TASK ORDER 72 UNDER CONTRACT EP-C-12-029

EXTERNAL PEER REVIEW OF A COMPARISON OF ALUMINUM AQUATIC LIFE CRITERIA APPROACHES

FINAL PEER REVIEW SUMMARY REPORT

January 13, 2017

Submitted to:
U.S. Environmental Protection Agency
Office of Water, Office of Science and Technology
Health and Ecological Criteria Division
1200 Pennsylvania Avenue, NW
Washington, DC 20460
Attn: Diana Eignor
Eignor.Diana@epa.gov

Submitted by:
Eastern Research Group, Inc.
110 Hartwell Avenue
Lexington, MA 02421

CONTENTS

1.0 INTRODUCTION

This report documents the results of an independent peer review of a Comparison of Aluminum Aquatic Life Criteria Approaches for the U.S. Environmental Protection Agency (EPA), Office of Water (OW). Eastern Research Group, Inc. (ERG, a contractor to EPA) organized this review and developed this report. Sections 2.1 to 2.5 of this report present, for each charge question, the individual reviewer comments and a summary of those comments. New information (e.g., references) provided by reviewers is presented in Section 3. Appendices A and B provide, respectively, the charge to reviewers and the complete set of comments submitted by each reviewer.

1.1 Background

The EPA OW is charged with protecting ecological integrity and human health from adverse anthropogenic, water-mediated effects, under the purview of the Clean Water Act (CWA). In support of this mission, EPA is working to update water quality criteria to protect aquatic life and aquatic-dependent wildlife from the presence of aluminum in freshwater and estuarine/marine environments.

The effects of water chemistry on the aquatic toxicity of metals have been an area of research and regulatory advancement since water quality criteria were developed. A partnership among industry, academia, and government utilized research advancements in the aquatic toxicity of metals to effectively develop a national water Biotic Ligand Model (BLM) to address the bioavailability of metals in aquatic systems and their acute toxicity to fish. The BLM adjusts the water concentration that causes acute toxicity of metals to aquatic organisms by calculating the relative binding affinity of cations in the water to the biotic ligand (e.g., gills or any other active site). The BLM was developed as a method to develop water quality criteria (WQC) for metals.

The purpose of this peer review was to provide a comparison of several approaches to generating aluminum criteria that reflect water quality condition impacts on toxicity. Approaches compared included existing complete 10 parameter BLM(s), a simplified BLM approach (e.g., pH, hardness, dissolved organic carbon, temperature), and regression-based equation approaches (e.g., hardness equations) to facilitate evaluation of the most appropriate approaches to consider across aluminum modeling approaches.

1.2 Peer Reviewers

ERG identified, screened, and selected five qualified experts who met technical selection criteria provided by EPA and who had no conflict of interest in performing this review:

- Gretchen K. Bielmyer-Fraser, Ph.D.; Associate Professor, Department of Biology, Valdosta State University.
- Tham C. Hoang, Ph.D.; Assistant Professor, Department of Environmental Sciences, Loyola University Chicago.
- Gregory Möller, Ph.D.; Professor of Environmental Chemistry and Toxicology, University of Idaho.
- Scott Smith, Ph.D.; Associate Professor, Department of Chemistry and Biochemistry, Wilfrid Laurier University, Canada.
- Randall Wentzel, Ph.D.; Senior Managing Scientist, ExPonent.

ERG provided reviewers with instructions, the aluminum BLM, a BLM User's Guide, an Excel spreadsheet that provided multilinear regression equations, and the charge to reviewers (Appendix A of this report)

prepared by EPA. Reviewers worked individually to develop written comments in response to the charge questions. After receiving reviewer comments, ERG summarized reviewers' responses to the charge questions, noting areas of agreement and disagreement, where relevant (see Section 2).

2.0 SUMMARY OF REVIEWER COMMENTS ORGANIZED BY CHARGE QUESTION

This section presents summaries of reviewer comments organized by charge question. Each summary is followed by a table presenting the individual reviewer comments by charge question. Individual comments are copied directly from written comments as submitted by each reviewer and presented in Appendix B.

2.1 Using the scenarios provided in the table below, do a side-by-side comparison of the results of Full BLM, Simplified BLM, and Excel results from the 2 MLR approaches (highlighted in yellow in Excel file).

- **Please draw some conclusions regarding the differences in the values generated and explain your rationale.**
- **Would the simplified BLM be sufficiently protective?**
- **Would the pH and hardness MLRs be appropriate?**
- **Please provide appropriate suggestions.**
- **Feel free to try your own scenarios to see differences and provide with your review.**

- **Please draw some conclusions regarding the differences in the values generated and explain your rationale.**

After evaluating each of the scenarios, all five reviewers noted that the full and simplified BLM approaches yielded similar results. Four reviewers (1, 2, 4, 5) also commented that the results from MLR approach 1 were similar to those from the full and simplified BLM approaches, while all five reviewers observed that the results from MLR approach 2 were consistently lower than those from MLR approach 1 or from the BLM methods.

Four reviewers (1, 2, 3, 5) discussed other trends in the results of the different methods. Reviewer 2 noted that after the elimination of MLR approach 2, the coefficient of variation across all model results for Scenarios 1-4 ranged from 10.7% to 20%, demonstrating acceptable model consistency. Reviewer 2 also observed that the full BLM with estimated HC5 based on a truncated triangle distribution (TRI) was the "least protective" method across Scenarios 1-4, while the simplified BLM with estimated HC5 based on a log normal distribution (LN) was the "most protective," based on trends in each method's difference from the mean. Reviewer 3 commented that all four types of criteria determination show the same general trend with criteria values highest when hardness is high and pH is high. Reviewer 3 wrote that lower hardness and lower pH waters are less protective, such that the estimated criteria is lower, and that pH has a more significant impact on the results than hardness.

Regarding pH and hardness, Reviewer 1 commented that the full BLM and the simplified BLM approaches produced similar HC5 values when the pH was varied, and that MLR approach 1 produced similar values to the full and simplified BLM approaches at pH 6 and 7 at a hardness value of 100 mg/L. However, at lower hardness concentrations and acidic pH, the MLR approach 1 had higher aluminum values than either of the BLM approaches. Reviewer 5 noted that the chronic values for aluminum generated by the two BLM methods were very similar for scenarios 1 and 2, but that the simplified BLM results were lower in scenarios 3 and 4.

- **Would the simplified BLM be sufficiently protective?**

Four reviewers (1, 2, 3, 5) wrote that the simplified BLM would be sufficiently protective (Reviewer 4 did not comment on this part of the question). Reviewer 1 commented that the simplified BLM produced aluminum HC5 values that are very similar to the full BLM, such that the simplified BLM may be considered protective if the full BLM is considered protective.

Reviewer 2 commented that the simplified BLM results cluster well with the other considered models, and that the simplified BLM (LN) is the most protective of the approaches considered. Reviewer 5 stated that the simplified BLM would be sufficiently protective in all cases except those of high hardness and pH, though this case would not be as common in most aquatic systems.

Reviewer 3 wrote that the provided scenarios were not sufficient to assess if the simplified BLM will be protective or not, because protectiveness can only be assessed by comparing the results to actual toxicity measurements, not to the results from other models. To this end, Reviewer 3 ran the 23 lowest values in the chronic dataset provided with the BLM as test cases and compared the predicted criteria values from the simplified BLM with the measured effects concentrations. Based on the results of these comparisons, Reviewer 3 found that the simplified BLM was sufficiently protective.

Three reviewers (2, 3, 4) commented that the input data needed for the simplified model is much less than that required for the full model. As a result, the simplified model would be more economical and easier to use. However, Reviewer 1 noted that the simplified BLM requires a DOC, but many state and local government water quality programs do not typically measure this value. Reviewer 1 recommended consideration of increased variability of the estimation of DOC in the application of the simplified BLM.

- **Would the pH and hardness MLRs be appropriate?**

In general, reviewers differed in their responses to this question. All five reviewers noted that the results of MLR approach 1 and MLR approach 2 were not consistent. Both Reviewer 2 and Reviewer 5 stated that the use of Final Acute Value (FAV) and Final Acute-Chronic Ratio (FACR) methods in approach 2 may be biasing the Criteria Continuous Concentrations (CCC) low, such that MLR approach 2 is not as accurate as the other three. Reviewer 4 questioned how the equation factor of each approach was derived. Reviewer 4 stated that the CCC derived from the most sensitive species should be lower than the CCC derived from all species, but noted that the opposite seemed to be occurring here.

Reviewer 2 stated that MLR approach 1 results appear to cluster well with the other considered models, and can thus be considered "sufficiently protective." Reviewer 2 also stated that MLR approach 1 would be appropriate for water quality criteria development, and that the low data input requirements of MLR approach 1 are attractive in managing costs. Regarding MLR approach 2, Reviewer 2 noted that this approach shows the lowest values of all models for all scenarios. This reviewer observed that MLR approach 2 appears to be less responsive to scenario hardness and pH inputs, and its results appear to be "overly protective" across the pH and hardness ranges explored in the modeling exercise, such that elimination of this approach from further consideration in the review was justified.

Reviewer 3 discussed how it is important to compare the model results with actual toxicity results rather than other model results. This reviewer tested the 23 lowest effect concentration samples in the chronic database against the two MLR approaches and found that approach 2 was more conservative than approach 1, and that the estimated values for approach 1 were not dramatically different from the measured effects. Reviewer 3 found the MLR approach 2 to be protective in all instances, but thought that many of these values were overly conservative. However, because approach 1 sometimes generates values that are not protective, the simplified or full BLM were the most appropriate methods to estimate chronic criteria.

Reviewer 1 wrote that it is difficult to determine if the MLRs are appropriate. In a comparison of the BLM and MLR approaches, Reviewer 1 noted that the MLR approach 1 produced values that were similar to the BLM values at a hardness of 100 mg/L, but higher under acidic pH and a hardness of 50 ppm. MLR approach 2 produced values that were lower than the BLM approaches. Reviewer 1 recommended that US EPA reach out to see if the scientific community has published or plans to publish alternative MLR approaches to estimate metal criteria.

Four reviewers (1, 3, 4, 5) had concerns about the accuracy of MLR approaches in high DOC environments. Reviewer 5 noted that differences between the MLR and BLM approaches are more apparent when DOC values are higher, as the BLM uses DOC as an input variable and the MLR approaches do not. Reviewer 1 wrote that DOC is a critical parameter for estimating aluminum water quality criteria, and Reviewer 3 commented that the use of only hardness and pH as criteria is potentially limiting if DOC is significant in the receiving waters. Reviewer 4 noted that 54% of the data used to derive the hardness and pH equation are missing DOC, and that other data have very low DOC concentrations. Reviewer 4 stated that the hardness and pH approaches should be used to predict toxicity only for the range of water quality that was used to develop the approach. In addition, Reviewer 4 had concerns about the low DOC concentrations in the database.

- **Please provide appropriate suggestions.**

Reviewer 1 wrote that the key issue concerning the utility of the BLM is that state and local governments typically do not measure the parameters required for the full BLM. To increase utilization of these methods, Reviewer 1 recommended that the US EPA OW develop tiered approaches beginning with what states and local governments are currently doing to eventually include more complex approaches like BLM. US EPA could then put forward tiered approaches in different areas, based on each area's risk of aluminum toxicity.

Reviewer 3 recommended presenting test cases and test organisms specific for different states or ecoregions that span several states, and giving states guidance in how the BLM would be run for specific example scenarios. Reviewer 3 also wrote that comparison between measured effects concentrations and water chemistry-based estimates of criteria values would be a reasonable exercise to see if criteria applied to laboratory solutions are actually protective.

Two reviewers (1, 2) commented on the functionality of the spreadsheets. Reviewer 1 wrote that the spreadsheet of the regression approaches was somewhat cryptic. The reviewer could not determine some of the acronyms or see R squared or goodness of fit values. Reviewer 2 wrote that the provided MLR spreadsheets lack user interface utility and appear as a demonstration of proof-of-concept. Reviewer 2 provided a link to an example of a spreadsheet approach with excellent user utility from the California Department of Toxic Substances Control.

- **Feel free to try your own scenarios to see differences and provide with your review.**

Three reviewers (1, 3, 5) included additional scenarios testing for the influence of DOC. Reviewer 3 performed simulations using the simplified BLM at a fixed hardness with varying pH and DOC concentrations, and for comparison performed the same modeling with MLR approach 2 without DOC as an input parameter. Reviewer 3 found that pH had a much more significant impact than DOC, and stated that these results demonstrate that MLR approaches as currently formulated are dramatically conservative for less sensitive receiving waters. As a result, Reviewer 3 commented that the conservative results from MLR approaches could potentially waste resources trying to address aluminum exceedances even when there are no potential impacts.

In contrast, both Reviewer 1 and Reviewer 5 found that HC5 values increased significantly with the BLM methods when DOC was raised. Reviewer 1 observed that the HC5 values for aluminum were several factors higher using both the full and simplified BLM approaches when the DOC was 5.0 ppm versus 1.0 ppm. Reviewer 5 observed that using the simplified BLM at a DOC of 10 mg/L yielded chronic aluminum values that were substantially higher than those generated using an input value of 1 mg/L DOC for each scenario tested. Also, Reviewer 5 found that changes in pH and hardness did not influence the chronic aluminum value when DOC was 10 mg/L as much as they did when DOC was 1 mg/L.

Regarding other additional scenarios performed, Reviewer 1 found that there is good agreement between the results of the BLM approaches at acidic pH values and lower hardness or Ca, Mg, and alkalinity values, but that the MLR approach 1 had higher aluminum values. Reviewer 1 also stated that a non-linear relationship may be a more appropriate fit to the species sensitivity distribution data at lower pH values, and that this relationship would result in higher HC5 values. Reviewer 5 observed that there were no differences in the chronic values generated when using the full BLM with constant pH and DOC but variable hardness. When using the simplified BLM for the same conditions, however, the chronic values did decrease somewhat.

Reviewer 4 noted that temperature was not presented in the database, and questioned the temperature ranges at which each method would apply. Reviewer 4 used the full BLM in simulations with low temperature, and found that the BLM predictions were very different from those in MLR approach 1. However, the reviewer noted that comparison is difficult when the actual DOC concentrations in the test waters are unknown.

Reviewer	Comments	Response to Comments
Reviewer 1	• **Please draw some conclusions regarding the differences in the values generated and explain your rationale.** 1a. In the scenarios addressed in Table 1, the full BLM and simplified BLM approaches produced similar HC5 values when pH was varied from 7 to 6. I did not see a confidence interval around the SSD regressions, so when using a HC5 as a protective value, it is difficult to determine the variability around the HC5 to distinguished it from background levels of	

<table>
<tr><th>Reviewer</th><th>Comments</th><th>Response to Comments</th></tr>
<tr><td></td><td>
aluminum in water. The spreadsheet of the regression approaches was somewhat cryptic. I could not determine some of the acronyms or see R squared or goodness of fit values. That said, the MLR approach 1 produced similar values to the full and simplified BLM approaches at pH 6 and 7 at a hardness value of 100 mg/L. At a lower hardness concentrations and acidic pH, the MLR approach 1had higher aluminum values than either of the BLM approaches. While MLR approach 2 had aluminum values that were lower than the other three approaches.

Table 2 contains comparisons of the approaches with the pH varied and Ca and Mg concentrations or hardness values halved. The data indicate that there is good agreement between the BLM approaches at acidic pH values and lower hardness or Ca, Mg, and alkalinity values. The MLR approach 1 had higher aluminum values. In looking at the species sensitivity distribution (SSD) plots, at pH 6.0 and 5.5, the data points indicate that a non-linear relationship may be a more appropriate fit to the data. This would result in higher HC5 values.

Table 2. Comparison of HC5 values for aluminum (ppb) for the BLM and MLR 1 approaches with pH 5.5 – 7.0 and lower hardness or Ca, Mg and alkalinity values.
<table>
<tr><td></td><td>pH 7.0</td><td>pH 6.0</td><td>pH 5.5</td></tr>
<tr><td>Full BLM with Ca 13 ppm; Mg 4.0 ppm and Alkalinity 27 ppm</td><td>157</td><td>50</td><td>32</td></tr>
<tr><td>Simple BLM with hardness 50 ppm</td><td>159</td><td>51</td><td>33</td></tr>
<tr><td>MLR approach 1, hardness 50 ppm</td><td>170</td><td>89</td><td>64</td></tr>
</table>
</td><td></td></tr>
</table>

<table>
<tr><th>Reviewer</th><th>Comments</th><th>Response to Comments</th></tr>
<tr><td></td><td>
Table 3 is a brief comparison at the effect of increased DOC concentrations of 5.0 ppm for the BLM approaches at acidic pH values and low hardness or Ca, Mg, and alkalinity concentrations. The HC5 values for aluminum were several factors higher at a DOC of 5.0 ppm versus a DOC of 1.0 ppm. Once again, there was good agreement between the full BLM and simplified BLM approaches.

Table 3. Comparison of HC5 Al values (ppb) and BLMs with acidic pH and DOC of 5.0 ppm
<table>
<tr><td></td><td>pH 6.0</td><td>pH 5.5</td></tr>
<tr><td>Full BLM DOC 5.0 ppm</td><td>182</td><td>158</td></tr>
<tr><td>Simplified BLM, DOC 5.0 ppm</td><td>185</td><td>156</td></tr>
</table>
• Would the simplified BLM be sufficiently protective?

1b. The simplified BLM produced aluminum HC5 values that are very similar to the full BLM for the variations in water quality parameters that were conducted. So, if the full BLM is protective than it appears that the simplified BLM is protective too. For the utilization of the simplified BLM, it still requires a DOC concentration and many states and local government water quality programs typically do not measure this value. Increased variability of the estimation of DOC using the EPA document on estimation methods (EPA, 2016) should be considered in the application of the simplified BLM.

• Would the pH and hardness MLRs be appropriate?

1c. It is difficult to determine if the MLRs are appropriate for use as a tool for aluminum water quality criteria. I cannot determine the suitability of the R squared values or determine if a sensitivity analysis of the inclusion of additional parameters, such as DOC,
</td><td></td></tr>
</table>

Reviewer	Comments	Response to Comments
	was conducted. The BLM approaches use DOC, which as table 2 shows, is a critical parameter to estimate aluminum WQC. Judging the MLR approaches versus the BLM, the MLR approach 1 produced values that were similar at a hardness of 100 mg/L, but under acidic pH and a hardness of 50 ppm, MLR approach 1 values were higher than the BLM values. MLR approach 2 produced values that were lower than the BLM approaches. The use of MLRs to support metals water quality criteria is an active research area. US EPA should reach out to the scientific community, possibly through a workshop or public meeting to see if the scientific community has published or plans to publish alternative MLR approaches to estimate metal criteria. • **Please provide appropriate suggestions.** 1d. The full BLM incorporates the water chemistry parameters that have a major influence on metal bioavailability in aquatic systems. It is the best science put forward to assess effects of metals to biota in aquatic systems (Di Torro et al., 2001; US EPA, 2007). However, the key issue concerning the utility of the BLM is that the users (states and local governments) typically do not measure the parameters to run the full BLM. The US EPA Office of Water (OW) is trying to address this gap through the simplified BLM approach using 4 parameters (DOC still is a problem for regulators to measure), multiple linear regression methods, water effects ratios (WER) and in 2016 the EPA produced a technical document on estimation of water quality parameters (EPA, 2016). The document provides methods to develop estimates for missing parameters using data analysis approaches for water quality parameter default values. To increase utilization of the above methods for aluminum WQC, I would recommend that the US EPA OW develop tiered approaches, with input from the states and the scientific community, to build from what the states and local governments are currently doing in a tier 1 to inclusion of more complex approaches (e.g. BLM) at higher tiers. Using ecoregions, the US EPA could identify where aquatic systems are likely to be at increased risk from aluminum toxicity. They could put forward tiered approaches for areas where there is likely increased risk from aluminum toxicity and therefore an increased level of	

<table>
<tr><th>Reviewer</th><th>Comments</th><th>Response to Comments</th></tr>
<tr><td></td><td>
effort is warranted from the states. Conversely, for aquatic systems where aquatic chemistry would indicate a lower risk to aquatic life, a lower level of effort tiered approach could be used.

• Feel free to try your own scenarios to see differences and provide with your review.

I included these as tables 2 and 3.

<u>Table 1. SCENARIOS</u>
<table>
<tr><th></th><th>Scenario 1</th><th>Scenario 2</th><th>Scenario 3</th><th>Scenario 4</th></tr>
<tr><td>INPUTS</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Temperature (°C)</td><td>21</td><td>21</td><td>21</td><td>21</td></tr>
<tr><td>pH</td><td>7</td><td>6</td><td>7</td><td>6</td></tr>
<tr><td>Hardness (mg/L Ca CO_3)</td><td>100</td><td>100</td><td>50</td><td>50</td></tr>
<tr><td>DOC (mg C/L)</td><td>1.0</td><td>1.0</td><td>1.0</td><td>1.0</td></tr>
<tr><td>HA (%)</td><td>10</td><td>10</td><td>10</td><td>10</td></tr>
<tr><td>Ca (mg/L)</td><td>26</td><td>26</td><td>26</td><td>26</td></tr>
<tr><td>Mg (mg/L)</td><td>8.0</td><td>8.0</td><td>8.0</td><td>8.0</td></tr>
<tr><td>Na (mg/L)</td><td>12</td><td>12</td><td>12</td><td>12</td></tr>
<tr><td>K (mg/L)</td><td>1.4</td><td>1.4</td><td>1.4</td><td>1.4</td></tr>
<tr><td>SO_4 (mg/L)</td><td>56</td><td>56</td><td>56</td><td>56</td></tr>
<tr><td>Cl (mg/L)</td><td>3.8</td><td>3.8</td><td>3.8</td><td>3.8</td></tr>
<tr><td>Alkalinity (mg/L Ca CO_3)</td><td>55</td><td>55</td><td>55</td><td>55</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td></tr>
</table>
</td><td></td></tr>
</table>

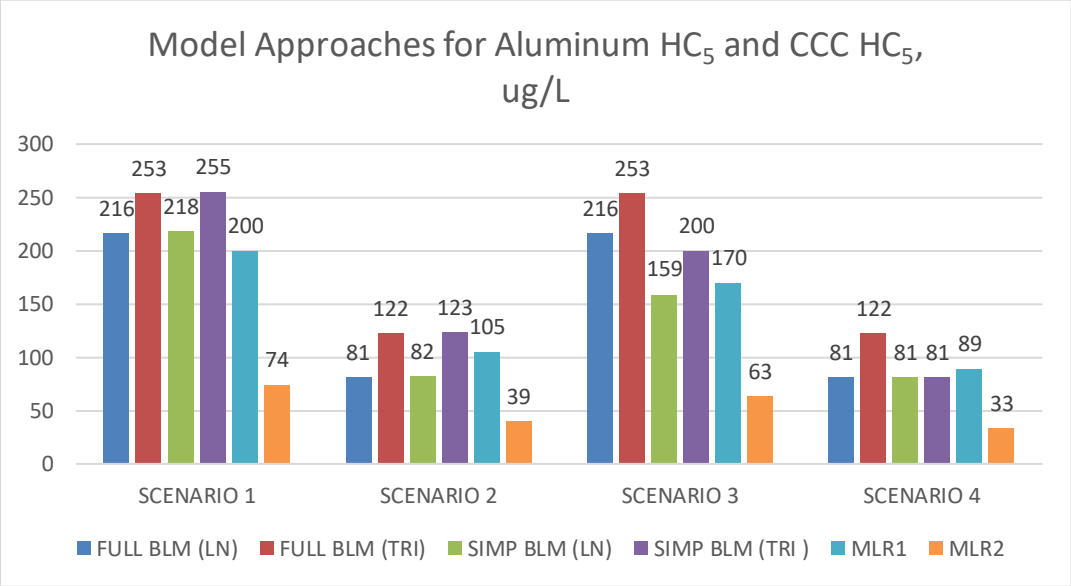
<table>
<tr><th>Reviewer</th><th>Comments</th><th>Response to Comments</th></tr>
<tr><td></td><td>
<table>
<tr><td>APPROACHES</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Full Aluminum BLM</td><td>216</td><td>81</td><td>216
Ca 14;Mg 4 = 157</td><td>81
Ca 14;Mg 4= 50</td></tr>
<tr><td>Simplified Aluminum BLM</td><td>218</td><td>82</td><td>159</td><td>51</td></tr>
<tr><td>MLR COMPARISON Approach 1</td><td>200</td><td>105</td><td>170</td><td>89</td></tr>
<tr><td>MLR COMPARISON Approach 2</td><td>74</td><td>39</td><td>63</td><td>33</td></tr>
</table>
</td><td></td></tr>
<tr><td>Reviewer 2</td><td>

- **Please draw some conclusions regarding the differences in the values generated and explain your rationale.**

As instructed, the values for water quality were input into the Chronic Al BLM “Site Chemistry” and “Simplified Site Chemistry” modules for the four identified scenarios. In addition, the Chronic Criteria Comparison result table calculated by the two MLR approaches in the provided spreadsheet were identified for the relevant four scenarios in the multiple regression model results of the two identified MLR1 and MLR2 approaches. The four scenarios varied only in the pH (6 or 7) and Hardness (50 or 100, mg/L $CaCO_3$) generating four unique combinations. pH, Hardness and their derived or constituent parts (e.g. alkalinity, Ca/Mg) are well understood factors in aluminum bioavailability and influence on aquatic toxicity.

The results from the modeling exercise are shown in Figure 1 as a scenario clustered bar graph for the six modeling approaches. Initial visual exploration of the results shows that MLR2 (green bar) consistently yield the lowest HC_5, and in a pattern outside the other results. While models have no guarantee of protection of all species for all conditions and co-exposures, MLR2 shows the lowest values of all models for all scenarios. This can be interpreted as “most protective” or alternately “overly protective,” relative to the other

</td><td></td></tr>
</table>

Reviewer	Comments	Response to Comments
	considered models. With an average MLR2 scenario result of 52 ug/L, with a range of 33 to 74 ug/L, compared to the current 87 ug/L chronic aluminum value, this result would require an average Water Effects Ratio (WER) multiplier of 0.6, even at higher pH levels where Al toxicity and bioavailability is known to decrease in many toxicity trials. MLR2 appears to be less responsive to scenario hardness and pH inputs. The MLR2 results appear to be "overly protective" across the pH and hardness ranges explored in the modeling exercise, and thus there is justification for elimination from further consideration in this review.  **Figure 1. Aluminum HC_5 model approaches.** The model result clusters observed in Figure 1 can be further explored for variability. Table 1 shows the comparative analysis of aluminum aquatic life criteria model results, ug/L, for Scenarios 1-4 detailing the mean, standard deviation, and coefficient of variation of all model results after elimination of MRL2. The coefficient of variation across all model results for Scenarios 1-4 ranges from 10.7% to 20% demonstrating acceptable model consistency. The magnitude and direction from mean, as difference and as a percent	

Reviewer	Comments	Response to Comments

across Scenarios 1-4 is also explored in Table 1. For visual reference in the table, negative differences from the mean are shaded red and positive deviations are shaded blue. The FULL BLM (TRI) model showed consistently positive (blue) difference from the mean and the SIMP BLM (LN) model showed consistently negative (red) difference from the mean. In overall percent magnitude and direction from mean, the pattern demonstrated in Table 1 is:

FULL BLM (TRI) > SIMP BLM (TRI) > FULL BLM (TRI) > MRL1 > SIMP BLM (LN)

While this magnitude and direction from mean of the model result cluster cannot identify a "correct" result, it can be stated that within these model results SIMP BLM (LN) is the "most protective" and FULL BLM (TRI) is the "least protective" across Scenarios 1-4. However, after eliminating MRL2, all model results are within the 20% maximum coefficient of variation found across Scenarios 1-4. This limited variability suggests the tested models, not including MRL2, are performing with satisfying consistency with pH changes from 6 to 7 and hardness changes from 50 to 100 mg/L $CaCO_3$.

Table 1. Comparative analysis of Aluminum Aquatic Life Criteria model results, ug/L, for Scenario 1-4.

	Scenario 1	MDM	%MDM	Scenario 2	MDM	%MDM	Scenario 3	MDM	%MDM	Scenario 4	MDM	%MDM
FULL BLM (LN)	216	-12	-5.4	81	-22	-21.1	216	16	8.2	81	-10	-10.8
FULL BLM (TRI)	253	25	10.8	122	19	18.9	253	53	26.8	122	31	34.4
SIMP BLM (LN)	218	-10	-4.6	82	-21	-20.1	159	-41	-20.3	81	-10	-10.8
SIMP BLM (TRI)	255	27	11.6	123	20	19.9	200	0	0.2	81	-10	-10.8
MLR1	200	-28	-12.4	105	2	2.3	170	-30	-14.8	89	-2	-2.0
	228	24.4		103	20.5		200	37.6		90.8	17.8	
	AVG	SD		AVG	SD		AVG	SD		AVG	SD	
	CV	10.7		CV	20.0		CV	18.8		CV	19.6	
CV = Coefficient of variation												
MDM = Magnitude and direction from mean												
%MDM = Percent magnitude and direction from mean												

Reviewer	Comments	Response to Comments
	• **Would the simplified BLM be sufficiently protective?** In the analysis of model results shown in Figure 1 and Table 1, the SIMP BLM results cluster well with the other considered models within 20% variability. SIMP BLM (LN) is the "most protective" in magnitude and direction of the five approaches considered in Table 1. SIMP BLM yields a result (81 ug/L) that is lower than the current CCC (87 ug/L) when modeled with pH=6 and hardness at 50 mg/L $CaCO_3$ with both log normal and triangle distribution and thus can be considered "sufficiently protective." The SIMP BLM model only requires input of pH, hardness, dissolved organic carbon, and temperature if the water resource under analysis is known to compare well with the default North American ion ratios. Thus the cost basis for collecting a larger data pool to better understand aluminum hazard in a water resource versus the 10 parameter FULL BLM would allow better, more robust, characterization of risk. The increased data, for example during high-flow and low-flow periods of the hydrograph of a water resource, could be valuable knowledge in managing risk at a lower cost for the regulated community. Thus SIMP BLM would be "sufficiently protective" under the requirements of the CWA. • **Would the pH and hardness MLRs be appropriate?** For the reasons stated above, MLR2 appears to generate relatively "over protective" results when compared to the other models and existing CCC WQC. pH and hardness MRL1 model results appear to cluster well with the other considered models within 20% variability and thus can be considered "sufficiently protective." I did not explore the MRL1 and MRL2 model result difference in detail, however it appears that use of Final Acute Value (FAV) and Final Acute-Chronic Ratio (FACR) in the MRL2 model produces an "over protective" result compared to all of the other model approaches in the present analysis. MRL1 produces a result (89 ug/L) close to the current CCC (87 ug/L) when modeled with pH=6 and hardness at 50 mg/L $CaCO_3$. MLR1 would be appropriate for WQC development, however MLR2 would not. The low data input requirements of MLR1 (pH and hardness), are attractive in managing cost basis for compliance in the regulated community in addition to allowing the development of a larger data pool to better characterize risk of a body of water during annual hydrograph and water quality changes. At present, the provided MLR spreadsheets lack user interface utility and appear as a demonstration of proof-of-concept. If advanced into full user utility, pH and hardness MRL1 spreadsheet approaches could be "sufficiently protective" under the requirements	

<table>
<tr><th>Reviewer</th><th>Comments</th><th>Response to Comments</th></tr>
<tr><td></td><td>of the CWA in the requested scenario analysis. An example of a spreadsheet approach with excellent user utility is the California Department of Toxic Substances Control LEADSPREAD 8 tool for assessing adverse human health effects from lead that is found at: http://www.dtsc.ca.gov/AssessingRisk/leadspread8.cfm</td><td></td></tr>
<tr><td>Reviewer 3</td><td>The results of this exercise are included in the table below (Table 1). And graphically in the figure below that (Figure 1). The Multiple Linear Regression (MLR) values were read directly from the lookup table and the Biotic Ligand Model (BLM) results were calculated using the default fitting of the sensitivity distribution (i.e., Estimated HC5 (5-percentile) based on Log Normal Distribution Model) and no chronic toxicity data sets were excluded in the modelling. The US “run” button was used.

Table 1 Scenarios from the original charge question with the modelling results inserted.

<table>
<tr><th></th><th>Scenario 1</th><th>Scenario 2</th><th>Scenario 3</th><th>Scenario 4</th></tr>
<tr><td>INPUTS</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Temperature (°C)</td><td>21</td><td>21</td><td>21</td><td>21</td></tr>
<tr><td>pH</td><td>7</td><td>6</td><td>7</td><td>6</td></tr>
<tr><td>Hardness (mg/L Ca CO_3)</td><td>100</td><td>100</td><td>50</td><td>50</td></tr>
<tr><td>DOC (mg C/L)</td><td>1.0</td><td>1.0</td><td>1.0</td><td>1.0</td></tr>
<tr><td>HA (%)</td><td>10</td><td>10</td><td>10</td><td>10</td></tr>
<tr><td>Ca (mg/L)</td><td>26</td><td>26</td><td>26</td><td>26</td></tr>
<tr><td>Mg (mg/L)</td><td>8.0</td><td>8.0</td><td>8.0</td><td>8.0</td></tr>
<tr><td>Na (mg/L)</td><td>12</td><td>12</td><td>12</td><td>12</td></tr>
<tr><td>K (mg/L)</td><td>1.4</td><td>1.4</td><td>1.4</td><td>1.4</td></tr>
<tr><td>SO_4 (mg/L)</td><td>56</td><td>56</td><td>56</td><td>56</td></tr>
<tr><td>Cl (mg/L)</td><td>3.8</td><td>3.8</td><td>3.8</td><td>3.8</td></tr>
</table>
</td><td></td></tr>
</table>

Reviewer	Comments					Response to Comments
	Alkalinity (mg/L Ca CO_3)	55	55	55	55	
	APPROACHES					
	Full Aluminum BLM	216	81	216	81	
	Simplified Aluminum BLM	218	82	159	51	
	MLR COMPARISON Approach 1	200	105	170	89	
	MLR COMPARISON Approach 2	74	39	63	33	

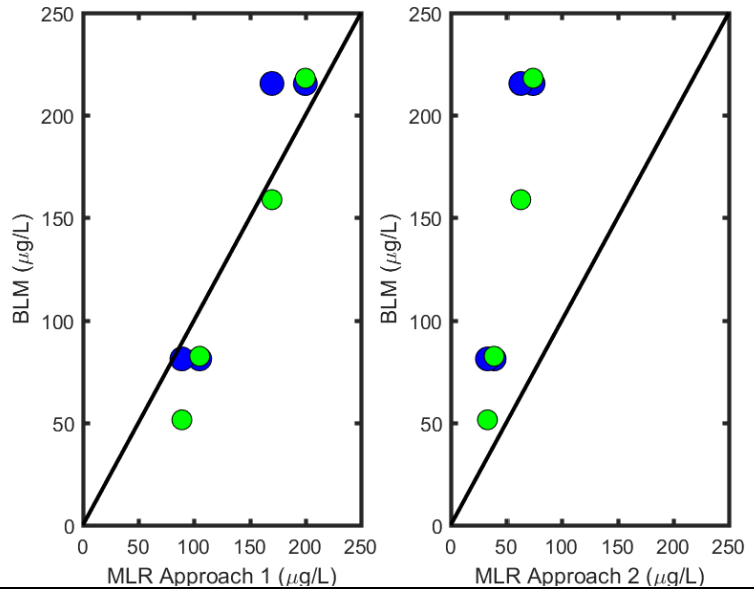


Figure 1: Comparison between MLR and BLM approaches to estimate aluminum chronic criteria. MLR results are on the x-axis; the left plot is using "Approach 1" and the right plot is using "Approach 2". The BLM results are presented on the y-axis with blue symbols for the full BLM and green symbols for the simplified BLM.

Reviewer	Comments	Response to Comments
	• **Please draw some conclusions regarding the differences in the values generated and explain your rationale.** As expected, all 4 types of criteria determination show the same general trend with criteria values highest when hardness is high and pH is high. Lower hardness and lower pH waters are less protective and thus the estimated criteria is lower. pH has a more significant impact on the results than hardness. The numerical values from MLR using Approach 1 are much more similar to the BLM estimates than for Approach 2. In Figure 1 the left plot (Approach 1) shows data centered around the 1:1 line. By contrast, Approach 2 estimates much lower criteria values than BLM or Approach 2 MLR. Finally, the simplified and full BLM estimates are very similar with the green and blue dots on Figure 1 almost overlapping. • **Would the simplified BLM be sufficiently protective?** The provided scenarios are not sufficient to assess if the simplified BLM will be protective or not. The simulations so far show that for these four cases the simplified BLM agrees very closely with the full BLM. To assess protection though it is necessary to compare to actual toxicity measurements, not to other models. To this end the 23 lowest values in the chronic dataset provided with the Al BLM were run as test cases and the predicted criteria values compared to the measured effects concentrations in Figure 2. To run this simplified BLM the water chemistry values for temperature, pH, DOC and hardness were input. Hardness was estimated according to the equation hardness=2.5*Ca+4.1*Mg where hardness is in mg $CaCO_3$/L and Ca and Mg are both in mg/L. The chronic criteria estimated from the simplified BLM are within a factor of two of the measured values except for the lowest measured effect concentration at 20 µg/L for *C. Dubia* in low DOC, low hardness, slightly acidic pH water. Based on this result the simplified BLM is sufficiently protective.	

Reviewer	Comments	Response to Comments

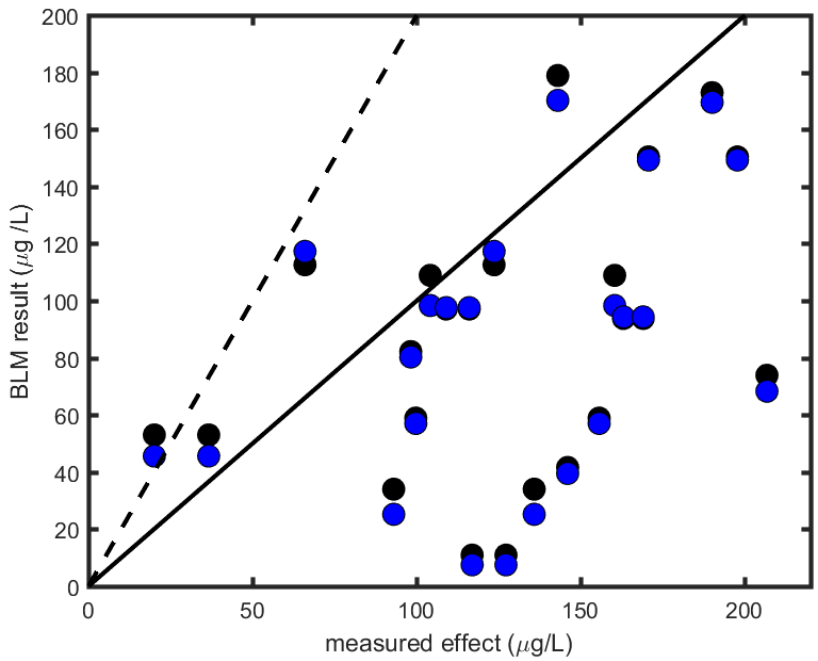
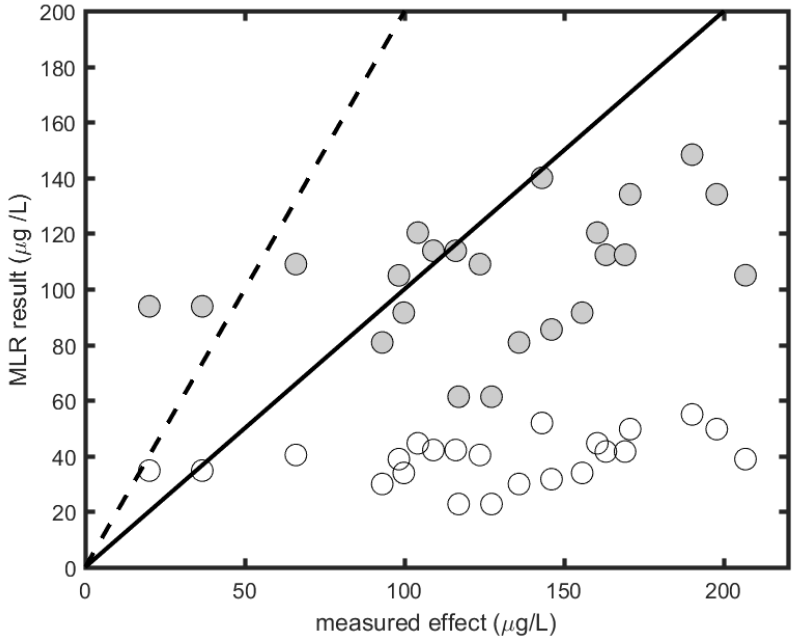


Figure 2: Simplified BLM modelling results of the 23 lowest chronic endpoint values in the chronic BLM database for aluminum provided with the Windward BLM model, versus the measured effect concentration. The solid black line corresponds to the 1:1 line and the dashed line is two times the one to one value. The blue circles are the results of running the full BLM calculation on the given water chemistry.

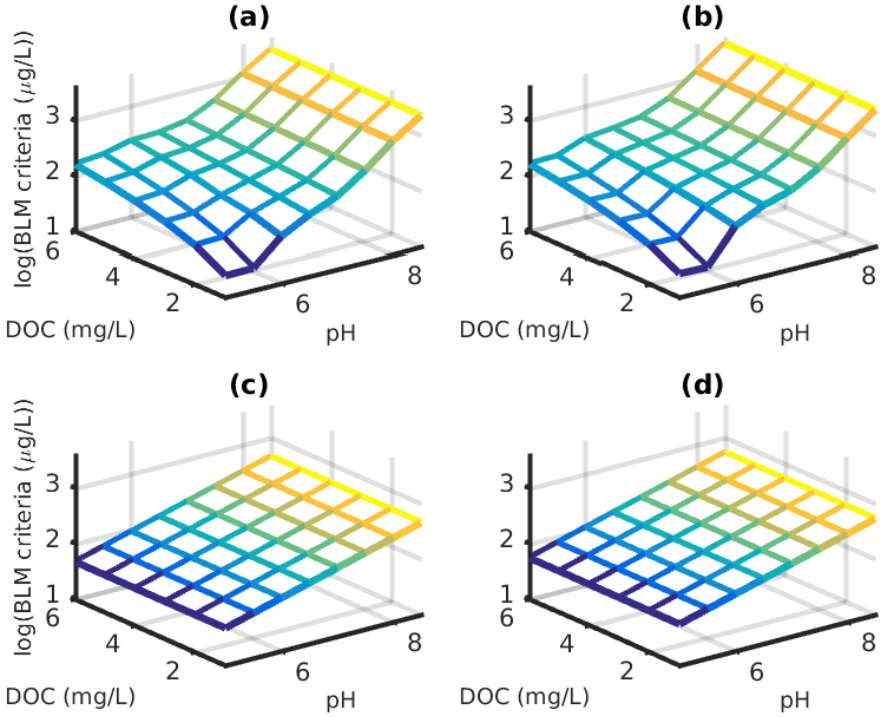
For comparison, Figure 2 shows full BLM runs on these same water chemistry values. The data for the two approaches almost overlap with full BLM values tending to be slightly lower than the simplified BLM values. Using the full BLM, even the lowest chronic toxicity value is now almost within a factor of two of the measured effect concentration.

- **Would the pH and hardness MLRs be appropriate?**

Again to decide appropriateness of the model predictions it is important to compare to toxicity results and not simply compare to other model results. For the same 23 lowest

Reviewer	Comments	Response to Comments
	effect concentration samples in the chronic database the two MLR approaches were tested. The results are presented in Figure 3.  Figure 3: Results of MLR estimates of chronic criteria values compared to measured values. The open symbols correspond to "Approach 2" and the filled symbols correspond to "Approach 1". The solid line is the 1:1 line and the dashed line is a line with a slope of 2 and an intercept of zero. In MLR estimation of the criteria values, Approach 2 is more conservative than Approach 1 with criteria values well below the measured effects. For Approach 1 the MLR estimated values are not as dramatically different from the measured effects. Both approaches yield values no greater than a factor of two above measured toxicity values, except for the two lowest measured values in Approach 1 which are a factor of approximately 4 greater than the measured toxicity value.	

Reviewer	Comments	Response to Comments
	• **Please provide appropriate suggestions.** Comparison between measured effects concentrations and water chemistry based estimates of criteria values is a reasonable exercise to see if criteria applied to laboratory solutions would be protective or not. In all instances the MLR Approach 2 is protective but many of these values are overly conservative with MLR estimated values as much as 5X lower than observed effects concentrations. This conservative tendency would be even more dramatic for more protective, i.e., harder and higher pH, waters. Approach 1 sometimes generates values that are not protective thus the simplified or full BLM are the most appropriate methods to estimate chronic criteria. The full BLM is slightly more protective than the simplified BLM but the simplified BLM performs almost identically with much reduced data requirements. I understand that EPA sets guidelines and the States use these as a starting point for developing their specific approaches. To increase "buy in" by the States it might make sense to present test cases and test organisms specific for different States, or ecoregions that span several States. Essentially give the States some guidance in how the BLM would be run for specific example scenarios. I expand on the suggestion of including different scenarios in the BLM documentation in my response to Charge Question 5 below. • **Feel free to try your own scenarios to see differences and provide with your review.** Hardness and pH only estimation of criteria are potentially limiting if DOC is significant in the receiving waters. To test for the influence of DOC, simulations were performed using the simplified BLM at a fixed hardness of 50 or 100 mg $CaCO_3$/L over a pH range (5 to 8.5) for DOC from 1 to 6 mg C/L. These conditions are within the reported range of the BLM calibration (Table 7-1 in the Users' Guide). For comparison the same modelling was performed with MLR Approach 2, but obviously without DOC as an input parameter. What is apparent from these models is that pH has by far the most significant impact; DOC does increase the criteria value but more gradually compared to pH (Figure 4). At pH values below 6 the shapes of the MLR and BLM surfaces are very different, as DOC has a more significant impact, but is not included in MLR modelling. In comparing the MLR and BLM approaches though it is apparent that MLR tends to have much lower concentrations (the z-axis on Figure 4 are all at the same scale, and the scale is logarithmic). To be protective, the most effort and attention is focused on the low range of species sensitivity	

Reviewer	Comments	Response to Comments
	and for sensitive receiving waters, but these results demonstrate that MLR approaches as currently formulated are dramatically conservative for less sensitive receiving waters. This could potentially waste resources trying to address aluminum exceedances even when there are no potential impacts. BLM estimated concentrations are much higher than MLR estimates for high pH waters, by as much as an order of magnitude.  Figure 4: Criteria values from the simplified BLM (a and b) and from Approach1 MLR (c and d) for hardness of 50 (a and c) and 100 (b and d) mg $CaCO_3$/L. pH was varied in both models and concentration of DOC was also varied in the BLM simulations. The z-axis is shown on a logarithmic scale in order to differentiate values.	

<table>
<tr><th>Reviewer</th><th>Comments</th><th>Response to Comments</th></tr>
<tr><td>Reviewer 4</td><td>
<u>SCENARIOS</u>
<table>
<tr><th></th><th>Scenario 1</th><th>Scenario 2</th><th>Scenario 3</th><th>Scenario 4</th></tr>
<tr><td>INPUTS</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Temperature (°C)</td><td>21</td><td>21</td><td>21</td><td>21</td></tr>
<tr><td>pH</td><td>7</td><td>6</td><td>7</td><td>6</td></tr>
<tr><td>Hardness (mg/L Ca CO_3)</td><td>100</td><td>100</td><td>50</td><td>50</td></tr>
<tr><td>DOC (mg C/L)</td><td>1.0</td><td>1.0</td><td>1.0</td><td>1.0</td></tr>
<tr><td>HA (%)</td><td>10</td><td>10</td><td>10</td><td>10</td></tr>
<tr><td>Ca (mg/L)</td><td>26</td><td>26</td><td>26</td><td>26</td></tr>
<tr><td>Mg (mg/L)</td><td>8.0</td><td>8.0</td><td>8.0</td><td>8.0</td></tr>
<tr><td>Na (mg/L)</td><td>12</td><td>12</td><td>12</td><td>12</td></tr>
<tr><td>K (mg/L)</td><td>1.4</td><td>1.4</td><td>1.4</td><td>1.4</td></tr>
<tr><td>SO_4 (mg/L)</td><td>56</td><td>56</td><td>56</td><td>56</td></tr>
<tr><td>Cl (mg/L)</td><td>3.8</td><td>3.8</td><td>3.8</td><td>3.8</td></tr>
<tr><td>Alkalinity (mg/L Ca CO_3)</td><td>55</td><td>55</td><td>55</td><td>55</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>APPROACHES</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Full Aluminum BLM</td><td>216</td><td>81</td><td>216 (157)</td><td>81 (49)</td></tr>
<tr><td>Full Aluminum BLM at 10oC</td><td>101</td><td>21</td><td>61</td><td>21</td></tr>
</table>
</td><td></td></tr>
</table>

Reviewer	Comments	Response to Comments

Full Aluminum BLM at 5ppmDOC	285	208	285	208
Simplified Aluminum BLM	218	82	159	51
MLR COMPARISON Approach 1	200	105	170	89
MLR COMPARISON Approach 2	74	39	63	33

Note: The hardness and Ca and Mg concentrations of the four scenarios are not corresponding. Concentrations of Ca and Mg in scenarios 3 and 4 for the case of lower hardness should be lower than those for scenarios 1 and 2. The highlighted data in parentheses (yellow) were the model predicted values when concentrations of Ca and Mg were reduced to a half concentrations of scenarios 1 and 2 (i.e., Ca = 13 mg/L, Mg = 4 mg/L). Other chemistry was the same as scenarios 1 and 2.

In general, the predictions by the full and simplified BLMs are pretty much similar. However, the water chemistry needed for the simplified model is much less than that for the full model. Therefore, if the BLM is used for setting water quality guidelines, the simplified model is more economic and easier to use. The model also takes into account the important bioavailability modifying factors for metals, such hardness, DOC, and pH.

The two hardness and pH approaches gave different predictions. The predictions of approach 2 was about 35% predictions of approach 1. The difference between the two approaches is the equation factor of -0.3325 (approach 1) vs -1.326 (approach 2). I am not sure how these equation factors were derived but only approach 2 was noticed in the Chronic Ranked tap of the excel file (cell B17, C52). The Chronic Ranked tap also indicates that the equation factor of -0.3325 in cell G46 was derived from the most sensitive species (Lampsilis, chironomus, salvelinus, salmo) while the factor of -1.326 in cell D20 was below the whole data set for all species. Was this factor derived from the data for all species? The CCC derived from most sensitive species should be lower than the CCC derived from all species including less sensitive species. The Criteria Comparison tap shows opposite outcomes. It seems to be conflict? Ultimately, the approach with lower

Reviewer	Comments	Response to Comments
	CCC is more protective. However, it needed clarification here before answering the question which approach is more protective? In comparison between the hardness/pH approaches and the BLM, the predictions by approach 1 and the BLMs are very much similar. Approach 1 only take effects of hardness and pH into account while the BLM includes effects of other factors such as temperature and DOC-an important modifying factor of metal bioavailability and toxicity. It is important to mention here is that 54% of the data used to derive hardness and pH equation are missing DOC. Other data have very low DOC concentrations, basically at DOC of ID water. The hardness and pH approaches should be used to predict toxicity for the range of water quality that was used to develop the approach. I am not sure how well the hardness and pH approaches will predict for water with higher DOC which is likely the case for the natural environment. Also temperature was not presented in the data base (might be in the published papers) so don't know what temperature range would be applied. I tried the full BLM with low temperature, such as 10oC which is more realistic for cold water fish like trout and higher DOC. The BLM predictions are much different from approach 1 (red data in table). It is difficult to compare when we don't know the actual DOC concentrations in the test waters. In addition, I don't understand about the low DOC concentrations in this data base (C_FW Core Chronic tap of the excel file), especially for chronic tests at which the tested organisms were fed. Organic materials would be released from the food. Our chronic tests with *D. magna* and snails in many years have shown a DOC concentration of about 4 mg/L in test water collected from the test chambers. A DOC of 0.5 mg/L for *C. dubia* in this data base seems to be the DOC of test water with no food. This is true for acute tests with no feeding but not for chronic tests. I was trying to run the simplified model with higher DOC and different temperature but got technical problems (see presentation in question 5 below).	
Reviewer 5	The Full BLM, Simplified BLM and the Excel results from MLR Comparison Approach 1 generated similar chronic values for aluminum in the five scenarios examined (See Table 1). Using MLR Comparison Approach 2, the chronic values were much lower than those generated by the other three methods (approximately 37% of those using MLR Approach In comparing the Full BLM and Simplified BLM, the chronic values using scenarios 1 and 2 (Hardness=100 mg/L; pH of 7 and 6, respectively) were very similar. However, when a	

Reviewer	Comments	Response to Comments
	lower hardness value of 50 mg/L was used in scenarios 3 and 4, the Simplified BLM generated chronic values below those from the Full BLM, which would be more protective of aquatic life. Alternatively, when a higher hardness value of 200 mg/L and a higher pH of 7.5 was used in scenario 5, the Simplified BLM generated a chronic value well above that from the Full BLM**. The Simplified BLM would be sufficiently protective in all cases except scenario 5 (higher hardness and pH).** However, this scenario (conditions) would be less commonly encountered of the five examined in most aquatic systems. Using the Full BLM, when keeping pH constant and changing only hardness at a DOC of 1 mg/L, there were no differences in the chronic values generated. When using the Simplified BLM for the same conditions, the chronic values did decrease somewhat; however, these values would still be sufficiently protective. Using the Simplified BLM, at a higher DOC of 10 mg/L the chronic aluminum values generated were substantially higher than using an input value of 1 mg/L DOC with all of the scenarios (Table 1). Also, at 10 mg/L DOC, change in pH from 7 to 6 and change in hardness from 100 to 50 mg/L, did not influence the chronic aluminum value as much as at 1 mg/L DOC. Using MLR Approach 2, where CCC is calculated by the FAV/FACR, generated much lower chronic aluminum values than the other approaches. Application of the FACR can bias the CCC low, which seems to be occurring here. **In my opinion, MLR Approach 2 is not as accurate as the other three.** Alternatively, MLR Approach 1 generated chronic aluminum values that were very similar to those calculated by the Full and Simplified BLMs. The MLR Approach 1 uses an equation which is based largely on pH and hardness, both of which greatly influence predicted aluminum toxicity in the BLM. Differences between these approaches are more apparent when DOC values are higher, as the BLM uses DOC as an input variable and the MLR Approaches do not.	

Reviewer	Comments	Response to Comments

Table 1. Hypothetical Scenarios for Water Chemistry at Different Sites

	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5
INPUTS					
Temperature (°C)	21	21	21	21	21
pH	7	6	7	6	7.5
Hardness (mg/L Ca CO_3)	100	100	50	50	200
DOC (mg C/L)	1.0	1.0	1.0	1.0	1.0
HA (%)	10	10	10	10	10
Ca (mg/L)	26	26	26	26	26
Mg (mg/L)	8.0	8.0	8.0	8.0	8.0
Na (mg/L)	12	12	12	12	12
K (mg/L)	1.4	1.4	1.4	1.4	1.4
SO_4 (mg/L)	56	56	56	56	56
Cl (mg/L)	3.8	3.8	3.8	3.8	3.8

Reviewer	Comments						Response to Comments
	Alkalinity (mg/L Ca CO_3)	55	55	55	55	55	
	APPROACHES						
	Full Aluminum BLM	215.546	81.10529	215.5462	81.10529	355.24786	
	Simplified Aluminum BLM	218.11829	82.3856	158.82442	51.36953	465.27683	
	Simplified Aluminum BLM (+10 mg/L DOC)	375.98963	369.5081	334.78604	368.44722	570.46629	
	MLR COMPARISON Approach 1	200	105	170	89	326	
	MLR COMPARISON Approach 2	74	39	63	33	121	

2.2 Do you believe the scientific and theoretical foundation of the chronic aluminum BLM is sound and defensible? Does the aluminum BLM improve the Agency's ability to predict toxicity to water column organisms due to aluminum in comparison to the currently applied dissolved aluminum concentration criterion?

Four reviewers (1, 2, 3, 5) wrote that the scientific and theoretical foundation of the chronic aluminum BLM was sound and defensible, and Reviewer 4 affirmed the theory of the BLM. Three of these reviewers (1, 2, 3) commented on historical successes in the application of the BLM. Reviewer 1 discussed a partnership among industry, academia, and government that used the BLM to develop a national water quality criterion for copper. The reviewer noted that the partnership used the BLM to address the bioavailability of metals in aquatic systems and their acute toxicity to fish. Reviewer 1 also noted that the BLM has been peer-reviewed in the open literature and by the EPA Science Advisory Board. Reviewer 2 made similar comments, noting that the BLM has a significant history of use in the regulatory arena and general acceptance in the scientific community. Reviewer 3 commented that the BLM for acute and chronic toxicity predictions and risk assessment for many metals is a mature area of science. The reviewer commented that BLM approaches have a long history of success, and that they have demonstrated an ability to protect the environment while also respecting the economic and social benefits of the metals industry.

Regarding the scientific foundation of the BLM method, Reviewer 3 noted that BLMs are based on fundamental physical chemistry, that the relationship between toxicity and short-term bioaccumulation is well established, and that numerous papers have linked accumulation to toxic effects. In addition, Reviewer 3 noted that knowledge of an underlying mechanism of toxicity is very important for establishing BLMs as an approach for criteria determination. The reviewer commented that because the BLM is based on a known mechanism, the model should apply for all conditions within the ranges of data used to derive the model parameters. Reviewer 3 stated that the toxic mechanisms for aluminum seem to be fairly well established in the literature, and this literature base helps establish a scientifically sound foundation for a chronic aluminum BLM. Reviewer 4 noted that the BLM takes the chemistry of metal speciation, bioavailability, and interaction of metals with organisms into account. Reviewer 4 commented that the BLM is a scientific base model, but that the regression approach is more statistically sound and that the details of other chemistry such as hardness and pH are not quantitatively explained.

Three reviewers (1, 2, 5) stated that the aluminum BLM represented an improved tool for EPA. Reviewer 1 commented that the aluminum BLM provides the US EPA with a state-of-the-science tool to predict the toxicity of aluminum to aquatic organisms. The reviewer also stated that the additional research on simplified BLM and estimation methods gives users additional tools for using the BLM for water quality criteria. Reviewer 1 commented that the efficacy of simplified BLM and other approaches is an active research area, and stated that US EPA may need to evaluate current studies of these methods for their application in aluminum criteria. Reviewer 1 also recommended that the aquatic chemistry of aluminum be discussed prominently in any criteria. Reviewer 2 stated that the BLM provides a robust site-specific assessment tool, and that in light of the BLM's treatment of fundamental water chemistry thermodynamics and multiple organism toxicity endpoints, the aluminum BLM is an advancement from the current dissolved aluminum CCC. Reviewer 5 noted that a strength of the BLM is that it uses a large database with chronic values for algae, inverts, and sensitive fish species. Additionally, the BLM uses more water chemistry parameters to determine the chronic aluminum value than the current dissolved aluminum concentration criterion. Based on both these factors, Reviewer 5 believed that the BLM will improve the Agency's ability to predict toxicity of aluminum to water column organisms.

Reviewer	Comments	Response to Comments
Reviewer 1	2a. Yes, the BLM has a solid scientific and theoretical basis (Di Torro, et al., 2001; Santore et al., 2001; Paquin et al., 2002). A partnership among industry, academia, and government utilized research advancements in the aquatic toxicity of metals to fish to effectively develop a national water quality criterion for copper to protect aquatic organisms. This effort used the BLM to address the bioavailability of metals in aquatic systems and their acute toxicity to fish. The BLM adjusts the water concentration that causes acute toxicity of metals to aquatic organisms by calculating the relative binding affinity of all anions in the water to the biotic ligand of the gill. The partnership completed a BLM-based freshwater quality criterion document for copper in 2007 (US EPA, 2007) (http://www.epa.gov/waterscience/criteria/aqlife.html). The BLM has been peer reviewed in the open literature and by the EPA Science Advisory Board. 2b. The aluminum BLM provides the US EPA with a state of the science tool to predict the toxicity of aluminum to aquatic organisms. The additional research to incorporate a simplified BLM and the US EPA document on estimation methods provides users with additional tools to utilize the BLM approach for WQC. The US EPA OW deserves to be commended for these advancements in the application of these approaches to improve and protect our Nation's water bodies. As I stated in response to earlier questions, the evaluation of various approaches to assess the risks of metals to aquatic systems is an active research area and current studies and their associated publications using MLR, simplified BLM, estimations methods and other approaches may need to be evaluated by US EPA for their application in aluminum criteria. A tiered approach that would have input from the users of WQC (states and local governments) could assist in focusing the more complex tools, e.g. the BLM, at the aquatic systems most at risk from aluminum toxicity. The aquatic chemistry of aluminum should be discussed prominently in any criteria. Aluminum in acidic aquatic systems is a major factor in causing toxicity to aquatic organisms. Aluminum is more soluble under more acidic and more alkaline conditions and relatively insoluble at pH 6 to 8 (Gensmer and Playle, 1999).	
Reviewer 2	The BLM addresses the formation of metal complexes and competitive binding at biotic ligand sites in the assessment of bioavailability, exposure, and hazard potential. The	

Reviewer	Comments	Response to Comments
	inorganic environmental chemistry of aluminum in water is complex but straight forward with several oxyhydroxides dominating typical water chemistry. The demonstrated capability of BLM to model free ion and complex concentrations is sound and defensible, with a significant history of use in the regulatory arena and general acceptance in the scientific community. The complex water chemistry of aluminum, including complex formation, ligand binding competition and equilibria, the BLM provides a robust site specific assessment tool. With embedded fundamental water chemistry thermodynamics and multiple organism toxicity endpoints, the aluminum BLM is an advancement from the current dissolved aluminum CCC, and thus improves and broadens the Agency's ability to predict toxicity and thus manage CWA responsibilities.	
Reviewer 3	Yes the scientific and theoretical foundation of the chronic aluminum BLM is sound and defensible. The BLM for acute and chronic toxicity predictions and risk assessment for many metals is a mature area of science. I think it is fair to say that most scientists are so convinced of the foundation of the BLM that they are surprised the adoption is not more wide-spread in regulatory application. BLM approaches have a long history of success and represent a distinct advantage over application of the precautionary principle. They have demonstrated an ability to protect the environment while also respecting the significant economic and social benefits of having a strong metals industry. Biotic ligand models are based on very old and fundamental physical chemistry; that is to say the application of solving for the equilibrium position of multiple simultaneous reactions. This practice was well established by Garrels and co-workers in the 1960s and is often referred to as geochemical modelling. The BLM does geochemical modelling with the application of additional simultaneous reactions for metal complexation and cation competition at the biotic ligand, or site of toxic action. The observation that toxicity is proportional to short term bioaccumulation is well established and numerous papers have linked accumulation to toxic effects. The link between bioaccumulation at fish gills and the mode of toxicity for silver and copper is well established (see citations in the aluminum BLM User's Guide). They are acute ionoregulatory toxicants. Mechanistic information exists for other metals as well. The knowledge of the underlying mechanism of toxicity is the final piece of information that really establishes BLMs as an excellent approach to criteria determination based on	

Reviewer	Comments	Response to Comments
	receiving water chemistry. Since there is a known mechanism, and the model is mechanistic, it should apply for all conditions (within constraints of the data ranges used to derive the parameters). I am less familiar with the aluminum toxicology literature to know if the mechanism of toxicity is as well established. To investigate the state of the art though, I did consult a recent review chapter (Wilson, R. W. "Aluminum" in Homeostatis of Toxicology of Non-essential Metals, Wood, C.M., Farrell, A.P. and Brauner, C. J. (Eds.), Elsevier, pp. 68-104) and I see that ionoregulation, interference with Ca^{2+} tight junctions and respiratory impairment are all known mechanisms of aluminum toxicity to aquatic organisms. Thus, it seems the toxic mechanisms for aluminum are fairly well established and this is the final consideration in establishing the scientifically sound foundation for a chronic aluminum BLM.	
Reviewer 4	I do believe the theory of the BLM. It takes the chemistry that change metal speciation and bioavailability and interaction of metal with organisms into account. It is a scientific base model while the regression approach is more statistic sound and the details of other chemistry rather hardness and pH is not quantitatively explained.	
Reviewer 5	The BLM is used to predict a site specific chronic hazard concentration for aluminum. A strength of the BLM is that it uses a large database with chronic values for algae, inverts (mainly *C. dubia* and *D. magna*, as well as others), and sensitive fish species. Additionally, the BLM considers various water chemistry parameters (i.e. temperature, pH, DOC, hardness, alkalinity, CO_2, and major cations and anions) to determine the chronic aluminum value. The premise is that cations will compete with aluminum ion to bind the biotic ligand (i.e. binding site on fish gill) and anions could complex aluminum, thereby changing its speciation and (usually reducing) toxicity. The scientific and theoretical foundation of the chronic aluminum BLM is both sound and defensible. Because a larger database of chronic toxicity data is used, and more water quality parameters are used than in the currently applied dissolved aluminum concentration criterion, I believe that the BLM will improve the Agency's ability to predict toxicity of aluminum to water column organisms.	

2.3 Currently, States use Water Effects Ratio (WER) adjustment to the 87 ug/L chronic aluminum value. Do you think the application of the BLM or MLR as a site-specific adjustment reduces uncertainty associated with metals bioavailability and toxicity? Please explain.

Four reviewers (2, 3, 4, 5) thought that using the BLM or MLR reduced uncertainty compared with using the WER. Reviewer 5 wrote that there is more uncertainty using the WER adjustment than in using the BLM or MLR, as the BLM and MLR methods incorporate additional water chemistry parameters and consider multiple species. Similarly, Reviewer 4 commented that while WER matches some important water quality parameters, many others that can influence metal speciation and bioavailability are not included. In contrast, the BLM takes everything into account, thus is more realistic and provides less uncertainty.

Reviewer 2 and Reviewer 3 also commented on the comparatively greater testing requirements associated with WER. Reviewer 2 noted that WER adjustment requires comparative toxicity testing. However, running multiple toxicity tests is costly, time consuming, and impractical, especially if multiple species are included. Reviewer 2 wrote that BLM and MLR approaches offer the ability to accommodate multiple scenarios with a minimum number of water quality parameter inputs, and thus help to reduce the uncertainty associated with aluminum bioavailability and toxicity at specific sites. Similarly, Reviewer 3 wrote that WERs are a reasonable approach to account for differences in receiving water chemistry, but that the main disadvantages of WERs are the additional experimentation required and the use of test animals. Further, Reviewer 3 stated that WER relies on a standard toxicity test which inherently has large statistical uncertainty. The reviewer noted that the BLM and MLR approaches are based on many toxicity tests, which together provide improved statistical confidence in the results compared to a WER test. In addition, the reviewer commented that BLM and MLR approaches allow for virtual testing of many different organisms and endpoints, which help make criteria derived using those approaches more statistically robust than those derived from WER tests.

Reviewer 1 commented that in the last several years, it has become apparent to the US EPA that the data requirements for the BLM may be too rigorous for most state water quality criteria monitoring programs, and that there is a need to consider the utility of alternative approaches to calculating metals toxicity under various water chemistry conditions. Reviewer 1 recommended consideration of a tiered approach to focus users' efforts on sites where aquatic systems may be at the most risk from aluminum concentrations. Reviewer 1 provided an example of a tiered approach from the United Kingdom, in which the BLM was used only at specific sites where more detailed study was needed.

Reviewer	Comments	Response to Comments
Reviewer 1	In the last several years it has become apparent to the US EPA that the data requirements for the BLM may be too rigorous for most state water quality criteria (WQC) monitoring programs. There is a need to consider the utility of alternative approaches to calculating metals toxicity under various water chemistry conditions, including a consideration of complex versus simplified BLMs, MLR hardness equations, and other multiple liner	

Reviewer	Comments	Response to Comments
	regression approaches. This would promote an overarching, coherent approach to the development of metals criteria and facilitate their adoption into water quality standards. To increase utilization, a tiered approach should be considered to focus a user's level of effort on sites where aquatic systems may be at the most risk from aluminum concentrations. The inclusion of the aquatic chemistry of aluminum into a tiered process may enable the US EPA to focus user efforts and "Do simple better" (Maddon, 2016). As was stated in 2a, the US EPA has tools at increasing levels of complexity that could be formatted into an approach that would be more acceptable to state regulators and other users. In the United Kingdom (UK) a tiered approach was put forward to assess aquatic systems and the BLM was used in tier 3 at specific sites where more detailed study was needed (Simpson, et al., 2014). Its components were: Tier 1: first tier: This tier compares the dissolved metal concentration against the generic "bioavailable metal" environmental quality standards (EQS). An exceedance requires a tier 2 assessment. Tier 2: The second tier of the assessment uses the simplified bioavailability tools along with additional data (pH, DOC, and Ca) to provide a refined assessment of the potential risk. Sites which pass at this tier do not require any further action, Tier 3: More detailed local investigations, including the use of the full BLM models, and consideration of local background concentrations.	
Reviewer 2	The CCC for aluminum proposed by U.S. EPA was 87 mg/L, and protects only two species, brook trout and striped bass (Stephen et al., 1985). Water Effects Ratio (WER) adjustment requires comparative toxicity testing in standard test waters and in the waters associated with targeted management for Al. The ratio multiplier can be < 1 or > 1 depending on the constituents in the target water and the mixing/dilution of managed discharges. In general, chronic toxicity testing — especially if multiple species are involved — is costly and time consuming. In addition, it is impractical to run multiple toxicity tests that reflect the dynamic change potential of Al toxicity as pH, hardness, temperature and other significant factors change in a water resource. Hence, uncertainty is built into the current	

Reviewer	Comments	Response to Comments
	WER adjusted CCC. BLM/MLR approaches offer the ability to accommodate multiple scenarios with a minimum number of water quality parameter inputs, and thus serve as a site specific reduction in uncertainty associated with aluminum bioavailability and toxicity.	
Reviewer 3	Water Effects Ratios (WER) are certainly a reasonable approach to take differences in receiving water chemistry into account. The main disadvantage of WERs is the additional experimentation required and the use of test animals. To reduce animal testing, and experiments, and cost, WERs could be performed by simulation using this chronic aluminum BLM. A WER test relies on performance of a standard toxicity test and will have the, generally large, statistical uncertainty inherent in such tests. The BLM approach is based on many toxicity tests, all obviously with associated uncertainty as well, but an increase in "n" will allow for improved statistical confidence in the results compared to a WER test. In addition BLM allows for virtual testing of many different organisms and end points which would be impossible to implement in a wide scale way for all different receiving waters. Similarly MLR is based on trends across numerous toxicity tests and ranges of water chemistry, as such the statistics of criteria derived using this approach is more robust than WER tests relying on individual test results.	
Reviewer 4	WER is an approach to apply laboratory results to field environments. Using WER approach, only some important water quality of laboratory and filed waters are matched but many others that can influence metal speciation and bioavailability are not included. BLM takes all into account and therefore could be more realistic and less uncertainty.	
Reviewer 5	There is more uncertainty using the WER adjustment to the 87 µg/L chronic aluminum value than in using the BLM or MLR approach. These methods incorporate additional water chemistry parameters and have considered multiple species.	

2.4 The chronic aluminum BLM has been validated using EU validation procedures. We currently do not have validation procedures in the EPA 1985 Guidelines document ("*Guidelines for Deriving Numerical National Water Quality Criteria for the Protection of Aquatic Organisms and Their Uses*" (Stephan et al. 1985)). The EU requirements are for validation studies at three trophic levels including alga, an invertebrate, and a fish. Validation studies have been conducted for the alga

(*Pseudokirchneriella subcapitata*), the cladoceran (*Ceriodaphnia dubia*) and fathead minnow (*Pimephales promelas*). Preliminary analyses indicate that the observed EC20 of the toxicity studies and the predicted BLM values are within a factor of 2 (100% for *P. subcapitata*, 100% for *P. promelas*, and 96% for *C. dubia).*

Are the results from the validation of the BLM sufficient to support the incorporation of the BLM directly into the aluminum criteria document for regulatory use? Please explain. Do you have any suggestions?

Three reviewers (2, 3, 5) wrote that the results from the EU validation of the BLM are sufficient to support the incorporation of the BLM directly into the aluminum criteria document for regulatory use. Reviewer 2 stated that the EU's three trophic levels approach meets a common sense benchmark in setting water quality standards. Reviewer 2 also noted that the validation data set could be expanded to include more species at each trophic level even without revision of the EPA 1985 Guidelines document, since the data inputs for the BLM are basic water quality parameters, and there appears to be sufficient controlled aluminum toxicity study data. Reviewer 3 commented that the fact that the aluminum BLM has been validated for Europe means that it should be valid for North America, though the reviewer acknowledged that there may be special cases of unique water chemistry or organisms on each continent. However, Reviewer 3 affirmed that the theoretical basis of the BLM will apply independent of geography. Reviewer 3 also wrote that it is essential to select appropriate species and receiving water chemistry if the BLM is to be used as a regulatory tool, and that the current version of the software makes these selections extremely easy. Reviewer 5 noted that the organisms chosen are generally sensitive to contaminants and are commonly used in toxicity testing. The reviewer commented that if analyses have confirmed that the observed toxicity values and the predicted BLM values are within a factor of 2 for these organisms, then the results from the validation of the BLM are sufficient. Reviewer 5 also noted that *C. dubia* is usually the most sensitive of the three species tested, and questioned whether the BLM over-predicted or under-predicted toxicity in 4% of the *C. dubia* data. The reviewer wrote that If the BLM over-predicted toxicity to *C. dubia,* then the generated chronic aluminum value would theoretically still be protective.

In contrast to the other three reviewers, Reviewer 4 wrote that the data are not sufficient, and Reviewer 1 wrote that the question was difficult to answer. Reviewer 4 strongly encouraged using the BLM for setting aluminum criteria, but wrote that the data used for the BLM development and calibration seemed weak. The reviewer affirmed that DOC is a very important factor for BLM, but noted that half of the data were from studies conducted 30 to 40 years ago in which DOC was not measured. Reviewer 4 commented that more data, especially data with higher DOC, are needed to calibrate the BLM before using it for regulatory purposes. Reviewer 1 wrote that validation efforts of the chronic aluminum BLM should be included in a criteria document, but that it is difficult to answer the question on application for regulatory use based on the information provided in question 4. Reviewer 1 recommended the inclusion of field observations for validation, and discussed a study regarding the validation of predicted versus measured water quality parameters for the BLM. The study found that WQC generated using measured values in two ecoregions were much greater that those generated using estimated values, but that WQC generated using measured values in a third ecoregion were similar to those generated using estimated values. Reviewer 1 noted that DOC appeared to be the primary parameter affecting the estimated values in the first two ecoregions, while ionic parameters were the key parameters affecting the correlation with estimated values in the third ecoregion. The reviewer noted the study's recommendation of a sensitivity analysis on water quality parameters to determine drivers for a given ecoregion.

Reviewer	Comments	Response to Comments
Reviewer 1	4a. Validation studies can have varied purposes, such as, laboratory toxicity to field toxicity, predicted model concentrations with observed concentrations in the field, and in this case, predicted BLM concentrations with laboratory effects concentrations. I think validation efforts of the chronic aluminum BLM should be included in a criteria document. However, from this small amount of information provided in question 4 it is difficult for the reviewer to answer the question on application for regulatory use. Is correlation to an EC20 the acceptable standard for a protective criteria? Are we comparing a SSD HC5 with a toxicity test that produces an EC20? 4b. I would recommend inclusion of field observations in a validation discussion. Kovach et al. (2014) gave a presentation on validation of predicted versus measured water quality parameters for the BLM. Estimated values were from the draft EPA2012 document on estimation tools, finalized in 2016 (US EPA, 2016). Kovach et al. (2014) discussed their analysis of the use of estimation methods for three ecoregions in CO. They found that in Ecoregions 21 and 25 WQCs generated using the full suite of measured values are often much greater that those using a full suite of estimated values. For the two ecoregions, DOC appears to be the primary parameter affecting the estimated values. In ecoregion 26, WQCs generated using a full suite of measured values versus a full suite of estimated values were similar. The ionic parameters were the key parameters affecting the correlation with estimated values. They recommended conducting a sensitivity analysis on water quality parameters to determine drivers for a given ecoregion.	
Reviewer 2	The preliminary analyses suggest the EU procedures used in validating the chronic aluminum BLM support the incorporation of the BLM directly into the aluminum criteria document for regulatory use. In my experience and in general, EU water quality directives meet or exceed US standards. The EU three trophic levels approach meets a common sense benchmark in setting water quality standards. Without revision of the EPA 1985 Guidelines document, the validation data set could be explored and expanded to include more species at each trophic level since the data inputs for the BLM are basic water quality parameters and there appears to be sufficient controlled Al toxicity study data. US/EU harmonization of validation procedures in the current aluminum BLM is a solid	

Reviewer	Comments	Response to Comments
	approach and supports the incorporation of the BLM directly into the aluminum criteria document for regulatory use.	
Reviewer 3	Yes, the results from the EU validation of the BLM are sufficient to support the incorporation of the BLM directly into the aluminum criteria document. The principles of chemistry and biology that link geochemical speciation and toxicity obviously know no borders. The fact that the aluminum BLM has been validated for Europe means that it should still be valid for North American application. There are obviously special cases of "unique" water chemistry on both continents so the calibration range of the BLM should be carefully verified for specific application but in general the same principles and framework in Europe should work in the United States. There will be organism differences as well but the current implementation of the chronic aluminum BLM allow for selection of a subset of appropriate species from the entire database. Fundamentally the mechanism of toxicity, be it ionoregulatory or interference with tight junctions, is going to be common across species anyway so the theoretical basis of the BLM applies independent of geography. It is essential to judiciously pick appropriate species and receiving water chemistry to use the BLM as a regulatory tool, but the current version of the software makes this extremely easy.	
Reviewer 4	I strongly encourage using the BLM for setting Al criteria. However, the data used for the BLM development and calibration here seem to be weak. As mentioned in question 1, DOC is very important factor for BLM. However, half of the data were from studies that were conducted 30-40 years ago and DOC was not measured. The concepts of DOC and dissolved metals likely were not concerned at that time. The low DOC concentrations in the recent studies are still my questions as discussed in question 1. Therefore, I think more data, especially with higher DOC to represent the natural environmental conditions are needed to calibrate the BLM before using it for regulation purposes. Although the user manual said the model was calibrated with a wide range of DOC (0.5 to 30 mg/L), I don't see any high DOC concentration in this data base. Toxicity data with higher DOC are necessary for model validation.	
Reviewer 5	The organisms chosen are representative of three trophic levels, generally sensitive to contaminants and are commonly used in toxicity testing. Past versions of the BLM with	

Reviewer	Comments	Response to Comments
	other metals have been considered acceptable if the observed toxicity values and the predicted BLM values are within a factor of 2. If the analyses confirm that is the case for *P. subcapitata*, *P. promelas* and 96% of the data collected for *C. dubia*, in my opinion the results from the validation of the BLM are sufficient to support the incorporation of the BLM directly into the aluminum criteria document for regulatory use. Further, when multiple toxicity tests are performed (especially by different laboratories) EC20 values can be produced that differ by more than a factor of 2. *C. dubia* is usually the most sensitive of the three species tested. My only question would be whether the BLM over predicted or under predicted toxicity in 4 % of the *C. dubia* data. If the BLM over predicted toxicity (generated a lower chronic value) to *C. dubia,* then the chronic aluminum value generated would still theoretically be protective.	

2.5 Ease of Use: One of the comments we hear from states is that the BLM is difficult to use and they also are unclear as to how to put it into standards.

- **In your opinion, which approach is easiest to use?**
- **Do you have any suggestions as to how to make an approach easier for a stakeholder (i.e., State) to use- examples such as improvements to user manual, better upfront input design, etc.?**
- **Do you have suggestions to improve the ease of use of the BLM?**
- **It is possible to develop an MLR equation for the BLM. Do you believe that would be helpful? Please explain.**

- **In your opinion, which approach is easiest to use?**

Three reviewers (3, 4, 5) wrote that the MLR approaches are easiest to use. Reviewer 4 noted that hardness and pH are the only two required parameters in the regression approach, and stated that anyone can plug these parameters into the equation to calculate the criteria. Reviewer 4 commented that running the BLM requires more chemistry input and technical knowledge, but that it is a scientific base model and is more environmentally relevant. Reviewer 3 noted that modelling the scenarios for this charge was easier with the MLR approach, as it just involved looking up the values from a table. However, Reviewer 3 also wrote that the chronic BLM software was extremely easy to use and was well-designed. The reviewer commented that using the BLM was not much more difficult than using the MLR, as it was easy to simply cut and paste from a spreadsheet, hit run, and then read the results off the screen for different scenarios.

Two reviewers (2, 5) reported having difficulty opening or installing the BLM, but stated that the method was fairly easy to use once it was installed. Reviewer 2 discussed difficulties which led to hours of frustration while trying to install the BLM software. However, Reviewer 2 noted

that the software allowed quick data entry and results generation once it was working. The reviewer commented that the result clustering of the full BLM and the simplified BLM encourages use of the simplified BLM due to reduced costs associated with water quality testing for the reduced input data set. Reviewer 2 wrote that the BLM could be considered the easiest to use if augmented with additional user manual and help screen information. The reviewer also stated that MLR approach 1 appeared solid with respect to clustering with the BLM approaches. However, Reviewer 2 commented that the MLR approach 1 Excel spreadsheet seemed raw and appeared to have been developed to demonstrate capability, rather than as a final distribution version.

Reviewer 1 wrote that if appropriate water chemistry data are available and data quality issues are met, the simplified BLM is the easiest to use. Reviewer 1 commented that US EPA may need to develop side by side comparisons of the full BLM, simplified BLM, and MLR approaches with current methods used by the states. The reviewer stated that discussions with the user community concerning data availability, data quality, and the use of different methods will build a knowledge base with the potential users of these approaches. Reviewer 1 also listed out information presented in a study discussing barriers to adaptation of the BLM by the states, including issues of data quality, complexity, and limited resources.

Two reviewers (2, 3) commented on the flexibility and site-specific nature of the approaches. Reviewer 3 wrote that the BLM approach has much greater long-term flexibility than MLR. The reviewer wrote that adding and removing species from the chronic database is extremely easy with the BLM software, and that these tasks are not readily accomplished with the MLR approach. Reviewer 3 noted that the BLM graphical output allows the user to assess endpoint quality in the context of the normalized species sensitivity distribution, but that the MLR does not allow for such an assessment of goodness of fit compared to measured toxicity. The reviewer stated that the advantages of BLM clearly outweigh the few additional keystrokes and clicks. Reviewer 2 wrote that site-specific regulatory tools and guidance can help mitigate compliance costs for the regulated community. However, the reviewer noted that a downside to site-specific approaches is that they require development of site-specific tools such as the BLM and MLR. Reviewer 2 stated that while the regulation text may be more of an approach than a number, this upfront accommodation could lead to less cost of compliance for the CWA-regulated community in the absence of a modeled risk result demonstrating hazard.

- **Do you have any suggestions as to how to make an approach easier for a stakeholder (i.e., States) to use- examples such as improvements to user manual, better upfront input design, etc.?**

All five reviewers provided recommendations for improvements to the user manual. Reviewer 2 stated that the BLM user manual was relatively poorly developed, and that it appeared to have had limited review and input from stakeholders and the user community. The reviewer recommended various modifications to the manual including step-by-step examples and familiarization exercises. Reviewer 3 recommended documenting scenarios such as running the model either with a specific subset of species or with a few different water chemistries. Reviewer 3 noted that the document could walk the user through inputting the scenarios, clearly show what output they will see, and then interpret the output for the user. Reviewer 3 also suggested making the documentation specific to aluminum, as the document focuses more on copper examples. In addition, Reviewer 3 recommended adding appendices that include the specifics of the geochemical modelling, provide supporting information on aluminum toxicity mechanisms, and discuss the link to bioaccumulation. Reviewer 3 also had numerous specific comments

regarding the User's Guide document and the models themselves. Among other topics, the reviewer discussed a message that appears during the BLM run, the use of a pulldown menu when running the model, and various specific text changes to sections of the user manual.

Regarding additional content for the user's manual, Reviewer 5 suggested more detailed instructions in downloading and opening the BLM in the user manual, as the reviewer did not find that area particularly helpful. Reviewer 1 suggested that the user manual include a discussion of the US EPA estimation methods document, a discussion of the environmental chemistry of aluminum, a sensitivity analysis for water chemistry parameters, and approaches for ecoregions that may be of increased risk from aluminum toxicity. Reviewer 4 wrote that it would be helpful to have a section in the manual that describes the steps for running the model. Reviewer 4 discussed only seeing results for three scenarios in the model's output files after running the four scenarios from question 1. The reviewer discussed obtaining the results for scenario 4 from the probability plot instead of the output files, and noted that there needs to be an explanation for CCC and HC5 somewhere in the user manual or in the EPA technical guidelines. Reviewer 4 wrote that the model gives estimated HC5 based on both the Log Normal Distribution Model and the truncated Triangle Distribution, but that they are not very different. The reviewer wrote that users would appreciate knowing which HC5 is the CCC value. Reviewer 4 also noted that the model discussed options for calculating metal speciation and toxicity, but the reviewer didn't see the speciation option in this version.

Reviewer 2 discussed other tools for risk modeling used by students in the reviewer's environmental toxicology courses, and commented on the high level of technical documentation, usability, and software quality of these tools. The reviewer stated that the BLM and MLR tools in the current evaluation are less user-friendly and have poorer technical documentation compared to these other tools. Also, Reviewer 2 commented that the user interface of both tools could be made significantly more user-friendly, and recommended that help and information callouts be implemented in both tools as an example.

Reviewer 1 wrote that in addition to the science itself, issues related to the use of an aluminum WQC by state and local governments need attention too. The reviewer noted that increased dialogue with users via electronic forums can enhance the communication of needs and problems with the adoption of a BLM-centered WQC for aluminum.

- **Do you have suggestions to improve the ease of use of the BLM?**

Two reviewers (2, 4) discussed a need to address problems with the software's performance. Reviewer 2 wrote that the BLM software required more installation finesse than other similar tools. The reviewer noted that the user manual was of little help in addressing installation challenges, and stated that additional step-by-step installation guidance would be helpful. Reviewer 2 also recommended demonstration examples and a common errors-and-fix listing in the manual. As an example, the reviewer discussed range error and floating point error messages during early installation, and stated that these were frustrating due to the lack of resources for fixing the errors. Reviewer 4 wrote that although the current BLM version is more advanced than previous versions, it is still more complicated than the regression approach. The reviewer discussed having technical problems when running the model, such as an error that occurred while inputting data and issues with runtime and the model freezing. Reviewer 4 noted that a BLM version with a better Excel interface would be ideal, and provided two examples of applications with better interfaces.

Reviewer 3 commented that if it is essential to use an equation or look-up table approach, the BLM could be run across all variables to generate multidimensional surfaces that could be fit to equations. The reviewer noted that the simplified BLM would be even easier to use, as it has fewer input parameters. Reviewer 3 did not see the advantage to any "improvement" like this due to the loss of extra features and flexibility of the BLM as it is currently formulated.

Reviewer 1 wrote that the adoption of new methods takes time. The reviewer praised US EPA and scientific researchers' efforts in communicating the science behind the BLM and getting states to include it as an option for WQC for metals. Reviewer 1 also praised US EPA and the scientific researchers' judgement in evaluating simpler approaches for WQC for metals. The reviewer noted that these activities will produce better scientific tools for the user community to consider and adopt in their WQC programs.

- **It is possible to develop an MLR equation for the BLM. Do you believe that would be helpful? Please explain.**

Two reviewers (2, 5) were not certain of the answer to this question. Reviewer 5 wrote that while developing an MLR equation for the BLM should be possible, the reviewer was unsure of the answer to this question. The reviewer did not think that developing an MLR equation for the BLM was needed, as the chronic aluminum values generated by the BLM were very similar to the ones generated by MLR approach 1. Reviewer 2 wrote that the results of the present work support the ability of MLR approach 1 to yield "sufficiently protective" site-specific chronic criteria. The reviewer also noted that the relative "over protective" results of MLR approach 2 using FAV and FACR data should be explored further in a sensitivity analysis of the model parameters. The reviewer recommended further testing and development using more scenarios for MLR approach 1 to demonstrate robustness, and recommended comparing the output from MLR approach 2 to known examples of species toxicity to verify the relative "over protective" performance of MLR approach 2 compared to the other models. Reviewer 2 noted that validation documentation such as that provided for the BLM is not shown for the MLR approach, and stated that assessing whether MLR development would be helpful requires this documentation.

Reviewers 1 and 3 each offered recommendations regarding the form of the MLR equations. Similar to Reviewer 5, Reviewer 3 wrote that an MLR equation for the BLM would be possible. Reviewer 3 noted that surfaces can be fit to an empirical function, and commented that higher dimension surfaces could be fit spanning all the input chemistry variables. Reviewer 3 urged anyone pursuing this option to think beyond linear functions and consider a multiple nonlinear regression approach, as this approach would provide additional flexibility and a better match to BLM values. The reviewer noted that spreadsheets can perform nonlinear calculations just as readily as linear ones. Reviewer 1 wrote that MLR equations that include DOC data appear to offer the best correlation to metal toxicity data. The reviewer referenced a study that discussed the use of MLR to derive WQC for metals. The study determined that the species-specific MLR models performed as well as the BLM and somewhat better than the hardness-based MLRs. The reviewer stated that the MLR approach, with the DOC parameter included, appears to be a good alternative to the BLM WQC for metals.

Reviewer	Comments	Response to Comments
Reviewer 1	• **In your opinion, which approach is easiest to use?** 5a. For the full BLM, simplified BLM and MLR approaches the paradigm shift is a barrier to user adaptation. The US EPA may need to develop side by side comparisons of these approaches with current methods used by the states. Discussions with the user community concerning data availability, data quality, use of estimation methods, when to use more complex techniques, etc., will build a knowledge base with the potential users of these approaches. If appropriate water chemistry data is available and data quality issues are met, the simplified BLM is the easiest to use. Gensemer et al. (2014) presented information on barriers to adaptation of the BLM by the states. They included: – Insufficient water quality data (e.g. DOC) to run BLM – Question on how much can be measured vs. estimated – Lack of understanding of BLM and limited resources – Concern over complexity (10 parameters vs. 1) – Lack of need • **Do you have any suggestions as to how to make an approach easier for a stakeholder (i.e., States) to use- examples such as improvements to user manual, better upfront input design, etc.?** 5b. Research excellence is a function of technical quality and the relevance of research products/outputs produced by the research organization. Most of the questions in this review have focused on high quality science, however, that is just part of what a scientific product or criteria should address. Discussions to increase the utilization of an aluminum WQC by the users (states and local governments) needs attention too. Increased dialogue with the users via electronic forums can enhance the communication of needs and problems with adoption of a BLM centered WQC for aluminum. As I have suggested, the utilization of a criteria document may need to use a tiered approach that moves from	

Reviewer	Comments	Response to Comments
	state-of-the-practice methods to partial BLM approaches or MLR and then, if required to a full BLM for a given site or system. Suggested improvement is the manual: include a discussion of the US EPA estimation methods document (US EPA, 2016), a discussion of environmental chemistry of aluminum, a sensitivity analysis for water chemistry parameters, and approaches for ecoregions that may be of increased risk from aluminum toxicity. • **Do you have suggestions to improve the ease of use of the BLM?** 5c. Around 20 states include the copper Biotic Ligand Model (BLM) as a site-specific standard option and another 15 – 20 are considering the use of the BLM. Adoption of new methods takes time and I think the US EPA and scientific researchers have done well to communicate the science behind the BLM and to get states to act in including it as an option in assessing WQC for metals. As stated in 5a and b, there are barriers to the adoption of the full BLM. The US EPA and the scientific researchers in metal WQC are using good judgement to evaluate simpler approaches for WQC for metals. These activities will produce better scientific tools for the user community to consider and adopt in their WQC programs. • **It is possible to develop an MLR equation for the BLM. Do you believe that would be helpful? Please explain.** 5d. Similar to a simplified BLM, MLR equations can incorporate key water quality parameters to correlate to toxicity data. In this current research area, MLR equations that include DOC data appear to offer the best correlation to metal toxicity data. Brix, et al. (2015) discussed the use of MLR to derive WQC for metals. They included parameters of pH, hardness, and DOC. Using MLR, they presented correlations between predicted and measured LC50s for several metals. The correlation of the data points, that were within a factor of two, ranged from 67 – 97%. Most of the correlations were greater than 80%. Brix et al. (2015) determined that the species specific MLR based models performed as well and the BLM and somewhat better than the hardness based MLRs. The MLR approach, with the DOC parameter included, appears to be a good alternative to the BLM WQC for metals.	

Reviewer	Comments	Response to Comments
Reviewer 2	• **In your opinion, which approach is easiest to use?** The BLM software version provided was "clunky" and had several error reports during initial installation scenario trials. In these initial trials, I ran it from a remote server and several difficulties created hours of frustration. I reinstalled a new copy in a single folder on my desktop and after some initial challenges, the SIMP BLM worked with the supplied water quality parameters. I then found that saving and renaming the file allowed the FULL BLM to work with modified 10 parameter entry. Although this "work-around" functioned, I did not explore further if direct data entry would work. Once I had it working the software allowed quick data entry and results generation. The result clustering of the FULL BLM and the SIMP BLM within 20% CV encourages use of the SIMP BLM due to reduced water quality testing costs in the reduced input data set. The MLR1 approach appeared solid with respect to clustering with BLM approaches, however the Excel spreadsheet seemed raw and recently developed to demonstrate capability, not as a final distribution version. The MLR produced a grid of result data that could be interpolated for specific pH and hardness data. In this regard, after the initial challenges, the BLM worked and could be considered easiest to use if augmented with additional user manual and help screen information. State management of their water resources allows development of a knowledge base on the applicability of simplified tools such as SIMP BLM and MLR1 in risk assessment, for example the applicability of default ion ratios in a particular water resource. Site specific regulatory tools and guidance can help mitigate compliance costs for the regulated community relative to a one size fits all (or "what's the number") approach. Unfortunately, a site specific approach requires the development and use of site specific tools such as the BLM and MRL. While the regulation text may be more an approach than a number, this upfront accommodation yields a site specific result that could yield less cost of compliance for the CWA regulated community in the absence of a modeled risk result demonstrating hazard. EPA often works to draft model language for states to harmonize and simplify regulatory management in the development of site specific criteria approaches.	

Reviewer	Comments	Response to Comments
	• **Do you have any suggestions as to how to make an approach easier for a stakeholder (i.e., State) to use- examples such as improvements to user manual, better upfront input design, etc.?** In the university environmental toxicology courses I teach, students regularly download and use IEUBK (EPA), LEADSPREAD (CA DTSC), and Benchmark Dose Software (BMDS, EPA) to model risk. The high level of technical documentation, usability and software quality of these tools allow new student users to learn and apply state of the art modeling approaches in the assessment of risk. Many of these students track to careers involving management of environmental risk with regular use of risk modeling. The BLM and MLR tools in the current evaluation have relatively overall poor software user friendliness and technical documentation compared to the above mentioned tools. For example, the HELP menu in the BLM product only brings up the user manual. The BLM user manual is relatively poorly developed and appears to have had limited review and input by stakeholders and the user community. The writing style in the User Guide and Reference Manual (VERSION 3.2.2.38) often appears as one written by an advanced user for an advanced user rather than a novice user. I would recommend modifications to include step-by-step examples and familiarization exercises. Such modifications would help stakeholder and states' utility of the powerful BLM approach. The addition of help and information callouts in both the BLM and MLR tools is recommended (see LEADSPEAD 8 as an example). The user interface of both tools could be made significantly more user friendly. • **Do you have suggestions to improve the ease of use of the BLM?** The BLM software required more installation finesse than other similar tools that I use regularly such as IEUBK. I installed the BLM software on two different computers and experienced similar faults. The user manual was of little help in guiding me through the installation challenges and that was frustrating. Additional step-by-step installation user manual guidance would be helpful. As well, demonstration examples and a common errors-and-fix listing in the manual is recommended. In my early installation, range error and floating point error messages, along with occasional errors found in the Input Check were frustrating because of the lack of resources or feedback to fix the errors. Because of my experience, I can fully understand why some of the user community finds the BLM	

Reviewer	Comments	Response to Comments
	software difficult to use. Enhancement of help tools and user manual resources would advance user friendliness. • **It is possible to develop an MLR equation for the BLM. Do you believe that would be helpful? Please explain.** The results of the four scenario exercise in the present work support the ability of MLR1 to yield "sufficiently protective" site specific chronic criteria. It is a preliminary proof-of-concept exercise. The relative "over protective" results of MLR2 using Final Acute Value (FAV) and Final Acute-Chronic Ratio (FACR) data should be explored further in a sensitivity analysis of the model parameters. There is increasing comparative use and exploration of MLR approaches in the scientific literature. The facile use and set-up of advanced spreadsheet approaches has merit in practicality, rapid adaptability to new data as it develops, and user friendliness because of the relatively common knowledge base of spreadsheet use. When given the choice of using IEUBK or LEADSPREAD in risk assessment challenges on exam problems where they have to defend their choice of risk assessment tool, about half of the students use LEADSPEAD, justifying its use by relative usability and simplicity to develop a reference point for quantitative risk assessment rather than the deeper probability distributions available by IEUBK. Considerably more testing and development using more scenarios is recommended for MLR1 to demonstrate robustness, in addition to MLR2 output to known examples of species toxicity to verify the relative "over protective" performance compared to the other models; this validation may demonstrate that MLR2 is "sufficiently protective" and that the other five models, with clustered results, are not protective. This is unlikely since the BLM model has been validated using the three trophic level EU approach. Validation documentation, such as that provided for the BLM, is not shown for the MLR approach in this comparison exercise of aluminum aquatic life criteria approaches. This would be required to assess whether MLR development would be helpful.	
Reviewer 3	The chronic BLM software was extremely easy to use. I have actually never run any implementation of BLM software before. I am very familiar with the theory and I have done my own modelling but never actually run the packaged software. It was extremely easy to use and well-designed although I do have some minor suggestions (see below). As an additional minor point I was able to run the software in the Linux operating system	

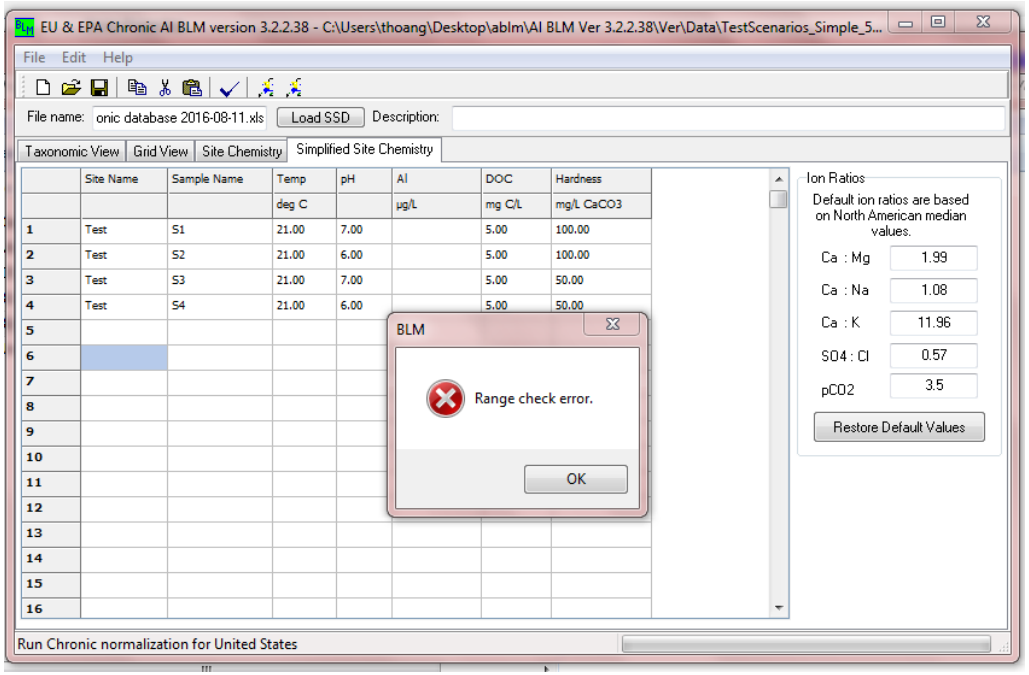
Reviewer	Comments	Response to Comments
	using windows is not an emulator (WINE). I just did this out of curiosity but it worked "out of the box". • **In your opinion, which approach is easiest to use?** From the point of view of filling out the table of required scenarios for this charge the MLR approach was easier. It literally just involved looking up the values from the table, but even if the conditions had not exactly matched the pre-calculated entries it is a trivial matter to substitute numbers into the provided equations; in fact, that is how I did the MLR modelling I included in this document. That being said the BLM was not all that more difficult to use. For different scenarios it is easy to simply cut and paste from a spreadsheet and hit run and then read the results off the screen. The BLM approach has much greater long-term flexibility than MLR though. Imagine scenarios of adding in new species, or removing non-relevant species from the chronic database. This is extremely easy with the BLM software, and can be implemented by the end-user. To do this with MLR approach is not readily available at the "permit writer" level. As new data becomes available it would be simple, even for the end-user, to add into the chronic database. The graphical output at the end of the run allows the end-user to assess the quality of the determined endpoint in the context of the normalized species sensitivity distribution. MLR does not allow for such an assessment of ``good-ness-of-fit" compared to measured toxicity. In my opinion, clearly the advantages of BLM far out weight the few additional keystrokes and clicks. • **Do you have any suggestions as to how to make an approach easier for a stakeholder (i.e., State) to use- examples such as improvements to user manual, better upfront input design, etc.?** I do have some suggestions to improve the documentation. My main suggestion is to make it specific to aluminum – aluminum is right in the title of the document but the document focuses more on Cu examples. This makes sense because of the greater amount of development of Cu BLMs but for an aluminum document Al should be front and centre. As a chemist I would love to see an appendix associated with this document that includes the chemical "nuts and bolts" of the geochemical modelling. Are polynuclear aluminum complexes included for example? Also an appendix showing	

Reviewer	Comments	Response to Comments
	supporting information such as a brief review of aluminum toxicity mechanisms and the specific link to bioaccumulation. This would not be material the typical end-user would care about but having it available if questions arise would be beneficial. From an end-user perspective the manual should have a few worked examples. Just providing scenarios, such as running the model with a specific subset of species, or a few different water chemistries. The document could walk the user through inputting the scenarios and then clearly show what output they will see and interpret the specific output for the end-user. In this way anyone can follow along the examples and by seeing the same result confirm that they know how to use the software. This will provide them with confidence when running the software for themselves in the future on their own scenarios. Some specific comments regarding the User's Guide document … a) A message appears on the screen when the BLM is run indicating "36 entries will be included 138 not included". The specific numbers might change with each run I did not notice. It is not at all clear what this refers to, it is not the chronic database because I see this message when I have not deselected any of the datasets. So some explanation of this should be given in the User's Guide (or remove that message if the end-user does not need to see it). b) When running a lot of different water chemistries it is a bit tedious to use the pulldown to select each test to see the result in the graphical window, or to search through the text file of the output. Perhaps an option for simplified output would be appropriate? If the user just wants to see the HC5 criteria they could get a simple output that just shows them those values – with the other details still available for when they are required. c) On page 9 section 4.1.1 the final line says ``… relevant thermodynamic reactions rates." is misleading. At least as I understand it the model is based on equilibrium and not kinetics; time is never input as a variable. This is a picky point but could confuse some readers.	

Reviewer	Comments	Response to Comments
	d) Section 4.1.4 seems to have some errors in the writing. I get what the authors are trying to convey but they should rewrite the start of that section. e) Again a picky point but when ions are mentioned, as in section 4.1.5 the charge should be given. For total amounts of Ca and Mg they can be referred to without charge. Perhaps the authors are trying to avoid confusion? Anyway, I suggest as a possible consideration for revision. f) Section 4.2.1 should give the relationship between hardness and measured Ca and Mg concentrations. Similar to how 4.1.7 explains the estimation of DIC from alkalinity. g) Section 4.1.8 seems really unnecessary. I can see the idea of maintaining it from a developers' perspective but not for the general user. That section is one of the longest in the document and that field is not used at all. Especially for aluminum, sulfide is not going to be an important ligand. h) Page 19 section 6.3 the described fields do no match what is in the software. A small point but this could confuse the novice user – Biomass is referred to in the example but two Biomass options are available from the pulldown menu but neither is simply called "Biomass". i) Section 6.7.4 refers to prediction of metal toxicity or speciation but this version of the BLM does not give the option for speciation output and the focus of the output is more about HC5 numbers than actual toxicity predictions (although those are available in the output). j) Section 6.12 the authors should make it clear when they will get the option to see this screen. When I first read the manual I did not know how to access this screen – it was only after running that it became obvious. Also, a small point but it would be great if the fitted line in the graphic window extended down to the HC5 value. The intention of this window is visualization and that would really help in visualizing where the specific HC5 value comes from. Again, a small point. • **Do you have suggestions to improve the ease of use of the BLM?**	

Reviewer	Comments	Response to Comments
	If somehow it is essential to use an equation, or look-up table approach, the BLM could be run across all variables to generate multidimensional surfaces and these surfaces could be fit to equations. Obviously the simplified BLM, having fewer input parameters, would be even easier to use. Personally I do not see the advantage to any "improvement" like this because you would lose all the extra features, and in particular the flexibility, of the BLM as currently formulated. • **It is possible to develop an MLR equation for the BLM. Do you believe that would be helpful? Please explain.** Yes an MLR equation for the BLM would be possible. I show surfaces in this document they could be fit to an empirical function. Higher dimension surfaces could be fit as well spanning all the input chemistry variables. It would not be possible to visualize these surfaces but that does not limit the ability to fit them to arbitrary functions. I would urge anyone pursuing this option to think beyond linear functions though and consider a multiple nonlinear regression approach. This would provide additional function flexibility and a better match to rigorously calculated BLM values. If in the end, all that is required is an equation, I do not see what advantage sticking to a linear equation has; spreadsheets are just as capable of doing nonlinear calculations.	
Reviewer 4	I agree with the comments that BLM is more difficult to use than regression approach. With the regression approach, hardness and pH are the only two required parameters and anyone can plug these parameters in the equation to calculate the criteria. Running the BLM requires more chemistry input, technical and knowledge. However, it is a scientific base model and can be used for wide range of water chemistry and more environmental relevant. Although the current BLM version is more advanced than previous versions, such as it incorporated the probability plot tool into the program that give the users options to play around with output graphs and selection, it is still more complicate than the regression approach. In addition, I got some technical problems when running the model, such as an error showed in picture below when I changed the input data to 5 mg/L DOC to predict for different scenarios. To resolve this problem I had to close the model and reopen it. Also a few times the model was freezing towards the end of the run that did not allow me	

Reviewer	Comments	Response to Comments
	to further change the input data to run for other scenarios. Onetime I left the model ran over night and it was Ok in the morning. This indicates running time issue. These problems happened with my both desktop and laptop computers. I am not sure this is technical problems from my end? Given that said, it would be ideal and much easier to use if the modeler can develop a BLM version with more interface with excel like Toxcalc or its recent replaced CETIS of Tidepool Inc. In the output files, I only see results for three scenarios while I ran the model for four scenarios given in the table above. The results for scenarios 4 (pH 6, hardness 50) were obtained from the probability plot but not from the output files. It needed an explanation for CCC and HC5 somewhere in the user manual or in EPA technical guidelines. The model gives estimated HC5 based on Log Normal Distribution Model and estimated HC5 based on truncated Triangle Distribution (US EPA FAV). They are little different. The users would appreciate knowing which HC5 is the CCC value. The values I presented in the table above are estimated HC5 based on Normal Distribution Model. The user manual describes the principles of the BLM, input and out data, explains the functions of each menu item, … and example application. I still had some difficulties for getting the model run at the beginning. It would be more helpful to have a section in the manual that describes the steps for running the model. Although the model said about	

Reviewer	Comments	Response to Comments
	the options for calculating metal speciation and toxicity but I don't see speciation option in this version. 	
Reviewer 5	The MLR approaches are easiest to use. I had problems mainly in opening the BLM. I downloaded the file and saved it to dropbox and could not open the BLM. When I downloaded the file to my desktop; however, I was eventually able to open the BLM. I had to disable my anti-virus software to do so. Once I was able to open the BLM, it was fairly simple and straightforward to use. I would suggest more detailed instructions in downloading and opening the BLM in the user manual. I did not find that area particularly helpful. I would think that developing an MLR equation for the BLM would be possible, but I do not know the answer to that question. I don't really think that developing an MLR	

Reviewer	Comments	Response to Comments
	equation for the BLM is needed, given that the chronic aluminum values generated by the BLM were very similar to the ones generated by MLR Approach 1.	

3.0 NEW INFORMATION PROVIDED BY REVIEWERS

This section presents all new information that reviewers provided in addition to or within their specific responses (presented in Section 2, above) to the charge questions.

Reviewer	Comments	Response to Comments
Reviewer 1	**REFERENCES** Brix, K., D. DeForest, L. Tear, R. Santore, A. Ryan, and W. Adams. 2015. The use of multiple-linear regression to derive water quality criteria for metals. EPA Water Quality Criteria Workshop. Crystal City, VA. Di Toro, D.M.; Allen, H.E.; Bergman, H.L.; Meyer, J.S.; Paquin, P.R.; Santore, .RC. Biotic ligand model of the acute toxicity of metals. 1. Technical basis. *Environ Toxicol Chem,* **2001,** 20,2383–2396. Santore, R.C.; Di Toro, D.M.; Paquin, P.R.; Allen, H.E.; Meyer, J.S. Biotic ligand model of the acute toxicity of metals. 2. Application to acute copper toxicity in freshwater fish and Daphnia. *Environ ToxicolChem,* **2001,** 20,2397–2402. Gensemer. R., and R. Playle. 1999. The Bioavailability and Toxicity of Aluminum in Aquatic Environments. *Critical Reviews in Environmental Science and Technology,* 29(4):315–450 (1999). Gensemer, R., C.A. Claytor, J. Gondek, D.K. DeForest, J.S. Meyer, and J.W. Gorsuch. 2014. Regulatory Implementation of the Coppe BLM: What Have We Learned, and How are We Doing? SETAC NA 35th Annual Meeting. Vancouver, BC, Canada. November 11, 2014.	

Reviewer	Comments	Response to Comments
	Kovach, A., Canton, S., Claytor, C., Gondek, J., Gensemer, R., and J.Gorsuch. 2014. Investigating the Effects of Using Estimated Water Quality Parameters in Generating Copper Water Quality Criteria Using the Biotic Ligand Model. SETAC NA 35th Annual Meeting. Vancouver, BC, Canada. November 12, 2014. Maddon, J. 2016. Quote from Chicago Cubs manager. Chicago, IL. Paquin, .PR.; Gorsuch, J.W.; Apte, S.; Batley, G.E.; Bowles, K.C.; Campbell, P.G.; Delos, C.G.; Di Toro, D.M.; Dwyer, R.L.; Galvez, F.; Gensemer, R.W.; Goss, G.G.; Hogstrand, C.; Janssen, C.R.; McGeer, J.C.; Naddy, R.B.; Playle, R.C.; Santore, R.C.; Schneider, U.; Stubblefield, W.A.; Wood, C.M.; Wu, K.B. The biotic ligand model: A historical overview. *Comp Biochem Physiol.* **2002,**133:3–35. Simpson, P., F. Verdonck, P. van Sprang, A. Peters, and G. Merrington. 2014. A systematic comparison of user friendly tools used to estimate the bioavailability of copper, nickel and zinc in the aquatic environment of Europe. Draft Report. Arche (Assessing Risk of Chemicals) Belgium, and wca, UK. April, 2014. US EPA 2007. Aquatic life ambient freshwater quality criteria – copper, 2007 Revision. February 2007. U.S. Environmental Protection Agency, Office of Water, Office of Science and Technology, Washington DC: EPA-822-R-07-001, 2007. US EPA (2016) Draft Technical Support Document: Recommended Estimates for Missing Water Quality Parameters for Application in EPA's Biotic Ligand Model. EPA 820 R-16-106. OW EPA Washington, DC.	

APPENDIX A

CHARGE TO REVIEWERS

Technical Charge to External Peer Reviewers

Contract No. EP-C-12-029
Task Order 72
October 2016

External Letter Peer Review of a Comparison of Aluminum Aquatic Life Criteria Approaches

BACKGROUND

The U.S. Environmental Protection Agency (EPA) Office of Water is charged with protecting ecological integrity and human health from adverse anthropogenic, water-mediated effects, under the purview of the Clean Water Act (CWA). In support of this mission, EPA is working to update water quality criteria to protect aquatic life and aquatic-dependent wildlife from the presence of aluminum in freshwater and estuarine/marine environments.

The effects of water chemistry on the aquatic toxicity of metals have been an area of research and regulatory advancement since water quality criteria were developed. A partnership among industry, academia, and government utilized research advancements in the aquatic toxicity of metals to effectively develop a national water (BLM) to address the bioavailability of metals in aquatic systems and their acute toxicity to fish. The BLM adjusts the water concentration that causes acute toxicity of metals to aquatic organisms by calculating the relative binding affinity of cations in the water to the biotic ligand (eg., gills or any other active site). The Biotic Ligand Model (BLM) was developed as a method to develop water quality criteria (WQC) for metals.

The purpose of this peer review is to provide a comparison of several approaches to generating aluminum criteria that reflect water quality condition impacts on toxicity. Approaches to be compared include existing complete 10 parameter BLM(s), a simplified BLM approach (e.g., pH, hardness, dissolved organic carbon, temperature), and regression-based equation approaches (e.g. hardness equations) to facilitate evaluation of the most appropriate approaches to consider across aluminum modeling approaches.

You have been provided with:

- A zipfile named "**Al BLM Ver 3.2.2.38.zip**". After unzipping, you will have the following:
  (1) Current version of the aluminum chronic BLM user's manual. It is located in the folder "Ver" and is titled "**BLM_Manual 3 22 38_EAA-Database_2016-02-29.pdf**".
  (2) Aluminum BLM model is located in the folder "Ver" and is titled "**BLM_UI_Al.exe**". There are two versions of the BLM model in this file:
    A. **Full BLM** that uses input for all the parameters. Input page is located in the tab named "Site Chemistry"
    B. **Simplified BLM** uses inputs for temperature, hardness, DOC, and pH and default values for the others. Input page is located in the tab named "Simplified Site Chemistry"

NOTE: This might not be completely clear in the User's Manual. You enter your parameters on the tabs named "Site Chemistry" or "Simplified Site Chemistry" and then you must save and give it a file name. When you run the program, make sure you hit the Running Man that is labelled US (the EU version is also included).

- An Excel spreadsheet titled "**Aluminum Tables_Multiple Parameter Regression Approach 2_4.29.16.xlsx**" that provides multilinear regression equations (MLR) using pH and hardness.
    - The first tab is labeled "Criteria Comparison" and the 2 approaches for calculating are highlighted in yellow
        - **COMPARISON Approach 1** - CCC is calculated directly with Other Data used to Fulfill Missing MDR Group
            - CCC= e^((0.2385*ln(hardness))+(0.6475*pH)-0.3325)
        - **COMPARISON Approach 2** - CCC is calculated by the FAV / FACR
            - CCC = e^((0.2385*ln(hardness))+(0.6475*pH)-1.326)

Please review and familiarize yourself with these materials.

CHARGE QUESTIONS

1. Using the scenarios provided in the table below, do a side-by-side comparison of the results of Full BLM, Simplified BLM, and Excel results from the 2 MLR approaches (highlighted in yellow in Excel file).
    - Please draw some conclusions regarding the differences in the values generated and explain your rationale.
    - Would the simplified BLM be sufficiently protective?
    - Would the pH and hardness MLRs be appropriate?
    - Please provide appropriate suggestions.
    - Feel free to try your own scenarios to see differences and provide with your review.

SCENARIOS

	Scenario 1	Scenario 2	Scenario 3	Scenario 4
INPUTS				
Temperature (°C)	21	21	21	21
pH	7	6	7	6
Hardness (mg/L Ca CO_3)	100	100	50	50
DOC (mg C/L)	1.0	1.0	1.0	1.0
HA (%)	10	10	10	10
Ca (mg/L)	26	26	26	26
Mg (mg/L)	8.0	8.0	8.0	8.0
Na (mg/L)	12	12	12	12
K (mg/L)	1.4	1.4	1.4	1.4
SO_4 (mg/L)	56	56	56	56
Cl (mg/L)	3.8	3.8	3.8	3.8
Alkalinity (mg/L Ca CO_3)	55	55	55	55

	Scenario 1	Scenario 2	Scenario 3	Scenario 4
APPROACHES				
Full Aluminum BLM				
Simplified Aluminum BLM				
MLR COMPARISON Approach 1				
MLR COMPARISON Approach 2				

2. Do you believe the scientific and theoretical foundation of the chronic aluminum BLM is sound and defensible? Does the aluminum BLM improve the Agency's ability to predict toxicity to water column organisms due to aluminum in comparison to the currently applied dissolved aluminum concentration criterion?

3. Currently, States use Water Effects Ratio (WER) adjustment to the 87 ug/L chronic aluminum value. Do you think the application of the BLM or MLR as a site-specific adjustment reduces uncertainty associated with metals bioavailability and toxicity? Please explain.

4. The chronic aluminum BLM has been validated using EU validation procedures. We currently do not have validation procedures in the EPA 1985 Guidelines document ("*Guidelines for Deriving Numerical National Water Quality Criteria for the Protection of Aquatic Organisms and Their Uses*" (Stephan et al. 1985)). The EU requirements are for validation studies at three trophic levels including alga, an invertebrate, and a fish. Validation studies have been conducted for the alga (*Pseudokirchneriella subcapitata*), the cladoceran (*Ceriodaphnia dubia*) and fathead minnow (*Pimephales promelas*). Preliminary analyses indicate that the observed EC20 of the toxicity studies and the predicted BLM values are within a factor of 2 (100% for *P. subcapitata*, 100% for *P. promelas*, and 96% for *C. dubia).*

   Are the results from the validation of the BLM sufficient to support the incorporation of the BLM directly into the aluminum criteria document for regulatory use? Please explain. Do you have any suggestions?

5. **Ease of Use:** One of the comments we hear from states is that the BLM is difficult to use and they also are unclear as to how to put it into standards.
   - In your opinion, which approach is easiest to use?
   - Do you have any suggestions as to how to make an approach easier for a stakeholder (i.e., State) to use- examples such as improvements to user manual, better upfront input design, etc.?
   - Do you have suggestions to improve the ease of use of the BLM?
   - It is possible to develop an MLR equation for the BLM. Do you believe that would be helpful? Please explain.

APPENDIX B

INDIVIDUAL REVIEWER COMMENTS

COMMENTS SUBMITTED BY

Reviewer 1

External Peer Review of a Comparison of Aluminum Aquatic Life Criteria Approaches

The purpose of this peer review is to provide a comparison of several approaches to generating aluminum criteria that reflect water quality condition impacts on toxicity. Approaches to be compared include existing complete 10 parameter BLM(s), a simplified BLM approach (e.g., pH, hardness, dissolved organic carbon, temperature), and regression-based equation approaches (e.g., hardness equations) to facilitate evaluation of the most appropriate approaches to consider across aluminum modeling approaches.

CHARGE QUESTIONS

1. **Using the scenarios provided in the table below, do a side-by-side comparison of the results of Full BLM, Simplified BLM, and Excel results from the 2 MLR approaches (highlighted in yellow in Excel file).**
    - **Please draw some conclusions regarding the differences in the values generated and explain your rationale.**

1a. In the scenarios addressed in Table 1, the full BLM and simplified BLM approaches produced similar HC5 values when pH was varied from 7 to 6. I did not see a confidence interval around the SSD regressions, so when using a HC5 as a protective value, it is difficult to determine the variability around the HC5 to distinguished it from background levels of aluminum in water. The spreadsheet of the regression approaches was somewhat cryptic. I could not determine some of the acronyms or see R squared or goodness of fit values. That said, the MLR approach 1 produced similar values to the full and simplified BLM approaches at pH 6 and 7 at a hardness value of 100 mg/L. At a lower hardness concentrations and acidic pH, the MLR approach 1had higher aluminum values than either of the BLM approaches. While MLR approach 2 had aluminum values that were lower than the other three approaches.

Table 2 contains comparisons of the approaches with the pH varied and Ca and Mg concentrations or hardness values halved. The data indicate that there is good agreement between the BLM approaches at acidic pH values and lower hardness or Ca, Mg, and alkalinity values. The MLR approach 1 had higher aluminum values. In looking at the species sensitivity distribution (SSD) plots, at pH 6.0 and 5.5, the data points indicate that a non-linear relationship may be a more appropriate fit to the data. This would result in higher HC5 values.

Table 2. Comparison of HC5 values for aluminum (ppb) for the BLM and MLR 1 approaches with pH 5.5 – 7.0 and lower hardness or Ca, Mg and alkalinity values.

	pH 7.0	pH 6.0	pH 5.5
Full BLM with Ca 13 ppm; Mg 4.0 ppm and Alkalinity 27 ppm	157	50	32
Simple BLM with hardness 50 ppm	159	51	33
MLR approach 1, hardness 50 ppm	170	89	64

Table 3 is a brief comparison at the effect of increased DOC concentrations of 5.0 ppm for the BLM approaches at acidic pH values and low hardness or Ca, Mg, and alkalinity concentrations. The HC5 values for aluminum were several factors higher at a DOC of 5.0 ppm versus a DOC of 1.0 ppm. Once again, there was good agreement between the full BLM and simplified BLM approaches.

Table 3. Comparison of HC5 Al values (ppb) and BLMs with acidic pH and DOC of 5.0 ppm

	pH 6.0	pH 5.5
Full BLM DOC 5.0 ppm	182	158
Simplified BLM, DOC 5.0 ppm	185	156

- **Would the simplified BLM be sufficiently protective?**

1b. The simplified BLM produced aluminum HC5 values that are very similar to the full BLM for the variations in water quality parameters that were conducted. So, if the full BLM is protective than it appears that the simplified BLM is protective too. For the utilization of the simplified BLM, it still requires a DOC concentration and many states and local government water quality programs typically do not measure this value. Increased variability of the estimation of DOC using the EPA document on estimation methods (EPA, 2016) should be considered in the application of the simplified BLM.

- **Would the pH and hardness MLRs be appropriate?**

1c. It is difficult to determine if the MLRs are appropriate for use as a tool for aluminum water quality criteria. I cannot determine the suitability of the R squared values or determine if a sensitivity analysis of the inclusion of additional parameters, such as DOC, was conducted. The BLM approaches use DOC, which as table 2 shows, is a critical parameter to estimate aluminum WQC. Judging the MLR approaches versus the BLM, the MLR approach 1 produced values that were similar at a hardness of 100 mg/L, but under acidic pH and a hardness of 50 ppm, MLR approach 1 values were higher than the BLM values. MLR approach 2 produced values that were lower than the BLM approaches.

The use of MLRs to support metals water quality criteria is an active research area. US EPA should reach out to the scientific community, possibly through a workshop or public meeting to see if the scientific community has published or plans to publish alternative MLR approaches to estimate metal criteria.

- **Please provide appropriate suggestions.**

1d. The full BLM incorporates the water chemistry parameters that have a major influence on metal bioavailability in aquatic systems. It is the best science put forward to assess effects of metals to biota in aquatic systems (Di Torro et al., 2001; US EPA, 2007). However, the key issue concerning the utility of the BLM is that the users (states and local governments) typically do not measure the parameters to run the full BLM. The US EPA Office of Water (OW) is trying to address this gap through the simplified BLM approach using 4 parameters (DOC still is a problem for regulators to measure), multiple linear regression methods, water effects ratios (WER) and in 2016 the EPA produced a technical document on estimation of water

quality parameters (EPA, 2016). The document provides methods to develop estimates for missing parameters using data analysis approaches for water quality parameter default values.

To increase utilization of the above methods for aluminum WQC, I would recommend that the US EPA OW develop tiered approaches, with input from the states and the scientific community, to build from what the states and local governments are currently doing in a tier 1 to inclusion of more complex approaches (e.g. BLM) at higher tiers. Using ecoregions, the US EPA could identify where aquatic systems are likely to be at increased risk from aluminum toxicity. They could put forward tiered approaches for areas where there is likely increased risk from aluminum toxicity and therefore an increased level of effort is warranted from the states. Conversely, for aquatic systems where aquatic chemistry would indicate a lower risk to aquatic life, a lower level of effort tiered approach could be used.

- **Feel free to try your own scenarios to see differences and provide with your review.**

I included these as tables 2 and 3.

Table 1. SCENARIOS

	Scenario 1	Scenario 2	Scenario 3	Scenario 4
INPUTS				
Temperature (°C)	21	21	21	21
pH	7	6	7	6
Hardness (mg/L Ca CO_3)	100	100	50	50
DOC (mg C/L)	1.0	1.0	1.0	1.0
HA (%)	10	10	10	10
Ca (mg/L)	26	26	26	26
Mg (mg/L)	8.0	8.0	8.0	8.0
Na (mg/L)	12	12	12	12
K (mg/L)	1.4	1.4	1.4	1.4
SO_4 (mg/L)	56	56	56	56
Cl (mg/L)	3.8	3.8	3.8	3.8
Alkalinity (mg/L Ca CO_3)	55	55	55	55

	Scenario 1	Scenario 2	Scenario 3	Scenario 4
APPROACHES				
Full Aluminum BLM	216	81	216 Ca 14;Mg 4 = 157	81 Ca 14;Mg 4= 50
Simplified Aluminum BLM	218	82	159	51
MLR COMPARISON Approach 1	200	105	170	89
MLR COMPARISON Approach 2	74	39	63	33

2. Do you believe the scientific and theoretical foundation of the chronic aluminum BLM is sound and defensible? Does the aluminum BLM improve the Agency's ability to predict toxicity to water column organisms due to aluminum in comparison to the currently applied dissolved aluminum concentration criterion?

2a. Yes, the BLM has a solid scientific and theoretical basis (Di Torro, et al., 2001; Santore et al., 2001; Paquin et al., 2002). A partnership among industry, academia, and government utilized research advancements in the aquatic toxicity of metals to fish to effectively develop a national water quality criterion for copper to protect aquatic organisms. This effort used the BLM to address the bioavailability of metals in aquatic systems and their acute toxicity to fish. The BLM adjusts the water concentration that causes acute toxicity of metals to aquatic organisms by calculating the relative binding affinity of all anions in the water to the biotic ligand of the gill. The partnership completed a BLM-based freshwater quality criterion document for copper in 2007 (US EPA, 2007) (http://www.epa.gov/waterscience/criteria/aqlife.html). The BLM has been peer reviewed in the open literature and by the EPA Science Advisory Board.

2b. The aluminum BLM provides the US EPA with a state of the science tool to predict the toxicity of aluminum to aquatic organisms. The additional research to incorporate a simplified BLM and the US EPA document on estimation methods provides users with additional tools to utilize the BLM approach for WQC. The US EPA OW deserves to be commended for these advancements in the application of these approaches to improve and protect our Nation's water bodies. As I stated in response to earlier questions, the evaluation of various approaches to assess the risks of metals to aquatic systems is an active research area and current studies and their associated publications using MLR, simplified BLM, estimations methods and other approaches may need to be evaluated by US EPA for their application in aluminum criteria. A tiered approach that would have input from the users of WQC (states and local governments) could assist in focusing the more complex tools, e.g. the BLM, at the aquatic systems most at risk from aluminum toxicity.

The aquatic chemistry of aluminum should be discussed prominently in any criteria. Aluminum in acidic aquatic systems is a major factor in causing toxicity to aquatic organisms. Aluminum is more soluble under more acidic and more alkaline conditions and relatively insoluble at pH 6 to 8 (Gensmer and Playle, 1999).

3. **Currently, States use Water Effects Ratio (WER) adjustment to the 87 ug/L chronic aluminum value. Do you think the application of the BLM or MLR as a site-specific adjustment reduces uncertainty associated with metals bioavailability and toxicity? Please explain.**

In the last several years it has become apparent to the US EPA that the data requirements for the BLM may be too rigorous for most state water quality criteria (WQC) monitoring programs. There is a need to consider the utility of alternative approaches to calculating metals toxicity under various water chemistry conditions, including a consideration of complex versus simplified BLMs, MLR hardness equations, and other multiple liner regression approaches. This would promote an overarching, coherent approach to the development of metals criteria and facilitate their adoption into water quality standards.

To increase utilization, a tiered approach should be considered to focus a user's level of effort on sites where aquatic systems may be at the most risk from aluminum concentrations. The inclusion of the aquatic chemistry of aluminum into a tiered process may enable the US EPA to focus user efforts and "Do simple better" (Maddon, 2016). As was stated in 2a, the US EPA has tools at increasing levels of complexity that could be formatted into an approach that would be more acceptable to state regulators and other users.

In the United Kingdom (UK) a tiered approach was put forward to assess aquatic systems and the BLM was used in tier 3 at specific sites where more detailed study was needed (Simpson, et al., 2014). Its components were:

Tier 1: first tier: This tier compares the dissolved metal concentration against the generic "bioavailable metal" environmental quality standards (EQS). An exceedance requires a tier 2 assessment.

Tier 2: The second tier of the assessment uses the simplified bioavailability tools along with additional data (pH, DOC, and Ca) to provide a refined assessment of the potential risk. Sites which pass at this tier do not require any further action,

Tier 3: More detailed local investigations, including the use of the full BLM models, and consideration of local background concentrations.

4. **The chronic aluminum BLM has been validated using EU validation procedures. We currently do not have validation procedures in the EPA 1985 Guidelines document ("*Guidelines for Deriving Numerical National Water Quality Criteria for the Protection of Aquatic Organisms and Their Uses*" (Stephan et al. 1985)). The EU requirements are for validation studies at three trophic levels including alga, an invertebrate, and a fish. Validation studies have been conducted for the alga (*Pseudokirchneriella subapitata*), the cladoceran (*Ceriodaphnia dubia*) and fathead minnow (*Pimephales promelas*). Preliminary analyses indicate that the observed EC20 of the toxicity studies and the predicted BLM values are within a factor of 2 (100% for *P. subcapitata*, 100% for *P. promelas*, and 96% for *C. dubia).***

   Are the results from the validation of the BLM sufficient to support the incorporation of the BLM directly into the aluminum criteria document for regulatory use? Please explain. Do you have any suggestions?

4a. Validation studies can have varied purposes, such as, laboratory toxicity to field toxicity, predicted model concentrations with observed concentrations in the field, and in this case, predicted BLM concentrations with laboratory effects concentrations. I think validation efforts of the chronic aluminum BLM should be included in a criteria document. However, from this small amount of information provided in question 4 it is difficult for the reviewer to answer the question on application for regulatory use. Is correlation to an EC20 the acceptable standard for a protective criteria? Are we comparing a SSD HC5 with a toxicity test that produces an EC20?

4b. I would recommend inclusion of field observations in a validation discussion.

Kovach et al. (2014) gave a presentation on validation of predicted versus measured water quality parameters for the BLM. Estimated values were from the draft EPA2012 document on estimation tools, finalized in 2016 (US EPA, 2016). Kovach et al. (2014) discussed their analysis of the use of estimation methods for three ecoregions in CO. They found that in Ecoregions 21 and 25 WQCs generated using the full suite of measured values are often much greater that those using a full suite of estimated values. For the two ecoregions, DOC appears to be the primary parameter affecting the estimated values. In ecoregion 26, WQCs generated using a full suite of measured values versus a full suite of estimated values were similar. The ionic parameters were the key parameters affecting the correlation with estimated values. They recommended conducting a sensitivity analysis on water quality parameters to determine drivers for a given ecoregion.

5. <u>Ease of Use:</u> One of the comments we hear from states is that the BLM is difficult to use and they also are unclear as to how to put it into standards.

- **In your opinion, which approach is easiest to use?**

5a. For the full BLM, simplified BLM and MLR approaches the paradigm shift is a barrier to user adaptation. The US EPA may need to develop side by side comparisons of these approaches with current methods used by the states. Discussions with the user community concerning data availability, data quality, use of estimation methods, when to use more complex techniques, etc., will build a knowledge base with the potential users of these approaches. If appropriate water chemistry data is available and data quality issues are met, the simplified BLM is the easiest to use.

Gensemer et al. (2014) presented information on barriers to adaptation of the BLM by the states. They included:

- Insufficient water quality data (e.g. DOC) to run BLM
- Question on how much can be measured vs. estimated
- Lack of understanding of BLM and limited resources
- Concern over complexity (10 parameters vs. 1)
- Lack of need

- **Do you have any suggestions as to how to make an approach easier for a stakeholder (i.e., States) to use- examples such as improvements to user manual, better upfront input design, etc.?**

5b. Research excellence is a function of technical quality and the relevance of research products/outputs produced by the research organization. Most of the questions in this review have focused on high quality science, however, that is just part of what a scientific product or criteria should address. Discussions to increase the utilization of an aluminum WQC by the users (states and local governments) needs attention too. Increased dialogue with the users via electronic forums can enhance the communication of needs and problems with adoption of a BLM centered WQC for aluminum. As I have suggested, the utilization of a criteria document may need to use a tiered approach that moves from state-of-the-practice methods to partial BLM approaches or MLR and then, if required to a full BLM for a given site or system.

Suggested improvement is the manual: include a discussion of the US EPA estimation methods document (US EPA, 2016), a discussion of environmental chemistry of aluminum, a sensitivity analysis for water chemistry parameters, and approaches for ecoregions that may be of increased risk from aluminum toxicity.

- **Do you have suggestions to improve the ease of use of the BLM?**

5c. Around 20 states include the copper Biotic Ligand Model (BLM) as a site-specific standard option and another 15 – 20 are considering the use of the BLM. Adoption of new methods takes time and I think the US EPA and scientific researchers have done well to communicate the science behind the BLM and to get states to act in including it as an option in assessing WQC for metals. As stated in 5a and b, there are barriers to the adoption of the full BLM. The US EPA and the scientific researchers in metal WQC are using good judgement to evaluate simpler approaches for WQC for metals. These activities will produce better scientific tools for the user community to consider and adopt in their WQC programs.

- **It is possible to develop an MLR equation for the BLM. Do you believe that would be helpful? Please explain.**

5d. Similar to a simplified BLM, MLR equations can incorporate key water quality parameters to correlate to toxicity data. In this current research area, MLR equations that include DOC data appear to offer the best correlation to metal toxicity data. Brix, et al. (2015) discussed the use of MLR to derive WQC for metals. They included parameters of pH, hardness, and DOC. Using MLR, they presented correlations between predicted and measured LC50s for several metals. The correlation of the data points, that were within a factor of two, ranged from 67 – 97%. Most of the correlations were greater than 80%. Brix et al. (2015) determined that the species specific MLR based models performed as well and the BLM and somewhat better than the hardness based MLRs. The MLR approach, with the DOC parameter included, appears to be a good alternative to the BLM WQC for metals.

REFERENCES

Brix, K., D. DeForest, L. Tear, R. Santore, A. Ryan, and W. Adams. 2015. The use of multiple-linear regression to derive water quality criteria for metals. EPA Water Quality Criteria Workshop. Crystal City, VA.

Di Toro, D.M.; Allen, H.E.; Bergman, H.L.; Meyer, J.S.; Paquin, P.R.; Santore, .RC. Biotic ligand model of the acute toxicity of metals. 1. Technical basis. *Environ Toxicol Chem,* **2001,** 20,2383–2396.

Santore, R.C.; Di Toro, D.M.; Paquin, P.R.; Allen, H.E.; Meyer, J.S. Biotic ligand model of the acute toxicity of metals. 2. Application to acute copper toxicity in freshwater fish and Daphnia. *Environ ToxicolChem,* **2001,** 20,2397–2402.

Gensemer. R., and R. Playle. 1999. The Bioavailability and Toxicity of Aluminum in Aquatic Environments. *Critical Reviews in Environmental Science and Technology,* 29(4):315–450 (1999).

Gensemer, R., C.A. Claytor, J. Gondek, D.K. DeForest, J.S. Meyer,and J.W. Gorsuch. 2014. Regulatory Implementation of the Coppe BLM: What Have We Learned, and How are We Doing? SETAC NA 35th Annual Meeting. Vancouver, BC, Canada. November 11, 2014.

Kovach, A., Canton, S., Claytor, C., Gondek, J., Gensemer, R., and J.Gorsuch. 2014. Investigating the Effects of Using Estimated Water Quality Parameters in Generating Copper Water Quality Criteria Using the. Biotic Ligand Model. SETAC NA 35th Annual Meeting. Vancouver, BC, Canada. November 12, 2014.

Maddon, J. 2016. Quote from Chicago Cubs manager. Chicago, IL.

Paquin, .PR.; Gorsuch, J.W.; Apte, S.; Batley, G.E.; Bowles, K.C.; Campbell, P.G.; Delos, C.G.; Di Toro, D.M.; Dwyer, R.L.; Galvez, F.; Gensemer, R.W.; Goss, G.G.; Hogstrand, C.; Janssen, C.R.; McGeer, J.C.; Naddy, R.B.; Playle, R.C.; Santore, R.C.; Schneider, U.; Stubblefield, W.A.; Wood, C.M.; Wu, K.B. The biotic ligand model: A historical overview. *Comp Biochem Physiol.* **2002,**133:3–35.

Simpson, P., F. Verdonck, P. van Sprang, A. Peters, and G. Merrington. 2014. A systematic comparison of user friendly tools used to estimate the bioavailability of copper, nickel and zinc in the aquatic environment of Europe. Draft Report. Arche (Assessing Risk of Chemicals) Belgium, and wca, UK. April, 2014.

US EPA 2007. Aquatic life ambient freshwater quality criteria – copper, 2007 Revision. February 2007. U.S. Environmental Protection Agency, Office of Water, Office of Science and Technology, Washington DC: EPA-822-R-07-001, 2007.

US EPA (2016) Draft Technical Support Document: Recommended Estimates for Missing Water Quality Parameters for Application in EPA's Biotic Ligand Model. EPA 820 R-16-106. OW EPA Washington, DC.

COMMENTS SUBMITTED BY

Reviewer 2

External Peer Review of a Comparison of Aluminum Aquatic Life Criteria Approaches

The purpose of the peer-review is to compare model approaches that can generate aluminum water quality criteria (WQC) in the presence of highly variable water quality conditions that influence aquatic toxicity. The biotic ligand model (BLM), a tool to assess the bioavailability and potential toxic impacts of metals in freshwater and marine aquatic ecosystems, has been integrated into regulatory requirements of the Clean Water Act (CWA) for several metals. The model approaches in this present review compared existing complete 10 parameter BLM(s), a simplified BLM approach (e.g., pH, hardness, dissolved organic carbon, temperature), and regression-based equation approaches (e.g. hardness equations).

The BLM model software "EU/EPA Chronic Al Version 3.2.2.38" (Windward Environmental, LLC) and the Excel calculating spreadsheet "Aluminum Tables_Multiple Parameter Regression Approach 2_4.29.16. xlsx" were provided for this review. Both the complete 10 parameter BLM approach and the simplified BLM approach generate two results: 1) Estimated HC_5 (5-percentile) based on Log Normal Distribution Model, and 2) Estimated HC_5 based on truncated Triangle Distribution (US EPA Final Acute Value, FAV). The Multiple Linear Regression (MLR) spreadsheets contained two approaches for Criteria Continuous Concentrations (CCCs):

COMPARISON Approach 1 where CCC is calculated directly with other data used to fulfill missing model deviation ratio, MDR group

CCC= e^((0.2385*ln(hardness))+(0.6475*pH)-0.3325)

and, COMPARISON Approach 2 where CCC is calculated by the Final Acute Value, FAV / Final acute-chronic ratio, FACR.

CCC = e^((0.2385*ln(hardness))+(0.6475*pH)-1.326)

In this review, the following abbreviations are used:

FULL BLM (LN) - complete 10 parameter biotic ligand model with Estimated HC5 (5-percentile) based on Log Normal Distribution Model.

FULL BLM (TRI) - complete 10 parameter biotic ligand model with Estimated HC5 based on truncated Triangle Distribution (US EPA Final Acute Value, FAV).

SIMP BLM (LN) - simplified 4 parameter biotic ligand model with Estimated HC5 (5-percentile) based on Log Normal Distribution Model.

SIMP BLM (TRI) - simplified 4 parameter biotic ligand model with Estimated HC5 based on truncated Triangle Distribution (US EPA Final Acute Value, FAV).

MLR1 - regression-based equation where CCC is calculated directly with other data.

MLR2 - regression-based equation CCC is calculated by the Final Acute Value, FAV / Final acute-chronic ratio, FACR.

Review Charge Questions and Responses

1. **Using the scenarios provided in the table below, do a side-by-side comparison of the results of Full BLM, Simplified BLM, and MLR and MLR 2 approaches.**
    - **Please draw some conclusions regarding the differences in the values generated and explain your rationale.**

As instructed, the values for water quality were input into the Chronic Al BLM "Site Chemistry" and "Simplified Site Chemistry" modules for the four identified scenarios. In addition, the Chronic Criteria Comparison result table calculated by the two MLR approaches in the provided spreadsheet were identified for the relevant four scenarios in the multiple regression model results of the two identified MLR1 and MLR2 approaches. The four scenarios varied only in the pH (6 or 7) and Hardness (50 or 100, mg/L $CaCO_3$) generating four unique combinations. pH, Hardness and their derived or constituent parts (e.g. alkalinity, Ca/Mg) are well understood factors in aluminum bioavailability and influence on aquatic toxicity.

The results from the modeling exercise are shown in Figure 1 as a scenario clustered bar graph for the six modeling approaches. Initial visual exploration of the results shows that MLR2 (green bar) consistently yield the lowest HC_5, and in a pattern outside the other results. While models have no guarantee of protection of all species for all conditions and co-exposures, MLR2 shows the lowest values of all models for all scenarios. This can be interpreted as "most protective" or alternately "overly protective," relative to the other considered models. With an average MLR2 scenario result of 52 ug/L, with a range of 33 to 74 ug/L, compared to the current 87 ug/L chronic aluminum value, this result would require an average Water Effects Ratio (WER) multiplier of 0.6, even at higher pH levels where Al toxicity and bioavailability is known to decrease in many toxicity trials. MLR2 appears to be less responsive to scenario hardness and pH inputs. The MLR2 results appear to be "overly protective" across the pH and hardness ranges explored in the modeling exercise, and thus there is justification for elimination from further consideration in this review.

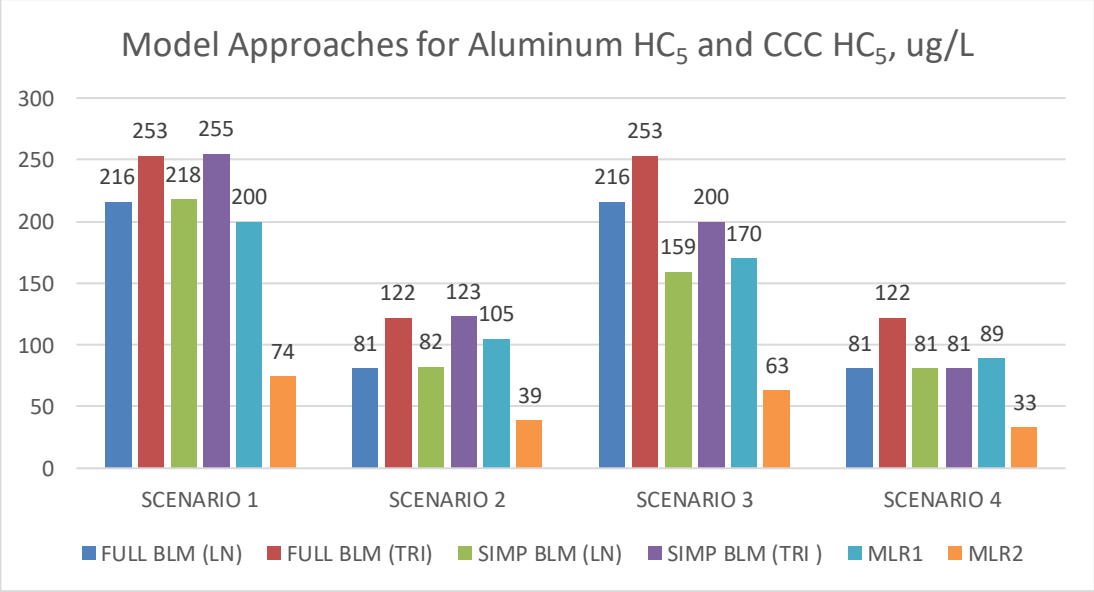


Figure 1. Aluminum HC_5 model approaches.

The model result clusters observed in Figure 1 can be further explored for variability. Table 1 shows the comparative analysis of aluminum aquatic life criteria model results, ug/L, for Scenarios 1-4 detailing the mean, standard deviation, and coefficient of variation of all model results after elimination of MRL2. The coefficient of variation across all model results for Scenarios 1-4 ranges from 10.7% to 20% demonstrating acceptable model consistency. The magnitude and direction from mean, as difference and as a percent across Scenarios 1-4 is also explored in Table 1. For visual reference in the table, negative differences from the mean are shaded red and positive deviations are shaded blue. The FULL BLM (TRI) model showed consistently positive (blue) difference from the mean and the SIMP BLM (LN) model showed consistently negative (red) difference from the mean. In overall percent magnitude and direction from mean, the pattern demonstrated in Table 1 is:

FULL BLM (TRI) > SIMP BLM (TRI) > FULL BLM (TRI) > MRL1 > SIMP BLM (LN)

While this magnitude and direction from mean of the model result cluster cannot identify a "correct" result, it can be stated that within these model results SIMP BLM (LN) is the "most protective" and FULL BLM (TRI) is the "least protective" across Scenarios 1-4. However, after eliminating MRL2, all model results are within the 20% maximum coefficient of variation found across Scenarios 1-4. This limited variability suggests the tested models, not including MRL2, are performing with satisfying consistency with pH changes from 6 to 7 and hardness changes from 50 to 100 mg/L $CaCO_3$.

Table 1. Comparative analysis of Aluminum Aquatic Life Criteria model results, ug/L, for Scenario 1-4.

	Scenario 1	MDM	%MDM	Scenario 2	MDM	%MDM	Scenario 3	MDM	%MDM	Scenario 4	MDM	%MDM
FULL BLM (LN)	216	-12	**-5.4**	81	-22	**-21.1**	216	16	**8.2**	81	-10	**-10.8**
FULL BLM (TRI)	253	25	**10.8**	122	19	**18.9**	253	53	**26.8**	122	31	**34.4**
SIMP BLM (LN)	218	-10	**-4.6**	82	-21	**-20.1**	159	-41	**-20.3**	81	-10	**-10.8**
SIMP BLM (TRI)	255	27	**11.6**	123	20	**19.9**	200	0	**0.2**	81	-10	**-10.8**
MLR1	200	-28	**-12.4**	105	2	**2.3**	170	-30	**-14.8**	89	-2	**-2.0**
	228	**24.4**		**103**	**20.5**		**200**	**37.6**		**90.8**	**17.8**	
	AVG	**SD**		**AVG**	**SD**		**AVG**	**SD**		**AVG**	**SD**	
	CV	**10.7**		**CV**	**20.0**		**CV**	**18.8**		**CV**	**19.6**	
CV = Coefficient of variation												
MDM = Magnitude and direction from mean												
%MDM = Percent magnitude and direction from mean												

- **Would the simplified BLM be sufficiently protective?**

In the analysis of model results shown in Figure 1 and Table 1, the SIMP BLM results cluster well with the other considered models within 20% variability. SIMP BLM (LN) is the "most protective" in magnitude and direction of the five approaches considered in Table 1. SIMP BLM yields a result (81 ug/L) that is lower than the current CCC (87 ug/L) when modeled with pH=6 and hardness at 50 mg/L $CaCO_3$ with both log normal and triangle distribution and thus can be considered "sufficiently protective." The SIMP BLM model only requires input of pH, hardness, dissolved organic carbon, and temperature if the water resource under analysis is known to compare well with the default North American ion ratios. Thus the cost basis for collecting a larger data pool to better understand aluminum hazard in a water resource versus the 10 parameter FULL BLM would allow better, more robust, characterization of risk. The increased data, for

example during high-flow and low-flow periods of the hydrograph of a water resource, could be valuable knowledge in managing risk at a lower cost for the regulated community. Thus SIMP BLM would be "sufficiently protective" under the requirements of the CWA.

- **Would the pH and hardness MLRs be appropriate?**

For the reasons stated above, MLR2 appears to generate relatively "over protective" results when compared to the other models and existing CCC WQC. pH and hardness MRL1 model results appear to cluster well with the other considered models within 20% variability and thus can be considered "sufficiently protective." I did not explore the MRL1 and MRL2 model result difference in detail, however it appears that use of Final Acute Value (FAV) and Final Acute-Chronic Ratio (FACR) in the MRL2 model produces an "over protective" result compared to all of the other model approaches in the present analysis. MRL1 produces a result (89 ug/L) close to the current CCC (87 ug/L) when modeled with pH=6 and hardness at 50 mg/L $CaCO_3$. MLR1 would be appropriate for WQC development, however MLR2 would not. The low data input requirements of MLR1 (pH and hardness), are attractive in managing cost basis for compliance in the regulated community in addition to allowing the development of a larger data pool to better characterize risk of a body of water during annual hydrograph and water quality changes. At present, the provided MLR spreadsheets lack user interface utility and appear as a demonstration of proof-of-concept. If advanced into full user utility, pH and hardness MRL1 spreadsheet approaches could be "sufficiently protective" under the requirements of the CWA in the requested scenario analysis. An example of a spreadsheet approach with excellent user utility is the California Department of Toxic Substances Control LEADSPREAD 8 tool for assessing adverse human health effects from lead that is found at: http://www.dtsc.ca.gov/AssessingRisk/leadspread8.cfm

2. Do you believe the scientific and theoretical foundation of the chronic aluminum BLM is sound and defensible? Does the aluminum BLM improve the Agency's ability to predict toxicity to water column organisms due to aluminum in comparison to the currently applied dissolved aluminum concentration criterion?

The BLM addresses the formation of metal complexes and competitive binding at biotic ligand sites in the assessment of bioavailability, exposure, and hazard potential. The inorganic environmental chemistry of aluminum in water is complex but straight forward with several oxyhydroxides dominating typical water chemistry. The demonstrated capability of BLM to model free ion and complex concentrations is sound and defensible, with a significant history of use in the regulatory arena and general acceptance in the scientific community. The complex water chemistry of aluminum, including complex formation, ligand binding competition and equilibria, the BLM provides a robust site specific assessment tool. With embedded fundamental water chemistry thermodynamics and multiple organism toxicity endpoints, the aluminum BLM is an advancement from the current dissolved aluminum CCC, and thus improves and broadens the Agency's ability to predict toxicity and thus manage CWA responsibilities.

3. Currently, States use Water Effects Ratio (WER) adjustment to the 87 ug/L chronic aluminum value. Do you think the application of the BLM or MLR as a site-specific adjustment reduces uncertainty associated with metals bioavailability and toxicity? Please explain.

The CCC for aluminum proposed by U.S. EPA was 87 mg/L, and protects only two species, brook trout and striped bass (Stephen et al., 1985). Water Effects Ratio (WER) adjustment requires comparative toxicity testing in standard test waters and in the waters associated with targeted management for Al. The ratio multiplier can be < 1 or > 1 depending on the constituents in the target water and the mixing/dilution of managed discharges. In general, chronic toxicity testing — especially if multiple species are involved — is costly and time consuming. In addition, it is impractical to run multiple toxicity tests that reflect the dynamic change potential of Al toxicity as pH, hardness, temperature and other significant factors change in a water resource. Hence, uncertainty is built into the current WER adjusted CCC. BLM/MLR approaches offer the ability to accommodate multiple scenarios with a minimum number of water quality parameter inputs, and thus serve as a site specific reduction in uncertainty associated with aluminum bioavailability and toxicity.

4. The chronic aluminum BLM has been validated using EU validation procedures. We currently do not have validation procedures in the EPA 1985 Guidelines document ("Guidelines for Deriving Numerical National Water Quality Criteria for the Protection of Aquatic Organisms and Their Uses" (Stephan et al. 1985)). The EU requirements are for validation studies at three trophic levels including alga, an invertebrate, and a fish. Validation studies have been conducted for the alga (Pseudokirchneriella subcapitata), the cladoceran (Ceriodaphnia dubia) and fathead minnow (Pimephales promelas). Preliminary analyses indicate that the observed EC20 of the toxicity studies and the predicted BLM values are within a factor of 2 (100% for P. subcapitata, 100% for P. promelas, and 96% for C. dubia).

Are the results from the validation of the BLM sufficient to support the incorporation of the BLM directly into the aluminum criteria document for regulatory use? Please explain. Do you have any suggestions?

The preliminary analyses suggest the EU procedures used in validating the chronic aluminum BLM support the incorporation of the BLM directly into the aluminum criteria document for regulatory use. In my experience and in general, EU water quality directives meet or exceed US standards. The EU three trophic levels approach meets a common sense benchmark in setting water quality standards. Without revision of the EPA 1985 Guidelines document, the validation data set could be explored and expanded to include more species at each trophic level since the data inputs for the BLM are basic water quality parameters and there appears to be sufficient controlled Al toxicity study data. US/EU harmonization of validation procedures in the current aluminum BLM is a solid approach and supports the incorporation of the BLM directly into the aluminum criteria document for regulatory use.

5. <u>Ease of Use:</u> One of the comments we hear from states is that the BLM is difficult to use and they also are unclear as to how to put it into standards.

- **In your opinion, which approach is easiest to use?**

The BLM software version provided was "clunky" and had several error reports during initial installation scenario trials. In these initial trials, I ran it from a remote server and several difficulties created hours of frustration. I reinstalled a new copy in a single folder on my desktop and after some initial challenges, the

SIMP BLM worked with the supplied water quality parameters. I then found that saving and renaming the file allowed the FULL BLM to work with modified 10 parameter entry. Although this "work-around" functioned, I did not explore further if direct data entry would work. Once I had it working the software allowed quick data entry and results generation. The result clustering of the FULL BLM and the SIMP BLM within 20% CV encourages use of the SIMP BLM due to reduced water quality testing costs in the reduced input data set. The MLR1 approach appeared solid with respect to clustering with BLM approaches, however the Excel spreadsheet seemed raw and recently developed to demonstrate capability, not as a final distribution version. The MLR produced a grid of result data that could be interpolated for specific pH and hardness data. In this regard, after the initial challenges, the BLM worked and could be considered easiest to use if augmented with additional user manual and help screen information.

State management of their water resources allows development of a knowledge base on the applicability of simplified tools such as SIMP BLM and MLR1 in risk assessment, for example the applicability of default ion ratios in a particular water resource. Site specific regulatory tools and guidance can help mitigate compliance costs for the regulated community relative to a one size fits all (or "what's the number") approach. Unfortunately, a site specific approach requires the development and use of site specific tools such as the BLM and MRL. While the regulation text may be more an approach than a number, this upfront accommodation yields a site specific result that could yield less cost of compliance for the CWA regulated community in the absence of a modeled risk result demonstrating hazard. EPA often works to draft model language for states to harmonize and simplify regulatory management in the development of site specific criteria approaches.

- **Do you have any suggestions as to how to make an approach easier for a stakeholder (i.e., State) to use- examples such as improvements to user manual, better upfront input design, etc.?**

In the university environmental toxicology courses I teach, students regularly download and use IEUBK (EPA), LEADSPREAD (CA DTSC), and Benchmark Dose Software (BMDS, EPA) to model risk. The high level of technical documentation, usability and software quality of these tools allow new student users to learn and apply state of the art modeling approaches in the assessment of risk. Many of these students track to careers involving management of environmental risk with regular use of risk modeling. The BLM and MLR tools in the current evaluation have relatively overall poor software user friendliness and technical documentation compared to the above mentioned tools. For example, the HELP menu in the BLM product only brings up the user manual. The BLM user manual is relatively poorly developed and appears to have had limited review and input by stakeholders and the user community. The writing style in the User Guide and Reference Manual (VERSION 3.2.2.38) often appears as one written by an advanced user for an advanced user rather than a novice user. I would recommend modifications to include step-by-step examples and familiarization exercises. Such modifications would help stakeholder and states' utility of the powerful BLM approach. The addition of help and information callouts in both the BLM and MLR tools is recommended (see LEADSPEAD 8 as an example). The user interface of both tools could be made significantly more user friendly.

- **Do you have suggestions to improve the ease of use of the BLM?**

The BLM software required more installation finesse than other similar tools that I use regularly such as IEUBK. I installed the BLM software on two different computers and experienced similar faults. The user manual was of little help in guiding me through the installation challenges and that was frustrating. Additional step-by-step installation user manual guidance would be helpful. As well, demonstration examples and a common errors-and-fix listing in the manual is recommended. In my early installation, range error and floating point error messages, along with occasional errors found in the Input Check were frustrating because of the lack of resources or feedback to fix the errors. Because of my experience, I can fully understand why some of the user community finds the BLM software difficult to use. Enhancement of help tools and user manual resources would advance user friendliness.

- **It is possible to develop an MLR equation for the BLM. Do you believe that would be helpful? Please explain.**

The results of the four scenario exercise in the present work support the ability of MLR1 to yield "sufficiently protective" site specific chronic criteria. It is a preliminary proof-of-concept exercise. The relative "over protective" results of MLR2 using Final Acute Value (FAV) and Final Acute-Chronic Ratio (FACR) data should be explored further in a sensitivity analysis of the model parameters. There is increasing comparative use and exploration of MLR approaches in the scientific literature. The facile use and set-up of advanced spreadsheet approaches has merit in practicality, rapid adaptability to new data as it develops, and user friendliness because of the relatively common knowledge base of spreadsheet use. When given the choice of using IEUBK or LEADSPREAD in risk assessment challenges on exam problems where they have to defend their choice of risk assessment tool, about half of the students use LEADSPEAD, justifying its use by relative usability and simplicity to develop a reference point for quantitative risk assessment rather than the deeper probability distributions available by IEUBK. Considerably more testing and development using more scenarios is recommended for MLR1 to demonstrate robustness, in addition to MLR2 output to known examples of species toxicity to verify the relative "over protective" performance compared to the other models; this validation may demonstrate that MLR2 is "sufficiently protective" and that the other five models, with clustered results, are not protective. This is unlikely since the BLM model has been validated using the three trophic level EU approach. Validation documentation, such as that provided for the BLM, is not shown for the MLR approach in this comparison exercise of aluminum aquatic life criteria approaches. This would be required to assess whether MLR development would be helpful.

COMMENTS SUBMITTED BY

Reviewer 3

External Peer Review of a Comparison of Aluminum Aquatic Life Criteria Approaches

1. Using the scenarios provided in the table below, do a side-by-side comparison of the results of Full BLM, Simplified BLM, and Excel results from the 2 MLR approaches (highlighted in yellow in Excel file).

The results of this exercise are included in the table below (Table 1). And graphically in the figure below that (Figure 1). The Multiple Linear Regression (MLR) values were read directly from the lookup table and the Biotic Ligand Model (BLM) results were calculated using the default fitting of the sensitivity distribution (i.e., Estimated HC5 (5-percentile) based on Log Normal Distribution Model) and no chronic toxicity data sets were excluded in the modelling. The US "run" button was used.

Table 1 Scenarios from the original charge question with the modelling results inserted.

	Scenario 1	Scenario 2	Scenario 3	Scenario 4
INPUTS				
Temperature (°C)	21	21	21	21
pH	7	6	7	6
Hardness (mg/L Ca CO_3)	100	100	50	50
DOC (mg C/L)	1.0	1.0	1.0	1.0
HA (%)	10	10	10	10
Ca (mg/L)	26	26	26	26
Mg (mg/L)	8.0	8.0	8.0	8.0
Na (mg/L)	12	12	12	12
K (mg/L)	1.4	1.4	1.4	1.4
SO_4 (mg/L)	56	56	56	56
Cl (mg/L)	3.8	3.8	3.8	3.8
Alkalinity (mg/L Ca CO_3)	55	55	55	55
APPROACHES				
Full Aluminum BLM	216	81	216	81
Simplified Aluminum BLM	218	82	159	51
MLR COMPARISON Approach 1	200	105	170	89
MLR COMPARISON Approach 2	74	39	63	33

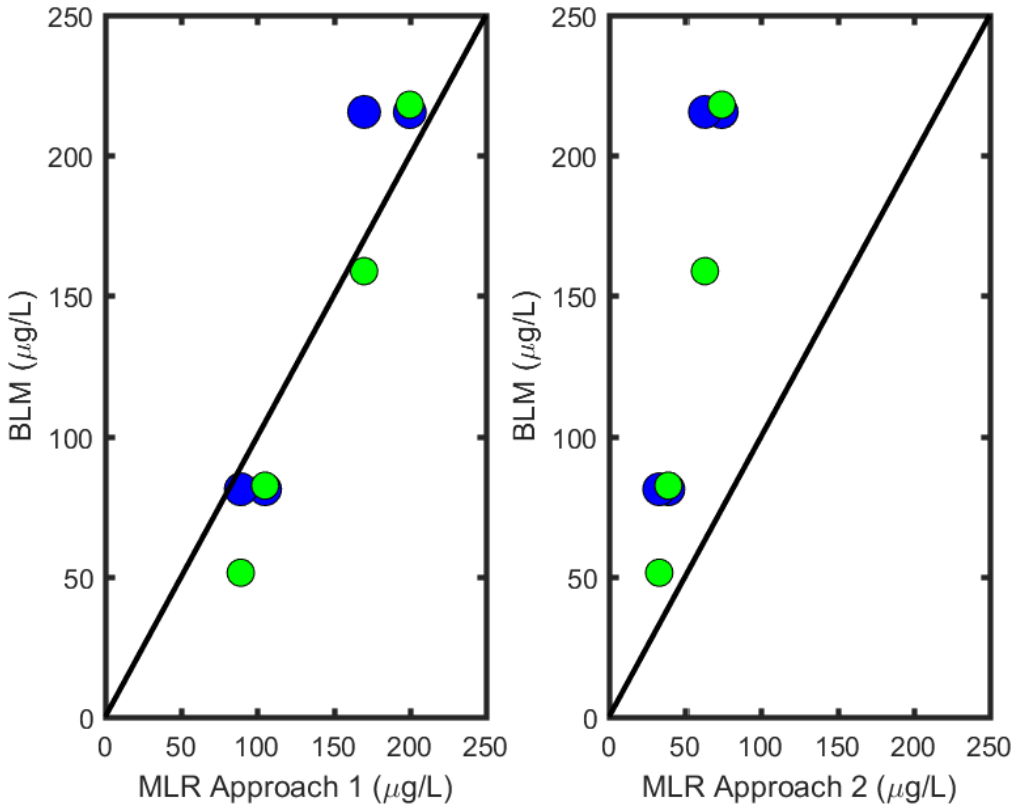


Figure 1: Comparison between MLR and BLM approaches to estimate aluminum chronic criteria. MLR results are on the x-axis; the left plot is using "Approach 1" and the right plot is using "Approach 2". The BLM results are presented on the y-axis with blue symbols for the full BLM and green symbols for the simplified BLM.

- **Please draw some conclusions regarding the differences in the values generated and explain your rationale.**

As expected, all 4 types of criteria determination show the same general trend with criteria values highest when hardness is high and pH is high. Lower hardness and lower pH waters are less protective and thus the estimated criteria is lower. pH has a more significant impact on the results than hardness. The numerical values from MLR using Approach 1 are much more similar to the BLM estimates than for Approach 2. In Figure 1 the left plot (Approach 1) shows data centered around the 1:1 line. By contrast, Approach 2 estimates much lower criteria values than BLM or Approach 2 MLR. Finally, the simplified and full BLM estimates are very similar with the green and blue dots on Figure 1 almost overlapping.

- **Would the simplified BLM be sufficiently protective?**

The provided scenarios are not sufficient to assess if the simplified BLM will be protective or not. The simulations so far show that for these four cases the simplified BLM agrees very closely with the full BLM. To assess protection though it is necessary to compare to actual toxicity measurements, not to other models. To this end the 23 lowest values in the chronic dataset provided with the Al BLM were run as test cases and the predicted criteria values compared to the measured effects concentrations in Figure 2. To run

this simplified BLM the water chemistry values for temperature, pH, DOC and hardness were input. Hardness was estimated according to the equation hardness=2.5*Ca+4.1*Mg where hardness is in mg $CaCO_3$/L and Ca and Mg are both in mg/L. The chronic criteria estimated from the simplified BLM are within a factor of two of the measured values except for the lowest measured effect concentration at 20 µg/L for *C. Dubia* in low DOC, low hardness, slightly acidic pH water. Based on this result the simplified BLM is sufficiently protective.

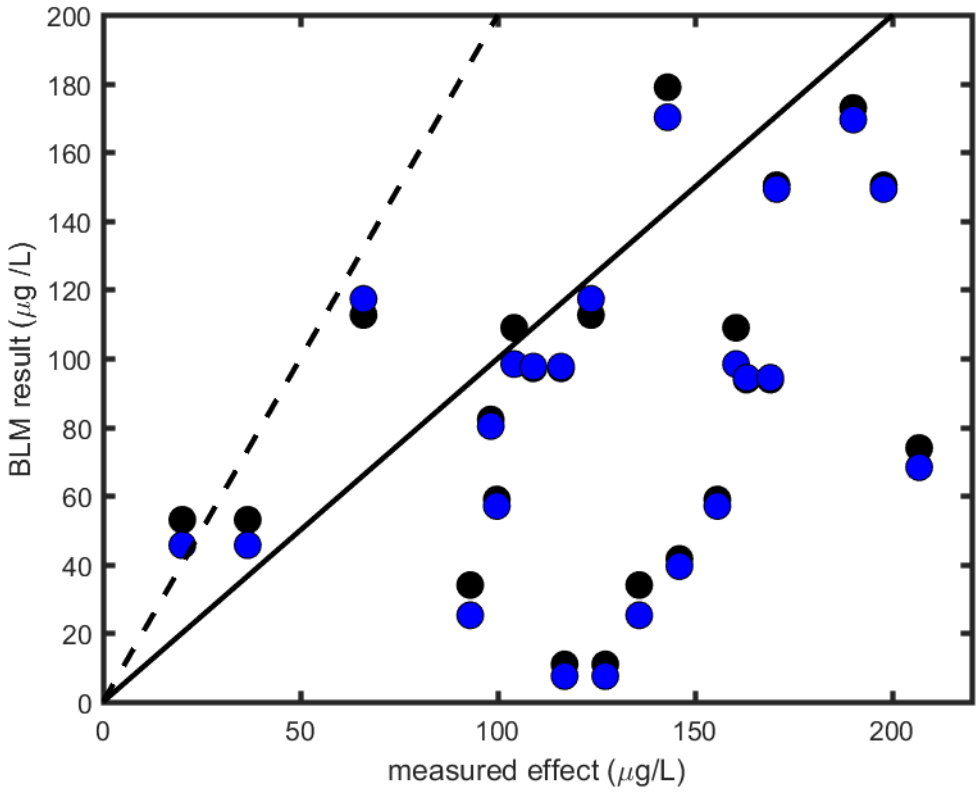


Figure 2: Simplified BLM modelling results of the 23 lowest chronic endpoint values in the chronic BLM database for aluminum provided with the Windward BLM model, versus the measured effect concentration. The solid black line corresponds to the 1:1 line and the dashed line is two times the one to one value. The blue circles are the results of running the full BLM calculation on the given water chemistry.

For comparison, Figure 2 shows full BLM runs on these same water chemistry values. The data for the two approaches almost overlap with full BLM values tending to be slightly lower than the simplified BLM values. Using the full BLM, even the lowest chronic toxicity value is now almost within a factor of two of the measured effect concentration.

- **Would the pH and hardness MLRs be appropriate?**

Again to decide appropriateness of the model predictions it is important to compare to toxicity results and not simply compare to other model results. For the same 23 lowest effect concentration samples in the chronic database the two MLR approaches were tested. The results are presented in Figure 3.

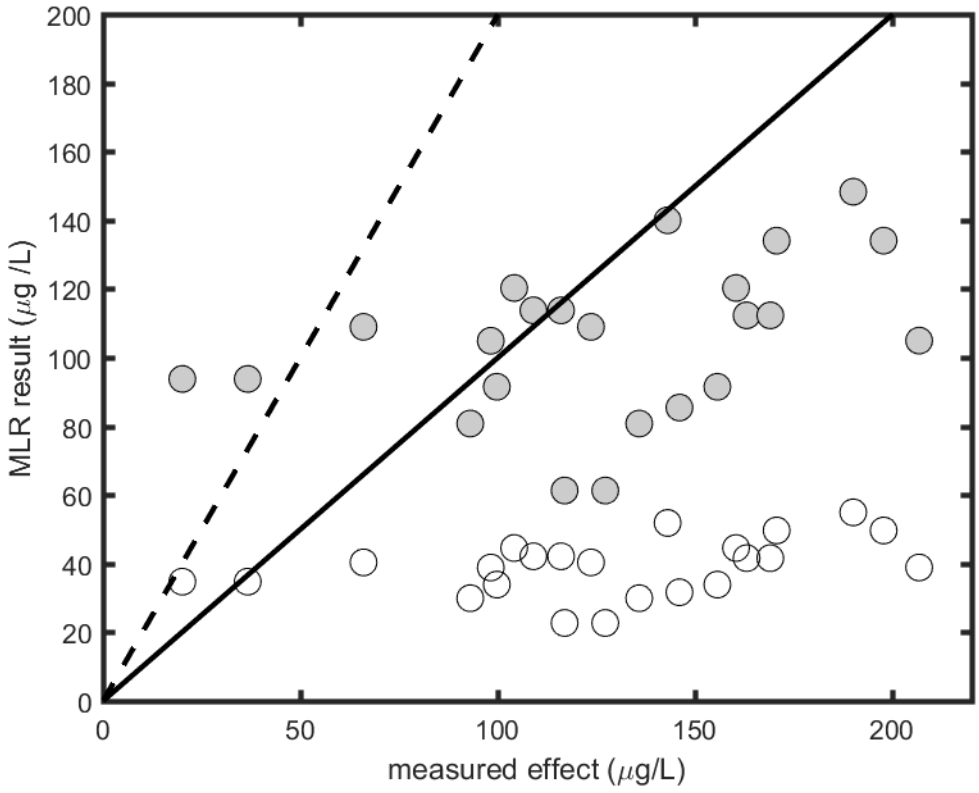


Figure 3: Results of MLR estimates of chronic criteria values compared to measured values. The open symbols correspond to "Approach 2" and the filled symbols correspond to "Approach 1". The solid line is the 1:1 line and the dashed line is a line with a slope of 2 and an intercept of zero.

In MLR estimation of the criteria values, Approach 2 is more conservative than Approach 1 with criteria values well below the measured effects. For Approach 1 the MLR estimated values are not as dramatically different from the measured effects. Both approaches yield values no greater than a factor of two above measured toxicity values, except for the two lowest measured values in Approach 1 which are a factor of approximately 4 greater than the measured toxicity value.

- **Please provide appropriate suggestions.**

Comparison between measured effects concentrations and water chemistry based estimates of criteria values is a reasonable exercise to see if criteria applied to laboratory solutions would be protective or not. In all instances the MLR Approach 2 is protective but many of these values are overly conservative with MLR estimated values as much as 5X lower than observed effects concentrations. This conservative tendency would be even more dramatic for more protective, i.e., harder and higher pH, waters. Approach 1 sometimes generates values that are not protective thus the simplified or full BLM are the most appropriate methods to estimate chronic criteria. The full BLM is slightly more protective than the simplified BLM but the simplified BLM performs almost identically with much reduced data requirements.

I understand that EPA sets guidelines and the States use these as a starting point for developing their specific approaches. To increase "buy in" by the States it might make sense to present test cases and test

organisms specific for different States, or ecoregions that span several States. Essentially give the States some guidance in how the BLM would be run for specific example scenarios. I expand on the suggestion of including different scenarios in the BLM documentation in my response to Charge Question 5 below.

- **Feel free to try your own scenarios to see differences and provide with your review.**

Hardness and pH only estimation of criteria are potentially limiting if DOC is significant in the receiving waters. To test for the influence of DOC, simulations were performed using the simplified BLM at a fixed hardness of 50 or 100 mg $CaCO_3$/L over a pH range (5 to 8.5) for DOC from 1 to 6 mg C/L. These conditions are within the reported range of the BLM calibration (Table 7-1 in the Users' Guide). For comparison the same modelling was performed with MLR Approach 2, but obviously without DOC as an input parameter. What is apparent from these models is that pH has by far the most significant impact; DOC does increase the criteria value but more gradually compared to pH (Figure 4). At pH values below 6 the shapes of the MLR and BLM surfaces are very different, as DOC has a more significant impact, but is not included in MLR modelling. In comparing the MLR and BLM approaches though it is apparent that MLR tends to have much lower concentrations (the z-axis on Figure 4 are all at the same scale, and the scale is logarithmic). To be protective, the most effort and attention is focused on the low range of species sensitivity and for sensitive receiving waters, but these results demonstrate that MLR approaches as currently formulated are dramatically conservative for less sensitive receiving waters. This could potentially waste resources trying to address aluminum exceedances even when there are no potential impacts. BLM estimated concentrations are much higher than MLR estimates for high pH waters, by as much as an order of magnitude.

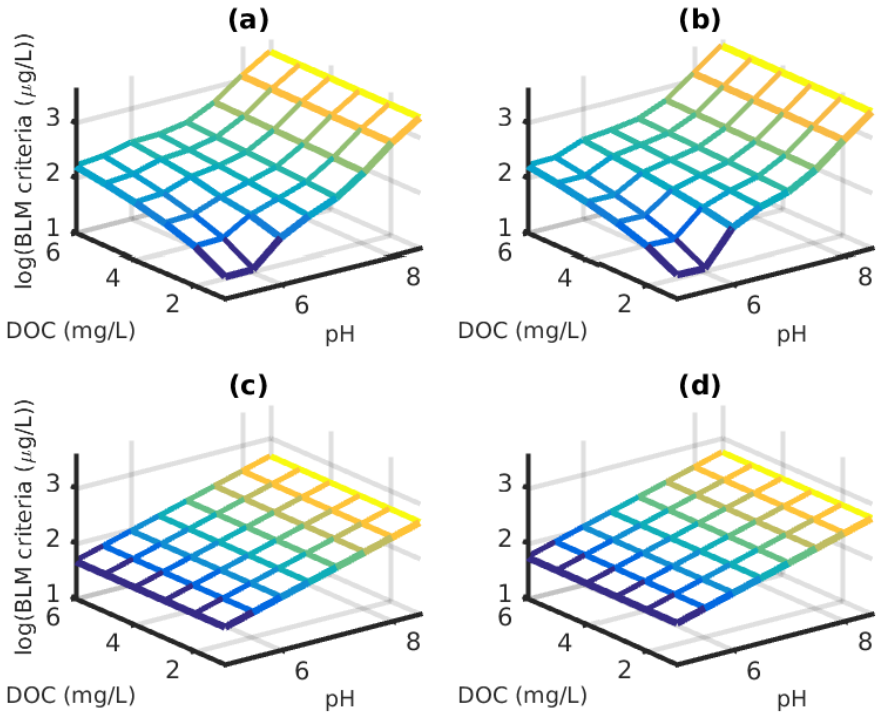


Figure 4: Criteria values from the simplified BLM (a and b) and from Approach1 MLR (c and d) for hardness of 50 (a and c) and 100 (b and d) mg $CaCO_3$/L. pH was varied in both models and concentration of DOC was also varied in the BLM simulations. The z-axis is shown on a logarithmic scale in order to differentiate values.

2. Do you believe the scientific and theoretical foundation of the chronic aluminum BLM is sound and defensible? Does the aluminum BLM improve the Agency's ability to predict toxicity to water column organisms due to aluminum in comparison to the currently applied dissolved aluminum concentration criterion?

Yes the scientific and theoretical foundation of the chronic aluminum BLM is sound and defensible. The BLM for acute and chronic toxicity predictions and risk assessment for many metals is a mature area of science. I think it is fair to say that most scientists are so convinced of the foundation of the BLM that they are surprised the adoption is not more wide-spread in regulatory application. BLM approaches have a long history of success and represent a distinct advantage over application of the precautionary principle. They have demonstrated an ability to protect the environment while also respecting the significant economic and social benefits of having a strong metals industry.

Biotic ligand models are based on very old and fundamental physical chemistry; that is to say the application of solving for the equilibrium position of multiple simultaneous reactions. This practice was well established by Garrels and co-workers in the 1960s and is often referred to as geochemical modelling. The BLM does geochemical modelling with the application of additional simultaneous reactions for metal complexation and cation competition at the biotic ligand, or site of toxic action. The observation that toxicity is proportional to short term bioaccumulation is well established and numerous papers have linked accumulation to toxic effects.

The link between bioaccumulation at fish gills and the mode of toxicity for silver and copper is well established (see citations in the aluminum BLM User's Guide). They are acute ionoregulatory toxicants. Mechanistic information exists for other metals as well. The knowledge of the underlying mechanism of toxicity is the final piece of information that really establishes BLMs as an excellent approach to criteria determination based on receiving water chemistry. Since there is a known mechanism, and the model is mechanistic, it should apply for all conditions (within constraints of the data ranges used to derive the parameters). I am less familiar with the aluminum toxicology literature to know if the mechanism of toxicity is as well established. To investigate the state of the art though, I did consult a recent review chapter (Wilson, R. W. "Aluminum" in Homeostatis of Toxicology of Non-essential Metals, Wood, C.M., Farrell, A.P. and Brauner, C. J. (Eds.), Elsevier, pp. 68-104) and I see that ionoregulation, interference with Ca^{2+} tight junctions and respiratory impairment are all known mechanisms of aluminum toxicity to aquatic organisms. Thus, it seems the toxic mechanisms for aluminum are fairly well established and this is the final consideration in establishing the scientifically sound foundation for a chronic aluminum BLM.

3. Currently, States use Water Effects Ratio (WER) adjustment to the 87 ug/L chronic aluminum value. Do you think the application of the BLM or MLR as a site-specific adjustment reduces uncertainty associated with metals bioavailability and toxicity? Please explain.

Water Effects Ratios (WER) are certainly a reasonable approach to take differences in receiving water chemistry into account. The main disadvantage of WERs is the additional experimentation required and the use of test animals. To reduce animal testing, and experiments, and cost, WERs could be performed by simulation using this chronic aluminum BLM. A WER test relies on performance of a standard toxicity test and will have the, generally large, statistical uncertainty inherent in such tests. The BLM approach is based on many toxicity tests, all obviously with associated uncertainty as well, but an increase in "n" will allow for

improved statistical confidence in the results compared to a WER test. In addition BLM allows for virtual testing of many different organisms and end points which would be impossible to implement in a wide scale way for all different receiving waters. Similarly MLR is based on trends across numerous toxicity tests and ranges of water chemistry, as such the statistics of criteria derived using this approach is more robust than WER tests relying on individual test results.

4. **The chronic aluminum BLM has been validated using EU validation procedures. We currently do not have validation procedures in the EPA 1985 Guidelines document ("Guidelines for Deriving Numerical National Water Quality Criteria for the Protection of Aquatic Organisms and Their Uses" (Stephan et al. 1985)). The EU requirements are for validation studies at three trophic levels including alga, an invertebrate, and a fish. Validation studies have been conducted for the alga (Pseudokirchneriella subcapitata), the cladoceran (Ceriodaphnia dubia) and fathead minnow (Pimephales promelas). Preliminary analyses indicate that the observed EC20 of the toxicity studies and the predicted BLM values are within a factor of 2 (100% for P. subcapitata, 100% for P. promelas, and 96% for C. dubia).**

   Are the results from the validation of the BLM sufficient to support the incorporation of the BLM directly into the aluminum criteria document for regulatory use? Please explain. Do you have any suggestions?

Yes, the results from the EU validation of the BLM are sufficient to support the incorporation of the BLM directly into the aluminum criteria document. The principles of chemistry and biology that link geochemical speciation and toxicity obviously know no borders. The fact that the aluminum BLM has been validated for Europe means that it should still be valid for North American application. There are obviously special cases of "unique" water chemistry on both continents so the calibration range of the BLM should be carefully verified for specific application but in general the same principles and framework in Europe should work in the United States. There will be organism differences as well but the current implementation of the chronic aluminum BLM allow for selection of a subset of appropriate species from the entire database. Fundamentally the mechanism of toxicity, be it ionoregulatory or interference with tight junctions, is going to be common across species anyway so the theoretical basis of the BLM applies independent of geography. It is essential to judiciously pick appropriate species and receiving water chemistry to use the BLM as a regulatory tool, but the current version of the software makes this extremely easy.

5. <u>Ease of Use:</u> **One of the comments we hear from states is that the BLM is difficult to use and they also are unclear as to how to put it into standards.**

The chronic BLM software was extremely easy to use. I have actually never run any implementation of BLM software before. I am very familiar with the theory and I have done my own modelling but never actually run the packaged software. It was extremely easy to use and well-designed although I do have some minor suggestions (see below). As an additional minor point I was able to run the software in the Linux operating system using windows is not an emulator (WINE). I just did this out of curiosity but it worked "out of the box".

- **In your opinion, which approach is easiest to use?**

From the point of view of filling out the table of required scenarios for this charge the MLR approach was easier. It literally just involved looking up the values from the table, but even if the conditions had not exactly matched the pre-calculated entries it is a trivial matter to substitute numbers into the provided

equations; in fact, that is how I did the MLR modelling I included in this document. That being said the BLM was not all that more difficult to use. For different scenarios it is easy to simply cut and paste from a spreadsheet and hit run and then read the results off the screen.

The BLM approach has much greater long-term flexibility than MLR though. Imagine scenarios of adding in new species, or removing non-relevant species from the chronic database. This is extremely easy with the BLM software, and can be implemented by the end-user. To do this with MLR approach is not readily available at the "permit writer" level. As new data becomes available it would be simple, even for the end-user, to add into the chronic database. The graphical output at the end of the run allows the end-user to assess the quality of the determined endpoint in the context of the normalized species sensitivity distribution. MLR does not allow for such an assessment of ``good-ness-of-fit" compared to measured toxicity. In my opinion, clearly the advantages of BLM far out weight the few additional keystrokes and clicks.

- **Do you have any suggestions as to how to make an approach easier for a stakeholder (i.e., State) to use- examples such as improvements to user manual, better upfront input design, etc.?**

I do have some suggestions to improve the documentation. My main suggestion is to make it specific to aluminum – aluminum is right in the title of the document but the document focuses more on Cu examples. This makes sense because of the greater amount of development of Cu BLMs but for an aluminum document Al should be front and centre. As a chemist I would love to see an appendix associated with this document that includes the chemical "nuts and bolts" of the geochemical modelling. Are polynuclear aluminum complexes included for example? Also an appendix showing supporting information such as a brief review of aluminum toxicity mechanisms and the specific link to bioaccumulation. This would not be material the typical end-user would care about but having it available if questions arise would be beneficial.

From an end-user perspective the manual should have a few worked examples. Just providing scenarios, such as running the model with a specific subset of species, or a few different water chemistries. The document could walk the user through inputting the scenarios and then clearly show what output they will see and interpret the specific output for the end-user. In this way anyone can follow along the examples and by seeing the same result confirm that they know how to use the software. This will provide them with confidence when running the software for themselves in the future on their own scenarios.

Some specific comments regarding the User's Guide document …

a) A message appears on the screen when the BLM is run indicating "36 entries will be included 138 not included". The specific numbers might change with each run I did not notice. It is not at all clear what this refers to, it is not the chronic database because I see this message when I have not deselected any of the datasets. So some explanation of this should be given in the User's Guide (or remove that message if the end-user does not need to see it).

b) When running a lot of different water chemistries it is a bit tedious to use the pulldown to select each test to see the result in the graphical window, or to search through the text file of the output. Perhaps an option for simplified output would be appropriate? If the user just wants to see the HC5

criteria they could get a simple output that just shows them those values – with the other details still available for when they are required.

c) On page 9 section 4.1.1 the final line says ``… relevant thermodynamic reactions rates." is misleading. At least as I understand it the model is based on equilibrium and not kinetics; time is never input as a variable. This is a picky point but could confuse some readers.

d) Section 4.1.4 seems to have some errors in the writing. I get what the authors are trying to convey but they should rewrite the start of that section.

e) Again a picky point but when ions are mentioned, as in section 4.1.5 the charge should be given. For total amounts of Ca and Mg they can be referred to without charge. Perhaps the authors are trying to avoid confusion? Anyway, I suggest as a possible consideration for revision.

f) Section 4.2.1 should give the relationship between hardness and measured Ca and Mg concentrations. Similar to how 4.1.7 explains the estimation of DIC from alkalinity.

g) Section 4.1.8 seems really unnecessary. I can see the idea of maintaining it from a developers' perspective but not for the general user. That section is one of the longest in the document and that field is not used at all. Especially for aluminum, sulfide is not going to be an important ligand.

h) Page 19 section 6.3 the described fields do no match what is in the software. A small point but this could confuse the novice user – Biomass is referred to in the example but two Biomass options are available from the pulldown menu but neither is simply called "Biomass".

i) Section 6.7.4 refers to prediction of metal toxicity or speciation but this version of the BLM does not give the option for speciation output and the focus of the output is more about HC5 numbers than actual toxicity predictions (although those are available in the output).

j) Section 6.12 the authors should make it clear when they will get the option to see this screen. When I first read the manual I did not know how to access this screen – it was only after running that it became obvious. Also, a small point but it would be great if the fitted line in the graphic window extended down to the HC5 value. The intention of this window is visualization and that would really help in visualizing where the specific HC5 value comes from. Again, a small point.

- **Do you have suggestions to improve the ease of use of the BLM?**

If somehow it is essential to use an equation, or look-up table approach, the BLM could be run across all variables to generate multidimensional surfaces and these surfaces could be fit to equations. Obviously the simplified BLM, having fewer input parameters, would be even easier to use. Personally I do not see the advantage to any "improvement" like this because you would lose all the extra features, and in particular the flexibility, of the BLM as currently formulated.

- **It is possible to develop an MLR equation for the BLM. Do you believe that would be helpful? Please explain.**

Yes an MLR equation for the BLM would be possible. I show surfaces in this document they could be fit to an empirical function. Higher dimension surfaces could be fit as well spanning all the input chemistry

variables. It would not be possible to visualize these surfaces but that does not limit the ability to fit them to arbitrary functions. I would urge anyone pursuing this option to think beyond linear functions though and consider a multiple nonlinear regression approach. This would provide additional function flexibility and a better match to rigorously calculated BLM values. If in the end, all that is required is an equation, I do not see what advantage sticking to a linear equation has; spreadsheets are just as capable of doing nonlinear calculations.

COMMENTS SUBMITTED BY

Reviewer 4

External Peer Review of a Comparison of Aluminum Aquatic Life Criteria Approaches

1. **Using the scenarios provided in the table below, do a side-by-side comparison of the results of Full BLM, Simplified BLM, and Excel results from the 2 MLR approaches (highlighted in yellow in Excel file).**

    - **Please draw some conclusions regarding the differences in the values generated and explain your rationale.**
    - **Would the simplified BLM be sufficiently protective?**
    - **Would the pH and hardness MLRs be appropriate?**
    - **Please provide appropriate suggestions.**
    - **Feel free to try your own scenarios to see differences and provide with your review.**

SCENARIOS

	Scenario 1	Scenario 2	Scenario 3	Scenario 4
INPUTS				
Temperature (°C)	21	21	21	21
pH	7	6	7	6
Hardness (mg/L Ca CO_3)	100	100	50	50
DOC (mg C/L)	1.0	1.0	1.0	1.0
HA (%)	10	10	10	10
Ca (mg/L)	26	26	26	26
Mg (mg/L)	8.0	8.0	8.0	8.0
Na (mg/L)	12	12	12	12
K (mg/L)	1.4	1.4	1.4	1.4
SO_4 (mg/L)	56	56	56	56
Cl (mg/L)	3.8	3.8	3.8	3.8
Alkalinity (mg/L Ca CO_3)	55	55	55	55
APPROACHES				
Full Aluminum BLM	216	81	216 (157)	81 (49)
Full Aluminum BLM at 10oC	101	21	61	21
Full Aluminum BLM at 5ppmDOC	285	208	285	208

	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Simplified Aluminum BLM	218	82	159	51
MLR COMPARISON Approach 1	200	105	170	89
MLR COMPARISON Approach 2	74	39	63	33

Note: The hardness and Ca and Mg concentrations of the four scenarios are not corresponding. Concentrations of Ca and Mg in scenarios 3 and 4 for the case of lower hardness should be lower than those for scenarios 1 and 2. The highlighted data in parentheses (yellow) were the model predicted values when concentrations of Ca and Mg were reduced to a half concentrations of scenarios 1 and 2 (i.e., Ca = 13 mg/L, Mg = 4 mg/L). Other chemistry was the same as scenarios 1 and 2.

In general, the predictions by the full and simplified BLMs are pretty much similar. However, the water chemistry needed for the simplified model is much less than that for the full model. Therefore, if the BLM is used for setting water quality guidelines, the simplified model is more economic and easier to use. The model also takes into account the important bioavailability modifying factors for metals, such hardness, DOC, and pH.

The two hardness and pH approaches gave different predictions. The predictions of approach 2 was about 35% predictions of approach 1. The difference between the two approaches is the equation factor of -0.3325 (approach 1) vs -1.326 (approach 2). I am not sure how these equation factors were derived but only approach 2 was noticed in the Chronic Ranked tap of the excel file (cell B17, C52). The Chronic Ranked tap also indicates that the equation factor of -0.3325 in cell G46 was derived from the most sensitive species (Lampsilis, chironomus, salvelinus, salmo) while the factor of -1.326 in cell D20 was below the whole data set for all species. Was this factor derived from the data for all species? The CCC derived from most sensitive species should be lower than the CCC derived from all species including less sensitive species. The Criteria Comparison tap shows opposite outcomes. It seems to be conflict? Ultimately, the approach with lower CCC is more protective. However, it needed clarification here before answering the question which approach is more protective?

In comparison between the hardness/pH approaches and the BLM, the predictions by approach 1 and the BLMs are very much similar. Approach 1 only take effects of hardness and pH into account while the BLM includes effects of other factors such as temperature and DOC-an important modifying factor of metal bioavailability and toxicity. It is important to mention here is that 54% of the data used to derive hardness and pH equation are missing DOC. Other data have very low DOC concentrations, basically at DOC of ID water. The hardness and pH approaches should be used to predict toxicity for the range of water quality that was used to develop the approach. I am not sure how well the hardness and pH approaches will predict for water with higher DOC which is likely the case for the natural environment. Also temperature was not presented in the data base (might be in the published papers) so don't know what temperature range would be applied. I tried the full BLM with low temperature, such as 10oC which is more realistic for

cold water fish like trout and higher DOC. The BLM predictions are much different from approach 1 (red data in table). It is difficult to compare when we don't know the actual DOC concentrations in the test waters. In addition, I don't understand about the low DOC concentrations in this data base (C_FW Core Chronic tap of the excel file), especially for chronic tests at which the tested organisms were fed. Organic materials would be released from the food. Our chronic tests with *D. magna* and snails in many years have shown a DOC concentration of about 4 mg/L in test water collected from the test chambers. A DOC of 0.5 mg/L for *C. dubia* in this data base seems to be the DOC of test water with no food. This is true for acute tests with no feeding but not for chronic tests. I was trying to run the simplified model with higher DOC and different temperature but got technical problems (see presentation in question 5 below).

2. **Do you believe the scientific and theoretical foundation of the chronic aluminum BLM is sound and defensible? Does the aluminum BLM improve the Agency's ability to predict toxicity to water column organisms due to aluminum in comparison to the currently applied dissolved aluminum concentration criterion?**

I do believe the theory of the BLM. It takes the chemistry that change metal speciation and bioavailability and interaction of metal with organisms into account. It is a scientific base model while the regression approach is more statistic sound and the details of other chemistry rather hardness and pH is not quantitatively explained.

3. **Currently, States use Water Effects Ratio (WER) adjustment to the 87 ug/L chronic aluminum value. Do you think the application of the BLM or MLR as a site-specific adjustment reduces uncertainty associated with metals bioavailability and toxicity? Please explain.**

WER is an approach to apply laboratory results to field environments. Using WER approach, only some important water quality of laboratory and filed waters are matched but many others that can influence metal speciation and bioavailability are not included. BLM takes all into account and therefore could be more realistic and less uncertainty.

4. **The chronic aluminum BLM has been validated using EU validation procedures. We currently do not have validation procedures in the EPA 1985 Guidelines document ("*Guidelines for Deriving Numerical National Water Quality Criteria for the Protection of Aquatic Organisms and Their Uses*" (Stephan et al. 1985)). The EU requirements are for validation studies at three trophic levels including alga, an invertebrate, and a fish. Validation studies have been conducted for the alga (*Pseudokirchneriella subcapitata*), the cladoceran (*Ceriodaphnia dubia*) and fathead minnow (*Pimephales promelas*). Preliminary analyses indicate that the observed EC20 of the toxicity studies and the predicted BLM values are within a factor of 2 (100% for *P. subcapitata*, 100% for *P. promelas*, and 96% for *C. dubia).***

   Are the results from the validation of the BLM sufficient to support the incorporation of the BLM directly into the aluminum criteria document for regulatory use? Please explain. Do you have any suggestions?

I strongly encourage using the BLM for setting Al criteria. However, the data used for the BLM development and calibration here seem to be weak. As mentioned in question 1, DOC is very important factor for BLM. However, half of the data were from studies that were conducted 30-40 years ago and DOC was not measured. The concepts of DOC and dissolved metals likely were not concerned at that time. The low DOC

concentrations in the recent studies are still my questions as discussed in question 1. Therefore, I think more data, especially with higher DOC to represent the natural environmental conditions are needed to calibrate the BLM before using it for regulation purposes. Although the user manual said the model was calibrated with a wide range of DOC (0.5 to 30 mg/L), I don't see any high DOC concentration in this data base. Toxicity data with higher DOC are necessary for model validation.

5. **Ease of Use: One of the comments we hear from states is that the BLM is difficult to use and they also are unclear as to how to put it into standards.**

    - **In your opinion, which approach is easiest to use?**
    - **Do you have any suggestions as to how to make an approach easier for a stakeholder (i.e., State) to use- examples such as improvements to user manual, better upfront input design, etc.?**
    - **Do you have suggestions to improve the ease of use of the BLM?**
    - **It is possible to develop an MLR equation for the BLM. Do you believe that would be helpful? Please explain.**

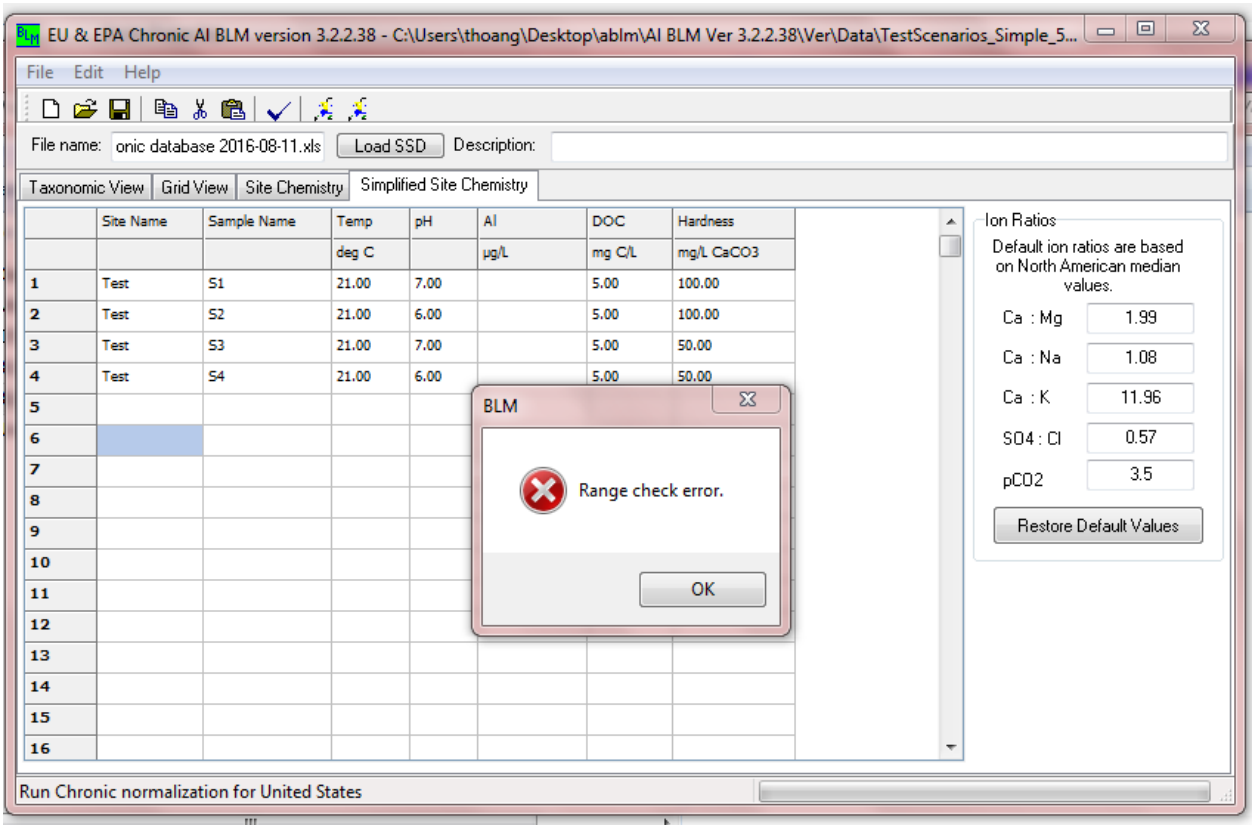
I agree with the comments that BLM is more difficult to use than regression approach. With the regression approach, hardness and pH are the only two required parameters and anyone can plug these parameters in the equation to calculate the criteria. Running the BLM requires more chemistry input, technical and knowledge. However, it is a scientific base model and can be used for wide range of water chemistry and more environmental relevant.

Although the current BLM version is more advanced than previous versions, such as it incorporated the probability plot tool into the program that give the users options to play around with output graphs and selection, it is still more complicate than the regression approach. In addition, I got some technical problems when running the model, such as an error showed in picture below when I changed the input data to 5 mg/L DOC to predict for different scenarios. To resolve this problem I had to close the model and reopen it. Also a few times the model was freezing towards the end of the run that did not allow me to further change the input data to run for other scenarios. Onetime I left the model ran over night and it was Ok in the morning. This indicates running time issue. These problems happened with my both desktop and laptop computers. I am not sure this is technical problems from my end? Given that said, it would be ideal and much easier to use if the modeler can develop a BLM version with more interface with excel like Toxcalc or its recent replaced CETIS of Tidepool Inc.

In the output files, I only see results for three scenarios while I ran the model for four scenarios given in the table above. The results for scenarios 4 (pH 6, hardness 50) were obtained from the probability plot but not from the output files. It needed an explanation for CCC and HC5 somewhere in the user manual or in EPA technical guidelines. The model gives estimated HC5 based on Log Normal Distribution Model and estimated HC5 based on truncated Triangle Distribution (US EPA FAV). They are little different. The users would appreciate knowing which HC5 is the CCC value. The values I presented in the table above are estimated HC5 based on Normal Distribution Model.

The user manual describes the principles of the BLM, input and out data, explains the functions of each menu item, … and example application. I still had some difficulties for getting the model run at the beginning. It would be more helpful to have a section in the manual that describes the steps for running the

model. Although the model said about the options for calculating metal speciation and toxicity but I don't see speciation option in this version.



COMMENTS SUBMITTED BY

Reviewer 5

External Peer Review of a Comparison of Aluminum Aquatic Life Criteria Approaches

CHARGE QUESTIONS

1. **Using the scenarios provided in the table below, do a side-by-side comparison of the results of Full BLM, Simplified BLM, and Excel results from the 2 MLR approaches (highlighted in yellow in Excel file).**
    - **Please draw some conclusions regarding the differences in the values generated and explain your rationale.**
    - **Would the simplified BLM be sufficiently protective?**
    - **Would the pH and hardness MLRs be appropriate?**
    - **Please provide appropriate suggestions.**

The Full BLM, Simplified BLM and the Excel results from MLR Comparison Approach 1 generated similar chronic values for aluminum in the five scenarios examined (See Table 1). Using MLR Comparison Approach 2, the chronic values were much lower than those generated by the other three methods (approximately 37% of those using MLR Approach 1).

In comparing the Full BLM and Simplified BLM, the chronic values using scenarios 1 and 2 (Hardness=100 mg/L; pH of 7 and 6, respectively) were very similar. However, when a lower hardness value of 50 mg/L was used in scenarios 3 and 4, the Simplified BLM generated chronic values below those from the Full BLM, which would be more protective of aquatic life. Alternatively, when a higher hardness value of 200 mg/L and a higher pH of 7.5 was used in scenario 5, the Simplified BLM generated a chronic value well above that from the Full BLM. **The Simplified BLM would be sufficiently protective in all cases except scenario 5 (higher hardness and pH).** However, this scenario (conditions) would be less commonly encountered of the five examined in most aquatic systems.

Using the Full BLM, when keeping pH constant and changing only hardness at a DOC of 1 mg/L, there were no differences in the chronic values generated. When using the Simplified BLM for the same conditions, the chronic values did decrease somewhat; however, these values would still be sufficiently protective. Using the Simplified BLM, at a higher DOC of 10 mg/L the chronic aluminum values generated were substantially higher than using an input value of 1 mg/L DOC with all of the scenarios (Table 1). Also, at 10 mg/L DOC, change in pH from 7 to 6 and change in hardness from 100 to 50 mg/L, did not influence the chronic aluminum value as much as at 1 mg/L DOC.

Using MLR Approach 2, where CCC is calculated by the FAV/FACR, generated much lower chronic aluminum values than the other approaches. Application of the FACR can bias the CCC low, which seems to be occurring here. **In my opinion, MLR Approach 2 is not as accurate as the other three.**

Alternatively, MLR Approach 1 generated chronic aluminum values that were very similar to those calculated by the Full and Simplified BLMs. The MLR Approach 1 uses an equation which is based largely on pH and hardness, both of which greatly influence predicted aluminum toxicity in the BLM. Differences between these approaches are more apparent when DOC values are higher, as the BLM uses DOC as an input variable and the MLR Approaches do not.

Table 1. Hypothetical Scenarios for Water Chemistry at Different Sites

	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5
INPUTS					
Temperature (°C)	21	21	21	21	21
pH	7	6	7	6	7.5
Hardness (mg/L Ca CO_3)	100	100	50	50	200
DOC (mg C/L)	1.0	1.0	1.0	1.0	1.0
HA (%)	10	10	10	10	10
Ca (mg/L)	26	26	26	26	26
Mg (mg/L)	8.0	8.0	8.0	8.0	8.0
Na (mg/L)	12	12	12	12	12
K (mg/L)	1.4	1.4	1.4	1.4	1.4
SO_4 (mg/L)	56	56	56	56	56
Cl (mg/L)	3.8	3.8	3.8	3.8	3.8
Alkalinity (mg/L Ca CO_3)	55	55	55	55	55
APPROACHES					
Full Aluminum BLM	215.546	81.10529	215.5462	81.10529	355.24786
Simplified Aluminum BLM	218.11829	82.3856	158.82442	51.36953	465.27683
Simplified Aluminum BLM (+10 mg/L DOC)	375.98963	369.5081	334.78604	368.44722	570.46629
MLR COMPARISON Approach 1	200	105	170	89	326
MLR COMPARISON Approach 2	74	39	63	33	121

2. Do you believe the scientific and theoretical foundation of the chronic aluminum BLM is sound and defensible? Does the aluminum BLM improve the Agency's ability to predict toxicity to water column organisms due to aluminum in comparison to the currently applied dissolved aluminum concentration criterion?

The BLM is used to predict a site specific chronic hazard concentration for aluminum. A strength of the BLM is that it uses a large database with chronic values for algae, inverts (mainly *C. dubia* and *D. magna*, as well as others), and sensitive fish species. Additionally, the BLM considers various water chemistry parameters (i.e. temperature, pH, DOC, hardness, alkalinity, CO_2, and major cations and anions) to determine the chronic aluminum value. The premise is that cations will compete with

aluminum ion to bind the biotic ligand (i.e. binding site on fish gill) and anions could complex aluminum, thereby changing its speciation and (usually reducing) toxicity. The scientific and theoretical foundation of the chronic aluminum BLM is both sound and defensible. Because a larger database of chronic toxicity data is used, and more water quality parameters are used than in the currently applied dissolved aluminum concentration criterion, I believe that the BLM will improve the Agency's ability to predict toxicity of aluminum to water column organisms.

3. **Currently, States use Water Effects Ratio (WER) adjustment to the 87 ug/L chronic aluminum value. Do you think the application of the BLM or MLR as a site-specific adjustment reduces uncertainty associated with metals bioavailability and toxicity? Please explain.**

   There is more uncertainty using the WER adjustment to the 87 µg/L chronic aluminum value than in using the BLM or MLR approach. These methods incorporate additional water chemistry parameters and have considered multiple species.

4. **The chronic aluminum BLM has been validated using EU validation procedures. We currently do not have validation procedures in the EPA 1985 Guidelines document ("*Guidelines for Deriving Numerical National Water Quality Criteria for the Protection of Aquatic Organisms and Their Uses*" (Stephan et al. 1985)). The EU requirements are for validation studies at three trophic levels including alga, an invertebrate, and a fish. Validation studies have been conducted for the alga (*Pseudokirchneriella subcapitata*), the cladoceran (*Ceriodaphnia dubia*) and fathead minnow (*Pimephales promelas*). Preliminary analyses indicate that the observed EC20 of the toxicity studies and the predicted BLM values are within a factor of 2 (100% for *P. subcapitata*, 100% for *P. promelas*, and 96% for *C. dubia).***

   Are the results from the validation of the BLM sufficient to support the incorporation of the BLM directly into the aluminum criteria document for regulatory use? Please explain. Do you have any suggestions?

   The organisms chosen are representative of three trophic levels, generally sensitive to contaminants and are commonly used in toxicity testing. Past versions of the BLM with other metals have been considered acceptable if the observed toxicity values and the predicted BLM values are within a factor of 2. If the analyses confirm that is the case for *P. subcapitata*, *P. promelas* and 96% of the data collected for *C. dubia*, in my opinion the results from the validation of the BLM are sufficient to support the incorporation of the BLM directly into the aluminum criteria document for regulatory use. Further, when multiple toxicity tests are performed (especially by different laboratories) EC20 values can be produced that differ by more than a factor of 2.

   C. dubia is usually the most sensitive of the three species tested. My only question would be whether the BLM over predicted or under predicted toxicity in 4 % of the *C. dubia* data. If the BLM over predicted toxicity (generated a lower chronic value) to *C. dubia,* then the chronic aluminum value generated would still theoretically be protective.

5. **<u>Ease of Use:</u> One of the comments we hear from states is that the BLM is difficult to use and they also are unclear as to how to put it into standards.**
    - **In your opinion, which approach is easiest to use?**
    - **Do you have any suggestions as to how to make an approach easier for a stakeholder (i.e. State) to use- examples such as improvements to user manual, better upfront input design, etc?**
    - **Do you have suggestions to improve the ease of use of the BLM?**
    - **It is possible to develop an MLR equation for the BLM. Do you believe that would be helpful? Please explain.**

The MLR approaches are easiest to use. I had problems mainly in opening the BLM. I downloaded the file and saved it to dropbox and could not open the BLM. When I downloaded the file to my desktop; however, I was eventually able to open the BLM. I had to disable my anti-virus software to do so. Once I was able to open the BLM, it was fairly simple and straightforward to use.

I would suggest more detailed instructions in downloading and opening the BLM in the user manual. I did not find that area particularly helpful.

I would think that developing an MLR equation for the BLM would be possible, but I do not know the answer to that question. I don't really think that developing an MLR equation for the BLM is needed, given that the chronic aluminum values generated by the BLM were very similar to the ones generated by MLR Approach 1.