RioTinto

KMP SO_2 EEM Program – Technical Memo W04

Summary Report on Power Analyses

March 2016

Prepared for:

Rio Tinto Alcan
1 Smeltersite Road, P.O. Box 1800,
Kitimat, BC, Canada V8C 2H2

Prepared by:

ESSA Technologies Ltd.
Suite 600 – 2695 Granville St.
Vancouver, BC, Canada V6H 3H4
(Using data provided by Rio Tinto Alcan)

Table of Contents

Table of Figures

Table of Tables

1 Introduction

As part of the Kitimat Modernization Project Sulphur Dioxide Environmental Effects Monitoring Program (EEM, ESSA Technologies et al. 2014b), we conducted power analyses to assess our ability to correctly detect changes of interest in water chemistry in the sensitive EEM lakes. That is, we asked how confidently will the established monitoring program be able to identify lakes that have exceeded the thresholds for changes in ANC, SO_4^{2-}, and/or pH?

1.1 The Environmental Effects Monitoring Program

The aquatic ecosystem component of the EEM focuses on the monitoring of 7 acid-sensitive lakes and 4 acid-insensitive lakes[1]. These lakes are monitored in order to be able to detect any potential changes in the Key Performance Indicators (KPIs) identified in the EEM (see Table 14 of the EEM), which include ANC, SO_4^{2-}, and pH. Table 27 in the EEM reports the critical thresholds for ANC, SO_4^{2-}, and pH for each of the lakes. One of the objectives of the EEM is to determine whether any of these thresholds are exceeded in any of the lakes subsequent to the implementation of KMP.

Initial results from continuous pH monitors installed in 3 of the EEM lakes indicate that variation in pH within these lakes during the fall index period is substantially greater than was assumed during the collaborative development of the EEM Plan by Rio Tinto Alcan (RTA), the BC Ministry of Environment (MOE), ESSA Technologies (ESSA), the Haisla First Nation, and Limnotek. Year-to-year variation in ANC, SO_4^{2-} and pH is also substantial.

High natural variability in lake chemistry reduces the power to detect a true change in the condition of a lake. Appropriate management decisions depend on a concrete understanding of the power to detect a true change of > 0.3 pH units, which could potentially affect aquatic biota. Power analyses can be used to quantify our ability to correctly determine whether or not a true change of > 0.3 pH units has occurred given the natural variability in lake chemistry. Power analyses can also be used to quantify our ability to correctly determine that a change in lake chemistry has not occurred.

One of the overall goals in the implementation of the aquatic section of the EEM is to correctly determine whether the KPIs have or have not been exceeded. The EEM then specifies the actions required based on those results, which may include more detailed monitoring or mitigation actions to reduce emissions.

---

[1] The 4 insensitive lakes are not considered in the present power analyses because they are not predicted to experience significant changes in lake chemistry. These insensitive lakes are not true controls for the sensitive lakes because: 1) they have a different level of acid-sensitivity, and 2) they are also within the KMP deposition zone and therefore can be expected to show changes in sulphate concentrations, though not in ANC or pH. However, ongoing monitoring of the insensitive lakes is still very valuable for other aspects of the EEM evidentiary framework (ESSA Technologies et al. 2015). The October 2015 sampling program included 3 true control lakes outside of the KMP deposition zone, lakes that were also sampled in October 2013 as part of the Kitimat Airshed Assessment (ESSA et al. 2014a).

1.2 Overall Guiding Questions

The following four questions have guided our overall approach to the power analyses. The first two questions represent the essence of power analyses, prospectively assessing the probability of the monitoring system generating erroneous conclusions about potential changes in lake chemistry. The third question is about potential improvement to the lake monitoring design to increase the power to correctly detect changes. The fourth question focuses on whether or not lake chemistry changed as a result of smelter emissions, an issue to be addressed annually.

Power Analysis Questions

1. What is the probability of falsely concluding that the KPI has been exceeded when there has not been a true change in the underlying condition of the lake?
    - This is commonly called a "false positive"
    - In statistics this is called a "Type 1 error"
    - e.g., observed changes are within the range of natural variability but are mistakenly attributed to a fundamental change in lake chemistry due to emissions from the smelter, even though such changes have not occurred

    Relevance of this question: A false positive would trigger increased monitoring to confirm that the change was real. False positives (if they occurred in more than 2 lakes of moderate to high importance) might trigger mitigation actions that were unnecessary because there had not been a true change in the underlying condition of these lakes. Such mitigation actions could include substantial economic costs that were not required.

2. What is the probability of falsely concluding that the KPI has not been exceeded when there has in fact been a true change in the underlying condition of the lake?
    - This is commonly called a "false negative"
    - In statistics this is called a "Type 2 error"
    - e.g., smelter emissions have caused biologically significant changes in lake chemistry, but these changes have been masked by natural variability

    Relevance of this question: A false negative would fail to trigger mitigation actions that are in fact necessary, which would result in adverse ecological impacts because there had truly been a change in the underlying condition of the lake.

Monitoring Program Design Question

3. Given the level of variability observed in pre-KMP monitoring data, an expanded literature review, and the results of these power analyses, is the current design of the EEM water chemistry monitoring program appropriate? Or, does the monitoring program need to be modified to better determine whether or not KMP emissions have caused KPI thresholds to be exceeded?

Post-KMP Monitoring Question

4. What is the probability that Lake X has exceeded the EEM KPI thresholds?

1.3 Specific Project Questions

Within the context of the four guiding questions, we conducted a series of power analyses to answer the following questions about the aquatic monitoring portion of the EEM:

1. How well will we be able to detect changes in water chemistry indicators that are greater than the identified limits (i.e., KPI thresholds)?
2. How many years of monitoring are needed to be able to confidently detect these changes?
3. What is the benefit of collecting multiple samples within a year?
4. What is the benefit of monitoring control lakes?
5. Is our ability to confidently detect changes in lake chemistry improved by using multiple metrics rather than just pH, the KPI selected in the EEM for its biological relevance?
6. Would increasing the threshold for detecting a change in lake chemistry reduce the chance of a false positive without significantly affecting our ability to correctly detect a true change in lake chemistry?"

1.4 Impact Pathways to be Detected by Monitoring

The impact pathways of concern are as follows:

1. increased sulphur dioxide emissions from KMP could cause
2. increased sulphate deposition within a lake's watershed that could cause
3. transport of acidifying ions into the lake (if watershed neutralization is limited), which could cause
4. acidification of the lake (if the acid neutralizing capacity of the lake is insufficient to buffer the acidic deposition), which could cause
5. adverse impacts on the aquatic ecosystem (if acidification results in a significant change in pH).

This is the pathway of concern that would result in acidification, but neutralization of acidifying ions could happen in either the watershed or lake, depending on geochemical attributes. The 7 sensitive EEM lakes are vulnerable to proceeding fully along this pathway, or at least through steps 1 to 4, whereas most of the lakes assessed in the SO_2 Technical Assessment Report (STAR) were not vulnerable to acidification (ESSA Technologies et al. 2013).

There are three primary metrics of interest for measuring changes in lake chemistry due to acidification– ANC (i.e., Gran ANC), SO_4^{2-} concentration and pH. Secondary metrics are also monitored[2]. pH is the primary metric because it is most closely associated with ecological responses. ANC is a measure of the acid neutralizing capacity of the lake. ANC varies positively (but non-linearly) with pH. Both ANC and pH decrease when lakes acidify. SO_4^{2-} is a corroborating indicator of this acidification pathway. On its own, increases in SO_4^{2-} concentration would show that SO_4^{2-} deposition and transport through the watershed is occurring; however, if there is no associated decrease in pH and ANC, then it indicates that neutralization of the acidifying ions is occurring, in both the watershed and the lake. Conversely, if pH and ANC are decreasing in the lake with no associated increase in SO_4^{2-}, it indicates that the acidification is being driven by

---

[2] Secondary metrics of interest include aluminum, dissolved organic carbon, and the sum of base cations minus strong acid anions (a derived metric).

another mechanism (i.e., not from KMP-related sulphur emissions). These potential scenarios are explored in the documentation of the Evidentiary Framework in the EEM (ESSA Technologies et al. 2014b).

In the EEM, the critical threshold for the aquatic ecosystems' KPI is a decrease in pH of 0.3 units or greater. The EEM also proposes critical thresholds for ANC and SO_4^{2-} (corresponding with a decrease in pH of 0.3 units), consistent with the multi-metric approach of the Evidentiary Framework in the EEM.

The EEM focuses on changes to individual lakes. During the development of the EEM, the BC Ministry of Environment (BC MOE) stressed that the public was concerned about the potential impacts to individual lakes, which led to a lake ranking process (see lake ratings in Appendix D of the EEM). One can assess changes over the entire group of lakes to infer the regional extent of impacts, but as emphasized by BC MOE, it is even more important to detect ecologically relevant impacts on individual lakes.

The EEM has limited scope for improving the monitoring design. In general, the power of a monitoring design can be improved by adding more locations (lakes), collecting more samples within a lake (higher sampling frequency), or improving measurement quality by using better equipment and/or sampling techniques. However, these options are severely restricted in the present case. It is not possible to measure additional lakes because we are already monitoring all of the lakes of interest as identified in both the STAR (ESSA et al. 2013) and subsequent Kitimat Airshed Assessment (ESSA et al. 2014a). Collecting additional samples within each year is only logistically feasible at three road-accessible lakes where such intensive monitoring is already occurring (and is evaluated in this study). The actual measurements are already being conducted by highly trained technical staff with decades of collective experience, using high quality field equipment and accredited analytical laboratories. Some error in pH measurements is unavoidable due to the difficulty of precisely measuring pH in lakes of low ionic strength (Covington et al. 1985).

2 What We Did – The Analyses

2.1 Introduction to Power Analyses

Power analyses are about determining the probability of correctly detecting:

a change in an **indicator**,
in terms of a particular **individual or group**,
of an explicit magnitude (**effect size**),
over a specified **period of time**,
relative to a defined **baseline**,
given the existing **variability** in the data.

Indicators – We evaluated changes in Gran ANC, SO_4^{2-} and pH.

Individual or group – We analyzed each lake individually, using lake-specific thresholds for that particular lake. Each lake is independently important, and conclusions and decisions will be made independently for each lake.

Effect size – The effect sizes of interest are the EEM thresholds for changes in ANC, SO_4^{2-} and pH, as identified in Table 27 of the EEM. The pH effect size is the same for all of the lakes (i.e., a decrease of 0.3 pH units). However, the ANC and SO_4^{2-} effect sizes are lake-specific. The decrease in ANC associated with a decrease of 0.3 pH units depends on the non-linear pH-ANC titration curve for each lake and the lake's current state (i.e., where is the lake on the titration curve). The change in SO_4^{2-} associated with a decrease of 0.3 pH units is estimated based on the change in ANC. These effect sizes are also shown in the Technical Appendix (Technical Memo W05).

Period of time – We explored changes in statistical power and false positives over 1 to 20 years of post-KMP monitoring.

Baseline – We defined the baseline for each lake chemistry variable in each lake as the average of the values observed in annual sampling over three years (August 2012, October 2013 and October 2014. While recognizing that sampling in August (2012) will be different from sampling in October (2013 and 2014), we observed that the benefit of having 3 years of data instead of 2 years of data outweighed the extra variation induced by variable times of sampling.

Variability in lake chemistry – Variability in lake chemistry is a central challenge to our ability to correctly detect change in the EEM KPI. We considered the following sources of variability:

Inter-annual variability is natural variability from year-to-year in lake chemistry for a single lake even if deposition remained constant. It is important to quantify and account for this variability because it can mask or exaggerate underlying changes in lake chemistry due to the KMP. The only way to reduce the influence of this type of variability is to increase the number of years of lake chemistry observations or to have controls.

Intra-annual variability is variability in lake chemistry for a single lake within a season (i.e., day-to-day or week-to-week) due to fluctuations in physical, chemical and biological processes. We can reduce the influence of this source of variability by increasing the frequency of sampling within a year and by selecting a period of time that is characterized by lower intra-annual variability than others. Sampling in the fall should reduce the influence of this variability because it is typically a period when lake chemistry is most stable (ESSA Technologies et al. 2013, 2014b).

Measurement error is variability due to imperfect estimation of lake chemistry as a result of the method and / or instrument used. Measurement error leads to differences between what is observed and the true underlying lake chemistry state.

2.2 Lakes and Data

We used empirical data from three sets of lakes.

Study Lakes

There are 7 sensitive EEM lakes, as identified in the EEM (ESSA Technologies et al. 2014b). These lakes had annual samples collected in August 2012 under the STAR (ESSA Technologies et al. 2013) and October 2013 and October 2014 under the EEM (ESSA Technologies et al. 2015). Three of the lakes (LAK006, LAK012, LAK023) had multiple samples taken during the fall of 2014, as well as continuous monitoring of pH values.

Control Lakes

Three lakes (DCAS14A, NC184 and NC194) that were sampled in 2013, as part of the Kitimat Airshed Assessment (ESSA Technologies et al. 2014a), were identified as appropriate control lakes due to being similar to the sensitive EEM lakes (i.e., low ANC and generally similar runoff) but predicted to not be influenced by KMP (i.e., they are located well outside the KMP deposition zone and predicted to receive very low levels of acidic deposition and to show negligible changes in pH under KMP (< 0.01 pH units)). These lakes were also sampled in October 2015, as part of the annual EEM sampling, though these data were not available for this analysis.

Supplemental Lakes

Due to the short baseline and high variability in the study lakes, we acquired a supplementary data set to help improve our estimates of predicted lake chemistry variability over a longer time period. This data set includes 10 lakes in southwestern BC that were sampled by Environment Canada from 2005 to 2014 (P. Shaw, pers. comm.; “Environment Canada lakes”). These lakes are not perfect matches to the study lakes, but they are generally similar enough to provide information on the predicted magnitude of natural inter-annual and intra-annual variability in lake chemistry.

2.3 Our Approach

Overview

We used simulation modelling based on the observed data to evaluate our ability to correctly detect changes in ANC, SO_4^{2-} and pH. For each lake, and each of the lake chemistry metrics, we quantified the power (the ability to detect a change that has truly occurred) and the false positive rate (the probability of falsely detecting changes when they had not occurred).

The following steps provide a general overview of our approach. The details of these analyses are documented in the Technical Appendix (Technical Memo W05).

1. Develop a lake chemistry model to simulate alternative, realistic post-KMP time series of each metric for each lake.
2. Run the lake chemistry model to simulate one potential future time series based on a specified “KMP effect”.
3. Simulate “sampling” of this time series to generate “observations” during the post-KMP period.
4. Apply a statistical test to the simulated “observation” to test for a KMP effect (as would be done in the future with multiple years of post KMP monitoring data).
5. Compare the results of the statistical test with the known “true” effect used in the simulation.
6. Score the outcome of the statistical test (e.g., a true positive, false positive, etc.).
7. Repeat steps 2-7 1000 times with each trial representing a random permutation of how lake chemistry and sampling might turn out based on known sources of variability in lake chemistry and measurement error.
8. Compile and present the results across the 1000 trials.

Table 1 summarizes the key components of the simulations, which are further described below.

Table 1: Summary of the key components of the simulations explored in combination with one another. Note that not all possible combinations were explored (see Figure 1).

Component	Description
Post-KMP period	Number of years of post-KMP monitoring (1-20 years)
KMP effect	A Δ pH of 0 or -0.3 pH units and corresponding predicted change in ANC and SO_4^{2-} occurring either (1) abruptly one year after the KMP or (2) gradually (i.e., linearly) over 10 years.
Variability	The magnitude of inter- and intra-annual variability in the EEM lakes either: (1) based on empirical estimates (as described in Section 3); or (2) with inter- and intra-annual variability reduced by an amount equal to the ratio of average Environmental Canada to EEM lake chemistry variability.
Analytic test	Either a statistical test of: (1) a difference in mean lake chemistry between pre- and post-KMP periods; or (2) a difference in the rate of change in lake chemistry between EEM and control lakes.
Frequency of within year sampling	Number of observations of lake chemistry during fall monitoring period (1, 3, 5 or continuous samples; only applicable to LAK006, LAK012 and LAK023).
Control lakes	The inclusion of three sensitive lakes that were outside the predicted KMP deposition area.
Alternative detection thresholds	Lake specific detection thresholds (0, -0.1, -0.2 or -0.3 pH unit changes and corresponding changes in gran ANC and SO_4^{2-}) used to statistically test for a -0.3 pH unit KMP “effect”.

The KMP Effect

We do not know for sure what the effect of the KMP will be on each of the sensitive EEM lakes. We have previously conducted modelling to predict the potential effects (ESSA Technologies et al. 2013, 2015), but even with the best available data, scientific understanding and models, those predictions still have uncertainties. However, in the present context, we want to assess our ability to correctly detect a decrease of 0.3 pH units (as well as corresponding changes in ANC and SO_4^{2-}) if such a change were to really occur. We chose a decrease of 0.3 pH units because changes of this magnitude are associated with ecological changes to lakes including changes in community structure and function (Fölster et al. 2007). We also assess our ability to correctly conclude there has been no change in lake chemistry. Additionally, we consider the possibilities that the effect of the KMP could occur immediately or gradually over a decade. We will not know what the true temporal pattern is until we have sufficient years of post-KMP observations to assess empirically.

Lake Chemistry Model and Baseline Period

The lake chemistry model starts with the baseline lake chemistry observations, adds a defined hypothetical KMP effect, and then adds assumed variability over a 20-year period. We defined the baseline conditions as the average of the observed values in 2012-2014. This period does not represent an ideal, stable baseline because KMP emissions were already declining during this period (and for several years previously), and therefore were lower than longer-term pre-KMP baseline conditions. Given the pre-KMP data available, alternate baselines could have included the average of 2013-2014 (i.e., only use fall samples) or just 2014 (i.e., the last year prior to KMP); however, both of these options would reduce the number of pre-KMP observations and decrease the power to detect changes in the primary metrics. We opted to utilize all of the pre-KMP data in defining the baseline.

Variability

We incorporated several sources of variability in our analyses: shared inter-annual variation (i.e., a common, regional effect); lake-specific inter-annual variation; and lake-specific intra-annual variation. We used the empirical observations from 2012 to 2014 to estimate the lake-specific inter-annual variability (based on annual sampling data) and intra-annual variability (based on the three lakes with multiple within-season measurements). We used the Environment Canada lakes to provide an alternate estimate of inter-annual variability (based on a longer time series) and to estimate the proportion of inter-annual variability attributable to common, shared effects. The ratio of the variability in the two data sets was also used to provide an alternative, reduced estimate of the level of inter-annual as well as intra-annual variability because the short EEM time series could potentially be overestimating the “true” longer-term variability.

Observations

We simulated lake chemistry observations for each lake in each year by taking the lake chemistry value from the simulated time series and adding measurement error. Our assumptions about the magnitude of measurement error are based on the estimates reported by Perrin and Bennet (2015) in their technical report on the EEM sampling performed in 2014.

Analysis

We used two alternative statistical models to test for KMP effects in the simulated lake chemistry time series. The first model tested whether there was a difference in mean lake chemistry between pre- and post-KMP observations (i.e., 2012-2014 vs. 2015 onward). The second statistical model we used tested whether there was a difference in the rate of change in lake chemistry over time between control and EEM lakes.

In each of our simulations, we know whether the KMP effect is an abrupt or gradual change in water chemistry and can therefore choose to apply the right statistical test for the pattern. However, in the real world, until we have collected sufficient post-KMP observations to validate the pattern, we will not know whether a potential KMP effect would manifest itself as an abrupt change or a gradual change over some period of time, and so we cannot know prospectively which statistical test is best to apply. We therefore also explored the influence of applying the wrong statistical test (i.e., testing for a difference in means when the KMP effect is gradual).

Multiple Samples per Year

We explored the benefit of multiple within-year sampling events for the three lakes where multiple samples are feasible and have been previously collected. The observation and analysis components of the analyses were modified as appropriate to account for multiple samples within each year. For pH only, we also explored the influence of continuous monitoring during the fall period using the Manta pH meters (as performed in 2014).

Thresholds

The main power analyses are based on evaluating whether or not we can detect a change of any magnitude (i.e., a detection threshold of 0) given an actual KMP effect size of either 0 or -0.3 pH units (or corresponding changes in ANC or SO_4^{2-}). We also explored the influence of increasing the threshold for detecting a KMP effect. Specifically, we considered detection thresholds of -0.1, -0.2 and -0.3 pH units (and corresponding detection thresholds for changes in ANC and SO_4^{2-}).

This is an important but complex component of the way the analyses are structured. The default threshold of 0 means that the when there is a simulated underlying change of -0.3 pH units, the analysis is actually identifying any case in which it detects a decrease in pH of any magnitude and "counting" it as a detection. This is typically how power analyses are structured. Exploring alternative detection thresholds (i.e., up to the level of effect actually simulated) is then conducted to determine whether there is any improvement in performance (i.e., higher power and/or lower false positive rates). Of course in the future, with actual observations, if a change >0.0 pH units is observed, that does not at all imply that there has been a change of greater than 0.3 pH units. However, with actual observations in the future, the analyses will no longer be structured as a prospective power analyses but rather as an estimate of the probability of a change in lake chemistry of a certain magnitude (see Section 4.2).

Scenarios

We did not explore all possible combinations of the components described above. Instead we identified a subset of scenarios that would allow us to understand the influence of each component on our ability to detect changes in lake chemistry. These scenarios are schematically represented in Figure 1.

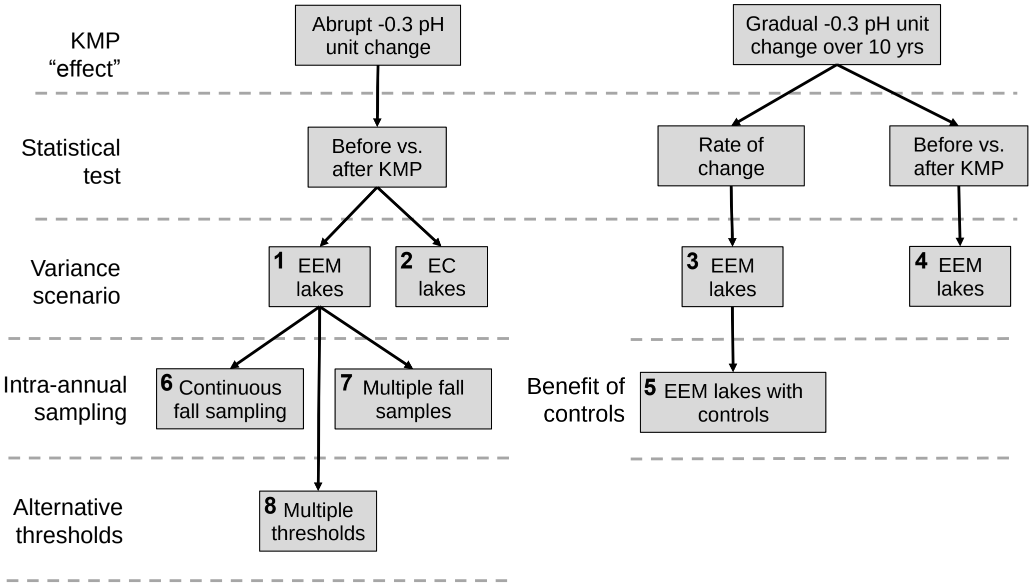

Figure 1: The eight separate scenarios considered in the power analyses. Each scenario (denoted by a bold number in the top left corner of the boxes) was run with and without a KMP "effect" to quantify power and false positive rates with 1 to 20 years of post-KMP observations.

Quantifying Error Rates

We quantified power, the rates of false positives and the rates of false negatives across a range of observation windows (i.e., number of years of observations) and the eight scenarios covering alternative combinations of KMP effect, variance and statistical models to test for KMP effects. For all scenarios we ran 1000 Monte Carlo trials and scored the outcome of each trial as one of 4 possible classifications (Figure 2):

True positive: the analysis detected an effect when an effect was simulated

True negative: the analysis did not detect an effect when no effect was simulated

False positive: the analysis detected an effect when no effect was simulated

False negative: the analysis did not detect an effect when an effect was simulated

To quantify power for scenarios that involved simulating KMP effects on lake chemistry we calculated the proportion of Monte Carlo trials that were true positives. To quantify false positive rates for scenarios that involved no simulated KMP effect on lake chemistry we calculated the proportion of Monte Carlo trials that were false positives.

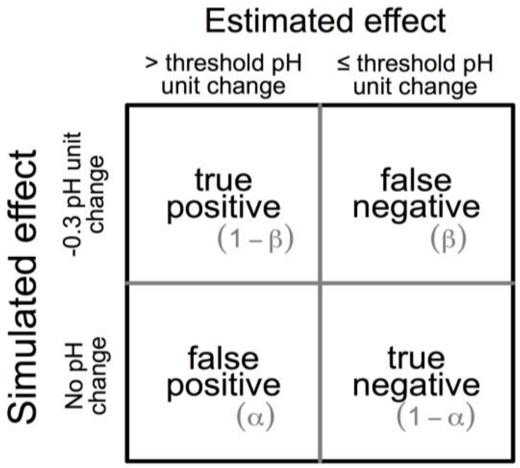

Figure 2: The four possible classifications for each Monte Carlo trial based on a simulated -0.3 pH unit KMP effect. The terms α, β, and 1 − β are equivalent to type I error, type II error, and power in traditional hypotheses testing.

3 What We Learned – Results Interpretation

In this section we provide an overview of the results of the power analyses. We focus on communicating how the knowledge gained from the results has improved our understanding of the problem, especially with respect to the specific project questions described in Section 1.3. The full results of the power analyses are presented in detail in the Technical Appendix (Technical Memo W05). The results of Scenario 1 and Scenario 2 are shown in Figure 3 to illustrate how the results of the power analyses are visualized and interpreted. The Technical Appendix includes similar visualizations of the results for the other scenarios, as well as tables detailing the results for all of the lakes and scenarios, for 3, 5 and 10 years post-KMP.

Specific project questions are addressed below, but some of the major results from Scenario 1 (e.g., the base case) include:

- On average, power is highest for ANC, though it is very low for two lakes; false positive rates are very low for ANC
- On average, power is lowest for pH, though false positive rates are still very low
- False positive rates for SO_4^{2-} are very high for several lakes within the first five years
- On average, power is lower for the combined set of metrics than each of them individually, but false positive rates are zero
- Across all of the metrics, LAK022 and LAK023 consistently have among the highest power
- LAK028 and LAK042 have very low power for ANC
- LAK028 has very low power for SO_4^{2-}
- LAK012 and LAK042 have low power for pH

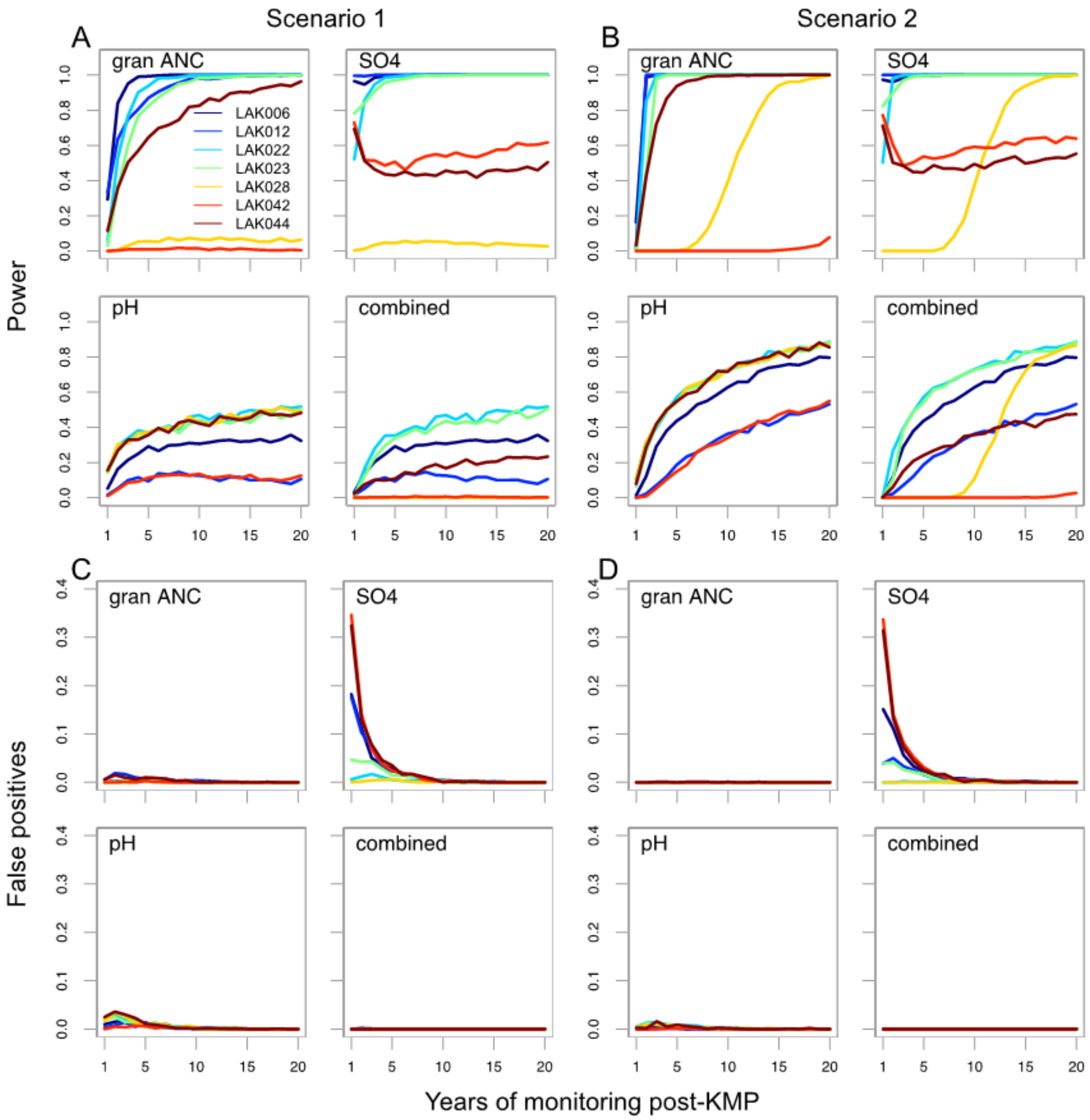

Figure 3: Power to detect a -0.3 unit change in pH, and corresponding changes in gran ANC and SO_4^{2-}, both individually and in combination, under alternative variance scenarios (A: inter-annual EEM lake variability; B: inter-annual EC lake variability). Power is defined here as the proportion of 1000 Monte Carlo trials where a statistical model that tested for a pre- vs. post-KMP difference in mean lake chemistry correctly estimated a significant change (Scenarios 1 and 2 in Figure 1). False positive rates for the same scenarios but for simulations without a KMP effect are presented in panel C (assuming EEM lake variability) and D (assuming EC lake variability), respectively.

3.1 Lake Chemistry Variability

Lake chemistry for the EEM lakes was generally more variable among years than within (i.e., for those with both inter-annual and intra-annual observations). Variability tended to be highest for

SO_4^{2-}[3], then ANC, then pH. Variability in the Environment Canada lakes was much lower than the EEM lakes, particularly for ANC (Gran ANC) and SO_4^{2-}.

3.2 Specific Project Questions

Power to Detect KMP-induced Changes in Lake Chemistry

How well will we be able to detect changes in water chemistry indicators greater than the identified limits? How long will it take to detect these changes?

Changes in ANC and SO_4^{2-} were easier to detect than changes in pH. Most lakes had high power[4] to detect abrupt changes in ANC and SO_4^{2-} within 5 years. Lakes with the greatest variability and/or relatively modest changes in ANC and SO_4^{2-} changes associated with $\Delta pH=-0.3$ had low or moderate power over any time frame. All lakes had low to moderate power to detect an abrupt decrease in pH of 0.3 units. False positive rates were high for SO_4^{2-} in the first few years, but otherwise false positive rates were very low for all lake chemistry variables.

Although pH had the lowest variability of the three metrics, it also consistently had the least power. In general, lower variability will support higher power, but in the present case the more dominant factor is the difficulty of having a measurement error (+/- 0.2 pH units) that is very large relative to the magnitude of change we are trying to detect (0.3 pH units).

If we assume that natural variability is better approximated by the Environment Canada lakes than the empirical EEM observations (for which we only had a few years), then natural variability would be lower and changes in lake chemistry would be easier to detect. Under this alternate variance scenario, power was higher but false positives did not increase appreciably.

Gradual changes in lake chemistry were harder to detect than abrupt ones. For example, after 10 years, the power to detect gradual changes in pH was approximately half as much as the power to detect an abrupt change. False positive rates were higher when the underlying change was gradual because natural variation more frequently led to the appearance of gradual change when none in fact had occurred – this was especially true for SO_4^{2-} because it has the highest variability. For almost all of the lakes and metrics, at 5 years the power to detect gradual changes was considerably less than the power to detect abrupt change; however, because the gradual change is simulated as a linear change over 10 years, the effect being detected at 5 years is only half of the full KMP effect size.

To assess the influence of applying the wrong statistical test, we tested for a difference in means when lake chemistry had been simulated as a gradual change. Applying the wrong statistical test resulted in reduced power. However, it also resulted in reduced false positives because natural variation in lake chemistry was less likely to be misinterpreted as a true underlying change. We did not evaluate the consequences of applying the wrong statistical test when lake chemistry changed abruptly; however we expect that power under such a scenario would also be lower than if the correct statistical test were applied.

---

[3] Some of the year-to-year variability in sulphate is due to the fact that sulphur dioxide emissions were decreasing during the period from 2012 to 2014 (ESSA et al. 2015).

[4] Qualitative classifications of the level of power: >0.8 = "high", 0.2-0.8 = "moderate", and <0.2 = "low".

Intra-annual Samples

What is the benefit of collecting continuous / intra-annual samples?

Increasing the number of samples collected during the fall increased power to detect change in lake chemistry. The benefit of increased sampling was greatest for pH, where, on average, increasing from a single sample to five samples doubled power. This is an important result because power is consistently much lower for pH than either ANC or SO_4^{2-} across the scenarios we considered and so the gain in power associated with multiple fall samples is particularly beneficial. However, these increases in power came at the cost of increased false positive rates when there were less than 5 years of post-KMP observations. In the first few years, additional samples add variability in lake chemistry that can be misinterpreted as a true underlying change. This points to the value of verifying (through more intensive or repeated sampling) that an apparently significant change in pH has indeed occurred, and is not merely due to natural variability in pH. Increasing the number of samples also increased the power to detect changes in ANC and SO_4^{2-}; however, the three lakes with intra-annual sampling already had high to very high power for these metrics after 5 years.

An additional benefit of collecting multiple samples is the continued improvement of our estimates of intra-annual variability. Although not as large as inter-annual variability, intra-annual variability is still an important source of variability (both with respect to these power analyses and our broader understanding of the dynamics of the system). However, our estimates of intra-annual variability are only based on observations from three lakes within a single year (2014). We cannot increase the number of lakes, but additional years will help us better characterize within year variability which can inform future power analyses and interpretations of monitoring data.

Continuous monitoring of lake pH led to an increase in the power to detect true changes in pH for the three lakes for which continuous monitoring is currently possible. However, the relative increases in power (1.1-1.4 times greater power) were less pronounced than the benefits of increasing sampling from 1 to 3 samples per fall (1.5-2.6 times greater power) and they came at the cost of increased false positives rates relative to only a few fall samples of lake chemistry. Continuous sampling however has the additional benefit of detecting storm or snowmelt driven acidic episodes, should they occur.

Control Lakes

What is the benefit of monitoring control lakes?

The inclusion of control lakes increased power to detect changes in lake chemistry for the first few years of post-KMP observations but thereafter generally had only a modest benefit relative to analyses without controls. At 5 years, the inclusion of controls resulted in a large relative increase in power for ANC, but power remained low to low/moderate. For SO_4^{2-} the controls resulted in relatively small changes to power, with the notable exception of LAK028 that showed an increase in power from 0.15 to 0.57, which is especially beneficial since it had the lowest power for SO_4^{2-}. For pH, the addition of controls resulted in large relative gains but small absolute gains, with the power remaining very low after 5 years.

The generally modest benefit of control lakes occurred because the degree of regional coherence in lake chemistry variability (i.e., shared year "effects") has been assumed to be relatively small, though we do not truly know the degree of shared variability within this region. We did not have sufficient data to estimate the shared effect within the study region and instead estimated the magnitude from the Environment Canada lakes where there was sufficient data to try to quantify shared regional variation in lake chemistry. If the true degree of shared regional variation in lake chemistry is higher than that assumed in our analyses, then the benefit of controls would be expected to be more pronounced.

An additional benefit of the control lakes is that their continued monitoring will improve our estimates of natural inter-annual variability in lake chemistry in this region, in the absence of any significant changes in sulphur deposition. Estimates of lake chemistry variability used in the power analyses were based on the EEM lakes, which could be confounded by the effects of changes in pre-KMP emissions. However, estimates of natural variation from the control lakes will not have such confounding effects, and will better inform future power analyses and interpretations of monitoring data.

Multiple Metrics

Is our ability to confidently detect changes in lake chemistry improved by using multiple metrics?

Considering all three lake chemistry variables simultaneously when testing for a KMP effect, led to lower or equal power relative to any single lake chemistry variable, but also lower false positives (often zero). Because false positives were already very low in most cases there does not appear to be much benefit to considering all three lake chemistry variables together. However, given differences in power and false positive rates among the metrics when evaluated individually, it will be beneficial to continue to monitor and analyze all three metrics independently.

Alternate Thresholds of Detection

What is the benefit of using alternate thresholds of detection?

Increasing the threshold for detection led to significant losses in power but only minimal gains in terms of the false positive rate because false positives were so low to begin with. In general detection thresholds can be changed to reduce false positives (i.e., setting the bar for detection higher) and we would expect that using a very small detection threshold would in increase power (proportion of true positives) but at the cost of increased false positive rates (i.e., because the lower threshold is more likely to be exceeded even when there is no effect). However, in the present case the false positive rate is already very low for ANC and pH, and for SO_4^{2-} after 5 years, meaning that increasing detection thresholds is not necessary.

3.3 Additional Conclusions

For some lakes we are going to have to accept that we are not likely to ever have high power to detect KMP induced changes of a -0.3 pH unit magnitude within the context of the sampling “levers” under our control. This is primarily the case for LAK042 and LAK028. This is a concern because both of these lakes are predicted to experience decreases in pH of greater than 0.3 pH units[5]. It is important to consider both our power to detect a decrease in pH of 0.3 units (and/or corresponding changes in ANC and SO_4^{2-}) as well as the magnitude of the actual predicted change for a given lake under the KMP. For example, we may be much less concerned with low power to detect a change in pH of 0.3 units for a lake with a predicted change of only 0.1 pH units. The power analyses and the predicted changes in lake chemistry can both be updated annually as more post-KMP observations are collected and we gain an improved understanding of natural variation in lake chemistry, as well as actual changes in sulphur deposition.

4 Implications for the EEM

4.1 Recommendations

The monitoring and analysis recommendations that emerge from our analyses include:

1. Continue taking continuous and multiple samples from those lakes where it is logistically feasible.
2. Continue to use multiple metrics to assess potential KMP effects, but analyze the metrics separately.
3. Continue to sample control lakes annually, primarily to characterize natural variability (among years and among lakes) in the system. The control lakes will also help estimate the level shared variation (e.g., common year effect) across the region over time. If the level of shared variation is substantially higher than what we have assumed (based on a different region), then the control lakes will become even more important in isolating and detecting KMP induced changes in EEM lake chemistry.
4. Consider using Gran ANC as the primary indicator of KMP induced change in lake chemistry. Gran ANC had a higher power to detect true changes than pH but lower false positive rate than SO_4^{2-}.
5. Always review the raw monitoring data itself – there is no replacement for carefully looking at the raw data prior to analysing it. Although visual assessment should not be relied upon alone, it is a critical piece of the overall analysis that should not be forgotten. Visual assessment can suggest important questions to investigate if the observed patterns are different than anticipated or do not appear to align with the conclusions of the formal statistical analyses.
6. Incorporation of autocorrelation into both simulated data and the error structure of the statistical model fit to the data to test for a KMP effect should be considered in subsequent power analyses and analyses of empirical data (refer to Technical Appendices for further discussion).

---

[5] The ESSA-DFO has previously been used to predict post-KMP changes in pH based on the 2012 sampling data (ESSA Technologies et al. 2013) and the 2013 sampling data (ESSA Technologies et al. 2015). Both of these lakes had a predicted pH change of greater than 0.3 units in analyses based on sampling data from one of those years but not the other, so neither has a uniform prediction with respect to exceeding the 0.3 pH unit threshold.

7. Explore cloud computing options to improve the efficiency of running the simulations for the power analyses if they are revised in the future. With desktop computing power, some of the analyses take many days to run.
8. Wait until having collected 5 years of post-KMP monitoring data before drawing conclusions about potential changes to lake chemistry, due to the predicted low power and higher false positives (for some scenarios) in the first few years of post-KMP monitoring. At a minimum, wait until the end of the initial phase of the EEM program (3 years of post-KMP monitoring data).
9. Due to very low power to detect a change of 0.3 pH units in LAK042 (as described above), it would be beneficial to have intra-annual monitoring for this lake. We recommend the following phased approach:
    a. Do further reconnaissance of the feasibility of increasing the number of samples taken per year for LAK042.
    b. If increased sampling is feasible, then simulate the potential benefits of increasing intra-annual samples (applying assumptions of intra-annual variability from other lakes). We expect there to would be benefit, but less than other lakes due to the lack of pre-KMP intra-annual samples for the baseline. Evaluate the tradeoffs between the predicted values of the additional information versus the costs of acquiring it.
    c. If increased sampling is feasible and beneficial, implement the increased level of sampling for these lakes.

4.2 How Monitoring Results will be Interpreted in the Future

Our recommendation for the analyses of the EEM sampling data is to use the same statistical test(s) as we have applied in the present analyses. Until we have a sense of whether changes in lake chemistry are best described by a step function or gradual rate of change we will need to apply both tests. We recommend applying these statistical tests within a Bayesian framework, in order to be able to generate a probability distribution of the estimated effect size for each lake chemistry variable and lake. We recommend this approach because probability distributions are more intuitive to interpret and explain than frequentist confidence intervals (e.g., for Lake X, in year Y we can say that the probability that pH has declined by more than 0.1 units = 50% and by more than 0.3 units = 15%). As more years of post-KMP data are collected, the study's conclusions will be more precise, and statistical power will increase (Figure 3).

The EEM considers KPI thresholds in a deterministic way, assuming that it will be possible to conclude that a lake either has or has not decreased by greater than 0.3 pH units. However, our analyses suggest that the EEM will need to consider such effects within a probabilistic framework due to the high level of variability relative to the effect size of interest. This raises some very important questions that are not currently addressed in the EEM, namely the acceptable level of certainty required to conclude that a lake has exceeded its thresholds. It could likely be agreed that a 95% probability is sufficient to conclude that a particular lake has exceeded its threshold, but what about, for example, a lake with a 60% probability of having exceeded its threshold?

Ultimately, the interpretation of the level of certainty that is acceptable for drawing a particular conclusion and any associated decisions (e.g., more intensive monitoring and mitigation options) is a question of risk tolerance, which is fundamentally a policy question (with scientific input). Such policy decisions require further discussion and consideration.

4.3 Future Learning

As more post-KMP observations are collected, we may be able to refine some aspects of the power analyses, adding to the strength of prospective analyses regarding potential effects as well as starting to allow retrospective analyses to evaluate whether actual effects are or are not occurring. With more data, we will be able to better evaluate the temporal patterns of potential effects (i.e., abrupt vs. gradual). We will also be able to improve our estimates of inter-annual variability (especially from the control lakes) and intra-annual variability (from the intensively monitored lakes) as more data are collected. We will also start to be able to estimate the common, shared inter-annual variability within this region and determine whether it differs from the assumptions based on the Environment Canada lakes. Lastly, we will start to be able to directly evaluate the relationship between changes in emissions and changes in water chemistry (including assessing potential temporal lags).

5 Literature Cited

Covington, A.K., P.D. Whalley and W. Davison. 1985. Recommendations for the Determination of pH in Low Ionic Strength Fresh Waters. International Union of Pure and Applied Chemistry, Analytical Chemistry Division, Commission on Electroanalytical Chemistry. Pure and Applied Chemistry 57(6):877-886.

ESSA Technologies, J. Laurence, Limnotek, Risk Sciences International, Rio Tinto Alcan, Trent University, Trinity Consultants and University of Illinois. 2013. Sulphur Dioxide Technical Assessment Report in Support of the 2013 Application to Amend the P2-00001 Multimedia Permit for the Kitimat Modernization Project. Vol.2: Final Technical Report. Prepared for RTA, Kitimat, BC. 450 pp.

ESSA Technologies, J. Laurence, Risk Sciences International, Trent University, and Trinity Consultants. 2014a. Kitimat Airshed Emissions Effects Assessment. Report prepared for BC Ministry of Environment, Smithers, BC. 205 pp. + appendices.

ESSA Technologies, J. Laurence, Limnotek, Risk Sciences International, Trent University, and Trinity Consultants. 2014b. Sulphur Dioxide Environmental Effects Monitoring Program for the Kitimat Modernization Project. Program Plan for 2013 to 2018. Prepared for Rio Tinto Alcan, Kitimat, B.C. 99 pp.

ESSA Technologies Ltd. 2015. Sulphur Dioxide Environmental Effects Monitoring for the Kitimat Modernization Project – 2013 and 2014 Annual Reports. Prepared for Rio Tinto Alcan, Kitimat, B.C. 27 pp.

Fölster, J., C. Andrén, K. Bishop, I. Buffam, N. Cory, W. Goedkoop, K. Holmgren, R. Johnson, H. Laudon, and A. Wilander. 2007. A Novel Environmental Quality Criterion for Acidification in Swedish Lakes – An Application of Studies on the Relationship Between Biota and Water Chemistry. Water Air Soil Pollut: Focus 7:331-338.

Perrin, C.J and S. Bennett 2015. Rio Tinto Alcan Kitimat Modernization Project: Environmental effects monitoring of lake water quality in 2014. Data report prepared by Limnotek Research and Development Inc. for Rio Tinto Alcan Ltd. 20p

Technical Appendix

Methods and Results of Power Analyses

The technical appendix with detailed descriptions of the methods and results of the power analyses is provided in Technical Memo W05.