MONITORING ACID MINE DRAINAGE

MEND Project 4.7.1

This project was funded by Energy, Mines and Resources Canada and the British Columbia Ministry of Energy, Mines and Petroleum Resources under the Canada/British Columbia Mineral Development Agreement.

August 1990

MONITORING
ACID MINE DRAINAGE

Prepared by:

Emily Robertson, Biometrician
1525 - 200th Street
Langley, B.C. V3A 4P4
(604) 530-1080

In Association With:

Steffen Robertson and Kirsten (B.C.) Inc.
800 - 580 Hornby Street
Vancouver, B.C. V6C 3B6

August 1990

This project was funded by Energy, Mines and Resources Canada and the British Columbia Ministry of Energy, Mines and Petroleum Resources under the Canada/British Columbia Mineral Development Agreement.

Titles in this series:

Draft Acid Rock Drainage Technical Guide Volume 1

Literature Review for Biological Monitoring of Heavy Metals in Aquatic Environments

Hydrogeological Assessment and Development of AMD Control Technology for Myra Falls Waste Rock

Kutcho Creek Project Acid Generation Testwork Phase II

Geochemical Assessment of Subaqueous Tailings Disposal in Buttle Lake, British Columbia

Acid Drainage from Mine Walls: The Main Zone Pit at Equity Silver Mines

The Effect of Treated Acid Mine Drainage on Stream Macroinvertebrates and Periphytic Algae: An in situ Mesocosm Experiment

In 1990, the Task Force continued to support seven ongoing projects and coordinated the start up of four new projects. The eleven projects included the following:

2.30 Underwater Disposal of Waste Rock and Tailings

2.60 Blending and Segregation (Kutcho Creek Project)

3.30 Mount Washington Evaluation

3.40 Dry Covers on Waste Rock

3.52 Constructed Wetland: Bell Mine

4.30 Aquatic Invertebrates Monitoring

4.40 Sediment Monitoring Techniques

1.30 Cinola Gold Project Kinetic Tests (Rock Pads)

3.31 Britannia (AMD) Toxic Leachate

4.50 Ion Speciation Model

For availability and purchase of any of the above research reports contact:

BiTech Publishers Ltd.

Distributed and sold by:
BiTech Publishers Ltd.
903 - 580 Hornby Street
Vancouver, British Columbia
Canada V6C 3B6
Tel:(604)669-4280 Fax:(604)669-1779

Printed and bound in Canada

1991

ISBN O-921 095-I 7-1

Foreword

This report on Monitoring Acid Mine Drainage prompted much discussion and debate within the Monitoring Subcommittee. It contains concepts and suggestions for improving the way we do environmental monitoring in the mining industry. Any document that calls for new directions will inevitably spark debate.

In the end, it was decided that the report was too important not to be released. This decision was made with the full knowledge that many readers will have useful reactions, comments and questions about its contents. I would ask that these be directed to the Task Force, and not the author, since the project contract is over, and she has already donated well beyond a reasonable limit of free time.

The Monitoring Subcommittee has resolved that the concepts and suggestions made in this report need to be illustrated in demonstration projects at actual mine sites. The planning for these is now underway, and should form part of the workplan for 1991-1992.

Brian Wilkes, R.P. Bio.
Chairman, Monitoring Subcommittee,
B.C. Acid Mine Drainage Task Force

SUMMARY

This report has been prepared in response to the Acid Mine Drainage Task Force's request for a review of the effectiveness of the monitoring programs at existing mines, and the design of optimum monitoring programs for the B.C. context. Although conceived as a statistical exercise using existing data sets, the exercise immediately foundered due to the inadequacy of the available data. This report documents our conclusion that the existing fixed-frequency data sets are suited only for the description of very long-term trends; accurate estimates of mean concentrations, loads and peak values require different sampling methods. Having very little data to work with and an obvious need for education regarding monitoring design, the emphasis of the project shifted to writing a mini-text on monitoring design for ARD sites.

Section 1 begins with an examination of the monitoring methods currently used in Waste Management permits for mines with acid drainage potential. The unreplicated fixed-frequency samples are shown to be inaccurate in estimating mean concentrations and completely inadequate to indicate peak values and short-term fluctuations. Alternate methods of monitoring are reviewed from the monitoring and statistical literature, each with its own advantages and disadvantages. Different monitoring goals (e.g. detecting long-term trends, accurately measuring excursions) are discussed with reference to the different monitoring methods available. The point is made that no monitoring program can be optimized statistically without clearly stated goals: a program that efficiently measures monthly means would not also efficiently catch peak values. Rather than burden industry with monitoring programs that attempt to measure all possible variations for all possible contingencies, it is recommended that the Acid Mine Drainage Task Force engage in an 'Environmental Audit' process to determine the specific goals of monitoring for each site. This discussion leads to the first and most important recommendation in the report: to critically examine the information needed for management at each mine: accuracy, threshold concentrations, time lags, cost constraints and risks for each ARD component. Monitoring results should be 'defensible', both in the scientific and enforcement senses.

Chapter 2 is a review of basic sampling statistics as they are applied to water quality data. The problems of dealing with rapidly fluctuating values are emphasized. The technique of performing a preliminary sampling study of a site is described. Preliminary studies determine the variances in different components of a site, and thus permit the calculation of predicted accuracies of different sample sizes, selection of optimum strata, and the allocation of future samples to optimize sampling efficiency. The lack of proper preliminary sampling at any of the B.C. mines examined in this study made it impossible to perform one of the initial goals of this project, which was to design optimum monitoring methods for specific sites. Sampling design requires measures of variance, which are lacking in unreplicated fixed-frequency data.

Understanding the process of the generation and release of ARD helps to focus a monitoring program on critical time periods. Chapter 3 illustrates how the process

affects water quality sampling, with an emphasis on seasonal and flow-related effects. The critical importance of good flow data at AKD sites is emphasized.

Chapter 4 is an exploration of the best monitoring data set available; a year's worth of almost daily data from a coastal mine. Day-to-day variations in concentration are high and greatly exceed the analytical error of the mine's environmental lab; i.e. the speed with which a sample can be analyzed many be more important for getting an accurate reading than the usual 'quality assurance' concerns of laboratory technique. Daily data are compared with the monthly official monitoring record to illustrate the short-comings of monthly sampling in a rapidly fluctuating system. Three different monitoring schedules are designed for this mine to suit three different monitoring goals: peak values, mean values and loads. For example, the error of the estimated annual zinc load could be decreased by more than 60% by taking 6 additional samples (18 instead of 12). This improvement is accomplished by allocating the samples according to the observed seasonal variance pattern instead of fixed monthly intervals.

Chapters 5 and 6 contain general guidelines for the monitoring of untreated mine water and monitoring in the receiving environment. This discussion was limited to generalities because there were no data sets available that supported proper monitoring design or even a rigorous determination of general confidence intervals or accuracy. The use of experimental design to ensure that proposed field studies (both regular monitoring and special studies) are more likely to have conclusive and useful results is very strongly recommended. Section 6.6.3 illustrates what can happen when more effort is put into trying to sample 'everything' rather than carefully identifying the information goals of the monitoring program.

A brief discussion of biological monitoring as an alternative to water quality monitoring is the main topic of Section 7. Biological samples integrate water quality over time, and thus contain much more information than an accurate measure of an ephemeral quantity such as dissolved concentrations. Any discussion of optimum water quality monitoring would be incomplete if it did not point out the value of biological monitoring.

The theme of this document is that improved statistical methodology for monitoring rests on defining the information needed for good management. Too much emphasis has been put on laboratory analysis techniques and on trying to apply statistics to squeeze something out of existing data sets; not enough emphasis has gone into answering hard questions about how defensible the monitoring data is. What degree of certainty is needed on estimates? Does the data alert us when an environmental risk threshold has been breached? Is it available in time to permit useful management responses? What could we do better if we had the information? These are not statistical questions, but they are of the greatest priority in optimizing ARD monitoring.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

This report was prepared at the request of the Acid Mine Drainage Task Force who recognized the need for improved water quality monitoring methods in order to better address environmental concerns and to support good regulatory supervision.

Acid rock drainage (ARD) is caused by the natural oxidation of sulphide minerals contained in rock that is exposed to air and water. The source of most new acid generating rock is ore and waste rock exposed by mining; ARD caused by mining is also called acid mine drainage (AMD). There are at least 6 active and 5 abandoned mines in British Columbia currently generating ARD (Steffen Robertson & Kirsten, 1989).

The sampling methods used to monitor ARD at active mines are the same as those used to monitor other liquid mining effluent: single samples are taken quarterly or monthly, and are used to represent the average values for that time period. These unreplicated samples are used as estimates* of average concentration, they are compared to other unreplicated data sets for impact assessment, and they are scanned over time to look for trends. Many important sources of variation+' and error* in single samples have been ignored or assumed to cancel out over time (Oguss & Erlebach, 1976).

As our understanding of ARD has increased, it has become clear that it is characterized by high frequency variations and seasonal effects. As the analysis of data in this report demonstrates, quarterly or monthly single samples may provide very inaccurate estimates of true* mean concentrations for the time period, especially in streams and rivers. Furthermore, instead of focusing on mean concentrations, it may be the range* or frequency of changes in concentration of ARD components that are the most relevant to impacts on aquatic organisms. These variations are not monitored at a11 by the sampling regime in current permits.

Through cooperation and good communication between industry and government agencies, additional (extra to the permit) sampling has become the rule at many mines, with 'gentlemen's agreements' governing the sampling, analysis and sharing of information. Thus good management of affected water resources has generally been accomplished, though by somewhat irregular means.

1.1 Clarifying the Purposes of Monitoring

Improvements in monitoring usually focus on increasing accuracy' or precision*, without concern for whether the information being collected is optimally useful. No increase in accuracy or precision is valuable if the wrong phenomenon or quantity is being measured.

---

* Words designated by this symbol are defined in the Glossary of Statistical Terms, Appendix I.

The permits currently written for ARD mines accept very rough estimates of monthly or quarterly means (*i.e.* single grab samples) as the most relevant measure of water quality. This reduces all the variations, the peak values and sudden changes that may have occurred in the month to a single parameter, the mean. Unfortunately, even an accurate9 monthly mean, by itself, is not a good predictor of environmental impact except at grossly polluted levels. Therefore to focus on more accurately estimating monthly means is to risk missing a more relevant measure, *i.e.* one that would detect short-term, subtle or incipient changes in the environment. The first principle of designing a monitoring program is to be sure that the most useful types of information will be collected.

1 .1 .1 Demonstrating Compliance

Of course, the purpose of monitoring is to demonstrate compliance to a permit...but what is the purpose, ***exactly***, of the permit? The permit officially states the means by which government managers shall protect public interests: by ensuring drinking water quality, fisheries resources and general environmental protection. In Systems Operations language, the permit defines the feedback and control mechanisms by which the manager makes management decisions. The permits issued by the Waste Management Branch set out concentrations, the 'objectives', for each ARD component which the observed samples shall not exceed. Both statistically and in Systems Operations terms, this is a very poorly defined regulatory mechanism.

The first problem is the inaccuracy of the monitoring samples. Analytical accuracy is not the issue here, since compliance monitoring samples are analyzed by independent labs with more than adequate accuracy. The accuracy in question is the accuracy with which each sample represents the true mean for the location and time period in which it was taken. Variation is the key to sampling design [see Section 2.1]. Without an estimate of variation, it is impossible to say how accurate or inaccurate a single sample is. The variation can only be established by taking multiple samples (over time &/or space) and examining their frequency distribution. (This is the purpose of a preliminary study, to identify the variances in the system against which future samples can be compared.) Thus a single sample's accuracy is unknown (Oguss & Erlebach, 1976).

Therefore, the certainty with which the manager can be sure that concentrations are compliant is as wide as the range of the true values, and is unknown. Furthermore, when a noncompliant sample is observed, it is left to the manager's 'judgement call' to determine how serious (frequent, long lasting, high risk, etc.) the excursion was.

[Monthly or quarterly single samples collected over many years may eventually demonstrate that there is little or no variation in a given parameter", or that all the variation occurs well below the 'alarm' level. In these cases sampling can probably continue unchanged, using the accumulated old data as the reference for accuracy. However, for

all parameters whose record shows substantial variation or whose values approach the 'alarm' level, the accuracy of single samples must be assumed to be poor (Section 2)].

The second problem with the 'control mechanism' defined in permits is the difference between the objective concentration and the lowest known toxicity or impact concentration. The gap between the two values represents a safety zone, the width of which is not a standard or rigorously determined distance. The objectives are based on considerations of background levels, interactions with other possible pollutants, best practicable technology, etc. The degree to which a model of risk helps to define this safety zone is not clear: do risks increase linearly from the objective to the toxic concentration? or in steps with thresholds? logarithmicly? While the exact relationship may be unknown, the assumptions made should be clearly stated, because the manager's response should be based on them.

These two problems compound each other when compliant observations are drawn from a site where the range of real values exceeds the objective some small proportion of the time: the excursions are undetected and the associated risk is unknown.

The current wording used in permits has us 'shooting' at an arbitrary 'target' (the objective) with a very inaccurate 'gun' (the fixed-frequency single sample). In order to get useful information, the managers are routinely put in a position of having to supplement the compliance data with additional sampling.

It is clear that monitoring to demonstrate compliance to a permit is a thankless task if the permit does not directly link the information collection process to clearly defined assessments of risk to water resources. Ward, et ***al*** (1986) have called this the 'Data Rich but Information Poor' syndrome in water quality monitoring, and say that it typifies the great majority of monitoring programs currently in operation in North America.

1.1.2 Serving Management Goals

At a conference of the Ecological Society of America entitled, 'New Approaches of Monitoring Aquatic Ecosystems' (Boyle, T.P. 1987), the following comment was made regarding environmental data and information:

> ***"Water quality monitoring has concentrated on data collection efforts while largely neglecting information issues. Information is extracted from data when trends are quantified or correlations through time or space are validated. Simply collecting more and more data with little regard to its information content wastes valuable resources. To assure management or regulatory success, more attention must be paid to methods for precisely specifying the information required from a data set before the data are collected- If this is done well, sufficient funds may be saved to support environmental rehabilitation and resource conservation."*** (Perry, ***et al., 1987)***

We know that ARD is characterized by seasonal variation and fluctuating values, not by steady-state values. Before we can design a trend detection program with the right level of sensitivity', or an early warning system that allows us to respond fast enough to a short-term problem, it is necessary to consider what is at risk, what sorts of chronic and acute impacts are possible, the likely time-frame of the impact, and the manager's response strategies. While this may seem like a major digression from the statistics of designing monitoring programs, it is in fact central.

Is the mean concentration the most important determinant of response in aquatic organisms? Perhaps the range, the peak values, the variance, the rate of change, or the load are more important determinants. Are there different vulnerabilities in the system at different times of the year ? Are there concurrent stresses on this system from other sources (natural or man-made) that might influence their impact?

Monitoring programs cannot be optimized unless the objectives of the program are clearly spelled out in terms of meaningful chronic and acute thresholds, the accuracy needed and the response time (Lettenmaier, *et al,* 1978, and Ward, *et al,* 1986). This is a clear contrast to the retrospective 'What do these data tell us?' approach. Once the goals are clearly stated, it is a straight forward task to design a program that efficiently and economically produces the required data.

Government agencies and academics throughout the industrialized world have been developing methods for improving their ability to accurately anticipate impacts and to flag important trends. It is widely accepted that the design of monitoring programs needs to be an interactive multi-stage process.

Whitfield (1988) recommends a 5 step process for each site-specific design: 1) establishment of a monitoring goal; 2) selection of a sampling strategy to meet the goal; 3) periodic review of adequacy of sampling including quality control studies; 4) optimization of sampling related to the goal over time; and 5) review of adequacy of monitoring goal.

Mar, *et al.* (1986) recommend a 4 step process: 1) identification the environmental changes of interest and the effects that would most likely manifest these changes; 2) selection of variables and sampling techniques, formulation of cause and effect hypotheses, and search for alternate or proxy variables; 3) design, in particular exploring the tradeoffs between improved discrimination and added cost; and 4) integration of the monitoring program into the overall management goal. Mar emphasises cost factors as a primary element in the design process, because the exploration of the tradeoffs helps to focus the investigation on the necessary level of accuracy needed for good management.

These approaches are convergent with Holling's 'Adaptive Environmental Assessment and Management' (Holling, 1978) techniques which focus on 3 issues: 1) determining the best strategy to sample the quantity of interest; 2) determining the statistical basis for the

sampling design (i.e. the preliminary investigation); and 3) estimating the cost of such observations.

The interactive planning process that appears to be best suited to the ARD situation is called the 'Environmental Audit', developed by Perry, Schaeffer and Herricks (1987). They emphasize the distinction between surveillance (trend monitoring) and management monitoring. Trends can be detected within historical data records and within fixed-frequency data records that span many years, without benefit of a *priori* hypotheses or design. In contrast, regulatory monitoring is only valid when it is planned to produce information for decision making. The Environmental Audit process begins by translating 'management questions' into formal, quantifiable statements called 'Audit Objectives'. Management questions are generally concerned with perceived damage, criteria for exceedances, and consequences of taking action. From these concerns the Audit Objectives are derived: quantifiable statements of what will be measured in order to support management decisions. This process requires decisions on the resolution needed for detecting changes or exceedances. Perry et *al* suggest that cost concerns and the limitations of resources for monitoring should be considered a separate issue from information needs, in order to avoid confusing the two. Once the information needs are listed along with the necessary sampling design for each component, management can allocate resources based on their perception of the risks. Almost inevitably this process identifies 'tension points' where an exhaustive data set would be too expensive, and precision must be sacrificed for economy. The advantage of the process is that such tradeoffs have been identified clearly and that the choices made are defensible in comparison to the alternatives.

Incorporating an Environmental Audit, or similar planning process, into' existing ARD monitoring programs would require preliminary intensive studies of the variance patterns at each site, decisions regarding the resolution needed for each variable, design of the optimum sampling schedule to achieve this level of resolution, and *rewriting the permit to include these information goals* or 'Audit Objectives' *(not* specific sampling methods which may quickly become obsolete). This guarantees that the information required for good management will be available, and also makes the rationale and priorities behind the sampling methods clear to all interested parties.

1.1.3 The Goal of 'Defensible' Numbers

If monitoring is to provide management with reliable information on which to base important decisions, the rigour with which the numbers are collected and evaluated is of utmost importance. A truly optimum monitoring program will produce numbers (e.g. concentrations) that are 'defensible' in three ways:

▸ Defensible observations, in the sense of being true and accurate representations of the values that really occurred. There should always be a calculated confidence interval' associated with each estimated value (e.g. means) showing the reliability of the estimate.

▸ Defensible criteria for judgement, *i.e.* the thresholds and limits enforced should be ones which represent valid criteria of risk or environmental response.

▸ Defensible source, in the sense of accurately identifying the mine as the cause of the problem (as opposed to background, other sources, or random environmental changes).

Of all the data sets offered for examination in this study, including 'official' and internal monitoring programs, not one was producing defensible numbers in any of these senses.

1.2 What is Wrong With Single, Fixed-Freuuencv Samples?

Before reviewing the alternate methods of monitoring, it is valuable to examine the limitations of the existing data that have been collected as fixed-frequency single samples.

Data of this type are ideally suited for only one type of analysis: trend monitoring over long periods of time. Using time-series analysis" (e.g. Whitfield and Woods, 1984), it is possible to detect very small trends in water quality, or to measure small impacts due to upstream changes, despite seasonal changes and annual cycles. For example, using monthly data for 13 years from the Kootenay River, Whitfield and Woods were able to give rigorous estimates of the nature and magnitude of changes in water quality resulting from the construction and operation of the Libby Dam, even though each month's data was affected in a sightly different manner. Unfortunately, the number of years of data required by time-series methods (generally at least 10 years for monthly data) makes this type of analysis a poor management tool.

For short-term comparisons, the fixed-frequency single samples have very severe limitations. They are: inaccurate in representing the time interval, non-random*, and drawn from heterogeneous" and heteroscedastic" time strata. To simplify this discussion, we will use monthly data as an example, with the understanding that the same problems apply to annual, quarterly, or weekly data.

1.2.1 Inaccuracy of the Single Sample

A single monthly sample is taken to represent the average concentration during the entire month. How well it does this depends on how variable the concentration was during the month. The greater the variation in concentrations, the less is the likelihood that the single sample 'caught' a value close to the true mean.

When many samples are taken in the month, it is possible to statistically calculate a mean and to calculate the 'confidence' of that estimate: we can say, for example, that the mean concentration was 26mg/l with a 95% confidence interval of ±4mg/l. If we want to compare this to a different location's data, we now know how different they have to be in order to be confident that the difference is real. For instance, if the upstream mean was 24mg/l ± 3mg/l, we see that these values overlap and there is no real difference; the higher downstream mean is not significantly higher. In a different situation we might want to compare one mean of 26.00 ±0.04 with another of 23.00 ±0.03; these are very significantly' different.

Without replicate samples, there can be no calculated mean with its calculated confidence interval [see Section 1.3.1].

Note that there are three 'dimensions' to natural variation within each month: instantaneous", temporal" and spatial*. Instantaneous variation is the observable differences between samples taken at the same time; e.g. if you filled 6 bottles simultaneously, the differences between them would be a measure of the instantaneous variation. Temporal variation refers to the day to day or moment to moment changes during the month. Spatial variation refers to the observable differences between sampling locations.

A good preliminary study estimates each of these components of variation within the month, because differences 'between' can only be demonstrated by comparison to differences 'within' (Green, 1979). Once each component of variation has been examined, a good monitoring design will allocate replicates so as to most efficiently improve the accuracy of the estimate'. For instance, in well-mixed flowing water, there may be virtually no instantaneous or spatial variation, but very high temporal variation; therefore single samples could be taken in one location, with the number of replicates per month being determined by the resolution needed by management.

Unfortunately, years' worth of monthly single samples taken without any measure of variation cannot be used to make valid comparisons between sites or between years because the variances of the underlying populations+' are unknown. There is no way to

[1] What we have referred to as 'accuracy' is technically 'precision: the reproducability of observations. Unless there is bias in the measuring method, precision will lead to accuracy. The accuracy of laboratory analysis of water samples is a possible source of bias in AMD data, but it is very small compared to sampling error. Therefore we have used the terms 'accuracy' and 'precision' synonymously.

be certain that apparent differences are not due to chance alone; there is no way to distinguish between barely significant differences and highly significant differences.

1.2.2 Heterogeneity Between Samples.

Given good monthly data, we might want to calculate annual averages and use them to compare sites, using simple statistical methods such as t-tests". Parametric" methods assume homogeneity+', *i.e.* they assume that each sample is taken randomly'+' from a well-mixed population. Seasonal changes in water quality create heterogeneity? the means and variances of some months will be different from those of other months. When this pattern of changes is overlooked and unlike samples are grouped together, the result is to greatly increase the variance of the annual samples, which in turn means that differences between two such samples would have to be much greater in order to be distinguished using t-tests or analysis of variance' (ANOVA). (The appropriate analysis would be a non-parametric Paired Comparisons Test.) In many cases the few high samples that may be taken in the year will bias the annual mean.

1.2.3 Different Variances Between Samples.

Equality of sample variances is another basic requirement of most parametric tests. The unreplicated data available from most ARD sites provide no measure of the variances from which each was taken, but it is very likely that the variances associated with high values are greater than those associated with low values. The precision of samples drawn from high variability months is much less than the precision of samples from low variability months. As with non-homogeneity, heteroscedasticity" tends to obscure distinctions that might otherwise be made. This has important consequences for inferences drawn from old ARD data sets. For instance, a comparison between a contaminated site and an uncontaminated site is made less powerful" when different variances obscure the differences between the means.

ANOVA and t-tests are parametric analyses that assume homogeneity and homoscedasticity" of each sampled population, as well as normal" distributions (see Section 2) within each subdivision of the design (e.g. within years or within sites). The unreplicated single samples that constitute the main record at ARD sites probably violate all of these assumptions.

In practical terms, t-tests and ANOVA are fairly robust when applied to data that fails to meet the strict assumptions under which the tests were derived and tested. The results obtained by using these tests on single monthly grab samples may not be seriously misleading. But the significance tests" applied to the results will be incorrect, and there is no direct method of calculating exact significances. Therefore the analysis of single monthly samples should be restricted to non-parametric methods.

1.2.4 Fixed Intervals of Sampling

Fixed-frequency samples are obviously not random with respect to time. This lack of randomness might bias the data if there is any source of variation that is also on a fixed monthly schedule, e.g. equipment maintenance. Ideally all possible sources of variation in mine operations and water treatment should be identified so that sampling can be randomized with respect to their schedules.

Another problem with fixed interval sampling is that it misses shorter term changes and events completely, or over-represents the importance of brief occurances. For example, if data are collected in fixed quarterly intervals, shorter-term seasonal events are easily missed. For monthly data sets this risk is reduced but still present.

Finally, there is the problem that a fixed-frequency data set cannot be used to estimate lag effects shorter than the sampling interval. Therefore many years of such data contain no information about the duration of excursions or other phenomenon that last a shorter time than the sampling interval. In contrast, a long (e.g. 10 year) ***randomly*** sampled monthly data set would probably contain useful information on the autocorrelation+' in the system, since many time intervals would be well represented.

When single samples are taken within set time intervals (e.g. monthly) they should be taken at random times within each time interval.

1.2.5 Composite Samples Aren't Much Better

Some permits require the collection of composite samples in an attempt to better represent the true mean value; e.g. 'weekly composite of daily samples, 2 per day, 7 days per week.' When samples are taken more than once a month, individual samples are allowed to exceed the permit limits as long as the arithmetic monthly mean is compliant. The Metal Mining Liquid Effluent Regulations and Guidelines give different values for maximum authorized monthly concentrations depending on whether these are based on single grab samples (e.g. the maximum for copper is 0.6 mg/l), composite samples (0.45 mg/l), or arithmetic mean of several samples (0.3 mg/l) [see MMLERG Schedule 1, part 1: Authorized Levels of Substances]. This schedule acknowledges the variation that is present in the data. A single grab sample truly represents only the instantaneous concentration at the moment it was taken; the composite sample blends together the concentrations of several moments ***so that information about their variation is lost.*** Only the mean of several samples can be accepted as an estimate of the true mean, because the information about the differences within the sample are preserved and can be used to calculate confidence limits for the estimated mean.

1.3 Improved Monitoring Methods

Since the shortcomings of fixed-frequency single samples have been known for a long time, there have been many publications in recent years devoted to improving the quality and efficiency of sampling while reducing the risks of undersampling. Following a preliminary intensive study (essential to determine variances), there are many choices of sampling methods, each suited to a monitoring goal and particular type of variation (Whitfield, 1988; Liebetrau, 1979). The methods most relevant to sampling mining waste water are briefly reviewed below.

1.3.1 Replicated Sampling

Replication is the process of taking a pre-determined number of samples which then jointly represent the 'population' (e.g. time interval) from which they were drawn. When the variance has been determined by preliminary study, the desired precision for an estimated mean period can be achieved by taking replicate samples. The number of samples needed is a function of the precision needed in the resulting estimate of the mean, and constitutes an important element of the monitoring design. The samples should be taken randomly from the 'population' they represent. Replicate sampling is useful when variances are predictable and the mean (rather than peak values) is the focus of the monitoring. Replicated data are ideal for use with parametric statistical methods. (Green, 1979)

1.3.2 Sequential Sampling

Sequential sampling is a highly efficient method of estimating a mean value to a pre-determined level of precision, and it is especially valuable in cases where the variance is not known in advance. This method requires that the sampler keep taking additional single samples until the desired level of precision (of the estimate of the mean) has been reached. Unnecessary and redundant sampling are avoided, which is especially valuable in cases where individual sample costs are very high. The method is inappropriate if there is a long time lag between sample collection and the availability of the laboratory readings; it is also inappropriate if the mean or variance of the water being sampled is changing during the process and therefore instable. Data sets collected in sequential sampling episodes can be compared using parametric statistical methods if the samples are normally distributed and their variances are comparable. (Wald, 1947)

1.3.3 Exceedance Driven Sampling

This is a modified form of fixed-frequency sampling in which the frequency can be increased when observed levels exceed predetermined thresholds; it is intended to enhance the surveillance capability of a monitoring program. The monitoring method outlined in the Metal Mining Liquid Effluent Regulations and Guidelines (MMLERG) is an exceedance driven formula. The strength of the method is its improved tracking of

rising values, and therefore increased likelihood of flagging non-compliant values. The efficiency with which exceedance-driven sampling can 'catch' peaks and avoid oversampling during stable periods is determined by the flexibility and feedback time built into the exceedance driven schedule. For instance, the schedule in the MMLERG is based on the running 6 month average, and can only increase sampling to a weekly schedule. This particular design is very ineffective in flagging short-term peak values, and very slow to return to infrequent sampling after a high value has been 'caught'. However, an exceedance driven program could be devised that was more efficient in rapidly fluctuating situations. (Valiela and Whitfield, 1989)

This method has most of the disadvantages of the fixed-frequency method (unreplicated non-random samples), and is especially inappropriate for estimating means because the sampling frequency increases as the observed values increase, creating significant bias. The data are unsuitable for parametric analyses.

1.3.4 Markovian Sampling

Markovian sampling also is a method of sampling more frequently when the observations rise above threshold levels; both the sample size and time interval are adjustable based on 'alert levels' (e.g. complying, marginal, warning and alert) determined by the previous set of samples. In highly variable or episodic systems, Markovian sampling responds more quickly than the equivalent exceedance driven program. The process of defining the alert levels with their corresponding sample intervals and levels of replication is a valuable exercise for clarifying management strategies. Markovian sampling cannot guarantee a predetermined level of precision in estimating peak values and, like exceedance driven programs, it produces biased estimates of mean values. (Arnold, 1970; Smeach and Jernigan, 1977)

1.3.5 Stratified Sampling

Stratifying* the 'population' into units of homogeneous* variance is a major improvement over unstratified sampling when there are areas or time units within the population that have different variances. It allows greatly increased sampling efficiency because sampling effort can be distributed according to the variance within each strata'. For instance, if variance is correlated* with flow conditions, the year could be divided into time units representing flow conditions (rather than calendar months), and the means within each flow stratum could be very efficiently estimated with equal accuracy. Stratified sampling must, of course, be based on a preliminary study to determine the strata. When there are significantly different strata in the water body being monitored, this method is the most economical way to produce accurate estimates of mean values. The resulting data can be used in parametric statistical tests which allow unequal sample sizes. Stratified sampling is not suitable for tracking peak values. (Green, 1979)

1.3.6 Continuous Monitoring

Some ARD components, such as pH and conductivity, can be measured continuously with probes, and the result digitally recorded in intervals as short as fractions of a second. Computerized data loggers designed for this purpose can provide a complete record of changing values, and can telemetrically alert people if preset thresholds are exceeded. Data of this sort is no longer a sample in the usual sense, but a complete record. It provides the most accurate calculated means and the most accurate tracking of peak and minimum values. It also provides valuable information regarding the frequency of changes and the duration of peak values, both of which may have important biological effects. Continuous monitoring is the only method which does not require a preliminary study; in fact, it is the best method of doing the intensive preliminary study. The disadvantages of continuous monitoring are high initial costs for equipment and calibration, and the lack of suitability for monitoring many ARD components such as dissolved metals.

1.3.7 Continuous Proxy Monitoring

Observed concentrations of many ARD components are correlated with variables that can be monitored continuously. When these correlations are strong, the record of the continuously monitored variable can be used as a proxy for the correlated variable, accepting a measurable error of the estimate. When the error of estimating from the proxy is unacceptably high, the correlation relationship can be used for prompting direct sampling of a target variable when the continuously monitored proxy variable exceeds preset threshold levels. For instance, dissolved heavy metal concentrations are likely to be negatively correlated with pH; the data logger can monitor pH continuously and can 'call' a technician to take samples for metals when the pH falls below a certain value. A preliminary study is needed to determine the correlations, the number of replicates needed to estimate peak values, and the appropriate threshold values. If properly designed, such a system can give both accurate mean values and good tracking of peak values.

1.4 Designing New Monitoring Programs

The variety of monitoring methods listed above makes it clear that there are good methods available to suit a wide variety of situations. But substantive improvements in ARD monitoring will not come just from substituting one of these methods for the fixed-frequency single samples. The design of new programs should include the following steps for each mine:

. Preliminary study of variances and seasonal patterns. [Section 2.6]

- Interactive planning to clarify monitoring objectives, to determine the necessary resolution of the data and the response times needed, and to make choices regarding cost vs. precision trade-offs.
- Selection of appropriate monitoring methods.
- Rewriting of the permit to incorporate new methods.
- Periodic review.

Of these tasks, only the first and third require statistical input. The major challenge of optimizing monitoring programs remains the problem of anticipating what sort of 'news' from the site would prompt someone to take some action.

2.0 SOME BASIC SAMPLING STATISTICS

This chapter is a brief primer of some of the theory behind sampling design, as it relates to ARD monitoring. It is offered for the convenience of the reader and is not intended as a substitute for a good textbook. Readers wishing more detailed explanations arc referred to any of the following: Sokal and Rohlf, 1969; Steele and Torrie, 1960; Cochran, 1963; Green, 1979.

The basic model of impact assessment is the comparison of before and after impact samples, or impact vs. control samples, or both. The ideal situation is one in which the baseline data ('before') provides a permanent record for future comparisons *and* a well-chosen control provides an ongoing comparison to account for independent effects such as acid rain, increasing recreational uses, logging, etc. Statistically this model leads to a two-way factorial design:

	Impact	Control
Before (Baseline)		
After		

Each square or 'cell' of this model represents a set of samples for one variable, such as dissolved zinc concentrations. The most powerful statistical method for analyzing this sort of design is Analysis of Variance*, or ANOVA, with which the interaction between the two-way differences can be evaluated. Thus if both the control stream and the impacted stream are affected by, let's say, a road nearby, the comparison between before and after conditions can still be made, and the conclusions are much more valid than either l-way comparison would be. In the ANOVA results, the F-test of the Interaction SS ('sum of squares') is the primary measure of significance of an impact. This basic model can be expanded to incorporate covarying factors (making it an Analysis of Covariance, ANCOVA) or to deal with multiple variables simultaneously (Multiple Analysis of Variance, MANOVA), each suited to special applications.

The number of samples in each cell of the design determines the degrees of freedom that will be available for significance testing, and this fact deserves attention in all baseline and on-going monitoring programs. The more samples in each cell, the more accurately the within-cell variances are known, and therefore the stronger the test. While samples sizes can vary, the overall strength of the test is largely determined by the smallest cell, ie. the one with the fewest samples. No amount of extra sampling in the impacted zone can compensate for inadequate sampling of the baseline or control. The actual number of samples needed in each cell can be determined during preliminary sampling (in fact, this is the primary purpose of preliminary sampling): it is influenced by the amount of variation present and also by the resolution needed, *i.e.* how small a

difference or change should be detectable? This resolution should always be chosen with great care, since the cost of achieving higher resolutions increases very rapidly.

The basic ANOVA design has some statistical assumptions that should be considered carefully for application to water quality data. ANOVA assumes that the within-cell data represent homogeneous and normally distributed 'populations'. Water quality data is rarely normally distributed and there are often time or flow related factors causing non-homogeneity; these are discussed in the following subsections.

2.1 Freauencv Distributions of Water Quality Data

Water quality is a phenomenon that exists in real time, changing perhaps from moment to moment. The spatial and instantaneous components of variation (*i.e.* differences from place to place in a lake, or differences between simultaneous samples taken from poorly mixed water) correspond to the classic types of data (e.g. 'weights of newborn babies' or 'bushels of grain per acre') that are used to illustrate basic statistical applications. You can think of these data as numbers waiting to be randomly sampled in the same way that you might sample needle lengths within a stand of trees. However the changes in water quality *over time* constitute a different sort of 'population' for sampling.

If we could see a continuous record of the concentration of one variable measured at one spot over a period of, let us say, a month, it would show periods of no change, periods of increasing concentration and periods of decreasing concentration. The 'population' is the infinite number of instantaneous values that occurred during the month. This population has a true maximum and minimum, a true range, and a true mean value.

For practical purposes we reduce this population to a sequence of samples; each one is a 'snip' out of the continuous record. For instance, the continuous record for a month could be represented by 744 hourly samples. These data could be used to calculate an estimated mean, standard deviation*, etc., and to draft a frequency distribution, which would indicate the relative frequency of occurrence of concentrations within the range observed during the month. The frequency distribution tells us whether the scatter of the data is symmetrical around the mean, whether the data 'fit' a standard parametric distribution (e.g. a normal distribution), and identifies irregularities that may have importance in sampling design. The variance that can be calculated from the samples is a descriptive statistic indicating how widely scattered the data are. [Note that we are ignoring the fact that the samples occurred in a specific sequence and are therefore not random samples -- more on this below.]

Figure 1 illustrates some types of continuous records that might occur in a stream, and the frequency distributions that would be produced by an intensive sampling program. Groundwater or effluent from a closely controlled process might be constant over long

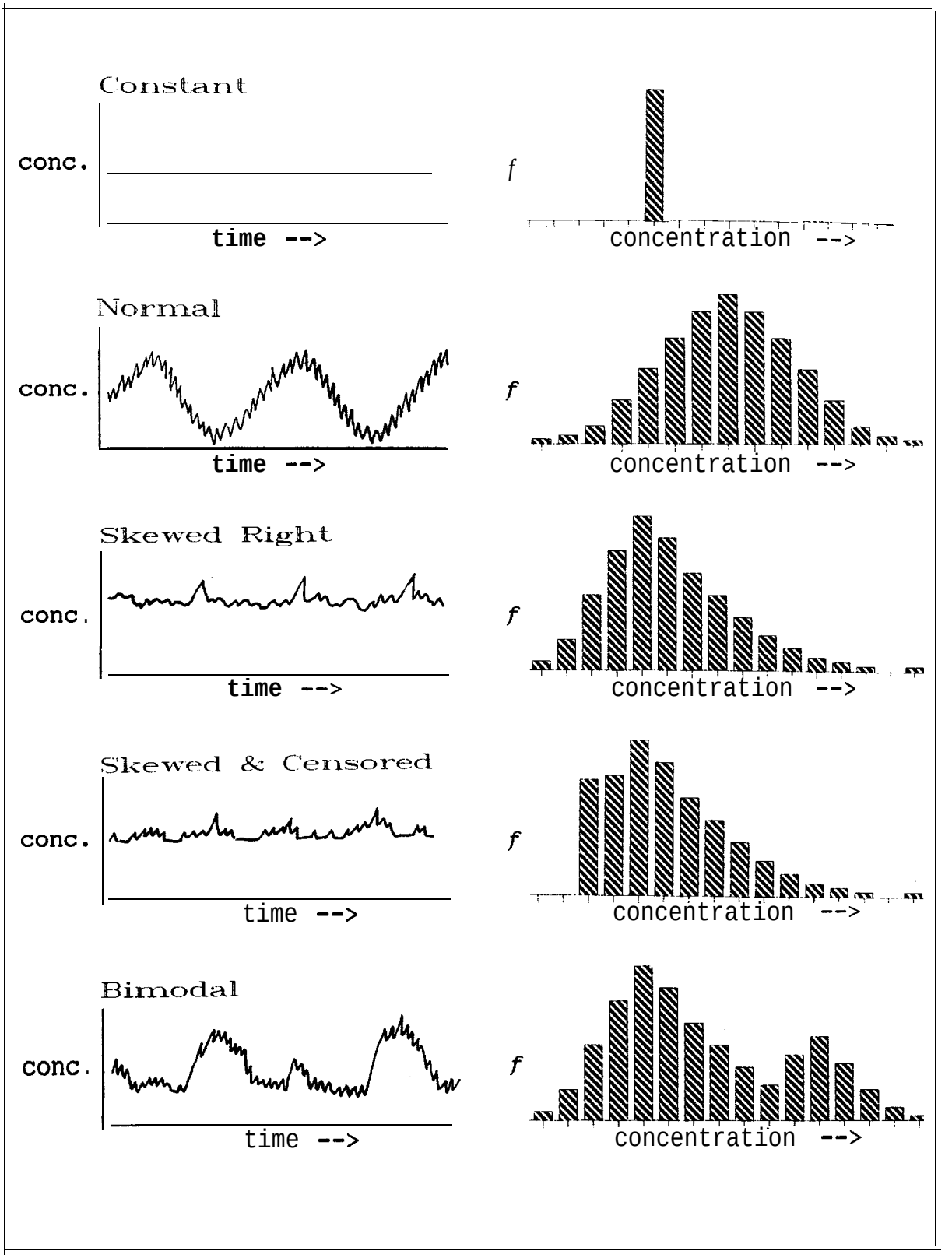


Figure 1: Types of Continuous Records and Their Frequency Distributions.

periods of time. A normal distribution might be found for a variable like hourly surface temperature measurements. Concentration data is likely to show a distribution that is skewed to the right, and may be censored' by a detection limit. Another pattern that we frequently see in water quality data is the bimodal' or multi-modal' distribution, indicating that data from time periods with different distributions have been grouped together.

The sample variance and the shape of the frequency distribution are of primary concern to a sampling design exercise, because they determine the number of samples needed in order to reach a predetermined confidence level for estimates of the mean and peak values. The smallest sample sizes are needed when the variance is small (relative to the desired confidence) and when the frequency distribution is normal. Normal distributions are not common in water quality data; right-skewed distributions are typical.

2.2 Coping with Non-Normal Distributions

Many sampling programs are designed using the assumption that the distribution is normal, without ever checking to see whether it actually is. When data are normally distributed, the well-described properties of the normal curve can be used to calculate estimated means, ranges and confidence limits from a very small number of samples. Unfortunately these calculations give biased estimates when small samples are drawn from a population that is not normal. Calculating a mean, for instance, gives equal weight to all samples because the distribution is assumed to be symmetrical around the mean, and therefore the samples are equally likely to lie on either side of the true mean. But if the distribution is asymmetrical, then the samples are more likely to have come from one side than the other, and the mean calculated from a small sample is likely to be biased.

2.2.1 Transformation

Often the data from a skewed population distribution can be mathematically trans-formed" to a set of numbers that has a normal distribution. The population parameters estimated this way are unbiased and can be back-transformed to the original units. Lognormal distributions are often found in water quality data and respond very well to this treatment (Niku, et.al. 1981; Shaarawi and Kwiatkowski, 1986). Transformed data can be used in any parametric statistical procedure that requires normally distributed data (e.g. t-tests, ANOVA), and is therefore the most easily analyzed and understood method of dealing with non-normal data.

2.2.2 Monte Carlo Techniques

Skewed and irregular populations that do not lend themselves to transformation can be dealt with by using modelling techniques. One modelling method, called Monte Carlo simulation", uses a selected set of baseline data to describe an observed frequency distribution, and then assumes that any future data will be drawn from populations with the same distribution. This has the advantage of requiring no parametric assumptions about the shape of the distribution, but it is only 'good' for as long as the baseline data truly represent the current population. The preliminary investigation must check for different distributions in different strata in order to be sure that the appropriate distribution is used in future applications. Although Monte Carlo techniques are often used when baseline data are 'thin', the actual data requirements to reach the same power as an equivalent parametric test are higher.

2.2.3 Adjusting for Detection Limit Effects

Often the only irregular feature of the distribution is caused by the effect of detection limit censoring. Using data from a good preliminary study, it is possible to estimate the frequency distribution of the missing left tail' of the curve, and thus to accurately adjust future samples in order to calculate unbiased means, etc. (Gleit, 1985).

All of these techniques require an initial study that collects enough samples to adequately characterize the underlying frequency distribution.

2.3 Stratification

Frequency distributions are very likely to be different in different locations and time periods. Exploring the water body for locations that have different means and/or variances is a well-understood basic principle of sampling. These areas are called strata, and they must be sampled separately. If data from different strata are lumped together, the effect on the overall frequency distribution is to greatly increase the variance. In some cases, a bimodal or multimodal frequency distribution will result. Leaving the data lumped would greatly increase the uncertainty associated with each sample, and therefore greatly increase monitoring sample requirements.

Different frequency distributions can occur over time strata as well as locations. For example, there may be seasonal effects such as lake turn-over or spring run-off that create distinctly different variances and frequency distributions during different times of the year. These also need to be sampled separately in order to optimize a monitoring program. Redistributing sampling effort so that all time strata are sampled with equal efficiency will result in equal confidence intervals for all strata, and an economical design.

2.4 Autocorrelation: the Lag Effect

No matter how accurately we can characterize the frequency distribution of water quality data, it is necessary to account for the fact that the data do not occur randomly but in a sequence over time. Each observation is in part a reflection of the concentration or value that could have been observed in earlier time intervals. There are important lag effects 'built in' to the body of water, that are determined by the mixing and flushing rates, and sometimes by physical or chemical interactions (e.g. buffering). Because each observation is partially dependent on it's own prior values, statisticians refer to it as 'serially dependent' or 'autocorrelated'.

Autocorrelation is an important factor in determining optimum sampling frequencies, especially when an accurate mean is the goal of the monitoring program. When there is a high level of autocorrelation, as in groundwater or a well-mixed lake, concentration values are slow to change, and the optimum frequency of sampling is low. These systems are very economical to sample because each sample remains a good indicator for a long time. Conversely, when there is very little autocorrelation, such as in a small stream draining a watershed with low retention, the optimum sampling frequency is higher.

To estimate a mean using the fewest samples, the sampling frequency should be long enough so that each sample is independent of the previous one. When the lag effects are very short, the sampling frequency does not have to be very high; it only needs to be high enough to collect enough samples to achieve the desired confidence limits of the mean.

A good preliminary study should determine the autocorrelation in each water body to be monitored. In addition to permitting the most economical design for monitoring mean values, this also will indicate the duration of peak or minimum values, which may be an important aspect of environmental impact. The lag effects can only be measured by sampling more frequently than the duration of the lags; thus a preliminary study should include some intensive temporal sampling in each strata.

2.5 The Interpretation of Small Samples

2.5.1 The Mean

Each sample taken from a body of water is like a 'snip' out of the continuous record of all values that occurred. How useful is it as an indicator of the mean value? The situation is analogous to taking one sample from any population: the sample is likely to be a good indicator if the population variance is low, and a poor indicator if the variance is high.

Water monitoring programs that rely on infrequent single samples are assuming that the variance is low (compared to the desired confidence limits) and that autocorrelation is high.

If a proper preliminary study has been done, the value of a single sample or small set of samples can be determined statistically (e.g. Oguss and Erlebach, 1976). For instance, if the data are normally distributed (or can be transformed to such), the variance of the initial data can be used to calculate the upper and lower bounds of a 95% confidence limit for a new estimated mean based on a single new sample. The Z-score of a new observation can be calculated to indicate the likelihood that the new observation comes from the previously described population or represents a changed population (e.g. a rising or declining mean). This presentation of the results provides much more certainty about the quality of the new information, and the distinctions that can or cannot be made with it.

Unfortunately the confidence limits of single samples are often very wide in comparison to the certainty needed for management decisions. If the frequency distribution is normal, a very small number of replicates will usually 'tighten' the confidence interval to an acceptable level of certainty. Non-normal, irregular distributions have a higher data requirement to reach the same levels of confidence as a normal distribution. The worst scenario is the situation where the distribution has not been studied in advance; in this case the confidence limits of a single sample cannot be determined.

There are two important factors to note:

- The confidence intervals of old data cannot be calculated without frequency distribution and autocorrelation information from an intensive study.
- The target confidence interval should be based on the certainty needed for decision making, not on a convenient sample size or standard procedure.

2.5.2 The Range

When several samples are taken from a large population, it is safe to assume that none of them is the true maximum or minimum of the population. How, then, do we estimate important peak values if they were not directly sampled? One of the advantages of the normal distribution is the ease with which the tails of the curve can be estimated. Given normally distributed data and an estimated mean and variance, we can use Z-scores to calculate a probability of a certain value (let us say, the maximum permit value) occurring in that population. As an illustration, a manager could use Z-scores to calculate the odds of values ≥0.1mg/l occurring during a sampled period. In another situation, a manager might receive sample data which is all below the permit value, but indicating by its variance and the Z-score of the permit value that exceedences probably *did* occur during the time period.

When the frequency distribution is not normal and cannot be made normal via transformation, the preliminary study data can be used to generate a model of the tails of the distribution, and this can be used to estimate the probabilities of exceeding specified values in subsequent monitoring data.

2.6 Preliminary Sampling: A Prerequisite

The importance of preliminary sampling is probably the most underemphasized principle of field studies. There is no substitute for it; decades of archived data cannot be analyzed to estimate the necessary parameters.

The following sequence of steps is appropriate to situations in which ARD contamination is actively occurring or suspected.

2.6.1 Preliminary Stratification

A preliminary study begins with the identification of all factors that might influence the mean or variance of ARD variables. Sampling sites should be established at all locations where there is any rationale for different values. Time dependent variables such as seasonal effects, flow relationships, etc. should be anticipated. When the delimiters of the strata aren't known in advance (e.g. the flow rate at which there are substantial changes in variance), it is wise to use smaller/shorter strata in the preliminary study. Some of these space and time strata may be grouped together in the final design, but they need to be explored separately first.

2.6.2 Cofactors

Any independent factors, such as flow rate or temperature, that may influence the mean or variance of the data, should always be measured during the preliminary study. These 'cofactors' may be found to account for a substantial amount of the variation in the ARD variables, and tracking them may permit a substantial reduction in water sampling in the final monitoring design.

2.6.3 Instantaneous Variation

Within each combination of location and time strata, 4 to 6 replicate samples should be taken to determine instantaneous variation. The results will allow an exact determination of the number of replicates needed for future sampling. Instantaneous variation includes 'real' variation as well as analytical error and field or processing errors (Quality Assurance). If instantaneous variation is overlooked, it becomes a hidden part of short-term temporal variation, and might lead to a design where sampling has to be done more frequently, thus raising monitoring costs. The data from this part of the preliminary

study may also be useful for determining correlations between ARD variables, which may allow important economies in the final design.

2.6.4 Autocorrelation

Short-term temporal variation is explored by taking many samples within each time strata. First, an intensive study is needed to determine the time-lag between independent samples within each strata. Monitoring several key variables, such as pH and conductivity, with a continuous probe would give the most exact measure of autocorrelation within each stratum. This should then be corroborated for the other ARD variables by collecting sets of samples at shorter time intervals.

The smallest unit of time such that subsequent samples are independent becomes the minimum sampling interval for future random sampling.

The data collected in this segment of the preliminary study may also be used to characterize hysteresis relationships between flow and concentration (see Section 3.2.5).

2.6.5 Frequency Distribution & Variance

Having determined the time interval between independent samples, a set of at least 30 samples should be taken randomly during each time stratum to determine the frequency distribution of the data. Thirty samples is usually adequate to demonstrate. that a normal distribution is normal; more will be needed to adequately characterize an irregular distribution. In the interest of speed, it would be acceptable to sample the first 30 independent intervals and then examine the data to determine if more samples are needed. To be entirely correct, however, random sampling over the entire time stratum is recommended. This may be especially important if independent cofactors influence the mean or variance.

If seasonal or flow influences are not understood well enough in advance to establish strata before the preliminary sampling, it is wise to sample key variables daily (or at the minimum independent sample frequency if it is greater than daily), for a year to establish these strata. The designer can use regression* methods to select strata based on continuous cofactors, such as flow, and can fit seasonal strata more accurately.

2.6.6 Design for Future Monitoring

With the complete set of preliminary data *and* clearly stated control criteria (e.g. confidence limits, threshold values) the designer can optimize the monitoring program. One of the most important tasks is to reexamine the strata used in preliminary sampling: the final stratification should be- reduced to only those strata that have different means, variances and/or frequency distributions. Sampling effort can be allocated to strata so as

to equalize efficiency. Economies in sampling can be gained by making use of variable/variable correlations and/or relationships with cofactors. The autocorrelation in the system determines the frequency of sampling to be used in accurately tracking peak values and also the minimum frequency between independent samples. If variable sampling frequencies are to be 'built in' to the monitoring program (e.g. exceedance-driven or markovian sampling), these can be set up with suitable feedback times, sampling frequencies and replication. If the data are not normally distributed, the designer can choose the appropriate transformation, or can select appropriate non-parametric methods for analyzing and reporting incoming monitoring data.

A monitoring design should always include an outline of suitable statistical analysis methods for the resulting data, and an explanation of the limitations of the data.

2.6.7 Reanalysing Old Data

Data collected before the preliminary sampling program was done can sometimes be reanalysed using the results of the preliminary sampling program. For example, the confidence limits of single samples can usually be estimated. Unfortunately, in most cases very little can be gained with retrospective analysis because the data are too incomplete, lacking proper time stratification and frequency, replication, and measurement of cofactors.

2.6.8 Preliminary Studies at Uncontaminated Sites

For preliminary studies prior to active ARD generation, the aims of monitoring are different, and the preliminary study is much less elaborate. An uncontaminated water body cannot show the patterns of variance that will occur with ARD contamination. The baseline study should identify strata and cofactors, and complete an instantaneous and a short-term temporal study in each strata and at a range of values of each cofactor. This will provide information on the autocorrelation in the system, the natural ranges and frequency distributions of ARD component variables, and the correlations between them. With this information a low-intensity monitoring program can be designed to optimize the flagging of changes that may be due to ARD contamination.

The cost of doing a thorough preliminary study is small compared to the cost in the future of having inadequate baseline data. It guarentees that questions regarding change and impact can be answered efficiently and with good certainty. A proper preliminary study should be a requirement for all new mining developments with ARD potential.

3.0 ARD DISCHARGES: WEATHER AND FLOW RELATIONSHIPS

An understanding of the generation and release of ARD, as it occurs in B.C. mines, is essential for designing optimum sampling methods. This section contains descriptions of the patterns of variance and correlations between ARD components, how the climate at the mine site affects the seasonal pattern of concentration, and some considerations for the resulting impacts on different receiving environments.

3.1 Generation and Release of ARD

The most complete source of information regarding acid rock drainage in B.C. is the Draft Acid Rock Drainage Technical Guide (Steffen Robertson & Kirsten, 1989); the following comments are taken from this source.

ARD is produced by natural oxidation of sulphide minerals when rock bearing these minerals is exposed to air and water. Mining is not the only source of ARD, but the tonnes of porphyry and massive sulphide ores that are brought to the surface by metal mining constitute the major increment in ARD. There is a time lag between exposure of the sulphur bearing rock and the release of ARD which depends upon pH, temperature, oxygen availability, degree of saturation with water, surface area exposed, the presence of acid neutralizing minerals, and the presence of bacteria (*Thiobacillus ferro-oxidans* and others).

The acid produced by oxidation mobilizes heavy metals and other soluble constituents contained in the rock. The acid may subsequently be buffered or neutralized by the receiving waters, but the high metal loadings remain in solution and may seriously harm aquatic organisms. Components of ARD include sulphate, acid, iron, manganese, copper, aluminum, lead, cadmium, zinc, arsenic and nickel.

Each ore produces a unique mix of acid and heavy metal leachate, and because these components have different mobilities, ARD does not have a constant composition. The site-to-site differences and variabilities within each site are demonstrated in Table 1. Table la compares the coefficients of variation (CV) of ARD components from different sources. [The coefficient of variation, being the ratio of the standard deviation over the mean, expressed as a percent, is a measure of relative variation that allows us to make a simple comparison between sites having very different means, or when different units of measurement were used. A CV of 100% indicates that the standard deviation and the mean are equal. The lower the CV, the more narrow is the observed variation relative to the mean value.]

Table lb compares correlations between ARD components within three sites. Note that most of the relationships are weak ($-.7 < r < .7$), and that they vary from site to site.

Table la: Coefficients of Variation at Several ARD Sites,

	Equity Silver	Westmin Myra	Island Copper	Mt. Washington	Kindrat
pH	10.8	5.9	19.1	6.9	6.1
SO_4	72.1		81.0	70.1	82.3
As	108.3		159.1		68.5
cu	483.0	90.0	80.5	28.5	12.5
Fe	372.7	99.1	262.3	54.6	50.0
Zn	229.9	130.4	54.0	42.6	46.9
Al	464.8	68.3		39.7	28.6
Cd	94.1	207.7	56.2		

Table lb: Correlations Between ARD Components at Three Sites.

Equity Silver: Bessemer Creek at Siltcheck Dam n=396

	pH	SO_4	As	CU	Fe	Zn
SO_4	0.1229					
As	-0.0312	-0.2268				
cu	-0.4286	-0.03 13	0.0098			
Fe	-0.3311	-0.1537	0.1686	0.8523		
Zn	-0.5065	-0.0244	-0.0013	0.9179	0.6772	
Al	-0.4081	-0.08 11	0.0298	0.9649	0.9449	0.8391

Island Copper: North Drainage Ditch n=55

	pH	SO_4	Fe	Cd	cu	Zn
SO_4	-0.6770					
Fe	-0.0183	-0.1541				
Cd	0.0057	0.0730	-0.0918			
cu	0.0677	-0.0957	-0.0365	0.6547		
Zn	-0.0321	0.0614	-0.0838	0.9278	0.6338	
Mn	-0.2667	0.3204	0.0204	0.7807	0.3029	0.7095

Westmin: Old Tailings Line Road Seepages n= 11

	pH	SO_4	Al	cu	Fe	Mg
SO_4	-0.4553					
Al	-0.3990	0.9886				
cu	-0.3815	0.9830	0.9955			
Fe	-0.5585	0.7332	0.6316	0.6294		
Mg	-0.3851	0.9889	0.9962	0.9958	0.6357	
Zn	-0.3919	0.9754	0.9966	0.9951	0.5838	0.9926

ARD is found in underground mine workings, open pit drainage, and in waste rock piles, tailings and ore stockpiles which are exposed to precipitation, runoff and seepage. When ARD problems are identified, mines are required to collect runoff from contaminated areas (usually the entire mine site) and treat it to neutralize the acidity and remove heavy metals.

3.2 Seasonal Patterns of ARD Release

Most ARD sites have a seasonal pattern in the concentrations of ARD components in the drainage water. Exceptions are found in the constant concentrations of adit waters, where ARD is released under relatively constant conditions of flow (B. Godin, *pers. comm.*). More commonly, the acid and metal salts generated in the ARD process will accumulate as long as there is enough water to support the oxidation process and not enough to wash them out of the rock. Therefore during dry periods, frozen conditions or light precipitation, there may be very little evidence of ARD contamination in surface waters. The first rain (or snow melt) that is heavy enough to wash through the rock will carry a very high concentration of acid and heavy metals. Subsequent rains may wash out equal or even higher concentrations if the first rain left many salts behind, or may carry lower concentrations if earlier washings were relatively thorough. Thus the basic 'model' of ARD release is that it is proportional to water flow through the rock and to the quantity of accumulated salts remaining to be washed out.

Noting a seasonal pattern is important in monitoring design because it permits the seperation of the data into seasonal strata; without stratification the data are often multimodal. Both examples below have bimodal patterns in the unstratified data.

3.2.1 Coastal (no snowpack) Mines.

When there is no snowpack, the fall rams following a dry summer are the time of greatest ARD release. Figure 2 illustrates this phenomenon by comparing a surface flow hydrograph with dissolved copper and zinc concentrations in untreated water that has been collected from a waste rock dump. This is 1989 data from Mine 'A' which is in a mild coastal climate and receives little snow or freezing temperatures.

3.2.2 High Elevation or Interior (heavy snow) Mines.

When there is a significant snow pack or prolonged period of freezing temperatures, as would occur in high elevations along the coast or in interior locations, the spring snowmelt period is the time of greatest ARD release into runoff waters. This pattern is illustrated in Figure 3 with data from Mine 'B' which receives significant snow 7 months of the year. Fall freshet concentrations are marginally higher, but the load is clearly greatest in the spring.

Figure 2 Seasonal Pattern of Flow and Metal Concentrations at a Coastal Mine.

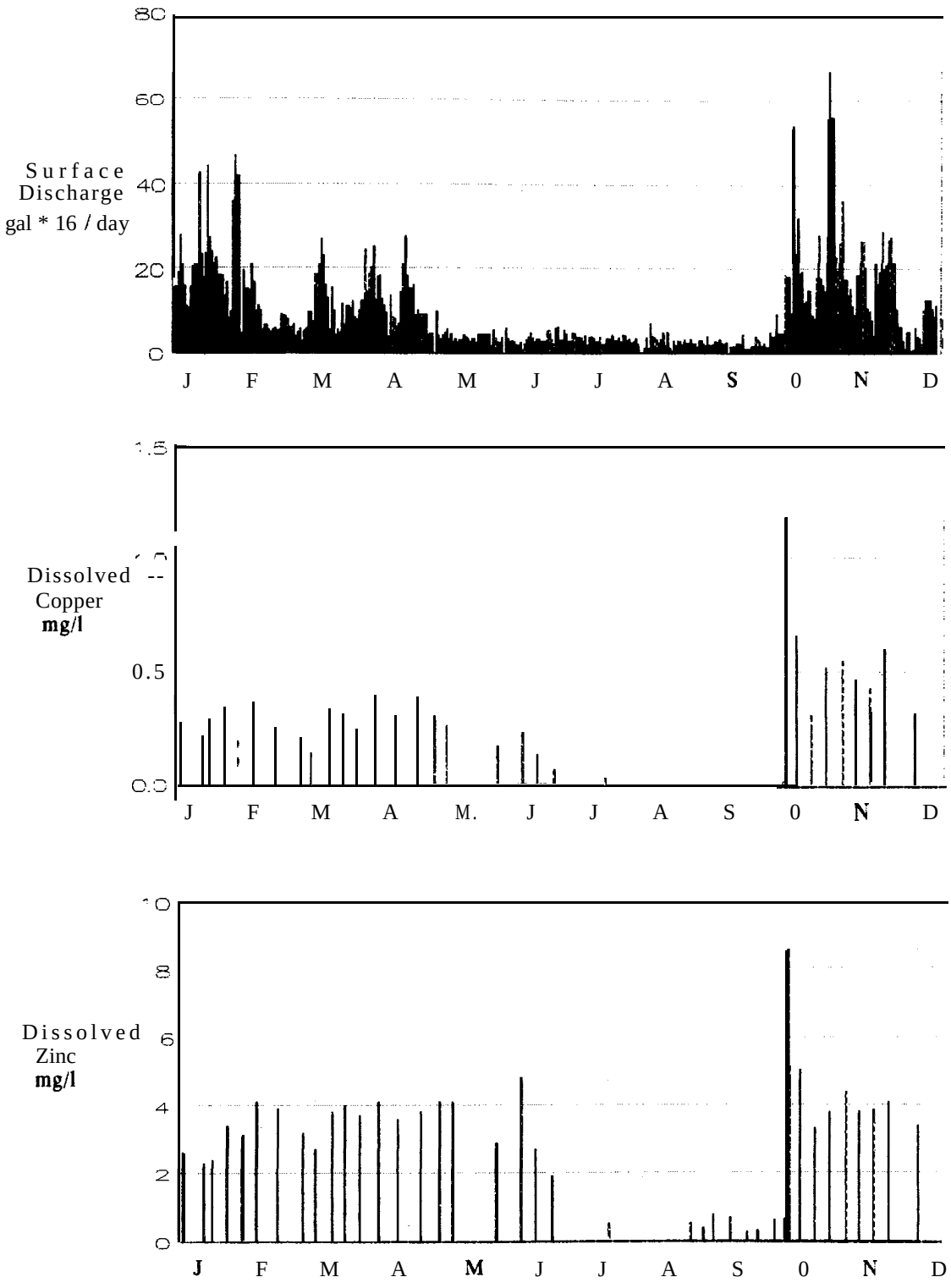
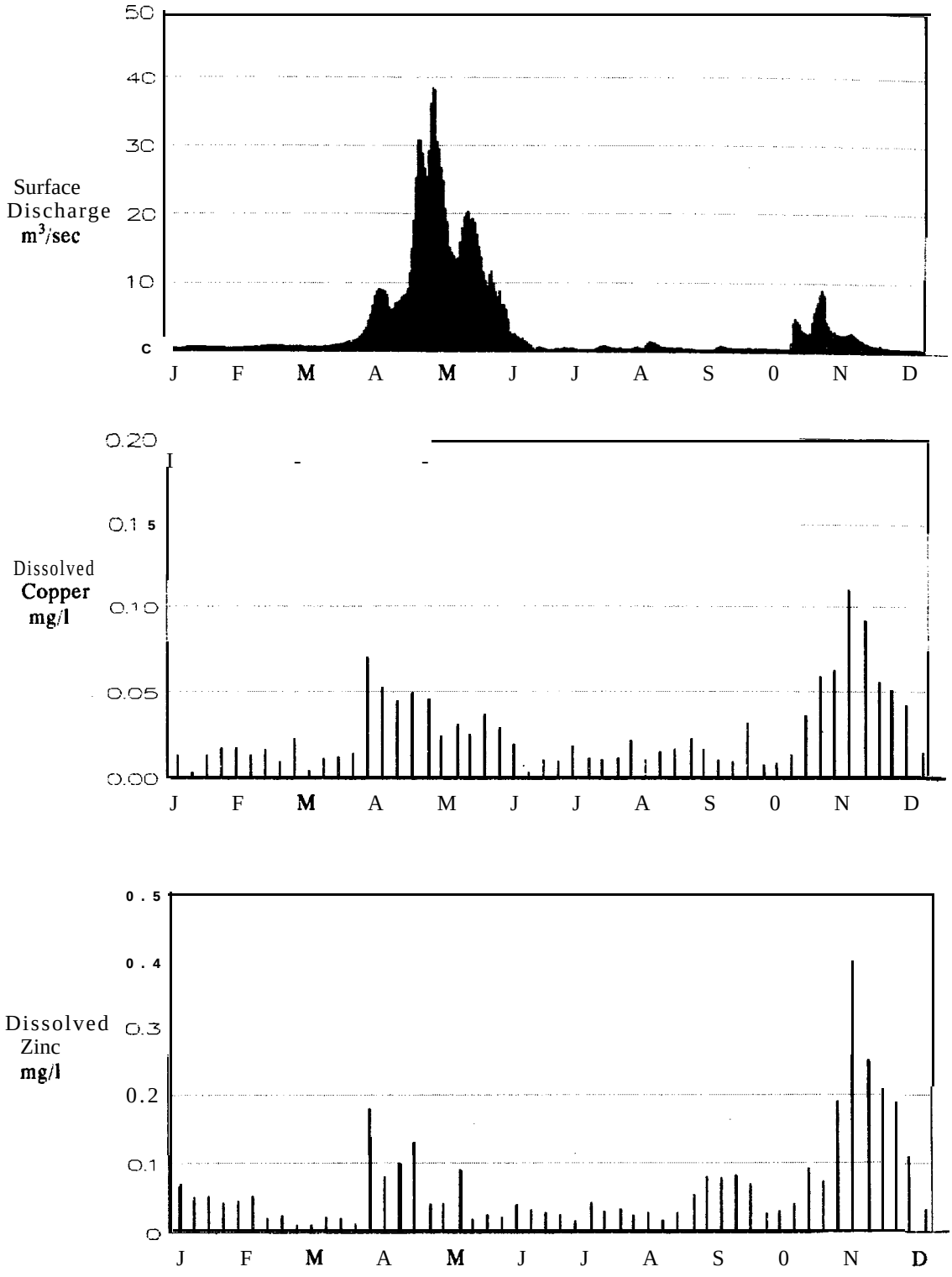


Figure 3 Seasonal Pattern of Flow and Metal Concentrations at an Interior Mine.



A mine site could show an intermediate pattern, where either fall rains or spring rain/melt produces the highest peak of ARD contaminants, depending on the amount of water washing through the rock and the amount of salts that have accumulated since the last thorough rinsing. The pattern of seasonal variation must be determined for each site.

3.23 Background and Baseline Monitoring Sites.

Many mines are in watersheds that contain other low-grade ore bodies and/or abandoned mines which may produce significant background ARD contamination. Weathering of naturally exposed rock usually produces relatively low amounts of ARD because the surfaces have been heavily oxidized and the process has slowed down to a very low rate. However, rock slides and road building can produce new actively generating ARD sites. Old mine sites may actively generate ARD for hundreds of years. Therefore the monitoring for background and baseline studies should include the possibility that ARD is present in the watershed from sources other than the active or proposed mine in question.

Background and baseline studies are often pursued with quarterly sampling or one or two intensive short-term programs. Given the seasonal patterns of ARD release that we observe from mine sites, it is very likely that quarterly or semi-annual sampling of background and baseline sites would miss the presence of significant ARD.

For example:

- The baseline water quality investigation for a mine located on the North Coast region of B.C. used water samples collected on August 9, 1987 (Godin, '1988) and again on August 9, 1988 (Godin and Chamberlain, 1990). Such limited sampling during the same season in both years reduced the liklihood of revealing natural contamination.

- The pre-operational study for a silver mine in the Nechaco region used samples taken on July 14 and October 13, 1973 and July 26, 1974. These sampling dates miss the snowmelt period that would be expected to carry the highest concentrations of ARD.

- For many years a mine on Vancouver Island monitored the baseline water quality in an upstream creek 4 times a year: March, June, September and December. In 1989 they switched to monthly monitoring and found elevated metal concentrations in October. (See Figure 4.)

The importance of timing the investigation to the seasonal pattern of ARD release has been recognized by several researchers who have adapted their programs accordingly (e.g. the Mount Washington investigations by B. Godin). But there are numerous other examples of Stage 1 and 2 site investigations, research on abandoned mine sites, and background monitoring schedules that were timed in a way that would have missed the

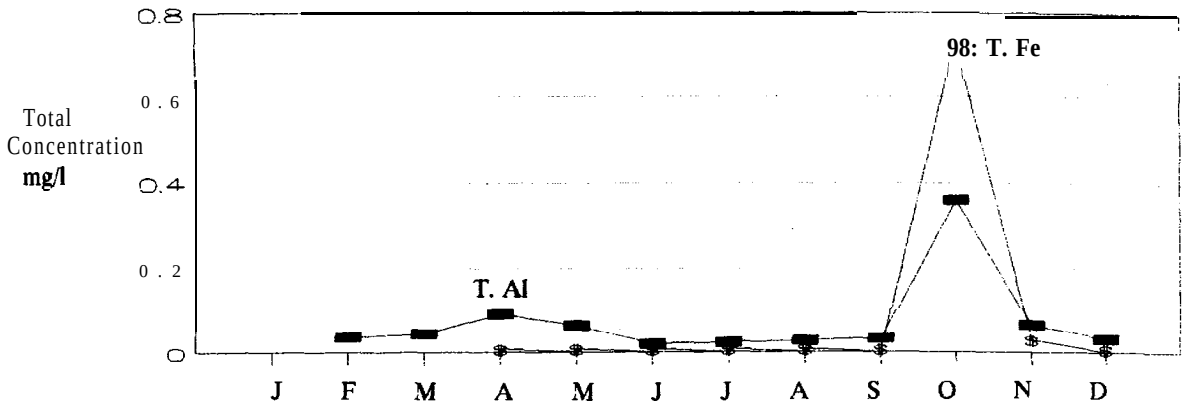


Figure 4 Monthly Background Metal Concentrations in a Vancouver Island Creek.

peak ARD release if it occurred. The method for preliminary studies outlined in Section 2.6.8 would prevent these problems.

Background monitoring should also anticipate seasonal patterns of variation. Although the impact of natural ARD is likely to be very small compared to that of a mine, it is wise to be able to separate the mine's impact from other sources. For instance, road building for forestry development in the same watershed could conceivably expose acid generating rock and produce substantial ARD over time. Natural sources and various man-made sources should be distinguished accurately for optimum management. A preliminary study identifies the optimum times for monitoring background concentrations and/or loads.

3.2.4 Groundwater

ARD contaminated surface water may seep into subsurface drainage before it can be collected for treatment, and may eventually find its way into the water table: consequently there is concern that treatment of surface ARD will not prevent the contamination of groundwater and the transport of ARD to other areas and/or times. In addition, the adit waters (underground tunnel drainage) in sulphur bearing rock are often highly contaminated, and there is concern that some of this water may elude the pumping system and seep back into the groundwater.

Although several mines sample groundwater once or twice a year, there have been no groundwater data sets available with which to explore different monitoring strategies. Presumably the surface water recharging the groundwater will show seasonal ARD patterns of contamination, although the process is complicated by lag effects and

chemical reactions within the ground. Semi-annual or quarterly sampling could easily miss evidence of contamination. Since groundwater contamination could cause serious problems persisting long after a mine is closed, a higher priority should be given to determining whether or not this is a problem at each mine. A preliminary study of variance in groundwater would be simple because there are no cofactors or alternate sampling locations to explore, and short-term variability is likely to be very low.

32.5 The Hysteresis Effect

Despite the observation that acidity and metal concentrations are released in a seasonal pattern that corresponds to the rainfall and surface discharge pattern, there is only a weak correlation between instantaneous concentrations and instantaneous flow. For example, Erickson and Deniseger (1987) found apparently constant copper concentrations at different flows in weekly samples during spring snowmelt at the Mt. Washington site, and concluded that snowmelt did not significantly dilute ARD. Table 2 shows some correlations between daily flow and ARD components at two different mines; these low r values indicate that the relationship between flow and concentration is not linear.

Table 2: Correlations of ARD Components with Daily Flow.

	Mine 'A'	Mine 'C
P H	-.1813	-.2146
Copper	.3564	.3163
Zinc	.2821	.3155
Cadmium	.2860	n/a

The cause of these weak correlations between flow and concentration is that there are different relationships with rising vs. stable or falling flows. Hysteresis describes the cyclic relationship of concentration with flow. The acids and dissolved metals that are generated during low precipitation periods (summer dry months and winter frozen periods) accumulate on the surfaces of the rocks as salts. The small amount of water percolating through the rock pile will be bound to rock surfaces and will be evaporated off rapidly due to the heat of the oxidation process. Thus little water reaches the toe of the pile, and ambient surface drainage will show little impact of ARD. The first rains or melt waters to flow out the toe of the pile carry very heavy loads of these salts. As more water flushes through the rock, more rock surfaces will be washed more thoroughly, and concentrations will increase. If flows are stable for a while, concentrations may remain stable or may begin to decrease as reserves of salts decline. Eventually concentrations begin to decline with steady or rising flows because of declining reserves. At the end of the hysteresis cycle the rock is well rinsed; continued flushing will only carry the ARD being generated at the time. Data from Mine 'A' shows (Figure 5a) a single hysteresis loop during the first week of fall freshet in a coastal mine's drainage ditch; note that the peak concentration occurs before the peak

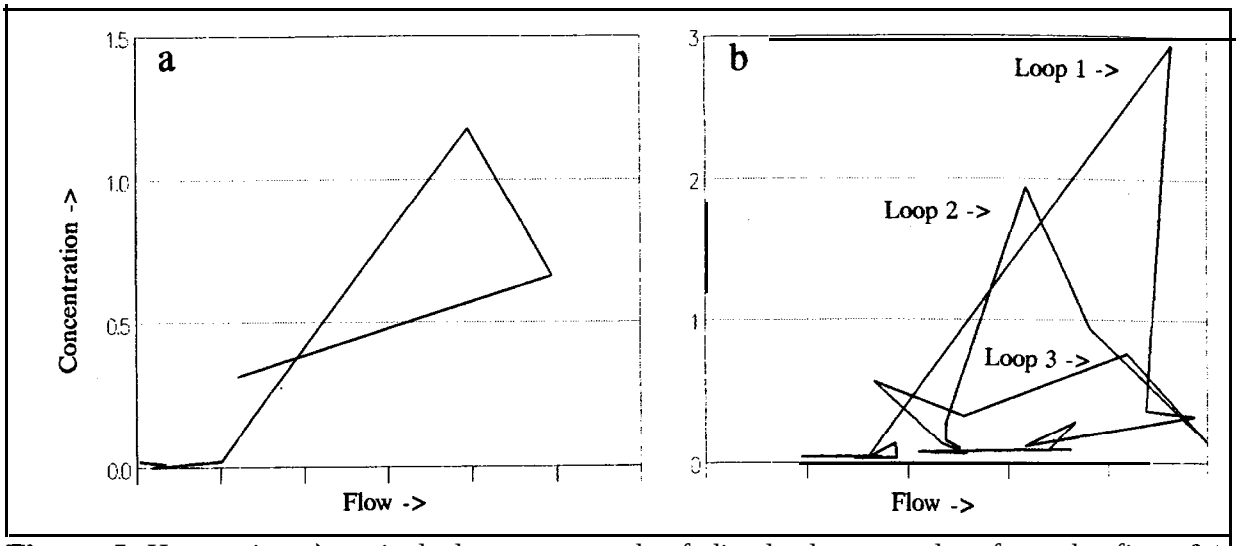


Figure 5. Hysteresis: **a)** a single loop: one week of dissolved copper data from the first fa rain at Mine 'A'; **b)** several diminishing loops: one month of total zinc data from Mine 'C'.

flow and that the concentration did not return to the starting value when flows dropped. A more complex pattern of diminishing loops seen in one month's data from Mine 'C' is typical of the flow vs. concentration relationship in ARD data, and illustrates why the linear correlations calculated from such data are often insignificant.

Hysteresis makes it difficult to 'catch' peak concentrations of ARD, because their timing is determined by the history of recent weather conditions as well as immediate rainfall and flow conditions.

3.2.6 Conclusions

The main loads of ARD contaminants should be expected to wash out during moderate to high flows that follow low flow periods. The seasonal pattern of ARD release is therefore generally predictable for any particular site. A good monitoring program must incorporate this seasonal pattern, and some understanding of hysteresis, in order to detect an ARD problem or background load, and to adequately record the peak releases.

3.3 Seasonal Patterns of ARD Impact on Receiving Waters.

Water quality monitoring cannot be optimized without assessing risks to the receiving environment. Since we know that the release of ARD is likely to be seasonal, and the vulnerabilities of the receiving environment are also seasonal, it is important to anticipate specific impact events that are likely to occur at each site due to these seasonal patterns.

If there are different levels of risk during different seasons, which is probable, then it is important to 'fine tune' the monitoring progam so that the information feedback time and the level of confidence in the monitoring results are appropriate to the risks.

For example, a migratory fish might be present in the receiving environment for a short but critical period of its lifecycle each year, during which sudden changes in water chemistry might cause sublethal impacts. It is appropriate to monitor differently during this seasonal event because the information needs are different; higher confidence levels and faster feedback are needed for good regulatory supervison.

The best monitoring program for an AKD site should include annual information goals (e.g. tracking trends in means and loads) as well as seasonal information goals that have direct bearing on specific environmental risks. It is worth repeating that a good permit should specify the information goals of the monitoring program (Section 1.1.2), which should be based on an assessment of site-specific risks, some of which may be seasonal.

The following discussion of seasonal impacts in different receiving waters is offered to illustrate the points above and to prompt the reader to consider the wide range of possible ARD impacts in the context of seasonal timing.

3.3.1 Streams and Rivers

Streams and rivers are highly vulnerable to sudden high releases of ARD contamination, especially when they have relatively little dilution or buffering capacity; resident organisms have no way to escape exposure. If organisms are stressed due to chronically elevated heavy metal levels, the impact may be compounded by the high and fluctuating concentrations that occur during the peak release season. Streams and 'even rivers can be depopulated by a single acute toxicity event, and it may take a long time for fish and their key food species to recolonize the stream.

In cases where the main ARD release comes with snow melt, and when the minesite melts before or after the rest of the watershed, the impact is exacerbated by lack of dilution. This occurs in Murex Creek and the Tsolum River where the abandoned Mt. Washington minesite melts relatively quickly after lower elevations are free of snow, causing copper concentrations downstream to rise sharply (Erickson and Deniseger, 1987). The site of Mine 'B' melts sooner than the surrounding watershed due to disturbance from operations and heat generated by the oxidation process in the waste rock; the creek receives this runoff while the rest of its watershed is in winter low flow conditions (Patterson, 1989). In both cases the lack of dilution during peak ARD release probably amplifies the environmental stress caused by the contaminant load.

One special concern with ARD in streams and rivers frequented by fish is that the seasonal releases of ARD may coincide with crucial salmonid life stages. Upstream migration of spawning adults during fall freshet might be affected by a change in acidity

or metals concentrations. Fry emerging from gravel during spring freshet may also be very vulnerable.

3.3.2 Lakes

Lakes dilute, buffer and damp the seasonal input from contaminated streams, but then different seasonal patterns affecting the contaminant load may arise due to physical and biological events in the lake. Basic limnology must be considered in determining the validity of comparisons between sites in lakes or between lakes. Thermal stratification, turnover events and flushing rates should be expected to affect the concentrations and distribution of ARD contamination in the water column. ARD components measured in the water column may be of less consequence than the contamination taken up by aquatic organisms during different stages of their lives. Seasonal cycles of plant and animal growth and tissue breakdown can be important factors in the uptake, release and recycling of contaminants. The different chemical species (i.e. dissolved, extractable and total) of each contaminant are not equally involved in the biological cycles of the lake, and should therefore be monitored separately in the case of high risk contaminants.

Any water quality data from a lake must be taken to represent only a small piece of a very complex total picture, in which seasonal changes contribute one dimension of change. Water quality comparisons between lakes, a problematic task at best, should be limited to data taken in the same season.

3.3.3 Marine Water

ARD contamination in saltwater is so overwhelmingly diluted and buffered that it is extremely unlikely that seasonal changes in marine water chemistry due to ARD would be observed outside the immediate zone of influence. Ion concentrations too low or variable to be monitored reliably through water quality sampling may still be sufficient to cause impacts on plants and animals. (Although submersion in seawater is considered to be the most successful means of halting acid generation in tailings and waste rock, metal ions mobilized in freshwater may remain biologically active in seawater.) Biological monitoring, which is inherently seasonal, is a more efficient means of observing ARD impact in marine water than water sampling.

4.0 MONITORING TREATED EFFLUENT

From a statistical point of view, the monitoring of treated ARD water should be much like the monitoring of any industrial effluent. The variations that occur are determined primarily by the water treatment process and not by natural forces. When the treatment is closely controlled and operating smoothly, effluent water can have metal levels even lower than the water in background streams. But the treatment is not always consistent, and the quality of effluent water may be highly variable.

Permits issued under the Waste Management Act require effluent water to be sampled at fixed intervals, usually monthly, to demonstrate compliance to the levels of acid and metals specified. As discussed above, monthly monitoring assumes that the quantities monitored are accurately represented by a single sample (i.e. little instantaneous or short-term temporal variation), change only slowly (i.e. are highly autocorrelated over time), and are not affected by seasonal influences. In other words, they should behave like other industrial effluents. Two factors may interfere with consistent performance of the treatment system: high variations in flow, and variations in the concentrations of acid and metal in the incoming water. If the treatment is not adjusted to accommodate these variations, the effluent water quality will reflect them. How well does monthly monitoring detect such problems when they occur?

The following exercise is an illustration of ways in which existing data can be explored statistically to evaluate monitoring programs. The data set used was the best available, having almost daily records for several variables. Much more could be done with a proper preliminary data set. Unfortunately there were no other data sets from B.C. mines with sufficient data to permit additional examples.

4.1 An Example -- Mine 'C'.

Mine 'C' collects pit dewatering flows and surface drainage from waste dumps, an abandoned open pit, and the entire active area of the mine site. This water is gravity fed to a treatment facility where it is mixed with highly alkaline processing water. At most times of year the processing water dominates, and there is a problem with high pH in the effluent. Their permit requires the pH to be between 6.0 and 11.0. Metal sulphates and carbonates are precipitated as a sludge from the mixed waters; the sludge is disposed of with tailings. Monthly monitoring has shown that the resulting water is generally of very good quality, for effluent; the mine passes quarterly LC, bioassays regularly, acknowledging a bit of a problem keeping the pH below 11.0.

The data used to evaluate this monitoring program (Appendix 2) consisted of almost daily samples (weekends often missing) taken by the mine and analyzed in their environ-

mental lab (pH, dissolved and total zinc and copper), and the almost daily outflow record. Data from the single monthly grab samples, taken for official monitoring and analyzed by an independent laboratory, were compared to the daily data set. The dates of monthly sampling were not randomized but were not fixed either, generally falling in the middle of the month. The daily data for dissolved metals was often below the environmental labs' detection limit, and therefore very incomplete. For total metals the data record for 1988 was the most complete.

4.1.1 Flow Record

The flow record consisted of calculations based on a flow counter that was read daily 5 days a week: weekends or holiday periods were assigned the average flow based on the accumulated count. Figure 6 shows the daily flows as recorded: note the runs of identical flow on days that were assigned average values. The volume of discharge clearly reflects a typical seasonal pattern, with high and variable flows in the fall and winter, and less variation and lower flows in the spring and summer.

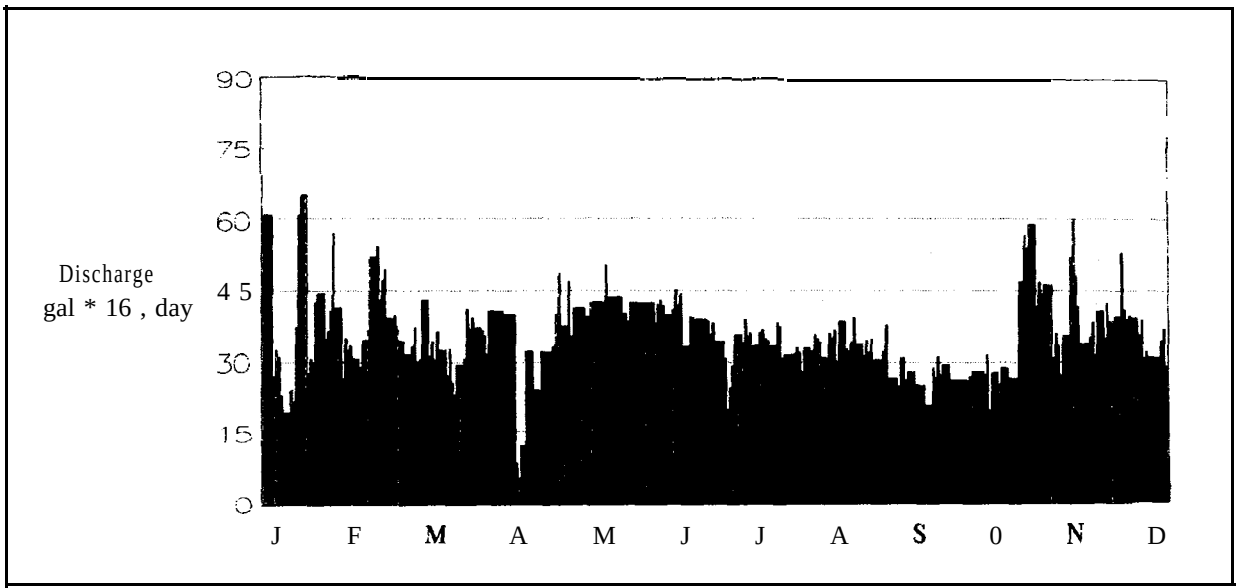


Figure 6 Daily Treated Effluent Volumes Recorded at Mine "C".

Since relationships with flow are crucial to understanding AKD data, the first question to be asked concerned the importance of the information missing from the days with averaged (instead of measured) values for flow. Selecting only days with measured flow and previous day's measured flow (n=217), the daily differences between flows were examined (Figure 7). The mean difference of 145 m^3/day is very small compared to the range, and the distribution of daily differences is very symmetrical and acceptably normal. The standard deviation of 7688 m^3/day is high, The range of observed differ-

ences (-38310 to 23480) shows clearly that day to day differences in flow are common and may be very large. Therefore the use of averaged values does not bias the data, but much information is lost. If loads are to be calculated from the concentrations observed in this effluent, and if concentrations vary from day to day, then daily flow data is necessary for accurate estimates.

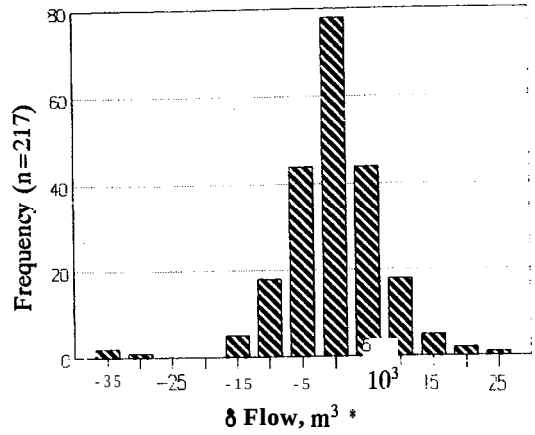


Figure 7 Day-to-Day Flow Changes, Mine 'C'.

4.1.2 pH Values

The pH data is summarized in Table 3. For this variable there is no independent data from the analytical laboratory because pH is not stable; therefore the accuracy of the environmental laboratory's results cannot be determined. The usefulness of accurate pH readings would in any case be limited by the apparently brief time period they represent. The daily record (Figure 8) shows a high frequency of variation in the range of 10 to 12. This series **was** evaluated for autocorrelation using Box-Jenkins time series methods and the findings were that the daily pH levels are independent; less than 10% of the

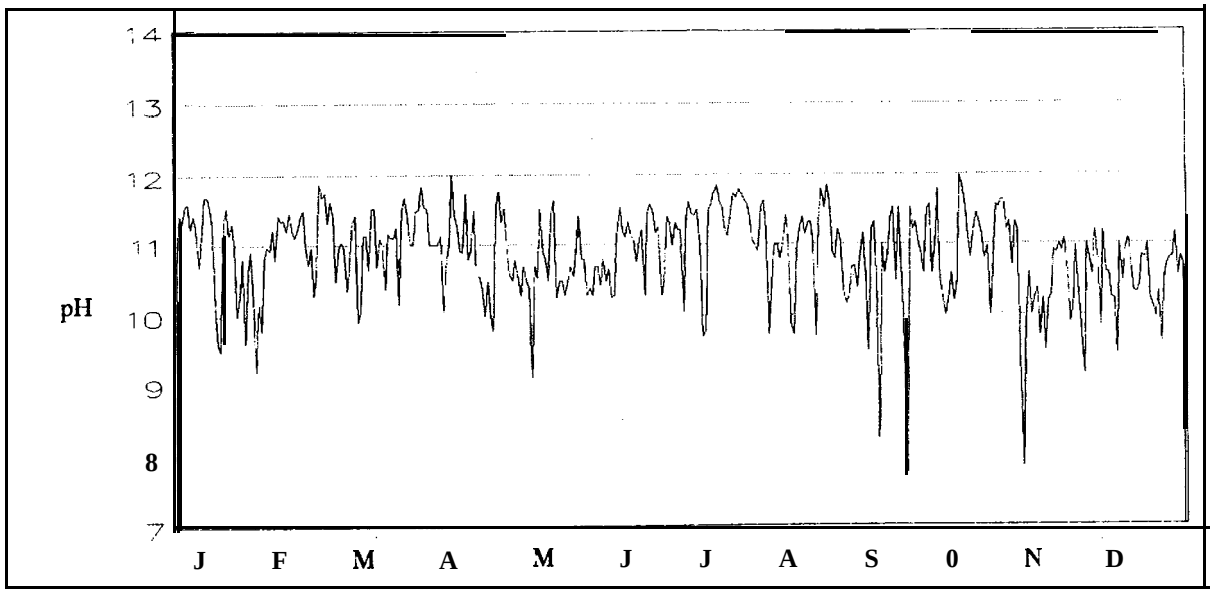


Figure 8 Daily Record of pH in Treated Effluent from Mine 'C'.

differences between daily pH levels can be explained by lag effects. In other words, the pH observed today has no significant relationship to the pH observed yesterday or tomorrow. This result was surprising because the average retention time in the treatment ponds is claimed to be 1 day, and a significant 1 day lag would be expected for outflow from such a system.

Table 3: Comparison of Daily and Monthly pH Data from Mine 'A'.

	Monthly Mean	Monthly s.d.	Single Grab	Monthly Mean vs. Single
Jan	10.82	0.716	11.31	0.49
Feb	11.11	0.444	11.66	0.55
Mar	11.05	0.481	11.13	0.08
Apr	10.97	0.555	10.55	-0.42
May	10.63	0.463	10.71	0.08
Jun	10.96	0.402	11.13	0.17
Jul	11.28	0.546	11.55	0.27
Aug	10.94	0.594	10.96	0.02
Sep	10.66	0.921	11.50	0.84
Oct	11.00	0.557	11.84	0.84
Nov	10.30	0.746	9.17	-1.13
Dec	10.53	0.434	10.34	-0.19
Mean	10.85		10.96	0.11

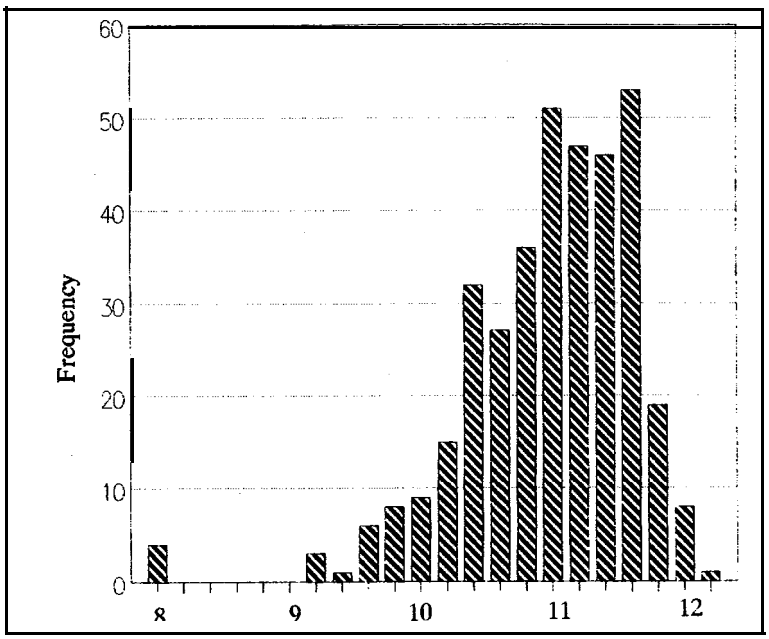


Figure 9 Frequency Distribution of pH Values in Treated Effluent at Mine 'C'.

A frequency distribution graph of the daily pH values (Figure 9) shows a curve skewed to the left. The mean of 10.8 is influenced by 4 values below 8.2, which probably represent episodes of higher volume of ARD during heavy rains.

The monthly grab samples are shown in Figure 10 (dark bars) and compared to monthly averages based on the complete data record (light bars). Three of the 12 monthly grabs were more than one standard deviation from the corresponding monthly means (calculated from daily data); these occurred in February, October and November, when flows are highly variable. Based on daily data, the 95% confidence interval for the true mean ranged from ±1.8 pH units in the most variable month to ±0.8 pH units in the least variable month. This indicates that the 9 grabs that were very close to true monthly means were 'lucky'; an average grab would be expected to have an error of > 1 pH unit.

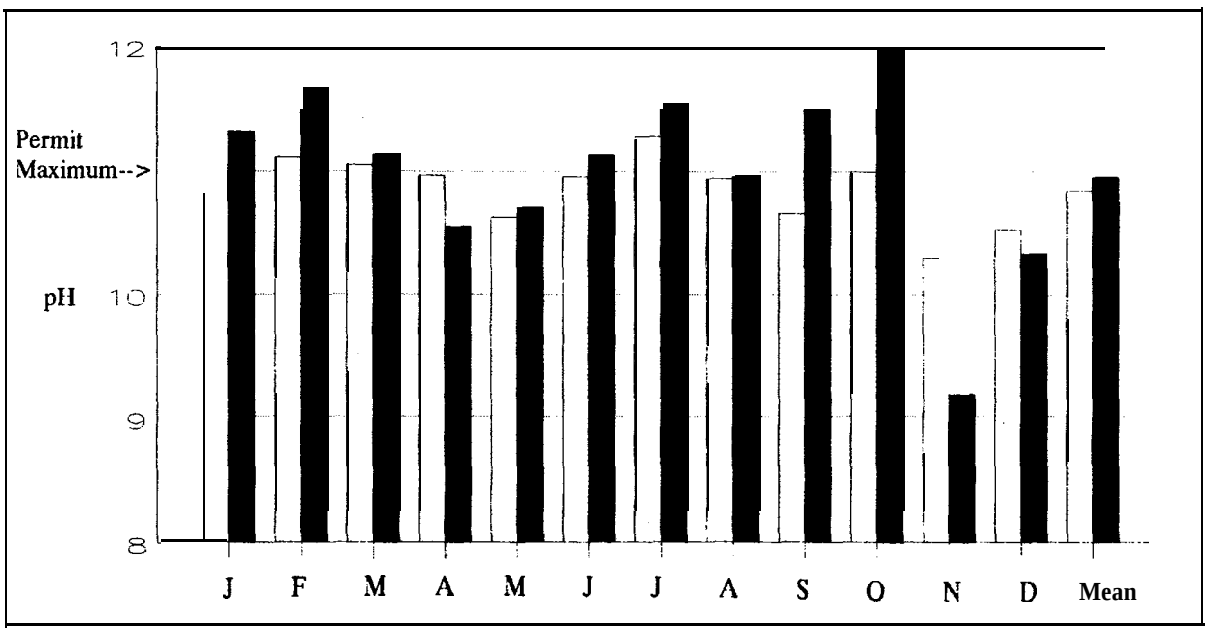


Figure 10 Monthly Means ▒ and Single Monthly Samples ▉ of pH at Mine C.

Using the monthly values to monitor for compliance, we note that in 9 out of 12 months the 'true' mean is ≤11.0, while the monthly grabs were compliant only 5 times in 12. Clearly the mine exceeded its permit value of 11.0 frequently; the difference in distinction between grab and daily data is probably not important.

What is perhaps of greater importance is the fact that the episodes of lower pH are completely missed by the monthly samples. Although the low pH's that occurred during these episodes are compliant, they probably flag very different results in the treatment facility, which may be associated with high levels of other ARD components. Note that these episodes are of very short duration: only single days at pH's below 9.0. In fact, the daily data record is consistent with the possibility that sharp dips in pH occur quite frequently for durations of less than half a day.

4.1.3 Total Zinc

The daily record of total zinc (n=215) is graphed in Figure 11. Here again a high level of daily variation is evident. Time-series analysis could not be done on this fragmented data set, but a regression of daily total zinc concentrations on the previous day's concentration was insignificant ($p<.9$, $r^2 = .23$), meaning that each day's concentration is essentially independent of the previous or following day's concentration. Total zinc concentrations are positively (but weakly) correlated with flow (r=.32) and negatively correlated with pH (r=-.59). The permit value of 1.0 mg/l was exceeded in 5 of the 215 daily samples; these exceedances occurred on days with low pH's and moderate to high flows. These data show a highly skewed non-normal distribution.

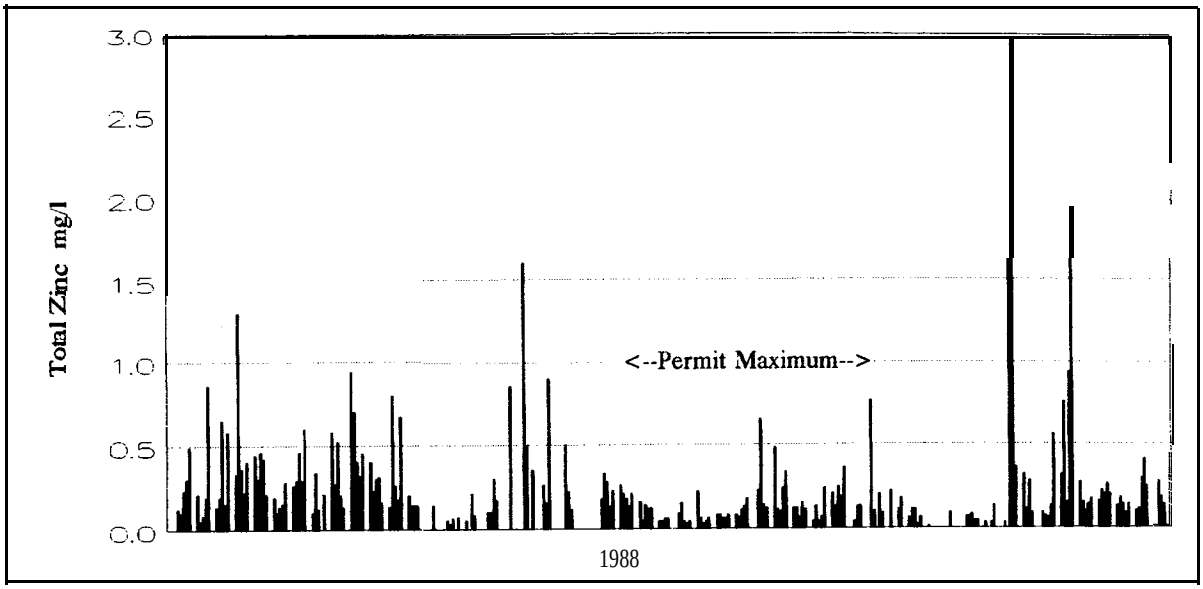


Figure 11 Daily Record of Total Zinc, Mine 'C'.

A check on the mine's analytical accuracy shows that the mine's results are consistently lower than the independent lab's results for the same day (Table 4). This difference averages 0.02 mg/l for the twelve pairs of samples, and is approximately 9% of the mean of the 215 sample set. The daily data must therefore be assumed to be low by about 9% each day. It should be noted, however, that for the highest sample of the twelve pairs (November's sample) the mine recorded a higher value than the independent lab. It is impossible to distinguish between analytical error and instantaneous variation as the cause for this difference.

Monthly average concentrations, based on daily samples, are compared to monthly single grab samples in Figure 12. Most months are close in terms of mg/l differences, but the proportional errors [('true'/grab) - 1] averaged 0.47; ie. each grab was likely to be almost 50% high or low. Because positive and negative errors cancel eachother, the mean of the monthly grabs, 0.36 mg/l, is closer to the adjusted annual mean of the 215 samples, 0.28 mg/l, but still represents and error of 22%.

Table 4: Total Zinc: Monthly Means and Single Samples from Mine 'C'.

	Single Grab	Same Day Envir.	Lab	Diff	Monthly Mean	Adjusted Mean(9%)	Single Grab Adj. Mean
Jan	0.18	0.13		-0.05	0.35	0.38	-0.20
Feb	0.39	0.34		-0.05	0.30	0.32	0.07
Mar	0.42	0.32		-0.10	0.35	0.38	0.04
Apr	0.23	0.21		-0.02	0.11	0.12	0.11
May	0.63	0.50		-0.13	0.55	0.59	0.04
Jun	0.15	0.13		-0.02	0.15	0.17	-0.02
Jul	0.10	0.03		-0.07	0.08	0.09	0.01
Aug	0.12	0.10		-0.02	0.18	0.20	-0.08
Sep	0.22	0.22		0.0	0.17	0.18	0.04
Oct	0.03	co.01		-0.03	0.05	0.06	-0.02
Nov	1.70	1.94		0.24	0.58	0.64	1.06
Dec	0.20	0.18		-0.02	0.19	0.21	-0.01
Annual	0.36			-0.02*	0.26	0.28	0.09

*9% of annual mean of 0.26

November's high concentration was a real event, not an 'outlier'; the mine's record for that day showed a concentration of 1.94 mg/l. However, the mine's record also shows the previous day's concentration at only 0.94, and three days later (after two days with no samples) the concentration had dropped to 0.27 mg/l. The daily record also shows an even greater exceedance three week's earlier: concentrations of 2.04 and 2.94 for two days, flanked by 0.03 and 0.36 mg/l. The year's data shows two other exceedances, in January and May, in which the high concentrations were sustained for only one day, flanked by concentrations well below the permit value.

Since these high values occur during above average flows, the inadequacy of monthly grab samples is even greater when used for calculating monthly or annual loads. Figure 13 compares loads calculated from monthly grab sample concentrations and monthly flows (dark bars) vs. 'true' loads calculated from daily concentration and flow data (light bars). The 'true' loads are of course not exactly accurate: the error due to averaged flows has already been mentioned; estimates for the missing days' zinc data were the averages of the two flanking values. These two estimates of annual load are quite different; the monthly grab data has over-estimated the annual load by almost 1.2 tonnes of total zinc.

Since zinc concentrations were >1.0 in 5 days of the 215 sampled, it seems likely that other exceedances may have occurred during the other 151 days. Such events would have an impact on annual loads.

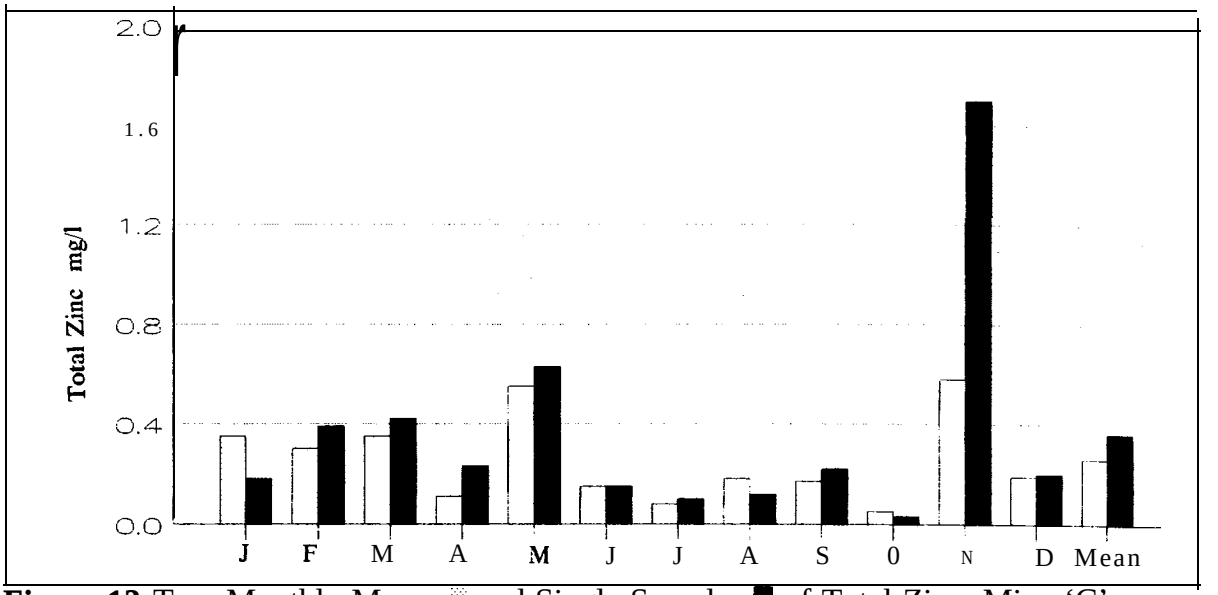


Figure 12 True Monthly Means ▒ and Single Samples ▉ of Total Zinc, Mine 'C'.

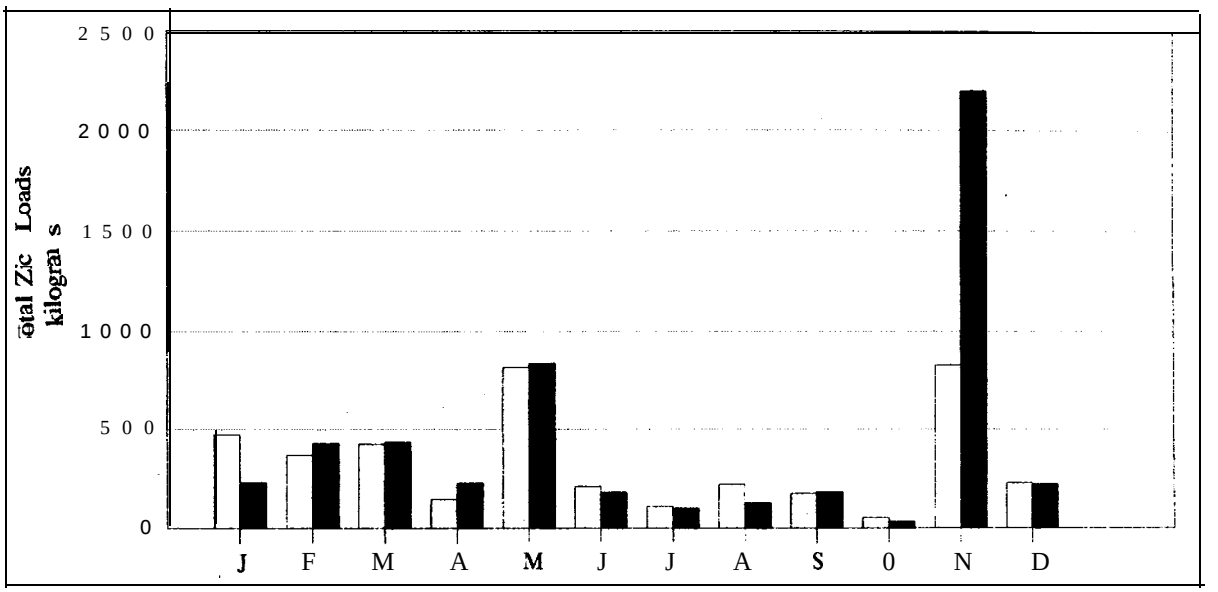


Figure 13 Zinc Loads Based on Mean Concs. ▒ and on Single Samples ▉.

An attempt to estimate the missing daily zinc concentrations, based on a multiple regression of zinc concentration on flow and pH, was not good enough to use (R^2=.38). These relationships are probably not linear, due to the hysteresis effect on rising and falling flows. A more sophisticated model incorporating the hysteresis effect can't be built on the fragmented data available. However the link between high levels of zinc and low pH's is strong enough to invite the following observation: the pH fell below 9.6 on 13 days, 10 of which were sampled for zinc concentration and 3 of which were not.

Of the 10 measured, 5 had zinc concentrations >1.0 mg/l. So the unsampled 3 days may have missed one or two additional peak values. We would conclude that the calculated annual load shown in Figure 13 may be low because of this missing data.

4.1.4 Total Copper

The daily record for total copper (n=212) is graphed in Figure 14. The concentrations are much lower than total zinc, and much more compliant than pH, but the pattern of variation is very similar. As with pH and total zinc, the daily total copper concentrations were found to be independent ($p<.9$ for a lag effect, $r^2=.17$). Daily concentrations of total copper never exceeded the permit value of 0.6 mg/l. Total copper concentrations are positively correlated with flow (r=0.32) and negatively correlated with pH (r=-.56). The mine's analytical accuracy was poor for the total copper samples, which is to be

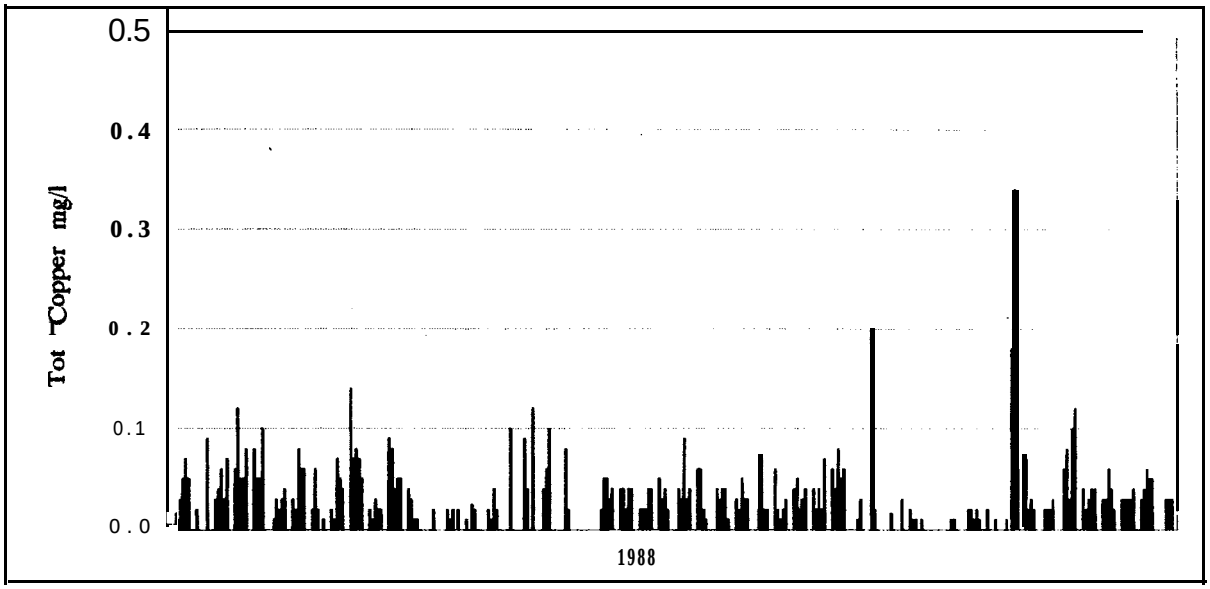


Figure 14 Daily Record of Total Copper, Mine 'C'.

expected in this low range. Comparing the mine's results with those of the independent lab for the same day (Table 5), and accounting for rounding errors, the mine's results are high by an average of 0.016 mg/l, which is approximately 40% of the mean (n=212).

Monthly average concentrations, based on adjusted daily samples, are compared to monthly single grab samples in Figure 15. Very few months have similar mean values, although the differences are much smaller in mg/l than those observed for total zinc. The annual means are almost identical: 0.024 mg/l for the single grab samples and 0.025 for the mean of the adjusted monthly means. Thus, even though the monthly samples taken for zinc overestimated the true mean, the copper samples taken on the same days have given an accurate estimate.

Table 5: Total Copper: Monthly Means and Single Samples, Mine 'C'.

	Single Grab	SameDay Envir.Lab	Accuracy Difference	Monthly Means	Adjusted Mean(37%)	Grab vs.Mean
Jan	0.021	0.03	0.009	0.054	0.034	-0.013
Feb	0.009	0.06	0.051	0.039	0.025	-0.016
Mar	0.055	0.07	0.015	0.047	0.030	0.025
Apr	0.015	0.025	0.010	0.020	0.013	0.002
May	0.052	0.04	-0.012	0.065	0.041	0.011
Jun	0.014	0.03	0.016	0.034	0.022	-0.008
Jul	0.004	0.02	0.016	0.035	0.022	-0.018
Aug	0.000	0.01	0.010	0.037	0.023	-0.023
Sep	0.016	0.016	0.000	0.027	0.017	-0.001
Oct	0.000	0.01	0.010	0.012	0.008	-0.008
Nov	0.094	0.12	0.026	0.064	0.041	0.053
Dec	0.009	0.03	0.021	0.037	0.023	-0.014
Ann.	0.024		0.014"	0.039	0.025	-0.00 1

*37% of annual mean of .039

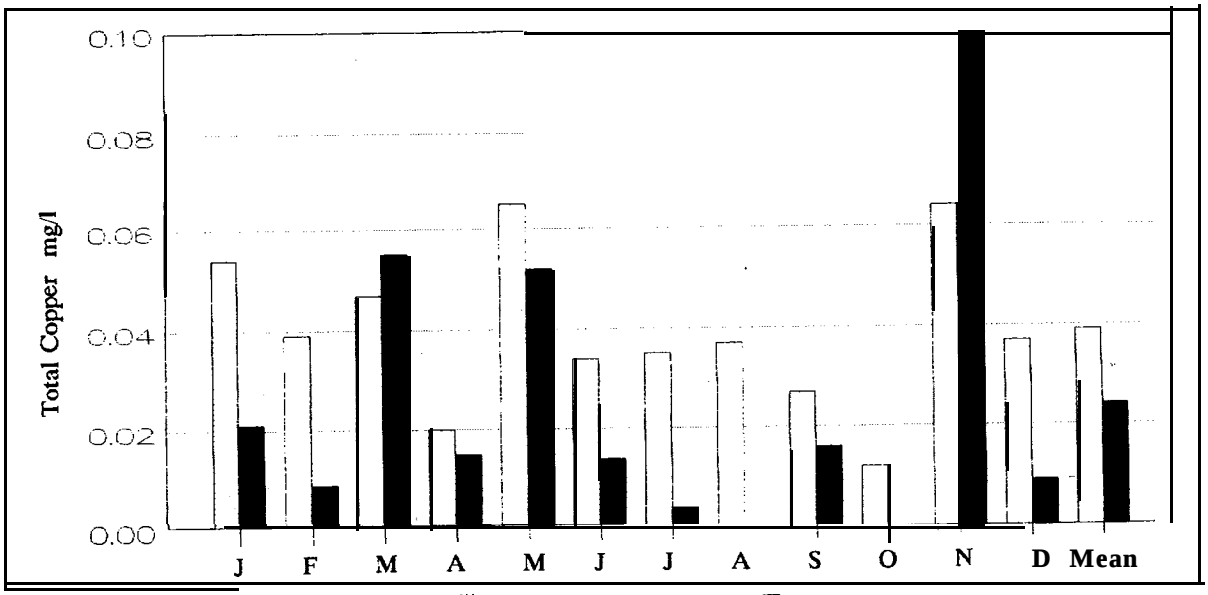


Figure 15 True Monthly Means ▒ and Single Samples ▉ of Total Copper, Mine 'C'.

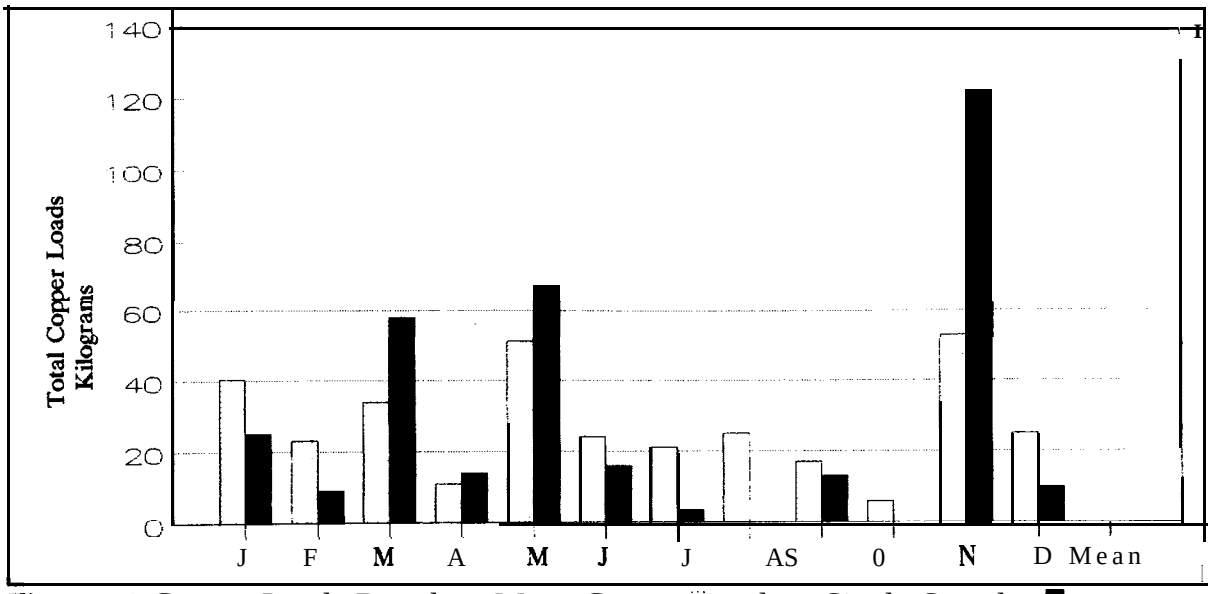


Figure 16 Copper Loads Based on Mean Cones. ▒ and on Single Samples ▉.

Calculated monthly loads based on daily samples versus monthly samples are compared in Figure 16. (Note the low values compared to zinc loads.) The monthly grabs have underestimated the annual load of total copper by only 14 kilograms. If additional peak values of copper discharge were missed by the daily data, as was postulated for zinc, then the 'true' annual load would have been a little higher.

4.1.5 Conclusions

Bearing in mind the weaknesses of the data set used, we can nevertheless conclude:

Analytical Accuracy: The differences between the mine's environmental lab and the independent lab, while substantial, were less than the day-to-day variations in metal concentration.

Single Samples as Estimators of Monthly Mean Values: Daily values for pH, zinc and copper were found to be independent and appear to change rapidly, making a single sample a weak indicator of monthly means. Monthly single grab samples for pH had 95% confidence limits >1 pH unit for most months. For total zinc, monthly grab samples had errors averaging ±0.14mg/l (47% of the 'true' monthly means). The errors of monthly grab samples for total copper averaged ±0.016mg/l.

Single Samnles as Estimators of Exceedances: Monthly single grabs are virtually useless for indicating the frequency or magnitude of short-term exceedances that occurred in this data set. The duration of these high values appeared to be less than 24 hours.

4.2 Optimum Sampling of Treated ARD Effluent

The fixed frequency monitoring schedule in Mine 'C's permit is suited to the monitoring of most industrial liquid effluents, where the industrial and treatment processes are tightly controlled. Mine 'C's effluent data appears to represent a poorly controlled treatment process: the rapidly fluctuating pH levels indicate both that the treatment is not being adjusted for varying inputs of acid drainage, and that the mixing and settling time before discharge is less than 24 hours. When too little acid drainage is available, the pH of their effluent is > 11.0, which happens very frequently. When a large pulse of ARD comes into the system, non-compliant concentrations of total zinc are released.

The key to better effluent monitoring and better control of the treatment process is to monitor incoming and outgoing pH closely. Although pH does not correlate tightly with metals, it is a good indicator of the system state. The mine needs incoming pH data in order to adjust the treatment procedure. Improvements in treatment would reduce the variance in the outflow, allowing compliance monitoring to be kept to a minimum.

Given a treatment system that seemed to function in a manner similar to Mine 'C's in the above data, what would be the optimum monitoring strategy? First, of course, it is essential to do a preliminary study, in order to confirm variance patterns. Secondly, it is necessary to specify whether it is peak values, monthly means, or loads that must be accurately detected.

4.2.1 Monitoring for Peak Values

We have seen that peak metal discharges are associated with low pH values, and that the pH can change so rapidly that a daily sequence of measures are virtually independent. Clearly the pH in this system should be monitored continuously as the best simple indicator of effluent quality.

For metal analysis, single water samples should be taken on days when the pH drops below 10.0. In 1988, this sampling program would have required 30 samples and would have 'caught' all pulses of metal in the effluent. (Note: if the system were in better control, neutralizing the high pH values of the effluent more consistently, the pH of incoming acid drainage would be a better signal to initiate sampling of the effluent.)

The exact peak concentration is not easy to measure, since it occurs so briefly. For many ARD components, the low risk associated with very short term exceedances reduces the need to measure them precisely. Unfortunately it is impossible to know at the start of an incident of rising concentration how high it will go or how long it will last. And there are some ARD components that are very dangerous in elevated concentrations. In situations where high values are associated with high risk to the environment, it is necessary to monitor them more closely. The best way to do this is to take hourly water samples during the exceedance episode, and subsequently choose which ones to

analyze, based on the pH and flow records. Lacking an hourly data set, it is impossible to estimate the number of such samples that would be needed to adequately characterize the peak values. Note that the cost of obtaining such information is high and should be assessed with regard to the environmental risk.

4.2.2 Monitoring for Mean Values

Since pH values are so variable in this system, the best and easiest way to get accurate pH means (weekly, monthly, annually) is to monitor continuously and use a data logger to calculate mean values.

For metals, which cannot be monitored continuously, a choice must be made regarding the level of accuracy required. The single monthly samples taken in 1988 gave monthly estimates that averaged 0.56 standard deviations from the 'true' monthly means for total zinc. However, 1988 was actually a 'lucky' year: using the 215 daily zinc measures in a Monte Carlo random resampling simulation, the average error (i.e. 50% likelihood) of single monthly samples is 0.80 standard deviations from the 'true' monthly means.'

In calculating annual means from the monthly samples, positive and negative errors cancel each other, so that calculated annual means should converge toward the 'true' annual means. However, this convergence is biased by the fact that the single monthly samples are 'drawn' from populations of different variances. Using Monte Carlo methods, we find that the standard deviation of annual zinc means calculated from single monthly samples in the 1988 data is ±0.39 mg/l. With a permit value of only 1.00 mg/l, this is a very high level of error to work with.

The corresponding figures for total copper were that the single samples in 1988 averaged 0.61 standard deviations from the 'true' monthly means. The Monte Carlo estimate for the average performance of single monthly samples in calculating the annual mean is ±.026 mg/l, which is only 4% of the permit level of 0.60 mg/l. The error is smaller for copper because its variation is lower; it is easier to monitor accurately.

In general, monitoring programs for mean values should be based on the most variable metal, which for Mine 'C' is zinc. (The exception to this rule is when the receiving environment is relatively insensitive to the most variable contaminant, in which case the program should be based on the most variable of the high risk contaminants.)

The accuracy of the total zinc monitoring can best be improved by reallocating the sampling effort according to the pattern of variation. It is a basic principle in statistics to

---

[2] If the data were normally distributed, the average error of single monthly samples would be expected to be 0.68 standard deviations, corresponding to 50% of the area under the normal curve. The higher finding in this case is caused by the non-normal distribution of the data and illustrates the wisdom of choosing the non-parametric Monte Carlo method.

allocate samples so that subunits are sampled with equal and adequate efficiency; variation in efficiency of sampling from time unit to time unit can seriously bias the results (Green, 1979), as we have seen in the total zinc data from Mine 'C'.

One way of selecting appropriate time strata is shown in Table 6. There are 8 months with standard deviations <.2, three with standard deviations between .2 and .5, and one with a standard deviation of .8. If there were no further information about the variation, these values could be used to define three time strata. There *is* additional information in the climatological record for 1985-1989, which indicates that October in 1988 was unusually dry. Additionally, we know that fall rains (Oct & Nov) and spring runoff (Apr & May) are times when peak ARD flows occur.

Table 6: Information for Stratification

	Zinc s.d.	Precip. 1988	Precip. 85-89	Season
Jan	0.306	50.5	59.1	Winter
Feb	0.154	23.7	51.4	Winter
Mar	0.233	33.2	45.5	Winter
Apr	0.079	41.7	22.0	Snowmelt
May	0.461	21.5	17.2	Snowmelt
Jun	0.085	8.0	9.9	Summer
Jul	0.051	2.2	3.7	Summer
Aug	0.147	2.8	2.2	Summer
Sep	0.175	11.3	16.7	Summer
Oct	0.037	5.3	26.7	Fall Freshet
Nov	0.799	80.5	66.0	Fall Freshet
Dec	0.079	58.1	40.0	Winter

Three strata are proposed, consisting of: 1. summer months (Jun, Jul, Aug & Sept) as the low variation strata; 2. winter months (Dec, Jan, Feb and Mar) of moderate variation; and 3. the freshet and snow melt months of Apr, May, Oct and Nov with the highest variation.

Monte Carlo techniques have been used to estimate the standard deviations that would be found with increasing sample sizes within these seasonal strata (Figure 17). The current sampling frequency corresponds to 4 (once each month for 4 months) in this graph. At this frequency, the standard deviation of observed means (Le. mean of the four samples) around the 'true' mean (based on available daily data) is ±.063 mg/l for the summer months, ±.091 mg/l for the winter months, and ±.283 mg/l for the freshet months. Therefore the sampling efficiency during in the summer and winter months is 3 to 4 times greater than the sampling efficiency of the freshet months.

To sample with equal efficiency in all strata, sample numbers corresponding to the intersection of one horizontal line on the graph should be chosen. For instance, a standard deviation of ±.135 mg/l achieved by sampling 14 times during the freshet months is close to the standard deviation of sampling 3 times during the winter months, and (±.128 mg/l), and once or twice during the summer (±.165 or ±.086). Using this allocation of sampling effort (14 freshet, 3 winter and 1 summer) to calculate an annual mean would have an expected standard deviation of ±.15 mg/l. Thus by increasing the

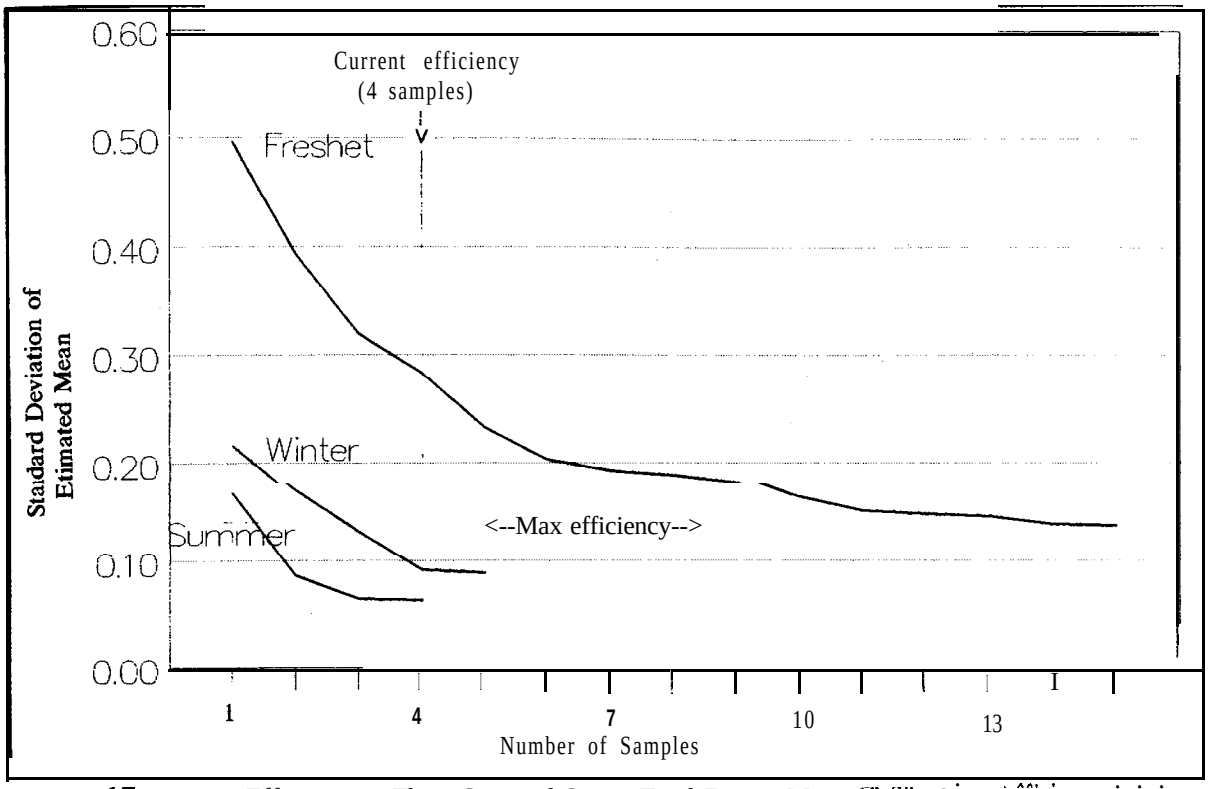


Figure 17 Sampling Efficiency in Three Seasonal Strata, Total Zinc at Mine 'C'. The best efficiency for the Freshet stratum (14 samples) corresponds to 1+ samples for the Summer stratum and ~3 for the Winter stratum.

official monitoring of this system by only 6 samples per year, the standard deviation of the estimates of annual means can be more than halved, from ±0.39 mg/l to ±.15 mg/l.

This is still 15% of the permit value, and may be judged to be too wide a confidence interval. However it is clear that the curve for improvements in efficiency with higher sample numbers has levelled out for the freshet months; increased sampling frequency will not narrow the confidence interval significantly. Also note that additional sampling during the winter and summer seasons will increase the accuracy of these seasonal estimates somewhat, but would not improve the estimate of the annual mean. A second summer sample *is* recommended, in order to have minimal replication.

The timing of these samples within time strata should be random, although the finding that the day-to-day records are virtually independent reduces the sensitivity of the design to randomness. There is no virtue in spacing samples evenly across a time stratum. A schedule for sampling according to this design might look like this:

Summer	Winter	Freshet
July 18	Jan 30	Apr 17, 18, 25
Sept 8	Dec 10, 20	May 4, 6, 12, 28
		Oct 10, 12, 17
		Nov 1, 19, 26, 29

Such a schedule should be established before the start of each year. If weekends represent additional expense or delays in analysis, then days can be randomly selected from weekdays only (assuming that there are no operational differences in the mine on weekends that would thus be missed). Once the schedule is set it should be adhered to. A different random schedule should be established each year. While the results of this monitoring would constitute the 'official record', it would additionally be very valuable for the mine to continue monitoring more frequently for its own information.

The last step is to be sure that all parties with an interest in the monitoring program are aware of the expected accuracy of the results and are comfortable with it.

The statistical methods used in the exercise above are not the only means of optimizing sampling effort. For instance, Monte Carlo methods used to define the sampling efficiency curves (Figure 17) were used in this case because proper preliminary data was not available. When a good preliminary study has been done, and assuming that the data distributions within strata are acceptably normal, the preferred method of choosing sample sizes is the standard parametric sample size calculation (see, for example, Sokal and Rohlf, 1979).

The method of stratification chosen for the example is also not the only method. For instance, if daily flow events were the driving force behind variation differences, it would be more efficient to stratify based on daily flow levels. Thus, instead of using broad calendar month groupings to define seasons, the actual flow record could be watched, and freshets and snowmelt periods could be identified as they occurred. The choice of stratification method should consider whether the advantage in sampling efficiency warrents the additional man-hours or equipment needed for operation. In any case, the the stratification scheme cannot be expected to 'fit' the 'real' variance pattern of a site very accurately if the preliminary data on which it is based is weak.

To recap the basic steps in designing this monitoring strategy:

- Examine the preliminary study data or fullest data record to determine appropriate strata (in this case time-strata [seasons] were used, but other data might be stratified according to flow levels or pH).
- Determine the sampling efficiency in each stratum (e.g. Figure 17).
- Allocate samples to strata in such a way as to give approximately equal accuracy in each one. Estimate the overall accuracy of the monitoring program.
- Discuss the accuracy level with all interested parties, so that the limitations of the program are understood in advance.

4.2.3 Monitoring for Accurate Loads

The error of a calculated load is greater than the errors of the flow and concentration measurements on which it is based. The comments made above regarding estimating mean concentration apply to the concentration data used to calculate load: a stratification scheme that yields more accurate concentration estimates will yield more accurate load estimates. Similarly, flows need to be measured, not estimated from watershed size or annual rainfall. The more frequently flow is monitored, and the more accurately the corresponding concentrations can be assigned, the more accurate the load estimate will be. Calculations of load should always include a calculation of standard error.

One important source of error in estimating ARD loads is short term peak concentrations of metals, which are associated with above average flows. While these peak values may be of very short duration, they can contribute very substantially to the load. Thus we have three challenges in accurately estimating loads:

- Accurate measurements of flow,
- Accurate estimates of concentration, and
- Assessing the contribution of short term peak values.

The accuracy of Mine 'C's flow record cannot be assessed because there is no independent information. The standard deviations given in the following discussion are based on concentration data only, and would be larger if the errors in flow data were known. In any case, continuous (as opposed to cumulative) monitoring of discharge volumes would be a definite asset to any load monitoring program.

As discussed above, the seasonal estimate of mean zinc concentrations for Mine 'C' cannot be improved much beyond the level achieved with 14 samples during the 4 freshet months. Using Monte Carlo random resampling from the 1988 data, we find that the standard deviation of loads calculated from random monthly single samples (*i.e.* the current monitoring schedule) is ±2712 kg. This can be improved substantially to ±1043 kg when seasonal means are multiplied by accumulated seasonal discharges. Such an estimate will always be biased low because it will not accurately reflect the contribution of short term peak values.

In the daily 1988 data from Mine 'C', 12% of the estimated annual load was contributed during short term episodes of high values: 5 days in which concentrations were > 1 .O mg/l. Obviously it is important to measure the concentrations and discharges during these peak occurrences. The best way to monitor for peak values is to monitor pH continuously, as described in Section 4.2.1, taking a water sample when the pH drops below a threshold value. The standard deviation of the annual mean load can be reduced to ±663 kg when peak values are sampled separately.

The instantaneous concentration can change very rapidly during an episode of peak release. Due to the hysteresis effect, the maximum concentration usually occurs while

the flow rate is increasing. For variables that cannot be monitored continuously, catching the 'average' concentration that should be attributed to a short term peak is an elusive task. The number of samples taken during an excursion should be determined by considering the associated risk. For many ARD components there may be little environmental risk associated with the short term peak values that occur at a specific site; therefore a single sample for each episode of high values probably gives adequate accuracy. (The actual accuracy thus achieved cannot be estimated without reference to an hourly data set.)

The case would be different for the most toxic ARD components, or any ion for which the peak concentrations reached and the durations of the episodes could have severe impact on fauna. The amount of information necessary (i.e. frequency of sampling) should be determined by assessing environmental risk. As mentioned in section 4.2.1., the greatest accuracy is obtainable by collecting water samples hourly during such an episode, and choosing which and how many to analyze afterwards, based on continuous pH or flow records.

In systems that have extreme variability in effluent flows, and where concentrations of ARD components are strongly correlated with flow, the accuracy of load estimates can also be improved by measuring concentrations on peak flow days. The increased accuracy obtainable this way in the Mine 'C' data set is small because the correlation between total zinc and flow is weak.

Three options for load monitoring are compared in Table 7 below; flows have been assumed to be measured without error in this exercise done with Mine 'C's daily total zinc data.

Table 7: Sampling Options for Monitoring Zinc Loads at Mine 'C'.

Options	Num. of Samples	Standard Deviation	Bias of the Mean
1. Monthly single grab samples	12	±2,712	low
2. Seasonal means	18	±1,043	low
3. Seasonal means plus daily peak data	48	± 663	none
4. As above plus daily data during high flows	66	± 639	none

4.3 Bioassays

Regardless of the accuracy and completeness of **ARD** effluent water quality monitoring records, it is often impossible to base defensible decisions regarding management and regulation on them. This is because the limit values in permits do not relate directly to environmental cause and effect relationships. Such a link will probably remain elusive for many decades to come. In the meantime, bioassays provide an important 'reality check' to the problem of monitoring ARD effluent for environmental hazards.

The mines whose data were reviewed for this study are required to conduct quarterly static LC, tests on undiluted effluent. The mines pass these bioassay tests almost without exception. These results indicate that the ARD treatment methods are working at least to the extent that samples of the undiluted effluent are not acutely lethal to fish.

In general, quarterly static LC_{50} tests are best suited to monitoring industrial effluent which is the result of closely controlled processes. We have seen in the data record for Mine 'C' that high day-to-day variability in water quality is a feature of this effluent. If water for the tests is collected in a very short period of time, it will be relatively uniform, and the static LC_{50} conducted with this water will not reflect the stress of frequent and rapid changes in water chemistry. Flow-through LC_{50}'s conducted with water samples collected in a short time period also will not reflect the effects of the high day-to-day variability of this water.

Two other pollution stresses not evaluated by LC_{50}'s are the effects of infrequent episodes of short term high concentrations of dissolved metals or pH changes, and the chronic effects (over years and generations) of small elevations in heavy metal levels.

Other sorts of bioassays would probably give a more rigorous evaluation of ARD effluent. The use of *in situ* bioassays (e.g. fish cages) or streamside 'on line' flow-through bioassays would be much more relevant to the real impact of the effluent on the environment.

As noted in Section 3, the timing of peak ARD discharges can coordinate with critical salmonid life stages. It would be extremely valuable to conduct *in situ* bioassays with the appropriate species and life stages of fish in order to evaluate the permit levels now being used.

In summary, the LC_{50}'s currently used to monitor ARD effluents are evaluating short term acute lethality, but they are not testing the effects of the particular sorts of stresses that can result from highly variable ARD discharges: high frequency of water chemistry changes, short term episodes of high concentrations, and chronic effects.

5.0 MONITORING UNTREATED SURFACE WATER, SEEPS AND GROUNDWATER

5.1 At the Mine Site

Surface water and seeps should never be assumed to be 'clean' unless they have been monitored during the seasons and flow levels that correspond to peak ARD releases. In the absence of other indications of a problem, such as staining or elevated metals at a downstream station that is not otherwise accounted for, it should be adequate to monitor surface water and seeps only at the peak times. The resulting single annual figure does not represent an annual mean, but rather the highest concentration. Decisions regarding interception and treatment of contaminated water at the mine site generally should be based on the peak value and not the mean.

Mines could pursue this monitoring on their own, with analyses done in their own labs, even though their accuracy at very low levels is poor. The purpose is to be able to sample frequently enough to flag *high* values. A lot of 'noise' around the detection limit is no problem in this case. Analytical error is easy to determine via comparison with independent laboratory analyses, and it is likely to be smaller than day to day differences observed for elevated variables.

Groundwater may not show any of the seasonal patterns we see in surface water, or it may show them with a lag. In any case, annual or semi-annual monitoring cannot be considered adequate to prove that there is no problem with groundwater contamination. A preliminary study would be needed in order to estimate the autocorrelation in the groundwater data, Without one, it is impossible to say what frequency of monitoring would be appropriate.

5.2 Background and Contiguous Watersheds

For background monitoring of upstream water, unimpacted areas of the mine's watershed, and other watersheds that drain into the same receiving water, annual monitoring timed to coincide with the peak ARD release will give the most valuable indication of whether an ARD problem exists or not. If evidence of ARD is found, accurate estimates of mean concentration or estimates of load may be needed. These waters should then be sampled on a time-stratified schedule (such as the one devised for Mine 'C's effluent), and seasonal means calculated from these samples. This represents a significant increase in monitoring, but it needs to be done only when annual peak monitoring has shown that contamination exists. It is in the mine's interest to have accurate information on the magnitude of other sources of contamination to downstream water bodies.

Again, there is no reason to send all these samples to an independent lab unless the confidence limits of the mine's lab results exceed the needed confidence, or overlap an important threshold value.

6.0 MONITORING THE RECEIVING ENVIRONMENT -- WATER QUALITY

6.1 Streams and Rivers.

Monitoring streams and rivers has much in common with monitoring treated effluent: exceedances and lesser variations may occur in short time intervals. Fixed-frequency, exceedance-driven or Markovian sampling are not suited to the very high variability that may occur with ARD contamination unless the sole purpose of monitoring is to provide crude estimates of mean concentration. The design of a monitoring program must be guided by decisions on the required accuracy for estimates of mean and peak values; the choice of accuracy level should be based on the magnitude of the minimum changes or differences that must be detected. These are management and biological decisions, not statistics.

The most efficient sampling designs are those which are stratified according to the dominant pattern of variation in the data, which for ARD contaminated streams and rivers would be seasonal or flow related. Therefore preliminary studies for sampling design should determine the best stratification scheme for each study site.

When the seasonal pattern of ARD release from the mine site is the dominant variance pattern, and if accurate seasonal means are the identified management goal of the monitoring, receiving streams or rivers should be monitored on a time-stratified schedule (see Section 4.2.2). Other schemes of stratification (e.g. flow levels) may be superior in different situations. Random sampling (i.e. sampling days selected randomly) within strata is necessary to avoid bias.

The virtues of randomly timed sampling within time strata apply to estimates of mean concentrations, not to the detection of exceedences. No program designed to efficiently estimate means will do a good job of surveillance in a rapid flow-thru situation. If the preliminary study shows that ARD components are strongly correlated with **pH** in the stream, a continuous **pH** probe can provide a continuous indicator of concentration changes, and could be used to signal alert conditions. The technology of remote data logging has advanced to the point where such installations can be relatively trouble-free and reliable. The advantage of having a continuous record with a signalling capability is that it gives the mine an early warning of excursions before they are at their worst. In some cases, quick feedback prompting operational adjustments or intervention could prevent an acute toxicity event from occurring. Extra data collected as a result of such signalling should not be combined with routine monitoring data in the calculation of mean values.

There are cases where the concentrations of ARD contaminants rise downstream from the effluent discharge point, despite collection and treatment of surface water from the minesite. This is evidence of contaminated seeps and groundwater entering the stream.

The best way to pinpoint the location of the source is to collect replicated samples along the two sides of the stream during the peak ARD release (for seeps) and/or during summer low flows (for groundwater). The number of replicates needed would depend on the relative magnitude of the increase one wished to be able to detect.

6.2 Lakes

Lake monitoring tends to expand to ever more stations, depths, sampling dates, etc. because the interpretation of water chemistry data can be very elusive (see Section 3.3.2). Good monitoring design requires that managers first identify the nature of the impacts that are likely from ARD, the water quality changes that might preceed or accompany such impacts, the resolution (i.e. smallest magnitude of change) needed, and make careful selections of the chemical species to be targeted. In particular, the manager should consider what s/he would do with data that showed a small increase in mean concentration of some ARD component but without an identifiable environmental response. Water quality data is likely to serve its most useful role as corroborative evidence when changes in living tissue and/or species distribution and abundance are noted. Therefore an economical lake monitoring program seeks evidence for the expected impacts of contamination, which are mostly in the sphere of biological monitoring (see Section 7.1), rather than attempting to infer impact from water quality data. This allows water quality monitoring to be reduced to the estimation of annual and/or peak loads and means of biologically active chemical forms.

For example, a manager concerned about cadmium contamination in a lake could set up a program to monitor the inflow load (dissolved, extractable and total Cd) and the outflow load, as well as sampling the most likely ‘sinks’ of cadmium in the lake: sediments, and aquatic organisms. At the end of the first year of monitoring the manager knows approximately how much cadmium was retained in the lake, what chemical form it is in, and where it is accumulating. This plan for monitoring is efficient because it allows the manager to identify low risk situations, such as lakes with nil accumulation of biologically active forms of Cd. It also alerts the manager to increase biological monitoring if the lake is accumulating more Cd than can be accounted for in the sampled sediments and biota. In other words, it allows the manager to assess the risk to the lake in a simple way that is free of hypotheses about changes over time (which require good baseline data) or relative impact (which require good control data).

There are several ARD threatened lakes in B.C. that have been studied intensively and/or frequently but with inconclusive results for the following reasons: lack of baseline data, lack of controls, insufficient replication, failure to sample appropriate chemical species, failure to sample at the right time of year, failure to sample ‘cause’ data and ‘effect’ data in appropriate times and quantities, etc. Most of these problems were failures of experimental design, and could have been corrected in the proposal stage. The Waste Management Branch should give consideration to retaining the services of a

statistician for experimental design consultation. The following comments are offered as guidelines for the planning of water quality studies in lakes.

Proper baseline data (*i.e.* including estimates of variance) and the monitoring of control lakes are very important references in interpreting the validity of any trends in water quality over time. What appear to be short term trends may turn out to be random variations within the pre-operational variance of the lake. Comparisons made only over time (instead of to control lakes) assume that the mine is the only agent of change affecting the lake. Climate variation, acid precipitation, and random factors could also have caused the observed changes over time. Only proper control and baseline data allow the researcher to distinguish subtle changes [Section 2.0].

Since few lakes have a perfectly matched pristine 'twin available to serve as a control, often the best strategy is to sample several of the most similar lakes available. Cluster analysis or more sophisticated multivariate statistical methods can then be used to indicate the degree of differences between the impacted lake and the multiple controls.

Monitoring of background and uncontaminated drainages that flow into the lake is essential. An impact, if observed, can be more securely 'sourced' to the mine if other sources of contamination are found to be insignificant.

For standard univariate comparisons of concentrations over time, or comparisons between different lakes, samples should be taken only during corresponding seasons (e.g. spring turnover). These samples should always be replicated to estimate the variation of the 'population' from which they are drawn. Since the variation between replicate samples may vary with seasonal influences, the number of samples needed to reach a predetermined level of confidence may also vary with the seasons. Determining the optimum sampling schedule and effort is a simple experimental design problem. A second year of sampling on a lake should never be undertaken without using the first year's data for design. [Time series analysis can make use of unreplicated fixed-frequency samples, if sufficient data (e.g. 10+ years of monthly data) is available, but the 'noisier' the data are, the higher the data requirements become for conclusive results.]

For monitoring impacts based on water quality data, accurately measured input and outflow loads may be more valid than seasonal 'spot' concentrations. An annual contamination load 'budget' (*i.e.* input minus output) can demonstrate that contaminants are accumulating, even if the biological or physical 'sinks' are not identified. The different chemical forms (dissolved, extractable and total) of a contaminant may have very different fates in a lake, and should be monitored separately. In order to optimize the accuracy of the 'budget', the export load should be measured at least as accurately as the inflow load, since the errors on these estimates are additive when calculating their difference. In most cases the concentration variables in the outflow stream will be more

highly autocorrelated and therefore will require fewer samples to achieve the same level of accuracy as the inflow stream(s).

Accurately estimating import and export loads requires a major investment of sampling effort, since both flows and concentrations must be measured frequently. For situations that do not warrent this level of expenditure, taking replicated samples of concentration during turnover can provide reliable and consistent representations of (admitedly limited) information. These data can be used in detecting trends over many (>10) years, and are especially valuable if pre-operational samples were taken, control lakes are also monitored, and appropriate chemical species are monitored.

In summary, lake monitoring should produce data that is consistent (in terms of accuracy) and interpretable with reference to valid comparisons. Enlarging the 'grab bag' of variables (chemical species, sampling times, locations, depths, etc.) usually does not improve the experimental design. Detecting trends and making comparisons require good baseline and background data and good controls.

6.3 Marine

Water influenced by tidal flushing or ocean currents can absorb huge amounts of contamination before elevated concentrations can be detected with water samples. This is just as well, since observing changes in water chemistry may be the 'booby prize' of the investigation:

> ***"Contaminant concentrations in the physical environment*** (for ***example, sediments and water) have in many instances been*** shown ***to bear little relationship to the uptake of contaminants by organisms and to biological effects. Therefore, managers gain only a limited quantity of useful inf or-mation concerning possible ecological effects or seafood contamination from monitoring contaminant concentrations in the physical environment."***
> (Segar, ***et al,*** 1987)

What, then? At the 'New Approaches to Monitoring Aquatic Ecosystems' conference, bioassays were the universal monitoring recommendation. The same design features emphasized for lakes apply to marine investigations: an adequately described baseline and proper controls. (Data on background is much harder to get because of the vast number of other possible sources.)

7.0 MONITORING THE RECEIVING ENVIRONMENT -- BIOLOGICAL

The challenge in designing good biological monitoring is to avoid the temptation to 'look for changes' in an open-ended fashion; a change can only be interpreted as evidence of contamination by comparison with a proper control or adequate baseline data, both of which are often lacking. Natural variation from site to site and from year to year is a major 'noise" factor in determining impact; it is unwise to interpret any change as being caused by any independent factor unless data from a control site (preferably several) is available for comparison. For example, the absence of metal sensitive species in a single year's study cannot be accepted as proof of metal pollution.

Biological monitoring that depends on relative abundance of species, such as diversity indexes, yield information only to the expert and experienced taxonomist, and are very subject to misinterpretation: a change may have occurred over time, but not necessarily because of heavy metal contamination. For instance, the removal of forest cover from long stretches of a stream due to **minesite** development will alter the temperature regime and siltation, and therefore the biota.

Tissue levels of metals provide a much more direct 'reading' of bioavailable metals, although there may be large differences between species and life stages; different organisms pick up and metabolize metals differently. The data will have more internal consistency if sampling is well replicated, and if carefully selected bioindicator species or tissues are used. Bioassay methods that can be interpreted directly, such as hepatic metallothionein as an indicator of zinc, copper and cadmium exposure, are the most likely methods of producing 'defensible' numbers.

In the marine environment where water quality and planktonic populations tend to be ephemeral, benthic sampling is more likely to give clear results. There has been great progress in the identification of 'sentinel organisms', tissue analyses, and strategies that serve to standardize results and improve their interpretability. (Segar, et ***al.***, 1987)

There are many sorts of bioassays available and it is not the purpose of this report to evaluate them. However, there are **some** statistical points to remember in planning bioassay monitoring and interpreting the results:

- It is not valid to compare results with pristine habitats, unless the pristine character of the pre-mine water body was well documented and shown to have been similar. Usually the only fair comparisons are with its own past record or with ***similar*** unimpacted habitats in the same (or near by) watershed(s). Using ***several*** controls can compensate for the lack of a perfect match, and is an excellent way to evaluate the relative condition of a water body.

. It is not legitimate to compare results with 'provincial averages' which were not random samples and are very likely to be biased. If there were a random sample available for comparison, the comparison should be made using a method that incorporates the variances of the two samples.

▸ Caging fish that normally are free to select their own microhabitats may subject them to temperature stress, dietary differences and other factors that would tend to bias the findings. Results of caged (and other contrived sorts of) bioassays should only be compared with data obtained under similar circumstances.

It is interesting to examine the required marine monitoring in the permits issued for Mine 'A' regarding the discharge of tailings into Rupert Inlet. The mine has been monitoring the inlet and adjacent waters in essentially the same way for 19 years. In addition to extensive monitoring of the physical oceanography of the inlet and tailings, they are required to sample 5 stations (4 depths each) quarterly for dissolved metals (80 samples/year); 28 stations 3 times annually for phytoplankton biomass; 16 stations quarterly for periphyton; 16 zooplankton tows semi-annually to 'determine the zooplankton population'; 4 night tows of zooplankton for metals analysis; 72 samples of benthic organisms annually 'to monitor the effect of tailings', and 6 stations annually for crabs, clams and mussels for tissue metal analysis and body condition. The expense of this program is enormous, and the resulting information 'thin'. The 32 zooplankton tows alone probably cost as much to collect and analyze as the entire water monitoring program at many mines, and are incapable of producing a 'defensible' number.

This program was originally designed in an attempt to provide early detection of a problem if one occurred, but without rigorous consideration to the certainty with which small changes could be detected. The water chemistry data, for the reasons mentioned above, is probably of very little value. The taxonomy and phytoplankton abundance measures are subject to enormous natural variation, and are not properly controlled. The metal concentrations in 'grabbed' plankton and benthos are of dubious value because of species and age/exposure differences.

The metal concentrations in bivalve and crab tissues are the most valuable components of the program; this data has the potential to reveal, for instance, that "the bioavailable abundance of contaminant W, as determined by bioindicators X, in area Y, at time T after startup, does (or does not) differ from the baseline mean concentration by more than Z%". Good monitoring programs can be designed using data of this type; they require the manager (or Environmental Audit Committee) to make decisions about the rate and magnitude of change that is necessary and worthwhile to detect.

The other types of data in Mine 'A's monitoring program have much less value for any kind of decision making; they are not 'defensible' in the sense of Section 1.1.3, and risk

producing fuzzy results indefinately. This constitutes an unfair bias in the monitoring, since observed differences may be more apparent than real, and it will almost certainly lead to ever escalating costs for inconclusive results.

8.0 CONCLUSIONS

1. Fixed-frequency single samples, as required in current Waste Management permits, are inaccurate for estimating mean concentrations of ARD contaminants. In the one data set in which their accuracy could be explored, single monthly samples were found to have an average error of 0.8 standard deviations from the true monthly means (Section 4.2.2), [ie. there is a 50% chance that the concentration observed was as close to the true mean as 0.8 standard deviations, and a 50% chance that it was higher or lower than that!].

2. For calculating contaminant loads, the inaccuracy of means estimated by single samples is compounded when accurate flow measurements are not available. Indirect and proxy estimates of flow (e.g. 'estimated annual discharge per hectare') are not accurate enough for load calculations, and loads based on such estimates should not be taken seriously in any management context.

3. Monthly single samples are virtually useless for indicating the frequency or magnitude of short-term exceedances. This is due largely to the rapid rate at which concentrations in surface water can vary, especially during short seasonal episodes. For example, in a record of a year's daily pH measurements in treated effluent, the daily values were found to be independent: each day's pH had *no predictive value* for the following day's pH.

4. The existing monitoring data sets do not contain the information needed to design improved monitoring schedules because they have not measured variation, which is the essence of monitoring design. Even data sets from many years of unreplicated samples do not provide the necessary information.

5. Monitoring programs can be developed for each mine that would greatly increase accuracy without large increases in sampling effort. For example, the error of the estimate of annual zinc load at Mine 'C' could be decreased by more than 60% (from ±2712 kg/yr to ±1043 kg/yr) by taking only 6 additional samples (18 instead of 12). This improvement is accomplished by allocating samples in accordance with the observed seasonal variance pattern instead of fixed monthly periods.

6. Analytical accuracy has been overemphasized as a priority in ARD monitoring: the day to day variations occurring in surface water are much greater than the mine's laboratory vs. analytical laboratory results. Slow feedback of sample results further reduces the value of sending samples to independent labs for analysis.

9.0 RECOMMENDATIONS

The need to revise and improve ARD monitoring is obvious; the goal of designing optimized monitoring schedules cannot be reached without two prerequisites. The first and most important requirement is to critically examine the information needed for management: the accuracy, threshold concentrations, time lags, cost constraints and risks associated with each ARD component that might have any bearing on management decisions (Section 1.1). If these choices and decisions are clear, the new monitoring program can truly enhance the management of ARD sites.

The second requirement is to conduct a proper preliminary sampling program at each site in order to measure variation and distinguish strata (Section 2.6). With these two requirements 'in hand', a statistician can easily determine how and when to sample in order to collect the information needed. It will be necessary to do both steps individually for each ARD site.

Given the rapid rate of variation in surface waters, the speed with which analytical results can be obtained is generally of more importance than analytical accuracy. Monitoring *via* probes and using proxy variables (Sections 1.3.6 and 1.3.7) should be encouraged in situations where surveillance is a high priority.

A much higher priority should be given to accurately measuring flows. Accurate load calculations, and therefore impact predictions, are impossible without accurate discharge records.

The permit for each site should be rewritten to incorporate specific *information goals* of monitoring (e.g. confidence limits for estimates) rather than the sampling methods. Submissions of data and annual reports should demonstrate that these goals have been met.

Preoperational studies at proposed mine sites should include measures of natural variation and should span the season(s) of anticipated high ARD discharge. There is no excuse for collecting inadequate baseline data for future projects.

Background monitoring also should measure natural variation and should span the season(s) of anticipated high ARD discharge.

The Waste Management Branch should retain the services of a statistician to review proposed monitoring methods and field studies, recommend appropriate data analysis methods, and ensure that final reports (both in-branch and those submitted by industry) are statistically correct.

REFERENCES

Arnold, J.C., 1970. A Markovian Sampling Policy Applied to Water Quality Monitoring of Streams. Biometrics 26:739-747.

Boyle, Terence P., editor, 1987. New Approaches to Monitoring Aquatic Ecosystems. ASTM Special Technical Publication 940. American Society for Testing and Materials, Philadelphia.

Cochran, W.G., 1963. Sampling Techniques, 2nd edition. Wiley, New York.

El-Shaarawi, A.H. and S.R. Esterby, editors, 1981. Time Series Methods in Hydrosciences.

Gleit, Alan, 1985. Estimation for Small Normal Data Sets with Detection Limits. Environmental Science and Technology 19:1201-1206.

Godin, B., 1988. Baseline Water Quality Monitoring at the Westmin Resources Limited Silbak Premier Project. Data Report published by Environment Canada, Environmental Protection Service, Pacific and Yukon Region.

Godin, B. and V. Chamberlain, 1990. Baseline Monitoring, Westmin Resources LTD, Silbak Premier Mine. Regional Data Report DR 90-01 published by Environment Canada, Environmental Protection Service, Pacific and Yukon Region.

Green, Roger H., 1979. Sampling Design and Statistical Methods for Environmental Biologists, Wiley, N.Y. 257pp.

Holling, C.S., editor, 1978. Adaptive Environmental Assessment and Management. International Institute for Applied Systems Analysis, Wiley & Sons, New York.

Lettenmaier, D.P., K.W. Hipel, and A.I. McLeod, 1978. Assessment of Environmental Impacts. Part Two: Data Collection. Environmental Management 2:537-554.

Liebetrau, A.M., 1979. Water Quality Sampling: Some Statistical Considerations. Water Resources Research 15:1717-1725.

Mar, B.W., R.R. Horner, J.S. Richey and R.N. Palmer, 1986. Data Acquisition: Cost-Effective Methods for Obtaining Data on Water Quality. Environmental Science and Technology 20:545-55 1.

Niku, Salar, F.J. Samaniego and E.D. Schroeder, 1981. Discharge Standards Based on Geometric Mean. Journal WPCF 53:471-473.

Oguss, E. and W.E.Erlebach, 1976. Limitations of Single Water Samples in Representing Mean Water Quality. I. Thompson River at Shaw Spring, British Columbia. Technical Bulletin #95, Inland Waters Dir., Water Quality Branch, Environment Canada.

Patterson, Robert J., 1989. Assessment of Acid Mine Drainage Control Measures and Resultant Impact on Streams Draining the Equity Silver Minesite, 1988. Published by Equity Silver Mines Ltd.

Perry, J.A., D.J. Schaeffer and E.E. Herricks, 'Innovative Designs for Water Quality Monitoring: Are We Asking the Questions Before the Data Are Collected?' in New Approaches to Monitoring Aquatic Ecosystems, ASTM STP 940, T.P. Boyle Ed., American Society for Testing and Materials, Philadelphia, pp 28-39.

Segar, D.A., DJ.H. Phillips and E. Stamman, 1987. 'Strategies for Long-term Pollution Monitoring of the Coastal Oceans' in New Approaches to Monitoring Aquatic Ecosystems, ASTM STP 940, T.P. Boyle, Ed., American Society for Testing and Materials, Philadelphia, pp 12-17.

Shaarawi, A.H.E. and R.E. Kwiatkowski, 1986. Statistical Aspects of Water Quality Monitoring. Developments in Water Science 27:

Smeach, S.C. and R.W. Jemigan, 1977. Further Aspects of a Markovian Sampling Policy for Water Quality Monitoring. Biometrics 33:41-46.

Sokal, Robert R. and F. James Rohlf, 1969. Biometry. W.H. Freeman and Company, San Francisco.

Steffen Robertson and Kirsten, 1988. Acid Mine Drainage in B.C. -Analysis of Results of Questionnaire from A.M.D. Task Force. Unpublished report.

Steffen Robertson and Kirsten, 1989. Draft Acid Rock Drainage Technical Guide. Volume 1. British Columbia Acid Mine Drainage Task Force Report. Crown Publications, Victoria.

Steele, R.G.D., and J.H. Torrie. 1960. Principles and Procedures in Statistics. McGraw-Hill, New York. 481 pp.

Valiela, D. and P.H. Whitfield, 1989. Monitoring Strategies to Determine Compliance with Objectives. Water Resources Bulletin 25:63-69.

Wald, A. 1947. Sequential Analysis. John Wiley, New York.

Ward, R.C., J.C. Loftis and G.M. McBride, 1986. The "Data Rich but Information Poor" Syndrome in Water Quality Monitoring. Environmental Management 10:291-297.

Whitfield, P.H., 1988. Goals and Data Collection Designs for Water Quality Monitoring. Water Resources Bulletin 19:115-121.

Whitfield, P.H. and P.F. Woods, 1984. Intervention Analysis of Water Quality Records. Water Resources Bulletin 20:657-668.

APPENDIX I: GLOSSARY OF STATISTICAL TERMS

a priori planned in advance

accuracy the closeness of a measured or computed value to its true value.

ANOVA analysis of variance; a parametric statistical test used to compare means.

autocorrelation serial dependency; see Section 2.4.

bimodal (of a distribution) having two peaks; representing the probable combination of two different populations.

censor with respect to frequency distributions, to cut off or mask variation.

coefficient of variation the standard deviation expressed as a percentage of the mean (CV = s * 100 / x); used to compare relative variation in different populations.

confidence interval the difference between the upper and lower confidence limit.

confidence limits the upper and lower boundries within which a population parameter (e.g. mean) is estimated to lie with a certain certainty (e.g. 95% confidence limits).

correlate vary in association with; no cause and effect or dependence relationship is implied.

Correlation the functional relationship between two covarying variables.

design to plan a course of data collection and analysis suited to testing a specific hypothesis at a chosen level of certainty.

error the difference between the observed or estimated value and the true value.

estimated derived by computation using data from a sample, as opposed to the true parametric value.

frequency distribution the curve tracing the outline of a histogram of the frequency of observations within equal divisions of the range of variation.

heterogeneity the property of being poorly mixed or composed of different subgroups.

heteroscedasticity inequality of variances; usually evaluated using an F test on their ratio.

homogeneity the property of being well mixed; randomly sorted.

homoscedasticity equality of variances.

hypothesis a proposition set forth as an explanation or description of some phenomena to guide investigation.

hysteresis the phenomenon exhibited by a system in which the reaction of the system to changes is dependent upon its past reactions to change.

instantaneous existing simultaneously.

mean the arithmetic mean, obtained by adding several quantities together and dividing the sum by the number of quantities.

model a mathematical construct designed to preserve the structure and characteristics of a natural phenomenon or population.

Monte Carlo technique consists of repeated random resampling from an existing data set to produce 'simulated samples' with the same frequency distribution as the original data. Applying the Central Limit Theorem, the means of hundreds of such samples are taken as estimates of population parameters.

multimodal (of a distribution) having several peaks.

noise random and/or uncontrolled variation which increases error in parameter estimates.

normal a bell-shaped curve describing the probability of occurance of different values of a variate.

parameter a population statistic describing the absolute characteristics of a population distribution; e.g. mean, variance.

parametric refering to the entire population; usually used with reference to populations with known frequency distributions, such as the normal distribution.

power of a test the ability of a statistical test to reject a false hypothesis.

population the finite or infinite number of individual items subject to a statistical study.

precision the closeness of repeated measurements of the same quantity.

random every possible sample that could be collected from the population has an equal probability of being selected.

range the limits between which variation exists or is possible; the maximum minus the minimum value.

regression the functional relationship between an independent variable and a dependent variable.

replication repeated measures of the same entity which jointly represent parametric values.

resolution the smallest difference between two entities that can be distinguished accurately.

sensitivity the resolution with which a change can be detected.

significance the probability that the conclusion drawn regarding the hypothesis being tested is correct.

simulation the use of a model or models to generate proxy data or parameters.

spatial existing across space.

standard deviation a measure of dispersion in a frequency distribution, equal to the square root of the mean of the squares of the deviations from the arithmetic mean.

Stratify to divide into strata.

stratum (pl. strata) a division within a larger population in which the data are homogeneous and homoscedastic.

t-test one of several tests using the Student's t distribution for significance testing.

tail with respect to frequency distributions, the far ends of the curve representing the extremes of the range.

temporal existing over time.

time-series analysis a set of statistical methods used on data that is serially dependent and may show cyclical patterns.

transform the reversible mathematical translation of a unit from a linear or arithmetic scale to a non-linear scale.

true the parametric value derived from the entire population, as opposed to 'estimated' which is derived from a sample.

variance the square of the standard deviation; a measure of dispersion.

variation the scatter or spread of observed values of a variable.

Z-score the number of standard deviation units from the mean of a normally distributed population.

APPENDIX 2

Treatment Ponds Effluent, Mine 'C', 1988

Day	January Flow	pH	Zn-T	Cu-T	February Flow	pH	Zn-T	Cu-T	March Flow	pH	Zn-T	Cu-T
1	60850	10.7			41410	9.8	0.44	0.08	33250	11.04	0.27	0.01
2	60850	11.4			25810	10.8	0.3	0.05	37150	10.95	0.52	0.07
3	60850	11.3			34750	10.97	0.46	0.05	30150	10.37	0.2	0.05
4	60850	11.55	0.12	0.01	32380	10.93	0.42	0.1	30560	10.71	0.13	0.04
5	26440	11.56	0.1	0.03	33410	11.22	0.21	0.0	43040	11.3		
6	32610	11.23	0.23	0.05	30580	10.8			43040	11.4		
7	31110	11.41	0.3	0.07	30580	11.4			43040	9.92	0.94	0.14
8	22450	11.2	0.49	0.05	30580	11.32	0.19	0.01	31150	10.05	0.7	0.07
9	18780	10.7			28960	11.34	0.1	0.03	34140	11.13	0.4	0.08
10	18780	11.1			34460	11.19	0.13	0.02	30330	11.13	0.32	0.07
11	18780	11.66	0.21	0.02	34630	11.44	0.15	0.03	36290	10.66	0.45	0.05
12	23550	11.66	0.05		36640	11.2	0.28	0.04	32450	11.5		
13	21090	11.50	0.08		51950	11.1			32450	11.5		
14	37180	11.27	0.19		51950	11.2			32450	10.7	0.4	0.02
15	60660	10.15	0.86	0.09	51950	11.4	0.26	0.03	26670	11.08	0.23	0.01
16	65260	9.6			54060	11.47	0.29	0.02	32860	11.0	0.3	0.03
17	65260	9.5			42870	10.98	0.46	0.08	25050	10.39	0.31	0.02
18	65260	11.31	0.13	0.03	46980	10.73	0.29	0.06	22650	11.15	0.16	0.02
19	26950	11.51	0.19	0.04	49290	10.96	0.6	0.06	29420	11.1		
20	30570	11.12	0.65	0.06	39020	10.3			29420	11.1		
21	30020	11.29	0.15	0.03	39020	10.7			29420	11.24	0.13	0.09
22	42500	10.98	0.58	0.07	39020	11.88	0.1	0.02	30700	10.16	0.8	0.08
23	44450	10.0			39590	11.66	0.34	0.06	41190	11.48	0.26	0.04
24	44450	10.3			37290	11.71	0.12	0.02	36710	11.65	0.18	0.05
25	44450	10.82	0.33	0.06	34210	11.3			39150	11.43	0.67	0.05
26	34560	9.56	1.29	0.12	34210	11.6	0.21	0.01	37110	11.0		
27	36300	10.64	0.36	0.05	31620	11.4			37110	11.0		
28	40770	10.91	0.22	0.05	31620	10.5			37110	11.46	0.2	0.04
29	56770	10.07	0.4	0.08	31620	10.96	0.58	0.02	36460	11.48	0.14	0.03
30	41410	9.2							35470	11.85	0.14	0.01
31	41410	10.2							31620	11.52	0.14	0.01

Treatment Ponds Effluent, Mine 'C', 1988

Day	April Flow	pH	Zn-T	Cu-T	May Flow	pH	Zn-T	Cu-T	June Flow	pH	Zn-T	Cu-T
1	40760	11.5			37490	10.6			42520	10.7		
2	40760	11.0			37490	10.5			42520	10.7		
3	40760	11.0			46920	10.8			42520	10.4		
4	40760	11.0			35390	10.57	0.86	0.1	42520	10.8		
5	40760	11.0			41420	10.3			42520	10.5		
6	40760	11.13	0.14	0.02	41420	10.7			42520	10.66	0.18	0.02
7	40020	10.05			41420	10.5			38030	10.27	0.33	0.05
8	40020	10.7			41420	10.4			42150	10.30	0.28	0.05
9	40020	11.0			41420	9.09	1.6	0.09	43120	11.13	0.13	0.03
10	40020	12.0			39390	10.71	0.5	0.04	41970	11.51	0.23	0.04
11	40020	11.41	0.05	0.02	39420	10.54			39970	11.2		
12	8350	11.23	0.03	0.01	42440	11.5	0.35	0.12	39970	11.1		
13	5260	10.92	0.06	0.02	42810	10.9			39970	11.31	0.26	0.04
14	11950	10.9			42810	10.8			40940	11.13	0.21	0.04
15	11950	11.7	0.07	0.02	42810	10.5			45050	11.08	0.18	0.02
16	32240	10.8			42810	11.45	0.26	0.04	42000	10.76	0.13	0.04
17	32240	10.9			42660	11.61	0.15	0.06	44330	11.05	0.21	0.04
18	32240	11.47	0.05	0.01	50350	10.25	0.9	0.1	33330	11.2		
19	23550	10.6			43780	10.5			33330	10.3		
20	23550	10.55	0.21	0.03	43780	10.5			33330	11.44	0.16	0.02
21	23550	10.39	0.08	0.02	43780	10.3			39380	11.56	0.05	0.02
22	32140	10.0			43780	10.5			39170	11.43	0.14	0.02
23	32140	10.5			43780	10.7			38760	11.15	0.12	0.04
24	32140	10.0			43780	10.56	0.5	0.08	38900	11.25	0.12	0.04
25	32140	9.8			40190	10.9	0.22	0.02	38810	10.3		
26	32140	11.55	0.1	0.02	40190	11.4	0.11	0.0	38810	10.5		
27	33210	11.75	0.1	0.01	38370	10.8			38810	11.38	0.04	0.05
28	40090	11.31	0.3	0.04	42520	10.8			38570	11.29	0.04	0.03
29	48670	11.5	0.17	0.02	42520	10.3			35690	10.98	0.06	0.04
30	37490	11.0			42520	10.4			38010	11.30	0.06	0.02
31					42520	10.3						

Treatment Ponds Effluent, Mine 'C', 1988

	July				August				September			
Day	Flow	pH	Zn-T	Cu-T	Flow	pH	Zn-T	Cu-T	Flow	pH	Zn-T	Cu-T
1	34180	11.2			31640	11.6			32010	10.28	0.19	0.05
2	34180	11.2			31640	11.20	0.23	0.02	34630	10.68	0.36	0.06
3	34180	10.05			33100	9.73	0.65	0.07	30560	10.7		
4	34180	11.38	0.09	0.04	32070	10.48	0.14	0.02	30560	10.4		
5	30740	11.58	0.15	0.03	27850	10.97	0.12	0.02	30560	10.8		
6	19400	11.42	0.04	0.09	33020	11.0			30560	11.15	0.04	0.00
7	24100	11.40	0.03	0.03	33020	10.8			32290	10.6	0.13	0.01
8	29280	11.48	0.04	0.04	33020	10.98	0.48	0.06	37560	9.5	0.13	0.03
9	35720	11.0			32260	11.40	0.11	0.02	26030	11.2		
10	35720	9.7			35750	10.96	0.10	0.01	26030	11.3		
11	35720	9.79	0.22	0.06	34790	9.89	0.24	0.02	26030	10.6		
12	33880	11.46	0.06	0.06	33920	9.72	0.34	0.03	26030	7.95	0.77	0.20
13	38870	11.55	0.03	0.02	30900	10.9			24420	11.0	0.1	0.02
14	35480	11.76	0.04	0.01	30900	11.2			30950	10.6		
15	36140	11.85	0.06	0.00	30900	11.38	0.12	0.04	30950	10.85	0.2	0.00
16	33570	11.6			36020	11.14	0.12	0.05	25650	11.35	0.09	0.00
17	33570	11.5			34900	11.30	0.06	0.02	27340	11.5		
18	33570	11.19	0.08	0.04	36630	11.30	0.15	0.03	27340	10.5		
19	36090	11.12	0.08	0.03	30190	11.00	0.11	0.04	27340	11.5	0.22	0.02
20	36650	11.43	0.06	0.04	38450	9.7			24550	10.5		
21	34720	11.70	0.06	0.04	38450	11.3			24550	9.6		
22	34290	11.64	0.08	0.01	38450	11.79	0.04	0.04	24550	7.75	0.11	0.0
23	33400	11.8			32760	11.49	0.13	0.02	24550	11.5	0.18	0.03
24	33400	11.7			32340	11.87	0.04	0.04	20420	11.2		
25	33400	11.61	0.08	0.03	33990	11.57	0.07	0.02	20410	11.3		
26	38120	11.55	0.07	0.02	39150	10.89	0.24	0.07	20410	11.0	0.06	0.02
27	37330	11.36	0.11	0.05	33670	10.8			28880	10.88	0.11	0.01
28	30930	11.06	0.13	0.03	33670	11.2			26130	10.60	0.11	0.01
29	31640	10.97	0.18	0.03	33670	11.04	0.21	0.06	31240	11.47	0.02	0.00
30	31640	10.9			31320	10.34	0.13	0.04	26630	11.55	0.06	0.01
31	31640	11.5			34310	10.17	0.25	0.08				

Treatment Ponds Effluent, Mine 'C', 1988

	October				November				December			
Day	Flow	pH	Zn-T	Cu-T	Flow	pH	Zn-T	Cu-T	Flow	pH	Zn-T	Cu-T
1	29430	10.6			46875	11.2			38270	9.84	0.14	0.04
2	29430	11.0			46875	9.01	2.04	0.18	31420	11.18	0.17	0.04
3	29430	11.8	0.01	0.0	56410	7.82	2.94	0.34	40700	10.6		
4	25558	10.5			53810	9.90	0.36	0.06	40700	10.6		
5	25558	10.2			58797	10.6			40700	10.25	0.16	0.03
6	25558	10.0			58797	10.0			34800	10.23	0.22	0.03
7	25558	10.3			58797	10.23	0.32	0.07	42390	9.45	0.2	0.06
8	25558	10.6			41700	10.39	0.11	0.02	38410	10.99	0.26	0.04
9	25558	10.2			46700	9.71	0.28	0.03	38410	10.49	0.20	0.02
10	25558	10.5			44080	10.25	0.09	0.02	39390	10.9		
11	25558	11.98	0.09	0.01	46218	9.5			39390	11.1		
12	26360	11.84	0.00	0.01	46218	10.2			39390	10.72	0.13	0.03
13	27912	11.5			46218	10.3			52860	10.34	0.18	0.03
14	27912	11.1			46218	10.91	0.09	0.02	40960	10.32	0.14	0.03
15	27912	10.81			30970	10.87	0.07	0.02	38630	10.42	0.09	0.03
16	27912	11.20			35790	10.99	0.06	0.02	39470	10.83	0.14	0.04
17	27912	11.43	0.07	0.02	33320	10.90	0.13	0.03	39137	10.8		
18	26750	11.31	0.07	0.02	26640	11.08	0.56	0.0	39137	11.0		
19	31660	11.12	0.08	0.01	35547	10.6			39137	10.23	0.1	0.03
20	19460	10.80	0.04	0.02	35547	9.9			35630	10.10	0.11	0.04
21	27280	10.96	0.04	0.01	35547	10.11	0.32	0.06	38720	9.95	0.3	0.06
22	27427	10.0			51890	10.98	0.76	0.08	31020	10.33	0.41	0.05
23	27427	10.6			59970	10.29	0.15	0.03	32320	9.59	0.25	0.05
24	24727	11.54	0.03	0.02	48110	9.86	0.94	0.10	31010	10.5		
25	28820	11.5			41760	9.17	1.94	0.12	31010	10.7		
26	28820	11.6	0.03	0.0	33758	11.0			31010	10.8		
27	28730	11.6	0.13	0.01	33758	10.8			31010	10.8		
28	25972	11.2			33758	10.56	0.27	0.04	31010	11.15	0.27	0.03
29	25972	11.3			33758	11.20	0.15	0.02	34460	10.56	0.18	0.03
30	25972	10.7			35110	10.75	0.10	0.03	36770	10.82	0.13	0.03
31	25972	11.3	0.03	0.01					29320	10.7		