RAMBOLL ENVIRON

ENVIRONMENT
& HEALTH

Comments on: The Proposed Amendment of the Sulfate Water Quality Standard Applicable to Wild Rice and Identification of Wild Rice Waters.

OAH Docket NO. 80-9003-34519,
Revisor NO. RD4324A
November 21, 2017

Comments by: Michael J Bock, PhD on behalf of the Iron Mining Association of Minnesota

Comments addressing: (1) Derivation of a porewater sulfide threshold, and
(2) Use of an equation to derive a sulfate standard

I am a professional environmental scientist and Senior Managing Consultant at Ramboll Environ. I have more than 20 years of professional experience in environmental science and statistical analysis. My educational background is in biogeochemistry, biological science, data analysis, and statistics. I hold a PhD in Oceanography from the University of Delaware. I am proving this written testimony to support and expand on the oral testimony I provided on the 23rd of October in St Paul.

My testimony addresses two aspects of the proposed sulfate rule (RULE 7.20-7.26) that are lacking in reasonableness and scientific basis: (1) the use of 120 microgram per liter (µg/l) as a porewater sulfide threshold value for the protection of wild rice (Rule 7.20; SONAR part 6E), and (2) the use of the MPCA equation to predict a waterbody specific sulfate standard based on sediment total organic carbon (TOC) and iron and the sulfide threshold (Rule 7.26-8.2; SONAR 6E p75-77)

Specifically, in my judgement the field data does not support a porewater sulfide threshold of 120 µg/l as a reasonable sulfide threshold due to deficiencies in the raw data and statistical analyses used to derive this value. A higher sulfide threshold is predicted by statistical analyses to be as protective as the 120 µg/l value. Furthermore, the equation used to translate the sulfide threshold to a surface water sulfate threshold does not provide reasonable predictions of the sulfate threshold and yields results that are often illogical and inconsistent with MPCAs conceptual model of the interaction between sulfate and sulfide.

SULFIDE THRESHOLD COMMENTS (RULE 7.20; SONAR PART 6E)

Reasonableness of the proposed sulfide threshold of 120 µg/l

An examination of the field data used by MPCA to support the rulemaking shows that there are a great many waterbodies in the MPCA dataset that exhibit porewater sulfide concentrations that exceed the MPCA proposed threshold (>120 µg/l) and also possess healthy stands of wild rice (Figure 1 and Figure 2). This finding calls into question the validity of MPCA threshold and suggests problems and deficiencies in how MPCA used the field data to derive a threshold. As MPCA has observed, the field data provides a useful dataset for assessing the potential influence of porewater sulfide on wild rice which can be supplemented by the laboratory mesocosms and hydroponics data. A discussion of the deficiencies in the MPCA sponsored hydroponics and mesocosm studies is beyond the scope of my comments but will be addressed by others. During my technical review of the data and analyses I identified and, when possible, corrected the deficiencies in the analyses conducted by MPCA. I describe my own derivation of a conservative sulfide threshold based on the field data. Based on my analyses, the conservative sulfide threshold I derived is as protective of wild rice health as the 120 µg/l MPCA standard.

Figure 1. Presence of Wild Rice versus Porewater Sulfide Concentration

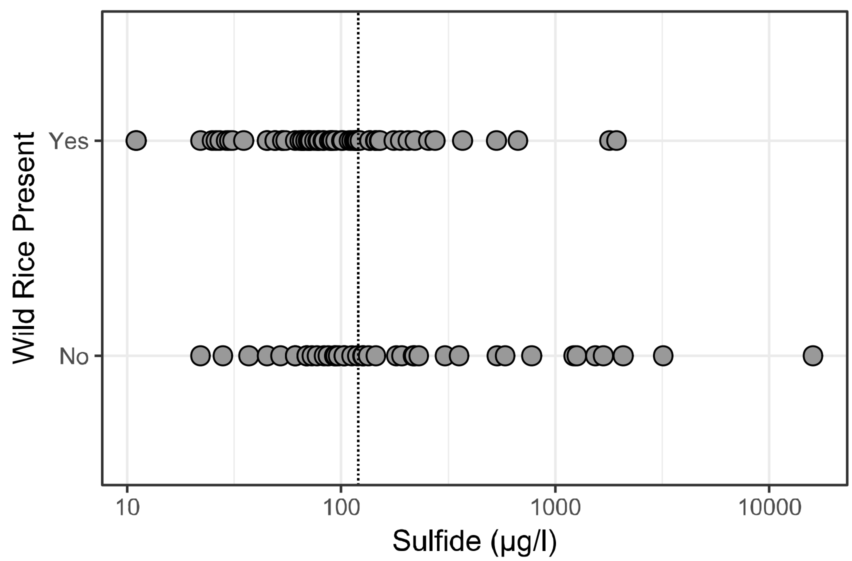
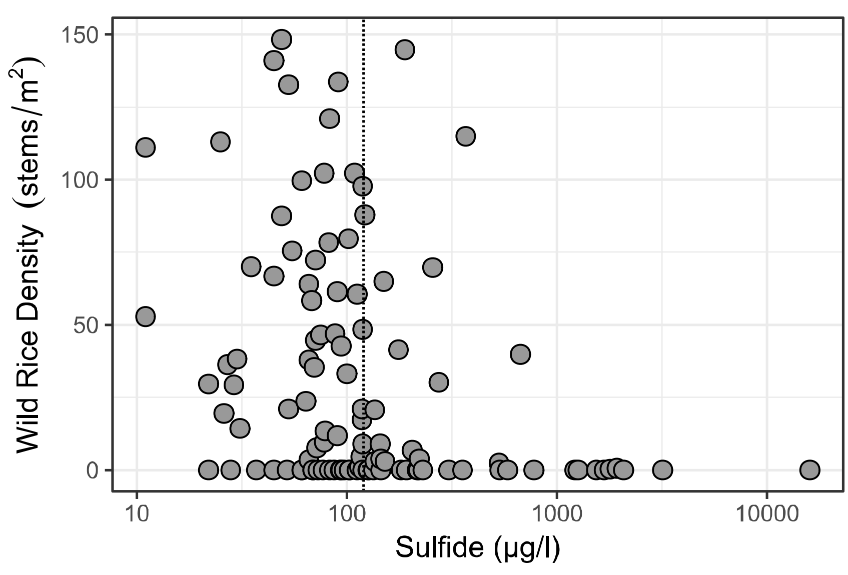


Figure 2. Wild Rice Stem Density versus Porewater Sulfide Concentration



Representativeness of the Field Data used by MPCA

MPCA generated a field dataset as part of the derivation of a new sulfate standard. Wild rice is known to be sensitive to a variety of factors unrelated to sulfide (e.g. water depth, winter temperature, the presence of competing species (TSD pages 27-28)). The field data used in MPCA's analysis was collected in 2012 (n=83) and 2013 (n=25). Significant flooding was reported in the Duluth region in 2012. This flooding occurred in June, a critical time for the germination of wild rice. MPCA did not discuss the possible importance of this flooding on in the 2012 data and the derivation of the sulfide threshold. In fact, more than 75% of the 25 samples with porewater sulfide between 100 µg/l and 150 µg/l were collected in 2012, the samples expected to have the most influence on the MPCA's 120 µg/l threshold. The potential biases associated with these data points has not been addresses by MPCA.

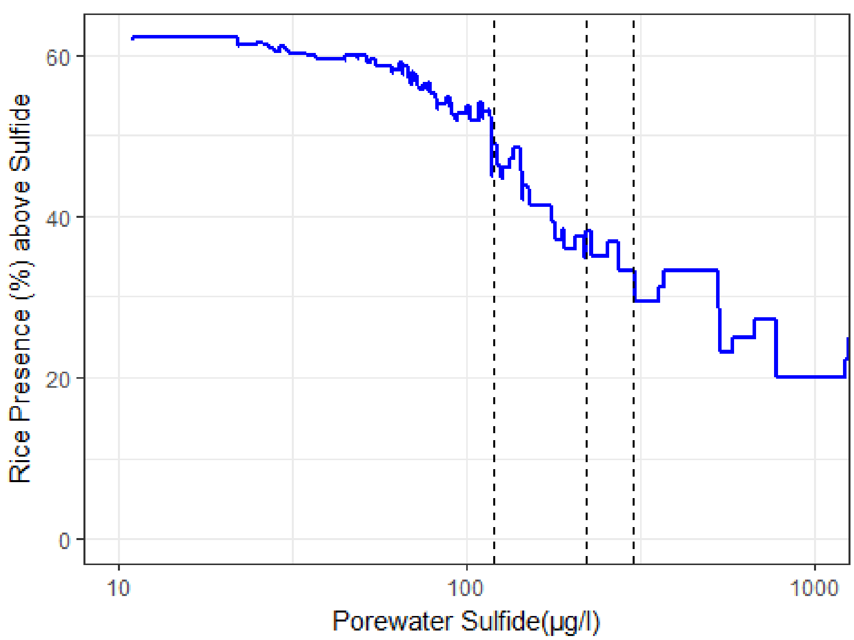
Appendix 4 of the TSD provides measurements of wild rice in 20 fixed sampling points over 12 years. These data show that there is tremendous annual variability in the density of wild rice stands. This variability calls into question MPCA's use of wild rice health metrics measured in a single year. Table A4-1 (page 112) describes annual variability of a factor of five or more in the same location. This variability complicates the interpretation of the field data.

FINDING: MPCA has not considered the impacts of the 2012 flooding on the representativeness of the field data from this year and has not addressed the importance of annual variability in interpreting wild rice health metrics.

Breakpoint Analysis

MPCA's first derivation of a sulfide threshold is based on the 'breakpoint' analysis of the field data described on page 69 of the SONAR and on pages 37 and 39 of the Final Technical Support Document (TSD; MPCA 2017). To derive this threshold MPCA visually examined a plot of porewater sulfide concentrations versus a measurement of wild rice health, specifically the presence of wild rice. MPCA identified a 'breakpoint' based on looking at a plot and identifying a drop in the presence of wild rice at 120 µg/l (SONAR 6E p 69). MPCA's plot is shown in Figure 3. This threshold is based entirely on MPCA's visual interpretation of the plot. In my professional judgement there is no visual evidence for a breakpoint at 120 µg/l and this value represents a visual artefact. The use of professional judgment, either MPCA's or my own, can easily lead to unconscious biases and has a high potential for erroneous conclusions.

Figure 3. MPCA Empirical Threshold Estimation



There are statistical methods that can be used to identify breakpoints, specifically piecewise regression (Seber and Wild, 1989). These methods avoid the biases associated with professional judgment and provide a statistical basis for decision making. When these methods are applied to the MPCA field dataset they indicate (1) if the lake with the highest sulfide is excluded (Bean)[1] the 'breakpoint' is more than twice the value identified by MPCA (2) if all water bodies are included[2] the breakpoint is more than 1000x the MPCA value and suggests no sulfide threshold in the field data. A summary of the statistically derived breakpoints in shown in Table 1. Note that the data was not log transformed prior to analysis as the unaltered data fit the statistical model without need for modifying the data.

Table 1. Sulfide Breakpoints Determined by Piecewise Regression

Data Set	Sulfide Breakpoint (µg/l)
All, n=108	16000
Exclude Bean, n=107	233
Transparency > 30 cm, n=96	13000
Transparency > 30 cm and exclude Bean, n=95	244

My re-evaluation of the MPCA 'breakpoint' analysis is based on well-defined statistical methods that avoid the biases and uncertainties associated with relying entirely on professional judgement. Based on these statistical analyses, the statistically determined breakpoint is at a minimum more than twice the MPCA value, and, if Bean is included, it is orders of magnitude higher than the MPCA value. However, it is important to recognize that only 16% of the data (n=17) exhibited porewater sulfide concentrations above 300 µg/l, meaning that there are insufficient data to confidently derive a breakpoint if the true breakpoint is higher than 300 µg/l. More simply put, the true threshold could be substantially higher than 300 µg/l.

FINDING: MPCA has not justified the use of an analysis based entirely on professional judgement. When statistical methods are applied the MPCA's value based on professional judgement is found to be incorrect. Statistical methods indicate a breakpoint as high as 16,000 µg/l.

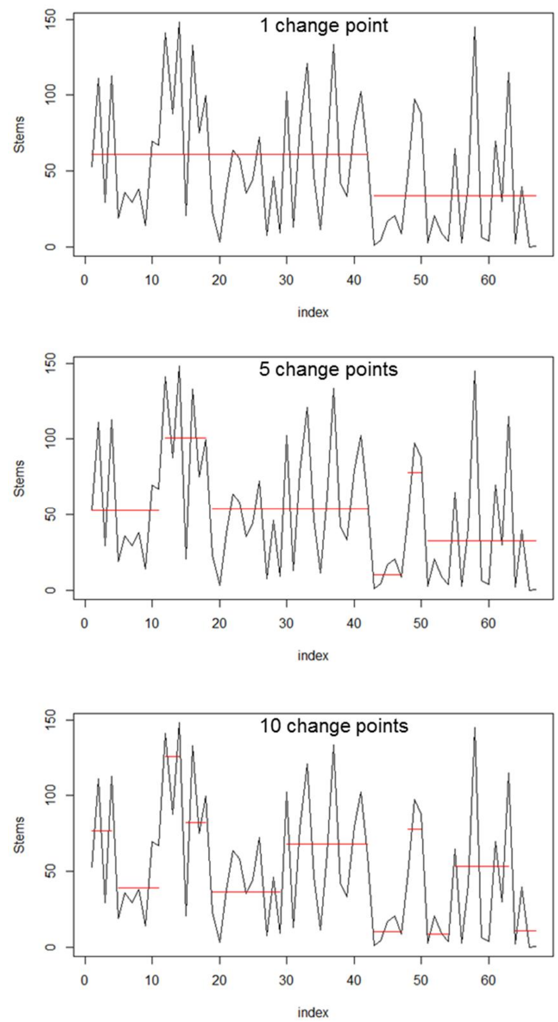
Change Point Analysis

MPCA conducted another analysis of the field data using 'change point' analysis to identify a threshold (SONAR 6E p 69). This method has been described in the peer-reviewed scientific literature (Hawkins 2001, Killick et al. 2016) and does not rely on professional judgement. MPCA analyzed the wild rice stem density data using this method. In brief, the observations are ordered from lowest to highest sulfide concentration and the algorithm searches for the value above which some measure of wild rice health, in this case stem densities, are different from those below the value. The algorithm was run using only waterbodies with wild rice present (n=67). A change point of 112 µg/l was identified. There are two potential issues with this analysis: (1) the confidence limits around the 112 µg/l value are 14 to 239 µg/l, meaning that there is low precision in the estimated change point, and (2) MPCA did not

[1] MPCA refers to this as the Class B n=107 dataset
[2] MPCA refers to this as the Class B n=108 dataset

test for the presence of multiple change points, which allows the change point to be further validated (Killick et al. 2016). I reran the analysis allowing for 1, 5, and 10 change points. The change point results are shown in Figure 4. The results of these analysis show that the single change point identified by MPCA is not unique and in fact does not represent a change point that can be associated with a change in wild rice density.

Figure 4. Sulfide Change Point Analysis, Stem Density



Although MPCA limited their change point analysis to stem density, this analysis can also be applied to the presence of wild rice and to the presence of dense stands of wild rice with a stem density of greater than 40 stems per square meter (TSD page 50). This differs from the previous analysis in that it is a binary change point analysis. The algorithms from Hawkins (2001) were used for this binary change point analysis. The results are shown in Table 2. Unfortunately, this algorithm does not permit the analysis of multiple change points and is limited to the analysis of a single change point. Therefore, these values have not been fully validated and should only be used to support other lines of evidence in the analysis of sulfide thresholds. These statistical change points are all substantially higher than the MPCA threshold of 120 µg/l, supporting the conclusion that 120 µg/l is below the true threshold.

FINDING: MPCA's change point analysis fails validation and should not be relied upon. When change point analyses are performed on additional wild rice metrics they result in values substantially higher than MPCA's value, further invalidating the MPCA analysis.

Table 2. Updated Sulfide Change Point Analysis, Wild Rice Presence and Stem Density

Data Set	Presence Breakpoint (µg/l)	Density Breakpoint (µg/l)
All, n=108	530	368
Exclude Bean, n=107	176	368
Transparency > 30 cm, n=96	274	368
Transparency > 30 cm and exclude Bean, n=95	274	368

Dose-Response Analysis

Although the methods described above are well founded in statistical theory and provide important information regarding the relationship between sulfide and wild rice health metrics, these methods are not typically used to derive protective thresholds. Typically a dose-response statistical model would be used for this sort of data, such as the relationship shown on pages 119-120 of the TSD for probability of wild rice presence versus pore water sulfide. However, the field data do not fit the requirements of such a model; specifically (1) there is no well-defined no-effect level due to high variability at all sulfide concentrations, and (2) sulfate is a nutrient required for plant growth (TSD page 53). At very low sulfate concentrations wild rice health may be reduced, a relationship that does not fit the statistical model used by MPCA. Thus, although MPCA does fit the field data to a dose-response curve, the data do not fit the assumptions of the statistical model and therefore any sulfide threshold derived using this method should not be used. Based on these challenges associated with a typical dose-response model and the fact that filed data do not fit the model, I selected binary analysis as a simple and robust method for analyzing these data for changes in wild rice health at higher sulfide concentrations.

FINDING: MPCA's dose response modeling does not meet the assumptions of the analysis, does not fit the data, and should not be used to derive an EC_{10}.

Binary Analysis

Binary analysis presents an alternative method for analyzing the relationship between wild rice and sulfide that is not affected by the issues that make the dose-response modeling of the field data invalid. Binary analysis is accomplished by combining data with similar sulfide concentrations into bins and analyzing wild rice health metrics for each bin. The bins were set by the statistical software such that the target number of samples in each bin is as close to 10 as possible. A sample size of 10 seeks to optimize the increased statistical power associated with a large sample size while maintaining sulfide ranges that are narrow enough to be useful for deriving a threshold. The resulting dataset was analyzed using two complimentary methods: (1) testing if there is any influence of sulfide on wild rice health metrics using a numeric data binning function (R function 'cut2'; Harrell 2017) and the chi-squared test, and (2) determining which concentration bins exhibit reduced wild rice health by comparing the confidence limits for the wild rice health metric in each concentration bin using the confidence intervals for binomial probabilities (R function 'binconf'; Harrell 2017). It is important to note that the analysis may be unable to determine which bins have reduced health metrics even if a significant difference between sulfide concentration bins is found (Zar 1984; Lyman Ott 2010)

Two wild rice health metrics were subjected to binary analysis: (1) the presence of wild rice, and (2) the presence of high density stands of wild rice (>40 stems per square meter). These same metrics were used by MPCA in their analyses (presence/absence of rice, presence/absence of high density stands of rice) (SONAR 2 page 69). My analyses are shown in Figure 5. In the top two plots of Figure 5, the data points represent the proportion of waterbodies with wild rice present for each sulfide concentration bin. The error bars represent the 95% confidence intervals. When the confidence limits for two concentration bins overlap then the difference between the percentages of locations with wild rice present in not significantly different using a typical 95% confidence level. The presence of high densities of wild rice is shown in the bottom two plots of Figure 5. These plots clearly show that there is no statistically significant difference in the wild rice health metrics for the sulfide concentrations bins. The plots do suggest some decrease at the top two concentrations bins, but the difference is minor and not statistically significant. Thus, these results indicate that the MPCA threshold of 120 µg/l is too low, and higher thresholds (2-3) are just as protective as the MPCA threshold. Furthermore, there are too few data points in the field data with porewater sulfide values high enough (300 µg/l or higher) to reliably determine a true upper threshold (Table 3).

Figure 5. Sulfide Binary Analysis, Wild Rice Presence and Stem Density

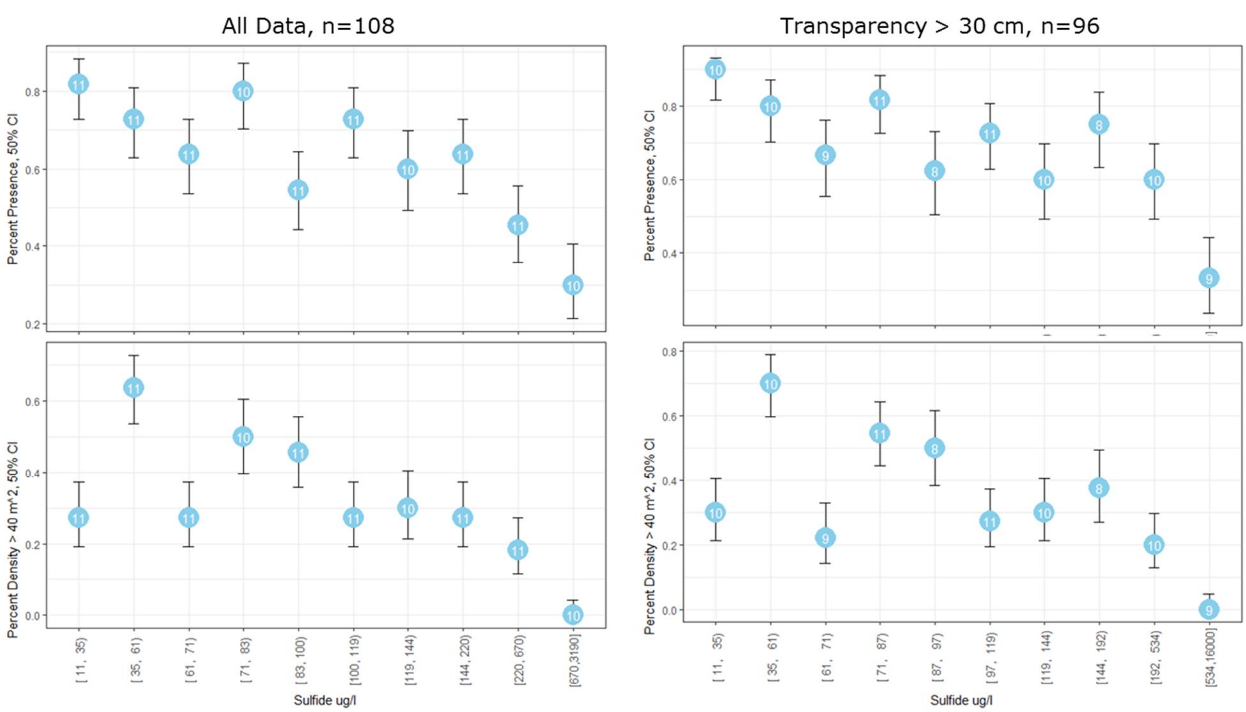


Table 3. Percentage of Sites with Sulfide >300 µg/l	
Data Set	Percent with Sulfide >300 µg/l
All, n=108	16%
Exclude Bean, n=107	15%
Transparency > 30 cm, n=96	14%
Transparency > 30 cm and exclude Bean, n=95	13%

FINDING: This analysis is more robust than the MPCA dose response analysis. The results of the binary analysis indicates that the MPCA threshold of 120 µg/L is not reasonable or valid. The MPCA value is too low and the true threshold is much higher.

Summary

When I analyzed the field data using appropriate methods I found no evidence that increasing the sulfide threshold to values 2-3 times the MPCA value would lead to a discernible decrease in the health of wild rice. There is insufficient data to reliably evaluate higher thresholds.

MPCA unreasonably excludes the alternative threshold of 300 µg/l in TSD Appendix 9. MPCA argues if you test one threshold, you cannot test another using the same data. MPCA's justification for excluding these data is unreasonable. MPCA indicates that a basic tenet of statistics is that a data point cannot be reused in a statistical analysis. While strictly speaking this is true, MPCA fails to acknowledge that the field data is being reused for a large number of statistical analyses. Typically the reanalysis of data in this manner represents exploratory data analysis. In this situation, the statistician must be very careful in ensure that they acknowledge the uncertainty in this approach. The statistician need not dismiss the analysis but rather should account for these factors by weighing each line of evidence when drawing conclusions. MPCA does offer an alternative analysis of the 300 µg/l threshold. The results indicate that 120 and 300 µg/l are similarly protective with respect to presence/absence, but density is reduced when 300 µg/l is compared to 120 µg/l.

FINDING: Based on the weight of evidence, I conclude that the 120 µg/l sulfide threshold proposed by MPCA is overly conservative and the true threshold is at least 2-3 times higher than the MPCA threshold.

SULFATE EQUATION COMMENT (RULE 7.26-8.2; SONAR 6E P75-77)

MPCA developed an equation to predict a waterbody specific sulfate standard based on sediment Total Organic Carbon (TOC) and iron and the sulfide threshold using multiple binary logistic regression (MBLR). I recalculated water body specific sulfate thresholds using the MPCA method and different sulfide thresholds to test the stability of the sulfate equation and its ability to predict wild rice health. I found a significant conflict in the performance of the equation that indicates that the equation does not provide sound predictions of the relationship between sulfide and sulfate and therefore is an unreasonable standard. Specifically, if the sulfide threshold is increased one would expect the sulfate threshold for a given water body to also increase. I found that in a large number of instances, when the sulfide threshold is increased the sulfate threshold decreases. A few examples are shown in Table

4. If elevated sulfate leads to high sulfide then we should not see these trends in sulfate thresholds. Thus, the MBLR truly is unable to predict sulfate from TOC and Fe for a porewater sulfide threshold and has limited predictive power. Any equation used to derive a sulfate standard must yield higher sulfate standards in a waterbody when higher sulfide thresholds are used. The fact that this is not true for the MBLR equation indicates that the equation is likely to lead to erroneous conclusions and is potentially a statistical anomaly. The use of such an equation presents a fundamental flaw in MPCAs approach and should not be used.

Table 4. Sulfate Standards from the MBLR Equation for Various Sulfide Thresholds

Waterbody	Sulfate Standard (mg/l) based on n=108			
	Sulfide 120	Sulfide 176	Sulfide 220	Sulfide 274
Pleasant	6.8	10.1	24.6	29.3
Second	148	806	1042	615
Lady Slipper	26.5	142.3	126.8	115.5
Duck Lake WMA	17.0	25.1	76.0	71.8
Sturgeon	69.6	596	409	300

FINDING: MPCA's equation does not fit the conceptual model of the relationship between sulfate and sulfide, TOC, and iron. The equation violates a basic rule that if the permissible sulfide threshold is increased the sulfate threshold should also increase. The equation does not accurately described the relationship between sulfate and sulfide and cannot be used for deriving sulfate thresholds.

VALIDATION AND ERROR RATES (RULE 7.26-8.2; SONAR 6E P75-77)

A critical component to any statistical model is model validation and an examination of the error rates. One component of model validation is 'does the model fit our understanding of geochemical processes?' In section D, pages 42-43 of the TSD, MPCA admits that the model used to derive the sulfate threshold are not based on the geochemical processes that relate sulfate to sulfide, TOC, and iron. In fact, the model is based solely on fitting a statistical model to the data. The reliance on a statistical model does not disqualify the MPCA equation, but in validating the model it is important to assess how well the model fits know geochemical processes, as demonstrated above in my comments on the MPCA equation, the model fails this simple test. Specifically, if TOC and iron are constant and the acceptable level of sulfide is raised, the model must derive a higher sulfate threshold. The model does not function in this way. In fact, when the entire dataset is examined for the example sulfide thresholds shown in Table 4, the highest sulfate threshold is not associated with the highest sulfide thresholds for the majority of waterbodies. Any statistical model that is inconsistent with known geochemical processes in this way should be rejected.

Even though the MPCA model fails the validation criteria given above, I also looked at error rates and how MPCA assessed the validity of the model. MPCA calculated error rates by looking at how often the equation correctly determines if the sulfide measured in porewater exceeds the sulfide threshold of 120 µg/l. MPCA also examine the error rates for a number of other thresholds (TSD page 57=58, Table 1-10). This is an incorrect way to validate the model and overestimates the accuracy of the model. If the purpose of the model is to protect the health of wild rice then a more appropriate determination of the error rate is how well the model predicts wild rice health. The following is an excerpt from the public hearing transcript from the 23 October hearing in St. Paul. Page 152, lines 14-25

MR. BOCK: Michael Bock. I think she already stated my question or clarifying. When you're talking about error rates, you're talking about the ability to predict the sulfide concentration above and below the threshold and not -- and that error [‘area’ in transcript’] has nothing to do with the presence or absence of wild rice if the equation predicts it's above or below the threshold based on the sulfate. Is that a correct interpretation?

MR. SWAIN: The error rates we've been discussing are how accurately the sulfide concentration is predicted and has nothing to do with wild rice presence and absence, I agree.

This is a critical observation. The standard should be validated against wild rice health metrics and NOT solely on predicted porewater sulfide. Given that wild rice health metrics are available in the dataset, there is no reason not to examine if the equation accurately predicts the health of wild rice.

The following tables show how two wild rice health metrics (presence/absence of wild rice or presence/absence of high density wild rice [> 40 stems m^2] are related to waterbodies that either (1) have porewater sulfide greater that a given threshold (120, 176, 220, and 274 µg/l) or (2) have sulfate in surface water exceeding the calculated standard based on a given sulfide threshold (120, 176, 220, and 274 µg/l). These statistics were calculated using the n=96 dataset (transparency > 30 cm), and the columns sum to 100%, but the trends are consistent with the results from the complete dataset.

Table 5. Wild Rice Health Metrics versus Exceedances of Sulfide and Sulfate Thresholds								
Percentage of Sites with Wild Rice Relative to Sulfide Threshold								
	120 µg/l		176 µg/l		220 µg/l		274 µg/l	
	< 120	>120	<176	>176	<220	>220	<274	>274
Wild Rice Present	24	45	25	52	28	50	27	62
Wild Rice Absent	76	55	75	48	73	50	73	38

Percentage of Sites with Wild Rice Relative to Sulfate Exceeding the Sulfide based Threshold								
	120 µg/l		176 µg/l		220 µg/l		274 µg/l	
	< 120	>120	<176	>176	<220	>220	<274	>274
Wild Rice Present	27	40	30	35	29	43	29	56
Wild Rice Absent	73	60	70	65	71	57	71	44

Percentage of Site with Dense Wild Rice Relative to Sulfide Threshold								
	120 µg/l		176 µg/l		220 µg/l		274 µg/l	
	< 120	>120	<176	>176	<220	>220	<274	>274
Dense Wild Rice Present	57	82	60	86	61	88	61	92
Dense Wild Rice Absent	43	18	40	14	39	13	39	8

Percentage of Site with Dense Wild Rice Relative to Sulfate Exceeding the Sulfide based Threshold								
	120 µg/l		176 µg/l		220 µg/l		274 µg/l	
	< 120	>120	<176	>176	<220	>220	<274	>274
Dense Wild Rice Present	27	40	30	35	29	43	29	56
Dense Wild Rice Absent	73	60	70	65	71	57	71	44

These results show that the error rate for sulfate is much larger than the error rate for sulfide. Meaning that the equation based sulfate thresholds do a poor job of predicting wild rice health. Furthermore, the error rates for wild rice health based on sulfide and sulfate are virtually identical for the various sulfide thresholds. Based on these data I conclude that the MPCA threshold of 120 µg/l is no more predictive of wild rice health that 176, 220, and 274 µg/l, and perhaps higher thresholds. These analysis confirm the conclusion that the MPCA standard does a poor job of predicting wild rice health.

Finally the body of work produced by MPCA suffers from a fundamental statistical flaw. One should not use the same dataset used to derive a model to validate the model. A statistical model should be validated using an independent dataset, a process known as cross validation (Seymour 1993). Throughout this process MPCA has proposed various datasets for cross validation but I only address the validation dataset described in the TSD in these comments. MPCA devotes very little discussion to the validation dataset (TSD page 44 and 59) but based on a review of the raw data and the TSD I have determined that:

- The validation dataset consists of 47 water bodies
- The data was collected from the SAME waterbodies used to develop the equation and the SAME years, but the specific year used to derive the equation was not used in the validation dataset.

This clearly does not represent an independent dataset and therefore the model has not been cross validated. In order to properly validate the model MPCA must use a truly independent dataset. They must sample waterbodies NOT included in the development of the model AND the must use data from year OTHER than 2012 and 2013. Based on these deficiencies, MPCA has NOT cross validated the model used to derive the standard. A model that has not been validated is inappropriate for rule making and represents an unreasonable standard.

FINDING: MPCA has not validated the components of the rule, has not provided a true validation dataset, and has not properly evaluated error rates. Because of these deficiencies MPCA has not properly evaluated the effectiveness and reasonableness of the rule.

RECOMMENDATIONS

Based on my analysis of the sulfide threshold (Rule 7.20; SONAR part 6E) and sulfate equation (Rule 7.26-8.2; SONAR 6E p75-77), I am recommending the following changes:

1. Establishing a toxic porewater sulfide threshold of 120 µg/l is not reasonable.

2. MPCA's equation to predict a waterbody specific sulfate threshold based on TOC, iron, and the sulfide threshold is not reasonable.

REFERENCES

Harrell, F.E. 2017. Harrell Miscellaneous (R package version 4.0-3). https://cran.r-project.org/web/packages/Hmisc/Hmisc.pdf.

Hawkins, D.M. 2001. Fitting Multiple Change-Point Models to Data. Computational Statistics and Data Analysis 37:323-341.

Killick, R., K. Haynes, and I.A. Eckley. 2016. changepoint: Methods for Changepoint Detection (R package version 2.2.2). https://cran.r-project.org/web/packages/changepoint/changepoint.pdf.

MPCA. 2017. Final Technical Support Document: Refinements to Minnesota's Sulfate Water Quality Standard to Protect Wild Rice. Minnesota Pollution Control Agency. Saint Paul, MN. August.

Ott, R.L. and M.T. Longnecker. 2015. An introduction to statistical methods and data analysis. Nelson Education.

Seber, G.A.F. and C.J. Wild. 1989. Nonlinear Regression. Wiley, New York.

Geisser, Seymour .1993). Predictive Inference. New York, NY: Chapman and Hall. ISBN 0-412-03471-9.

Zar, J.H. 1984. Biostatistical analysis. 2nd Edition. Prentice Hall, USA.