NCDP Scientific Advisory Council Agenda

10:00am – 3:30pm November 14, 2018 Agronomic Division Building Conference Room 4300 Reedy Creek Road, Raleigh, NC 27607

Desired Outcomes:

- Shared understanding of the HRL Schedule.
- Shared understanding of exceedance frequency versus confidence levels.
- Shared understanding and resolution of criteria development sequence.
- Shared understanding and resolution of Chlorophyll *a* criteria.

Time	Торіс	Speaker(s)
10:00	Convene Introductions Approval/Comments on meeting minutes September Administrative Business (2019 Spring/Summer schedule, December 3-4 Meeting) 	Jenny Halsey (facilitator)
10:15	Draft HRL Schedule	Brian Wrenn
10:30	Chlorophyll a Considerations on Use of Exceedance Frequency vs. Confidence Levels 	Clifton Bell Jenny Halsey (facilitator)
11:00	Break	
11:10	 Chlorophyll a Discussions Frequency/Duration drive Magnitude v. Magnitude drive Frequency/Duration 	Lauren Petter Jenny Halsey (facilitator)
12:00	Lunch	
12:30	Continue Chlorophyll <i>a</i> Discussions	Jenny Halsey (facilitator)
2:00	Break	
2:15	Continue Chlorophyll <i>a</i> Discussions	Jenny Halsey (facilitator)
3:15	Wrap-up, closing remarks, and adjourn	Jenny Halsey (facilitator)
3:30	Adjourn	

High Rock Lake Nutrient Criteria Schedule					
Task Date Comment					
Complete development of Chla criteria	December 3, 2018	Draft criteria for Chla agreed to by SAC			
Complete development of N criteria	February 2019	Draft concentration/loading rate as criteria or "action level" for bioconfirmation process agreed to by SAC			
Complete development of P criteria	February/April 2019	Draft concentration/loading rate as criteria or "action level" for bioconfirmation process agreed to by SAC			
Complete development of any bioconfirmation criteria	April/June 2019	Draft bioconfirmation methodology agreed to by SAC			
Complete revisits of other response variables previously discussed	June 2019	Draft criteria for any response variables previously discussed agreed to by SAC			
Draft criteria proposal documents	August 2019	Completion of draft documents for review by SAC			
Submit final documents to CIC	October 2019	Final HRL criteria package submitted to the CIC			



North Carolina Nutrient Science Advisory Council

Considerations for the Frequency Component of CHLa Criteria

Clifton F. Bell | November 14, 2018



Frequency is a Fundamental Component of Water Quality Criteria



- Acknowledges:
 - Natural variability
 - Use resiliency
- Allowable exceedance rates also part of assessment guidance (e.g., 10% rule for toxics)
- Acknowledges uncertainty in:
 - Representativeness
 - Persistence
 - Sampling/analysis uncertainty

Assessment Error Types

Assessment	"True" Status		
Decision	Impaired	Attaining	
Impaired	Correct	Type I Error "False Positive"	
Attaining	Type II Error "False Negative"	Correct	

Various Frequency Concepts/Rules Have Been Applied, Most with Short-Duration Criteria in Mind

	l	1
Critical Exceedance Rate	Source	Application
\leq 1 sample in 3 years	EPA, 1997**	Acute criteria.
>10%	EPA, 1997**	Acute criteria not supporting beneficial uses. Sampling and measurement error accounted for.
0.09% (1 sample out of 1,095)	EPA, 2002	Acute criteria. A 1-in-3-year frequency of daily averages.
0.36% (1 sample out of 274)	EPA, 2002	Chronic criteria. A 1-in-3-year frequency of 4-day averages.
5% plus a 15% effect size	EPA, 2002	Toxics criteria, equivalent to a 1-in-3-year frequency.
10%	EPA, 2003	For chronic criteria, and acute if justified. Binomial or raw score test.

Binomial Test Well-Suited for Not-to-Exceed Criteria

- Requires:
 - Critical exceedance rate (e.g., 10%)
 - Confidence level (e.g., 90%, to limit Type I error rate to 10%)
 - Minimum sample size (i.e., minimum power, to address Type II errors)

$$\mathbf{P}(\mathbf{x}) = \frac{n!}{(n-x)! \, x!} \cdot \mathbf{p}^x \cdot \mathbf{q}^{n-x}$$

Other Methods for Reducing Type II Error (False Negatives)

- Set conservative magnitude
- Incorporate additional information into assessment



Two Potential Frequency Approaches for Seasonal Average Criteria

Explicit 1-in-3 (ala Florida)

- Calculate seasonal average for each year of data
- More than one exceedance \rightarrow consider criteria exceeded
- Requires at least:
 - Two years of consistent results
 - Third "tiebreaker" year if first two years have mixed results

Running (multi-year) average

- Data pooled from multiple years
- Option of using a statistical test such as t-test, confidence interval on geomean, etc.
- Still likely to require minimum number of years of data and/or data per year

Statistical Basis of the 1-in-3 Approach – Following Florida DEP (2012)

- Annual geometric mean CHLa criteria
- Type II errors addressed by using a reference condition approach to set criteria magnitude
- Water body just in attainment has 20% exceedance probability in any given year.
- Binomial formula shows Type I errors $\leq 10\%$

X (# exceedances)	P(X) (probability of X exceedances in 3 trials)	Cumulative P (probability of ≤X exceedances)
0	0.512	0.512
1	0.384	0.896
2	0.096	0.992
3	0.008	1.000

Considerations for applying the 1-in-3 Statistical Considerations to a NC CHLa Criterion

- Option of addressing Type II error rate both on criteria magnitude and narrative assessment elements.
- 20% exceedance probability would mean attainment required in most years, not just as a long-term average.
 - Considered when setting magnitude
- Same binomial math would apply as in FDEP (2012) $\leq 10\%$ Type I error rate

Examples of Multi-Year Average CHLa Criteria

- Minnesota
 - Long-term average
 - Use up to 10 years of data
 - No statistical test
- Missouri
 - Running average, minimum of 3 years
 - No statistical test
- Kansas
 - Long-term average
 - Minimum 4 samples over a 12-year period
 - No statistical test







Warning, small datasets limit power of statistical test to reject null hypotheses

CHLa examples from 2016 IR Database

Station	n	-90% c.i.	-70% c.i.	Geom. Mean	+70% c.i.	+90% c.i.
YAD181G (Bunch Lake)	9	11	16	17	29	33
CPF055C2 (Jordan Lake)	13	31	35	36	45	48
208458892 (Lake Mattamuskeet)	13	67	75	76	99	107
YAD181E (McCrary Lake)	9	14	17	19	29	33

Final Thoughts on Pros & Cons of 1-in-3 vs. Multi-Year Statistical Test

1-in-3

- Allowance is to base attainment on second highest year.
- Magnitude set to be attained in ≥80% years.
- Requires 2-3 years of data

Multi-Year Average

- Allowance is based on averaging out of low and high years.
- Magnitude set to be attained in ≥50% of years
- Requires ? years of data
- Option to include explicit statistical test.
- Stat. test might make it harder to list or delist with small datasets.

Extra Slides



Results of different approaches for calculating CHLa in High Rock Lake

Results of Theoretical HRL Assessment Using 2008, 2009, & 2011 Data

[The controlling value is either the second highest CHLa value (if means calculated for individual years) or the multi-year average (if means calculated for multiple years).]

Method	Months	Type of Mean	No. of Yrs Averaged	Spatial Averaging	Controlling Value	Controlling Station/Zone
1	Apr-Oct	Geometric	1	No	50.1	YAD152C
2	May-Oct	Geometric	1	No	53.4	YAD152C
3	Apr-Oct	Arithmetic	1	No	53.6	YAD152C
4	May-Oct	Arithmetic	1	No	56.1	YAD152C
5	Apr-Oct	Geometric	3	No	50.4	YAD152C
6	May-Oct	Arithmetic	3	No	56.1	YAD152C
7	Apr-Oct	Geometric	1	Yes	47.3	Transitional
8	May-Oct	Geometric	3	Yes	51.2	Transitional
9	Apr-Oct	Geometric	2	No	53.3	YAD152
10	May-Oct	Arithmetic	2	No	58.0	YAD152

Relation between Mean and 90th Percentile CHLa, 2002 - 2016

50 $R^2 = 0.8597$ 45 Mean CHLa (ug/L) 00 55 05 $R^2 = 0.6383$ 25 20 30 35 40 45 50 55 60 65 70 75 80 90th Percentile CHLa (ug/L) Geometric Mean Arith. Mean Linear (Geometric Mean) Linear (Arith. Mean)

HRL Mainstem, All stations

Text Example Using 2016 IR Data, 25-40 ug/L Tier Approach

Effect	Proportion of Stations
Flip from non-attainment to attainment	2.5%
Flip from non-attainment to requiring narrative evaluation	13.9%
Flip from attainment to requiring narrative evaluation	7.8%
Stays in non-attainment	17.2%
Stays in attainment	58.7%

















From: To:	Petter, Lauren <u>Wrenn, Brian L; Astrid Schnetzer; William Hall; Dr. Katie Martin; Hans Paerl; jdbowen (jdbowen@uncc.edu);</u> <u>Linda Ehrlich; Marcelo Ardon (mlardons@ncsu.edu); Martin Lebo; Michael O"Driscoll; Nathan Hall; Petter, Lauren;</u> <u>Clifton Bell; Deanna Osmond (deanna osmond@ncsu.edu)</u>
Cc:	Banihani, Qais; Behm, Pamela; Brower, Connie; Deamer, Nora; Fensin, Elizabeth; Hawhee, Jim; Hill, Tammy; Hong, Bongghi; Lin, Jing; Manning, Jeff; Stevenson, Leigh E; Templeton, Mike; Ventaloro, Christopher
Subject:	[External] Follow up to 9/24 SAC Meeting
Date:	Tuesday, October 16, 2018 10:10:22 AM
Attachments:	Scenario Pathways Petter 101618.docx

CAUTION: External email. Do not click links or open attachments unless verified. Send all suspicious email as an attachment to <u>Report Spam.</u>

Hey everyone,

I have attached the scenarios compilation I offered to prepare at our last meeting. I highlight questions to be answered and cautions to take to ensure we are getting what we need for criteria derivation out of this exercise. Before reading the attachment, the following summaries best articulate my reasons for thinking of magnitude along with duration and frequency instead of doing it in two parts, because doing it in two parts may lead to criteria derivation that is not appropriately based on science.

Duration and Frequency First Approach

Taking a random scenario from the attachment, Scenario 1MULTI, we end up with a format for a criterion that looks like:

"Instantaneous Chlorophyll value of X shall not be exceeded more than 10%, in some TBD unit of time \leq 1 calendar year, *with/without confidence interval*, no more than once in a TBD year period."

Does the pre-selection of a given scenario influence the magnitude step or is it disconnected? If it is disconnected then do we plan to independently develop a screening range and set of metrics for use in a narrative procedure? Typically such thresholds are developed from some "higher water quality" starting point, which I think we would agree is not the case here, and at a less site specific scale. If pre-selection does influence magnitude, we will have to be careful we don't fall into a derivation process where the magnitude development is influenced by what is attainable and/or consider affordability.

Magnitude, Duration, and Frequency Together Approach

Alternatively, by not using a scenario format and since the group has already discussed the possible criteria derivation methods for chlorophyll, the group could focus on documenting which method(s) could be considered for all lakes in NC. This would include methods like reference condition, modified/least impacted conditions, modeling to meet known quantitative endpoints, literature (including comparable state examples or risk based information tied to toxins), etc. After documenting the defensibility of applying any of these methods to lakes in North Carolina (through showing comparable geologic make up or other quantifiable similarities), the group could select the most applicable method outcome (retaining that outcome's specific magnitude, duration, and frequency) to ensure the designated uses are met. This ensures the resulting recommendation is consistent with how the criteria was derived and then we can use the conversations to date

regarding the components of criteria to add commentary on how future assessments of the criterion can be handled.

Lauren Petter, Environmental Scientist Water Quality Standards Section Water Protection Division Phone: (404) 562-9272

Compilation and Write Up for Possible Flow Chart Scenario Formats

Many of the questions posed in the flowchart diagram and companion write up are focused on how the criteria will be assessed. While these will need to be answered at some point, first we have to re-focus our attention on deriving protective criterion values.

I have taken the scenarios that I believe result from working through the various paths of the flowchart and provided them below to directly focus on developing a strong scientific justification for the criteria. Along those lines, I have also added questions under the main two scenario types that should be our focus for those who want to discuss duration and frequency in the absence of magnitude.

This exercise may create more work than starting by selecting a protective magnitude, duration, and frequency at the same time, but I wanted to be responsive to all of the members of the group. The questions focus the discussion on why certain duration and frequency, as well as statistical considerations, are appropriate and scientifically defensible, rather than on which one(s) individuals prefer. When we select a magnitude in the end, it can't be contradictory to how it was derived and we can pull the support we have created in this exercise to bolster our rationale for the criterion eventually selected by the SAC. Additionally, any assessment related questions in the flow chart will benefit the state as they consider moving forward with a given criteria recommendation.

SCENARIO 1: Exceedance Based Criterion [per left pathway at 0A1]. This first "scenario" type results in the 10+ scenarios below, with example criterion text.

Scenario 1A1i: Exceedance Based Criterion that uses instantaneous chlorophyll a concentration with 0% exceedances and no confidence interval

"Instantaneous Chlorophyll value of X shall not be exceeded"

Scenario 1A1a: Exceedance Based Criterion that uses average chlorophyll a concentration with 0% exceedances and no confidence interval

"Averaged Chlorophyll value of X shall not be exceeded"

Scenario 1A2i: Exceedance Based Criterion that uses instantaneous chlorophyll a concentration with 0% exceedances and a TBD confidence interval

"Instantaneous Chlorophyll value of X shall not be exceeded, with a TBD confidence interval"

Scenario 1A2a: Exceedance Based Criterion that uses average chlorophyll a concentration with 0% exceedances and a TBD confidence interval

"Averaged Chlorophyll value of X shall not be exceeded, with a TBD confidence interval"

Scenario 1B1i: Exceedance Based Criterion that uses instantaneous chlorophyll a concentration with 10% exceedances and no confidence interval

"Instantaneous Chlorophyll value of X shall not be exceeded more than 10% of the time, in some TBD unit of time \leq 1 calendar year"

Scenario 1B1a: Exceedance Based Criterion that uses average chlorophyll a concentration with 10% exceedances and no confidence interval

"Averaged Chlorophyll value of X shall not be exceeded more than 10% of the time, in some TBD unit of time \leq 1 calendar year"

Scenario 1B2i: Exceedance Based Criterion that uses instantaneous chlorophyll a concentration with 10% exceedances and a TBD confidence interval

"Instantaneous Chlorophyll value of X shall not be exceeded more than 10% of the time, with a TBD confidence interval, in some TBD unit of time \leq 1 calendar year"

Scenario 1B2a: Exceedance Based Criterion that uses average chlorophyll a concentration with 10% exceedances and a TBD confidence interval

"Averaged Chlorophyll value of X shall not be exceeded more than 10% of the time, with a TBD confidence interval, in some TBD unit of time \leq 1 calendar year"

Scenario 1C1i: Exceedance Based Criterion that uses instantaneous chlorophyll a concentration with TBD% exceedances and no confidence interval

"Instantaneous Chlorophyll value of X shall not be exceeded more than TBD% of the time, in some TBD unit of time \leq 1 calendar year"

Scenario 1C1a: Exceedance Based Criterion that uses average chlorophyll a concentration with TBD% exceedances and no confidence interval

"Averaged Chlorophyll value of X shall not be exceeded more than TBD% of the time, in some TBD unit of time < 1 calendar year"

Scenario 1C2i: Exceedance Based Criterion that uses instantaneous chlorophyll a concentration with TBD% exceedances and a TBD confidence interval

"Instantaneous Chlorophyll value of X shall not be exceeded more than TBD% of the time, with a TBD confidence interval, in some TBD unit of time ≤ 1 calendar year"

Scenario 1C2a: Exceedance Based Criterion that uses average chlorophyll a concentration with TBD% exceedances and a TBD confidence interval

"Averaged Chlorophyll value of X shall not be exceeded more than TBD% of the time, with a TBD confidence interval, in some TBD unit of time \leq 1 calendar year"

Scenario 1MULTI: Exceedance Based Criterion that uses instantaneous chlorophyll a concentration with 1A1, 1B1, or 1C1 % exceedances, with/without confidence interval, but including a multi year consideration as part of the criterion. For simplicity "with/without confidence interval" is indicated for non-zero rate examples.

"Instantaneous Chlorophyll value of X shall not be exceeded, in some TBD unit of time \leq 1 calendar year, no more than once in a TBD year period." [Included to keep parallel structure of examples, but should be equivalent to 1A1i and 1A2i, right?]

"Instantaneous Chlorophyll value of X shall not be exceeded more than 10%, in some TBD unit of time \leq 1 calendar year, *with/without confidence interval*, no more than once in a TBD year period."

"Instantaneous Chlorophyll value of X shall not be exceeded more than TBD%, in some TBD unit of time \leq 1 calendar year, *with/without confidence interval*, no more than once in a TBD year period."

Scenario 1MULTIa: Exceedance Based Criterion that uses average chlorophyll a concentration with 1A1, 1B1, or 1C1 % exceedances, with/without confidence interval, but including a multi year consideration as part of the criterion. For simplicity "with/without confidence interval" is indicated for non-zero rate examples.

"Averaged Chlorophyll value of X shall not be exceeded, in some TBD unit of time \leq 1 calendar year, no more than once in a TBD year period." [Included to keep parallel structure of examples, but should be equivalent to 1A1a and 1A2a, right?]

"Averaged Chlorophyll value of X shall not be exceeded more than 10%, in some TBD unit of time \leq 1 calendar year, *with/without confidence interval*, no more than once in a TBD year period."

"Averaged Chlorophyll value of X shall not be exceeded more than TBD%, in some TBD unit of time ≤ 1 calendar year, with/without confidence interval, no more than once in a TBD year period."

Scenario 1NOMULTI: See language used in scenarios above 1MULTI, subpart i and a.

Details to Capture/Questions to Address for Scenario 1:

Exceedance Rate of 0%, 10%, TBD % - Explain the reasoning and defensibility for choosing a percentage rate. This should include statistical and scientific bases.

Confidence Interval – Explain the reasoning and defensibility of choosing a confidence interval, including the interval value chosen. This should include statistical and scientific bases. Will the confidence level be built into the criterion or used in assessments only?

Multiple Year Usage - How many years can be considered to reflect recent conditions? What is the scientific basis for allowing multiple years to be considered for an exceedance based criterion, that is presumably reflective of a shorter term effect? Data sufficiency is always a consideration for implementing programs, however, there needs to be a link between criteria derivation and an assessment that measures that in an appropriate way.

Over manipulation of data – In order to address concerns with too many "allowances," resulting from a combination of averaging decisions and statistics, explain how the above decisions do not result in too many manipulations of the data. Care should be given to not create a compounding effect that impacts the

protectiveness desired from the criterion. Also, answers should be clear when the consideration is for an assessment reason versus protection related to criteria derivation reason.

Excerpts heard from the SAC members during 9/24 meeting:

- Advantage of current exceedance format does deal with high values "meaning something" but second piece of how it is applied includes statistical recognition for variability
 - Or does the tendency for high chlorophyll values (as opposed to outlier "highs" in other systems) make this concern matter less?
- CB noted exceedance basis makes more sense when chlorophyll a is tied to short term effects
 - If the mean is lowered this would also bring down the tail (of the distribution)

SCENARIO 2: Average Based Criterion [per right pathway at 0A1]. This second "scenario" type results in the 6 scenarios below, with example criterion text.

Scenario 2GWCL: Geometric mean value with confidence level

"Geometric mean chlorophyll of X shall not be exceeded, with a TBD confidence level"

Scenario 2AWCL: Arithmetic mean value with confidence level

"Arithmetic mean chlorophyll of X shall not be exceeded, with a TBD confidence level"

Scenario 2GNTE: Geometric mean value not to be exceeded

"Geometric mean chlorophyll of X shall not be exceeded"

Scenario 2ANTE: Arithmetic mean value not to be exceeded

"Arithmetic mean chlorophyll of X shall not be exceeded"

Scenario 2GMULTI: Geometric mean value not to be exceeded at some frequency over X years

"Geometric mean chlorophyll of X shall not be exceeded greater than 1in3/2in5/TBD years"

Scenario 2AMULTI: Arithmetic mean value not to be exceeded at some frequency over X years

"Arithmetic mean chlorophyll of X shall not be exceeded greater than 1in3/2in5/TBD years"

Details to Capture/Questions to Address for Scenario 2:

Averaging Format – What are the pros/cons for each averaging type (discussing pros and cons gets more at the conversational aspect of why it sounds like the appropriate choice)? Do certain distributions of chlorophyll a data inform whether a geometric mean or arithmetic average should be used? Are there other bases for determining which type of averaging format should be used? The answers should include statistical and scientific bases.

Confidence Interval – Explain the reasoning and defensibility of choosing a confidence interval, including the interval value chosen. This should include statistical and scientific bases. Will the confidence level be built into the criterion or used in assessments only?

Multiple Year Usage - How many years can be considered to reflect recent conditions? What is the scientific basis for allowing multiple years to be considered? What is the basis for the selected frequency? Data sufficiency

is always a consideration for implementing programs, however, there needs to be a link between criteria derivation and an assessment that measures that in an appropriate way.

Over manipulation of data – In order to address concerns with too many "allowances," resulting from a combination of averaging decisions and statistics, explain how the above decisions do not result in too many manipulations of the data. Care should be given to not create a compounding effect that impacts the protectiveness desired from the criterion. Also, answers should be clear when the consideration is for an assessment reason versus protection related to criteria derivation reason.

Excerpts heard from the SAC members during 9/24 meeting:

- Averaging is preferable because experience shows there are transient/anomalous spikes and these are avoided with averaging (Linda)
- Average should also consider peaks because higher values will bring up average (ML)
- Seasonal average is most important and to shelve the concern with extremes (CB)
- Discussion that models are better at averages than instantaneous values (CB, NH, HP)
- Geomeans help with outliers that are not manageable through correctable actions (NH/HP)
- Geomeans are useful if the central tendency is deemed more important (Linda)
- Arithmetic means are good if statistical confidence is not an option
 - Alternatively, geo means are the preference is statistical test/confidence is considered (JB)
- If data is log normal, then go with Geomean (JB)

Magnitude Thoughts on Flowchart:

Following up on what Jim and I alluded to at the 9/24 meeting, and building on what I said in my email, I still believe there is the potential for our combined document to include information on statewide lake information, along with the site specific HRL information. Given that we have spent over 3 years talking about lots of lake information I don't think it would be a far reach to add in some recommendations for lakes other than HRL. It would represent a recommendation just like that of HRL and could include additional commentary to reflect where additional study is needed, if appropriate. There is a greater benefit to developing a process that is more inclusive to a variety of lake water qualities, as well as simultaneously considering magnitude, duration, and frequency, as opposed to a process that too narrowly focuses on HRL as a sole example. The inclusion of statewide information also goes along with those members of the group who are more averse to risk in terms of ensuring all designated uses are protected. A state always has the ability to adopt criteria on a more localized scale, but adding in a statewide approach provides some additional output to show for all of our efforts to date.