THE USE OF ROUTINE MONITORING DATA
FOR ASSESSMENT AND PREDICTION OF WATER CHEMISTRY

Kevin A. Morin[1] and Nora M. Hutt[2]

[1] Morwijk Enterprises Ltd.
[2] Normar Enterprises, Suite 703, 1155 Harwood Street, Vancouver, British Columbia   V6E 1S1

<u>ABSTRACT</u>

Mining companies collect and analyze numerous water samples every year to protect the environment and to meet permit requirements. Through the years, the collective number of analyses at any minesite can extend into the thousands. This water-quality database is periodically reported to mine administration and regulatory agencies, and then often lies dormant until the next report.

A water-quality database can be used for much more than monitoring environmental conditions and meeting permit requirements. Through the years, the database becomes a valuable scientific tool for the further assessment and prediction of water chemistry. The scientific value is derived from several attributes, including (1) the abundance of data which permits valid statistical analyses and (2) the composite variability in chemistry due to natural processes such as temperature variations and artificial factors such as analytical error. This paper presents the concepts behind this approach for assessment and prediction, the limiting assumptions, and examples from an actual water-chemistry database in British Columbia. The mathematical validity of the approach is demonstrated by an accompanying paper at this symposium using an intensive database of daily and hourly sampling.

1. INTRODUCTION

Metals can be transported around and from a minesite with air, in water, and as solids. Each minesite will have a unique combination of pathways that warrant monitoring, but water transport typically receives much attention at most minesites. Testimony to this attention can be found in thick paper files and large computer spreadsheets at mines across Canada.

The monitoring of water quality around a minesite usually reflects the environmental concerns of the mining company and the permit/license requirements of regulatory agencies. The results of these water-quality analyses indicate how well the mining company is meeting its environmental objectives and the requirements of its permit/license. After a few years, a significant database of water-chemistry analyses begins to accumulate and continues to grow. Typically, little is done with this database after its initial interpretation.

This paper demonstrates how a minesite water-chemistry database can be valuable for detailed assessments and short-term predictions of water quality. The larger and more extensive the database, the more reliable will be the results. Through case studies involving up to thousands of historical analyses, this paper will illustrate the variability that typically exists in water-chemistry databases and how the factors contributing to the variability, such as artificial analytical error and natural seasonal fluctuation, can sometimes be accounted for. It is not the intent of this paper to critically comment on a mine's monitoring program, but to simply take the available database and show how it can be used. The value of this approach can, for example, be found in its use in (1) determining short-term peak concentrations that might bypass a monitoring program, (2) estimating future water-treatment costs, (3) refining water-retention times in holding ponds to obtain a particular range of metal concentration, (4) determining the acceptable degree of failure in water-quality control technologies such as clay covers, and (5) negotiating closure bonds with government agencies.
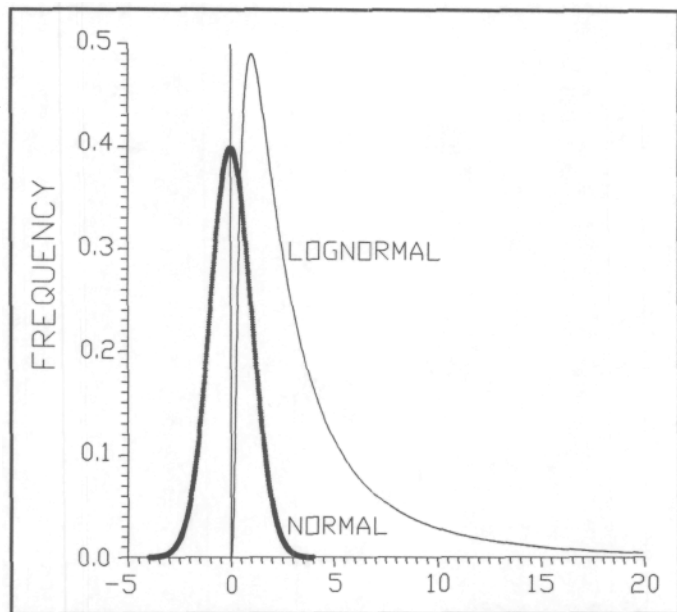
2. THEORY

The important step in reducing a water-chemistry database to a statistically defined model is to identify the independent variables or "predictor variables" from which all other water-chemistry parameters can be calculated. Common predictor variables for the databases are typically: pH, and/or sulfate, and/or flow rate. In reality, there are many potential candidates for predictor variables, such as temperature and rock type, but the intent of this procedure is to reduce the number of predictor variables without significantly increasing the variability in the reduced data. As an example to clarify this point, if pH and flow rate are valid predictors for a particular database, then the use of pH alone would increase the variability of the predictions because flow is now assumed to essentially have no influence. In effect, the use of pH alone in this case creates "scatter" in predictions caused by the changes in flow rate which are not being accounted for. In general, a predictor variable should only be eliminated from the analysis if (1) the increase in variability is not significant upon deleting the predictor from consideration or (2) sufficient data for the predictor are not available in the database. With the previous example, flow rate should be eliminated from consideration as a predictor (1) if there are few data on flow or (2) if flow measurements were not made at the same time water-chemistry
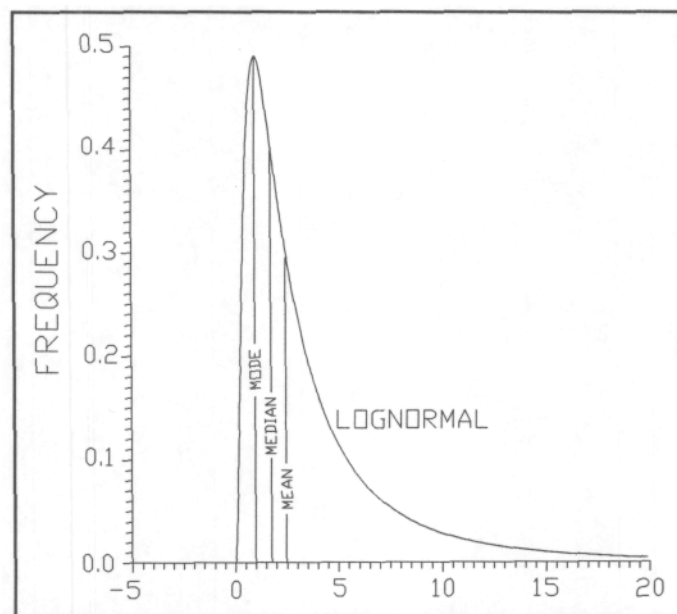
samples were collected.

Because we are working with large databases, it necessary to mathematically reduce their size to more manageable levels. A basic way to do this is through statistics, using concepts such as means and standard deviations. A review of basic statistics and a discussion of logarithmic transformations are useful at this point.

A "normal" frequency distribution is one that is symmetrical on both sides of the mean, similar to a sloping inverted "V" (Figure 1). Only if data distributions are normal can the basic concepts of means, standard deviations, and probability levels be used. If data are not normally distributed, then mathematical transformations such as logarithms are available to transform the data to a normal distribution. For example, if the values for the lognormal distribution in Figure 1 were to be transformed into logarithmic values, the new distribution would appear normal like the other curve in Figure 1. However, great care must be used in interpreting log-normal distributions because the means, medians, modes, and probabilities do not have the same meanings as in normal distributions (e.g., Figure 2). Nevertheless, these lognormal distributions are common in geochemistry and other geological sciences (Davis, 1986).

In the case of standard deviations, the logarithmic value (number of log cycles above or below the logarithmic mean) is added or subtracted from the logarithmic mean. For example, if a mean were 2 ($10^2 = 100$) and the standard deviation was 0.2 log cycles, then +1 standard deviations above the mean would be 158 ($10^{[2+(02*1)]}$). Two standard deviations above the mean would be 251 ($10^{[2+(02*2)]}$). Further discussion of this and related matters is presented in Aitchison and Brown (1976) or Davis (1986).

FIGURE 1. Examples of Normal and Lognormal Distributions

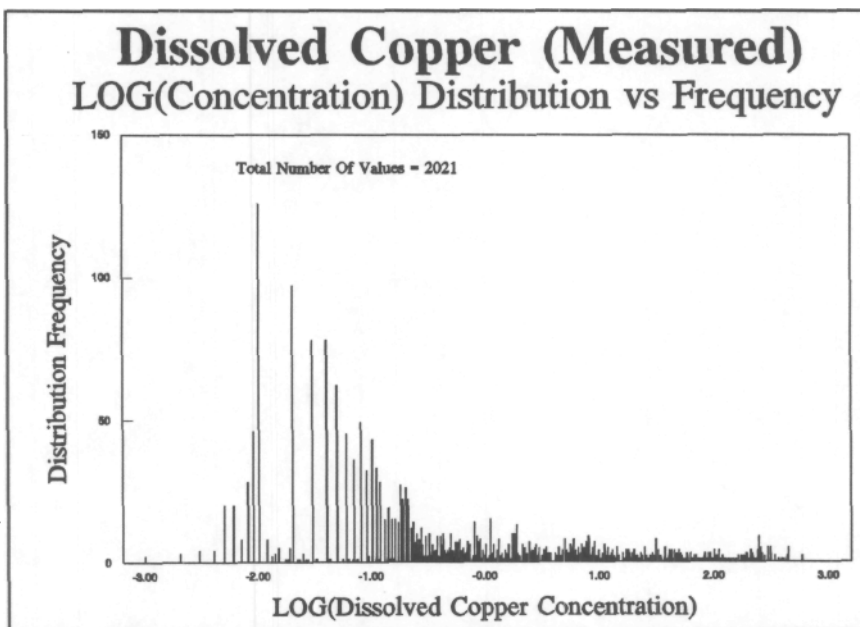FIGURE 2. Mode, Mean, and Median in Lognormal Distributions

3. PROCEDURE

The obvious first step is to compile all water-chemistry data no matter when or where it was collected, and no matter what its perceived reliability may be. Replicate analyses or proof of data integrity are not required. For example, at each of eight monitoring stations, an intensive monitoring database of up to six hundred analyses was formed over a six month period at Island Copper (Morin et al., 1993, these proceedings). Although no replicates or proof of data quality were included in the study, valid interpretations could still be drawn from the information. In fact, that intensive database has provided the statistical information to justify the procedure used in this paper.

At the Bell Minesite near Granisle, British Columbia, analyses of water chemistry had accumulated for over ten years prior to its recent closure. These analyses included the tailings supernatant, seepage through the tailings-impoundment dams, toe seepage from the mine-rock dumps, and pit-sump pumpage (Morin and Hutt, 1993). For dissolved copper alone, there were more than 2000 analyses in the database. This copper database will be used to illustrate the procedure.
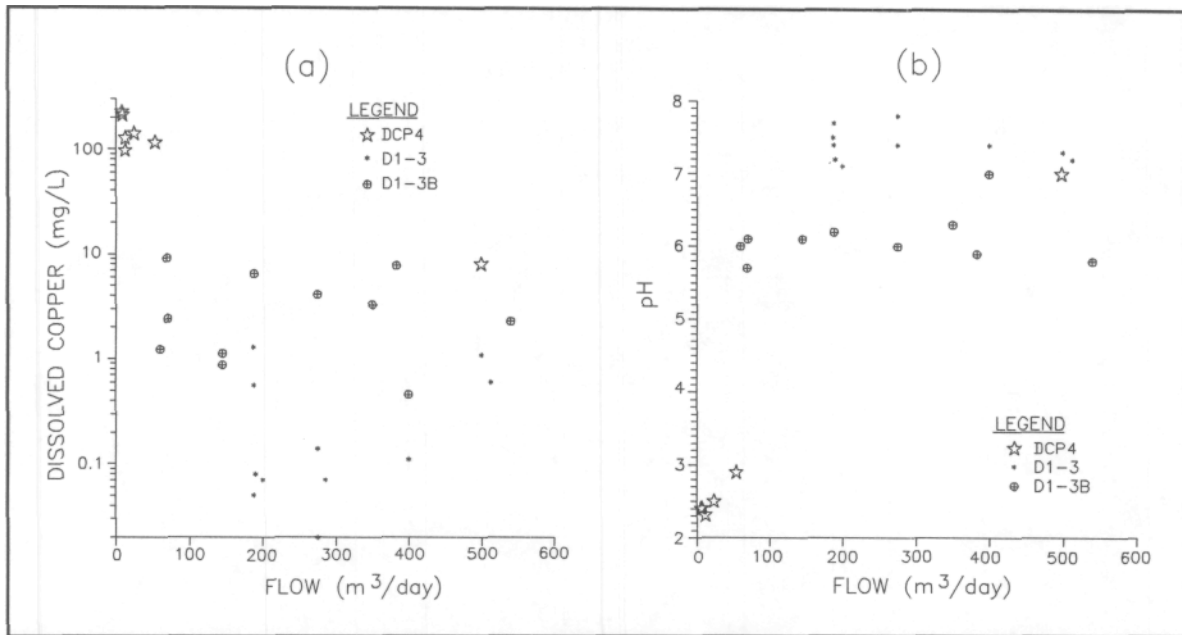
Concentrations in the cumulative dissolved copper database were found to span almost six orders of magnitude (a factor of 1,000,000) with the highest individual frequencies below 0.1 mg/L (Figure 3). With such a range of concentrations, calculations of statistical parameters such as means and standard deviations would be of little value and, because the data are not normally distributed, would not be statistically valid in any case. As a result, the search began for the predictor parameters which caused the large variation in the data.

There was little corresponding data for both flow rate and chemistry, but the few available data suggested copper was not affected by flow rate within a range of approximately one order of magnitude (Figure 4a). In fact, the changes with flow could actually be attributed to the changes in aqueous pH (Figure 4b). Similarly, variations in copper did not correlate well with time (Figure 5) and the same variations in concentration often appeared from year to year.
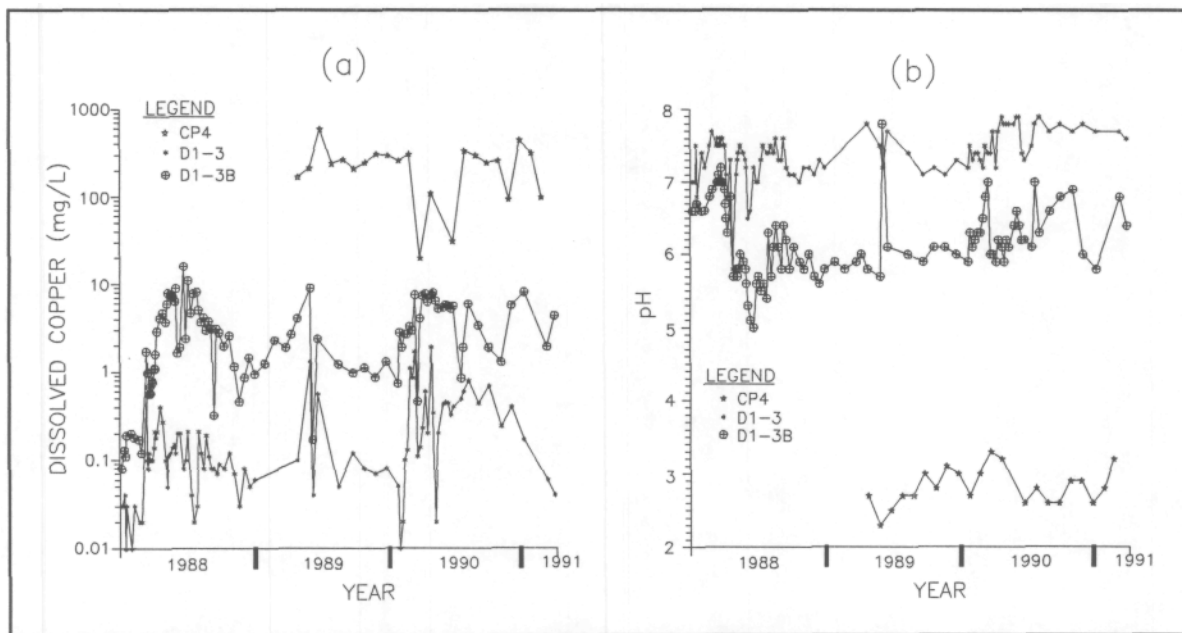
A scatterplot of copper with pH thus confirmed that pH was in

FIGURE 3. Logarithmic$_{10}$ Frequency Distribution of Historical Bell Mine Analyses for Dissolved Copper.

194

FIGURE 4. Comparison of Flow to (a) Dissolved Copper and (b) pH.

FIGURE 5. Comparison of Time to (a) Dissolved Copper and (b) pH.

fact the primary predictor parameter (Figure 6). As a result, the variations in pH could explain much of the variation noted in the entire Bell Mine copper database.

A line was iteratively fitted to the pH-copper distribution (Figure 6) until the frequency

195

distribution of the error (defined as predicted copper minus measured copper) was essentially centered at an error of zero (Figure 7). However, this approach yielded negative values for the error which would lead to negative predicted concentrations when a standard deviation was applied to the data. Therefore, this technique was considered unacceptable for detailed evaluations and predictions.
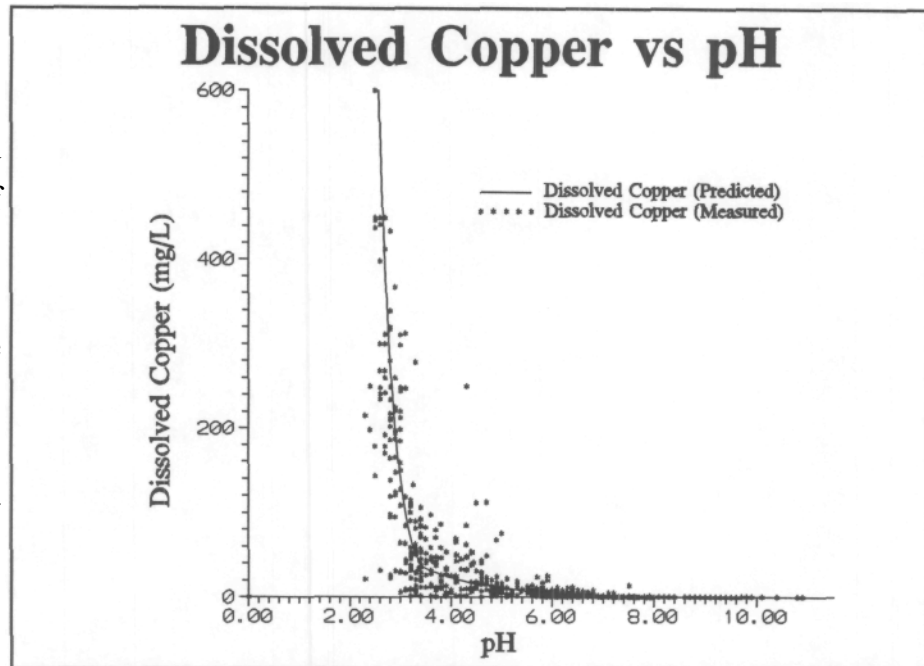

FIGURE 6. Arithmetic Scatterplot of Dissolved Copper vs. pH for Historical Bell Mine Analyses.

An alternative approach using an error defined by predicted copper divided by measured copper did not produce a statistically normal distribution (Figure 8). Consequently, this approach was also considered unacceptable for detailed evaluations and predictions.

A scatterplot of the logarithmic concentration of copper with pH showed greater promise than the other approaches, because it highlighted the distribution around neutral pH (Figure 9) which was not obvious on an arithmetic plot (Figure 6). The logarithmic plot also generated about the
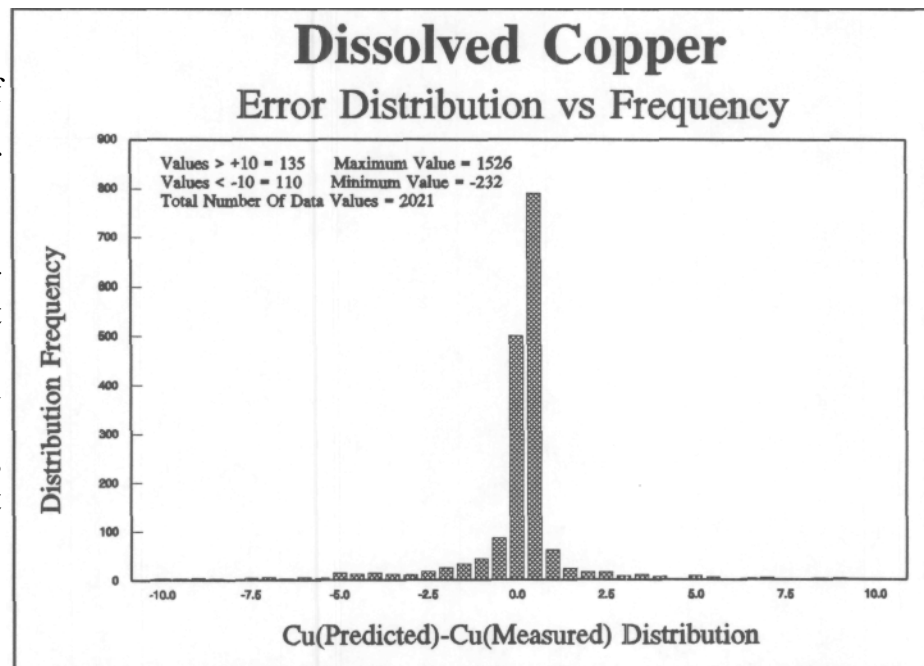

FIGURE 7. Prediction Error (Cu_{pred}-Cu_{meas}) vs. Frequency for Dissolved Copper.

same amount of scatter at any particular pH (along any vertical line on Figure 9), which is confirmed by calculated standard deviations in various ranges of pH (Figure 10). Finally, the logarithmic plot did not lead to any negative predicted concentrations because logarithms of negative values are undefined.

A three-segment line was visually placed through the data trends

FIGURE 8. Prediction Error (Cu_{pred}/Cu_{meas}) vs. Frequency for Dissolved Copper.

of log copper with pH (Figure 9). The error was then defined as the number of log cycles between the log of the predicted concentration and the log of the measured concentration at each value of pH. The frequency distribution of this error resembled a normal distribution (Figure 10) and this approach was then adopted as a valid procedure. The best-fit lines of Figure 9 were adjusted as necessary to minimize the logarithmic error to a minimal value (0.133 for Figures 9 and 10).

With this procedure established, similar evaluations were made for all other metals and non-metals. In fact, even the logarithm of composite parameters such as acidity provided good results (Figures 11 and 12). Also, comparisons of 1992 data with the older data
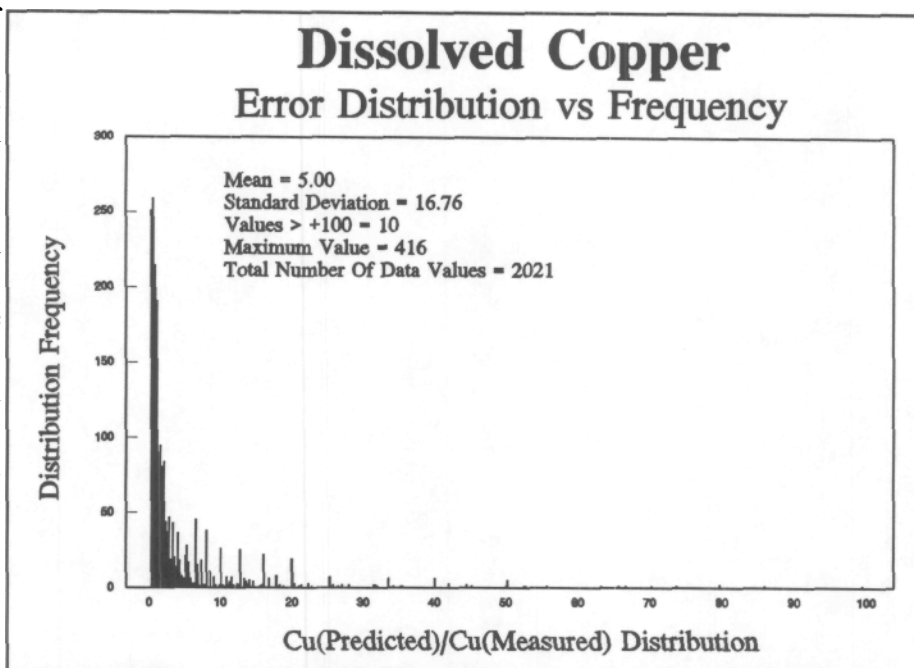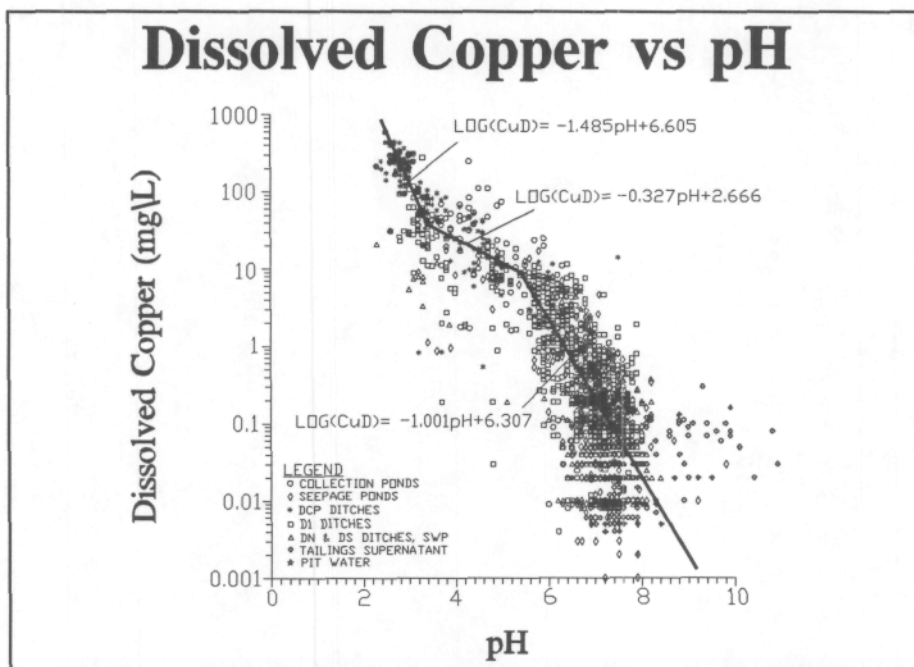

FIGURE 9. Logarithmic$_{10}$ Scatterplot of Dissolved Copper vs. pH for Historical Bell Mine Analyses.

197

demonstrated again that the same variations occurred yearly, reflecting a repetition of the same natural processes such as climatic variations and/or the same artificial factors such as analytical accuracy.

The compilation of all water-chemistry data following this procedure led to an empirical model which summarized the entire database (Table 1). From this water-chemistry model,

FIGURE 10. Prediction Error ($\log_{10}(Cu_{pred}/Cu_{meas})$) vs. Frequency for Dissolved Copper.

information could be obtained on (1) concentrations of short duration such as one hour, (2) the worst-case water chemistry that could appear in a holding pond with a residence time of one month, and (3) the predictions of future water chemistry.

For predictions of water chemistry, indirect tests of the empirical model (Table 1) showed that it would likely have reasonable accuracy even into time spans of centuries. The only requirement was that pH and acidity be predicted in each watershed through time. This requirement was met through extensive ABA analyses and kinetic tests (humidity cells and three scales of heap-leach testwork) of all relevant geologic materials at the minesite (Morin and
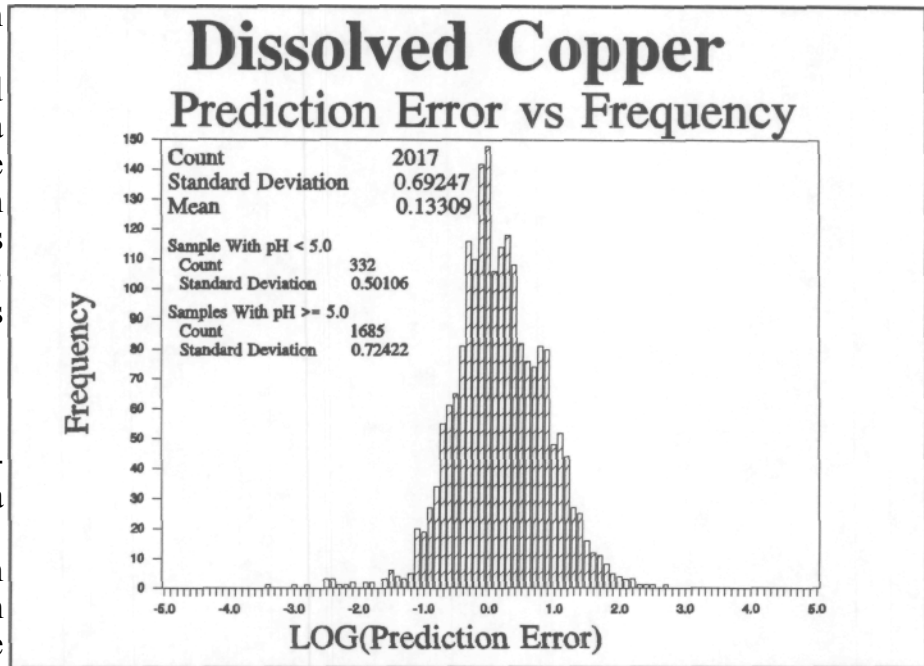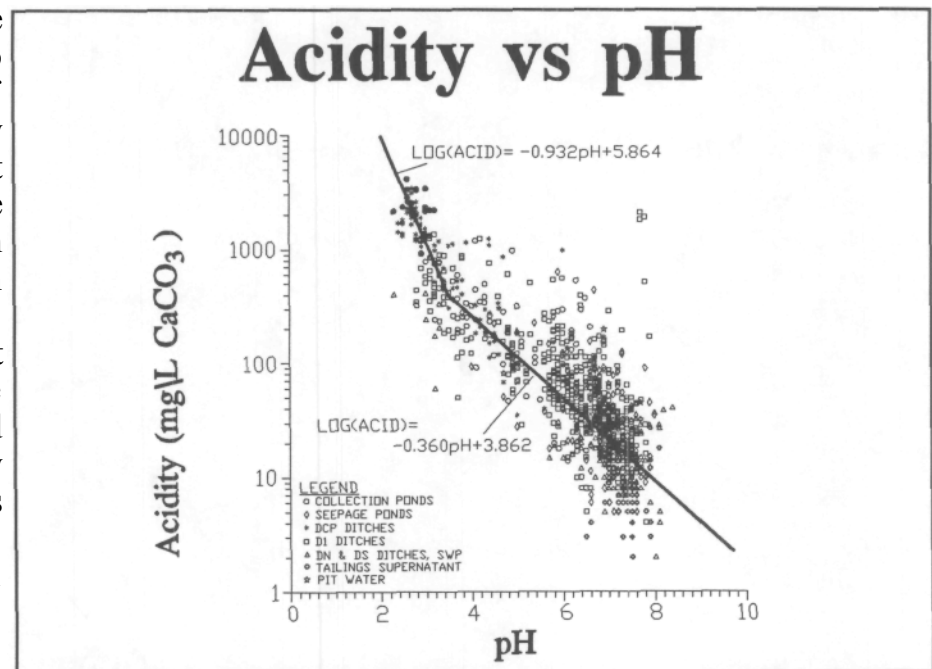

FIGURE 11. Acidity vs. pH.

Hutt, 1993).

For estimation of concentrations of various durations at the minesite, the approach described by Morin et al. (1993) was used in which each period of time can be related to a particular probability level (Table 2). According to Morin et al., there are probability levels which theoretically extend towards infinity at extreme low probabilities and at
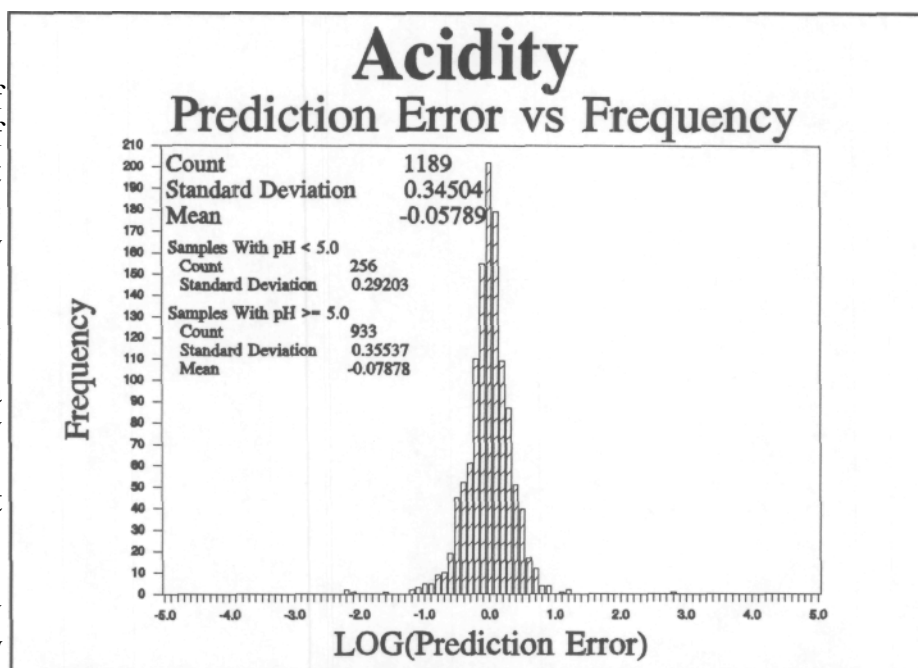

FIGURE 12. Acidity Prediction Error vs. Frequency.

extremely short times (less than a second). However, this study benefitted from three scales of heap-leach testwork (samples weighing approximately 23, 2500, and 30000 kg, respectively) in which sulfuric-acid-laced water was constantly recirculated through the rock for over a year. Aqueous pH reached levels as low as 0.8 and the corresponding aqueous concentrations reached high, steady-state levels which defined "worst-case" chemistry. This worst-case chemistry was composed of concentrations that were typically around +1 to +2 standard deviations above the means. According to Table 2, this chemistry corresponds to yearly repetitions of monthly to quarterly durations. Consequently, the monthly to quarterly sampling on which most of the database was established was appropriate for detecting maximum on-site concentrations.

ACKNOWLEDGEMENTS

TABLE 1

EXAMPLE OF A SITE-SPECIFIC EMPIRICAL WATER-CHEMISTRY MODEL

PARAMETER	pH RANGE	BEST-FIT EQUATION	LOG (STD DEV'TION)
Acidity	pH < 3.5	$\log(acid) = -0.932pH + 5.864$	0.345
	pH > 3.5	$\log(acid) = -0.360pH + 3.862$	
Alkalinity		$\log(alk) = +0.698pH - 3.141$	0.654
Diss Aluminum	pH < 6.0	$\log(Al) = -0.925pH + 4.851$	0.429
	pH > 6.0	$Al = 0.2$ mg/L	
Total Aluminum		If diss Al > 0.3 mg/L, total Al = diss Al	
Diss Arsenic		< 0.2 mg/L	
Diss Cadmium	pH < 3.0	$Cd = 0.07$ mg/L	
	pH > 3.0	$Cd = 0.015$ mg/L	
Diss Calcium		$\log(Ca) = +0.619\log(SO4) + 0.524$	0.375
Diss Chromium	pH < 3.0	$Cr = 0.08$ mg/L	
	pH > 3.0	$Cr = 0.015$ mg/L	
Diss Copper	pH < 3.4	$\log(Cu) = -1.485pH + 6.605$	0.692
	3.4 < pH < 5.4	$\log(Cu) = -0.327pH + 2.666$	
	pH > 5.4	$\log(Cu) = -1.001pH + 6.307$	
Total Copper		$\log(CuT) = +0.962\log(CuD) + 0.180$	0.230
Diss Iron	pH < 4.4	$\log(Fe) = -1.429pH + 6.286$	0.807
	pH > 4.4	$\log(Fe) = -0.455pH + 2.000$	
Total Iron		If diss Fe > 1.0, total Fe = diss Fe	
Diss Lead		$Pb = 0.05$ mg/L	
Diss Nickel		$\log(Ni) = -0.317pH + 0.853$	0.607
Total Nickel		total Ni = dissolved Ni	0.613
Diss Selenium		$Se = 0.2$ mg/L	
Diss Silver		$Ag = 0.015$ mg/L	
Diss Zinc		$\log(Zn) = -0.441pH + 1.838$	0.667
Total Zinc		total Zn = dissolved Zn	0.144

TABLE 2					
PROBABILITY LEVELS AND CORRESPONDING TIME INTERVALS					
TIME INTERVAL	1 YEAR	1 MONTH	1 WEEK	1 DAY	1 HOUR
PROBABILITY	100%	8.3%	1.9%	0.27%	0.011%
NO. OF STD DEV ABOVE MEAN[1]	+0.00	+1.73	+2.34	+3.00	+3.85

[1] Obtained from normal-distribution tables (e.g., Pollard, 1977) after dividing probability by 2

REFERENCES

Aitchison, J., and J.A.C. Brown. 1976. The Lognormal Distribution. University of Cambridge, Department of Applied Economics Monograph 5. Cambridge University Press. 176 p.

Davis, J.C. 1986. Statistics and Data Analysis in Geology, Second Edition. John Wiley & Sons, Toronto. 646 p.

Morin, K.A., and N.M. Hutt. 1993. Mine-Rock and Tailings Geochemistry, and Prediction of Water Chemistry. Support Document E, Bell '92 Closure Plan. 400 p.

Morin, K.A., I.A. Home, and D. Flather. 1993. The appropriate geochemical monitoring of toe seepage from a mine-rock dump. 17th Annual Mine Reclamation Symposium (these proceedings), May 4-7, Port Hardy, British Columbia.