in the reassessment of detection and quantitation limits.
- Invite comment on one or more alternative procedures for determining and describing test sensitivity.

Clause 6a also provides EPA with the option of proposing modifications to the existing procedures.

1.2.2 Clause 6b

Clause 6b requires that EPA obtain a peer review of its reassessment, and describes six specific topics that must be included in the charge to the peer reviewers. Specifically, Clause 6b requires EPA to:

- Submit the reassessment of existing procedures (including any proposed modifications thereof) and any evaluation of alternatives for peer review by experts in the field of analytical chemistry and the statistical aspects of analytical data interpretation.
- Conduct the peer review in accordance with EPA's peer review policies.
- Prepare a charge to the peer review panel that requests the peer reviewers to consider:
  - Criteria for selection and appropriate use of statistical models
  - Methodology for parameter estimation
  - Statistical tolerance and prediction
  - Criteria for design of detection and quantitation studies, including selection of concentration levels ("spiking levels")
  - Interlaboratory variability, and
  - Incorporation of elements of probability design.

1.2.3 Clause 6d

Clause 6d requires EPA to provide the Petitioners and Intervenor (the "litigants") with an opportunity for review of the Agency's assessment concurrent with the Clause 6b peer review.

1.2.4 Clause 6e

Clause 6e requires EPA to provide the litigants with:

- An opportunity to meet periodically (i.e., every six months) to discuss the Agency's progress during development of the assessment,
- A plan for performing the assessment on or before the second of these meetings, and
- Copies of relevant documents, where appropriate, in advance of these meetings.

1.2.5 Clause 6f

Clause 6f establishes a schedule and requirements concerning final action on the notice described in Clause 6a. Specifically:

- On or before September 30, 2004 (since amended to November 1, 2004), EPA is to sign and forward to the OFR a notice taking final action on the notice described in Clause 6a, and
- Coincident with publication of this notice of final action, EPA is to provide the litigants with an opportunity to meet and discuss the implications of the final notice and/or the need for any subsequent EPA action in light of the final notice.

Exhibit 1-1. Full Text of Clause 6 of the Settlement Agreement

6. Reassessment of Method Detection Limit and Minimum Level Procedures

a. On or before February 28, 2003, EPA shall sign and forward to the Office of the Federal Register for prompt publication a notice inviting public comment on a reassessment of the existing Agency procedures for determination of sensitivity of analytic test methods for aqueous samples, specifically, EPA procedures for determining the detection limits and levels of quantitation of contaminants in aqueous samples, including, at a minimum, the "Definition and Procedure for Determination of the Method Detection Limit" published at 40 C.F.R. Part 136, Appendix B, as well as the "minimum level" procedures, which is described in section 17.8 of Method 1631B. The notice shall invite comment on EPA's evaluation of one or more alternative procedures for determining and describing test sensitivity. The notice also may propose modifications to the existing procedures. The notice shall invite public comment for a period of no less than one hundred twenty (120) days.

b. Prior to publishing the notice inviting public comment on EPA procedures for determining test sensitivity, EPA shall submit its reassessment of existing procedures (including any proposed modifications thereof) and its evaluation of alternatives for peer review by experts in the field of analytical chemistry and the statistical aspects of analytical data interpretation. In its charge to the peer review panel, EPA shall request that the peer review consider: criteria for selection and appropriate use of statistical models; methodology for parameter estimation; statistical tolerance and prediction; criteria for design of detection and quantitation studies, including selection of concentration levels ("spiking levels"); interlaboratory variability; and incorporation of elements of probability design. EPA (or its authorized representative) shall conduct the peer review in accordance with EPA's current peer review policies in the January 1998 Science Policy Council Handbook (EPA 100-B-98-00) *[sic]*, including any subsequently-developed EPA peer review documents that may revise or amend that Handbook.

[Note - the correct document number for the Science Policy Council Handbook is EPA 100-B-98-001]

[c. ***Note - there is no clause "6.c" in the Settlement Agreement***]

d. During the peer review period, EPA shall also provide an opportunity for concurrent review and comment by the Petitioners and Intervenor.

e. In the development of the reassessment/assessment of alternatives, EPA shall provide the Petitioners and Intervenor with a periodic opportunity to meet (i.e., every six (6) months) on the Agency's progress. EPA shall prepare and present the Petitioners and Intervenor with the Agency's "plan" for conducting the reassessment/assessment of alternatives on or before the second such periodic meeting. Where appropriate, EPA shall provide the Petitioners and Intervenor with copies of relevant documents in advance of such meetings.

f. On or before September 30, 2004 (Note: since amended to November 1, 2004), EPA shall sign and forward to the Office of the Federal Register for prompt publication a notice taking final action on the notice described in subparagraph 6.a. Coincident with publication of the notice of final action, EPA shall provide Petitioners and Intervenor an opportunity to meet to discuss the implications of the final notice and/or the need for any subsequent EPA action in light of the final notice.

1.3 EPA's Approach to Conducting this Assessment

This document details the Agency's assessment of methodology for the determination of method sensitivity, specifically: detection and quantitation limits. This assessment is being conducted in accordance with a plan summarized in Section 1.3.1 and is based, in part, on an assessment of the data described in Section 1.3.2.

1.3.1 Study Plan

EPA developed a technical approach for 1) conducting the assessment, and 2) complying with all applicable requirements of the Settlement Agreement. The approach was documented in a draft study plan that has since formed the general framework for the assessment described in this Assessment Document. EPA also conducted a literature search to identify and review issues and concepts that should be considered when developing the plan. A summary of this literature review is provided in Appendix A to this Assessment Document.

The study plan described roles and responsibilities for implementing the plan, provided a background discussion of detection and quantitation limit concepts, including the MDL and ML, and outlined a series of 11 events associated with the Agency's assessment of detection and quantitation limit approaches. The relationship between those planned events and this Assessment Document is summarized in Exhibit 1-2 at the end of this chapter.

Although the Settlement Agreement did not require that EPA seek formal peer review on its draft plan, the Agency chose to conduct a peer review of the draft plan. The peer review was initiated in December 2001, conducted in accordance with EPA's peer-review policies, and performed by two statisticians and two chemists. EPA reviewed the comments and recommendations offered by these reviewers, and where appropriate, revised the plan to reflect the peer-review comments. EPA also reviewed, and where appropriate, revised the plan to reflect comments provided by the petitioners following their concurrent review.

1.3.2 Material and Data used in the Assessment

In order to perform the assessment described in this document, EPA sought to collect documentation describing existing detection and quantitation limit concepts and procedures and data that could be used to evaluate these concepts and procedures.

Documentation concerning the existing concepts and procedures was obtained by performing a literature search as described in Appendix A to this Assessment Document, and where appropriate, by purchasing copies of documents describing concepts or procedures from the organizations that published them.

In performing this assessment, EPA hoped to identify a substantial amount of data containing results of direct relevance to the determination of detection and low-level measurement capability. That is, measurement results in the low concentration region. To date, EPA has been able to identify only six data sets that were of use in fully evaluating variability in the range of analytical detection and quantitation. Three of the six were developed by EPA for the express purpose of studying the relationship between measurement variation and concentration across a wide variety of measurement techniques and analytes. EPA refers to these data sets as "EPA's ICP/MS Study of Variability as a Function of Concentration," "EPA's Multi-technique Variability Study" (also referred to as the "Episode

6000 study"), and "EPA's GC/MS Threshold Study" (also referred to as "the Episode 6184 study"). In all three cases, replicate measurement results from each combination of analyte and measurement technique were produced by a single laboratory over a wide range and large number of concentrations. The fourth data set was developed by the American Automobile Manufacturer's Association (AAMA) for the purpose of estimating one particular kind of quantitation value. That quantitation value is called an alternative minimum level (AML; see Gibbons *et al.*, 1997). In the AAMA study, replicate results were measured at a limited number of concentrations by multiple laboratories using EPA Method 245.2 (cold vapor atomic absorption; CVAA) for mercury and EPA Method 200.7 (inductively coupled plasma/atomic emission spectroscopy; ICP/AES) for twelve other metals. The final two data sets were jointly gathered by EPA and the Electric Power Research Institute (EPRI) to support interlaboratory validation of EPA Methods 1631 and 1638.

The studies from which these six data sets were obtained are summarized in sections 1.3.2.1 - 1.3.2.6 below. Additional information about these studies can be found in Appendices B and C to this Assessment Document.

In March 2003, EPA published an Assessment Document dated February 2003, and requested comments on the assessment and additional data to support continued evaluation of detection and quantitation limits. Three stakeholders commenting on the assessment also offered to provide EPA with data that would substantiate their views or aid EPA in further evaluating detection and quantitation procedures. These data are further described in Sections 1.3.2.7 - 1.3.2.8 and Section 1.3.3 below.

Although the petitioners offered specific suggestions for other data sets that they believed should be considered in this assessment, EPA found that these data sets did not include a sufficient number of results in the region of detection and quantitation to yield information for the assessment, overlapped with data already used in the assessment, or exhibited signs of significant contamination that made the data inappropriate for inclusion in the assessment. These data, and EPA's decisions regarding the data, are discussed in Section 1.3.3 below.

1.3.2.1 EPA's ICP/MS Study of Variability as a Function of Concentration

The objective of the ICP/MS study was to characterize variability as a function of concentration using EPA's draft Method 1638 for determination of nine metals by inductively coupled plasma with mass spectroscopy (ICP/MS). The nine metals were silver, cadmium, copper, nickel, lead, antimony, selenium, thallium, and zinc. The ICP/MS instrument used in this study averages triplicate scans to produce a single measurement of each element at each concentration. Such averaging is typical of ICP/MS design and use.

In preparation for the study, the ICP/MS was calibrated using triplicate scans averaged to produce a single measurement of 100, 1,000, 5,000, 10,000, and 25,000 nanograms per liter (ng/L) for each element. Originally, the instrument was calibrated using unweighted least squares estimates under the assumption of linearity. Subsequently, the analytical results were adjusted with weighted least squares estimates. Weighted least squares estimates are based on the knowledge that variability (expressed as the standard deviation) increases with increasing analyte concentration.

Although the instrumentation has the capability to provide intensity results for each of the three scans at each concentration, averaging the three scans to produce a single measurement is the normal operating mode, and the average was used to produce the measurements in this study. Draft Method 1638 specifies the use of average response factors rather than least squares estimation of a linear calibration, although it does allow for the use of such procedures.

All nine metals were spiked into reagent water to produce solutions at concentrations of: 0, 10, 20, 50, 100, 200, 500, 1,000, 2,000, 5,000, 10,000, and 25,000 ng/L. Each solution was divided into seven replicate aliquots for subsequent analysis. The aliquots were analyzed beginning with the blank (zero concentration) followed by analyses from the highest to the lowest concentration. This sequence was chosen to minimize carry-over effects and to allow the analyst to stop at the concentration that returned zero results. Carry-over is caused by residual sample remaining in the inlet system of the instrument, in this case, the ICP/MS. Carry-over can occur when analysis of a high-concentration sample is followed by analysis of a relatively low-concentration sample, as could occur if the replicates were analyzed in random order. Use of the highest to lowest analytical sequence ensured that each successive concentration analyzed was close enough to the previous concentration that any effects of carryover would be negligible and, therefore, would not compromise study results. (A more in-depth discussion of the randomized design and the effects of carry-over issues is provided in Chapter 3, Section 3.3.8.2).

Results at multiple mass-to-charge ratios, or m/z's, were reported for each metal, although draft Method 1638 specifies only one m/z for eight of the nine metals. For lead, m/z's 206, 207, and 208 are specified. Only data associated with m/z's specified in draft Method 1638 were used in the ICP/MS study.

1.3.2.2 EPA's Multi-technique Variability Study (the "Episode 6000 Study")

In 1997 and 1998, EPA conducted a study of variability vs. concentration for a number of analytical methods. Five laboratories were employed for the analyses; each analyte and method combination was tested by one of these laboratories. Details of the study design are described in EPA's *Study Plan for Characterizing Variability as a Function of Concentration for a Variety of Analytical Techniques* (July 1998). Based on the sampling episode number assigned to the study by the EPA Sample Control Center, the study and results have become known as the Episode 6000 study and data. The analytes and analytical techniques studied were:

- Total suspended solids (TSS) by gravimetry
- Metals by graphite furnace atomic absorption spectroscopy (GFAA)
- Metals by inductively-coupled plasma atomic emission spectrometry (ICP/AES)
- Hardness by ethylene diamine tetraacetic acid (EDTA) titration
- Phosphorus by colorimetry
- Ammonia by ion-selective electrode
- Volatile organic compounds by purge-and-trap capillary column gas chromatography with a photoionization detector (GC/PID) and electrolytic conductivity detector (GC/ELCD) in series
- Volatile organic compounds by gas chromatography with a mass spectrometer (GC/MS)
- Available cyanide by flow-injection/ligand exchange/amperometric detection
- Metals by inductively-coupled plasma spectrometry with a mass spectrometer (ICP/MS)

In this study, an initial (range finding) MDL was determined for each combination of analyte and analytical technique using minor modifications to the MDL procedure at 40 CFR part 136. Specifically, the modifications made the optional iterative step 7 of the MDL procedure mandatory and required the spike concentration to be no more than a factor of three times the determined MDL (instead of a factor of five times). During the study, however, two of the laboratories found that the reduction in the allowable spike range necessitated an unreasonably large number of iterations. In continuing the study, EPA returned to the spike-to-MDL ratio of five published in the 40 CFR part 136, Appendix B procedure.

After determining the initial MDL, each laboratory analyzed 7 replicate samples spiked at concentrations that were 100, 50, 20, 10, 7.5, 5.0, 3.5, 2.0, 1.5, 1.0, 0.75, 0.50, 0.35, 0.20, 0.15, and 0.10 times the initial MDL. In a few instances, laboratories analyzed more than 7 replicates. As often as possible, the replicate analyses at each concentration level were produced using the same calibration that was used in determining the initial MDL. Where laboratory reports indicated that multiple calibrations were conducted, each result was associated with its calibration in the data analysis.

Spiked aqueous solutions were analyzed in order from the highest concentration (100 times the MDL) to the concentration at which 3 or more non-detects (zeros) were encountered among the 7 replicates, or the lowest concentration specified (0.1 times the MDL), whichever occurred first. This analysis order (1) minimized carryover that could occur in some methods if a low-concentration sample had followed a high-concentration sample (as may happen when samples are analyzed in random order), and (2) prevented collection of a large number of zeros if the signal disappeared.

For methods that do not produce a signal for a blank, the signal will disappear somewhere below the MDL, i.e., a zero will be reported. Laboratories were instructed that when three nondetects (out of seven measurements) were reported, it was not necessary to move to the next lower concentration, because it would be of no practical value to have laboratories measure seven zeros, move to a lower level, measure seven zeros, etc.

A variant of the iterative procedure for determining the MDL was used for organic compounds determined by chromatographic methods. Methods for organics normally list many (15 to 100) analytes, and the response for each analyte is different. Therefore, to determine an MDL for each analyte, the concentration of the spike would need to be inversely proportional to the response. Making a spiking solution with 15 to 100 different concentrations is cumbersome and error prone. The approach used in the study was to run seven replicates at decreasing concentrations until signal extinction, then select the concentration(s) appropriate for the determining the MDL for each analyte according to the MDL procedure. In some cases, the laboratories selected the concentrations, in others cases, EPA did. This approach was generally applied for organics analysis. However, laboratories also had the option of using some combination of the monotonically decreasing concentrations described above and a few selected concentrations to achieve the desired spiking levels.

1.3.2.3 EPA's GC/MS Threshold Study (the "Episode 6184 Study")

Data from the Episode 6184 study of variability vs. concentration were used to evaluate the effect of GC/MS thresholds on the ability to identify semivolatile organic compounds at low concentrations. Details of the design of this study are described in EPA's *Study Plan for Characterizing Error as a Function of Concentration for Determination of Semivolatiles by Gas Chromatography/Mass Spectrometry* (December 1998). Data were generated for 82 semivolatile organic compounds using EPA Method 1625C (semivolatile organic compounds by GC/MS). MDLs were not determined for these compounds. Instead, solutions of the analytes were prepared and analyzed at concentrations of 50.0,

20.0, 10.0, 7.50, 5.00, 3.50, 2.00, 1.50, 1.00, 0.75, 0.50, 0.35, 0.20, 0.15, 0.10, 0.075 and 0.050 ng/μL (or μg/mL). Each solution was injected into the GC/MS in triplicate with the mass spectrometer threshold set to zero, and again in triplicate with the mass spectrometer threshold set to a level typical of that used in routine environmental analyses. As with the ICP/MS study and the Episode 6000 study, and for the same reasons described in Section 1.3.2.1, samples were analyzed in order from the highest to the lowest concentration.

1.3.2.4 AAMA Metals Study of Methods 200.7 and 245.2

The American Automobile Manufacturers Association conducted an interlaboratory study of EPA Method 200.7 (metals by ICP/AES) and Method 245.2 (mercury by CVAA). The study was designed to estimate a quantitation value based on a concept termed the alternative minimum level (AML) that had been described in the literature (Gibbons *et al.*, 1997). Nine laboratories participated in the study, and each reported data for the following 13 metals: aluminum, arsenic, cadmium, chromium, copper, lead, manganese, mercury, molybdenum, nickel, selenium, silver and zinc. Study samples were analyzed by EPA Method 200.7 for 12 of the metals. Mercury was determined by EPA Method 245.2.

As part of the study design, the nine laboratories were randomized prior to the start of the study. Five sample matrices (including reagent water) were studied, including four wastewater matrices that are representative of the automotive industry. Starting from a blank, or unspiked sample, all target analytes were spiked at four concentrations to yield a total of five concentrations per matrix. Concentrations ranged from 0.01 to 10 μg/L for mercury and selenium on the low end, and from 2.0 and 1000 μg/L for mercury and selenium on the high end. In addition, the concentrations were matrix-dependent. The same concentration ranges for each metal by matrix combination were used for all five weeks of the study.

Matrix A (reagent water) was analyzed in all nine laboratories, and three laboratories analyzed each of the other four matrices. All analyses were repeated weekly over a five-week period. As a result, a total of 6,825 observations were obtained, which includes 2,925 observations for matrix A (9 labs × 13 metals × 5 spike concentrations × 5 weeks), and 975 observations (3 labs × 13 metals × 5 spike concentrations × 5 weeks) for each of the other four matrices (6,825 = 2,925 + (975 × 4)). There were two missing values for chromium in matrix A from laboratories 1 and 9.

1.3.2.5 Method 1631 Interlaboratory Validation Study

The Method 1631 interlaboratory validation study was conducted by EPA to evaluate performance of the method and to gather data to evaluate existing performance specifications, including detection and quantitation limits. To accommodate stakeholder interests and expand the scope of the study, the Electric Power Research Institute (EPRI) funded the distribution of additional samples to study participants.

This jointly funded study involved an international community of twelve participating laboratories and one referee laboratory. Each participating laboratory analyzed four different matrices, each containing mercury at a concentration selected to allow for characterization of method performance across the measurement range of the method. Each of the 12 participating laboratories was provided with 13 sample pairs (a total of 26 blind samples). These included 1 filtered effluent pair, 1 unfiltered effluent pair, 4 filtered freshwater pairs, 1 filtered marine water pair, 1 unfiltered marine water pair, and 5 spiked reagent water pairs. All 12 laboratories received and analyzed the same sample pairs (a total of 312 analyses). To measure the recovery and precision of the analytical system, and to monitor matrix interferences, the laboratories were instructed to analyze matrix spike and matrix spike duplicate samples

on specified field samples for each filtered and unfiltered matrix, spiked at 1-5 times the background concentration of mercury determined by analysis of an unspiked aliquot of the sample. The laboratories were instructed to perform all other QC tests described in Method 1631, including the analysis of blanks, and to conduct MDL studies in reagent water following the procedure at 40 CFR part 136.

1.3.2.6 Method 1638 Interlaboratory Validation Study

The Method 1638 interlaboratory validation study was conducted by EPA to evaluate performance of the method and to gather data that would allow revision of existing performance specifications, including detection and quantitation limits. To accommodate stakeholder interests and expand the scope of the study, the Electric Power Research Institute funded the distribution of additional samples to study participants.

A total of eight laboratories (and a referee laboratory) participated in the study. The study was designed so that each participating laboratory would analyze sample pairs of each matrix of interest at concentrations that would span the analytical range of the method. Each laboratory was provided with 11 sample pairs (a total of 22 blind samples). These included 1 filtered effluent pair, 1 unfiltered effluent pair, 4 filtered freshwater pairs, and 5 spiked reagent water pairs. All eight laboratories received and analyzed the same sample pairs (a total of 176 analyses). To measure the recovery and precision of the analysis, and to monitor matrix interferences, the laboratories were instructed to analyze a matrix spike and matrix spike duplicate of specified field samples in each filtered and unfiltered matrix, spiked at 1-5 times the background concentration of the analytes determined by analysis of an unspiked aliquot of the sample. The laboratories were instructed to perform all other QC tests described in Method 1638, including the analysis of blanks, and to conduct MDL studies in reagent water following the procedure at 40 CFR part 136.

1.3.2.7 American Council of Independent Laboratories Data

The American Council of Independent Laboratories (ACIL) is a trade association representing independent, commercial scientific and engineering firms. Its members are professional services firms engaged in testing, product certification, consulting, and research and development. On behalf of its membership, ACIL submitted comments on EPA's proposal. To substantiate their comments, ACIL provided EPA with data summary tables consisting of blank analyses used to calculate detection limits. The data provided were performed by a single laboratory using Method 200.7 for five analytes. Because only blank sample analyses were available, not all detection and quantitation limit procedures could be assessed using the data. However, comparisons of the detection limit procedures submitted by ACIL and the US Geological Survey were performed based on these data and are discussed in Appendix C. In addition, because blanks were analyzed approximately two to three times per week, a comparison of long-term to short-term variability was also performed using these blank data. ACIL also submitted an alternative procedure for estimation of a detection limit, which is summarized in sect. 2.3.3 of this document.

1.3.2.8 U.S. Geological Survey Method Detection Limit Data

To assist EPA's assessment of their long-term MDL (LT-MDL) procedure, the US Geological Survey (USGS) provided data from blank sample analyses. These data represented a combination of 78 metals, methods and matrices, and were analyzed approximately twice per month. Unlike the blank data provided by ACIL, these blanks were collected in the field, and, therefore, include more sources of variability. As with the ACIL data, it was not possible to assess all detection and quantitation limit

procedures using the blank data set because some procedures require use of samples spiked at one or more concentrations. The LT-MDL procedure is summarized in sect 2.3.4 of this document.

USGS also submitted spiked sample results along with the blank data. These spikes were limited to a single concentration, and did not sufficiently characterize the region of interest to allow for full evaluation of detection and quantitation levels.

1.3.3 Data Considered but not Used in this Assessment

The Petitioners and Intervenor to the Settlement Agreement suggested ten specific data sets that EPA should consider in its assessment of detection and quantitation limits. EPA evaluated each of these data sets to determine if the design of the study, including the concentrations targeted in the study, would provide sufficient data for evaluating measurement variability in the region of interest (i.e., at concentrations below, at, and above the region of detection and quantitation). If such data were available, EPA further evaluated the data set to ensure that it was of sufficient quality to support the Agency's assessment. Four of the ten data sets met these requirements and were used in EPA's assessment. Table 1 identifies each of the data sets suggested by the petitioners along with a brief rationale for using or excluding the data from this assessment, additional discussion is in Appendix B.

After EPA published the February 2003 Assessment Document for comment, ACIL submitted data as described in Section 1.3.2.7 above, and three commenters offered to provide EPA with additional data that would enhance EPA's assessment. EPA requested the data offered by each of these organizations, but received a response from only two of the three (Laucks Testing Laboratories and USGS). After evaluating the data, EPA determined that the data from Laucks Testing Laboratories was not useful because it was incomplete. The Laucks data unfortunately did not include the data from extraction to detection which is needed to compare detection and quantitation approaches. Most of the data sent by USGS was useful and is described in Section 1.3.2.8.

Table 1. Data Sets Suggested by Petitioners and Commenters

Dataset Source and Year	Analytes and technology	EPA's Use of Datasets
AAMA 1996-1997	Metals by ICP/AES (200.7)	Used in this assessment and described in Section 1.3.2.4
AAMA 1996-1997	Mercury by CVAA (245.2)	Used in this assessment and described in Section 1.3.2.4
EPA/EPRI 1997-1998	Mercury by CVAF (1631)	Used in this assessment and described in Section 1.3.2.5
EPA/EPRI 1997-1998	Metals by ICP/MS (1638)	Used in this assessment and described in Section 1.3.2.6
ACIL 2002-2003	Metals by ICP/AES (200.7)	Used in this assessment and described in Section 1.3.2.7
USGS 2002-2003	Metals by ICP/MS and GFAA	Used in this assessment and described in Section 1.3.2.8
EPRI 1987	Metals by GFAA (EPA 200)	Not used in this assessment because of insufficient low-level data
EPRI 1990	Metals by ICP/AES (EPA 200.7)	Not used in this assessment because of insufficient low-level data

Dataset Source and Year	Analytes and technology	EPA's Use of Datasets
EPRI 1994	Al, Be, Tl by GFAA (EPA 200)	Not used in this assessment because of overlap with EPA's Episode 6000 Study, which provides data on the same analytes but covers a larger number of concentrations in the region of interest
AAMA 1996-1997	PCBs by GC/ECD (608.2)	Not used in this assessment because of overlap with EPA's Episode 6000 Study, which provides data on the same analytes but covers a larger number of concentrations in the region of interest
EPRI 1996	Cd, As, Cr by GFAA (EPA 200)	Not used in this assessment because of overlap with EPA's Episode 6000 Study, which provides data on the same analytes but covers a larger number of concentrations in the region of interest
MMA 2000-2001	Aroclors 1016 and 1260 by GC/ECD	Not used in this assessment because the maximum number of replicates (5) in the dataset, is less than the minimum number required (7) to calculate an MDL. Samples spiked with low levels of Aroclors exhibited average recoveries >500%, across 10 laboratories.
Laucks Testing Laboratory 2003	Mercury, 2,4-Dinitrophenol, Hexachlorocyclo-pentadiene, 4-Nitrophenol (OSW 8270)	Not used in this assessment because the dataset was incomplete. Data included only calibration data and not the extraction to detection data needed to compare det/quant procedures.

1.4 Peer Review of the Agency's Assessment

In August 2002, EPA conducted a formal peer review of the Agency's assessment. This peer review, which satisfied requirements in Clause 6b of the Settlement Agreement, was conducted in accordance with EPA's peer review policies described in the Science Policy Council Handbook (EPA 100-B-00-001). The review was performed by two experts in the field of analytical chemistry and two experts in the statistical aspects of analytical data interpretation. Each reviewer was provided with a draft version of this Assessment Document, which documented the Agency's approach to the assessment and the Agency's preliminary findings and conclusions. Reviewers also were provided with copies of all data evaluated in the assessment, statistical programs used to analyze the data, and copies of the detection and quantitation concepts and procedures evaluated by EPA. In accordance with the Agency's peer review policies, the reviewers were provided with a written 'charge' intended to ensure the evaluation would meet EPA needs.

In its charge to the peer reviewers, EPA requested a written evaluation of whether the assessment approach described by EPA is valid and conceptually sound. Reviewers also were asked to consider and address eight specific questions pertaining to the adequacy of the concepts and issues considered, the evaluation criteria developed by EPA, EPA's assessment and conclusions, the data used to perform the assessment, suggested improvements to the procedures discussed, and EPA's consideration of interlaboratory vs. intralaboratory issues. Comments from peer reviewers were generally supportive of EPA's assessment and its presentation of the assessment in the Assessment Document. Where appropriate, EPA revised that Assessment Document to reflect specific suggestions and comments offered by the peer reviewers. The revised version of the Assessment Document, reflecting peer reviewer comments, was completed in February 2003, and made available through a public notice on March 12, 2003 (see section 1.5 below). Copies of all materials associated with the peer review,

including the peer review charge, the materials provided to the peer reviewers for review, complete copies of the peer reviewers' comments, and detailed EPA responses to each of the comments were provided in the public docket supporting the Agency's March 2003 assessment.

1.5 Proposal and Request for Public Comments

On February 28, 2003, the EPA Administrator signed two notices for publication in the Federal Register. These notices fulfilled EPA's obligations under Clause 6(a) of the Settlement Agreement and were published in the Federal Register on March 12, 2003.

The first of these notices announced the availability of EPA's assessment of detection and quantitation procedures that are applied to analytical methods used under the Clean Water Act. It also announced that results of the assessment could be found in the "Technical Support Document for the Assessment of Detection and Quantitation Concepts" (EPA 821-R-03-005, February 2003), requested public review and comment on the assessment. The full text of this notice was published at 68 FR 11791, March 12, 2003.

The second notice requested comment on proposed revisions to the detection and quantitation definitions and procedures at 40 CFR part 136. The proposed changes were based on the assessment and on stakeholder comments received over the years. The full text of this notice was published at 68 FR 11770, March 12, 2003.

1.5.1 Summary of Changes Proposed in March 2003

EPA proposed a number of technical and editorial changes to the definitions specified at 40 CFR 136.2 and to the procedure specified at 40 CFR 136, Appendix B. A detailed description of those changes can be found in the March 12, 2003 public notice (68 FR 11770). Briefly, those proposed changes included:

- A revised definition of the term "detection limit" at 40 CFR 136.2(f) to explicitly equate the term with the "method detection limit" specified in 40 CFR 136, Appendix B; and a revised definition of the term "method detection limit" included in Appendix B to provide technical clarifications and more clearly equate the term with the "critical value" described by Currie (1968, 1995) and the Limit of Detection described by the American Chemical Society (Keith et al., 1980; McDougal et al., 1983). Those concepts are further described in Chapter 2 of this assessment document.
- An expanded Scope and Application discussion in the codified MDL procedure to recognize that there are a variety of purposes and analytical methods for which the MDL procedure may be employed. The proposed revisions provided examples of four common uses of the MDL procedure (*i.e.*, demonstrating laboratory capability with a particular method; monitoring trends in laboratory performance; characterizing method sensitivity in a particular matrix; and establishing an MDL for a new or revised method for nationwide use.) The proposed revisions also clarified that the procedure may not be applicable to certain test methods such as those used to measure pH or temperature.
- Proposed modifications to the considerations for estimating the detection limit in Step 1 of the codified MDL procedure and to the specifications for establishing the test concentration range in Step 3 of the codified procedure.
- Proposed deletion of the optional procedure for calculating a 95% confidence interval estimate for the MDL.

- Proposed changes to the iterative procedure to mandate its use when determining an MDL for a new or revised method or when developing a matrix-specific MDL, but allow it to remain optional when determining an MDL for other purposes, such as verifying lab performance.
- Proposed addition of a new procedural section to address the treatment of suspected outliers.
- Proposed deletion of the discussion of analysis and use of blanks included in Section 4(a) of the codified procedure.
- Proposed changes to the optional pre-test described in Section 4(b) of the procedure to improve the utility of results from this test.
- Editorial changes to the codified version of the MDL. Examples of these editorial changes include addition of a summary section, clarifications, reorganization of steps, simplified presentation of calculations, and deletion of the reporting section.
- Proposed addition of a definition of the ML at 40 CFR 136.2
- Proposed addition of a procedure (including a definition) of the ML to 40 CFR 136, Appendix B
- Explicit allowance of alternative detection and quantitation procedures, provided that the resulting detection and quantitation limits meet the sensitivity needs for the specific application. The objective of this proposed allowance was to provide greater flexibility in establishing or improving the sensitivity of methods for use under CWA and facilitate approval of analytical methods from other agencies or organizations that utilize alternate detection and quantitation concepts.

In addition to requesting comment on the assessment and the proposed revisions, EPA also specifically requested comment on several aspects of the proposal, including alternative actions that could have been taken. With respect to the ML, for example, EPA explicitly sought comment on the proposed addition of the ML definition to 40 CFR 136.2 and procedure to 40 CFR 136, Appendix B vs an alternative option of not incorporating the definition at 40 CFR 136.2, but instead continuing to specify the ML on a method-by-method basis. EPA encouraged commenters to support their views with data or information that would assist the Agency in making a final decision.

1.5.2 Impact of Comments on the Assessment

EPA provided a 120-day period following publication of the notices for submission of comments (from the date of publication of the notices to July 10, 2004). In response to requests from stakeholders, EPA re-opened this comment period on July 16, for an additional 30 days (68 FR 41988).

During the comment periods, EPA received comments from 126 individuals or organizations representing the diversity of the stakeholder community on this issue. They included 23 laboratories, 31 water treatment plants, 3 federal agencies, 11 state and county agencies, 23 industrial firms, 3 instrument manufacturers, 19 trade organizations, 4 consultants, 8 individuals, and the law firm representing the petitioners. Comments offered by these groups addressed more than 25 different issues. A complete summary of the comments and EPA's responses to those comments can be found in Appendix B to this Assessment Document. These comments are discussed at various locations throughout this document, and include discussion of:

- Additional detection and quantitation limit procedures suggested by commenters. (Chapters 2, 3, and 5)
- Public comments received on chemical, regulatory, and statistical issues, along with EPA's consideration of these issues in light of the comments received. (Chapter 3)
- Comments received on each of the evaluation criteria used in EPA's assessment and EPA's response to those comments. (Chapter 4)

- Potential process for additional stakeholder involvement on the evaluation of detection and quantitation limit procedures (Chapter 6)

Appendix C contains a detailed analysis of the detection and quantitation limit procedures evaluated through computation of limits using the data described in Section 1.3.2. This analysis has been revised to reflect new data and comments on the original version of the assessment, which was published as Appendix C to the February 2003 version of EPA's Assessment Document.

1.6 Terminology used in this Document

We use the term "quantitation" in this document because of its common usage among analytical chemists, even though we recognize that the term "quantification" (i.e., the act of quantifying) is the term listed in most dictionaries. Also, when referring to detection and quantitation, we use the words "approach" or "concept" to refer, generically, to the procedures used to establish detection and quantitation limits or the theories on which those procedures are based. We use the word "limit" rather than "level" to indicate that the detection and quantitation concepts are directed at the lowest concentration or amount at which an analyte is determined to be present (detection) or may be measured (quantitation). In choosing the word 'limit' we do not mean to imply any sense of permanence. We recognize that measurement capabilities generally improve over time, and that detection or quantitation 'limits' established today may be superseded by future developments in analytical chemistry.

Although the Settlement Agreement refers to the word "sensitivity" to describe detection and quantitation limits, we have avoided such use of the term "sensitivity" in this document because the term is widely used by analytical chemists to describe something other than detection and quantitation capabilities. Traditionally, analytical chemists have referred to the term "sensitivity" as meaning instrument signal units per concentration units, such as is given for a calibration slope or a response factor. For example, in ion selective potentiometry, the sensitivity is 59 millivolts per decade change in concentration for monovalent species and half that for divalent species. Sensitivity is a performance characteristic, but it differs from detection limits. For example, one might compare the sensitivity of instruments. Obtaining a sensitivity of 10,000 counts per ppb indicates a properly functioning Sciex 250, while a Perkin-Elmer 6000's sensitivity would be 100,000 counts per ppb. Another performance characteristic of sensitivity is that it may vary in an expected pattern as with mass to charge ratio in mass spectrometry or atomic number for x-ray fluorescence spectrometry.

Exhibit 1-2. Relationship of Assessment Document to Assessment of Detection and Quantitation Limit Approaches

Event 1, Develop a detailed plan for responding to Clause 6 the Settlement Agreement: This event was completed in April 2002 when the draft plan was revised to reflect peer review and Litigant comments.

Event 2, Identify and explore issues to be considered: The Settlement Agreement identified six specific issues that should be considered during the assessment of detection and quantitation limit concepts, and subjected to formal peer review. During development of the technical approach, EPA identified a number of other issues that should be considered during the assessment. EPA listed and described each of these issues in the study plan and noted that identification of issues is likely to be a dynamic process, in that as a suite of issues is identified and discussed, other issues may surface. Finally, EPA stated its intent to prepare an "issue paper" that fully explained and discussed each of the identified issues. Chapter 3 of this Assessment Document serves the function of the issue paper described in the plan.

Event 3, Develop criteria against which concepts can be evaluated: After fully considering all relevant issues, EPA developed a suite of criteria that could be used to evaluate the suitability of various detection and quantitation procedures for use in CWA programs. Chapter 4 of this Assessment Document provides and describes the criteria selected by EPA after its consideration of all pertinent issues.

Event 4, Evaluate existing procedures for establishing detection and quantitation levels: EPA evaluated existing detection and quantitation limit concepts used or advanced 1) by voluntary consensus standards bodies (VCSBs), 2) in the published literature, 3) by EPA. As per the terms of the Settlement Agreement, the MDL and ML were explicitly targeted for inclusion. EPA committed to evaluating concepts published by ASTM International and ISO and to consider approaches and procedures offered by other organizations such as the American Chemical Society (ACS) and the International Union of Pure and Applied Chemistry (IUPAC), as well as other approaches that have been adopted by EPA for use in other programs or that were identified during EPA's review of the published literature. Chapter 2 describes the concepts that EPA evaluated in the assessment. Where appropriate, these approaches also are discussed in context to the issues that are identified and discussed in Chapter 3. Chapter 5 presents the results of EPA's assessment of each approach against the evaluation criteria established in Chapter 4. Appendices B and C of this document present additional details of EPA's assessment of each approach, using the data described in Chapter 1, Section 1.3.

Event 5, Develop and evaluate alternative procedures: EPA planned to develop and evaluate alternative procedures and modifications to existing procedures only if the Agency's assessment of existing procedures suggested that modifications or alternatives to the existing procedures were needed. EPA noted that its primary objective in developing such alternatives (or modifications) would be to address deficiencies noted in Event 4 and improve the performance of the procedures that best meet the criteria established in Event 3. In accordance with the plan and with EPA's findings during the assessment, this Assessment Document includes suggested modifications to the existing MDL and ML procedures.

Event 6, Conduct peer review of the Agency's assessment: EPA documented results of the Agency's assessment in a draft Assessment Document that was completed in August, 2002. EPA conducted a formal peer review of the assessment in accordance with the Agency's peer-review policies and guidance. The peer review was performed by two experts in the field of analytical chemistry and two experts in the statistical aspects of analytical data interpretation.

Events 7 - 11, Actions taken following peer review. After considering peer review comments, EPA revised its assessment and the draft Assessment Document to reflect peer review comments. In March 2003, EPA published two FR notices that met the terms of Settlement Agreement Clause 6a. Comments were received on those notices over a 4 month period ending in August 2003. EPA evaluated all comments received, and revised its assessment as appropriate to reflect these comments. This document details this revised assessment.

Chapter 2
Overview and History of Detection and Quantitation Limit Approaches

It is not possible to measure the concentration of a substance in water all the way down to zero. As an analogy, consider the following example: imagine measuring an object less than 16th of an inch in length with a ruler marked in 1/16th-inch increments. How well can the length of the object be measured using only the ruler? Similar issues arise as chemists try to measure ever smaller concentrations of substances in water. In response to the challenges associated with measuring low concentrations, chemists have defined numerical values that provide points of reference for reporting and using measurement results. These values are usually referred to as detection and quantitation limits. This chapter provides an overview of detection and quantitation approaches and procedures in analytical chemistry and their use in Clean Water Act applications.

2.1 Currie's Call for Standardization

Since 1968, most of the literature regarding detection and quantitation has referenced the work of Dr. Lloyd Currie, recently retired from the National Institutes of Science and Technology (NIST, formerly the National Bureau of Standards). In 1968, Currie published a paper in which he reviewed the then current state of the art regarding detection and quantitation, presented a three-tiered concept, and demonstrated his concept with operational equations for a single laboratory. In his paper, Currie reviewed eight existing definitions for the concept of detection, and reported that when these eight operational definitions were applied to the same data, they resulted in numerical values that differed by nearly three orders of magnitude. These results made it impossible to compare the detection capabilities of measurement methods using available publications. Currie proposed standardizing the terminology using theoretical definitions that he called the *critical value*, the *detection limit*, and the *determination limit*. (In 1995, writing on behalf of International Union of Pure and Applied Chemistry (IUPAC), Currie used the term "quantification limit" instead of his original term "determination limit." Substantial agreement with the International Organization for Standardization (also known as "ISO") on the meaning and language of detection and quantitation was achieved later, although some "subtle differences in perspective" remain [Currie, 2000]). His purpose for these definitions was to create a system in which the standard documentation of any measurement method would include a statement of capabilities that were directly comparable to any other method for measuring the same substance.

Currie used terms from statistical decision theory as the basis for his three-tiered system. In 1968 and 1995, Currie defined the *critical value* as the measured value at which there is a small chance that the concentration in the sample is zero. Consequently, any measured result greater than or equal to the critical value is considered evidence that the sample contains the substance of interest. Currie was careful to emphasize that the decision as to whether the substance has been detected is made by comparing the measurement result to the critical value. Figure 2-1 shows a critical value selected such that measurements greater than the critical value have less than a 1%

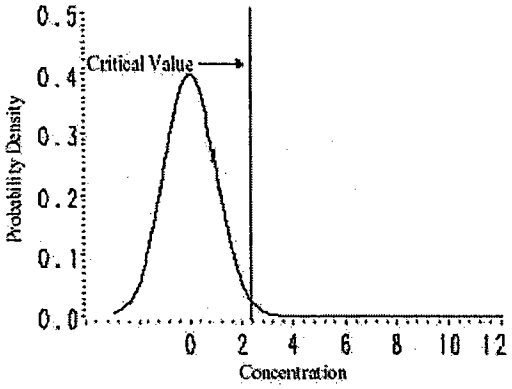

Figure 2-1

chance of being associated with a sample that does not contain the substance of interest. The area under the curve to the right of the critical value represents the probability that a measured value will exceed the critical value. The area under the curve to the left of the critical value represents the (much greater) probability of observing a value that is less than the critical value when the true concentration is zero.

Currie (1968 and 1995) used the term *detection limit* to refer to a true concentration that has a high probability of generating measured values greater than the critical value. That is, measurements on samples that contain concentrations equal to the *detection limit* have a high probability of exceeding the *critical value* and are, therefore, unlikely to result in a decision that the substance is not detected in the sample. In Currie's concept, the *critical value* and the *detection limit* are related and functionally dependent, but it is clear that the detection decision is made on the basis of comparing sample by sample measurements to the *critical value*. While Currie's terminology is consistent with standard statistical decision theory, it is in all likelihood responsible for a great deal of confusion among chemists and others who may associate the term 'limit' with some sort of decision point. Currie (1995) states: *"The single, most important application of the detection limit is for planning. It allows one to judge whether the CMP (Chemical Measurement Process) under consideration is adequate for the detection requirements."* Figure 2-2 shows a detection limit selected such that 99% of the measurements on a sample containing this concentration are expected to be above the critical value. The bell-shaped curve centered at the detection limit illustrates how likely various measurement responses are when the concentration of the substance in a sample is equal to the detection limit. That is, the figure shows the probability density of values measured in a sample with a true concentration equal to the detection limit. The area under the curve to the left of the critical value is equal to 1% of the total area, while the area to the right is equal to 99%.

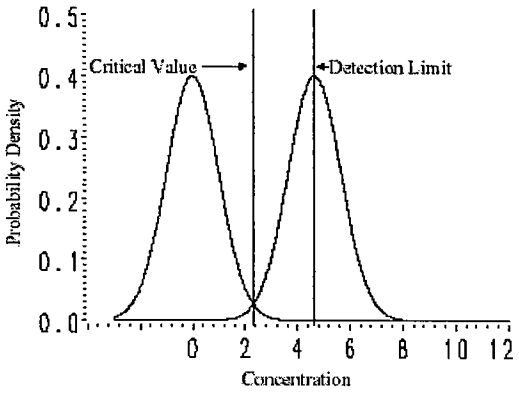

Figure 2-2

Currie (1968, 1995) defined the *determination limit*, later renamed the *quantification limit*, as (quoting Currie, 1995) *"performance characteristics that mark the ability of a CMP to adequately 'quantify' an analyte."* Quantification limits *"serve as benchmarks that indicate whether the CMP can adequately meet the measurement needs. The ability to quantify is generally expressed in terms of the signal or analyte (true) value that will produce estimates having a specified relative standard deviation (RSD) commonly 10%."* This translates into a quantification limit equal to a multiplier of 10 times the standard deviation (a measure of measurement variability) at the limit. The multiplier of 10 (equal to the inverse of the 10% RSD) is arbitrary, but has been used widely. IUPAC selected 10 as a "default value" (Currie, 1995), implying other values are possible. In papers published in 1980 and 1983, the American Chemical Society's Committee on Environmental Improvement also recommended the use of a multiplier of 10 for determining quantitation limits (see MacDougall, *et al.*, 1980 and Keith, *et al.*, 1983). Measured concentrations greater than the quantitation limit are considered to be reliable by chemists, although from a statistical perspective, any measured value, along with knowledge of the precision of the measurement, is useful.

Currie's goal of having method developers publish directly comparable descriptions of detection and quantitation capability remains elusive more than thirty years after publication of his first paper on this topic. Even if Currie's three-tiered concept were used, the treatment of related issues causes

difficulty in comparing methods. Some of these issues include interlaboratory variability, selection of appropriate statistical models, design of detection and quantitation capability studies, and statistical prediction and tolerance. These and other issues are discussed in Chapter 3 of this document.

2.2 Development of the MDL and ML as Practical Embodiments of Currie's Proposal

In 1981, staff at EPA's Environmental Monitoring and Support Laboratory in Cincinnati, Ohio, published a procedure for determining what they referred to as a method detection limit (MDL) (Glaser *et al.*, 1981). The MDL functions as a practical, general purpose version of Currie's *critical value*. The MDL was subsequently promulgated for use in CWA programs on October 26, 1984 (49 FR 43234) at 40 CFR part 136, Appendix B. Prior to formal development of the MDL in 1981, the EPA Office of Water had included the term "minimum level" (ML) or "minimum level of quantitation" in some methods for analysis of organic pollutants. These methods were proposed on December 3, 1979 and subsequently promulgated on October 26, 1984, along with the MDL. Additional information about the MDL and ML is provided below in Sections 2.2.1 and 2.2.2.

2.2.1 Method Detection Limit

Conscious of the definitions provided by Currie and others, Glaser *et al.* (1981) stated "[t]he fundamental difference between our approach to detection limit and former efforts is the emphasis on the operational characteristics of the definition. [The] MDL is considered operationally meaningful only when the method is truly in the detection mode, i.e., [the] analyte (the substance of interest) must be present." Expanding on this reasoning, Glaser *et al.* (1981) developed MDL estimates for methods that produce a result of zero for blanks, such as EPA Methods 624 and 625 for determination of organic pollutants by gas chromatography/mass spectrometry (GC/MS). Blank variability exists, whether or not it can be detected by measurement processes. Failure to detect this variability may be attributed to insufficient sensitivity of the measurement process or, as is the case with some measurement processes, thresholds that are built into equipment which censor measurements below certain levels. Currie's critical value is dependent on the ability to estimate measurement variability of blank samples. In cases where the substance is not detected in direct measurements on blanks, an alternative approach to estimating blank variability must be used. One option is to estimate measurement variability at concentrations that represent the lowest possible levels where a signal can be detected. This is the basic approach of the MDL, which provides a general purpose, straightforward, operational procedure for estimating a quantity analogous to the Currie critical value when measurement processes applied to blank samples do not produce detectable signals. More complex statistical procedures for estimating blank variability are possible and may be preferable from a rigorous statistical perspective, but the MDL has been found to be satisfactory by chemists in a wide range of applications.

In 1984, the MDL became a regulatory option for wastewater discharge permits authorized under the Clean Water Act. To determine the MDL, at least seven replicate samples with a concentration of the pollutant of interest near the estimated detection capabilities of the method are analyzed. The standard deviation among the replicate measurements is determined and multiplied by the *t*-distribution for n-1 degrees of freedom (in the case of 7 replicates, the multiplier is 3.143, which is the value for 6 degrees of freedom). The decision to base the MDL on a minimum of seven replicates reflected a consensus among EPA chemists and statisticians that a requirement of seven replicates is not overly burdensome for laboratories and that laboratories could reasonably be expected to perform the analyses in a single batch.

Both the MDL concept and the specific definition at part 136 have been used within EPA by the Office of Ground Water and Drinking Water (OGWDW), the Office of Solid Waste (OSW), the Office of Emergency and Remedial Response (OERR), and others. The MDL also has been used outside of EPA in *Standard Methods for the Examination of Water and Wastewater*, published jointly by the American Public Health Association (APHA), the American Water Works Association (AWWA), and the Water Environment Federation (WEF), and in methods published by the ASTM International, and elsewhere.

Some members of the regulated industry and others have criticized the MDL because:

- There are some inconsistencies between the definition and the procedure
- It does not account explicitly for false negatives
- It does not always yield a 1% false positive rate
- It does not sufficiently account for blank bias
- A prediction or tolerance limit adjustment is not provided
- It does not account for interlaboratory and temporal intralaboratory variability, and
- It allows discretion in the use of the optional iterative procedures

These issues are discussed later in this document.

2.2.2 Minimum Level of Quantitation

The minimum level of quantitation (ML) was originally proposed on December 5, 1979 (44 FR 69463) in footnotes to Table 2 of EPA Method 624 and to Tables 4 and 5 of EPA Method 625. The ML was defined as the "level at which the entire analytical system must give recognizable mass spectra and acceptable calibration points" (in the footnote to Table 2 in Method 624) and as the *"level at which the entire analytical system must give mass spectral confirmation"* (in the footnote s to Tables 4 and 5 in EPA Method 625).

Between 1980 and 1984, EPA also developed Methods 1624 and 1625 and promulgated these methods along with the final versions of EPA Methods 624 and 625 on October 26, 1984 (49 FR 43234). The definitions of the ML in the promulgated versions of EPA Methods 1624 and 1625 were the *"level at which the analytical system shall give recognizable mass spectra (background corrected) and acceptable calibration points"* (in footnote 2 to Table 2 in Method 1624) and as the *"level at which the entire GC/MS system must give recognizable mass spectra (background corrected) and acceptable calibration points"* (in footnotes 2 to Tables 3 and 4 in Method 1625).

As EPA developed additional methods over the next decade, the definition of the ML was generalized to *"the lowest level at which the entire analytical system must give a recognizable signal and acceptable calibration point for the analyte"* (see, e.g., Section 24.2 of EPA Method 1613 at 40 CFR part 136, Appendix A). In generating actual numerical values for MLs, the lowest calibration point was estimated from method development studies and included in the methods, although a specific calculation algorithm was not used. EPA methods that include the ML generally specify the number of calibration standards to be used and the concentrations of those standards. As a result, laboratories using those methods calibrate their analytical systems with a multi-point calibration (i.e., calibrate using a series of standards at different concentrations over the range of the instrument) that includes a standard at the lowest calibration point listed in the method (i.e., the ML).

In response to a need to establish a compliance evaluation threshold when the water quality-based permit limit is below the detection limit of the most sensitive analytical method published at 40 CFR part 136, EPA refined the definition of the ML in 1994 as 10 times the same standard deviation used to calculate the MDL[1]. Because the MDL is commonly determined as 3.14 times the standard deviation of seven replicate measurements, the ML was commonly calculated as 3.18 times the MDL. (The figure of 3.18 was derived by dividing 10 by 3.14; if more than 7 replicates were used to determine the MDL, both the MDL and the ML multipliers are adjusted accordingly, based on values from the *t*-distribution.) This calculation makes the ML analogous to Currie's quantification limit and the American Chemical Society's limit of quantitation (LOQ), which is defined as ten times the standard deviation of replicate or low concentration measurements (MacDougall, *et al.*, 1980 and Keith, *et al.*, 1983).

To simplify implementation of the ML, the definition also was expanded to state that the calculated ML is rounded to the whole number nearest to (1, 2, or 5), times 10^n, where n is an integer. The reason for this simplification is that calibration of an analytical system at some exact number (e.g., 6.27) is difficult and prone to error, whereas rounding to the whole number nearest to (1, 2, or 5) x 10^n provides a practicable value. The most recent definition of the ML is "*the lowest level at which the entire analytical system must give a recognizable signal and acceptable calibration point for the analyte. It is equivalent to the concentration of the lowest calibration standard, assuming that all method-specified sample weights, volumes, and cleanup procedures have been employed. The ML is calculated by multiplying the MDL by 3.18 and rounding the result to the number nearest to (1, 2, or 5) x 10^n, where n is an integer,*" and this definition was contained in the version of EPA Method 1631 that was promulgated on June 8, 1999 (64 FR 30417) (see Section 17.8 of EPA Method 1631 Revision B).

The ML will generally be somewhat lower than Currie's quantitation limit, even when similar sample sizes and estimation procedures are used. This is because the standard deviation used to calculate the ML will generally be smaller than the standard deviation at the lowest concentration at which the relative standard deviation is 10%. This is due to the fact that, in almost all cases, standard deviation is non-decreasing with increasing concentration, e.g., it generally tends to increase as concentration increases.

Some members of the regulated industry and others have criticized the ML because it:

- Does not account for interlaboratory and temporal intralaboratory variability, and
- Is based on a multiple of the estimated standard deviation which is assumed to be constant in the region of detection and quantitation, rather than a fitted model as suggested by the regulated industry.

These concerns are discussed later in this document.

---

[1]The refined definition of the ML first appeared in EPA's 1994 draft *National Guidance for the Permitting, Monitoring, and Enforcement of Water Quality-based Effluent Limitations Set Below Analytical Detection/ Quantitation Levels"*. The draft guidance was very controversial and never finalized. However, the refined definition of the ML has remained in use for newer analytical methods.

2.3 Other Detection and Quantitation Approaches

To expand somewhat on Currie (1968), standardizing the operational definitions of detection and quantitation would benefit society by making it easier to compare and select measurement methods based on low-level measurement capability and requirements in particular applications. Unfortunately, in spite of agreement on general principles and definitions advanced by Currie and his supporters, consensus on procedures that would result in comparable detection and quantitation estimates has been elusive. Sections 2.3.1 - 2.3.3, which are by no means an exhaustive list of the various approaches advanced to date, highlight approaches that have been most widely advanced for environmental applications.

2.3.1 EPA Approaches

Over the years, a number of detection and quantitation limit approaches have been developed, suggested, or used by EPA in responding to differing program mandates. In part, this situation reflects actual differences in the mandates, and in part, it reflects the fact that no concept advanced to date has emerged as a clear 'winner' that meets all needs for all situations. Approaches that have been used or suggested by EPA include the:

- MDL and ML (described in Sections 2.2.1 and 2.2.2)
- Instrument detection limit (IDL)
- Practical quantitation limit (PQL)
- Estimate quantitation limit (EQL)
- Contract-required detection limit (CRDL) and contract-required quantitation limit (CRQL)

Instrument Detection Limit: EPA methods for analysis of metals have historically included an instrument detection limit, or IDL. Functionally, the IDL is similar to the MDL except that the IDL includes temporal variability (it is determined on 3 non-consecutive days) and does not include all sample processing steps (the IDL characterizes the detection capabilities of the instrument as opposed to the method). Because IDLs do not reflect the entire measurement process and, for the most part, have been used only for measurement of metals, EPA did not consider the IDL as a potential alternate to the MDL when conducting the assessment described in this Assessment Document.

Practical Quantitation Limit: The practical quantitation limit, or PQL, was established in the 1980s by EPA's drinking water program as the lowest concentration at which reliable measurements can be made. The PQL is defined as *"the lowest concentration of an analyte that can be reliably measured within specified limits of precision and accuracy during routine laboratory operation conditions"* (52 FR 25690, July 8, 1987). The PQL is a means of integrating information on the performance of approved analytical methods into the development of a drinking water regulation. The PQL incorporates the following:

- Quantitation,
- Precision and bias,
- Normal operations of a laboratory, and
- The programmatic need to have a sufficient number of laboratories available to conduct compliance monitoring analyses of drinking water samples.

EPA uses two main approaches to determine a PQL for an analyte under the Safe Drinking Water Act (SDWA). One approach is to use the data from Water Supply (WS) studies (e.g., laboratory performance evaluation studies conducted by the Agency as part of the certification process for drinking

water laboratories). The PQL is established at the concentration at which at least 75% of the laboratories in the study, or the subset representing EPA Regional laboratories and state laboratories, obtain results within some predetermined percentage of the true value of the test samples (e.g., ±30%). This approach is used in most cases when WS data are available to calculate a PQL. The WS data approach was used to determine the PQLs for Phase V inorganic chemicals such as antimony, beryllium, cyanide, nickel and thallium (July 17, 1992; 57 FR 31776), as well as many other contaminants regulated under the SDWA.

In the absence of WS data, the second approach that EPA uses is the multiplier method. In this approach, the PQL is calculated by multiplying the EPA-derived MDL by a factor between 5 and 10. The exact multiplier varies and sometimes depends on the degree of concern about the specific contaminant (i.e., based on a human health risk assessment for consumption of drinking water).

Application of the PQL has been traditionally limited to drinking water. Furthermore, the PQL may not be related to the lowest quantitation limit because 1) the PQL is associated with the analyte and may have been determined irrespective of a specific analytical method (e.g., using data from a variety of methods approved for that analyte at 40 CFR part 141), 2) the performance evaluation (PE) samples from which it is derived contain pollutant concentrations that may be well above the true limit of quantitation, 3) the multiplier used to calculate a PQL when PE data are not available is somewhat dependent on concerns about risks from human exposure to contaminants in drinking water, and 4) the resulting PQLs may be too high for purposes other than the Safe Drinking Water Act (e.g., other EPA programs). In addition, because EPA has privatized the performance evaluation program for drinking water laboratory certification, it is not yet clear that appropriate data will be available in the future. Based on these facts, EPA did not conduct an assessment of the PQL for CWA applications.

In the late 1980s, EPA's Office of Solid Waste (OSW) adopted a different version of the PQL as a quantitation limit. No procedure for establishing the limits was given; instead values were extrapolated from the Contract Laboratory Program CRQLs (see below). Since 1994, OSW has actively removed the term "PQL" from its revised methods, replacing it with the term "estimated quantitation limit" (EQL). The term PQL and the original numerical values remain in a few older OSW guidance documents.

Lowest Concentration Minimum Reporting Level (LCMRL) and Minimum Reporting Level (MRL): Recognizing the potential for improvements over the PQL approach, and mindful that confidence in quantitation depends on measurement precision as well as accuracy, EPA's Office of Ground Water and Drinking Water has recently developed a standardized procedure for the determination of the "Lowest Concentration Minimum Reporting Level (LCMRL)" and a companion procedure for laboratories to establish their ability to quantify analytes at a "Minimum Reporting Level" (MRL).

The Lowest Concentration Minimum Reporting Level (LCMRL) is defined as the lowest true concentration for which the future recovery is predicted to fall, with high confidence (99%), between 50 and 150% recovery. A result below the LCMRL is an estimated value that does not satisfy these data quality objectives. However, it may be appropriate to report "estimated" data (i.e., below the LCMRL), depending upon the objectives of the study being conducted. The proposed LCMRL procedure is an iterative process that uses results from three or more different concentrations, of at least five to seven replicate reagent water samples at each concentration. The average recovery, standard deviation, number of replicates, and Student's t value are used to calculate a prediction interval of results that takes into account accuracy and precision at the level tested. For a concentration level to pass criteria, the prediction interval of results must be contained within the boundaries of a predefined quality control interval.

The Agency also has developed a procedure for use in the drinking water program which permits laboratories to confirm that they are capable of meeting a required MRL during their initial demonstration of capability. The MRL validation procedure will involve the analysis of one set of at least seven replicate reagent water samples spiked at the required MRL. To be validated at the MRL, the calculated prediction interval of results must be contained within the predefined quality control interval.

The Agency anticipates using standardized LCMRL/MRL procedures to support the monitoring required under the Safe Drinking Water Act for unregulated contaminants. Requirements for this monitoring are expected to be proposed in the *Federal Register* late in 2004. This proposal will include a full description of the LCMRL/MRL procedures.

Estimated Quantitation Limit: EPA's Office of Solid Waste has defined the EQL as:

> *"The lowest concentration that can be reliably achieved within specified limits of precision and accuracy during routine laboratory operating conditions. The EQL is generally 5 to 10 times the MDL. However, it may be nominally chosen within these guidelines to simplify data reporting. For many analytes the EQL analyte concentration is selected as the lowest non-zero standard in the calibration curve. Sample EQLs are highly matrix dependent. The EQLs in SW-846 are provided for guidance and may not always be achievable."* (see SW-846, Chapter 1).

As noted in most newer SW-846 methods, the EQLs are provided for guidance and may not always be achievable. Because the EQL is not rigorously defined and is guidance, because the EQL may be based on the MDL, and because the EQL can be the lowest calibration point and would, therefore, overlap the ML, EPA did not consider the EQL further in its assessment of detection and quantitation approaches.

Contract-Required Detection and Quantitation Limits: EPA's Superfund program has adopted the use of contractually-required limits that are based on consensus among analytical chemists about levels that can realistically be achieved in commercial laboratories using a contractually-specified method. Laboratories that participate in the Superfund Contract Laboratory Program (CLP) are required to demonstrate that they can achieve the specified CRDLs and CRQLs. The CRDLs are consensus values that apply to the analyses of metals using CLP methods. The CRQLs apply to organic analytes and are based on the concentration of the lowest non-zero calibration standard specified in the CLP methods, in a fashion analogous to the original derivation of the ML. Because few CWA applications involve the use of the CLP methods, EPA did not consider the CRDL or the CRQL as viable alternatives to the MDL and ML when conducting the assessment described in this document.

2.3.2 Industry-supported Approaches

The regulated community has demonstrated an interest in detection limit approaches since EPA first promulgated the MDL and ML for use in CWA programs in 1984 (49 FR 43234). As part of that rule, EPA promulgated Methods 601 through 613, 624, 625, 1624, and 1625 for organic compounds at 40 CFR part 136, Appendix A and EPA Method 200.7 for metals by inductively coupled plasma spectrometry (ICP) at 40 CFR part 136, Appendix C. EPA also promulgated the MDL procedure at 40 CFR part 136, Appendix B. The Virginia Electric Power Company (VEPCO) brought suit against EPA, challenging the Agency's use of the MDL in the promulgated methods. In a settlement, EPA agreed that

the MDL would be applicable only to the 600-series organic methods, as these methods already contained MDL values; i.e., it would not be applicable to EPA Method 200.7. The settlement agreement did not preclude future use of the MDL by EPA or the right of VEPCO to bring suit in such future use.

After the VEPCO settlement, the regulated community, mainly through efforts of the Electric Power Research Institute (EPRI), remained involved in detection and quantitation approaches to be used under EPA's CWA programs. The first approaches that industry advanced were the compliance monitoring detection level (CMDL) and compliance monitoring quantitation level (CMQL) (Maddalone, *et al.*, 1993). The CMDL/CMQL were variants of EPA's MDL/ML that attempted to adjust for interlaboratory variability.

The regulated community continued its efforts to develop alternative detection and quantitation approaches with development of the "alternate minimum level" (AML) in the mid-1990s (Gibbons *et al.*, 1997). The AML is based on statistical modeling of standard deviation versus concentration, which requires large amounts of data.

Most recently, the regulated community has funded development of the interlaboratory detection estimate (IDE) and interlaboratory quantitation estimate (IQE). The IDE and IQE have been balloted and approved by ASTM's Committee D-19 for water as Standard Practices D-6091 and D-6512, respectively. These approaches take into account nearly all sources of variability to arrive at detection and quantitation limits that are higher, on average, than the limits produced by other approaches (see Appendix C of this Assessment Document). Because the regulated community has shifted support from the CMDL/CMQL ?Why? and the AML to the IDE and IQE, and because EPA is not aware of other organizations that currently advocate the earlier approaches, EPA did not consider industry approaches other than the IDE/IQE in its assessment of possible alternatives to the MDL and ML.

As with all other approaches advocated to date, the IDE and IQE have fallen short of being ideal approaches for detection and quantitation for all organizations and applications. To date, EPA is not aware of a demonstrated implementation of the IDE or IQE in the development of an analytical method. Specific concerns that have been raised about the IDE and IQE are that:

- They contain an allowance for false negatives that may be inappropriate,
- The IDE and IQE are based on the use of prediction and/or tolerance intervals, which in some cases may yield conservative (high) estimates,
- The IDE and IQE require a large amount of data in order to be able to model variability versus concentration, including data generated in multiple laboratories, and
- The complexity and expense the statistical procedures involved in calculating an IDE and IQE could be a barrier to innovation and method development.

These concerns are discussed in detail later in this document.

In December 2002, the Inter-Industry Analytical Group (IIAG) submitted a proposal to EPA that recommends (1) a sensitivity test intended to "replace the MDL as a test of whether an individual laboratory is performing adequately," and (2) an interlaboratory validation study design intended to characterize precision and accuracy of methods used for regulatory compliance.

IIAG's proposed sensitivity test includes the provision that EPA first determine the lowest calibration point of a method, prescribe a dilution of that calibration point as the spike level (e.g., at one-half or two-thirds the lowest calibration point), specify a required number of replicates, and set a quality

control acceptance criterion. IIAG asserts that this test would provide all laboratories with a single spike level and an "unambiguous pass or no-pass test." EPA solicited comment on approaches that might be considered appropriate for such determinations (i.e., the lowest calibration point of a method, an appropriate dilution, a number of replicates, and an acceptance criterion for standard deviation between measurements of the replicates). IIAG's proposed "full range" validation study is intended to determine precision and bias across the entire working range of an analytical method (i.e., from a blank to the upper end of the working range) and would account for variability between laboratories. IIAG recommends that results of such a study be used to establish an interlaboratory method detection level.

At the time IIAG's submitted the sensitivity test and full range validation study, EPA did not have the opportunity to evaluate IIAG's proposal against the criteria discussed in Chapter 4, but included the complete text of the recommendations in the regulatory record supporting the February 2003 Assessment Document. EPA is including an assessment of this proposal in Chapter 5 of this document.

2.3.3 Approaches Advocated by the Laboratory Community and Voluntary Consensus Standards Bodies

In 1980 (MacDougall *et al.*, 1980) and 1983 (Keith *et al.*, 1983), the American Chemical Society's Committee on Environmental Improvement (CEI) advanced approaches for the Limit of Detection (LOD) and Limit of Quantitation (LOQ). The ACS LOD is defined as the lowest concentration level that can be determined to be statistically different from a blank. The recommended value for the LOD is three times the standard deviation of replicate measurements of a blank or low-level sample. The LOD is roughly equivalent to the MDL in numerical terms and conceptually equivalent to Currie's critical value.

The ACS LOQ is defined as the level above which quantitative results may be obtained with a specified degree of confidence. The recommended value for the LOQ is 10 times the standard deviation of replicate measurements of blanks or low-level samples. Because the LOD and LOQ are still used by the analytical community, they have been included in EPA's reassessment of detection and quantitation approaches.

In the mid-1980s, the ACS CEI introduced the concept of the Reliable Detection Limit (RDL) and the Reliable Quantitation Limit (RQL). The RDL and RQL were attempts at simplification of the LOD and LOQ. Both the RDL and the RQL involved applying a multiplier to the standard deviation derived from replicate measurements of a low-level sample. Neither concept received acceptance by the analytical community. Because the RDL and RQL are no longer being advanced by ACS, they were not considered for evaluation in EPA's assessment of detection and quantitation approaches.

In 1999 (Currie, 1999a and 1999b), IUPAC and ISO reached substantial agreement on the terminology and approaches documented by Currie (1995), although "subtle differences in perspective" of the organizations remain (Currie, 2000). IUPAC and ISO have not, to date, published methods that include limits reflecting these standards. Similarly, although ASTM International adopted the IDE in 1997 and the IQE in 2000, ASTM International has not included any IDE or IQE values in methods approved through the ASTM ballot process. On the other hand, ISO and ASTM International have published methods that employ the MDL. Because IUPAC and ISO have approved the critical value, detection limit, and quantification limit, and because ASTM International has approved through ballot the IDE and IQE, EPA has included these approaches in its assessment of detection and quantitation approaches.

At the ACS Annual Meeting held in August, 2002, CEI members discussed the issue of detection and quantitation, with the objective of determining if the LOD and LOQ approaches should be re-visited. At that meeting, several members suggested that the committee consider adopting a sample-specific detection limit approach in which the ratio of instrument signal to background noise is used to estimate a detection limit for each analyte in each sample analyzed. EPA did not include the signal-to-noise ratio concept in this assessment because its application is limited to specific types of measurement techniques, such as gas chromatography/mass spectrometry. Limitations of this concept for use in general environmental chemistry are best illustrated by the fact that it would not apply to any of the techniques traditionally used to determine the "conventional pollutants" cited in the Clean Water Act (the only pollutants cited by name in the Act), i.e., biochemical oxygen demand (BOD), total suspended solids (TSS), fecal coliforms, and pH.

During the comment period on the February 2003 assessment document, the American Council of Independent Laboratories (ACIL) submitted a procedure that was developed to address bias that may arise in the estimation of detection limits under certain conditions. The ACIL procedure separates estimation of a detection limit into two cases: analyses that always produce a numeric result, even so-called "blank" samples (i.e., zero analyte added), and analyses that do not always produce a numeric result, i.e. blank samples appear to produce no signal. We will call these Case I and Case II. Analysis of samples for metals with inductively coupled plasma optical emission spectroscopy (ICP-OES) is an example of ACIL Case I; analysis of samples for PCBs with gas chromatography and mass spectrometry (GC/MS) is an example of ACIL Case II.

For Case I analyses, ACIL suggests making use of the numeric results obtained from the analysis of blank samples which laboratories routinely run as a quality control measure. ACIL provided a detailed set of instructions for conducting the analyses and doing the MDL calculation. Differences between the ACIL Case I calculation and the EPA MDL calculation include: (1) use of blanks rather than low-level spikes to estimate standard deviation, (2) the calculation of both a critical level and a long-term MDL, where the MDL is based on adding the mean of the blank results to 2 times the product of the standard deviation and t-statistic, (3) a bias offset correction that adds the mean of the blank results to the calculated critical level and MDL, (4) recommends the use of results from a minimum of 20 analyses, and (5) analysis over the course of a year from routine daily operations (rather than on one day). ACIL's Case I procedure is similar, but not identical, to the USGS procedure that is described in Section 3.3.4 of this document. The ACIL Case I procedure has no explicit limits on the amount of contamination allowed in the blanks before a laboratory is considered to be "out of spec."

For Case II, blanks cannot be used to estimate the standard deviation because they do not provide a response. Thus, Case 2 recommends an iteration of multiple low level spikes somewhat similar to the requirements in the EPA MDL procedure. However, the calculation of an MDL from the results of these spiking experiments differs significantly from the EPA MDL procedure. The procedure also specifies a sensitivity check for which some of the details are not as explicit compared to the Case I part of the ACIL detection limit procedure.

2.3.4 Approaches Advocated by Other U. S. Government Agencies and Other Governments

Within the U.S., EPA found that other Federal agencies tend to rely on the detection and quantitation limit approaches described above or on variants of those procedures. For example, the USGS National Water Quality Laboratory (NWQL) began using the EPA MDL procedure in 1992. USGS

NWQL has since developed a variant of the MDL called the long-term MDL (LT-MDL) that has been in routine use since 1999. The LT-MDL determination ideally employs at least 24 spiked samples prepared and analyzed by multiple analysts on multiple instruments over a 6- to 12-month period at a frequency of about two samples per month.

Unlike EPA programs that rely on hundreds of commercial, Federal, State, and local laboratories for sample analysis, the samples collected for USGS water programs are analyzed by the USGS National Water Quality Laboratory in Denver, Colorado. As described by USGS, the long-term MDL is based on many of the same fundamental assumptions as the MDL, namely:

1. Normal data distribution,
2. Constant standard deviation from the spike concentration down to zero, and
3. Best-case detection condition (because LT-MDLs typically are determined by spiking the analyte in a clean matrix, e.g., reagent water).

The three primary differences between the EPA MDL and the USGS LT-MDL procedures are the (1) larger minimum number (24) of spike samples, (2) longer time period, and (3) mixing of instruments and analysts in the determination of the LT-MDL. Because the MDL and LT-MDL approaches otherwise are so similar, EPA did not evaluate the long-term MDL approach in the February 2003 assessment. Instead, EPA considered the underlying differences between the two approaches (namely the effects of temporal, instrument, and analyst variability) in its assessment of issues (see Chapter 3).

In the LT-MDL procedure the low-level spike used for each analyte and instrument is recalibrated at least once a year or when an anomaly occurs. USGS has enhanced the LT-MDL procedure by using their large volume of uncensored blind laboratory blank data, which also is collected yearly, as a reality-check on the spike-based LT-MDL. In cases where the standard deviation used to calculate an LT-MDL based on blind blank data is significantly different (especially when greater) than the standard deviation used to calculate the spike-based LT-MDL, the blank data are used to calculate the LT-MDL. Blind blank data also are used to evaluate whether the calculated LT-MDL requires an off-set correction for blank bias, i.e, LT-MDL = (s * Student t) + median or mean blank concentration. This offset is similar, but not identical, to the ACIL Case I procedure described in Sect. 2.3.3 of this document. The LT-MDL offset correction compensates for a blank distribution that is not centered on zero (as assumed by the EPA MDL formula).

The NWQL has found that this blank bias correction to the LT-MDL is especially important for blank-limited analytes, including some metals, total organic carbon, phenol, and nutrients. The NWQL also uses a data reporting convention that incorporates a higher reporting level (called the laboratory reporting level; LRL) that is set at two or more times the LT-MDL. However, this convention also includes reporting of data between the LT-MDL and LRL.

Outside the U.S., EPA found that the European Union (EU) relies on the terminology and conventions developed by Currie, IUPAC, and others (Eurachem, 2000). The EU advocates reporting all results along with an estimate of the uncertainty associated with each value. In its discussion of the issue, the EU indicates that use of the term 'limit of detection' only implies a level at which detection becomes problematic and is not associated with any specific definition. Instead, the EU focuses its attention on ways to estimate uncertainty, basing its approach on the ISO *Guide to the Expression of Uncertainty in Measurement* (1993). However, the EU also notes that the use of uncertainty estimates in

compliance statements and the expression and use of uncertainty at low levels may require additional guidance. The United Kingdom's Valid Analytical Measurement Programme (VAM) has adopted a similar approach that is based on both the ISO and the Eurachem guidance (Barwick and Ellison, 2000). Because these approaches are focused on estimating uncertainty rather than at establishing or defining limits for detection and quantitation, EPA did not consider the European approaches in this assessment.

Chapter 3
Issues Pertaining to Detection and Quantitation

As part of the Settlement Agreement concerning EPA's reassessment of detection and quantitation limit approaches, EPA considered several specific issues pertaining to these approaches. These issues included:

- Criteria for selection and appropriate use of statistical models,
- Methodology for parameter estimation,
- Statistical tolerance and prediction,
- Criteria for design of detection and quantitation studies, including selection of concentration levels ("spiking levels"),
- Interlaboratory variability, and
- Incorporation of elements of probability design.

In developing the plan for conducting this assessment, EPA identified other issues that should be considered. With the exception of the first issue, these issues are discussed in this chapter and include:

- Concepts of the lower limit of measurement (discussed in chapter 2),
- The need for approaches that can support CWA programs, including:
  - method performance verification at a laboratory,
  - method development and promulgation,
  - National Pollutant Discharge Elimination System (NPDES) applications,
  - non-regulatory studies and monitoring,
  - descriptive versus prescriptive uses of lower limits to measurement, and
  - use of a pair of related detection and quantitation procedures in all OW applications
- Censoring of measurement results,
- Sources of variability (including, but not limited to interlaboratory variability),
- False positives and false negatives,
- Measurement quality over the life of a method,
- Matrix effects,
- Background contamination,
- Outliers,
- Instrument non-response,
- Accepting the procedures of voluntary consensus standards bodies (VCSBs),
- National versus local standards for measurement,
- Ease of use (i.e., ability of study managers, bench chemists, and statisticians to do what is required by a detection or quantitation limit procedure),
- Cost to implement the procedures, and
- Laboratory-specific applications.

These issues are organized into three subsections that follow. Section 3.1 discusses the issues that are primarily driven by analytical chemistry concerns, Section 3.2 discusses the issues that are

primarily driven by CWA regulatory considerations, and Section 3.3 discusses issues that are primarily driven by statistical concerns. Table 3-1, at the end of this chapter, provides a summary of the issues discussed in Sections 3.1 - 3.3.

3.1 Analytical Chemistry Approaches to Detection and Quantitation

This section explains the key analytical chemistry issues involved in the development of detection and quantitation limits. These include: (1) nonzero sample blanks, (2) instrument censoring, (3) matrix effects, (4) analyte recovery, and (5) temporal variability of the measurement system.

3.1.1 Nonzero Sample Blanks

Analytical chemists rarely state that a sample contains zero concentration of a substance of interest. Even when the sample is created in a laboratory for the purpose of containing as little substance of interest as possible (a blank), analytical chemists recognize that some small residual amount of the substance may be present and contribute to the measurement result. The inability of a laboratory to reduce the concentration of a substance in the blank is often the limiting factor in attempts to make measurements at ever lower levels.

A classic example of this potential problem was published by Patterson in the late 1960s and 1970s (e.g., Patterson and Settle, 1976). Patterson demonstrated that the majority of concentrations of lead reported in the literature for such diverse matrices as urban dust, open ocean waters, and biological tissues were in error by several orders of magnitude. The source of the "gross positive errors" (or "positive bias" from blanks) was contamination introduced during sample collection, handling, and analysis. Interlaboratory studies of the day designed to determine consensus values for reference materials were, in fact, determining the consensus values for background contamination across laboratories. Patterson recognized the value in running blank samples (samples thought not to contain the substance of interest) to demonstrate that the sample collection, handling, and analysis processes were not introducing contamination. Patterson subsequently developed the techniques for "evaluating and controlling the extent and sources of industrial lead contamination introduced during sample collecting, handling, and analysis" that form the basis of the "clean techniques" used for metals analysis today, and that are incorporated in several EPA analytical methods, including EPA Method 1631 for measurement of trace-level mercury.

The most common analytes for which contamination problems are encountered in environmental measurements are metals, primarily zinc because of its ubiquity in the environment. With the exception of some volatile organic compounds, such as methylene chloride and acetone, that are used as solvents in laboratories, contamination in the measurement of organic compounds is less of a problem than contamination of samples for metals analyses. Therefore, for determination of metals, a blank is usually included or compensated in the calibration whereas, for organics, except for the solvents, the concentration in the blank is generally assumed to be zero and there is no compensation of the calibration.

Measurement methods designed to determine substances at very low concentrations may include requirements for the preparation and analysis of a variety of blanks that are designed to identify the extent and the sources of contamination. Analysts understand that "blank" does not necessarily mean zero concentration, and through careful control and evaluation, it is possible to make measurements for which the blank contribution is sufficiently small to be considered negligible.

In the February 2003 version of this document, EPA noted that useful detection and quantitation limit approaches should address the potential contribution of the blank through both the design of the study that generates the detection and quantitation limit estimates and the evaluation of the study results. Stakeholders commenting on EPA's 2003 assessment of these approaches added that, for many blank analyses, there is a measurable response (blank bias) that can be attributed to reagents, sample vessels, and other contamination sources, and that the MDL procedure failed to take these blank responses into account. Several commenters suggested that the mean of the blank results should be added to the formula used to calculate the MDL. The American Council of Independent Laboratories (ACIL) submitted a procedure to use blanks rather than spikes to estimate a detection limit. The USGS submitted a long term MDL procedure that uses either the mean or median of blank results as a lower bound reality check on the MDL whenever an MDL computed from the low level spiking experiments is sufficiently less than the blank results.

Following a careful evaluation of these comments and further consideration of this issue, EPA recognizes that, under certain conditions, it may be appropriate to account for blanks in establishing detection and quantitation limits with certain limitations. For example, a procedure to handle blanks should account for negative results, and should limit and control sample and laboratory contamination. Negative blanks are possible and can be caused by a blank-subtracted calibration in which the result for the calibration blank is greater than the results for the blanks used to establish the MDL. If such negative blanks were to result in a negative mean blank, adding the mean blank result to the formula could result in an unattainably low MDL. Conversely to eliminate unnecessarily high MDLs, laboratories also would need to ensure that the results of blank samples are not excessive. The laboratory would need to use "clean" and other techniques to control contamination to the lowest possible levels and/or use a second or higher order calibration function to preclude high results for a calibration blank from exerting undue influence on the sample results. In addition to working out some of the details of necessary bounds on blank correction and contamination, differences between the procedures submitted by USGS and ACIL need to be evaluated.

3.1.2 Analytical Instrument Thresholds: Data Censoring

Certain analytical instruments typically employ "thresholds" to eliminate spurious or background signals so that analysts can be relieved of the burden of removing or compensating these small signals. As a result, the instrument itself may return a response of zero to a blank (a "non-response"). As an example, gas chromatograph/mass spectrometer (GC/MS) instruments often contain thresholds below which no instrument signal is reported. With no instrument signal reported, no measurement result can

be reported, and the instrument will report zero to indicate the lack of a signal. To understand how instrument thresholds are used, it may be helpful to think of background static heard on a citizen-band (CB) radio or a walkie-talkie. The static is present, but it has no meaning. Turning the "squelch" knob to filter out the static also may make it impossible to hear the caller. In the context of detection, increasing the instrument threshold may cause the instrument to miss a substance of interest at a low level.

In 1997, EPA conducted a study of 82 semivolatile (acid and base/neutral) organic compounds measured by EPA Method 1625 in order to observe the performance of a GC/MS instrument both with and without application of an instrument threshold (sometimes known as the "Episode 6184 study"); see Chapter 1, Section 1.3.2.3). In the study, solutions at up to 17 concentration levels were analyzed with the threshold on (i.e., low-level signals are automatically eliminated) and with the threshold off (i.e., there is no suppression of signals). Samples were analyzed at decreasing concentrations, including a blank concentration level, with triplicate determinations at each concentration. With the threshold turned on, all of the measurements made on the blank were reported as zero. This is not surprising, given the purpose of the instrument threshold. Without the threshold off, only 27 of 230 measurements on the blank (11%) were reported as zero, and no negative results were reported.

Instrument thresholds have a direct and indirect impact on estimating detection and quantitation limits. The main direct impact is that it is not possible to estimate the standard deviation of measurements at zero. However, by definition, the standard deviation at zero is required to calculate the Currie critical value (CRV; L_c). The EPA MDL procedure was constructed to deal with this problem by providing a way to estimate a standard deviation at a low concentration, and including instructions for determination of a concentration as close to zero as is possible that will generate a measurement.

To calculate an MDL using the 40 CFR 136, Appendix B, MDL procedure, it is necessary to find the lowest concentration at which the analytical system will return results. Many laboratories have run repeated measurements in order to find this concentration. The challenge of finding this lowest concentration manifested itself in EPA's variability versus concentration (Episode 6000) studies. Technologies for determination of organic, conventional pollutants, and metal analytes were evaluated in the Episode 6000 studies. The MDL procedure suggests iteration until the calculated MDL is within a factor of five of the spike level. For the Episode 6000 studies, EPA instructed laboratories to use a factor of three instead of five in an attempt to more narrowly define the lowest spike level at which measurements could be made. This change to a factor of three also was suggested by one of the peer reviewers charged with evaluating EPA's 2003 assessment of detection and quantitation limits, who noted:

> *"However, the use of as much as five times the critical level for the spike concentrations could be problematic. The inflation of the MDL by using a spike at the critical level is only 25% for a method with a high-level CV of 20% (this and other calculations here are done with the Rocke and Lorenzato 1995 variance function assuming a sample size of 7). A spike concentration of 3 times the critical level inflates the MDL to a value 140% higher, which even there may be tolerable. Use of a value 5 times the critical level gives an inflation of over 280%. ..."*

Following some theoretical example calculations that are not reproduced here, the peer reviewer's comment continues with:

> *"Thus, I would recommend that the procedure be altered to use concentrations that are no more than 3 times the detection limit, and perhaps to permit concentrations lower then the critical level, including possibly blanks"* (Rocke, 2002).

The reviewer's calculations suggest that the MDL may be strongly inflated for a spike level of five times the MDL, but only moderately inflated at a spike level of three times the MDL. However, during the Episode 6000 studies, several laboratories asked for relief from the factor of three requirement because a large number of iterations were required to meet it. In response, EPA reinstated the factor of five for these laboratories. If the reviewer's example calculations are correct and a practical procedure for determining the MDL using the factor of three were implemented, it could exacerbate the concern from the regulated community that MDL values are too low.

Several stakeholders commenting on EPA's 2003 assessment suggested that approaches to detection and quantitation should address methods that do not always produce an instrument response, e.g. so-called blanks never produce a response because of electronic censoring by the instrument, and that EPA's approaches do not do so. These stakeholders prefer that the MDL not be applied to methods for which an identifiable analyte signal cannot be established using method blanks, where pattern recognition is required (e.g., Method 608 for PCBs as Aroclors), or where the method requires more than one signal for an analyte to be positively identified (e.g., the use of multiple ions in GC/MS methods). EPA recognizes that additional guidance needs to be developed for these methods. One commenter, the American Council of Independent Laboratories, submitted a draft set of procedures designed that partially addressed methods that do not produce an instrument response at zero concentration. EPA evaluated the ACIL procedure, which involves a complex set of iterative spiking experiments, and found that it needs further refinement. EPA agrees that this issue warrants further examination.

3.1.3 Matrix Effects

"Sample matrix" is a term used to describe all of the substances, other than the substance(s) of interest, present in an environmental sample. In the case of a wastewater sample, this would include the water itself, as well as any other dissolved or suspended materials.

"Matrix effect" is a term used to describe a situation in which a substance or combination of substances in the sample (other than the substance[s] of interest) influence the results of the measurement. Positive interferences may inflate the results for the substance or make it difficult to distinguish one substance from another. However, unless the positive bias from the matrix is consistent and predictable, the measurement result may be unreliable. Negative interferences may suppress the results for the substance to the point that the results cannot be distinguished from background instrument noise.

In some cases, finding a matrix effect indicates that the analyst should select a more appropriate method. For example, a colorimetric method for the measurement of sulfide may be a poor choice for the analysis of a sample that is very cloudy or darkly colored. In other cases, characteristics of the sample such as its pH may destroy the substance of interest, effectively preventing analysis for that substance.

Nearly all of the newer analytical methods approved at 40 CFR part 136 describe the preparation and analysis of quality control samples that are designed to indicate the presence of matrix effects (e.g., matrix spike and/or matrix spike duplicate samples). Many of these methods also contain techniques for addressing matrix effects. Further, EPA has developed guidance documents that amplify the discussions in those methods (e.g., *Guidance on Evaluation, Resolution, and Documentation of Analytical Problems Associated with Compliance Monitoring*, June 1993, EPA 821-B-93-001). For determination of mercury by EPA Method 1631 that is the subject of the Settlement Agreement, additional guidance on resolving matrix interferences to achieve specified detection and quantitation limits is provided in EPA's *Guidance for Implementation and Use of EPA Method 1631 for the Determination of Low-Level Mercury* (March 2001, EPA 821-R-01-023). Following the techniques in the methods and guidance will usually reduce adverse effects of the sample matrix on detection/quantitation limits and measurement results.

3.1.3.1 Allowance for Matrix Effects in Detection and Quantitation Limits

Some stakeholders have suggested that detection and quantitation limits should be determined in "real-world" matrices, rather than in reference matrices intended to simulate method performance in a particular medium (water, soil, biosolids, tissue). Some commenters on EPA's 2003 assessment believe that any form of a detection limit study should require demonstration of the lowest level of detection achievable in an interference-free matrix and should relate this to what is actually detectable in a highly complex matrix. One commenter stated that EPA should provide an objective set of procedures that a permittee can follow to avoid liability when faced with an MDL or ML it legitimately cannot achieve because of the unique nature of its effluent. EPA notes that permittee liability was not a goal or purpose of our assessment of detection and quantitation approaches and issues. Although EPA recognizes stakeholder concerns about the matrix effects, there are several problems associated with the approach suggested by some commenters. These problems include:

- Many "real-world" matrices contain the target pollutant at levels well above the detection or quantitation levels, making it impossible to characterize what can and cannot be detected at low levels. Diluting the sample to dilute the target pollutant concentration is an option. However, this also has the potential to dilute any interferences that might be present, thereby defeating the purpose of using the real-world matrix.

- Use of a reference matrix to establish detection and quantitation limits allows the results to be reproduced (i.e., confirmed) by an independent party; such a confirmation may not be possible with many real world matrices that may be subject to seasonal, diurnal, or other types of variability.

- Few environmental analyses are conducted on actual samples of reagent water or other reference matrices and there may be matrix-specific limitations to the sensitivity of any given analytical method. From a practical standpoint, it would be very impractical to evaluate method sensitivity in every possible matrix to which a method might be applied, or to establish a subset of all possible matrices that would satisfy the concerns of every regulated discharger.

- The cost of determining detection and quantitation limits in every possible matrix would be prohibitive.

Because of these difficulties a reference matrix (or reference matrices) is an appropriate and practical first choice to establish detection and quantitation limits. And the procedures for defining these limits should allow for evaluation of data collected in the specific matrices of concern. Laboratories or data users are most able to determine which matrices might be considered to be "highly complex" based on the matrices that are typically analyzed in a given laboratory. EPA's detection and quantitation procedures do not preclude laboratories from determining MDLs in matrices other than reagent or "blank" matrices, and the Agency encourages laboratories and others to determine matrix-specific MDLs when all efforts to resolve matrix interferences have been exhausted. The existing procedure at 40 CFR 136, Appendix B, includes a discussion regarding determination of matrix-specific MDLs for this reason. Laboratories usually are very capable of eliminating or compensating matrix interferences if tasked to do so. However, given the degree of concern about this issue it is appropriate for all parties to continue to search for additional solutions to the "real world" matrices issue.

3.1.3.2 Repository of Reference Matrices

Two of the four peer reviewers charged with evaluating EPA's assessment of detection and quantitation limit approaches suggested that EPA create a repository of reference matrices, similar to those developed by NIST, and that these reference matrices be used to challenge a test method and to establish detection and quantitation limits (Cooke, 2002 and Wait, 2002). EPA has considered such a repository from time to time and again in response to this suggestion, but has been unable to resolve all of the issues surrounding such a repository. Some of these issues are:

- The stability of aqueous samples,
- The holding times necessary to assure stability,
- The argument that no matrix from a given industrial discharge in an industrial category or subcategory reflects the characteristics of another discharge in that or other industrial categories or subcategories,
- The cost of maintaining such a repository, and
- The potential conflict with NIST and with non-governmental organizations that provide reference matrices.

Given these issues, it is appropriate to leave the development and maintenance of standard reference materials (SRMs) and certified reference materials (CRMs) to NIST and the commercial marketplace. These reference materials are a useful means of challenging a test method and EPA has suggested in recent methods that reference matrices be analyzed, when available, as an additional QC measure. For example, when EPA developed an appendix to Method 1631 for application to matrices other than water, EPA specified use of a quality control sample (QCS) with the statement that "many certified reference materials (CRMs) are available for total mercury in plants, animals, fish, sediments, soils, and sludge" and the requirement that "recovery and precision for at least one QCS per batch of samples must meet the performance specifications provided by the supplier."

Although SRMs and CRMs could be useful in establishing detection and quantitation limits, practical considerations are likely to preclude their use for this purpose in most situations. This is because the materials would need to contain the analytes of interest at levels that are near the detection limit (e.g., within 1 to 5 times the concentration of a determined MDL). Such concentrations are unlikely to occur in an SRM produced by NIST or a CRM produced by a vendor, and diluting the CRM/SRM would diminish matrix effects, as indicated in Section 3.1.3.1.

As an alternative to using standard reference materials, EPA commonly tests its analytical methods on a variety of real-world matrices, and allows for this variability in the QC acceptance criteria for the matrix spike (MS) and matrix spike duplicate (MSD) samples. For example, EPA published performance data in Table 3 of EPA Method 1631B for reagent water, fresh water, unfiltered and filtered marine water, and unfiltered and filtered secondary effluent, and allowed for the variability among these matrices in the QC acceptance criteria for the MS/MSD in the method. ASTM Committee D 19 allows this approach in development of QC acceptance criteria for methods (see Section 6.5.1.1 of ASTM D 5847: *Standard Practice for Writing Quality Control Specifications for Standard Test Methods for Water Analysis.*)

3.1.4 Recovery of Analytes from the Sample Matrix

In the preceding two sections, we discussed bias (errors) from blank contamination and matrix effects. Errors from recovery effects ("recovery bias") are discussed in this section. Recoveries are a measure of the amount (usually expressed as a percentage) of analyte that is recovered from the sample matrix and measured by the analytical system. Chemists sometimes use the phrase "accuracy of the method" when listing the percent recovery of an analyte. A goal of analytical chemistry is to achieve recoveries as close as possible to 100%. When this is not achieved, recovery correction may be used. The purpose of recovery correction is to adjust the measured concentration for the amount by which the measured concentration differs from the true concentration (if known). Recovery "factors" are initially determined by analysis of a sample containing a known (spiked) amount of the analyte. These factors are applied to measurements of samples with an unknown amount of the analyte in the same or a similar matrix. To illustrate the potential need for recovery correction, consider analytes, such as organic bases

(e.g., benzidine) and acids (e.g., phenols) in a water sample, that are either not totally (100%) recovered in the extraction process, or are adsorbed on the surface of a GC column at very low (nanogram) levels. As a result, the measured concentration of these analytes is always less than the true concentration in the water sample. These incomplete recoveries have led some developers of detection and quantitation limit approaches to believe that these limits should be recovery corrected (i.e., that the detection or quantitation limit should be adjusted inversely proportional to the recovery). For example, if an analyte is recovered at 50%, the detection and/or quantitation limit should be doubled, and the amounts measured in unknown samples also should be doubled to allow for recovery correction.

Several stakeholders have stated that understanding this "recovery bias" is particularly important when reporting results near the limit of detection, and is critical when reporting quantifiable results. These stakeholders believe that even if recovery bias is not controlled at the detection level, the approach to determining detection and quantitation limits should compensate for it.

Few of the traditional approaches to establishing detection and quantitation limits include procedures for recovery correction. For example, the issue was not addressed by Currie in his original proposal of a critical value or quantitation limit. Similarly, neither EPA's MDL and ML nor the American Chemical Society's LOD and LOQ, all of which are based on the approaches advanced by Currie, include a mechanism for recovery correction. When Currie introduced his critical value, he defined it as "the minimum significant value of an estimated net signal or concentration, applied as a discriminator against background noise" (Currie, 1995). Because the critical value is defined as a *measured* concentration rather than a *true* concentration, a recovery correction is not included.

The use of recovery correction, however, has been included in several of the most recently developed approaches for detection and quantitation. For example, the minimum detectable value (MDV) adopted by ISO and IUPAC, and the interlaboratory detection estimate (IDE) and interlaboratory quantitation estimate (IQE) adopted by ASTM include procedures for recovery correction. The IQE also contains a further correction that we have termed a "bias" correction.

In the ISO minimum detectable value approach, recovery is treated as a linear function versus concentration, and an extrapolation is used to estimate the recovery at zero concentration. This projection of the regression line to zero concentration can lead to errors because, depending on the intercept (in concentration units), the recovery at zero concentration can be positive, zero, or negative, resulting in an inflated minimum detectable value, an minimum detectable value very close to zero, or a negative minimum detectable value. For further details, see the section titled "Negative detection limits for the ISO/IUPAC MDV" in Appendix C to this Assessment Document, and the data in Table 2 of that appendix.

The IDE and IQE fit recovery versus concentration in a way analogous to the fitting in the minimum detectable value. The difference is that an unweighted model is used in the minimum detectable value, whereas the linear model in the IDE and IQE is weighted as determined by the model of standard deviation versus concentration that is used in calculating the IDE and IQE. (If this model is the constant model, the weighting is the same as for the minimum detectable value.) The IQE, but not the IDE, includes an additional correction for the bias associated with an estimate of the true standard deviation at each concentration as compared to the measured standard deviation at each concentration. In this context (a "bias" correction to the IQE), "bias" means the amount by which the estimated sample standard deviation differs from the true population standard deviation. This use should not be confused with a common use of "bias" in analytical chemistry measurements to denote the deviation of a result from the true value (usually expressed as percent.)

Recovery correction may be appropriate if (1) when developing method detection and quantitation limits, the recovery is consistent across laboratories, matrices, and conditions, and (2) the relative variability (as relative standard deviation) remains constant as the recovery decreases. These two conditions are rarely observed. The first requirement (consistent recovery) would need to be tested under a variety of conditions because, if the recovery varies among laboratories, matrices, and analytical conditions, then a detection and/or quantitation limit would need to be developed for each of these conditions. EPA's experience is that poor recovery is rarely consistent; i.e., if one laboratory measures a recovery of 40%, another laboratory may measure 20%, or 60%, but not exactly 40%.

Although some stakeholders disagree, EPA believes that the normal condition in environmental analytical measurements is that variability (as standard deviation) between sample results remains approximately constant as the recovery decreases (i.e., the relative precision [as RSD] is poorer at low recovery). For example, if the RSD is 10% at 100% recovery, the RSD may be 50% at 50% recovery, and may be 100% at 10% recovery. For examples of the effect of poor recovery on precision, see the quality control (QC) acceptance criteria for the semivolatile organic compounds in Table 8 of EPA Method 1625 (see 40 CFR part 136, Appendix A). This increase in relative variability is not the result of measurements being made at lower levels, as is the normal case, but is a result of variability in the extraction (partitioning) process. One stakeholder commenting on EPA's 2003 assessment stated that EPA's statement that variability remains approximately constant as recovery decreases may not hold true in all cases, and recommends that, if recovery falls within a specified level (e.g., less than 70%), detection limits should be adjusted accordingly. EPA acknowledges that there may be instances in which this general condition does not hold true.

EPA has traditionally viewed recovery correction with great caution, and has preferred to require that laboratories analyze quality control samples to demonstrate that analytes are recovered within an acceptable level. For example, EPA's Office of Water methods require that laboratories prepare and analyze both a reference matrix and a sample matrix that have been spiked with the analytes of interest, and that these analytes be recovered within method-specified acceptance criteria. If the recovery criteria are met, then samples analyzed in the batch are considered to be reliable within the overall level of error associated with the method, and results are reported without correcting for the recovery. Measurements of dioxins/furans, PCBs, and pesticides can be made to very low (femtograms per liter) concentrations, with no decrease in recovery compared to recoveries observed at much, much greater (microgram per liter) concentrations. (One microgram is equivalent to one million femtograms). The ability to measure dioxins/furans, PCBs or pesticides down to these low concentrations demonstrates that recoveries for these compounds do not decrease with decreasing concentration. There also are chemicals, such as the nitrophenols and benzidine, that are not recovered reliably at sub microgram per liter levels. But these instances are known and recognized in the instructions for conducting these measurements.

MDLs are established and listed in methods based on the determined (measured) concentration (not the spike concentration), and laboratories and others that are required to verify MDLs, verify based on the determined concentration. If EPA estimated and listed MDLs in methods based on the spike (true) concentration, logic would require that the true (recovery corrected) concentration be used for regulatory compliance with the result that all results, not just the MDL, would be greater than nonrecovery corrected results.

Recovery-correction techniques are employed in some Agency methods. Most notably are those methods that employ isotope dilution techniques, in which a stable, isotopically labeled analog of each target analyte is spiked into each sample. Because of their structural similarity to the analytes of interest, the labeled analogs are assumed to behave exactly like their unlabeled analogs (the target analytes). Because the recovery of the labeled analog will be similar to that of the target analyte, the technique allows for recovery correction of each target analyte and is particularly useful in highly complex matrices. In these methods, recovery correction techniques are specified as part of the procedures for calculating and reporting results and are dependent on the one-to-one relationship of the target analyte and the labeled analog. Inclusion of an additional procedure for recovery-correction in determining detection or quantitation limits for such methods could result in double-counting of analytical bias.

Another procedure for dealing with bias (errors) from blank contamination, matrix effects, or errors from recovery effects ("recovery bias"), is to assure that the detection or quantitation limits which is determined meet the data users data quality needs for both precision and accuracy, without any correction. As described previously in chapter 2, EPA's drinking water program is developing an approach to setting quantitation levels called the minimum reporting level or MRL. The MRL addresses these issues by setting a data quality objective for minimum and maximum permitted inaccuracy arising from these effects.

The issue of bias in determination of detection or quantitation limits be it from blanks, matrices or other than 100% recovery of an analyte is a longstanding issue. All parties should continue to collaboratively work to develop other solutions or approaches to mitigate bias effects.

3.1.5 Temporal Variability of Analytical Measurements

As with most other areas of technology, instruments continue to improve. Instrument manufacturers and laboratories are increasing data processing power, speed of analysis, and the reduction of chemical or electronic "noise." Any of these instrument improvements can be expected to improve the measurement of concentrations of environmental pollutants. This process can be illustrated for a variety of EPA methods. A case in point is EPA Method 1613 for determination of polychlorinated dibenzo-*p*-dioxins and polychlorinated dibenzofurans. Development of this method began in 1988. At the time, commercially available high resolution mass spectrometer systems were able to achieve a detection limit of approximately 4 pg/L and a ML of 10 pg/L. By the time that EPA proposed the method in 1991, the Canadian government published its own version that included a quantitation limit 5 pg/L. By the time EPA promulgated Method 1613 in 1997, many laboratories performing the analysis had replaced or supplemented their old instruments with newer models. As a result, many laboratories performing analyses using Method 1613 routinely measure sample results at levels 10 times lower than those analyzed routinely only 10 years earlier.

Although there is no such thing as a "perfect" measurement, the idea that "practice makes perfect" (i.e., analytical results get better with practice) applies to the quality of measurements made with a given method over time. We can demonstrate this using simple techniques like laboratory control charts. The improvements are a result of experience, as well as improvements in equipment over time. EPA expects changes in performance when new staff are trained. For this reason, many EPA methods specify that "start up tests" be repeated each time new staff arrive. It is not unusual to see slight increases in measurement variability as new staff are trained followed by a decrease back to normal level after analysts become sufficiently experienced with the analytical method. .

The use of quality control (QC) charts as a means of tracking method and laboratory performance as a function of time is described in EPA's *Handbook for Analytical Quality Control in Water and Wastewater Laboratories* (referenced in the 40 CFR part 136, Appendix A methods). Although these charts are instructive in tracking improvement or stability, they have two significant drawbacks: (1) they do not establish an absolute limit within which an analysis must be operated and (2) continued improvement can lead to unusually stringent limits that, eventually, will not be met. As long as absolute QC acceptance criteria (limits), such as those found in EPA methods, are established and as long as there is a recognition that stringent limits may be an artifact of improvement beyond what is routinely achievable, QC charts can be instructive in identifying statistically significant losses of, or improvements in, analyte responses in the region of interest. ASTM Committee D 19 adopted the philosophy of establishing absolute limits for analytical methods in approving Standard Practice D 5847.

Stakeholders commenting on EPA's 2003 assessment of procedures for characterizing the detection and quantitation limits of analytical methods expressed concern that EPA's MDL and ML are determined using a single batch of samples representing a "snapshot" in time, and do not account for the temporal variability that can occur in a laboratory from day to day (e.g., due to use of multiple analysts and instrumentation, changing laboratory conditions). Although the codified version of the MDL does not preclude laboratories from incorporating temporal variability into the procedure (e.g., it allows the use of more than 7 replicates and does not require that the replicates be analyzed in a single batch), many users understand the MDL to be a single batch procedure. EPA encourages, where appropriate, use of data gathered over an extended period of time to calculate an MDL because measurement capabilities tend to improve and laboratory conditions tend to vary. Detection and quantitation limit calculations can be supported by procedures that allow laboratories to affordably characterize such changes and improvements.

3.2 CWA Regulatory Issues Affecting Detection and Quantitation

Section 3.2.1 provides a brief overview of Clean Water Act activities that involve chemical measurements and are, therefore, directly impacted by detection and quantitation limit approaches. Specific issues to be considered in the context of these CWA applications and EPA's regulatory obligations are discussed in Sections 3.2.2 - 3.2.6.

3.2.1 Detection and Quantitation Limit Applications Under CWA

The Clean Water Act directs EPA, States, and local governments to conduct a variety of data gathering, permitting, and compliance monitoring and enforcement activities. Many of these activities depend directly on environmental measurements and, therefore, are affected, both directly and indirectly, by detection and quantitation limit approaches. Stakeholders commenting on EPA's assessment of detection and quantitation procedures stated that, because of the differing technical demands and regulatory and laboratory uses of detection and quantitation levels, the procedures for determining these values should be based on sound science. These stakeholders urged EPA to consider the implications of each technical decision it makes regarding determination of detection and quantitation values on the practical implementation of its regulations.

Several commenters believe that EPA, permit holders, and laboratories would be better served if different approaches to detection and quantitation were taken for each use. Commenters specifically cite uses as a start-up test in a single laboratory, as a value characterizing a given analytical method, as a test for approving a method modification or alternate test procedure (ATP), compliance monitoring, and as a permit compliance level. The Inter Industry Analytical Group, in particular, has recommended the following 3-part approach:

- A sensitivity test (as a test of start-up proficiency),
- A long-term MDL approach (for laboratory reporting),
- Full-range validation study (such as the ASTM IDE/IQE) for validation of new methods and for setting quantitation levels that will be used as permit compliance levels.

3.2.1.1 Method Development and Promulgation

Section 304(h) of the Clean Water Act (CWA; the "Act") requires EPA to promulgate test procedures (analytical methods) to be used for data gathering to support certification, permitting, and monitoring under the Act. These methods are promulgated at 40 CFR part 136, and include methods developed by EPA as well as those developed by other organizations, such as the publishers of *Standard Methods for the Examination of Water and Wastewater*, as well as AOAC-International, ASTM International, the U.S. Geological Survey, instrument manufacturers, and others. Upon request by a laboratory, permittee, instrument manufacturer, or other interested party, EPA also considers alternate testing procedures (ATPs). If EPA deems these ATPs to be acceptable for use, they may be published at 40 CFR part 136. A primary objective in promulgating methods developed by EPA and by other organizations is to provide the regulatory community, permittees, and laboratories with multiple options so that they may choose the method that yields the best performance at the lowest cost for the application.

In recent years, EPA has focused on developing methods for promulgation at 40 CFR part 136 where no other methods are available that meet an immediate or anticipated regulatory need. The National Technology Transfer and Advancement Act of 1995 (NTTAA) encourages government agencies to consider methods published by voluntary consensus standards bodies (VCSBs), such as Standard Methods and ASTM International, when VCSB methods are available. EPA accepts that many of these methods have been through a sufficient level of testing, peer review, and scientific acceptance to warrant proposal if they meet EPA's regulatory needs. When an individual laboratory, permittee, or other organization submits a request for approval of an alternate test procedure, however, EPA generally requires that the procedure be subjected to a level of testing that demonstrates that the method provides sensitivity, accuracy, and other measures of performance comparable to an approved method.

The lack of widespread consensus on detection limits has led organizations that develop methods to use different approaches, and many organizations have changed approaches over the years. Some stakeholders, who commented on the 2003 assessment, believe that method-specific detection and quantitation limits should account for interlaboratory variability, and therefore should be based on interlaboratory data. Other stakeholders believe that such a requirement would be overly restrictive and burdensome, resulting in fewer approved methods and technologies. The result is that a number of different approaches for detection and quantitation are embodied in the methods approved at 40 CFR part 136. The vast majority of the approved methods include the MDL which, as noted in Section 2.2.1, has been used by several EPA Offices, *Standard Methods*, AOAC, ASTM, and others. Other approaches embodied in the methods at 40 CFR part 136 include, but are not limited to:

1) a method "range" that is usually not defined, but is often interpreted as the lower end of the range in which pollutants either can be identified or quantified,
2) an "instrument detection limit" that has been defined by a variety of procedures, but is intended to capture instrument sensitivity only,
3) an "estimated detection limit" that may be based on best professional judgement, single laboratory data, or some other source of information,
4) a "practical quantitation limit," that has typically been determined according to one of the scenarios described in Section 2.3.1, and
5) "sensitivity" that is an undefined concept similar in result to the MDL.

A solution to this lack of consensus would be to require that all methods promulgated at 40 CFR part 136 contain uniform approaches for detection and quantitation. However, taking such action would be disingenuous and confound methods promulgation because:

- To date, no single detection and quantitation limit approach has emerged to meet the needs of all organizations for all applications.
- If EPA's were to select an approach that differs from those of other organizations, those organizations would be required to conform their method to accommodate the EPA approach. Doing so would mean that these organizations would have to invest additional laboratory resources to develop detection and quantitation limits that conform to OW definitions.
- If outside organizations decided against conforming their approaches to that of EPA, fewer methods would be promulgated at 40 CFR part 136. This would result in fewer options for the regulatory, permittee, and laboratory communities.
- If EPA selected an approach that has burdensome procedures for developing detection and quantitation limits, it could discourage development of innovative technology or method modifications.

Given these issues and EPA's desire to encourage the development of improved measurement techniques, and provide the stakeholder community with a variety of measurement options whenever possible, it would be counter-productive to allow method developers the choice of only one detection or quantitation limit approach, or to only promulgate those methods that contain this single approach. However, there are real benefits to standardization, all new methods developed by EPA for promulgation at 40 CFR part 136 should reflect such standardization, and EPA should strongly encourage outside organizations to include these standardized approaches in their methods. However; there was no clear consensus as to what this standardized approach should be. Industry advanced IDE/IQE procedures but others did not necessarily support them.

3.2.1.2 Verification of Laboratory Performance

Just as sensitivity is important for evaluating method performance, it is important to verify that a laboratory using a method can achieve acceptable levels of sensitivity for making measurements. Such demonstrations can take many forms and should be viewed in the context of the decision to be made. The analytical methods published at 40 CFR part 136 are designed for monitoring compliance with CWA permits. Most pollutants in permits have a numeric limit, and compliance with this limit is determined by laboratory analysis of samples from the waste stream or water body regulated by the limit. The laboratory that conducts such analyses must be able to demonstrate that its detection or quantitation limits are low enough to assure reliable measurements.

Thus, even where a method describes the sensitivity measured or estimated by the developer or the organization that published the method, some means are needed to demonstrate that a given laboratory can achieve sufficient sensitivity to satisfy the regulatory decision (e.g., monitoring compliance).

The EPA MDL procedure provides a means for verifying laboratory performance and has long been used in this fashion by EPA and various other Federal and State agencies as a measure of method sensitivity. Other procedures may be employed, including analysis of reference materials containing the analytes of interest at concentrations that are at or below the regulatory limits of interest, spiked samples that are similarly prepared (e.g., matrix spikes), or laboratory performance evaluation samples such as those used in laboratory accreditation studies. Several commenters on EPA's 2003 assessment recommended that a simple "sensitivity" test (e.g., determination of analyte recovery in a sample containing a low-spike concentration of the analyte) be used to evaluate or establish laboratory performance. Although at least two commenters submitted some ideas for conducting such a test, none were sufficiently or clearly detailed. However; EPA is open to consideration of approaches to verify lab performance.

The IDE and IQE were advanced by the regulated industry and subsequently approved by ASTM International as a means of characterizing the performance of a method in laboratories that participate in an interlaboratory study. These approaches were developed to establish detection and quantitation limits that could be met by any laboratory that participated in the study. However; the IDE/IQE cannot be used to verify individual laboratory performance.

Developers of the IDE/IQE have recognized that an analogous approach is desirable for single-laboratory application and have begun work on a within-laboratory detection estimate (WDE), to be followed by a within-laboratory quantitation estimate (WQE). As with the IDE/IQE, these approaches are intended to capture a wide range of sources of variability such as temporal variability, and include a prediction or tolerance limit (or both), but will not include interlaboratory variability. EPA would consider such single laboratory approaches if and when they are adopted by a voluntary consensus standards body, such as ASTM International. EPA will explore approaches for lab performance verification through the stakeholder process.

3.2.1.3 National Pollutant Discharge Elimination System

The National Pollutant Discharge Elimination System (NPDES) serves as a means by which EPA, States, and Tribes control point source releases of pollutants into the nation's waters. Under this system, individual facilities are issued NPDES permits that provide effluent limitations that restrict the quantities, discharge rates, and concentrations of pollutants that may be legally discharged. Typically, these limitations are based on technology-based standards. If, however, these technology-based limits are not adequate to protect the water quality standard designated for the facility's receiving water, stricter controls are warranted. In such cases, NPDES permits must contain "water quality-based" controls.

Development and Implementation of Technology-based Controls (Effluent Guidelines)

EPA promulgates national effluent limitations guidelines and standards under the authority of Clean Water Act Sections 301, 304, 306, 307, 308, and 501. The regulations allow the discharge of pollutants from normal industrial processes when the discharges have been treated using various levels of available and affordable treatment technologies. Functionally, these industry-specific guidelines establish standards for the quality of wastewater discharges to waters of the United States. They are generally stated in the form of concentration-based limits for selected substances that are not to be exceeded. For example, the maximum oil concentration in wastewater separated from oil pumped out of an offshore well and discharged on any single day shall not exceed 42 milligrams per liter (mg/L). This form is called a numeric effluent guideline limit or numeric limit.

Development and Implementation of Water Quality-based Controls

States designate water-quality standards for various bodies of water within their boundaries. Each standard consists of a designated use, criteria to support that designated use, and an anti-degradation policy. Examples of designated uses include public water supply, recreation, and propagation of fish and wildlife. A discharge that causes, has reasonable potential to cause, or contribute to an excursion of an applicable water quality standard requires a water-quality based limit. Such a water-quality based limit shall be established at levels that derive from and comply with applicable water-quality standards and must be consistent with the assumptions and requirements of any available waste load allocation for the discharge, approved by EPA pursuant to 40 CFR 130.7.

A special case occurs when the water quality-based effluent limit is less than the detection limit of the most sensitive analytical method. This case is addressed in Section 3.2.3 below, on compliance evaluation thresholds.

Permit Compliance Monitoring

Under Clean Water Act Sections 318, 402, and 405, NPDES permits are issued to owners of facilities that discharge wastewater to waters of the United States (e.g., coastal areas, lakes, rivers, streams, certain wetlands, etc.). Specific discharge limits are established either for individual facilities or for classes of facilities. Individual permits are established for industries with many site-specific issues that determine the substances discharged, such as the pharmaceutical industry in which the specific drugs produced could influence the water quality. NPDES permits generally specify the use of measurement methods promulgated at 40 CFR part 136 under the Clean Water Act Section 304(h) for purposes of compliance monitoring and other reports submitted under NPDES permits.

Detection plays a role in compliance monitoring because of concerns with measurement results at the low end of any analytical method. All measurement results are variable. At the low end of most measurement methods, there comes a point at which a particular measurement result is unacceptably likely (a policy decision) to have come from a sample in which the substance of interest is absent (zero concentration). Such a measurement result would be below the critical value defined by Currie (1995) and in common usage, would be called below detection. In practice, the reporting limit may be set equal to a critical value, detection limit, or quantitation limit. Assuming that the reporting limit is a detection limit of 1 mg/L of oil and grease, the measurement result would be reported as "less than 1 mg/L of oil and grease."

Stakeholders are particularly concerned with the use of the detection and quantitation limits for compliance purposes (e.g., judging whether a discharger is in compliance or whether a waterbody complies with its water-quality standards), for which a high level of reporting consistency and confidence in the data is required. Several commenters on EPA's assessment stated that procedures used to determine these limits should provide the certainty required to make regulatory decisions.

Several commenters suggested that there should be a single compliance benchmark for detection of each analyte that is independent of laboratory or method capabilities; laboratories used for compliance reporting would be required to demonstrate that they can detect at or below this level. These commenters note that such an approach would be particularly useful and appropriate for analytes with water quality (or other) standards set below current technological capabilities.

3.2.1.4 Non-Regulatory Studies and Monitoring

EPA conducts a variety of non-regulatory studies and monitoring activities to support its Clean Water Act programs. These activities range from long term surveys, such as the Great Lakes Water Quality Surveys that are conducted each spring and summer to monitor trends in water quality against established baselines, to short-term studies that are used to establish baselines, model pollutant cycles, and guide direction for future study and policy. Examples of such studies include the National Study of Chemical Residues in Fish that was conducted in the late 1980s (a follow-up to that study is currently underway), and the Lake Michigan Mass Balance Study conducted in the early 1990s.

When designing a study or monitoring program, EPA uses information about detection and quantitation limits, along with information on the risks associated with the pollutant(s) of interest and the cost of measurement, to select an appropriate method for measuring the pollutant. Accepting all positively valued measurement results and selecting a measurement method with a detection limit lower than the level of concern for the substance being measured would provide some assurance that measurement results associated with that concentration would be positively valued. Selecting a

measurement method with a quantitation limit lower than the level of concern for the substance being measured would generate measurement results that are easier to explain to the data user and the general public.

3.2.2 Descriptive versus Prescriptive Uses of Lower Limits to Measurement

The literature on detection and quantitation generally assumes that these procedures are descriptive, as opposed to prescriptive. In other words, detection and quantitation studies are described as characterizing the current performance of a laboratory or laboratories using a method to measure a substance rather than specifying specific performance benchmarks that a laboratory must meet to demonstrate and maintain proficiency. Two possible reasons for this treatment are: (1) the intended audience includes laboratory staff and measurement methods developers who wish to make new methods useable by as many laboratories as possible, and (2) the author may have an institutional reason for not attempting to control variability and thus lower detection and quantitation limits. On the other hand, the technology-based and water quality-based effluent limitations programs administered by EPA's Office of Water have an institutional goal of protecting human health and the environment. Providing this protection requires that the Agency be able to measure pollutants at ever lower concentrations. Establishing stringent standards and a compliance scheme for laboratories is one way to more rapidly develop the ability to measure at these concentrations. A prescriptive strategy concerning detection and quantitation limits would be to:

- Determine the detection and quantitation limits at multiple laboratories.
- Establish a detection limit and a quantitation limit for the method that is based on some performance of these laboratories. These limits could be established as the limits reported by the mean or median laboratory, or by some other criterion, such as the pooled value of the limits achieved by all laboratories or the limits that are met by a certain percentage of the laboratories.
- Use the established detection and quantitation limits as performance standards that must be demonstrated by laboratories that practice the method.

Such an approach is consistent with other performance standards included in EPA methods, such as standards for instrument calibration, recovery of spiked reference and matrix samples, etc.

The use of such an approach would help ensure that prescriptive detection and quantitation limits (i.e., performance standards) reflect the capabilities of multiple laboratories, rather than a single state-of-the-art research laboratory. Of course, it is possible that even when multiple laboratories are used to establish performance standards for detection and quantitation, some laboratories initially may not be able to achieve these standards. However, most laboratories facing this problem would try to improve and achieve these standards by investing in staff training, improved equipment, a stronger quality assurance program, or clean room practices and higher quality maintenance and operations.

There is a risk that some members of the laboratory community will not be able to meet the standard, either because they are not willing to invest the resources necessary to do so or for other reasons. That risk should be considered when using a prescriptive approach to detection and quantitation (i.e., establishing limits that act as performance standards). Conversely, the risk of using a descriptive approach is that it can result in detection and quantitation limits that reflect a broad community of laboratories, including those that have made little if any effort to control contamination and variability at these low levels, thus raising detection and quantitation limits to a level that is higher than desired for protection of public health and the environment.

3.2.3 Compliance Evaluation Thresholds

When technology-based effluent limitations are developed, the limits typically are at or above the quantitative measurement capabilities (e.g., the ML) of one or more analytical methods that are available to support compliance monitoring. Therefore, it is possible to monitor and evaluate permit compliance at concentrations with an accepted degree of measurement certainty.

A situation that arises frequently in addressing water quality-based limits is that permit limits may be set below the detection or quantitation limit of the most sensitive, approved analytical method. This is particularly true for pollutants that are toxic in extremely low concentrations or that bioaccumulate. A recommended approach for these situations is to include in the permit, the appropriate permit limit derived from the water quality model and the waste load allocation for the parameter of concern, regardless of the proximity of the limit to the analytical detection level, along with an indication of the specific analytical method that should be used for monitoring (See Technical Support Document for Water Quality-based Toxics Control, EPA/505/2-90-001, March 1991). Both the MDL and ML have been used as reporting limits or compliance evaluation thresholds in NPDES permits. EPA promulgated regulations for NPDES permits for dischargers to the Great Lakes basin that require the use of the ML for compliance assessment purposes (See Appendix F, Procedure 8, Part B of 40 CFR 132). EPA has recommended for most permitting situations that the compliance level be defined in the permit as the ML (See Technical Support Document for Water Quality-based Toxics Control, EPA/505/2-90-001, March 1991). Outside of the Great Lakes basin, it is important to note, however, that states that implement the NPDES permits program have not always followed EPA's guidance. The inconsistent use of the MDL and ML as reporting limits or compliance evaluation thresholds in NPDES limits suggest that EPA should develop additional implementation guidance.

From a technical standpoint, a one-sided limit that addresses false positives only, such as Currie's critical value or EPA's MDL, is the most appropriate approach for producing a compliance evaluation threshold for the situation in which the WQBEL is less than a detection limit in the most sensitive analytical method because the one-sided limit allows measurement to the lowest possible level while protecting a discharger from the risk of a false violation. For example, consider the situation in which 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (dioxin) is to be evaluated against the ambient water quality criterion of 13 parts-per-quintillion (ppqt). The most sensitive analytical method approved at 40 CFR part 136 is EPA Method 1613, with an MDL of 4 parts-per-quadrillion (ppq) and an ML of 10 ppq. The MDL is more than 300 times greater than the ambient criterion. Therefore, if dioxin is detected in the receiving water as a result of a discharge (i.e., the measurement result is greater than the MDL of 4 ppq), there has been an exceedance of the ambient criterion. Use of the ML as a compliance evaluation threshold is appropriate because it is the point at which the measurement could be considered reliable.

3.2.4 National versus Local Standards for Measurement

In accordance with the Settlement Agreement, EPA is re-examining the approaches of detection and quantitation used with methods approved for use at 40 CFR part 136. The Clean Water Act authorizes States and local governments to implement permits, with the requirement that they be at least as protective (stringent) as the national standards established by EPA. EPA recognizes that some States have implemented approaches to detection and quantitation that are either specific to that State, result in lower numeric limits in discharge permits, or both. Given that State and local governments use different approaches, a change by EPA with regard to this assessment of detection and quantitation procedures may have an unanticipated impact on those States and local governments.

3.2.5 Cost and Implementation Issues

Detection and quantitation limit procedures are typically employed by organizations that develop methods and by laboratories that use the methods. Method developers typically include governmental organizations such as EPA, NOAA, USGS, and DOE, or voluntary consensus standards bodies (VCSBs) such as the American Public Health Association (APHA), ASTM International, AOAC-International, and ISO/IUPAC. Method developers also may include manufacturers of instruments or supplies used in testing. Users of methods generally are the laboratories performing tests to assess and assure product quality, to support regulatory compliance monitoring, or to support scientific studies.

Method development requires a more diverse set of skills than method use because such development generally demands an understanding of quality systems, statistics, and analytical technologies. Personnel working for the method developer generally include a project manager, measurement analysts, who are experienced in several measurement technologies or very experienced in a specific, complex technology, and at least one statistician. Operating laboratories typically will not have a statistician, and the breadth and dept of the analyst experience may be less than in a method development laboratory, because an operating laboratory is focused on obtaining reliable results in the analysis of a given sample using a well-tested measurement technology.

3.2.5.1 Implementation of a Detection/Quantitation Limit Procedure by a Method Developer

The basic resources available to the method developer are time, money, and the technical skills of the developers's staff. The fundamental decision for implementing a detection or quantitation procedure is whether that procedure is intended to characterize the performance of the method at a well-performing laboratory or the performance of the method across a group of laboratories. If the procedure is intended to characterize the performance of the method across a group of laboratories, it is also necessary to decide if there will be some way to compare the performance of individual laboratories to the group performance standard. There are serious time, cost, and skill issues with each of these decisions. Ordering these decisions from the least resource intensive to the most, they are characterizing the performance of the method: (1) at a well-performing laboratory, (2) at a group of laboratories, or (3) at a group of laboratories with comparisons of individual laboratories. Other costs for the method developer could include planning, data management, reference laboratory services, and whether laboratories are willing to volunteer for the study or if their services must be purchased.

An independent decision is whether to assume a simple model for measurement variability and limit the number of test concentrations, iterate assuming a simple model, or to design a study of the relationship between measurement variation and the concentrations of the substances measured by the method. This decision will greatly influence the number of samples measured in the study and the required skill of the personnel who design, conduct and interpret the results of the study. If the laboratories do not volunteer for the study, then the direct cost for measuring these samples or blanks ranges from a few dollars per sample to more than $1,000 per sample for some analytes. Until the relationship between measurement results and standard concentrations becomes well known, such studies will require the active participation of professional statisticians in design, implementation, and analysis.

3.2.5.2 Implementation of a Detection/Quantitation Limit Procedure by a Laboratory

A laboratory may implement detection or quantitation procedures for its own quality control purposes, because of regulatory requirements, or to participate in a round robin study for a VCSB or some other organization. When participating in the study of another organization, the laboratory may

voluntarily accept some cost of the study for marketing purposes, professional development, or to benchmark the performance of the laboratory.

In each case, a detection or quantitation limit approach will be of little utility if it is not capable of being implemented by the laboratory. An advantage of straightforward approaches such as the EPA MDL, the ACS limit of detection, and the ISO/IUPAC critical value is that they can, in principle, be understood by analysts expected to use the approach. Likewise, the procedures described for implementing the MDL approach are straightforward and have been implemented by thousands of laboratories. In contrast, the ASTM IDE and IQE procedures are highly complex and, as a consequence, are beyond the capability of most environmental testing laboratories.

Highly complex procedures are usually more costly to implement than simple procedures. As noted in Section 3.1.5, method performance generally improves over time. This means that a detection and quantitation limit approach should be supported by procedures that will allow individual laboratories and other organizations to affordably characterize such improvement. Mandating interlaboratory studies using complex detection and quantitation procedures means that laboratories lacking statistical support staff, and seeking to develop new techniques or modify existing techniques to achieve improved measurement sensitivity would have to rely on, and perhaps even pay, other laboratories to demonstrate the sensitivity of their procedures. This limitation has the effect of hindering method development and improvement.

3.2.6 Use of a pair of related detection and quantitation procedures in all Clean Water Act applications.

In Section 3.2.1, we discussed several different applications for detection and quantitation limits under the Clean Water Act. To review, these applications are:

- Method development and promulgation,
- Method performance verification at a laboratory,
- Technology-based effluent guidelines development,
- Water quality-based effluent limits development,
- Permit compliance monitoring, and
- Non-regulatory studies and monitoring.

In the 2003 assessment, EPA argued that although EPA could develop a separate detection and quantitation approach for each of these applications and attempt to define and evaluate each of these approaches in our re-examination of detection and quantitation approaches, the resulting matrix of applications and approaches would cause confusion for stakeholders, such as regulators, permittees, and the laboratory community. To minimize this confusion, EPA suggested that a single pair of related detection and quantitation procedures could meet the needs of all CWA applications. Some commenters disagreed with this approach and recommended that at least two distinct procedures should be used, one for method development and one for verifying laboratory performance.

3.2.7 Accepting the Procedures of Voluntary Consensus Standards Bodies

In February 1996, Congress enacted Public Law 104-113 (15 *USC* 3701), the National Technology Transfer and Advancement Act (NTTAA). This act directs *"federal agencies to focus upon increasing their use of (voluntary consensus) standards whenever possible, thus reducing federal procurement and operating costs."* The Act gives Federal agencies discretion to use other standards

except where the use of voluntary consensus standards would be *"inconsistent with applicable law or otherwise impractical."*

The NTTAA is implemented by Federal agencies based on the policies described in Circular A-119 from the Office of Management and Budget (OMB). The current version of this OMB circular was published in the *Federal Register* on February 19, 1998 (63 FR 8546). Neither the NTTAA nor Circular A-119 require that agencies replace existing government standards with standards from a voluntary consensus standard body (VCSB). If EPA already has standards in place for detection and quantitation approaches, EPA is not obligated by NTTAA to replace these with VCSB standards. Although some stakeholders commenting on EPA's 2003 assessment encouraged EPA to allow use of alternative procedures for determining detection and quantitation levels, commenters in general did not support eliminating continued use of the MDL or ML.

Circular A-119 also discusses the effect of the policy on the regulatory authorities and responsibilities of Federal agencies. The circular states that:

> *"This policy does not preempt or restrict agencies' authorities and responsibilities to make regulatory decisions authorized by statute. Such regulatory authorities and responsibilities include determining the level of acceptable risk; setting the level of protection; and balancing risk, cost, and availability of technology in establishing regulatory standards. However, to determine whether established regulatory limits or targets have been met, agencies should use voluntary consensus standards for test methods, sampling procedures, or protocols."*

Thus, EPA is responsible for establishing the levels of risk and protection, not only for the regulatory limits applied to discharges, but also to the risks of decision errors (e.g., false negatives or false positives) in the detection and quantitation approaches applicable under the Clean Water Act.

Finally, Circular A-119 describes two types of technical standards: performance standards and prescriptive standards. A performance standard is defined as:

> *"a standard ... that states requirements in terms of required results with criteria for verifying compliance but without stating the methods for achieving required results."* In contrast, a prescriptive standard is one *"which may specify design requirements, such as materials to be used, how a requirement is to be achieved, or how an item is to be fabricated or constructed."*

Neither the NTTAA nor Circular A-119 direct agencies to favor performance standards over prescriptive standards, or vice versa. EPA believes that the current MDL procedure is a prescriptive standard, in that it specifies both the design of the MDL study and how the requirement to establish method sensitivity be achieved. There is some obvious flexibility or opportunity for judgement in employing the MDL procedure, and much of the historical debate over the utility of the MDL procedure would suggest that it may not be prescriptive enough. The alternative detection and quantitation approaches evaluated in this document, including the approaches submitted by commenters on the 2003 assessment, also are prescriptive, not performance, standards.

To effect a performance-based approach to estimating detection and quantitation limits, an option that EPA may consider is to allow method developers, laboratories, and others the choice of any one of a variety of approaches to establishing these limits, including the existing MDL procedure or a VCSB

standard. Thus, establishing method sensitivity could be considered a performance standard under NTTAA and Circular A-119, rather than a prescriptive standard. That these different approaches (prescriptive standards) yield different answers would be immaterial if EPA evaluates the answers relative to a specific decision, i.e. the benchmark becomes a performance rather than a prescriptive standard. That evaluation should not be divorced from knowledge of the decision to be made (e.g., the regulatory limit for a given pollutant).

3.2.8 Alternative Procedures

One of the peer reviewers who evaluated a draft version of the February 2003 assessment document noted that:

> *"EPA has stated in the TSD that one primary procedure is needed for clarity and to avoid confusion among stakeholders. If alternate procedures are needed, the EPA Clean Air Act system of reference and equivalent methods has worked well, and could be a model for EPA to follow under the Clean Water Act."* (Cooke, 2002)

EPA currently assesses and has approved at 40 CFR Part 136 methods that employ an alternative procedure for establishing method sensitivity. This approval process includes an overall evaluation of the suitability of the method in entirety and thus includes the detection or quantitation approach used to establish the performance specifications listed in the method.

The peer reviewer is referring to the system of reference methods used under the Clean Air Act. This system is similar to the existing "alternate test procedure" (ATP) program for analytical methods currently used within the Office of Water. The difference between the ATP program and the case of the procedures for establishing detection and quantitation limits is that in an ATP program, the goal is clear and agreed upon (i.e. is a method appropriate for CWA applications), whereas there remain fundamental theoretical issues surrounding the relative merits of the various detection and quantitation approaches that are the subject of this document.

For example, when a test procedure is developed for use in the Clean Air Act or Clean Water Act programs, the reference method is designed to measure Analyte X, in Matrix Y, at some concentration related to a regulatory need (e.g., a compliance limit) or environmental study. Alternative procedures may be capable of making measurements of Analyte X in Matrix Y, at the level of concern, using completely different instrumentation. Thus, the demonstration of equivalency between the reference method and a possible alternative method is judged using a metric that consists of Analyte X, Matrix Y, and the level of concern, as well as other aspects of method performance.

In contrast, the primary differences between the EPA MDL/ML approaches and potential alternatives such as the ASTM IDE and IQE are related to the philosophical differences of how detection and quantitation limits should be derived and applied. These differences are described at length in this chapter and the rest of the Assessment Document. Therefore, EPA does not believe that a variant of existing ATP programs is likely to be an effective model for assessing other detection and quantitation procedures.

A stakeholder commenting on EPA's 2003 assessment recommended that EPA adopt alternative procedures in Appendix B of 40 CFR 136 as site-specific alternatives to the MDL and ML when such an alternative is determined to be necessary by a discharger and/or regulatory agency (e.g., in special cases when more scientifically rigorous procedures are needed). As noted previously, EPA has reviewed and

approved at 40 CFR Part 136 methods that employ an alternative procedure for establishing method sensitivity as part of an overall evaluation of the suitability of the method. EPA has done so without need of any revisions to appendix B at 40 CFR part 136.

3.3 Statistical Issues

This section provides a brief explanation of the key statistical issues involved in the development of detection and quantitation limits.

3.3.1 Sources of Variability

Various known and unknown sources of variability will influence measurements made by a laboratory using a specific method. These sources may include random measurement error, differences in analysts, variations between different equipment manufacturers and models, variations in analytical standards, routine fluctuations in equipment performance, and variations in facility conditions (e.g., varying levels of background contributions).

There are several ways in which some of these sources of variability can be controlled. One is a strong quality assurance (QA) program that includes use of: 1) trained and qualified staff, 2) properly maintained equipment, 3) standards that are fresh and properly prepared and stored, 4) written standard operating procedures and methods for all sample handling, analysis, and data reduction/reporting activities, 5) procedures for monitoring ongoing laboratory performance, and 6) quality control (QC) samples and QC acceptance criteria to ensure that the laboratory systems are in control. The EPA methods promulgated at 40 CFR part 136 require the use of qualified staff, appropriately cleaned and calibrated equipment, and properly prepared standards. Many of these methods also provide detailed steps for performing all sample handling and analysis activities, and detailed requirements for analysis of specific quality control samples with corresponding QC acceptance criteria.

Even when prescribed EPA method requirements and guidance are used, however, it is not possible to completely eliminate all variability that can occur within or between laboratories. Even with procedures in place to control quality and reduce variability, it should be recognized that some laboratories, analysts, and instruments may achieve lower detection and quantitation limits than others. Ultimately, some laboratories may not be capable of meeting low-level measurement requirements without some effort to improve operations.

Many of these sources of variability are considered in establishing detection and quantitation limits for analytical methods used under EPA's Clean Water Act programs because these detection and quantitation limits are first established in single-laboratory studies, then evaluated or verified in multiple laboratories, and, where necessary, further evaluated in an interlaboratory study. These studies include evaluation of method performance characteristics, including detection and quantitation capabilities, in multiple laboratories using multiple matrices, analysts, instrumentation, reporting activities, standards, and reagents. Although detection and quantitation are not evaluated in the various matrices used in these studies, EPA's MDL procedure includes instructions for determination matrix-specific MDLs.

Some stakeholders commenting on EPA's assessment of approaches to detection and quantitation believe that accounting for these sources of variability when determining detection and quantitation limits is necessary because relative variability increases as the lower sensitivity limits of a method are approached. Some stakeholders believe, for example, that a methodology for detection and quantitation has to address the variability that occurs across laboratories (interlaboratory variability) using the same

analytical method. Other stakeholders believe, however, that interlaboratory variability is not an issue because detection and quantitation decisions are made in a single laboratory. Some stakeholders believe that procedures should address the long-term variability that can occur within a single laboratory over time. As discussed in section 3.1.5 of this chapter, EPA encourages, where appropriate, gathering data to address temporal variability. EPA acknowledges that interlaboratory variability is very important during the methods development process and should be incorporated, as appropriate, during the process. EPA also recognizes that within lab variability should be considered when establishing laboratory performance.

Over the years, stakeholders have noted that the variability that can result from application of analytical methods to different matrices also should be addressed by procedures for determining method-specific detection and quantitation limits. However, it has been EPA's experience that matrix effects typically can be overcome using various sample processing procedures. In EPA's interlaboratory validation studies of the 600-series wastewater methods, the recoveries of some organic analytes from real-world matrices were closer to 100% than were recoveries from a reagent water matrix. This effect is thought to be attributable to dissolved solids in the real-world matrix that, in effect, "salt out" the organic compounds. EPA does not believe it is appropriate or feasible to aggressively pursue matrix effects in establishing detection and quantitation limits (i.e., EPA has not attempted to find worst-case matrices in order to maximally exacerbate matrix effects). Instead, EPA considers the type of matrices that would be regulated under the Clean Water Act (e.g., the effluents that are discharged from properly designed and operated secondary treatment plants). Further discussion of matrix effects can be found in Section 3.1.3.

Because detection and quantitation limits focus exclusively on the capabilities of the measurement process, a source of variability that is *not* considered in any of the detection and quantitation limits is the variability that is associated with sample collection. If the sample is not representative of the population from which it was collected, then the variability associated with measurements made in the region of detection or quantitation may be immaterial. For example, EPA's Technology Innovation Office conducted a study to characterize the effects of sampling variability on measured results. In that study, results from seven discrete samples collected within a two-foot distance of one another were evaluated. Each sample was analyzed for the explosive TNT on-site using a colorimetric test kit, and in a laboratory using EPA SW-846 Method 8330 (high-performance liquid chromatography). Analysis of the results from these measurements indicated that 95% of the total variability was due to sampling location and only 5% was due to differences between the analytical methods. Put another way, differences in sampling location caused 19 times more uncertainty in the data results than did the choice of analytical method, over a distance of only 2 feet (Crumbling, 2002). While this result may not be typical, and EPA does not mean to diminish the importance of understanding measurement error in the region of detection and quantitation, EPA believes it is important to understand it in the context of the overall sampling and analysis error.

3.3.2 False Positives and False Negatives

3.3.2.1 False Positives and False Negatives in Making Detection Decisions

In this section, we discuss the impact of detection, quantitation, and reporting levels on false positive measurement results and false negative measurement results. The definitions of false positives and false negatives are directly related to the concepts of critical value and detection limit used by Currie (1995). These terms were adapted from statistical decision theory to establish the framework for decision making with regard to detection of analytes. The critical value (Lc), as defined by Currie, is the point at which the detection decision is made. That is, measured values that are less than the critical value are judged to be not statistically different from blanks ("not detected"). Measured values that are no less than

the critical value are judged to be statistically different from the blanks ("detected"). Denoting measured values that are less than the critical value as non-detects constitutes censoring and is discussed in more detail in Section 3.3.5.

The critical value is defined such that when the analyte is not present in a sample, there is a small possibility that a measurement will exceed the critical value. A measurement that indicates the critical value has been exceeded is, therefore, the result of one of two circumstances: (i) the analyte is present in the sample; or (ii) the analyte is not present in the sample and, by chance, the measurement has exceeded the critical value. The occurrence of (ii) is defined in statistics as Type I error ("false positive"). A measurement that is less than the critical value occurs when: (iii) the analyte is not present in the sample; or (iv) the analyte is contained in the sample at the hypothesized concentration but the measurement procedure fails to indicate its presence. The occurrence of (iv) is defined as the Type II error ("false negative").

The following table summarizes possible situations:

	State of the Sample	
Decision	*Concentration=C, where C>0* *Analyte Present*	*Concentration=0* *Analyte Not Present*
Concentration=C, where C>0	Correct (i)	Type I error (ii)
Concentration =0	Type II error (iv)	Correct (iii)

Calculating the probability of a Type I error only requires assumptions regarding the distribution of observations under the hypothesis that the concentration is equal to zero. In the terminology of statistical decision theory, Concentration = C, where C>0 corresponds to a true value is referred to as the "Alternative Hypothesis" (see, e.g. *Introduction to Mathematical Statistics*, by Hogg and Craig, 5th edition, [1995]). When C is hypothesized, assumptions need to be made about the distribution of observations at Concentration=C for the probability of Type II error to be evaluated.

In analytical chemistry, the probability of Type I error is often called the "false positive" rate and the probability of Type II error is often called the "false negative" rate. The statistical alternative hypothesis should be specified before introducing the false negative rate. An error common to some published discussions of false negative rates and detection and quantitation concepts is to state that use of Currie's detection limit as a reporting limit or action level will somehow "control" the rate of false negatives. This is both incorrect and counter-productive, because a single level cannot control false negative rates.

Currie introduced the idea of a Detection Limit, Ld, in place of a statistical alternative. The Detection Limit is *not* a part of the detection decision process (i.e., is the concentration in the sample statistically different from the blank?). The Detection Limit is defined such that when the true concentration of an analyte is equal to the Detection Limit, there is a small probability that a measured value will be less than the Critical Value (detection decision-making level in this case), and thereby result in the false negative decision.

One of the peer reviewers of EPA's 2003 Technical Support Document (the TSD) stated:

> *"Also, to reemphasize, the single most problematic issue when developing a detection limit is correction for false negatives. I took from the TSD (in §3.3.6) an implicit emphasis on LC-type values such as the MDL [when correctly calculated, as in (1)], as motivated by an underlying sort of practical/environmental conservatism that essentially removes false negatives from the estimator's development. I am willing to accept this interpretation. I suspect the fray will continue, however, since there seems to be a fair amount of confusion on the issue in the analytical chemistry literature. The bottom line from my reading of the TSD is that, in effect, we are calculating an LC, but using terminology that makes some readers think it's an LD. I can accept the argument that false negative errors are not the critical issue here, and hence that the approach is reasonable (once correct calculations are undertaken). But, the Agency should put forth an effort to overcome this confusion in terminology. (I expect they will ask me how, and in reply I'd suggest emphasizing that an LC calculation is a form of decision limit, not a detection limit. But here I suspect many users will still confuse the terms, or reverse their meaning, or not see the difference, or who knows what else? I don't know how winnable this battle is...)* " (Piegorsch, 2002)

To illustrate the intent of Currie's detection limit, consider a case where the detection decision-making level is set equal to Currie's critical value, and a sample is spiked at a true concentration equal to Currie's detection limit. Given a large number of measurements on this sample, about 99% of the measurement results will be reported as being measured above the detection decision-making level, and 1% of the measurement results will be reported as being measured below this level. Knowledge of the lowest true concentration that will routinely produce acceptable results (e.g., Currie's *detection limit)* can be used to determine if the measurement method meets the needs of a study. For instance, a study concerned with a wastewater treatment technology that is not expected to be effective at concentrations below 10 mg/L may call for a relatively inexpensive measurement method capable of detecting the analyte at 10 mg/L, rather than a more expensive measurement method capable of measuring a hundred times lower.

3.3.2.2: Effect of Bias on Rates of False Positives

The presence of bias in a method can have a strong effect on the rate of false positives associated with detection limit estimates. For example, in defining the critical level, Currie assumed that blank results follow a Normal distribution centered about zero (0). However, for some methods and analytes, this assumption may not hold due to factors that can and should be controlled, such as calibration errors and high background contamination. In many cases, bias can lead to either under- or over-estimation of detection limits. In cases such as these, not taking bias into account when determining detection and quantitation limits (using the mean or median of the results, for example) may influence false positive rates.

3.3.3 Use of Multiple Replicates

Existing detection/quantitation procedures are based on estimating the standard deviation of blank or spiked replicates. Statistical estimates tend to be less variable when the number of replicates increases. Some commenters on EPA's 2003 assessment believed that use of only seven replicates over a short period of time results in a substantial underestimation of the MDL. However EPA's MDL procedure does not limit the maximum number of samples that the laboratory may use to estimate the MDL; the procedure

simply sets a minimum number of seven replicates. Laboratories may choose to improve their estimates of the standard deviation that is used to calculate the MDL by analyzing more than seven replicates.

3.3.4 Statistical Prediction and Tolerance

To define a critical value, a detection limit, or a quantitation limit, different descriptive terminology is used to distinguish differences in the numeric value of the limit. The following example uses a critical value, but the questions motivating detection and quantitation limit decisions may be phrased in a similar fashion.

In setting a critical value, do we want a critical value that tells us how likely it is that:

- A measurement result was produced by measuring a blank sample,
- The next measurement result will be produced by measuring a blank sample, or
- The next [pick any number] of measurement results will be produced by measuring a blank sample?

In statistical terms, these three objectives may be addressed, respectively, by application of methodology for determining:

- Percentiles;
- Prediction intervals; and
- Tolerance intervals.

Percentiles are fairly straight forward to interpret, i.e., they specify the percentage of a distribution that falls below a given percentile value. Prediction and tolerance intervals are, in effect, confidence intervals on percentiles and can be somewhat more difficult to understand and apply. There are many excellent textbook and literature references that present the theory and application of prediction and tolerance intervals such as Hahn and Meeker, *Statistical Intervals*, 1991, and Pratt and Gibbons, *Concepts of Non-parametric Theory*, 1981. Hahn and Meeker describe at length the different statistical intervals including their properties, applications, and methodology for constructing the intervals. Pratt and Gibbons have an excellent discussion of tolerance intervals that is general in application due to the non-parametric perspective, i.e., no distributional assumptions are required for the results to be valid.

One of the peer reviewers of EPA's 2003 assessment stated:

> *"Tolerance intervals are inappropriate for environmental monitoring. The main issues here are 1) is the true concentration greater than some specified safe action level, with sufficient confidence, and 2) what interval of possible concentrations is consistent with one or a series of measurements, with a specified degree of confidence? Both are statements about a given sample or series of samples, and not about the hypothetical variability of future estimates. Suppose that one has a sample of 10 observations with mean concentration of 1 ppb and standard deviation of 0.5 ppb. Then the estimated 99% critical level is (2.326)(0.5) = 1.2 ppb. One may choose to use a t-score instead of a normal score so that the chance that a future observation will exceed this level is in fact 99%. In this case, the critical level estimate would be (3.250)(0.5) = 1.6 ppb. This does actually correspond to a prediction interval for future observations from a zero concentration sample.*

"If one asked instead for a 95% confidence interval for the .99 percentage point of the true distribution of measurements (assuming normality) when the true quantity is zero, this can be calculated approximately using a chi-squared distribution and covers the interval (0.9 ppb, 2.4 ppb). It does not, however, make sense to use 2.4 ppb as a threshold, since the chance of a future observation exceeding 2.4 ppb when the true mean concentration is 0 is about .0005, far smaller than the intended false-positive limit of .01." (Rocke, 2002)

Another of the peer reviewers of this assessment stated:

"the operational definition as taken from pp. 5-2/5-3 of

$$MDL = t\ 0.99\ (df)\ S$$

does not correspond to a confidence statement that I can interpret.... This should be replaced, although I agree that a number of statistical quantities could be used; this is where the "fray" seems to be most boisterous. (By the way, the TSD, and I, should be more careful in the use of statistical terminology. We both refer often to confidence "intervals," when in fact the quantity of interest is a confidence limit — or tolerance limit, etc. — on some underlying parametric quantity.)...

"If we accept the TSD's argument on p. 3-25 that the practical value of tolerance limits is limited, then the MDL should be viewed as a prediction limit. And if so, it must contain an additional term as per Gibbons (1994, p. 98):

$$t_{0.99} = (df)\ S \sqrt{1 + \frac{1}{n}}$$

"One caveat: although I think the prediction limit argument is acceptable, if the use of tolerance limits rather than prediction limits is in fact desired, then Gibbons' (1994, p. 99) presentation or an equivalent approach should be used instead to correct the MDL calculation." (Piegorsch, 2002)

Similarly, Hahn and Meeker describe situations in which the various intervals or limits are appropriate to use. (As noted by the peer reviewer, the terms "intervals" and "limits" are sometimes used interchangeably). They also give examples of the sort of applications that are suitable for each type of limit although the decision to use a particular type of limit in a given application is not determined strictly by theoretical considerations. It is also a matter of judgment.

Prediction intervals contain results of future samples from a previously sampled population with a specified level of confidence. Prediction limits are not estimators of parameters such as means or percentiles. For example, a prediction interval may be constructed to contain future sampling results expressed as a mean or standard deviation of a future sample or all of a certain number of individual future sampling results.

While the theoretical construct underlying Currie's critical level is clear and straightforward, EPA recognizes that estimating this level from limited data is less straightforward and the choice of an appropriate statistical methodology involves policy judgements that might legitimately differ for different uses of the MDL.

3.3.4.1 Tolerance Intervals

Tolerance intervals contain a specified proportion of a population of measured values with a given statistical confidence level. For example, we say that a proportion, P, of a population is contained within the intervals (L_1, L_2) with (1-α)100% confidence. Random variables that are the lower and upper ends of the interval, L_1 and L_2 , respectively, are referred to as tolerance bounds. A tolerance bound is therefore the endpoint of an interval of random length that is determined on the basis of having a specified probability of 1-α that its coverage of the population is at least equal to a specified value P. The quantity 1-α is referred to as the confidence level for the interval and P is the minimum proportion of the population contained in the interval. Tolerance bounds are not estimators of values such as a mean or a percentile but rather bounds that are always guaranteed to contain the desired value at some level of statistical confidence. Pratt and Gibbons discuss this and other properties that affect the utility of tolerance intervals and create difficulties in their interpretation and application.

In effect, the determination of what, if any, interval to use is a policy decision. The choice should consider how easy it is to estimate the interval you want under the conditions that exist. As Pratt and Gibbons point out, the interpretation of tolerance intervals (and analogously, prediction intervals) can be problematic, especially when issues of sample size and the choice of confidence level come into play. Pratt and Gibbons cite examples where the interplay of sample size and high percentile and confidence levels make tolerance intervals useless.

3.3.4.2 Use of Tolerance and Prediction in Setting Detection and Quantitation Limits

Statistical intervals can be, and have been by a number of authors, adapted for use in setting detection and quantitation limits. The basic approach requires a functional definition of detection or quantitation that includes a statistical term or terms. An interval could then be constructed about the statistical term which could be used to assess the detection or quantitation limit, or make an adjustment to a calculated value that would result in the detection or quantitation limit. For example, most detection limit estimators are functionally dependent on an estimate of standard deviation of measurement error. A statistical interval could be constructed about the standard deviation and the length of the interval could be used to assess the detection limit. The end points of the interval could be used as the basis for an adjustment (upward or downward) in the calculated limit.

The error rates in ASTM's IDE Standard Practice are based on statistical tolerance intervals (i.e., the nominal Type 1 error rate is 5% (5%=100%-95%), and the nominal Type 2 error rate is 10% (10% = 100%-90%)). Several stakeholders have commented that the use of a tolerance interval approach can protect, at a 99% level of confidence, against false positives and false negatives, and that tolerance intervals become increasingly important with a decreasing sample size. For example, if the sample standard deviation is determined with 7 measurements and all sources of variance are properly represented in the 7 measurements, then there is approximately a 5% chance that the true population standard deviation will be more than two times the sample standard deviation. For a typical ICP determination of 20 or more elements this means that at least one is likely to have a calculated MDL two times lower than it should be. Obviously the false positive rate for this element will be large.

The use of prediction and/or tolerance limits in setting detection and quantitation limits should be evaluated in the context of the specific application and policy considerations. In practice, the effect of adjustment of detection and quantitation limits by use of prediction and tolerance intervals can be quite large, depending on the amount of data available and the choices of percentiles and confidence levels.

3.3.5 Censoring Measurement Results

Measurement results are often reported as less than some detection, quantitation, or reporting limit (see Section 3.2.1.3, Permit Compliance Monitoring) without providing a single best estimate for the numeric result. For example, if a direct reading of the measurement results indicates a concentration of 3 mg/L and the reporting limit for the substance is 5 mg/L, the laboratory may only report that the measurement result is less than 5 mg/L. Statisticians call this suppression of results that are less than a specified amount "censoring." Reasons for the practice of censoring relate directly to issues surrounding the development of detection and quantitation limits (i.e., the premise that measurement results below certain low levels may not be useable for certain purposes).

Some data users prefer to use the actual measurement results (even if they are negative values), rather than to censor the results at a reporting or detection limit, because censoring data at such a limit loses information about low-level measurements and can introduce bias into the data set. If all low values are eliminated, then the average (mean) of the remaining data would have a positive bias. In other words, while negative or extremely low values may be considered problematic by some, they are of value to statisticians and modelers, because they convey useful information about the distribution of results.

Some programs, such as EPA's Superfund Contract Laboratory Program, require laboratories to report measurement results in conjunction with a qualifier that the result is below a specified detection, quantitation, or reporting level. In the example provided in the first paragraph of this section, the laboratory might report both a measured value of 3 mg/L and a reporting limit of 5 mg/L. Under certain assumptions, measurement results below the specified reporting level could then be used to calculate averages and statistical estimates that would be superior to estimates calculated using censored data.

Although the Superfund approach provides the greatest degree of flexibility for data users, it should be used with care. First, data users who choose to use values reported below a detection or quantitation limit need to have a firm understanding of the limitations of those data. Second, and as noted in Section 3.2.1.3, Permit Compliance Monitoring, reporting data below a detection or quantitation limit can lead to misinterpretation.

One of the peer reviewers that evaluated EPA's 2003 assessment of detection and quantitation limit approaches noted that European Union (EU) has adopted another variant for reporting or censoring data.

> *"In this case, the EU has adopted EPA Method 1613B (for analysis of dioxins and furans) as well as EPA's MDL approach. However, the EU has further specified that the MDL be used as an Upper Bound reporting limit where all non-detects are found in the analysis of human or animal foodstuff. This forces laboratories to achieve levels available with modern instrumentation, otherwise, the Upper Bound reporting level is above the regulatory compliance level, and the data (or foodstuffs) are rejected"* (Cooke, 2002).

EPA agrees that this approach, which yields a "worst-case" (or highest possible) estimate of the pollutant concentration, can serve as an incentive to the analytical and regulated community to pursue measurements at the lowest levels which analytical methods are capable of achieving. However, EPA also cautions that this approach effectively censors measurements made below the MDL and could yield an overestimate of the concentration of the analyte of concern.

Several stakeholders have requested that EPA provide specific guidance and procedures regarding data censoring and reporting, particularly when data are reported for compliance evaluation. EPA notes that the decision to censor data is a data reporting and data use policy issue, not a laboratory issue. This holds without regard to what detection or quantitation limit approach is used. The EU approach reflects a similar point of view, in that it relies on the MDL as a detection approach, but also establishes this limit as the reporting level for non-detects to better encourage development of lower MDLs. However, EPA also recognizes that laboratory methodologies and data reporting and use policies are interrelated.

3.3.6 Outliers

Outliers are extreme or aberrant measurement values that, on inspection, do not follow the characteristics of a set of data. Outliers may be generated by a number of causes, such as errors in following an analytical procedure, errors in recording results, or the result of extreme random variation in a properly operating process. For example, if a new measurement method is being tested but the laboratory fails to follow the procedure correctly when analyzing some samples, the associated measurement results may stand out as outliers. A graphic example is provided in Figure 3-1, which shows measured concentrations of aluminum versus spike concentrations for analytical results obtained using EPA Method 1620. At a spike concentration of 250 µg/L, one of the measured values is approximately 750 µg/L. This result visually stands out from the rest of the values, and may be an outlier.

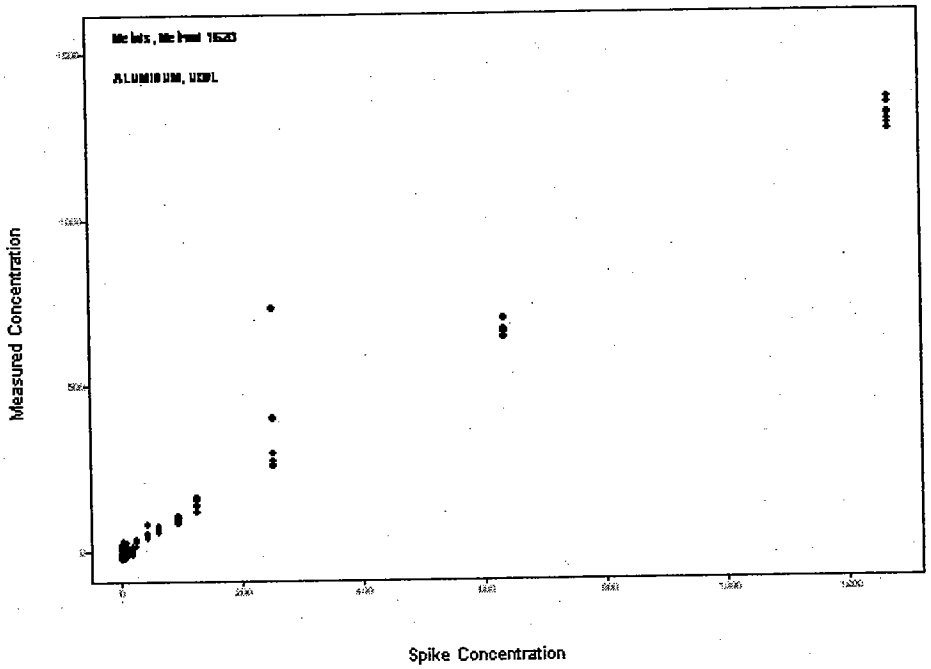

Figure 3-1

Stakeholders commenting on EPA's assessment of detection and quantitation procedures generally believed that outliers should be identified and removed from data used to determine detection and quantitation limits. Commenters added that, although it would be helpful to have specific instructions for identifying outliers, application of the instructions should be optional (i.e., to the discretion of the data user).

A common process for identifying potential outliers is to apply one or more statistical procedures for identifying values far from the mean (average) of the data. An example of such a procedure is described in ASTM Practice D-2777.

Because extreme values can be expected to occur on occasion, it may not be appropriate to exclude them from the results used to develop detection or quantitation values. As recommended in the ASTM procedure, the first step is to contact the laboratory to try to determine and resolve the cause. A review of the analyst's records associated with the measurement may establish whether the extreme value was caused by failure to follow the method or by some rare event associated with the method. If the method under study was not followed, or there is a known or suspected analytical error, it is appropriate to exclude the measurement result from the detection or quantitation analysis. If the measurement result is a rare event associated with the method under study it may also be appropriate to exclude the measurement result from the results in the study. EPA believes that results that are associated with spurious errors that cannot be corrected will invalidate the measurement and should not be incorporated into the MDL

determination.

3.3.7 Detection and Quantitation Studies

3.3.7.1 Study Design

The issues associated with the design of detection and quantitation studies include:

- how effectively a selection of spike concentrations can be used to correctly determine which model type should be used to model variability,
- the extent to which the distance between spike concentrations can impact estimates of detection and quantitation limits,
- how to reduce the influence of uncontrollable factors in the measurement process (probability design),
- how complete to make the design factors in terms of the physical measurement process, and
- how flexible to make the design factors in terms of the physical measurement process.

Spike Concentrations and Modeling

If a model under consideration cannot be described by the number of spike concentrations in the design, then it is not possible to tell if the model is appropriate. To take the simplest example, it is not possible to describe the slope of a line associated with linearly increasing variation from a single spike concentration. Two well-spaced spike concentrations would allow you to estimate a slope, but would provide no idea of the variability of the estimate. Three well-spaced spike concentrations represent the minimum requirement for estimating the linear relationship and the variability of that relationship.

Clayton *et al.* (1987) describe the relationship between the spread of the spike concentrations, the number of spike concentrations, and the number of replicate measurements with regard to estimated variability when a linear model is used. While the specific equation used in this paper does not apply to all models, it indicates principles that do apply. Increasing the number of replicate measurements and reducing the spread of the spike concentrations are all expected to reduce estimated variability along with the associated detection and quantitation limits. However, one of the components of variability associated with detection and quantitation is that associated with estimating the calibration relationship. To account for this source of variation, it may be appropriate to cover the entire calibration range. On the other hand, many replicates at a high concentration may improperly weight the data in favor of high detection and quantitation estimates.

It is also important to note that modeling of variability introduces modeling error, and direct measurements of the variance in the region of interest may provide a more appropriate estimate of variability, especially where the change in variance over this region is small.

Probability Design

The process known as randomization is an important statistical consideration in the design and interpretation of experimental studies. Randomization involves the allocation of experimental units to factors and treatments under study according to a design determined by probability. Randomization avoids bias and systematic errors that can occur in studies where randomization is not used. Randomization is discussed in classic texts such as *Statistics for Experimenters,* by Box, Hunter, and Hunter (1978).

In studies of measurement methods, randomization can be used in the process of creating spike concentration solutions and the ordering of analyses. However, randomization has practical drawbacks, particularly with regard to studies designed to establish detection or quantitation limits. For example, consider a simple study involving the analyses of samples spiked at five concentrations of the analyte of interest, with five replicate samples analyzed at each concentration. A total of 25 analyses are required for the study, and the analyses of the samples can be organized in a 5 by 5 matrix. A random number is assigned to each block in the matrix as a means of randomizing the order of the replicates at each concentration.

By virtue of this randomized design, a sample with a high concentration of the analyte of interest may end up being analyzed immediately prior to a sample with a very low concentration of the analyte. Unfortunately, this can lead to problems that result from the "carry-over" of analyte within the instrumentation from one analysis to the next. When carry-over occurs, the apparent concentration of the low-concentration sample can be inflated because some of the high-concentration sample 1 may be carried into the low-concentration sample 2. In the context of a study designed to establish "how low you can go" (i.e., establishing a detection limit), carry-over of the analyte into a low-concentration sample may compromise the results by inflating the result for low-concentration sample 2, but not inflating the results for other low-concentration samples because the randomized design did not cause them to be analyzed immediately following a high-concentration sample.

Analysts are aware of the potential for carry-over and generally take steps during routine analyses to minimize the chance that it will occur. Examples of steps that can minimize carry-over problems include analyzing "cleaner" samples before "dirtier" samples, and interspersing "blanks" between samples when possible or practical. Obviously, the intentional segregation of low and high concentration samples defeats the purpose of the randomized design. Interspersing blanks between the samples can be effective, as well as blocking similar concentrations together and randomizing blocks. But in order to ensure that the blanks do not have other effects on the results, blanks would be needed between each sample or block analysis, and this would greatly increase the cost of the study (e.g., 25 samples and 24 blanks would be required in case of pure randomization). Although this was done for the Episode 6000 study, this approach would not be practical in most cases. Therefore, despite the statistical benefits, in practice, randomization of the sample analysis sequence can be difficult to apply in detection and quantitation limit studies.

In the Agency's studies of variability as a function of concentration discussed in Sections 1.3.2.1 - 1.3.2.3 of this document, EPA chose to use a non-random design to avoid carry-over problems and to limit the potential difficulties with measurements at very low concentrations. For example, if there was no instrument response at concentration X, then it would be unlikely that there would be a response at a concentration of X/2. In the non-random design, EPA permitted the analyst to stop analyses of ever-lower concentrations, whereas a randomized design would have required that all the samples be analyzed, even when there was no instrumental response for many of those samples.

One of the peer reviewers evaluating EPA's 2003 assessment commented that the effects of carry-over could have been mitigated by studying variability around the calibration line rather than the mean of the replicates. However, carry-over affects subsequent samples differently. The effect of the carry-over cannot be mitigated, regardless of whether variability is studied around the calibration line or the mean of the replicates, unless the amount of carry-over is known and can be subtracted from the affected (low-concentration) sample. This subtraction has limitations because of error accumulation and because the amount of carry-over cannot be determined precisely without extensive studies at multiple concentrations.

Completeness

The physical measurement process can be studied using rough approximations or it can be studied more rigorously. A rough approximation could use the available components of a method as applied to convenient samples. A more rigorous study would use a complete, specific, and well-defined measurement method with all sample processing steps. The appropriate level of study will probably depend on the purpose of the study.

Measurement procedures (methods) may be more or less strictly designed. Variability in what is allowed in the procedures may add to variability in the measurement results. To the extent that permutations of a method's procedures are not expected to be used in a particular detection or quantitation study, EPA recommends that this information be included in the report on the study results. While there may be physical/chemical reasons for extrapolating the results of a variability study on one set of procedures to permutations of those procedures, there is no statistical basis for making such an extrapolation. Statistical theory by itself is only able to describe conditions that have been observed. On the other hand, a knowledge of the underlying physics of the measurement process can guide the completeness of the modeling process when statistical procedures fail. For example, the Rocke and Lorenzato model in the linear or log-log domain may be the best general characterization of a physical measurement process. Therefore, this model can be applied to data to produce a complete answer when statistical procedures fail to deduce the "correct" model.

3.3.7.2 Criteria for the Selection and Appropriate Use of Statistical Models

Detection and quantitation limits may be based on statistical models of the relationship between measurement variation and the concentration of a substance in the sample. Results are produced by adding varying known amounts of the substance to the sample ("spiking"), making replicate measurements at each concentration, and modeling the variability of the results as a function of concentration. This section summarizes the history of modeling variability versus concentration, considers criteria for selecting models, and discusses current practices with regard to available data.

3.3.7.2.1 Short History of Modeling Measurement Results

Over time, a number of different models have been used to estimate measurement variation. Currie (1968) modeled variation in radiochemical measurement methods using a procedure associated with counting large numbers of distinct objects which are appropriately modeled with the Poisson distribution. However, he relied on large sample sizes and standard normal distributions to describe all other types of measurement methods. Hubaux and Vos (1970) developed a procedure based on an estimated calibration relationship that uses smaller sample sizes to estimate Currie's detection and quantitation limits. Again, measurement results were assumed to follow standard normal distributions, but it was also assumed that measurement variation was constant throughout the range of interest. Similarly,

Glaser *et al.* (1981) suggested that measurement variation increases linearly with concentration, but they did not provide estimators under this theory because they believed that measurement variation is usually approximately constant in the range of detection. Glaser *et al.* (1981) did suggest that, when appropriate data were available, a linear regression analysis of the relationship over the analytical range be performed. Clayton *et al.* (1987) discussed transforming the measurement results (using logarithms or square root functions). Gibbons *et al.* (1991) suggested that measurement variability may be proportional to concentration. Rocke and Lorenzato (1995) proposed a model motivated by physical characteristics of measurement processes, in which measurement variability is approximately constant at low concentrations, but changes in a continuous mathematical manner to a relationship where variability increases as concentration increases.

Figure 3-2 illustrates the fundamental analytical measurement models in linear and logarithmic domains. The models are applicable to nearly all analytical measurements; we will not deal with the exceptions because they represent a small percentage of cases. As can be seen from the top two graphs, response is a linear function of concentration in both the linear and log domains. The middle two graphs and the bottom two graphs are those most pertinent to the discussion of detection and quantitation.

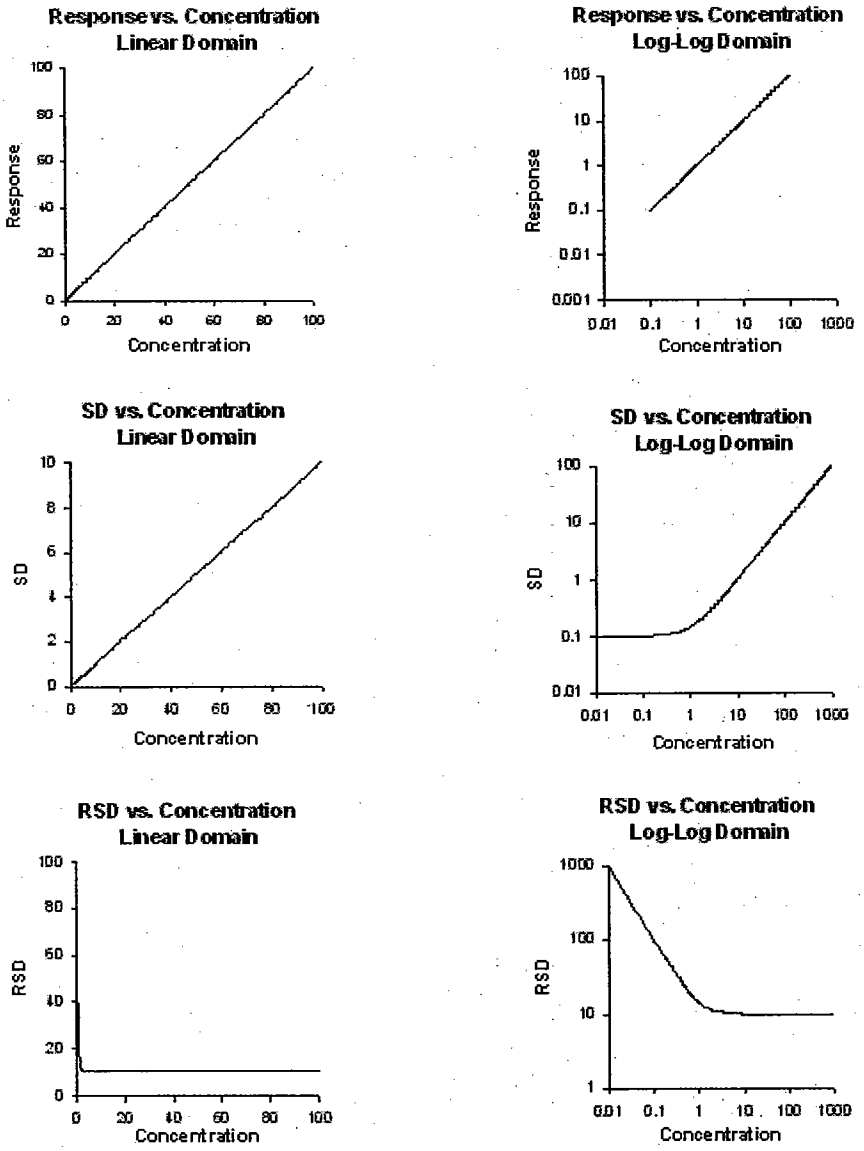

Figure 3-2

3.3.7.2.2 Detection Limits Using Variability at Low Concentrations

The middle two graphs in Figure 3-2 show variability versus concentration and show the model postulated by Rocke and Lorenzato. The flat (constant) portion of the graph in the linear domain is difficult to see because it occurs near the origin, but it can be seen easily in the log domain. Most detection approaches (e.g., Currie's critical value and detection limit; EPA's MDL; the ACS LOD) are constructed assuming that the flat (constant) region of the variability versus concentration relationship holds true, although the graph is rarely displayed (a horizontal line would be singularly uninteresting). Detection approaches such as Currie's critical value, detection limit, LOD, and MDL are constructed by multiplying the standard deviation in the flat region by some constant.

Contention and differences of opinion occur in determining how to arrive at an "appropriate" standard deviation and what to do with the standard deviation when you have it. Currie's critical value and EPA's MDL use a multiple of the standard deviation in a similar manner (a *t*-statistic adjusted for the number of replicates used for Currie's critical value; 3.14 for 7 replicates in EPA's MDL). The IDE uses an additional upward adjustment based on a statistical tolerance limit calculation.

3.3.7.2.3 Quantitation Limits Using Standard Deviation Multiples and Models of Standard Deviation versus Concentration and RSD versus Concentration

Both the limit of quantitation (LOQ) advanced by Currie and the American Chemical Society's Committee on Environmental Improvement and EPA's minimum level of quantitation (ML) result from multiplication of the standard deviation by a factor of 10, again assuming a flat portion of the variability versus concentration graph. This factor of 10 is directed at achieving a relative standard deviation (RSD) of 10 percent. An advantage of this approach is that a quantitation limit is produced, regardless of what the RSD turns out to be.

For example, it is known that the determination of 2,4-dinitrophenol by EPA Method 625 produces highly variable results and that 10 percent RSD cannot be achieved at any concentration level for this compound. Multiplying the standard deviation of replicate measurements of low-level samples results in a quantitation limit that is considerably higher than the quantitation limits for other compounds measured by Method 625. The RSD at this quantitation limit could be 30, 50, or 70 percent. Limiting the RSD associated with the quantitation limit to some arbitrary value (e.g., 30%, as with the ASTM IQE) could prohibit the use of EPA Method 625 for determination of 2,4-dinitrophenol. If 2,4-dinitrophenol were present at a high concentration in a discharge, it would not be reported. Although it could be argued that a more precise method should be used for determination of 2,4-dinitrophenol, determination of pollutants by a large suite of different methods would be quite costly with little meaningful benefit. Increasing precision (i.e., decreasing measurement error) would be critical only if the concentration at issue was near enough to a compliance limit that measurement error could influence the compliance determination. On the other hand, having widely varying RSDs for different analytes within the same method may be confusing to permitting and enforcement authorities who may not appreciate the subtleties of reporting violations in light of the underlying RSDs.

Another means of arriving at a limiting RSD is to graph RSD versus concentration, as shown in the bottom two graphs of Figure 3-2. This approach is used by the ASTM IQE. It has the advantage that a model is fit to data, rather than using a point estimate such as the Currie and ACS LOD or the EPA ML. However, this approach requires considerably more data than are necessary for approaches based on point estimates. In addition, how a model is selected can play a major role in the outcome.

3.3.7.2.4 Criteria for Selecting Models

Both statistical and graphical procedures have been proposed for selecting between models for predicting measurement results based on spike concentrations.

Statistical Criteria

While statistical criteria are available for choosing between models of similar types, the currently available criteria are not satisfactory for choosing between the wide variety of models considered for the relationship between measurement variation and spike concentration, based on EPA's studies. More technically, statistical criteria include using: (1) the simplest model to obtain statistical significance, (2)

the model with the smallest estimated variability, and (3) the model with the smallest likelihood ratio. Given the wide variety of models considered for detection and quantitation, there are problems associated with each of these procedures. Data that obviously do not follow the model may produce statistically significant results, variability may be estimated with weights that make the various estimates incomparable, and the likelihood function may not be comparable between models.

Graphical Criteria

Graphical criteria may be susceptible to some subjectivity in their application, but they are currently the best available method for choosing between models. At the most basic level, the primary graphical criteria is for the form of the model to be suggested by the available data. To consider the quality of the graphical analysis, it is useful to see if some small number of data are overly influential in determining if a model does or does not fit. Given the ability of the human eye to discern deviations from a straight line rather than from a curved line, a useful technique is to plot the data so that they will indicate a straight line if they follow the model of interest.

Both graphical and statistical criteria will be strongly affected by the number and choice of spike concentrations used to fit the different models. Too few spike concentrations will lessen the statistical power of significance tests for slope and curvature from which decisions on the type of model will be made. In addition, the amount of subjectivity with which decisions are made using graphs increases when fewer concentration levels are used. For example, the judgement of whether a residual plot depicts "random scatter" is essentially impossible when only five concentration levels are used (i.e., the residual plot will include only five points). The number of results from which standard deviations are calculated will also have an effect on how models are selected. This set of results may include analysis of multiple replicates at a single laboratory or analysis of one or more replicates from multiple laboratories. If data are obtained from too few laboratories or replicates, the standard deviation estimates will be less reliable, which could lead to incorrect model selection based on statistical or graphical criteria.

3.3.7.2.5 Assessment of Current Models

EPA plotted variability versus concentration data to evaluate the extent to which real data from measurement methods used under the Clean Water Act would conform to a number of different models. For details of how data sets were selected and how data were collected within the data sets, see Appendix B, *Characterizing Measurement Variability as a Function of Analyte Concentration for a Variety of Analytical Techniques*, of the February 2003 Technical Support Document (EPA-821-R-03-005, February 2003). Four sets of composite scatter plots for all combinations of analytical technique, analyte, and study were produced. These sets include:

1. Measurement versus Spike Concentration,
2. Log Measurement versus Log Spike Concentration,
3. Observed Standard Deviation versus Spike Concentration,
4. Log Standard Deviation versus Log Spike Concentration, and
5. Relative Standard Deviation (RSD) versus Log Spike Concentration.

There are hundreds of scatter plots in each set, sorted by the source, measurement technique, and study. The first set of scatter plots can be used to evaluate how well measurement results match the spiked concentration. If the assumed straight line model is true, then the relationship outlined by the plotted data will be approximately linear. These relationships are plotted using log-log plots so that small deviations from the straight line can be visualized easily. All the graphs are contained in attachments to

Appendix B of the Technical Support Document (EPA-821-R-03-005, February 2003).

The plot of observed standard deviations versus spike concentrations can be used to evaluate the reasonableness of the constant variation and/or linearly increasing variability models (Currie, 1968, Hubaux and Vos, 1970, and Glaser *et al.*, 1981). If the constant model for standard deviation is true, there would be no apparent relationship between the standard deviation and spike concentration. If the straight-line model for standard deviation is true, plots are expected to indicate an approximately linear relationship. Analogously, the standard deviation/spike concentration versus spike concentration is expected to show a straight-line relationship when variability is proportional to the spike concentration (Gibbons *et al.*, 1991). The log-log plots of standard deviation versus spike concentration are expected to indicate if log or square root transformations may be appropriate (Clayton *et al.*, 1987) or to display a shape that approximates a "hockey stick" when it is appropriate to use the model proposed by Rocke and Lorenzato (1995). With the Rocke and Lorenzato model, variability near zero will be approximately constant, but will increase proportionally with concentrations in the higher concentration range.

Because the large number of resulting plots makes it difficult to draw general conclusions, for the most part, conclusions must be considered on a case-by-case basis.

3.3.7.3 Methodology for Parameter Estimation

Along with various approaches of detection and quantitation and models for measurement, a number of specific procedures have been suggested for estimating model parameters. Maximum likelihood and least squares are two generally applicable statistical methods that can be used in estimating model parameters. There are advantages and disadvantages to both that must be weighed in particular cases. A standard statistical practice for evaluating the quality of an estimation procedure is to calculate the precision and bias, usually best understood by examining a plot of residuals from a fit to a function. All else being equal, the estimation procedure resulting in the greatest precision and least bias is preferred. In some cases, precision and bias can be calculated based on the assumptions behind the estimation procedure. In other cases, it is either necessary or convenient to estimate precision and bias using simulations. From a general theoretical perspective, the maximum likelihood estimation methodology is preferable because it generates estimates that are generally best with regard to properties of precision and bias (especially for larger sample sizes), while also being approximately normally distributed. Unfortunately, maximum likelihood methodology sometimes can be problematic because the method requires the solution of complex equations. Least squares estimation is generally more tractable, and thus is more generally applicable, although the estimates that result may not be as desirable from a theoretical statistical perspective.

What can sometimes be overlooked in considering estimation and model fitting is that direct measurement of variation of the blank or low level concentration may be the most cost-effective and least difficult method to implement especially where variability does not change much over the region of interest.