Mine Water Characterization for Probabilistic Modelling and Uncertainty Analysis

Nicolas Lauzon

Golder Associates Ltd., 500—4260 Still Creek Drive, Burnaby, BC, V5C 6C6, nlauzon@golder.com

Abstract Confidence in mine water or effluents concentrations predicted from water quality models is dependent on the validity of chemical loadings from water sources given as inputs to these modelling tools. This paper discusses a step by step process for fitting probability distributions to source water concentrations from observed water samples. This process is intended to ensure that time series generated from these distributions are composed of values within realistic ranges of source concentrations. The variability of the mean and standard deviation of these distributions and its incorporation into uncertainty analyses is also addressed in this paper.

Key Words Water samples, water quality modelling, probability distributions, uncertainty analysis

Introduction

Water quality models are used in mining for impact assessments and operational analyses in order to predict constituent concentrations in typical mine facilities (e.g. tailings water impoundments or pit lakes) or within a receiving aquatic environment affected by mine effluents. These models depend on water source loadings (i.e., mine and natural waters), which may be characterized through monitoring programs that include flow measurements and water samples analyzed for various constituents. However, water samples may be reduced in number and therefore represent only a subset of the range of possible concentrations from the source loadings. Probabilistic water quality modelling (fig. 1) is intended to address these cases with limited data. This modelling approach entails the generation of time series of source concentrations from probability distributions fitted on observed water samples. The length of the generated time series can be appreciably long (e.g., 50 years or more), so that the water chemistry predicted from the model may encompass a large number of combinations of climatic, chemical loading and flow conditions.

This paper discusses the step by step process for fitting of distributions so that they provide realistic ranges of source concentrations. These steps relate to the preliminary screening of water

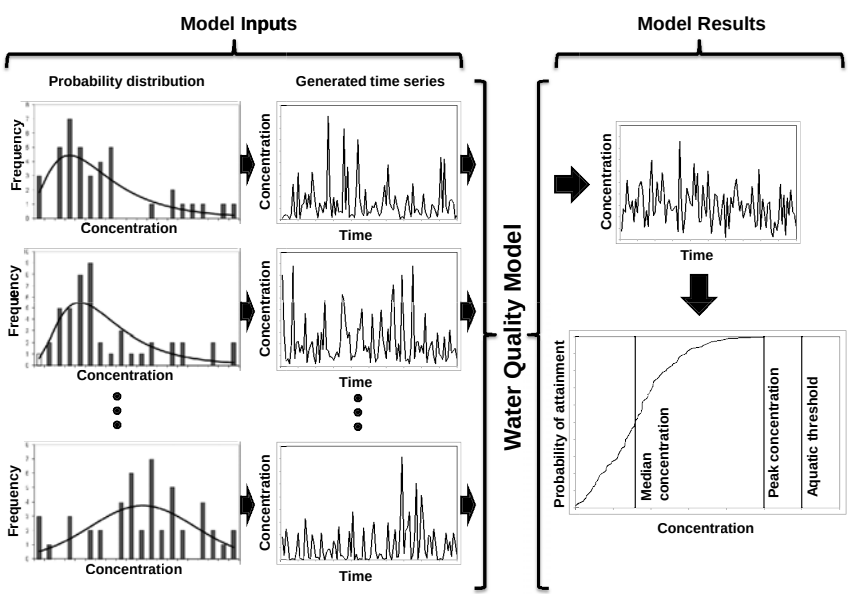

Figure 1 *Conceptual Diagram of the Probabilistic Water Quality Modelling Structure*

samples, detection of outliers, distribution fitting and assignment of realistic bounds on the distributions. The characterization of the variability of the distribution mean and standard deviation for uncertainty analysis is also addressed.

Probability Distributions for Water Sample Concentration Data

The process for distribution fitting (fig. 2) expands the method provided by EPA (1991), which assumed that concentrations from water samples typically follow normal or lognormal distributions, considering also into account the effect of non-detectable concentrations by accepting the delta lognormal as a possible distribution. The expansion addresses 1) special cases (e.g., limited or no data, or no detectable concentrations); 2) the screening of the water samples to prevent unusual concentration values from impacting the fitting of the probability distribution; and 3) the determination of an upper bound for the fitted distribution. The process was developed and described in details for application to water quality modelling in the assessment of impacts of oil sands mining developments in northern Alberta (Imperial 2005, Shell 2005, 2007).

The distribution fitting process must be undertaken for each water quality constituent of each water source. Fitting distributions on data groups separated into seasons is also recommended if seasonal trends are identified in a sufficiently large pool of water samples. Developing an application tool for implementing this process is recommended to avoid errors when the numbers of water sources and constituents are significant. The handling of special cases constitutes the preliminary steps to determine if the available pool of water samples is adequate to allow a distribution to be fitted. A surrogate water source may be used if no data is available for a given constituent of the source under study. Specific distributions would be assigned to cases with no detectable concentrations (i.e., constant at zero) or with limited data (i.e., lognormal or delta-lognormal, based on the number of detectable values and considering that such scarce values cannot be enough to estimate a standard deviation).

The most important aspect of the distribution fitting process is the screening of water samples, since it implies a decision whereby some samples may have limited or no interest. The objective of the screening is to identify concentration values that may induce a bias in the calculation of the distribution parameters, and ultimately generate times series of unreasonably or unrealistically high concentrations. Expert judgment is proposed as a preliminary screening (fig. 2) with the intent of discarding unusual samples, particularly high values from suspicious analytical results or non-detectable concentrations resulting from unreasonably high detection limits. Expert

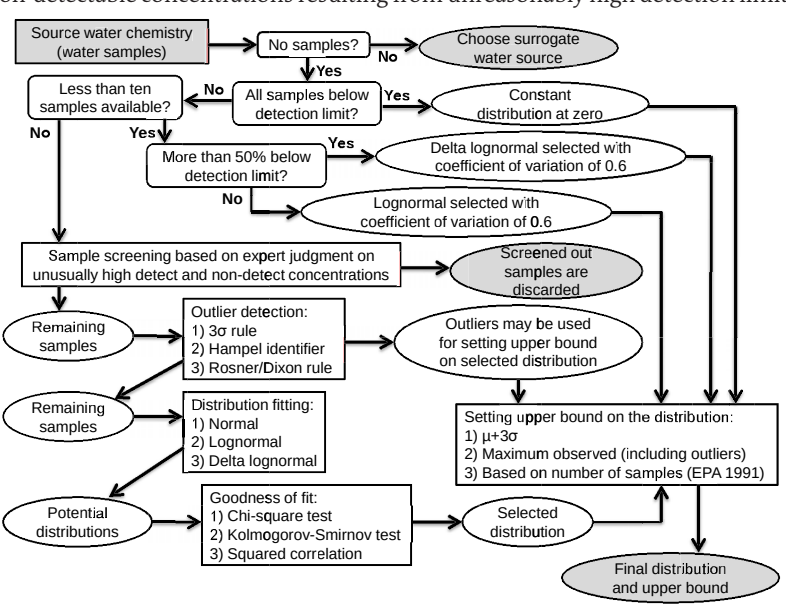

***Figure 2** Processing of Water Sample Concentration Data for any Given Constituent*

judgment must be based on comprehensive knowledge of the water sources considered. A secondary screening consists in the identification of high value using acknowledged outlier detection methods (fig. 2). Identification should be conservatively based on positive diagnostic from several detection methods. Furthermore, the identified values should still be considered for the setting of the upper bound of the distribution, even though they would not be used to calculate the distribution parameters (i.e., mean and standard deviation).

The second most important aspect of the distribution fitting process is the establishment of an upper bound on the fitted distribution, and no values in generated time series should be higher than that selected threshold. Several methods for the calculation of this upper bound should be considered (fig. 3). A final upper bound value must be selected based on expert judgment, supported by comprehensive knowledge of the water source considered. The upper bound must be conservatively high yet reasonable. This threshold is not expected to be higher than detectable values discarded from the preliminary screening, although it may be significantly higher than any identified outliers. Ultimately, the selected distributions must be able to generate time series with statistical characteristics (i.e., mean and standard deviation) that are similar to those of the concentrations from the water samples.

Uncertainty Analysis

Monte-Carlo simulations (Ayyub and McCuen 1997) are typically used as part of an uncertainty analysis on modelling predictions. This method requires the random generation of several realizations of the inputs fed to the model, in order to produce confidence limits on the modelling predictions applicable to a scenario designed for the mine development (fig. 3). The generation of the realizations would apply to all inputs to the model (i.e., processes, flow and water quality inputs), although this paper discusses only the processing of water quality inputs, namely concentrations of water sources.

Considering that the number of water samples might be small, the monitored data may represent only a subset of the data population. Consequently the statistical parameters, such as the mean and standard deviation of the distribution fitted to these data may be uncertain. Realizations of sets of statistical parameters describing uncertainty in the statistical distribution of the input data can thus derived by randomly generating a series of means and standard deviations.

The proposed method of generation assumes that the mean and standard deviation of the data population is unknown. Consider that $\bar{x}_{obs}$ and s_{obs} are the mean and standard deviation, re-

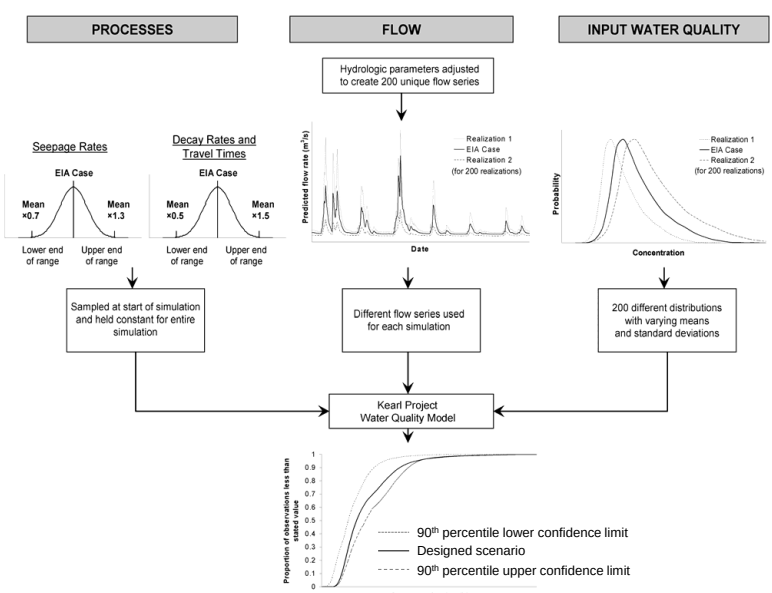

Figure 3 *Conceptual Diagram of the Water Quality Uncertainty Analysis Model*

spectively, of the observed sample concentrations for a given constituent, and $\bar{x}_{obs,i}$ and $s_{obs,i}$ are the mean and standard deviation, respectively, of realization *i* from the population that describe the constituent. The null hypotheses that apply for the mean and standard deviation are $\bar{x}_{obs} = \bar{x}_{obs,i}$ and $s_{obs} = s_{obs,i}$, respectively. The sequence of observed concentrations is of size n_{obs} and so is the size of the sequence of data that would be associated with $\bar{x}_{obs,i}$ and $s_{obs,i}$. If the sample size is n_{obs}, the statistical test relating s_{obs} to $s_{obs,i}$ provides the following relationship:

$$s_{obs,i} = \left(F s_{obs}^2\right)^{0.5} \quad (1)$$

where variable F is assumed to follow a Fisher distribution $f_p(v_1,v_2)$, of probability p and degrees of freedom v_1 and v_2. The degrees of freedom can be expressed as $v_1 = 2\, n_{obs} - 1$ and $v_2 = 2\, n_{obs} - 1$. A value for F is obtained by a random generator for each realization, according to the Fisher distribution. The statistical test relating $\bar{x}_{obs}$ to $\bar{x}_{obs,i}$ provide the following relationship:

$$\bar{x}_{obs,i} = \bar{x}_{obs} + T\left(\frac{s_{obs}^2 + s_{obs,i}^2}{n_{obs}}\right) \quad (2)$$

where variable T is assumed to follow a Student's distribution $t_p(v)$, of probability p and degree of freedom $v = 2\, n_{obs} - 2$. A value for T is obtained by a random generator for each realization, according to the Student's distribution.

Conclusions

Confidence in modelling predictions of concentrations within mine facilities or a receiving aquatic environment affected by mine effluents is dependent on a proper characterization of source water chemistry (i.e., mine and natural waters). The characterization described in this paper applies to probabilistic modelling of water concentrations, where probability distributions fitted from observed water samples are used to generate concentration time series as model inputs. The step by step process proposed for the fitting of the distributions is based on well established methods, which are however structured for the purpose of generating time series of water source concentrations within realistic ranges. In large part, the process calls on expert judgment on critical steps (i.e., screening of water samples and establishment of an upper bound) for the generation of concentrations that are conservative but not unrealistic given the existing knowledge of the source waters. Characterization of the variability of the mean and standard deviation of the concentration probability distributions was also addressed in this paper. This variability may be incorporated in uncertainty analyses, which are often required for determining confidence on modelling predictions of water concentrations.

Acknowledgements

I thank my colleagues at Golder Associates for their advice, suggestions, and review throughout the years of the content presented in this paper. I particularly extend my thanks to Ian Mackenzie, Jean-Philippe Bechtold, Jerry Vandenberg, Tammy Rosner, Dennis Kramer and Tasha Hall.

References

Ayyub BM, McCuen RH (1997) Probability, statistics, & reliability for engineers. CRC Press, Boca Raton, Florida, USA, 514 pp

EPA (1991) Technical support document for water quality-based toxics control. United States Environmental Protection Agency, EPA report number 50512-90-001

Imperial (2005) Kearl Oil Sands project - mine development. Submitted to Alberta Energy and Utilities Board and Alberta Environment, prepared by Imperial Oil Resources Ventures Limited, volumes 1 to 9, Calgary, AB

Shell (2005) Muskeg River Mine Expansion project application and environmental impact assessment. Submitted to Alberta Energy and Utilities Board and Alberta Environment, prepared by Shell Canada Limited, volumes 1 to 4, Fort McMurray, AB

Shell (2007) Jackpine Mine Expansion & Pierre River Mine project application and environmental impact assessment. Submitted to Alberta Energy and Utilities Board and Alberta Environment, prepared by Shell Canada Limited, volumes 1 to 5. Calgary, AB