Application of dynamic parametric sensitivity analysis for identifying contaminant sources in a watershed

H. Manolopoulos¹, B. E. Halbert¹, C. R. Paton² & J. M. Scharer³
¹SENES Consultants Limited, Richmond Hill, Ontario, Canada
²Cameco Corporation, Saskatoon, Saskatchewan, Canada
³University of Waterloo, Waterloo, Ontario, Canada

Abstract

Lakes of the Fulton Creek watershed in Northern Saskatchewan, Canada, possess constituents associated with uranium mining. Although mining activity ceased over 24 years ago and site decommissioning was largely completed in 1985, the discharge from this former tailings management watershed continues to have elevated levels of three specific constituents, namely dissolved radium-226, selenium and uranium, as well as total dissolved solids (TDS). The objective of the current study was to identify and rank the sources of constituents in the watershed. Constituent dispersion modelling was carried out employing a proprietary computer code called LAKEVIEW (developed by SENES Consultants Limited). Using the Metropolis-Hastings algorithm of a Markov chain computational procedure, parameter calibration was performed at two locations in the watershed with 25 years of more or less regular water quality and occasional sediment monitoring data. The calibrated model captured the time dependent trends for all variables. Employing the Metropolis-Hastings parameter sampler, dynamic parametric (source) sensitivity analysis was carried out for a 300-year period. The sensitivity responses showed strong temporal variability. Therefore, the normalized gradient sensitivity values were integrated and averaged over time for use as the measure of constituent source loads. Sediments in two lakes were shown to be both the current and future principal sources of all constituents. An adjacent surface tailings area was identified as a relatively minor contributor of all constituents but uranium.

Keywords: constituent source modelling, parameter distribution, constituent flux apportionment, normalized gradient sensitivity.
1 Introduction

The Fulton Creek watershed (15.75 km$^2$ area), a former uranium tailings management site, is situated East-Northeast of Beaverlodge Lake in Northern Saskatchewan, Canada and contains six lakes (Fookes, Marie, Meadow, Unnamed, Minewater and Greer lakes) affected by uranium mining (see Figure 1). The main flow of Fulton Creek is from Fookes Lake to Marie Lake, Meadow Lake and Greer Lake from where Fulton Creek drains into Fulton Bay on Beaverlodge Lake. A second, smaller branch flows from Minewater Lake east to Unnamed Lake, which drains into Meadow Lake. Beaverlodge Lake is a large sized water body of approximately 67 km$^2$ surface area and 19 m mean depth. It supports fresh water sport fisheries of lake white fish, lake trout and northern pike.

Previous studies identified dissolved radium-226, selenium, total dissolved solids (TDS) and uranium as constituents of concern. Constituent sources in the watershed include mill tailings deposited under water in Fookes and Marie lakes, and residual mine water slimes, mill tailings and barium-radium-sulphate co-precipitate sludge in sediments of Minewater Lake, Unnamed Lake, Meadow Lake and Greer Lake. Tailings deposited in Fookes and Marie lakes formed beaches at their points of discharge. The partially unsaturated tailings beach, which forms a delta in Fookes Lake, is sufficiently large to warrant separate evaluation.

Parametric sensitivity analysis is a study of system model response to variability of parameters. Traditionally, it is carried out if a parameter value is poorly known or uncertain. Most previous applications have been in risk analysis, financial planning, and model based policy assessment studies. The
principal application of parametric sensitivity analysis for watersheds has been to assess uncertainty and degree of confidence in existing data or models [1–3], hydrological parameter optimization [4], source and process identification in watersheds [5] and as an aid in watershed-scale water quality management [6]. In this paper, parametric sensitivity in response to perturbations in constituent fluxes from lake sediments and contiguous tailings was performed. To date, only a handful of published reports have pertained to formal sensitivity analysis related to surface water quality modeling in a watershed. The objective of the current study was to identify and rank the major sources of radium-226, selenium, TDS and uranium in the watershed. An immediate additional application of parametric sensitivity analysis is that, in combination with other statistical methods, it may also be used to outline an optimal remediation strategy.

2 The LAKEVIEW model

To address both current and future conditions, watershed dispersion modelling was carried out to predict levels of radium-226, selenium, TDS and uranium in water and sediments. These constituents were modeled using a proprietary computer code called LAKEVIEW that was developed by SENES Consultants Limited specifically for assessing mine waste effects on the environment. The code may be run in either a deterministic or a probabilistic manner and has been applied on numerous occasions in Canada, including in Northern Saskatchewan, to assess the impact of uranium and other mining operations.

LAKEVIEW is essentially a three-dimensional water quality transport and constituent speciation model for lakes and rivers in a watershed. Important processes incorporated into the LAKEVIEW include horizontal (lateral) and vertical transport of dissolved species, chemical and biochemical reactions, settling of particulate matter, and sediment exchange. Transport between segments (water column–sediment, for example) is computed by solving ordinary differential equations while constituent speciation within a segment is computed by thermodynamic stability criteria for equilibrium processes and kinetic considerations [7, 8]. Model output for lakes includes surface water, hypolimnion and sediment porewater concentrations of dissolved constituents, constituents adsorbed to suspended matter and solid phase concentrations in the sediment. In accordance with the modified Tessier’s extraction test [9], the constituents in the sediment are partitioned as being adsorbed onto calcite, quartz and organics (ion-exchangeable), complexed with iron hydroxide (reducible), embedded in sulphide matrix (oxidizable) as well as being present as solid solution, pure mineral and recalcitrant (un-extractable) fractions.

2.1 Parameter estimation

Data based environmental modelling often requires parameter estimates by some statistical methods. Markov Chain Monte Carlo (MCMC) methods such as the Metropolis-Hastings algorithm have been found to be powerful tools, especially when some prior knowledge about the parameters is available. The algorithm
has been employed to solve difficult, non-linear parameter estimation problems arising in various disciplines [10–12]. It involves one parameter at a time sampling and iteration to find the posterior parameter distribution.

The present application of the Metropolis-Hastings algorithm can be summarized as follows. First parameter “i” and its value ($\theta_i$) is assigned by random (Monte Carlo) draw from the prior distribution. A symmetric, triangular prior distribution based on minimum, mean and maximum parameter values offers a simple and convenient means for sampling the parameter space. It is important that the prior distribution is sufficiently wide to avert accidental exclusion of probable values. Using the sampled value, the predicted concentrations are computed by LAKEVIEW. The predictions are then compared with measurements by calculating the likelihood function ($p(y|\theta_i, \theta)$) by the normalized sum of squares between the observations and predictions:

$$p(y|\theta_i, \theta) = \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left(-\frac{\sum[y((\theta_i, \theta, t_k) - y(t_k)]^2}{2 \cdot n \cdot \sigma^2}\right)$$  \hspace{1cm} (1)

where:

- $\theta$ = vector of parameter values
- $\theta_i$ = most recently drawn parameter value
- $n$ = number of observation/prediction pairs
- $\sigma$ = standard error of observations
- $y(t_k)$ = observed water quality at time, $t_k$
- $y((\theta_i, \theta, t_k)$ = predicted water quality at time, $t_k$

It has been shown [10] that for a symmetric prior distribution such as the equilateral triangular distribution, the parameter acceptance criterion becomes the posterior probability ratio “r” defined as follows:

$$r = \frac{P(\theta_i | y)}{P(\Theta_i | y)}$$  \hspace{1cm} (2)

where:

- $\theta_i$ = proposed (most recently chosen) i-th parameter
- $\Theta_i$ = currently accepted i-th parameter value

It is evident from equation (2) that the probability of acceptance “r” in equation (2) is simply the ratio of likelihood functions evaluated with the newly sampled and currently accepted values of the parameters. The value of “r” is used to define the probability of acceptance criterion, $\xi(\theta, \Theta)$, for $\theta_i$ as follows:

$$\xi(\theta, \Theta) = \min(1, r)$$  \hspace{1cm} (3)

Finally, $\xi(\theta, \Theta)$ is compared with a random number ($\zeta$) drawn from a uniform distribution, $U(0,1)$. The proposed value, $\theta_i$ is accepted if $\xi(\theta, \Theta) \geq \zeta$. If accepted, $\theta_i$ becomes the current value $\Theta_i$ and the selection process is repeated. Accepted $\theta_i$’s naturally form a Markov chain; they are “binned” for generating the posterior probability density function, $P(\theta_i | y)$, and calculating its statistical properties such as the median, mean, and variance.
2.2 Parametric sensitivity analysis

An important objective of sensitivity analysis in environmental modelling is the quantitative assessment of parameter or process variation on some objective function. A number of different sensitivity measures (algebraic, integral, gradient) have been employed [2]. The current application involves the identification of constituent sources for which the water quality model variation is maximal. Since the absolute value of the sensitivity response varies substantially with respect to both constituent type and time, the normalized gradient sensitivity, also known as the logarithmic sensitivity, was used in this study. The “instantaneous” normalized gradient sensitivity of the objective function at any time, \( t \), in response to perturbation of the \( i \)-th parameter can be expressed as follows:

\[
\lambda(C, \theta_i) = \frac{\delta C/C}{\delta \theta_i} = \left[ \frac{C(\theta_i, t) - C(\theta_i^b, t)}{C(\theta_i^b, t)} \right] \left[ \frac{C(\theta_i, t)/C(\theta_i^b, t)}{C(\theta_i^b, t)/\theta_i^b} - 1 \right] - 1
\]  

(4)

where:

- \( \lambda(C, \theta_i) = \) normalized sensitivity gradient
- \( C(\theta_i, t) = \) response of the objective function with parameter \( \theta_i \) at time, \( t \)
- \( C(\theta_i^b, t) = \) response of the objective function to baseline parameter value \( \theta_i^b \)
- \( \theta_i = \) sample value of the i-th parameter
- \( \theta_i^b = \) baseline value of the i-th parameter

The objective function responses in the present case are the concentrations of radium-226, selenium, TDS, and uranium in Fulton Creek at the inlet to Fulton Bay. Baseline parameter values in equation (4) are the mean value of the posterior probability density for each parameter. The sensitivity value given by equation (4) quantifies the relative change in a calculated variable that is evoked by a relative change in a system parameter.

To evaluate the impact of a parameter on the objective function, the value of a given parameter is randomly generated by Monte Carlo draw within the distribution space defined by the inverse of the Metropolis-Hastings posterior probability density. The inverse is generated numerically as a rational function using MATLAB\textsuperscript{®} Curvefit toolbox. After parameter sampling, predicted concentrations for a specified time range are computed by the LAKEVIEW computer model. Parameters are selected and the concentration profiles solved repeatedly for 1500 trials. After completing the iterations, the “instantaneous” sensitivity, \( \lambda(C, \theta_i) \), is integrated to compute the time averaged normalized gradient sensitivity:

\[
\Lambda_{\text{ave}}(C, \theta) = \frac{1}{TN} \sum_{j=1}^{N} \int_{0}^{T} \frac{[C(j, \theta, t)/C(j, \theta^b, t)] - 1}{(\theta_j/\theta^b) - 1} \, dt
\]  

(5)

where:

- \( \Lambda_{\text{ave}}(C, \theta) = \) time averaged normalized gradient sensitivity
- \( T = \) time period of simulation
- \( N = \) total number of trials (iterations) for a given parameter
3 Results and discussion

Model parameter calibration was carried out for Fookes and Greer lakes for a 25-year period (1981 to 2005) using concentration data for the principal constituents of concern, namely dissolved radium-226, selenium, TDS and uranium. Parameters to be estimated included: sediment to water column mass transfer coefficients ($K_L$), reaction rate constants ($k_X$), overall sediment/porewater partition coefficients ($K_D$), partition coefficients for settling solids in the water column ($K_{DS}$), and external load factors ($f_L$) for the tailings areas. Parameter estimates were obtained through the application of the Metropolis-Hastings procedure using the squared difference between predicted and observed surface water concentrations as criterion. A summary of mean, median and the standard error of the parameter values is presented on Table 1.

The average external load factors for the Fookes Lake tailings delta and Minewater Lake (also treated as an external source) were 1.02 and 1.42, respectively. In most cases, the mean and median parameter values were close indicating symmetrical distribution. Generally, the “lake-to-lake” variability of most parameter estimates was surprisingly small. A noteworthy exception was the reaction rate constant ($k_x$) for selenium, particularly in Fookes Lake, which is significantly higher than the reaction rate constants for the other constituents. Evidently, selenium is being reduced from the highly mobile Se(VI) to the much less soluble Se(II) oxidation state. The lower reaction rate constants for radium-226 and uranium (as well as TDS, not shown) represent probable “sinking out” of the constituents from the sediment exchange zone as fresh, less contaminated solids are deposited over the more heavily contaminated zone. The high $K_D$ values for uranium and selenium indicate significant adsorption to iron mineral (iron hydroxide, hematite) surfaces and organic matter in both lakes [13, 14]. Furthermore, the statistical analysis resulted in nearly identical mean $K_D$ values in each of the two lakes. In fact, model improvements in the probabilistic sense were too small to be accepted by the Metropolis-Hastings discriminator function.

Table 1: Calibrated parameter estimates for the selected constituents in Fookes Lake and Greer Lake.

<table>
<thead>
<tr>
<th>Constituent</th>
<th>Parameter</th>
<th>Fookes Lake</th>
<th></th>
<th></th>
<th>Greer Lake</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>mean</td>
<td>median</td>
<td>Std. Dev.</td>
<td>mean</td>
<td>median</td>
<td>Std. Dev.</td>
</tr>
<tr>
<td>Selenium</td>
<td>$K_L$ (m yr$^{-1}$)</td>
<td>1.75</td>
<td>1.76</td>
<td>0.195</td>
<td>2.45</td>
<td>2.49</td>
<td>0.766</td>
</tr>
<tr>
<td></td>
<td>$k_X$ (yr$^{-1}$)</td>
<td>0.22</td>
<td>0.21</td>
<td>0.021</td>
<td>0.11</td>
<td>0.12</td>
<td>0.051</td>
</tr>
<tr>
<td></td>
<td>$K_{DS}$ (m$^3$ kg$^{-1}$)</td>
<td>5.87</td>
<td>6.01</td>
<td>0.779</td>
<td>6.89</td>
<td>6.84</td>
<td>1.001</td>
</tr>
<tr>
<td></td>
<td>$K_D$ (m$^3$ kg$^{-1}$)</td>
<td>4.82</td>
<td>4.98</td>
<td>0.753</td>
<td>4.82</td>
<td>4.83</td>
<td>1.018</td>
</tr>
<tr>
<td>Ra-226</td>
<td>$K_L$ (m yr$^{-1}$)</td>
<td>1.76</td>
<td>1.80</td>
<td>0.221</td>
<td>2.17</td>
<td>2.27</td>
<td>0.765</td>
</tr>
<tr>
<td></td>
<td>$k_X$ (yr$^{-1}$)</td>
<td>0.05</td>
<td>0.04</td>
<td>0.011</td>
<td>0.05</td>
<td>0.05</td>
<td>0.015</td>
</tr>
<tr>
<td></td>
<td>$K_{DS}$ (m$^3$ kg$^{-1}$)</td>
<td>2.51</td>
<td>2.56</td>
<td>0.048</td>
<td>2.44</td>
<td>2.57</td>
<td>0.769</td>
</tr>
<tr>
<td></td>
<td>$K_D$ (m$^3$ kg$^{-1}$)</td>
<td>3.77</td>
<td>3.96</td>
<td>1.151</td>
<td>3.70</td>
<td>3.76</td>
<td>0.586</td>
</tr>
<tr>
<td>Uranium</td>
<td>$K_L$ (m yr$^{-1}$)</td>
<td>1.67</td>
<td>1.69</td>
<td>0.201</td>
<td>2.39</td>
<td>2.39</td>
<td>0.582</td>
</tr>
<tr>
<td></td>
<td>$k_X$ (yr$^{-1}$)</td>
<td>0.02</td>
<td>0.01</td>
<td>0.005</td>
<td>0.03</td>
<td>0.03</td>
<td>0.011</td>
</tr>
<tr>
<td></td>
<td>$K_{DS}$ (m$^3$ kg$^{-1}$)</td>
<td>5.71</td>
<td>5.82</td>
<td>0.935</td>
<td>7.02</td>
<td>7.51</td>
<td>0.534</td>
</tr>
<tr>
<td></td>
<td>$K_D$ (m$^3$ kg$^{-1}$)</td>
<td>2.08</td>
<td>2.52</td>
<td>0.848</td>
<td>2.57</td>
<td>2.78</td>
<td>0.523</td>
</tr>
</tbody>
</table>
Between 2001 and 2005, an extensive water quality monitoring and sediment sampling program was carried out in the Fulton Creek watershed. Current source load contributions were calculated using this database. By assuming steady state for short time periods, the constituent source load balance can be written as follows:

\[ Q_{j,\text{out}} \left[ C_{j,\text{out}}^i - C_{\text{back}}^i \right] = \sum_{k} L_{j,k}^i + \varepsilon_j \]  (6)

where:
- \( Q_{j,\text{out}} \) = average annual outflow from lake “j”
- \( C_{j,\text{out}}^i \) = average concentration of constituent “i” in outflow from lake “j”
- \( C_{\text{back}}^i \) = background concentration of constituent “i”
- \( L_{j,k}^i \) = annual average load from source “k” affecting water quality in lake “j”
- \( \varepsilon_j \) = error in load estimate in outflow from lake “j”

The constituent loads at the outflow point “j” include all potential constituent loads from upstream sources. Constituent concentrations for modelling were annual averages (6-12 observations per year) over the five-year period. The background concentrations were from monitoring data in a neighbouring unaffected watershed. In equation (6), the source loads were regarded as unknowns. However, the resulting set of algebraic equations was underdetermined (more unknowns than equations). For this reason, the Minewater Lake load was evaluated independently by assuming it to be equal to the inflow to Unnamed Lake. Using iterative quadratic programming, the algebraic equations were solved simultaneously to give the “best” estimate of the source loads by minimizing the sum of the squared error terms (\( \min \sum \varepsilon_j^2 \)).

The apportionment of constituent source contributions is shown in Figure 1. Fookes Lake sediments were found to be the main source of uranium (53%), selenium (62%) and TDS (56%). In the case of radium-226, however, Greer Lake sediment was identified as the primary source contributing about 40% of the total load. Fookes Lake and Greer Lake sediments together accounted for 76% of the radium-226 export into Fulton Bay. The third most important source, Marie Lake sediment, was assessed to contribute 10% to 17% of the load of each of the four constituents. The combined contributions from the Fookes Delta tailings, Minewater and Unnamed lake sediments were estimated to be quite small (less than 10%) for all constituents excepting uranium, which was moderately affected (approximately 17%) by contributions from the Fookes Delta deposits.

Clearly, steady state source analysis is unsuitable for predicting long-term constituent source loads in the future, since the loads are expected to change with time. For predictive purposes, source load terms can be expressed as either internal (sediment) or external (adjacent tailings) fluxes. In the LAKEVIEW model, constituent flux from the sediment is given by:

\[ J_{j}^i = K_{L} \left( C_{j,s}^i - C_{j,w}^i \right) \]  (7)
where:

\[ J^i_j = \text{constituent “}i\text{” flux from lake “}j\text{” sediment (kg m}^{-2}\text{ y}^{-1}) \]

\[ K_L = \text{sediment to water column mass transfer coefficient (m y}^{-1}) \]

\[ C^i_{j,s} = \text{sediment porewater concentration of constituent “}i\text{” in lake “}j\text{” (kg m}^{-3}) \]

\[ C^i_{j,w} = \text{surface water concentration of constituent “}i\text{” in lake “}j\text{” (kg m}^{-3}) \]

The constituent flux from an external (tailings) source can be expressed analogously on a unit surface area basis (Flux = \( f_L J^i_E \)). Since the flux terms are linear functions of the parameters (either \( K_L \) or \( f_L \)) and the constituent load is directly proportional to the flux, it follows that any variability in the parameter values is reflected by the variability in the objective function. It may be argued that the maximum sensitivity response corresponds to the most important source.

The modelled system consisted of Fookes, Marie, Unnamed, Meadow and Greer lakes in the Fulton Creek watershed. Future “baseline” external constituent loads from the Fookes Lake tailings delta and Minewater Lake were predicted using the RATAP program [8], which was developed to assess the effects of decommissioning alternatives on tailings management facilities. The individual estimates of the calibrated parameters (\( k_x, K_D \) and \( K_{DS} \)) were averaged on a “per constituent” basis across all water bodies. The sediment-to- water column mass transfer coefficient (\( K_L \)), however, was averaged on a “per Lake” basis. Due to similarities in sediment composition, density, and porosity, the calibrated Fookes Lake \( K_L \) was applied to Marie Lake while the Greer Lake \( K_L \) value to Unnamed and Meadow lakes. The parameters in each lake were sampled

one at a time using the Monte Carlo procedure as described previously. The selected parameter value was then used as an input to the LAKEVIEW program. The predicted time series of water quality in Fulton Creek at the inlet to Fulton Bay was chosen as the objective function response to evaluate the overall effect of parameter variability. The normalized gradient sensitivity response \( [\lambda(C, \theta_i)] \) at the Fulton Bay inlet was calculated according to equation (6).

The normalized uranium gradient sensitivity response to sediment flux variability in Fookes Lake is shown in Figure 3 for a 300-year simulation period. The figure summarizes 96 trials. Similar responses were generated for all four constituents at 7 locations. A positive response value implies that the sediment remains a source of the constituent in question. It is noteworthy that all response values are either positive during the entire simulation period or decline to zero after 150 years or more. All sensitivity response patterns exhibit remarkable temporal variability. In a given lake, the trends are somewhat similar, but clearly not identical for each constituent. It appears that the response is non-linear with respect to the sediment flux variation (e.g. \( \delta C / \delta \theta_i \neq \text{constant} \)). The non-linearity is caused by the hyperbolic dependence of the equilibrium position between surface water and sediment porewater on the mass transfer coefficient (\( K_L \)). Consequently, the curves are dispersed (\( K_L \)-dependent) during at least a certain portion of the simulation period. In contrast, the normalized gradient sensitivity response was independent of the tailings load parameter (\( f_L \)) variations although the response curves were also strongly time varying. In this case, the normalized sensitivity response collapsed into an essentially single curve representing all possible parameter values (not shown).

To give an overall assessment of the long-term significance of constituent sources, the normalized gradient sensitivity curves were integrated and averaged over the 300 years (see equation (5)). Although sensitivity analysis has been applied previously for process and model identification, to our knowledge, the

Figure 3: Normalized gradient sensitivity response to uranium flux variation in Fookes Lake.
The time-averaged integrals of the gradient sensitivity responses of the four constituents are shown in Figure 4. The sensitivity values are a relative measure of the significance of a particular constituent source for the simulation time period. Evidently, Fookes Lake sediments continue to remain the dominant source of uranium, selenium and TDS. The relative contribution of Marie Lake sediments as the source of radium-226 increases with time and the radium-226 flux is expected to surpass the flux from Greer Lake as the barium/radium sulphate co-precipitate in Greer Lake sediment is exhausted. Beached tailings in the Fookes Lake delta are expected to become the second most important source of TDS. This is attributable to the oxidation of hydraulically unsaturated pyritic minerals that will continue to produce soluble sulphates and bicarbonates in the future. As shown in Figure 4, the decommissioned tailings area is a minor source of radium-226, selenium and uranium. This reflects the effectiveness of the remedial efforts taken some 30 years ago.


4 Conclusions

The Fulton Creek watershed contains a decommissioned surface tailings area (delta at Fookes Lake) and underwater lake tailings disposal areas (Fookes and Marie lakes in particular). Sediments of two lakes (Fookes and Greer) were
identified as current major sources (70% to 86%) of exported radium-226, selenium, TDS, and uranium. Using time averaged, normalized gradient parametric sensitivity as the measure, these sediments were predicted to remain the major sources for the next 300 years, although the relative importance of Greer Lake as the source of radium-226 and selenium is expected to decline. The tailings delta in Fookes Lake was shown to be a minor current and future source of all constituents accounting for 3% (selenium) to 18% (uranium) of the total load at the outlet of the watershed. Continued oxidation of pyritic tailings in the Fookes delta was predicted to contribute to future TDS loads. The normalized gradient parametric sensitivity method provided a convenient and unbiased comparison of the constituent sources for extended time periods. This study has also shown the importance of sampling the entire parameter distribution space particularly if the objective function response with respect to the parameter is expected to be non-linear.

References


