Proceedings of the 13th Annual British Columbia Mine Reclamation Symposium in Vernon, BC, 1989.
The Technical and Research Committee on Reclamation

ENVIRONMENTAL TESTING: WHAT DO THE RESULTS MEAN?

Allan W. Maynard, MSc.

Senior Partner

ASL Analytical Service Laboratories Ltd.

Presented at the 1989 British Columbia Mine Reclamation Symposium, June 8-9, 1989 Vernon, B.C.

1.0 INTRODUCTION

During the I960's and 1970's the contamination of water, soil, sediment and biota became increasingly well documented and reported. This coupled with a number of environmental accidents have resulted in a significant growth in media coverage about toxic chemicals.

Since about 1975, environmental legislation has been drafted and enacted to deal with specific toxic chemicals and central to these legislative documents have been specific contaminants lists. The earliest comprehensive list was that published by the Environmental Protection Agency and is now referred to as the "Priority Pollutants List". Other agencies such as the International Agency for Research on Cancer (IARC) and the United Nations have also published contaminants lists reflecting the

analytical service laboratories ltd.
CONSULTING CHEMISTS & ANALYSTS
1650 Pandora Street
Vancouver, B.C. • V5L 1L6
Fax (604) 253-6700 • Tel. (604) 253-4188

international concern over toxic chemicals. Recently in Canada/ the Ontario Provincial Government produced a list of 180 substances under its MISA (Municipal Industrial Source Abatement) program and Federally under the CEPA (Canadian Environmental Protection Act), a list of chemicals of national concern will be drafted (Environment Canada, 1988).

The preparation of such contaminants lists is indeed challenging given that over 800,000 known chemicals have been recorded with over 60,000 of these being presently used in commerce around the world. Of these, about 1500 are of possible concern (1).

As environmental regulatory agencies have built the inventory *of* "toxics" lists and continue to enhance the comprehensiveness of the legislation, the demand for high quality analytical laboratory capabilities has grown significantly. There is a dramatic trend towards multi-component determinations at very low concentrations in complex sample matrices. These requirements place an increasing burden on the laboratory and its ability to produce quality results.

For these reasons, it is important that analytical environmental chemists clearly understand and provide information concerning the limitations of their data. There have been numerous examples of misusing analytical data. It is usually not an intentional misuse but rather a general lack of understanding. Regulatory personnel especially, must be acutely aware about the limitations of environmental testing especially in placing concentration limits on specific parameters. It is important that this awareness and understanding be passed on to the public and to the press to prevent common misrepresentations such as the confusion between discussions of chemical contamination (measured levels of chemicals in the environment) with chemical toxicity (actual properties of the raw chemicals).

In this paper the question of analytical results and what they truly tell us will be addressed. This will be accomplished by discussing:

- Characteristics of a Method — including numerous definitions that are frequently used in discussing methods.

- Purposes and Types of Analyses — to differentiate between initial characterization work and final compliance monitoring.

- Some actual examples of analytical variability associated with environmental monitoring for conventional parameters, metals and organic substances.

- And, finally present a discussion of some external factors that can affect the ability of a laboratory to produce good results.

2.0 METHOD CHARACTERISTICS

The U.S. Dept of Food and Drugs Administration (FDA) evaluate methods on the basis of three characteristics (Horwitz, 1982).

- Reliability
- applicability - to a wide variety of sample types
- practicality - with respect to cost, time and training constraints

All of these are important.

Reliability is clearly the overriding consideration and is determined by establishing a method's.

- reproducibi1ity
- repeatability
- systematic error or accuracy
- specificity
- sensitivity or detection limit

These are listed in their approximate order of importance and each will be discussed separately.

Reproducibility is the total between laboratory precision of an analysis determined by inter-laboratory studies or round robins studies. It is a measure of the ability of different laboratories to check each other. Reproducibility is best expressed as a coefficient of variation (CV) which is the standard deviation around a mean/ and expressed as a percent. The real aim of inter-laboratory tests is to determine how much allowance must be made for variability among laboratories in order to make the values interchangeable.

Between laboratory precision is often confused with repeatability or within-laboratory precision. Within laboratory precision defines the ability of one laboratory to check itself. It is also expressed as a CV and should be better than total variability. In most AOAC studies, within-laboratory

variability is about 2/3 of the total variability (Horwitz, 1982). There are some methods for which the total variability is substantially higher than within-lab variability and such methods would be considered very personnel dependent and not good for certain purposes. An interesting experiment was recently carried out in New Zealand where a full assessment was made on the variability of blood alcohol analysis (Horwitz, 1982). Analysts were actually moved to other laboratories and it was determined that their repeatability or precision was higher than when they worked in their own laboratories. Many factors come into play in the overall variability of an analysis.

Accuracy or bias is the expression of how close a measured value is from the true value. Certain analyses can have a well defined error and this is termed a systematic error.

The relationship between precision and accuracy is ideally shown in Figure 1.

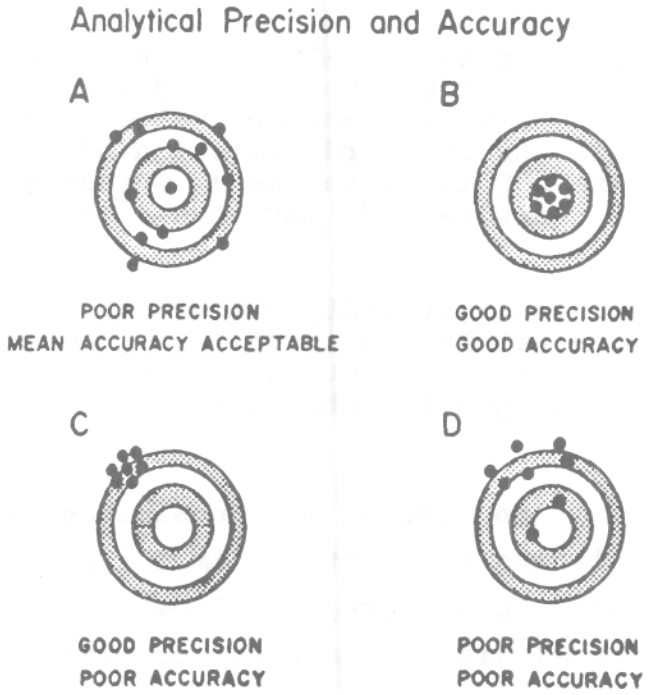

Figure 1. Relationship between precision and accuracy of analytical measurements

Another important consideration in determining the reliability of an analytical measurement is specificity which is the ability of a method to measure an analyte without interference from other chemicals.

Sensitivity is defined by the detection limit which is the smallest amount of an analyte that can be measured with a stated confidence. The American Chemical Society's Committee on Environmental Improvement (CEI) defined the following types of detection limits in an effort to standardize the reporting procedures of environmental laboratories (Keith et al 1983):

- Instrument Detection Limit (IDL) - the smallest signal above background noise that an instrument can detect reliably. Confusion often arises when IDL's are reported as method detection limits.

- Limit of Detection (LOD) - the lowest concentration level that can be determined to be statistically different from the blank. The recommended value of LOD is 3 S.D., where S.D. is the standard deviation of the blank in replicate analyses.

- Limit of Quantitation (LOQ) - the level above which quantitative results may be obtained with a specified degree of confidence. The recommended value for LOQ is 10 S.D., where S. D. is the standard deviation of blanks in replicate analyses. The LOQ tends to be a conservative estimate for reporting of results and may necessitate the rejection of valid data.

- Method Detection Limit (MDL) - the minimum concentration of a substance that can be identified, measured and reported with 99 or 95 percent confidence that the analyte concentration is greater than zero.

The latter definition (MDL) is considered the most realistic approach to calculating a detection limit since it is based on a complete analytical procedure, includes matrix effects, and is derived from the analysis of samples or standards rather than blanks. The calculation of MDL is fully discussed by scientists from the U.S. EPA (Glaser et al. 1981). The MDL method has now been adopted by EPA and the Ontario Ministry of Environment.

It should be realized that a method detection limit defines the ability of the method to detect the analyte in the sample presented to the lab. It doesn't necessarily reflect what the detection limit should be set to in terms of sampling problems. For instance in the measurement of zinc in water, 1 μg/litre or even less is relatively easy to detect. However an MDL of 5 μg/litre is more appropriate because of the difficulty to control contamination at levels below this, during sampling and sample storage.

3.0 PURPOSE AND TYPES OF ANALYSIS

In general, chemical analyses are required for three major purposes (Horwitz, 1982).

- To survey a field to determine the extent of a problem. This could be called characterization. A high degree of accuracy and precision is not necessarily required as the object is to define whether or not an analyte or analytes are present and whether the levels are high or low.

- Step 2 is to then monitor. Monitoring requires numerous determinations with practical methods that can provide sufficient information in the time alloted. Here, the emphasis should be on developing methods that will be rapid and relatively precise - precise enough to determine if significant concentrations are present on a significant number of occasions

- Finally compliance monitoring which requires a high degree of accuracy and precision. Here it is necessary to select a few parameters that are analysed by well defined methods.

In a consideration of environmental monitoring, there are three distinct classes of analytical measurements.

Class 1 - physical parameters that don't measure specific contaminants but rather define the condition of the water with respect to its Suspended Matter, Conductivity, Colour, Oil and Grease, Oxygen Demand, etc. Some of these can be candidates for compliance monitoring but the procedure must be defined. The results are method dependent.

Class 2 - methods for characterization - multi-component tests that are capable of measuring numerous parameters at one time.

i.e., Generalized extraction followed by qas chromatography/mass spectrometry (GC/MS) for organics; inductively coupled argon plasma (ICP) for metals.

Class 3 - specific analyses for parameters such as nutrients, anions, individual metals (usually by atomic absorption (AA) but some are well analysed by ICP) and individual organics by gas chromatography (GC) (or GC with a mass spectrometry type detector).

The selection of the methodology then, is dependent upon the purpose for which the results are required. As an example, consider the requirement of a regulatory agency that must set a comprehensive discharge permit for particular industry with unknown effluent quality. The first task would be to carry out a comprehensive characterization of the effluent. This would establish what is present and what is absent in the wastewater. The multi-component procedures would be best for this purpose. The second step would be to monitor at a defined frequency using more specific (Class 1 and 3) analysis methods. This would establish overall variability of discharge and establish possible relationships among parameters. Finally, a "short list" of parameters would be established for compliance monitoring using well defined specific methods.

The methods selected must be continuously defined by carrying out a comprehensive quality assurance program. A well managed laboratory will devote at least 15 - 20% of its effort on quality assurance. The main components of a quality assurance program involve the analysis of (Maynard et al 1986).

- field blanks - to quantify contamination in the sample collection and storage process
- laboratory blanks - to quantify contamination in the laboratory
- external standards - or "spikes" into distilled water, to ensure the daily standardization is correct
- spiked samples - to measure accuracy and precision and see how these are affected by the sample matrix
- replicates - to measure precision
- analysis of reference materials that are available from various agencies - to measure accuracy and precision

Each of these can be used to define within-lab and between lab analytical variability.

4.0 EXAMPLES OF ANALYTICAL VARIABILITY

Analytical variability can be summarized in an oversimplified way by plotting the determined mean coefficient of variation against the concentration. (Figure 2). The sources of these data are an examination of over 150 independent AOAC interlaboratory studies covering numerous topics from product analysis through to analysis for trace metals and pesticides (Horwitz, 1982). It can be seen that the lower the concentration the higher the analytical variability. As we move towards 1 ppb the variability approaches 60%

Figure 2

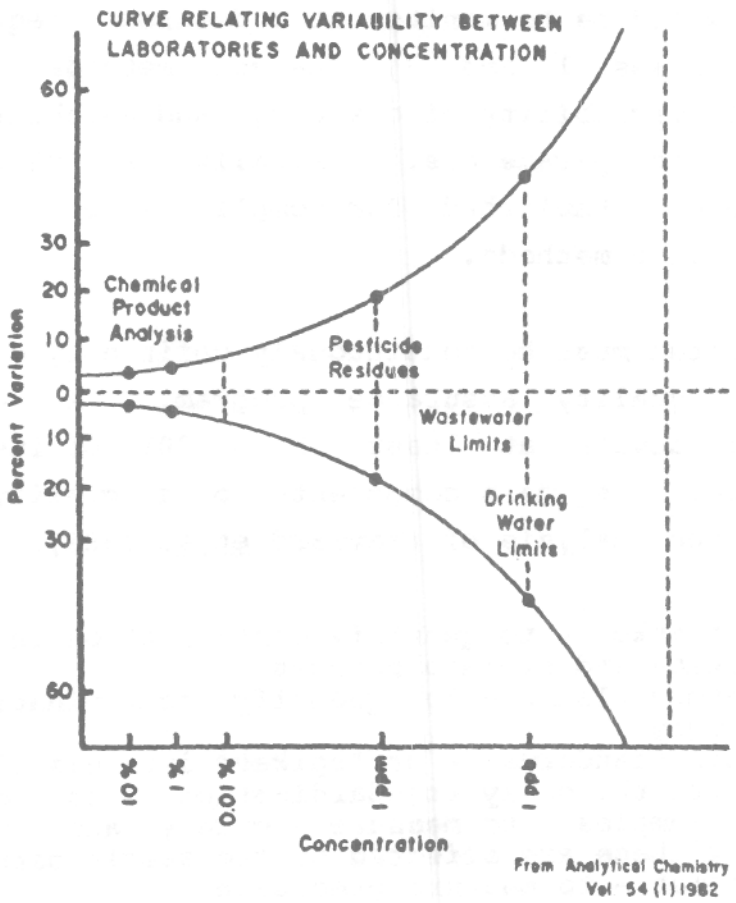

Figure 2 represents an ideal precision curve by well established AOAC methods. There are many analyses for which the precision curve deteriorates at the lower concentrations. There are other

procedures that are specifically designed for low levels and for these the precision is good at the 1 ppb level.

Some examples of analytical variability of various environmental monitoring methods are discussed further below.

4.1 Conventional Parameters

A recent study was carried out by Maynard and Daly (1988) on behalf of Petroleum Association for the Conservation of the Canadian Environment (PACE) to investigate the inter-laboratory variability associated with certain key conventional tests used to monitor refinery wastewaters. Thirteen laboratories participated in this program. The samples consisted of blanks, synthetic samples and actual effluents all submitted blind. They were submitted to the thirteen participants on the same day and the analyses were carried out over a tightly controlled schedule.

The results obtained are given in Table 1 which shows the comparison of the within-laboratory variability (these were run by the laboratory that prepared the samples) and the inter-laboratory variability (the results of the participating labs). The false positives column shows the number of times a positive result was obtained on a samples of distilled/deionized water that were submitted as samples.

The results indicated that some conventional and metal tests can provide relatively precise information on an inter-laboratory basis -

- Suspended Solids
- Organic Carbon
- Chemical Oxygen Demand
- Metals: Cr, Cu, Zn

TABLE 1 RESULTS SUMMARY FOR PACE INTER-LABORATORY STUDY

No.	Parameter	"True" Value	Within-Lab Summary (n=5) Mean (or Med)*1	S.D.	"False" Positives	Inter-Lab Summary (n=7-13)*3 Mean (or Med)*1	S.D.	"False" Positives
1	Suspended Solids	<1.	(<1.0)	-	0	(<1.0)*2	-	1/12
2	" "	30.	29.33	0.82	-	22.12	4.07	-
3	" "	-	39.40	3.85	-	30.78	5.66	-
4	Oil & Grease	<1.	(<1.0)	-	0	(<1.0)	-	5/12
5	" "	X+22	16.75	2.22	-	12.90	11.30	-
6	" "	9	5.40	2.07	-	6.24	5.09	-
7	Phenols	-	0.039	0.004	-	0.040	0.023	-
8	"	0.01	0.010	0.001	-	0.013	0.006	-
9	"	<0.001	(<0.001)	-	0	(<0.001)	-	6/12
10	DOC C	<0.5	0.5	-	3	(0.5)	-	5/9
11	" C	15.0	15.42	0.20	-	17.13	3.91	-
12	" C	-	11.60	0.22	-	18.189	9.12	-
10	COD	<10.	(<10.)	-	0	(<10.)	-	2/8
11	"	37.5	33.42	2.42	-	39.48	7.09	-
12	"	-	86.60	3.72	-	88.73	19.69	-
13	Ammonia N	-	1.23	0.11	-	1.68	0.73	-
14	"	<0.2	(<0.01)	-	-	<0.01	-	4/13
15	"	0.60	0.58	0.014	-	0.71	0.36	-
16	Chromium Cr	<0.01	(<0.01)	-	0	(<0.02)	-	1/8
17	"	<0.01	(<0.01)	-	0	(<0.02)	-	1/8
18	"	0.27	0.276	0.005	-	0.295	0.111	-
16	Copper Cu	<0.01	(<0.005)	-	0	(<0.01)	-	1/7
17	"	0.07	0.068	0.001	-	0.075	0.014	-
18	"	0.1	0.100	0.001	-	0.106	0.016	-
16	Zinc Zn	(<0.01)	<0.005	-	0	(<0.010)	-	1/8
17	Zinc Zn	0.09	0.080	0.004	-	0.096	0.023	-
18	Zinc Zn	(<0.01)	<0.005	-	0	(<0.010)	-	1/8

DOC = Dissolved Organic Carbon COD = Chemical Oxygen Demand
SD = Standard Deviation
*1 When the sample analysed was actually a blank, a median (med) is reported (in brackets) rather than a mean. A standard deviation does not accompany the median.
*2 When the median is actually a less than detection, the most commonly reported detection limit is given.
*3 n = 7 to 13 since not all labs analysed for each parameter, the n for each parameter is given in the last false positives column.
True Value - was known for synthetic spiked samples and the distilled water sent as samples. (from Maynard & Daly 1988)

For these tests, the coefficients of variation were 15 - 50 percent and the number of false positives on the "blind" blanks was minimal.

For the other parameters analytical variability is of a more serious nature -

Oil & Grease
Phenols
Ammonia

For these tests the coefficients of variation were 46 - 87 percent and a number of significant false positives were reported.

Because of the problems observed with these tests, it is suggested that lower level results be treated with extreme caution. There appears to be a significant probability that a total phenols result of, say 0.004 mg/L, could indeed be zero or on the other hand could be as high as 0.008 mg/L. This type of variability near the detection limit (required by the Ontario Government) of 0.002 mg/L needs to always be considered so the detection limit itself does not incorrectly reflect data reliability. A reviewer of data must be aware of the uncertainty associated with a measurement.

4.2 Metals

There have been many round robin studies involving the analysis of water, sediments and tissues for metals. Two recent programs that involved a large number of laboratories were co-ordinated by the Canadian National Research Council (NRC) to certify a number of sediment and tissue samples for eventual use as certified reference materials (Herman & Boyko 1985 a & 6). The results are presented in Tables 2 and 3 below. The relative standard deviations in these two examples, ranged from 7 to 34 percent after the elimination of outliers.

This information demonstrates that the reproducibility of analysing metals in sediment and freeze dried tissue can be very good. However, it must be emphasized that these results represent optimum conditions.

i.e., - a number of well qualified labs participated
- all the laboratories knew the samples were for an interlab study to certify the materials
- the sediments and tissue were "ideal" — well blended, etc.
- many outliers were eliminated (there were not many)

TABLE 2: FROM NRC STUDY ON METALS IN SEDIMENT*

ELEMENT		n	RANGE (µg/g)	MEAN (µg/g)	RSD
Copper	Cu	35	93.6 - 154	129	7%
Zinc	Zn	34	226 - 381	314	7%
Arsenic	As	20	6 - 28.7	20.2	34%
Cadmium	Cd	28	0.9 - 3	2.00	13%
Mercury	Hg	19	0.97 - 2.05	1.5	10%
Lead	Pb	32	119 - 255	211	10%

n = number of laboratories (after elimination of outliers)
* harbour sediment representative data only - from Berman & Boyko 1985a

TABLE 3: NRC STUDY ON METALS IN TISSUE*

ELEMENT	n	RANGE (µg/g)	MEAN (µg/g)	RSD
Copper Cu	38	3.0 - 11.2	6.1	17%
Zinc Zn	38	15 - 50	27.9	8%
Arsenic As	19	20.5 - 42.3	30.8	14%
Cadmium Cd	14	0.040 - 0.25	0.08	26%
Mercury Hg	19	0.88 - 2.49	1.84	15%
Lead Pb	35	3.7 - 16	9.0	24%

n - number of participating laboratories (after elimination of outliers)
* dogfish muscle tissue representative data only - from Berman & Boyko 1985b

4.3 Organic Parameters

With respect to organic parameters, analytical variability can be more acute especially when numerous parameters must be measured at one time. If a specific parameter or a shorter list is measured, the methodology can be more specific.

A study involving the inter-laboratory analysis for PCB in sediment was recently carried out (Alford-Stevens et al. 1985) with co-efficients of variation ranging from 29 to 37%.

TABLE 4: PCB IN SEDIMENT

	SAMPLE 1	SAMPLE 2
MEAN (µG/G)	52.	2.7
N	20	20
SD (µg/g)	15	1.0
CV (%)	29	37.

N = number of participating laboratories
SD = standard deviation
CV = coefficient of variation
Selected results from Alford-Stevens et al 1985

EPA methods for priority pollutant monitoring are usually assessed by carrying out spike and recovery experiments. EPA contract labs are required to prepare quality assurance samples by adding or "spiking" defined concentrations of contaminants to distilled water. Four such check samples must be initially analysed followed by a routine program involving 10% of the sample load. The EPA published acceptance criteria for all the priority pollutants (EPA, 1984) and some typical examples are presented in Table 2.

TABLE 5 EPA PRIORITY POLLUTANTS : RECOVERY OF SPIKED COMPOUNDS

PARAMETER	SPIKE LEVEL (µg/L)	RANGE (µg/L)
Vinyl Chloride	20	D - 43.5
Benzene	20	15.2 - 26.0
Pentachlorophenol	100	38.1 - 151.8
Phenol	100	16.6 - 100.0
Benzo-A-Pyrene	100	D - 195.
Crysene	100	D - 134.

Selected results from EPA, 1984 D = Detected

It can be seen that the EPA recognizes a high degree of variability for certain parameters. For instance for vinyl chloride, a tolerance of more than 100% variability is expected. For a 20 ppb spike sample/ a result from just detected (D) to 43.5 ^g/L is acceptable. Benzene on the other hand, is expected to be less variable at the 20 ppb level.

It should be emphasized that the above EPA recovery criteria reflect within-laboratory precision. Total variability (between laboratory) could be higher as could the variability of analysing actual samples, as opposed to spiked distilled water.

In a recent study carried out by the Chemical Manufacturers Association in the United States (Stanko & Fortini, 1985), three

highly reputable EPA contract labs were used to analyse spiked ground water. Some typical results are shown in Table 3 - two showing reasonable precision, two showing poor precision.

TABLE 6 : CMA ASSESSMENT STUDY

PARAMETERS	SPIKE LEVEL (µg/L)	RANGE (µg/L)		
Benzene	50	43.	-	86.
Ethylbenzene	200	ND	-	225.
Nitrophenol	300	11.	-	765.
Chlorophenol	300	200.	-	270.

Selected results from Stanko & Fortini, 1985

The most alarming aspect of the CMA study was that while quantitative agreement was poor for many parameters, there were some parameters for which qualitative agreement was not achieved. That is there were false positives and false negatives. Of 36 parameters spiked at levels of 50 - 300 ppb into 9 groundwater samples —

15 - were detected in all samples analysed by the 3 labs
6 - were detected in 80 - 99% of the samples
7 - were detected in 40 - 79% of the samples
and,
8 - were detected in less than 40% of the samples

This CMA study does demonstrate that full priority pollutant monitoring can be highly variable. However the monitoring is necessary and ideally suited to characterize wastewater and industrial wastes. Some procedures are definitely better than others and the purposes for the analytical programs must be properly assessed prior to their implementation. For instance GC/MS priority pollutant monitoring following a general base/neutral and acid extraction is designed to characterize with

respect to numerous contaminants at one time. This procedure is not for routine monitoring of one or even a few specific components. Many of the priority pollutant methods are not adequate for compliance monitoring and would not stand up in court.

Compliance monitoring must be carried out by well defined procedures for which the analytical variability is known and acceptable. And - most importantly - compliance limits must take into account this known analytical variability.

5.0 PROJECT PLANNING TO CONTROL EXTERNAL FACTORS

While the examples presented above demonstrate that there are definite limitations that must be known about analytical results, a well run program will generate meaningful data required for survey and regulatory purposes.

To ensure a program will provide the data that is needed, the project managers and field staff must communicate with the laboratory personnel to discuss the important components of the program such as:

- Sample collection and storage procedures
- Decisions regarding homogeneity of the samples - in some waste evaluation studies, non-homogenous samples must be collected.
- Parameter list - the list should be narrowed down as much as possible.
- Detection limit requirements
- Scheduling - good quality analytical work takes time to complete
- Report requirements

With these types of discussions at the project planning stage the overall program will be much more valid.

BIBLIOGRAPHY

Alford-Stevens, A.L.; W.L. Budde and T.A. Bellar. 1985. Interlaboratory study on determination of PCB in environmentally contaminated sediments. Anal. Chem. 57 (13) 2425-2458.

Herman, S. and V. Boyko, 1985, National Research Council of Canada Marine Analytical Chemistry Standards Program First Intercomparison Exercise for Trace Metals in Marine Sediments: NRC MS1/TM Preliminary Report. NRC, Ottawa, Canada

Berman, S. and V. Boyko, 1985. First Intercomparison Exercise for Trace Metals in Marine Biological Tissues NRC BT1/TM. Report II NRCC No. 25324. NRC, Ottawa, Canada

Environment Canada, 1988. Listing Toxics under CEPA - Is the Chemistry Right? Canadian Environmental Advisory Council, Environment Canada, Ottawa, Canada. May 1988.

EPA, 1984. Federal Register Part VIII Oct. 26, 1984. 40 CFR Part 136.

Glaser, J.A., D.L. Foerst, G.D. Mckee, S.A. Quave and W.L. Budde 1981. Trace analysis for wastewaters. Env. Sci. Tech. 15: 1426-1435.

Horwitz W., 1982 Evaluation of analytical methods used for regulation of foods and drugs. Anal. Chem. 54(1) pp 67A-76A.

Keith, L.H., W. Crummett, J. Deegan Jr., R.A. Libby, J.K. Taylor and G. Wentler. 1983. Principles of Environmental Analysis. Anal. Chem. 55: 2210-2220.

Maynard, A.W. and J. Daly, 1988. Interlaboratory Study to Evaluate the Analytical Variability of Conventional Parameters and Metals. Petroleum Association for the Conservation of the Canadian Environment (PACE), Ottawa, Ontario (in Press).

Maynard, A.W., John M. Park, and R.W. Deverall. 1986. "Quality Assurance in the Environmental Laboratory" ASL Analytical Service Laboratories Ltd., Vancouver, B.C.

Stanko, G.H. and P.E. Fortini, 1985. Inter and Intra laboratory assessment of selected SW-846 methods for analysis of Appendix VIII compounds in groundwatcr. Presented at U.S. EPA symposium on the Analysis of Pollutants in the