

## Probabilistic modelling applied to the mining industry to address water quality uncertainty

**N. Lauzon<sup>a</sup>, J.A. Vandenberg<sup>a</sup> and J.P. Bechtold<sup>a</sup>**

<sup>a</sup> *Golder Associates Ltd., Canada*  
Email: [nlauzon@golder.com](mailto:nlauzon@golder.com)

**Abstract:** Water quality models can be used for impact assessments and operational analyses of mining projects 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 effluent. 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 analysed for various constituents. However, water samples may be limited in number and therefore represent only a subset of the range of possible concentrations from the source loadings. A probabilistic water quality modelling approach is presented in this paper and is intended to address these cases with limited data. This approach entails the generation of time series of source concentrations from probability distributions fitted to observed water samples. The length of the generated time series can be 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.

An approach to characterise the variability of the parameters (i.e., mean and standard deviation) of the fitted probability distributions is then presented in this paper, and can be incorporated in the probabilistic modelling formulation to address uncertainty in water quality predictions. From this characterization, the model can be used to establish confidence bands on water quality predictions as part of an uncertainty analysis. Incorporation of climate scenarios to assess the impact of climate change to predicted constituent concentrations will also be discussed.

The modelling approach is briefly demonstrated with data and results from the environmental impact assessments of an oil sands mining development in Northern Alberta, Canada. The results presented herein illustrate the development of uncertainty bands around the results produced as part of the impact assessment (EIA). The results also show the uncertainty bands developed under different climate scenario as part of a climate change analysis. These bands, when compared to those developed for the uncertainty analysis under the EIA climate, show the possible shift in concentrations that may occur as a result of different climate conditions affecting the aquatic environment.

**Keywords:** *Water quality modelling, mining, water source loadings, probability distributions, uncertainty analysis*

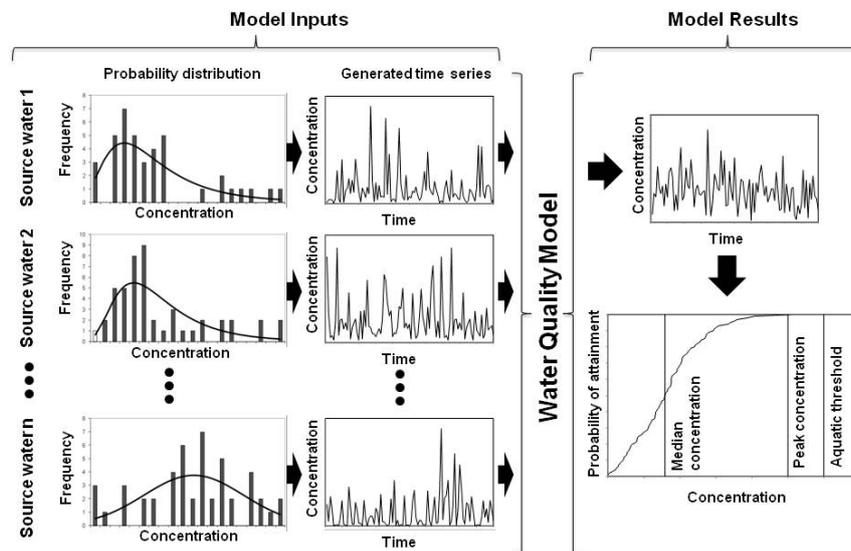
## 1. INTRODUCTION

Water quality models are used for impact assessments and operational analyses of mining projects 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 effluent. 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 analysed for various constituents. Concentrations of these constituents in mine and surface waters are innately variable. The modelling approach described in this paper involves using probability distributions estimated from observed data sets to describe the variability of these source concentrations. The approach entails the generation of time series of source concentrations from these probability distributions. The concentration time series are provided as inputs to a water quality model to predict constituent concentrations and associated variability for scenarios of interest (Figure 1). The length of the generated time series can be long (e.g., 50 years or more of daily inputs), so that the water chemistry predicted from the model may encompass a large number of combinations of climatic, chemical loading and flow conditions.

This probabilistic modelling formulation is then used as the premise to address uncertainty in water quality predictions. The approach proposed herein is the characterization of the variability in the mean and standard deviation of the source concentration distributions and its incorporation in the probabilistic model. From this characterization, the model can be used to establish confidence bands on water quality predictions as part of an uncertainty analysis.

The modelling and uncertainty analysis approaches presented in this paper have been developed and applied to several proposed and existing mining developments in the Athabasca oil sand region, Northern Alberta, Canada. The methodology and results of this approach were publicly reported in several environmental impact assessments, including those of Imperial Oil (2005), Shell (2005 and 2007) and Total (2010).

This paper discusses the step-by-step process for distribution fitting to provide realistic ranges of source concentrations. The characterization of the variability of the distribution mean and standard deviation for uncertainty analysis is then described. Incorporation of climate scenarios to assess the impact of climate change to predicted constituent concentrations is also discussed.



**Figure 1.** Conceptual diagram of the probabilistic water quality modelling structure (for a given water quality constituent).

## 2. PROBABILITY DISTRIBUTION FOR WATER SAMPLE CONCENTRATIONS

The process for distribution fitting (Figure 2) expands the method provided by U.S. EPA (1991), which assumes that concentrations from a set of water samples typically follow a normal, lognormal or delta-lognormal distribution. The expansion addresses: 1) special cases (e.g., limited or no data, or no detectable concentrations); 2) the screening of data to prevent unusual concentration values from impacting the fitting of

the probability distribution; 3) goodness of fit methods for selecting a distribution appropriate to the data, and 4) the determination of an upper bound for the fitted distribution.

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 automation tool for implementing this process is recommended to avoid errors when the numbers of water sources and constituents are substantial.

The handling of special cases constitutes the preliminary step to determine if the available pool of water samples is adequate to allow a distribution to be fitted. A surrogate (i.e., similar) water source may be used if no data are available for a given constituent of the source under study. Specific distributions would be assigned to cases with no detectable concentrations (i.e., uniform or constant distributions, or any symmetric distribution centred at half the detection limit) or with limited data (i.e., lognormal or delta-lognormal, based on the number of detectable values and considering that the production of time series from the assigned distribution will have a significant degree of uncertainty).

A critical aspect of the distribution fitting process is the screening of water samples, since it implies a decision whereby specific data may be removed from the population. 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 unrealistically high concentrations. Expert judgment is used as a preliminary screening (Figure 2) with the intent of discarding unusual samples, particularly high values from suspicious analytical results or non-detectable concentrations resulting from analytical methods with unreasonably high detection limits. Expert judgment must be based on a comprehensive knowledge of the water sources considered and the analytical techniques applied. A secondary screening consists of the identification of high values using acknowledged outlier detection methods, such as the  $3\sigma$  rule, the Hampel identifier and the Rosner/Dixon rule (Figure 2). Identification should be conservatively based on the 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 are not used to calculate the distribution parameters (i.e., mean and standard deviation).

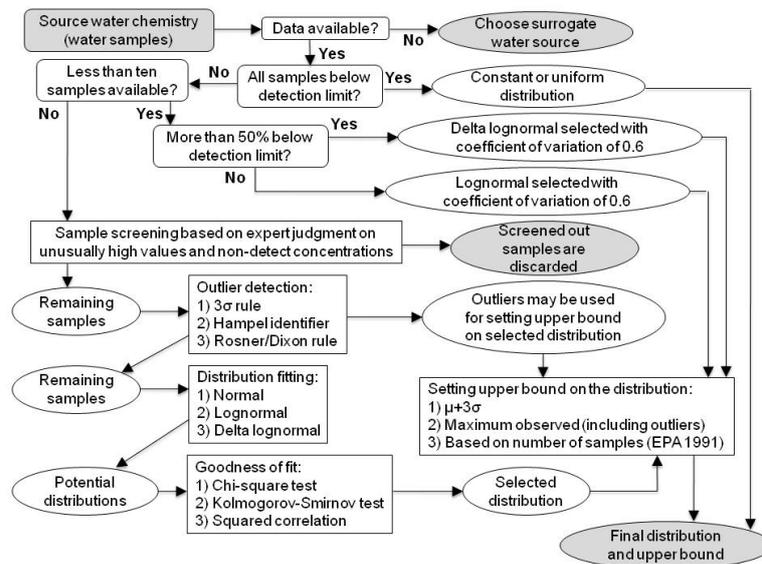


Figure 2. Processing of water sample concentration data for any given constituent.

Normal, lognormal and delta-lognormal distributions are fitted to the samples remaining from the screening process. One of these three distributions must then be selected as being the most appropriate to represent the samples, based on goodness of fit criteria. It is recommended that several criteria be used to evaluate the goodness of fit, with expert judgment providing the final decision on selected fitted distribution.

A second critical aspect of the distribution fitting process is the establishment of an upper bound on the selected fitted distribution, so that no values in the generated time series are higher than that selected threshold. Several methods for the calculation of this upper bound should be considered, such as the chi-square test, the Kolmogorov-Smirnov test and the squared correlation (Figure 2). 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 physically realistic. This threshold may be notably 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 underlying data upon which they are based.

The time series produced from source concentration probability distributions can be fed as inputs to a water quality model and used to predict in-stream concentrations at a set location in the receiving environment. This approach for developing time series purposely ignore temporal correlation with a water sources and cross-correlations between water sources. The intent is to allow the generation of a full array of combinations of concentration occurrences and flow conditions, in order to develop a full array of possible output concentrations in the aquatic environment from the modelling effort. Uncertainty is represented in the characteristics (i.e., mean and standard deviation) of these source concentration distributions, which induces uncertainty in the characteristics of the water quality model output. The following section gives a method to address uncertainty with this probabilistic modelling approach.

### 3. 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 to produce confidence limits around the predicted endpoint (Figure 3). One realization of inputs feeds one model simulation. The results of all model simulations are then used to generate confidence limits. The generation of the realizations would apply to all inputs to the model (i.e., processes, flow and water quality inputs), although this paper focuses particularly on the processing of the water quality inputs (i.e., the constituent concentrations associated with input streams or mine effluents).

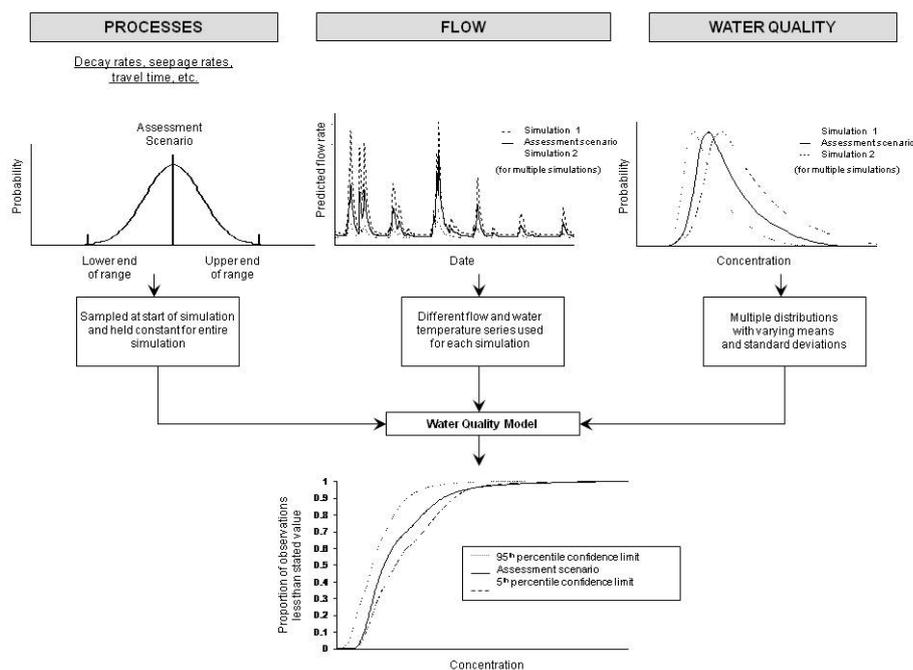


Figure 3. Conceptual diagram of the water quality uncertainty analysis model.

#### 3.1. Water Quality Inputs

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. Randomly-generated realizations of means and standard deviations (i.e., one pair for each water source and each water quality constituent) are therefore produced to describe uncertainty in the statistical distribution of the input data, assuming that the true mean and standard deviation of the underlying population are unknown. Each realization of means and standard

deviation are then used to generate time series of concentrations for each water sources serving as inputs for each Monte Carlo realization.

Consider that  $\bar{x}_{obs}$  and  $s_{obs}$  are the mean and standard deviation, respectively, 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 pairing of  $\bar{x}_{obs}$  with  $\bar{x}_{obs,i}$  and  $s_{obs}$  with  $s_{obs,i}$  can be established using standard statistical tests of hypothesis on the mean and standard deviation. It is assumed that the mean ( $\mu$ ) and standard deviation ( $\sigma$ ) of the data population are unknown. Furthermore, the sequence of observed concentrations is of size  $n_{obs}$ , which is also considered as 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, derived from Hines and Montgomery (1990):

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

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  $v_1 = v_2 = n_{obs} - 1$ . 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, derived from Hines and Montgomery (1990):

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

where variable  $T$  is assumed to follow a Student's distribution  $t_p(v)$ , of probability  $p$  and degree of freedom  $v = 2n_{obs} - 1$ . A value for  $T$  is obtained by a random generator for each realization, according to the Student's distribution. The value of  $n_{obs}$  is a key component of both Equations 1 and 2. The range of values for both  $\bar{x}_{obs,i}$  and  $s_{obs,i}$  is wider as  $n_{obs}$  is smaller.

### 3.2. Other Inputs to the Model

Examples of probabilistic formulations are given below for specific group of inputs:

- Process inputs: These inputs consist of model coefficients such as decay or settling rates. Variations of these inputs may be defined according to a standard distribution, such as a uniform and normal distribution. The central value of that distribution may consist of the average or median of values available in the literature for the associated parameters. The upper and lower bounds forced on the distribution may be based on extreme values found in the literature or other calibrated models or expert judgment.
- Flow inputs: If possible, flow time series should come from hydrological and hydrogeological models. Sets of time series would result from an uncertainty analysis conducted with these models, or an approach similar to that used for water quality inputs may be used as an alternative.

### 3.3. Example of Results

An example of results from the probabilistic modelling and uncertainty analysis presented in this paper was reported by Imperial Oil (2005) and is summarized herein. As part of an environmental impact assessment, this example evaluated constituent concentrations from the largest discharge end point (i.e., Muskeg River downstream North Pit Lake) of the Kearl Mine development in the Athabasca Oil Sands Region of Northern Alberta, Canada. The uncertainty analysis was undertaken for the closure phase of the mine development, when water is initially released from the mine area to the aquatic environment. HSPF (Bicknell et al. 1993), a hydrological and water quality model, was used for the predictions of concentrations as part of the uncertainty analysis and for the assessment of climate change (Section 4). The modelling effort incorporated uncertainty on water quality, processes, and flow inputs to the model.

Figure 4 illustrates the result of the uncertainty analysis, which assessed whole effluent acute and chronic toxicity, naphthenic acids, total dissolved solids and molybdenum. These constituents are signatures of mine waters and describe important water quality changes due to mine development. Furthermore, the five constituents represent different types of behaviours of constituents modelled, including degradable and non-degradable constituents.

The largest bands are predicted for acute and chronic toxicity (Figure 4), although their levels were predicted to be appreciably below aquatic life thresholds (i.e., 0.3 for acute toxicity and 1.0 for chronic toxicity, see AENV 1999). Relatively narrow confidence bands (i.e., the range of concentrations between the 5<sup>th</sup> and 95<sup>th</sup> percentile of predictions from the uncertainty analysis, as presented in Figure 4), coupled with conservative assumptions integrated into the water quality model formulation, produced a high level of confidence that predicted concentrations in the receiving environment would not be underestimated for this mine development.

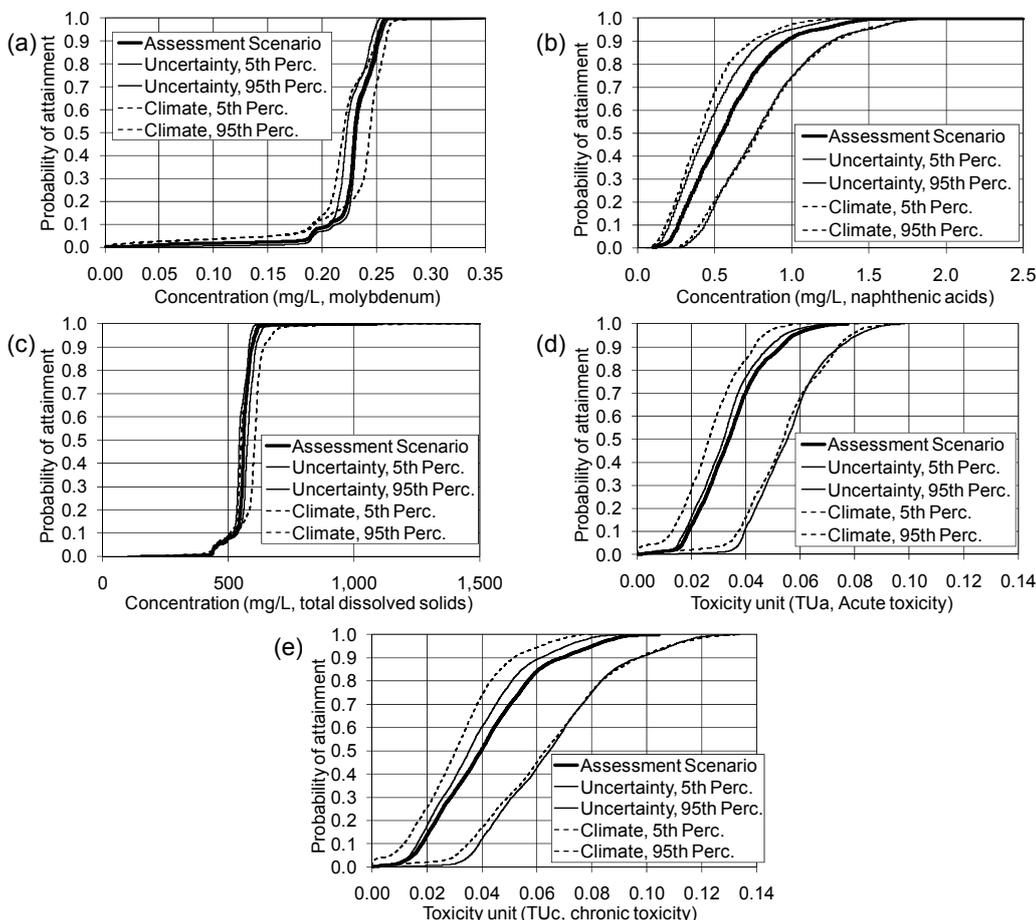


Figure 4. Distribution of concentrations with confidence bounds, as reported in Imperial Oil (2005).

#### 4. CLIMATE CHANGE ANALYSIS

The majority of the existing literature on climate change is focussed on how climate change will affect meteorological parameters, such as air temperature and precipitation. However, changes to air temperature and precipitation can lead to changes in infiltration, snow cover, evapotranspiration and, ultimately, streamflow, which could affect water quality (Chalecki and Gleick 1999; Murdoch *et al.* 2000). To estimate how climate change may affect in-stream water quality, the uncertainty analysis was repeated on two plausible climate scenarios in the example presented in Section 3.3 (Imperial Oil 2005). These simulations were completed assuming increased air temperatures (+ 3 °C), with varying effects on precipitation (i.e., one scenario involved increased amounts of precipitation by 5%, while the other involved decreased amounts of precipitation by 3%). The results were then used to develop confidence bands around model predictions as shown in Figure 4. These bands, when compared to those developed for the uncertainty analysis, show the possible shift in concentrations that may occur as a result of different climate conditions affecting the aquatic environment.

#### 5. CONCLUSION

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 characterisation described in this paper applies to probabilistic modelling of water concentrations, where probability distributions fitted from observed data are used to generate concentration time series as model inputs. The step-by-step process for fitting the distributions is based on established methods and is structured for the purpose of generating time series of water source concentrations within realistic ranges. The process uses both expert judgment and a systematic approach to generate concentrations that are conservative but realistic 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. Climate scenarios may also be incorporated into the uncertainty analysis to assess the effect of climate change on the modelling predictions. While the probabilistic model and uncertainty analysis presented in this paper has been developed from mine development applications, these approaches may also be applied to other water quality modelling studies that are intended to determine the fate of water quality constituents on a stream system or a waterbody.

## REFERENCES

- AENV (1999). Surface Water Quality Guidelines for Use in Alberta. Alberta Environment, Science and Standards Branch, Environmental Assurance Division, November 1999, Edmonton, AB, Canada, <http://environment.gov.ab.ca/info/library/5713.pdf>.
- Ayyub, B.M. and McCuen, R.H. (1997). Probability, statistics, & reliability for engineers. CRC Press, Boca Raton, Florida, USA.
- Bicknell, B.R., Imhoff, J.C., Kittle, J.L., Donigian, A.S. and Johanson, R.C. (1993). Hydrological Simulation Program – FORTRAN (HSPF): User’s Manual for Release 11.0. Report EPA/600/R-97/080, U.S. Environmental Protection Agency, Environmental Research Laboratory, Athens, GA, USA.
- Chalecki, E.L. and Gleick, P.H. (1999). A Framework of Ordered Climate Effects on Water Resources: A Comprehensive Bibliography. *Journal of the American Water Resources Association*, 35: 1657-1665.
- Hines, W.W. and Montgomery D.C. (1990) Probability and Statistics in Engineering and Management Science, Third Edition. John Wiley & Sons, New York, NY, USA.
- Imperial Oil (2005). Kearl Oil Sands Project - Mine Development. Volumes 1 to 9, submitted by Imperial Oil Resources Ventures Limited, to Alberta Energy and Utilities Board and Alberta Environment, Calgary, AB, Canada, <http://www.ceaa.gc.ca/050/document-eng.cfm?document=18778>.
- Murdoch, P.S., Baron, J.S. and Miller, T.L. (2000). Potential Effects of Climate Change on Surface-Water Quality in North America. *Journal of the American Water Resources Association*, 36: 347-366.
- Shell (2005). Muskeg River Mine Expansion Project Application and Environmental Impact Assessment. Volumes 1, to 4, submitted by Shell Canada Limited, to Alberta Energy and Utilities Board and Alberta Environment, Fort McMurray, AB, Canada, <http://www.ceaa.gc.ca/050/document-eng.cfm?document=16923>.
- Shell (2007). Jackpine Mine Expansion & Pierre River Mine Project Application and Environmental Impact Assessment. Volumes 1 to 5, submitted by Shell Canada Limited, to Alberta Energy and Utilities Board and Alberta Environment, Calgary, AB, Canada, <http://www.ceaa.gc.ca/050/document-eng.cfm?document=46923>.
- Total (2010). Joslyn North Mine Project Additional Information Project Update. Submitted by Total E&P Joslyn Ltd., to Alberta/Canada Joint Review Panel, Fort McMurray, AB, Canada, <http://www.ceaa.gc.ca/050/document-eng.cfm?document=41356>.
- U.S. EPA. (1991). Technical support document for water quality-based toxics control. United States Environmental Protection Agency, EPA report number 50512-90-001.