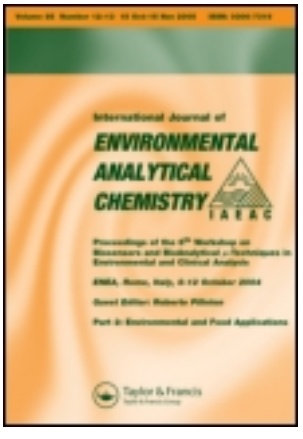

International Journal of Environmental Analytical Chemistry
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/geac20

An inter-laboratory study of detection limits in the analysis of water and wastewater for organo-chlorine pesticides by liquid/liquid extraction and gas chromatography-electron capture detector (USEPA Method 608)

David Eugene Kimbrough [a]
[a] Castaic Lake Water Agency, 27234 Bouquet Canyon Road, Santa Clarita, CA, 91350-2173, USA

Available online: 11 Jul 2011

To cite this article: David Eugene Kimbrough (2011): An inter-laboratory study of detection limits in the analysis of water and wastewater for organo-chlorine pesticides by liquid/liquid extraction and gas chromatography-electron capture detector (USEPA Method 608), International Journal of Environmental Analytical Chemistry, 91:10, 960-977

To link to this article: http://dx.doi.org/10.1080/03067310903359542

Intern. J. Environ. Anal. Chem.
Vol. 91, No. 10, 20 August 2011, 960–977

Taylor & Francis
Taylor & Francis Group

An inter-laboratory study of detection limits in the analysis of water and wastewater for organo-chlorine pesticides by liquid/liquid extraction and gas chromatography-electron capture detector (USEPA Method 608)

David Eugene Kimbrough*

Castaic Lake Water Agency, 27234 Bouquet Canyon Road, Santa Clarita, CA, 91350-2173, USA

(*Received 9 June 2009; final version received 22 September 2009*)

An inter-laboratory study was conducted to assess the Kaiser-Currie Model (KCM) for the determination of detection limits. Six laboratories participated in the analysis of samples prepared from distilled water, some containing organo-chlorine pesticides at a concentration of zero and other with a greater than zero concentration. The study consisted of three phases, the first of which was a study to assess the longer term variability of distilled water samples containing no organo-chlorine pesticides prepared by the participating laboratory analysed over a six month period. A second phase consisted of replicates of distilled water samples containing organo-chlorine pesticides prepared at a single concentration greater than zero by the laboratory and were analysed over several days. Finally, a third phase consisted of twelve distilled water samples, eleven containing organo-chlorine pesticides at a concentration of greater than zero and one with a concentration of zero prepared by a third party. Estimated detection limits were determined and then compared to the observed detection limits. Only in a minority of cases, where the distribution of results from samples containing a concentration of zero was normally distributed, did Currie's L_C accurately predict a concentration which corresponded to a 1% false positive rate in distilled water samples with a zero concentration of the study analyte. The USEPA's MDL performed more poorly. In the majority of cases, when any non-zero results were obtained from distilled water samples containing a concentration of zero, they were not normally distributed. Contrary to expectation, false negatives and false positives rarely occurred simultaneously on any given day. The variability between days of analysis and the use of noise reducing techniques proved to be a significant source of the observed non-normal distribution of distilled water samples. Conventional procedures based on the KCM and their underlying analytical and statistical assumptions did not provide useful predictions of laboratory sensitivity in most cases in this study.

Keywords: detection limit; critical level; drinking water; wastewater; regulatory compliance; gas chromatography; electron capture detector; USEPA 608; organo-chlorine pesticides; false positive; false negative

1. Introduction

One of the most contentious issues in analytical chemistry is that of reporting limits [1]. At very low concentrations, samples containing a non-zero concentration (here called an

*Email: dkimbrough@clwa.org

ISSN 0306–7319 print/ISSN 1029–0397 online

DOI: 10.1080/03067310903359542
http://www.informaworld.com

N sample) of analyte can be measured as having a zero (or negative) concentration and reported as being absent from the sample, called a false negative (FN). Conversely, samples containing a zero concentration of analyte (here called a Z sample) can erroneously be assigned a positive numeric result meaning it is present, which is called a false positive (FP). To preclude reporting FPs or FNs, analytical chemists have traditionally used detection limits (DL). A DL is a threshold concentration of a given analyte determined by a particular laboratory-method-analytic-combination (LMAC) below which results have a given probability of being either a FP or FN, typically 1% for FPs and 50% for FNs [2].

If a laboratory were to analyse a Z sample many times by a particular LMAC, it is generally expected to produce a normal distribution of non-zero results around a mean of zero. If a DL corresponds to a 1% FP rate in Z samples, the 99th percentile of these results would be the observed detection limit (ODL). Due to the impracticality of conducting such an exercise on a routine basis, statistical procedures have been developed to estimate the ODL from only a few Z or N samples. While estimating DLs is of interest to all analytical chemists and data users, it is of particular interest for the analysis of water for regulatory compliance in the United States. In some cases, the waters may be judged in or out of compliance with the Clean Water Act (CWA) based on whether results are measured above or below a detection limit. More broadly in other regulatory situations, important public health decisions may be made based upon the determination of a 'detected' analytes, i.e. present at a concentration higher than the DL.

The most widely used statistical procedures to estimate detection limits are the critical level (L_C), detection limit (L_D) [3], limit of detection [4,5], and the method detection limit (MDL) [6,7], although there other approaches that have been proposed [8,9]. While there has been extensive discussion in the literature over the last 30 years about the relative merits of these various statistical procedures on theoretical grounds, there has been a minimal amount of study of on detection limits from an experimental perspective in general [10] or in water in particular.

The purpose of this paper is examine the question of detection limits from an entirely empirical basis, specifically by measuring the ODLs for several LMACs using many Z and N samples and comparing those to the estimated DLs produced by different statistical procedures, i.e. do the estimated DLs come close to the ODLs. While a particular focus of this paper will be on methods and analytes important to regulatory compliance under the CWA and Safe Drinking Water Act (SDWA), the conclusions may be applied to other media and situations.

1.1 *Theory*

The basic theory upon which most published approaches for estimating detection limits is based on the work of Kaiser [11] and Currie [3]. Kaiser proposed that the detection limit ('*nachweisgrenze*') be defined as the concentration at which an analyte can be detected with 99% confidence that it is greater than zero and proposed to use hypothesis testing to find this quantity. Currie took this approach further and developed actual procedures for determining detection limits. The method was to make replicate measurements of Z samples and from these measurements determine the upper single-tailed 99% $(1-\alpha)$ tolerance interval, which Currie called the 'critical level' (L_C),

$$L_C = L_C = K_{1-\alpha}\sigma_0 \sim 1.6\sigma_0, \tag{1}$$

where σ_0 is the standard deviation of the results of the analysis of the Z samples and K is the tolerance factor. The theory is that any value greater than L_C exceeds the upper tolerance limit of results for Z samples and thus has some definite probability of actually being greater than zero, i.e. being a FP. Put a different way, if the null hypothesis is that a sample contains zero concentration of an analyte of interest, it would be rejected for any concentration greater than L_C. The approximation of K as being 1.6 for a 1% FP rate is based on an explicitly stated assumption of a normal distribution of results for the Z samples [3].

If an L_C were determined for a particular LMAC and if this L_C were used as a reporting limit, then if a sample analysed for this LMAC had a result less than the L_C, it would be reported as $<L_C$ and any result greater than the L_C would be assigned a numeric value. If a N sample containing a concentration exactly equal to L_C were analysed for this LMAC, there would be a 50% chance of this N sample being reported $<L_C$ and a 50% chance of being reported above, if the results were symmetrically distributed around L_C. L_C would provide, in theory, a 1% FP rate or less in a Z sample but a 50% FN rate in an N sample containing the analyte of interest at L_C.

To protect against FNs in N samples, Currie proposed a second threshold, the 'detection level' (L_D):

$$L_D = L_C + K_{1-\beta}\sigma_D, \tag{2}$$

(see Figure 1) with a probability of FN of β where σ_D is the standard deviation at concentration D. Currie assumes normal distributions for both N samples and Z samples and suggests that if it is also assumed that the standard deviations for both populations are the equal, then

$$L_D = 2L_C \sim 3.3\sigma_D, \tag{3}$$

This basic approach, with various modifications, has been adopted by various organisations, such as the International Union of Pure and Applied Chemistry (IUPAC) [4] the International Standards Organization (ISO) [12] and the American Chemical Society [5]. Likewise, when the United States Environmental Protection Agency (USEPA)

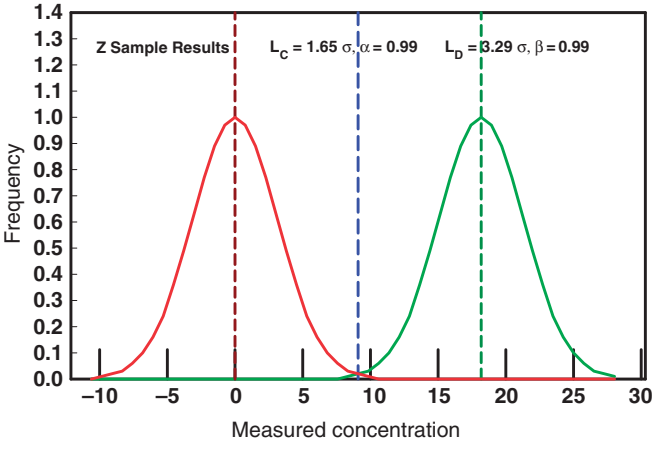

Figure 1. Currie's definition of L_C and L_D (after Currie 1968).

decided to develop a detection limit for both the CWA and the SDWA, the Agency used the KCM [7,12]. However the USEPA made some fundamental changes to the KCM. Their method detection limit (MDL) used multiple measurements of N samples, instead of Z samples as Currie and Kaiser had argued for, because, as the authors of the MDL wrote, 'MDL is considered operationally meaningful only when the method is truly in detection mode, i.e. analyte must be present' [7].

Following Kaiser, the USEPA's MDL used a confidence interval instead of Currie's tolerance interval and assumed a test population of N sample results has an unknown mean (x) and standard deviation (s) so the Student's t value was used instead of z and s instead of σ, so that this confidence interval is $x + t\ s$, where the t value is determined by the confidence level desired and the number of replicates (e.g., for $\alpha = 0.99$, $n = 7$, $t = 3.14$). If the concentration of the given N sample is at the correct value, as determined through an iterative procedure, the interval between the mean value and the zero concentration should be simply $t * s$ (Figure 2). Procedurally, the MDL is opposite of the L_C, although they both seek the same objective. The L_C seeks to determine the distribution of results produced by the analysis of Z samples and look upward in concentration to find the lowest concentration that has a fixed probability of not being a Z sample. One the other hand, the MDL seeks to determine the distribution of results produced by the analysis of N samples and looks downward in concentration to find the highest concentration that has a 99% probability of not being an N sample.

1.2 *Study design*

In 2005 the USEPA established a Federal Advisory Committee (called the Federal Advisory Committee on Detection and Quantitation Approaches and Uses in Clean Water Act Programs, or simply the 'FACDQ') [12] to assist in developing new approaches to the determining the MDL and minimum level of quantitation or ML. One component of that process was a large inter-laboratory pilot study to test the recommended statistical procedures that used the KCM as their basis. Many analytical methods and laboratories

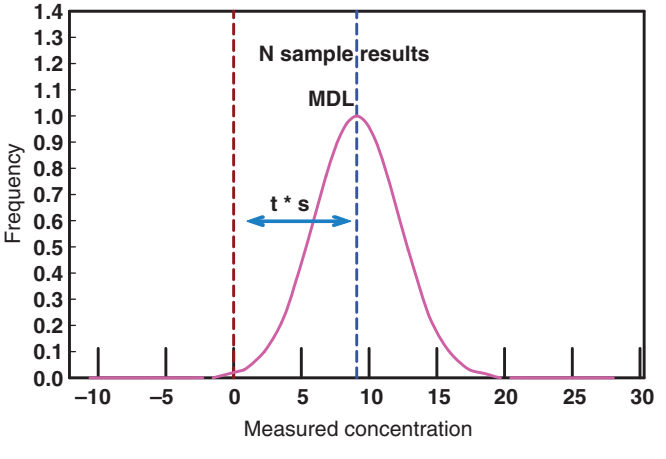

Figure 2. USEPA's definition of MDL (after Glaser *et al.* 1981).

were part of this study but for this paper data from six laboratories which were accredited for organo-chlorine pesticides was used. The data was not collected for the purposes of this paper but the author, a member of the FACDQ, used a portion of the data collect during this study for this paper.

1.2.1 *Study phases*

The portion of the study used in this paper consisted of three phases. The first phase required the participating laboratories to submit all Z sample data collected as part of their routine analytical activities from the six months prior to the beginning of the study (Phase 1). The second phase was a study of variability using at least seven replicate non-blind N samples which was used in this paper for the determination of the USEPA's MDL (Phase 2). Laboratories prepared an N sample at a concentration near where the laboratory anticipated its ability to accurately quantify would be and then analysed it at least seven times on seven separate days. Some laboratories analysed more replicates than seven and some at two different concentrations (for this paper, when this happened, MDL with the lower prepared concentration was used). The third phase had the laboratories analyse 12 blind samples consisting of 11 N samples and 1 Z sample, with different concentrations of analytes over several days (Phase 3).

1.2.2 *Laboratories, test methods and analytes*

The six participating laboratories that used gas chromatography with an electron capture detector (GC-ECD) after a liquid-liquid (aqueous/non-polar organic which could either be performed manually with a separatory funnel or with automated continuous extraction equipment) extraction by USEPA Method 608 [13] are identified here as Labs 29, 31, 32, 34, 35, and 37. Only those LMACs where data was available from all three phases, and had at least seven Z samples from the Phase 1 were used for this paper. There were 106 LMACs which met all of the requirements with a total of 3488 Z samples analysed in Phase 1. Phase 3 N samples were prepared in a matrix of distilled water, preserved, labeled, sealed, and shipped, as a whole volume so that the laboratory did not have to make any dilutions, to the laboratories by a third party company contracted for the FACDQ study. These contained concentrations of the various analytes listed in Table 1 with test resulting in a total of 5304 analyses.

Table 1. Analytes and expected concentrations in samples in the third phase study (all units in ng/L × 1000).

Sample ID	1	2	3	4	5	6	7	8	9	10	11	12
Analyte												
A	0	2	5	10	20	50	75	100	200	500	800	1000
B	0	1	2	5	10	20	50	100	200	500	800	1000

Notes: A = 4,4′-DDD, 4,4′-DDE, 4,4′-DDT, Dieldrin, Endosulfan II, Endosulfan Sulfate, Endrin, Endrin Aldehyde.
B = Aldrin, α-Chlordane, α-BCH, β-BCH, δ-BCH, Endosulfan I, γ-BCH, γ-Chlordane, Heptachlor, Heptachlor Epoxide.

1.2.3 *Assessment*

For each LMAC, the 99th percentile of the Phase 1 Z sample data was used to determine the ODL, as well as calculating the mean, skewness, kurtosis, standard deviation, and L_C. From the Phase 2 data, a standard deviation of the replicate N samples was determined and the MDL was calculated for the same LMACs. Finally, the number of FPs and FNs from the Z sample in Phase 3 were also determined for each LMAC. Then the ratio of the L_C and MDL were compared to the ODL for each LMAC. If the ratio of L_C or MDL to ODL was 1.0, then the estimate was accurate. The further the ratio is from 1.0, the less accurate the estimate was. In this manner, the observed and predicted concentrations where a 1% FP in Z samples would occur can be compared. To assess FNs in N samples, FN rates for each LMAC were determined from the Phase 3 N sample results. N samples that produced results less than or equal to zero were judged as FNs. For this paper, a population of results was considered Gaussian if the skewness and kurtosis were both between the values of 1 and −1.

2. Results and discussion

Tables 2a–r show the mean results from individual mean, standard deviation, count, and the ODL and Currie's L_C for the Z samples for each LMACs from Phase 1 and the MDLs from Phase 2. Table 3 presents a summary of the aggregate measures of accuracy of both MDL and L_C as compared to the ODL. Table 4 summarises the ratios of ODL to L_C and MDL as they correlate with the skewness and kurtosis of the Z samples from Phase 1.

Table 2a. Summary of the observed and estimated detection limits from Phase 1 and Phase 2 for 4,4′-DDD MDL published in method 608 = 11,000 (all units in ng/L).

Lab	Mean	SD	*n*	Kurtosis	Skew	L_C	MDL	ODL
29	43	220	30	25	5.1	351	4800	900
31	1500	5300	57	11	3.5	8600	14,000	24,000
32	2600	8200	30	18	4.3	13,000	3000	35,000
34	−8800	30,000	30	0.7	0.6	48,000	39,000	66,000
35	53	2500	30	12	−2.1	3900	1300	5500
37	0	0	30			0	20,000	0

Table 2b. Summary of the observed and estimated detection limits from Phase 1 and Phase 2 for 4,4′-DDE MDL published in method 608 = 4000 (all units in ng/L).

Lab	Mean	SD	*n*	Kurtosis	Skew	L_C	MDL	ODL
29	0	0	30			0	4700	0
31	450	2400	57	24	5.1	3800	6000	1300
32	−200	930	30	8.9	−2.3	1500	3000	1500
34	430	16,000	30	−0.1	0.5	25,000	19,000	37,000
35	−180	1000	30	25	−5.2	1700	1500	120
37	0	0	30			0	12,000	0

Table 2c. Summary of the observed and estimated detection limits from Phase 1 and Phase 2 for 4,4′-DDT MDL published in method 608 = 12,000 (all units in ng/L).

Lab	Mean	SD	*n*	Kurtosis	Skew	L_C	MDL	ODL
29	4400	24,000	30	25	5.2	3800	4600	9400
31	3600	16,000	57	24	4.9	26,000	8200	81,000
32	2600	8200	30	18	4.3	13,000	2200	35,000
34	−1800	33,000	30	−0.6	0.2	52,000	30,000	66,000
35	850	2600	30	6.6	2.9	4200	1500	9900
37	0	0	30			0	21,000	0

Table 2d. Summary of the observed and estimated detection limits from Phase 1 and Phase 2 for Aldrin MDL published in method 608 = 4000 (all units in ng/L).

Lab	Mean	SD	*n*	Kurtosis	Skew	L_C	MDL	ODL
29	12,000	5600	29	24	5.1	90,000	2100	230,000
31	2000	6300	57	8.8	3.1	10,000	8600	25,000
32	1500	3700	30	1.7	1.1	6000	9400	12,000
34	−12,000	30,000	30	−0.1	0.4	48,000	2500	57,000
35	0	0	30			0	1800	0
37	0	0	30			0	7500	0

Table 2e. Summary of the observed and estimated detection limits from Phase 1 and Phase 2 for α-chlordane MDL published in method 608 = 14,000* (all units in ng/L).

Lab	Mean	SD	*n*	Kurtosis	Skew	L_C	MDL	ODL
29	0	0	30			0	2900	0
31	0	0	57			0	31,000	0
32	−100	1800	30	9.2	1.1	2900	2800	4900
34	−8600	15,000	30	0.1	1.1	24,000	18,000	2800
37	0	0	30			0	22,000	0

Table 2f. Summary of the observed and estimated detection limits from Phase 1 and Phase 2 for α-BHC MDL published in method 608 = 3000 (all units in ng/L).

Lab	Mean	SD	*n*	Kurtosis	Skew	L_C	MDL	ODL
29	100	440	30	21	4.6	700	4900	1900
31	0	0	57			0	7200	0
32	730	3000	30	0.3	0.6	4800	8600	7900
34	4500	7000	30	0.2	0.5	11,000	13,000	22,000
35	0	0	30			0	2000	0
37	0	0	30			0	13,000	0

Table 2g. Summary of the observed and estimated detection limits from Phase 1 and Phase 2 for β-BHC MDL published in method 608 = 6000 (all units in ng/L).

Lab	Mean	SD	n	Kurtosis	Skew	L_C	MDL	ODL
29	50	210	30	19	4.4	340	3300	910
31	0	0	57			0	9100	0
32	−730	4400	30	−0.4	0.5	7100	11,000	8700
34	−3700	17,000	30	2.7	1.4	27,000	24,000	47,000
35	0	0	30			0	2400	0
37	0	0	30			0	11,000	0

Table 2h. Summary of the observed and estimated detection limits from Phase 1 and Phase 2 for δ-BHC MDL published in method 608 = 9000 (all units in ng/L).

Lab	Mean	SD	n	Kurtosis	Skew	L_C	MDL	ODL
29	93	360	30	20	4.6	570	3600	1500
31	0	0	57			0	7300	0
32	2300	4100	30	5.8	2.3	6500	5600	16,000
34	8800	16,000	30	0.2	0.1	25,000	17,000	45,000
35	0	0	30			0	1600	0
37	0	0	30			0	7600	0

Table 2i. Summary of the observed and estimated detection limits from Phase 1 and Phase 2 for dieldrin MDL published in method 608 = 2000 (all units in ng/L).

Lab	Mean	SD	n	Kurtosis	Skew	L_C	MDL	ODL
29	3.3	18	30	25	5.2	29	5500	73
31	200	1500	57	52	7.3	2400	5600	5200
32	670	2400	30	5.7	2.4	3800	4300	8700
34	1700	15,000	30	0.1	0.6	2500	19,000	39,000
35	180	770	30	21	4.6	1200	1500	3300
37	0	0	30			0	14,000	0

Table 2j. Summary of the observed and estimated detection limits from Phase 1 and Phase 2 for Endosulfan I MDL published in method 608 = 14,000 (all units in ng/L).

Lab	Mean	SD	n	Kurtosis	Skew	L_C	MDL	ODL
29	13	57	30	20	4.5	91	2700	250
31	140	1100	57	52	7.3	1700	5600	3700
32	−300	4800	30	9.1	2.6	7600	4000	16,000
34	−2600	16,000	30	−0.2	0.5	26,000	19,000	35,000
35	110	650	30	23	4.9	1000	1600	2700
37	0	0	30			0	6000	0

Table 2k. Summary of the observed and estimated detection limits from Phase 1 and Phase 2 for Endosulfan II MDL published in method 608 = 4000 (all units in ng/L).

Lab	Mean	SD	*n*	Kurtosis	Skew	L_C	MDL	ODL
29	13	73	30	25	5.2	120	4700	290
31	520	3300	57	47	6.9	5300	6600	14,000
32	67	3100	30	4.6	0.9	4900	2900	9500
34	−13,400	32,000	30	0.1	0.2	51,000	37,000	59,000
35	300	1000	30	9.2	3.3	1600	1500	4100
37	0	0	30			0	13,000	0

Table 2l. Summary of the observed and estimated detection limits from Phase 1 and Phase 2 for Endosulfan Sulfate MDL published in method 608 = 66,000 (all units in ng/L).

Lab	Mean	SD	*n*	Kurtosis	Skew	L_C	MDL	ODL
29	20	92	30	23	4.9	150	4900	390
31	160	1200	57	52	7.3	1900	6300	4100
32	4200	7200	30	0.2	1.1	12,000	8700	22,000
34	−7900	34,000	30	0.1	0.3	54,000	42,000	70,000
35	98	390	30	13	3.8	620	1600	1600
37	0	0	30			0	15,000	0

Table 2m. Summary of the observed and estimated detection limits from Phase 1 and Phase 2 for Endrin MDL published in method 608 = 6000 (all units in ng/L).

Lab	Mean	SD	*n*	Kurtosis	Skew	L_C	MDL	ODL
29	730	3700	30	25	5.2	5900	5200	15,000
31	950	4200	57	20	4.5	6800	8600	20,000
32	370	3000	30	6.9	2.5	4800	5500	11,000
34	770	16,000	30	−0.1	0.5	25,000	19,000	37,000
35	150	1100	30	7.7	1.2	1700	2400	3700
37	0	0	30			0	13,000	0

Table 2n. Summary of the observed and estimated detection limits from Phase 1 and Phase 2 for Endrin Aldehyde MDL published in method 608 = 23,000 (all units in ng/L).

Lab	Mean	SD	*n*	Kurtosis	Skew	L_C	MDL	ODL
29	70	210	30	6.2	2.8	330	5600	770
31	900	6400	57	52	7.3	1000	6100	24,000
32	1300	6000	30	10	3.3	9500	4900	24,000
34	−15,300	46,000	30	−0.1	−0.1	74,000	57,000	81,000
35	490	1600	30	9.6	3.2	2500	1500	6300
37	0	0	30			0	23,000	0

Table 2o. Summary of the observed and estimated detection limits from Phase 1 and Phase 2 for γ-BHC MDL published in method 608 = 9000 (all units in ng/L).

Lab	Mean	SD	n	Kurtosis	Skew	L_C	MDL	ODL
29	27	98	30	17	4.1	160	2500	420
31	1100	4100	57	17	4.1	6500	7400	19,000
32	900	2600	30	0.3	0.2	4200	12,000	6700
34	3700	6700	30	0.6	0.5	11,000	12,000	20,000
35	−1.3	7.3	30	25	−5.2	12	2000	0
37	0	0	30			0	7500	0

Table 2p. Summary of the observed and estimated detection limits from Phase 1 and Phase 2 for γ-Chlordane MDL published in method 608 = 14,000* (all units in ng/L).

Lab	Mean	SD	n	Kurtosis	Skew	L_C	MDL	ODL
29	340	1100	30	8.4	3.2	1800	2600	4400
31	2500	7600	57	26	4.8	12,000	5700	32,000
32	1200	5800	30	16	4.1	9300	9900	25,000
34	33	18,000	30	0.2	0.7	28,800	22,000	45,000
37	0	0	30			0	11,000	0

Table 2q. Summary of the observed and estimated detection limits from Phase 1 and Phase 2 for Heptachlor MDL published in method 608 = 3000 (all units in ng/L).

Lab	Mean	SD	n	Kurtosis	Skew	L_C	MDL	ODL
29	170	470	30	16	3.9	750	2700	2000
31	1600	3900	57	4.5	2.3	6200	8800	16,000
32	29,000	33,000	30	−0.4	1.0	52,000	8200	99,000
34	3900	16,000	30	−0.5	0.3	25,000	17,000	38,000
35	170	760	30	22	4.8	1200	1800	3200
37	0	0	30			0	6700	0

Table 2r. Summary of the observed and estimated detection limits from Phase 1 and Phase 2 for Heptachlor Epoxide MDL published in method 608 = 83,000 (all units in ng/L).

Lab	Mean	SD	n	Kurtosis	Skew	L_C	MDL	ODL
29	320	1500	30	24	5.0	2400	2500	6400
31	190	1400	57	52	7.3	2200	7700	4900
32	−1500	2200	30	−0.6	−0.9	3500	4300	730
34	−3900	15,000	30	0.1	0.6	24,000	18,000	31,000
35	−30	160	30	25	−5.2	260	1500	0
37	0	0	30			0	8800	0

Note: *USEPA method 608 has a single MDL for 'chlordane' and does not distinguish α from γ.

Table 3. Aggregate measures of accuracy L_C and MDL to the ODL for LMACs with ODL > 0.

Ratio	ODL/MDL	ODL/L_C	N
Mean	2.1	4.1	78
Median	2.3	1.8	
Minimum	0.07	0.01	
Maximum	3.3	110	
Group 1	1.5	0.8	22
Group 3	2.5	4.1	56

Table 4. Number of individual LMACs with ratios of ODL to either L_C or MDL within different limits.

Ratio	ODL/MDL	ODL/L_C
<0.50	43	4
0.50 > < 1.50	**11**	**13**
1.50 >	52	61
<0.75	48	4
0.75 > < 1.25	**5**	**6**
1.25 >	53	68
<0.90	49	4
0.90 > < 1.10	**3**	**2**
1.10 >	54	72

Table 5 has the number of FPs among the Phase 3 Z samples for each LMAC. Table 6 shows the number of FNs among the Phase 3 Z samples for each LMAC.

Of the 3665 Z sample results analysed by all laboratories as part of Phase 1367 were less than zero (10%), 528 were greater than zero (14%), and 2773 (76%) were equal to zero. For 22 LMACs, the Z sample results that were non-zero in value were distributed in a Gaussian fashion (Group 1). Another 28 LMACs that produced only zero values from the analysis of the Z samples (Group 2). The third group, which consisted of 56 LMACs, also produced non-zero values for the Z samples but had a non-Gaussian distribution, either kurtotic or skewed or both (Group 3) because the majority of the results were zero.

2.1 *Group 1*

When the results were normally distributed, the average values for L_C and MDL made relatively accurate estimates (i.e. +/−50%) as compared to the ODL. As can be seen in Table 4, the average ratio of ODL/L_C for this group was 1.5 but the average ratio for ODL/MDL was 0.8, which meant the MDL was usually lower than the actual 99th percentile of Z samples and L_C was generally higher. However, while the average estimate was not too far off, the actual number of estimates that fell within the acceptance limits for this paper was quite small, for the widest acceptance range, +50%, only 11% for the MDL and 12% for the L_C. If narrower acceptance limits are used the percentage of accurate

Table 5. Numbers of false positives among Phase 3 Z samples at different laboratories.

Analyte	Laboratory	29	31	32	34	35	37
4,4′-DDD		3	0	0	9	0	0
4,4′-DDE		1	1	0	9	0	0
4,4′-DDT		0	0	0	9	0	0
Aldrin		4	1	0	8	0	0
α-Chlordane	0	0	0	9	NA	0	
α-BHC	7	0	0	5	0	0	
β-BHC	4	0	0	6	0	0	
δ-BHC	4	0	0	6	0	0	
Dieldrin		0	0	0	10	0	0
Endosulfan I		2	0	0	9	0	0
Endosulfan II		0	0	0	9	0	0
Endosulfan Sulfate		1	0	0	8	0	0
Endrin		0	0	0	7	0	0
Endrin Aldehyde		0	0	0	9	0	0
γ-BHC	6	0	0	10	0	0	
γ-Chlordane	3	0	0	3	NA	0	
Heptachlor		3	0	2	7	0	1
Heptachlor Epoxide		0	0	0	3	0	0
Total false positives		38	2	2	136	0	1
Total determinations		190	190	190	190	170	190
Percent false positives		20	1	1	72	0	<1

Table 6. Numbers of false negatives among Phase 3 N samples at different laboratories.

Analyte	Laboratory	29	31	32	34	35	37
4,4′-DDD		4	3	11	0	0	0
4,4′-DDE		7	0	17	0	0	0
4,4′-DDT		4	2	11	0	0	0
Aldrin		2	2	19	0	0	0
α-Chlordane	5	1	31	1	NA	1	
α-BHC	0	2	30	0	0	0	
β-BHC	1	10	31	7	0	3	
δ-BHC	0	4	19	0	0	0	
Dieldrin		5	1	16	0	1	0
Endosulfan I		16	9	46	0	1	3
Endosulfan II		19	10	46	4	3	4
Endosulfan Sulfate		5	0	7	0	0	0
Endrin		5	0	19	0	1	0
Endrin Aldehyde		15	1	16	0	2	0
γ-BHC	0	2	26	0	0	0	
γ-Chlordane	7	0	30	12	NA	0	
Heptachlor		4	0	15	1	0	0
Heptachlor Epoxide		6	0	29	0	0	0
Total false negatives		105	47	419	25	8	11
Total determinations		2090	2090	2090	2090	1870	2090
Percent false negatives		5	2	20	1	<1	1

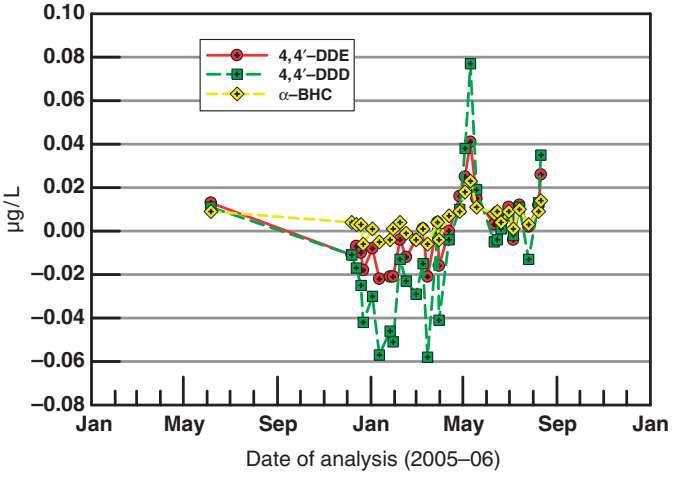

Figure 3. Phase 1 result over 14 months from Lab 34 for three pesticides.

results decline. It would appear that there were plenty of extremely inaccurate estimates of MDL and L_C which balanced each other out when they were averaged together, as the ranges of results in Table 4 might suggest.

Moreover, all of the Group 1 LMACs came from only two of the six labs, Lab 32 and Lab 34. In fact, almost all of the Group 1 LMACs were produced by Lab 34, 15 in all, while only seven came from Lab 32 and none from the others. Of the 12 LMACs in Group 1 where the estimated L_C was within 50% of the ODL, 11 were produced by Labs 32 and 35 (the results from Lab 35's analysis of 4,4′-DDD produced an ODL of 5500 and an L_C of 3900 ng/L) and had non-kurtotic and non-skewed results. This would suggest that when the results the analysis of Z samples are normally distributed, there is a much greater likelihood that the L_C will be within +/−50% of the ODL. In contrast, for the 11 LMACs where the MDL was within 50% of the ODL, only four had Z sample results that were normally distributed and only one was produced by Lab 34 (four were produced by Lab 32). Interestingly, this one LMAC from Lab 34 was the only one that produced an L_C and an MDL that were both within 50% of the ODL. It would appear that normally distributed Z sample results were not the norm and that they are associated with accurate estimates of the ODL but not always. If the acceptance criteria for accuracy of the MDL and L_C are narrowed, even fewer results were acceptable with fewer than 3% of either the MDL or L_C estimates within 10% of the ODL.

However this might actually be a somewhat misleading interpretation of the results. Consider the results from Lab 34's Phase 1 Z samples as shown in Figure 3. While the distribution of Z samples over a 14 month period did indeed produce 30 results that were neither kurtotic nor skewed, an examination of the individual results over time revealed a somewhat different story. The temporal trends of the Z sample results for 4,4′-DDE through 2006 show a definite pattern of increasing measured values over the course of 2006. All of the results before April 13 are less than zero and all but one of the results after this date were greater than zero. Plotting two other "normally" distributed analytes from Lab 34 showed similar albeit not identical patterns. Figure 4 shows three other normally distributed Z samples for three analytes analysed at Lab 32. In this case, the general trend

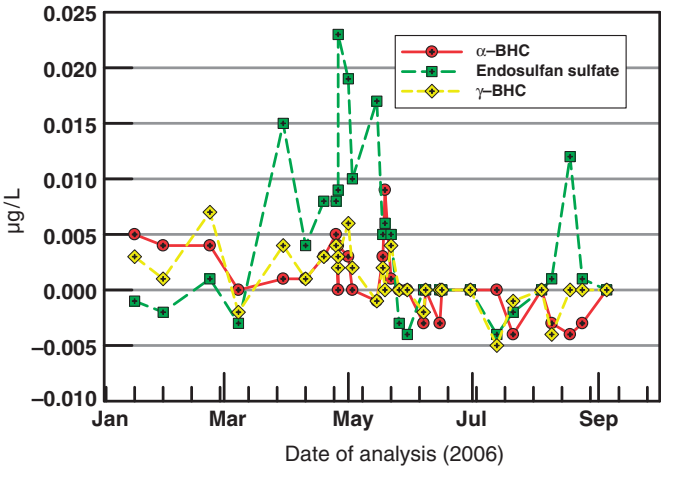

Figure 4. Phase 1 results over nine months from Lab 32 for three pesticides.

is to lower measured values. What is important about this observation is that if the L_C were calculated using all 30 Z sample data points, the calculated L_C is fairly accurate for that population. However, taking the results from 4,4′-DDE from Lab 34, if the L_C had been calculated using the results from before April 13, the L_C would not have been accurate for the results after 13 April, and vice versa. Indeed, had the L_C; been determine on a single day, as is common practice, it is not clear it would accurately estimate the ODL even a week later.

Perhaps of greater importance, Labs 32 and 34 had the poorest performance from a data quality perspective. Lab 32 had the highest number of FNs (419 or 20% of the N samples in Phase 3, more than all of the other laboratories combined) while Lab 34 had the highest number of FP (136 or 90% of the Z samples in Phase 3, likewise more than all of the other laboratories combined). Additionally, these two laboratories produced very significantly different estimates of the MDL and L_C; a difference of a factor of five to ten was typical. Ironically, the ODLs for these two laboratories tended to have the highest values for any given analyte as compared to the other laboratories.

2.2 *Group 2*

This group of the 27 LMACs had an ODL of zero, i.e. over the course of six months with not less than seven Z samples analysed (the mean number of Z samples in Phase 1 was 30), not a single non-zero value was determined. No L_C could be determined as the standard deviation of the Z samples (as well as the mean) was zero in all cases. While the MDLs were determined, the accuracy of the MDL could not be assessed against either the ODL or L_C except that it was very much larger, i.e. it was a positive non-zero value. As with Group 1, the LMACs in Group 2 are not randomly distributed among the laboratories with all 17 of the LMACs from Lab 37 having an ODL of zero, as did six LMACs from Lab 35, four LMACs from Lab 31, and two LMACs from Lab 29.

Interestingly, as a group, the calculated values of L_C and MDL were very inaccurate, i.e. they were not very close to the ODL, except when the L_C and ODL were both zero.

This was because Lab 35 and Lab 37 both produced among the lowest ODLs, including the lowest non-zero ODLs, but tended to have MDLs that were among those with highest values. Further, Lab 35 and Lab 37 had perhaps the best overall performance. Between the two of them, they had a combined 1 FP and 19 FNs which were, in both cases, less than the total number of FPs and FNs from any other individual laboratory. As with Lab 32 and Lab 34, Lab 35 and Lab 37, despite the similarity in the quality of their results, they produced very different L_Cs, MDLs, and ODLs for each given analyte.

2.3 *Group 3*

There were 58 LMACs which had an ODL that was greater than zero but the distribution of Z sample results was either kurtotic or skewed or both. In this case, both the MDL and the L_C produced very poor estimates of the ODL. The average ratio of L_C to ODL was 2.4 for this group while the average ratio of MDL to ODL was about 4. Both procedures produced estimated DLs that were much too low and ratios of ODL to L_C and MDL which were thus too high. This is no doubt a reflection of the non-normal distribution of the Z sample results. A more careful examination of the Phase 1 Z sample results reveals that the results were often bi-modally or poly-modally distributed. In some cases, such as the analysis of Heptachlor Epoxide at Lab 31, of the 57 Z samples analysed between January and September 2006, only one had a non-zero value. At the same laboratory analysing 4,4′-DDD, only 5 of 57 Z samples had a non-zero value, all positive. Indeed, of the 1098 Z samples analysed by Lab 31 during Phase 1, only 53 were non-zero and none were negative. Likewise, Lab 29, which analysed 599 Z samples in Phase 1, only produced 57 positive results and no negative results.

Further, 14 of the 17 LMACs analysed by Lab 31 were in Group 3 as were 16 of the 17 LMACs analysed by Lab 29. None of the LMACs analysed by Lab 37 and only one from Lab 34 were in Group 3. All of the negative results reported from Phase 1 Z sample analysis came from Labs 32, 34, and 35, mostly from the former two. The results from this group of LMACs suggests that these LMACs are producing two or more distinct populations of results, a zero result population and a positive result population for Labs 29 and 31, for example.

2.4 *Non-co-occurrence of false positives and negatives*

Using the Phase 3 results, the number of FP and FN were determined from the Z sample and N sample data as described above and the results summarised in Tables 5 and 6. The data from this paper might suggest that there is a general pattern that FP and FNs do not occur randomly but are rather associated with particular laboratories and/or methods, in particular those methods that tend to produce an ODL equal to zero produce few, if any, FPs and the majority of FNs while laboratories and/or methods that do not produce an ODL equal to zero tend to have the opposite relationship. This non-co-occurrence of FPs and FNs in the same LMAC is even heightened when it is noted that three quarters of those FPs produced were from a single laboratory and the four other laboratories performing Method 608 had only five FPs between them. The FP and FN rates seem to be linked to the question of bias, LMACs with positive results among the Z samples had a positive bias among the low concentration N samples and thus did not produce FNs. Conversely, LMACs with no non-zero values for the Z samples, and no FPs, had negative

biases among the N samples and thus produced FNs. Since an LMAC cannot simultaneously have a positive and negative bias, it is not likely to produce FNs and FPs simultaneously.

The results from Lab 29 also followed this pattern. Of the 105 FNs, almost half were from three analytes, Endosulfan I, Endosulfan II, and Endrin Aldehyde, but these same three analytes had only two FPs, both for Endosulfan I. Conversely, over half of the FPs Lab 29 produced were from four analytes, Alpha-, Beta-, Delta-, and Gamma-BHC which had all but one FP among them. Of the seven analytes that had both FPs and FNs, only two had both on a single day (beta-BHC, 10/23/2006 and Endosulfan Sulfate 10/7/2007).

Detection limits are often used to compare LMACs, such as laboratory to laboratory or method to method. Certainly the results of this study show that there are enormous differences between laboratories in terms of MDL, L_C, and even ODL. While these values were often in the same order of magnitude as the published MDL in Method 608, they generally covered two orders of magnitude. While some laboratories tended to produce lower MDLs, L_Cs, and ODLs than other laboratories, it was not clear that this actually corresponded to any meaningful measure of data quality. Examining the result on Tables 2j, 5 and 6, it can be seen that Labs 29 and 37 have significantly lower L_Cs and ODLs than the other laboratories and Lab 29 has the second lowest MDL. However, Lab 29 has 2 FPs and 16 FNs for Endosulfan I while Lab 37 has 0 FPs and only 3 FNs. So despite them have similar measures of DL, their performance was quite different. Conversely, Labs 34 and 35 have very different MDLs, L_Cs and ODLs (all three were about ten times higher in Lab 34 as compared to Lab 35) but nearly identical data quality measures, neither had any FPs and had 1 FN and 3 FNs respectively. Not only is there an enormous amount of inter-laboratory variability, both compared to the published values and compared to each other but there was little evidence that there was any correspondence between the value of the detection limit and laboratory performance. While this paper was limited to the analysis of organo-chlorine pesticides in water, the FACDQ study actually examined other LMACs such as inductively coupled plasma-atomic emission spectroscopy and ion chromatography and found similar results.

3. Conclusions

The L_C procedure for the estimation of detection limits only worked effectively, i.e. predicted concentration where a Z sample will produce a 1% FP rate within +/−50% of the ODL, when the distribution of Z sample results was normally distributed around a mean value close to zero. Relatively few LMACs studied had Z sample data that was normally distributed so the statistical procedures to estimate detection limits did not correspond to a 1% FP rate in Z samples. This is due to the fact that most Z samples analysed by laboratories in this paper produced results that were either all zeroes (the largest number), were non-normally distributed (either skewed or kurtotic or both), and/or had significant bias either positive or negative. Other alternatives to the L_C and MDL that assume normal distributions of Z sample results would probably suffer the same problem. The USEPA MDL did not fare as well as Currie's L_C, generally producing less accurate estimates than the L_C. Replicate analysis of N samples produce a distribution of results that are quite different from that of Z samples so using standard deviation produced by N samples to predict the future behaviour of Z samples does not seem supported by the results of this paper.

One major contributing factor to the non-normal distribution of the Z and N samples is the fact that sensitivity and overall performance of typical analytical instruments can change dramatically from day to day. The practice of determining a detection limit for a particular LMAC once (e.g. USEPA Method 608), as many older USEPA Methods require, or even annually, as a few newer USEPA Methods now require, does not ensure meaningful results. An MDL or L_C determined on one day does not appear, based on the results of this paper, to tell one anything useful about performance the next day.

The general assumption that FNs in N samples and FPs in Z samples will both occur for a given LMAC is not supported by the results of this paper. LMACs that produced no non-zero values for Z samples produced most, if not all of the false negatives in N samples in this paper. Likewise, LMACs that do produce non-zero values for Z samples do not produce very many, if any, FPs in N samples. Many analytical methods, e.g. chromatographic methods, have instrumental features to suppress signal noise, producing results of zero for Z samples and FNs among N samples. This may be a signal dampener, detector gain or peak evaluation software feature designed to reduce noise and false positives. This tends to create a negative bias that produces false negative results in low concentration samples N samples and no non-zero values in Z samples. Conversely, in methods or instruments where no such feature is used, there can be either a negative or positive bias in the measurements, often producing positive results among Z samples and no FNs among N samples. Generally, on a given day, a particular analytical procedure will either produce FPs in Z samples and no FN in N samples or vice versa, but not both FNs in Z samples and FN in N samples simultaneously. The two tiered approach of L_C for protection against FPs in Z samples and L_D for protection against FNs in N samples is probably unnecessary as both FPs and FNs are not likely to occur in the same LMAC, at least on given day.

Further, there was no correlation between either the estimated DL (L_C and MDL) or the ODL of any LMAC and the actual quality of results produced by that LMAC. An LMAC with a very low ODL may have a very high rate of FPs or FNs while another LMAC might also have a very low ODL but a very low rate of FPs or FNs or both. This would seem to imply that even if a statistical procedure were actually able to accurately estimate the ODL for a given LMAC, it is not clear that this provides any useful measure of performance.

Based on the results of this study, it does not appear that the KCM provides a realistic estimate of ODL for most LMACs. From a policy perspective, this study would also suggest that even if a realistic ODL could be calculated, i.e. a concentration corresponding to a 1% FP in Z samples for a given LMAC, this value does not appear to be a useful measure of laboratory performance.

Acknowledgements

The design of the FACDQ inter-laboratory study was conducted by a sub-committee of the FACDQ which included, Zonetta English, Louisville and Jefferson Counties Metropolitan Sewer District, Dr Richard Rediske, Annis Water Resources Institute, Dr Richard Burrows, Test America, Larry LaFleur, National Council for Air and Stream Improvement, Inc. Dick Reding, USEPA Office of Water, Engineering and Analysis Division, Bob Avery, Michigan Department of Environmental Quality, Environmental Laboratory Services. The author would like to offer special thanks to Ken Miller of Computer Science Corporation who was a technical consultant to the USEPA ODW and provided technical support to the FACDQ.

References

[1] P.S. Porter, R.S. Ward, and H.F. Bell, Environ. Sci. Technol. **22**, 856 (1988).
[2] P.F. Sanders, R.L. Lippincott, and A. Eaton, J. Am. Water Works Assoc. 104 (1996).
[3] L.A. Currie, Anal. Chem. **40**, 586 (1968).
[4] G.L. Long and J.D. Winefordner, Anal. Chem. **55**, 712 (1983).
[5] L.H. Keith and D.L. Lewis, *Proceedings of 203rd Meeting of the American Chemical Society, Committee on Environmental Improvement* (San Francisco, June 1992).
[6] Federal Register, Appendix B to Part 136 CFR 40, 26 October 1984, Vol. 49, No. 209, pp. 198–204.
[7] J.A. Glaser, D.L. Forest, G.D. McKee, S.A. Quave, and W.L. Budde, Environ. Sci. Technol. **15**, 1426 (1981).
[8] R.D. Gibbons, W. Taylor, F.H. Jarke, and K.P. Stoub, *Proceedings of Fifth Annual USEPA Symposium on Waste Testing & Quality Assurance* (USPO, Washington, DC, July 1989).
[9] A. Hubaux and G. Vos, Anal. Chem. **42**, 849 (1970).
[10] D.E. Kimbrough and J.R. Wakakuwa, Environ. Sci. Technol. **27**, 2692 (1993).
[11] H. Kaiser, Z. Anal. Chem. **5**, 1 (1965).
[12] USEPA, Technical Support Document for the Assessment of Detection and Quantitation Approaches, EPA-821-R-03-005 (Washington, DC, February 2003).
[13] CFR Part 136 Appendix A, *Methods for Organic Chemical Analysis of Municipal and Industrial Wastewater Method 608 – Organochlorine Pesticides and PCBs* (1984).