Azimuth Consulting Group Partnership

218-2902 West Broadway
Vancouver, BC
Canada V6K 2G8

Phone: 604-730-1220
www.azimuthgroup.ca

Technical Memorandum

Date: **December 20, 2017**
To: **Peter Madden, Nalcor Energy**
From: **Ryan Hill, Brian Pyper, Randy Baker**
Our File: **NE 16-01**

RE: Clarification of the Effect of Measurement Error on Analyses of Methylmercury Concentrations in the Lower Churchill River

In a technical memorandum prepared in November 2017, Azimuth analyzed spatial and temporal patterns in methylmercury (MeHg) concentrations (mostly dissolved; DMeHg) of Lower Churchill River water collected weekly over nearly one calendar year. We assumed that readers had read and understood the information in that memorandum, including details in the statistical appendix. However, it is clear to us from the review by Dr. Iris Koch (December 12 memo and associated slide presentation) and the summary memo by the Independent Expert Advisory Committee (IEAC, December 13) that aspects of the statistical analysis have been misunderstood and require clarification. The outcome of this clarification is our re-iteration and categorical conclusion that current data quality and sample replication within and among time periods is sufficient to provide strong and valuable inferences with respect to observed changes in mercury levels.

The technical memorandum by Dr. Koch to the Chair of the IEAC reviewed the “10+10” water chemistry data with the objective to “determine analytical and field variability” of the data. It was not a review of our statistical approach and analysis or its conclusions. Dr. Koch’s memorandum explained and used standard methods for characterizing “measurement uncertainty”, and concluded that estimates from the 10+10 study are in line with expectations. We also reached the same conclusion regarding the quality of the data provided by Flett Research. It is also worth pointing out that the observed level of uncertainty for MeHg is typical and would be expected for most parameters (e.g., metals) analyzed by commercial laboratories, and it is not unique to MeHg.

Dr. Koch then examined the implications of measurement uncertainty for detecting differences or changes in MeHg concentrations. Dr. Koch concluded that differences in

the measured concentrations of dissolved MeHg between N1 and downstream stations were often within the range of this measurement uncertainty. Azimuth agrees with this general observation, that conclusion pertains only to comparisons of individual samples. However, our interpretation of the monitoring data was *not* based on individual samples but rather on inferences derived across numerous *replicate* samples. The important distinction between examining individual samples versus groups of samples appears to have been overlooked. We note the following inter-related concluding statements made in the recent 13 December 2017 IEAC memo, which we address each of below:

1. "*Results from analytical labs should not be treated as single values. Instead there is a spread of results within which lies the actual concentration of methylmercury."*
2. *"The spread of values in this case is 40%. There is a 95% confidence that the 'true value' lies within this spread."*
3. *"Consideration of these measurement uncertainties when examining time trends at the reference station N1 (upstream of Muskrat Falls) showed that differences in dissolved methylmercury at station N4 (in the impoundment), N5 (just downstream of the impoundment), and N8 (near Goose Bay) were obscured by the measurement uncertainty, since most values at stations N4, N5 and N8 were within 40% of values at N1."*
4. *"In order to discern differences between a downstream station and the reference site, the methylmercury concentrations would have to increase by 80% (i.e. roughly double*)."

We agree with bullet #1. However, we did not compare single values to single values in our statistical analyses, nor did we assume that observed values were 'true' values.

We also agree with bullet #2 when referring to individual measurements, but this statement is not true when referring to a mean of two or more measurements.

The text of bullet #3, originating from Dr. Koch's memorandum, is very misleading because it ignores the important influence of sample replication on *reducing* uncertainty associated with measurement error.

In bullet #4, as with the second, this is true only when comparing a single value at a downstream station to a single value at a reference station. This is never the intent of a well-designed monitoring program and was not the approach taken in our analysis of the data as clearly laid out in the November 2017 memo.

What is missing in these concluding bullets, particularly in the last two, is an understanding of the role of sample replication, and how it affects our ability to detect differences in *mean* values in the presence of natural variation *as well as* measurement error. The following is a crucial point to understand: measurement error is not something that needs to be added to a group of replicate measurements, rather, measurement error is already a part (or the whole) of the variability observed among those samples[1].

[1] In this context, the figures in Dr. Koch's slide presentation that superimpose measurement uncertainty over monthly means are misleading (e.g., slides 27 and 32): these would only be correct if each month was represented by a single value. The true influence of measurement error is contained within the sample variation, and hence, the confidence interval for each monthly

Thus, the basic statistical principle for reducing uncertainty in an estimate (e.g., mean) holds regardless of whether variability is "natural" or due to measurement error – increasing sample sizes will increase the precision of estimates and increase the power to detect a difference. It is easy to illustrate, for example, the influence of sample size on the precision of a mean estimate given a specified level of measurement error[2].

Thus, the ability to discern differences between any two stations depends not only on measurement error, but also on sample size. No credible study design or statistical analysis would rely on a comparison of two individual measurements. Sample replication is required, and incrementally reduces the uncertainty associated with measurement error. All of the statistical analyses reported in our November 2017 technical memorandum accounted for this measurement error, automatically (implicitly), because we derived inferences across sample replicates, and the variation among replicates includes measurement error[3]. This is important – standard statistical tests *automatically* account for measurement errors in the "dependent" variable (e.g., measurements of DMeHg). For example, in a two-sample t-test, the "sample variation" is comprised of true differences among replicates (e.g., natural variation over space or time), as well as measurement error, if present.

Given the above, the notion that we cannot discern increases in MeHg of less than 80% is simply not true. The current monitoring program has sufficient sample replication to detect differences far less than 80% when evaluating data trends across several months, which is an appropriate time frame for basing inferences. Current data quality is more than sufficient to provide strong and valuable inferences with respect to potential changes in mercury levels because there is sufficient sample replication within and among time periods.

To illustrate with a simple example, we used the Bayesian analysis of monthly means for DMeHg from the appendix of our November 2017 memo; note that the analysis incorporates measurement error because individual values *include* that error. We estimated the mean differences at N4 to N8, relative to N1, across June to September, the four months of key difference in the study. Importantly, this analysis did not pool

---

mean would provide an unbiased estimate of uncertainty, which would be narrower than implied by these figures (much narrower if sample sizes were large).

[2] Assume a group of measurements is subject to natural variation with standard deviation σ_{nat} and measurement error with standard deviation σ_{err}. The standard deviation of the population of observed measurements will be $\sigma_{pop} = \sqrt{\sigma_{nat}^2 + \sigma_{err}^2}$, while the standard deviation of sample means across *n* replicates will be $\sigma_{means} = \sigma_{pop} / \sqrt{n}$. Thus, increasing sample size (*n*) reduces the influence of measurement error (which is automatically accounted for when estimating the population standard deviation) on the uncertainty of an estimated mean.

[3] In Dr. Koch's slide presentation, there is a statement on slide 33 that Azimuth's Bayesian analysis of monthly means "…took into account the variability within a month, but not measurement uncertainty." This is incorrect. Measurement error is part of the sample variance, which we accounted for, just as measurement error was the key component of sample variance in the "Ten by Ten" study designed to quantify measurement error.

replicates across months, but rather treated each month as a separate unit (i.e., the mean across posteriors for each month, treating each month equally).

The table below shows the mean proportional difference in DMeHg of downstream stations relative to N1, the standard deviation (SD) of the probability distribution for the mean difference, and the lower and upper 95% credibility intervals for the distribution.

Site	Mean	SD	Lower	Upper
N4	1.32	0.11	1.12	1.56
N5	1.06	0.11	0.86	1.29
N6	1.10	0.10	0.91	1.30
N7	1.11	0.12	0.91	1.37
N8	0.94	0.12	0.73	1.20

Because this analysis is based on probability distributions, we can make probability statements. First, we can be very confident that means at N4 were greater than at N1; the posterior mean suggests a 32% increase (relative to N1) across June-September, with a 95% credibility interval that this increase is between 12% and 56%. In other words, we detect a difference (a mean increase of only 32%) with high confidence. Estimated differences for N5-N8 were much smaller compared to N4. Here, we can state with reasonable confidence that there is a low probability (0.025) that mean levels of DMeHg at N5, N6, N7, or N8 exceeded those at N1 by more than 29%, 30%, 37%, or 20%, respectively, across months June-September.

Finally, while there are a variety of standard analyses (e.g., analyses of statistical power or precision) that we could perform using the existing dataset, we are confident that such analyses would further demonstrate that the current data quality and sample replication within and among time periods is sufficient to provide strong and valuable inferences to measure changes in mercury levels in the Lower Churchill River and estuary with good accuracy.

In conclusion, we hope that this memo clarifies to the IEAC that the water quality sampling program for the Lower Churchill River is a powerful tool for detecting changes to water column MeHg concentrations. Further, the current program will support future analyses that may explore possible relationships between mercury in water and mercury in biota, extending into Lake Melville.

We also wish to emphasize that we understand the importance of this program and others, that help to address the legitimate concerns of downstream community members and we take this responsibility seriously. We firmly believe that the water sampling program is rigorous and will enable the committee to draw appropriate inferences and conclusions that will ultimately assist in making good management decisions.