



United States
Environmental
Protection Agency

Office of Water
Mail Code 4304T

EPA 822-R-21-004
August 2021

**Response to Public Comments on US EPA's Draft
Ambient Water Quality Criteria
Recommendations for Lakes and Reservoirs of
the Conterminous United States: Information
Supporting the Development of Numeric
Nutrient Criteria**

Contents

Category 1 General Comments	5
Category 1.1 General Comments: Organization	5
Category 1.2 General Comments: Broader Scientific Input	5
Category 1.3 General Comments: Variability	6
Category 1.4 General Comments: Lakes Versus Reservoirs.....	8
Category 1.5 General Comments: Dual Nutrient Strategy.....	9
Category 1.6 General Comments: General Approach.....	10
Category 1.7 General Comments: Range of Allowable Values in Sliders	13
Category 1.8 General Comments: Clean Water Act Section 304(a) Regulation.....	13
Category 1.9 General Comments: General Compliments.....	15
Category 2 Problem Formulation	17
Category 2.1 Problem Formulation: Management Goals	17
Category 2.2 Problem Formulation: Assessment Endpoints in General	17
Category 2.3 Problem Formulation: Assessment Endpoints in General	19
Category 2.4 Problem Formulation: Aquatic Life Use Endpoints.....	20
Category 2.5 Problem Formulation: Assessment Endpoints for Recreational Criteria	22
Category 3 Analysis	23
Category 3.1 Analysis: National Lakes Assessment Data	23
Category 3.2 Analysis: Depth in the NLA	24
Category 3.3 Analysis: Other Data Sources	24
Category 3.4 Analysis: Field Collection Methodologies.....	25
Category 3.5 Analysis: Zooplankton Model – General Comments.....	26

Category 3.6 Analysis: Stressor-Response Approach – General	28
Category 3.7 Analysis: Zooplankton Model – Effect of Depth and Other Covariates.....	29
Category 3.8 Analysis: Zooplankton Model – Data Availability.....	30
Category 3.9 Analysis: Zooplankton Model – Threshold Selection	30
Category 3.10 Analysis: Deepwater Hypoxia – General Comments	31
Category 3.11 Analysis: Zooplankton Model – Applicability to Other Systems	34
Category 3.12 Analysis: Deepwater Hypoxia – Other Measures	34
Category 3.13 Analysis: Deepwater Hypoxia – Applicability.....	35
Category 3.14 Analysis: Deepwater Hypoxia – Temperature Thresholds	36
Category 3.15 Analysis: Microcystin Model – General	37
Category 3.16 Analysis: Deepwater Hypoxia – Stratification.....	37
Category 3.17 Analysis: Microcystin Model – Other Cyanotoxins	39
Category 3.18 Analysis: Microcystin Model – Detection Limit	39
Category 3.19 Analysis: Microcystin Model – Threshold Selection	40
Category 3.20 Analysis: Microcystin Model – Other Covariates and Measures	41
Category 3.21 Analysis: Phosphorus, Chlorophyll Model – General Comments.....	43
Category 3.22 Analysis: Phosphorus, Chlorophyll Model – Other Covariates.....	44
Category 3.23 Analysis: Nitrogen, Chlorophyll Model – General Comments.....	44
Category 3.24 Analysis: Model Statistics	45
Category 4 Characterization.....	48
Category 4.1 Characterization: Incorporating State Data – Need More Guidance	48
Category 4.2 Characterization: Incorporating State Data – General Comments	49
Category 4.3 Characterization: Limitations and Assumptions	49

Category 4.4 Characterization: Duration and Frequency	50
Category 5 Implementation	51
Category 5.1 Implementation: Management, Most Sensitive Use, and Other Issues	51
Category 6 Supporting State Criteria.....	53
Category 6.1 Supporting State Criteria: Candidate Criteria.....	53
Category 6.2 Supporting State Criteria: Training.....	55
Category 6.3 Supporting State Criteria: R Scripts	56
Category 6.4 Supporting State Criteria: Supporting Information for the Document	57
Category 6.5 Supporting State Criteria: Supporting Information for the R Shiny Apps	58
Category 6.6 Supporting State Criteria: Output Magnitudes.....	59
Category 6.7 Supporting State Criteria: Applicability Given Within-Lake Variability	60
Category 6.8 Supporting State Criteria: Applicability – National to Site-Specific.....	60
Category 6.9 Supporting State Criteria: Applicability – Unsampled Lake Types	62
Category 6.10 Supporting State Criteria: Constraints (Data).....	62
Category 6.11 Supporting State Criteria: Alternative Methods.....	63
Category 6.12 Supporting State Criteria: Derivation Efforts – Sampling Designs	63
Category 6.13 Supporting State Criteria: Combined Criteria.....	64
Category 6.14 Supporting State Criteria: Existing Criteria.....	66
Category 7 Editorial Comments.....	67
Category 7.1 Editorial Comments: General	67

Category 1 General Comments

Category 1.1 General Comments: Organization

Commenters: Wyoming Department of Environmental Quality.

Comment synopsis: The U.S. Environmental Protection Agency (EPA) should consider organizing the criterion document in a way that makes it easier to evaluate each model and its output independently.

Response: In keeping with the risk assessment guidelines used to develop the recommendations (U.S. EPA 1998), EPA provided detailed characterizations for each stressor-response model (see Chapter 3, Analysis, of the criterion document). For example, EPA documented for each model the source data that were used to populate them, provided the model parameters and equations, developed individual conceptual models as a way to visualize the configuration of the model's parameters, and provided the R code upon which each model was simulated. In addition, EPA developed visualizations for each stressor-response model using R Shiny apps as an added measure of clarity and usability for model users. For example, the visualizations facilitate conducting model simulations, making it easier to compare an individual model's outputs based on different parameter inputs, as well as facilitating inter-model comparisons.

Category 1.2 General Comments: Broader Scientific Input

Commenters: Wyoming Department of Environmental Quality; G. Hess, R.T. Angelo, and J. DeLashmit; Footprints in the Water, LLC; Coalition of Greater Minnesota Cities; and Water Environment Association of Texas and the Texas Association of Clean Water Agencies.

Comment synopsis: EPA received comments expressing the need for broader scientific input on the draft criteria recommendations. Some comments stated that EPA should have consulted states and developed the models in collaboration with state scientific experts prior to publishing the draft criteria recommendations. One comment suggested that EPA conduct additional external scientific peer reviews of the draft criteria recommendations.

Response: Both scientific rigor and state partnerships are very important to EPA. EPA's commitment to ensuring the scientific rigor with these Clean Water Act (CWA) Section 304(a) recommendations is reflected in the independent, scientific peer reviews of the manuscripts that describe the models in greater detail. This was a scientific oversight process that occurred over a 6-year period and distributed across five different scientific journals. The result was the publication of seven scientific journal articles in the open, peer-reviewed literature—articles that EPA made available to the public as part of its request for scientific views. In addition, EPA conducted an external, independent scientific peer review of its draft recommendations,

soliciting the advice from four national subject matter experts and revising its draft recommendations in response to their comments. As an added measure of transparency, EPA made the peer review charge questions, the peer review comments, and its responses to those comments available to the public as part of its request for scientific views on the draft criteria document.

EPA also conducted extensive outreach to states prior to the publication of the draft recommendations, engaging them directly over several years to garner valuable scientific and regulatory feedback. EPA conducted its outreach across a variety of outlets, including hosting technical webinars with state water quality criteria staff to discuss the draft models, conducting national level briefings for state water quality and drinking water programs (Association of Clean Water Agencies and Association of State Drinking Water Agencies), and providing electronic reprints of the scientific journal articles as they were published. In 2018, EPA also convened three different technical workshops, bringing together state and tribal water quality staff to discuss the latest science and increase the technical capacity for developing nutrient criteria. During the workshops, EPA discussed the draft lake models with state and tribal scientists, engaging them in technical discussions and benefiting from their advice and experience.

EPA also considers the recommendations a reflection of the cooperative federalism envisioned under the CWA. For example, all the water quality data EPA used to construct the stressor-response models are derived from two National Lakes Assessment (NLA) surveys (2007 and 2012), a national survey program (National Aquatic Resource Surveys) conducted as a coordinated partnership between EPA, and state water quality monitoring programs. Moreover, EPA engaged several states while developing the draft recommendations, either through their contribution of state water quality data or through their participation as a case study for the draft criteria recommendations (see Appendix A, criterion document).

Overall, EPA views its national recommendations to be a scientifically defensible foundation upon which states and authorized tribes can develop and adopt numeric nutrient criteria. Furthermore, EPA is committed to ensuring scientific rigor in the development of lake and reservoir numeric nutrient criteria by working in partnership with states and authorized tribes through its technical outreach and support program as they work through their water quality standards adoption process (N-STEPS, <https://www.epa.gov/nutrient-policy-data/n-steps>).

Category 1.3 General Comments: Variability

Commenters: Massachusetts Department of Environmental Protection; R.T. Angelo and J. DeLashmit; Clean Water Action et al.; Iowa League of Cities; City of Springfield, Missouri; Metropolitan St. Louis Sewer District; American Fisheries Society et al.; Water Quality, National Council for Air and Stream Improvement, Inc.; NEIWPC; Florida Department of Environmental

Protection; Wasatch Front Water Quality Council; Footprints in the Water, LLC; Kansas Department of Health and Environment; New Jersey Department of Environmental Protection; and Maryland Department of the Environment.

Comment synopsis: EPA received several comments on the subject of variability. Some comments asked that EPA explicitly consider variability in the context of productivity, or trophic state, with lakes exhibiting intra- and inter-lake variations in trophic state over space and time. Other comments focused on variability in a statistical sense, pointing out confounding factors and propagated model error as potential sources of variability in the models, or how the models generate a wide range of chlorophyll (Chl) *a* and nutrient concentration outputs with large credible intervals. Other comments requested that EPA take additional into account measures to either constrain model variability (e.g., through additional classification analyses) or more fully disclose the degree of variability in the models (e.g., sensitivity analysis and model validation). EPA also received comments suggesting that it consult and consider integrating into its models the large amount of lake-specific water quality data collected by states. Others suggested that by developing regional-scale models, variability in EPA's models could be improved.

EPA also received comments expressing concern that it did not consider lake attributes that can affect the stressor-response relationship between nutrients and the assessment endpoint it selected. Commenters suggested EPA revise its models by considering and accounting for the following attributes: color, flushing rates, turbidity, area, depth, ecoregion, water residence time, intra- and inter-annual thermal stratification, naturally high nutrient concentrations, and climate change.

Response: EPA recognizes that lakes exhibit natural variability in the water quality stressors and responses that it modeled in its recommendations. EPA also understands that the presence of certain co-occurring environmental factors, biotic and abiotic, may explain that variability and affect the accuracy of the relationships it modeled. EPA identified these factors, or covariates, in its recommendations, depicting them in conceptual models (e.g., see Figure 1 in the criterion document) and characterizing them in detail for a particular model (e.g., see Figure 23 in the criterion document and the associated text). For comments on specific covariates that might affect the individual models, see responses with the following titles: *Zooplankton Model – Effect of Depth and Other Covariates*; *Microcystin Model – Other Covariates and Measures*; and *Phosphorus, Chlorophyll Model – Other Covariates*. Residual variability is inherent to any environmental model; for the criterion models, EPA identified the primary factors that accounted for variability in the responses. These factors were included in the models. For discussions of response variability for particular models see responses with the following titles: *Phosphorus, Chlorophyll Model – General Comments* and *Nitrogen, Chlorophyll Model – General Comments*. Some variability can also be attributed to sampling and temporal variability, and

methods for accounting for these sources of variability are discussed in Appendix D of the document.

EPA will consult with states and authorized tribes on a case-by-case basis, working under its N-STEPS technical support program, to integrate additional water quality data in developing models that reflect regional and state level conditions. As discussed in the criterion document, EPA constructed a modeling framework that accommodates additional data, past or future, which is coincident in time and space with the models' underlying NLA data or is temporally and spatially complementary to the NLA data (e.g., high frequency data, non-growing season data).

Category 1.4 General Comments: Lakes Versus Reservoirs

Commenters: Upper Neuse River Basin Association; City of Springfield, Missouri; Metropolitan St. Louis Sewer District, North Carolina Water Quality Association; National Association of Clean Water Agencies; New Jersey Department of Environmental Protection; and Iowa Department of Natural Resources.

Comment synopsis: EPA received comments expressing concerns that the lake criteria recommendations do not differentiate between lakes and reservoirs. Commenters believe different models for nutrient criteria development should be used for each. Commenters pointed out that reservoirs typically have larger watershed sizes, more active physical circulation, more variable morphometry (e.g., depth, area), actively managed water elevation, and more variable water residence time as compared to natural lakes. Based on these differences and their potential to affect the stressor-response models that EPA developed, commenters suggested that EPA develop different models to guide and inform the development of numeric nutrient criteria for reservoirs.

Response: EPA agrees with commenters that the water quality of lakes can differ from the water quality in reservoirs, which may stem from actively managed water elevation. EPA also recognizes that water elevation in natural lakes may also differ in a functionally equivalent way due to natural differences in regional precipitation or other natural lake source water factors (e.g., seasonal snowmelt, groundwater inputs).

EPA refrained from any a priori distinctions between lakes and reservoirs (or between lakes of any kind) so that it could objectively examine stressor-response relationships through a data-driven, analytical process. First, EPA selected water quality data that were generated using a consistent methodology and were spatially representative of the lakes and reservoirs across the conterminous United States (the 2007 and 2012 NLA surveys sampled both lakes and reservoirs, see [National Lakes Assessment Design Documents](#)). Second, the unit of analysis for the total nitrogen (TN), total phosphorus (TP), zooplankton, and microcystin models was the water sample, in that relationships between Chl *a* and the response variables were estimated

within a water sample. Because of this formulation, the effect of factors such as retention time do not exert a strong effect on the model because their effects are already reflected in the observed measurements. Active water management can potentially affect the stratification status of a reservoir, and EPA recommends that these issues be considered on a site-specific basis when examining the hypoxia endpoint.

Category 1.5 General Comments: Dual Nutrient Strategy

Commenters: Ecosystem Consulting Service, Inc; Iowa League of Cities; City of Springfield, Missouri; Metropolitan St. Louis Sewer District; Minnesota Environmental Science and Economic Review Board; The Ohio Manufacturers' Association; and Coalition of Greater Minnesota Cities.

Comment synopsis: EPA received a few comments that dispute its dual nitrogen and phosphorus approach in modeling nutrient stressors and responses, deriving candidate criteria for both nutrient stressors, and in recommending that nitrogen and phosphorus criteria apply independently. Commenters urged EPA to consider a single nutrient stressor approach, one that reflects the limiting nutrient stressor at the time and place coincident with the biological response being observed.

Response: EPA disagrees with the comments that suggest EPA's recommendations should depart from a dual nitrogen and phosphorus water quality criteria approach. EPA refers commenters to its fact sheet, [*Preventing Eutrophication: Scientific Support for Dual Nutrient Criteria*](#) (2015), as its scientific basis for recommending that states adopt numeric nitrogen and phosphorus water quality criteria as regulatory measures to *prevent* water quality degradation and *protect* designated uses. EPA points commenters to the evidentiary record in the fact sheet that nutrient limitation of algal growth by nitrogen or phosphorus is not fixed in time or space, but rather it can vary between nitrogen and phosphorus; algal growth can also be co-limited by both nitrogen and phosphorus as a more widespread phenomenon. EPA also points to the risks that single nutrient prevention and control in upstream waters may pose in degrading the water quality downstream with the uncontrolled nutrient pollutant, which is an unintended outcome that could have multijurisdictional consequences.

In light of the scientific consensus that support EPA's recommendations for dual nitrogen and phosphorus water quality criteria, EPA does not dispute state management actions that target a single nutrient pollutant, either nitrogen or phosphorus, as a step in *remediating nutrient-impaired waters* and bringing those waters into compliance with state water quality standards. In some cases, single nutrient pollutant control may offer a critical path in reversing the effects of nutrient pollution, providing a timelier and environmentally effective measure within the broader remediation strategy. EPA's recommendations do not supplant existing single nutrient pollution remediation efforts undertaken by states (e.g., single nutrient total maximum daily

load [TMDL] for a point source, nonpoint source controls targeting phosphorus retention), nor do they preclude future single nutrient pollution remediation efforts if such conditions are scientifically justified.

Category 1.6 General Comments: General Approach

Commenters: Ecosystem Consulting Service, Inc; R.T. Angelo and J. DeLashmit; Footprints in the Water, LLC; National Wildlife Federation; American Farm Bureau Federation et al.; Iowa League of Cities; Central Valley Clean Water Association; National Association of Clean Water Agencies; New Jersey Department of Environmental Protection; Florida Department of Environmental Protection; Iowa Department of Natural Resources; Public Employees for Environmental Responsibility; Maryland Department of the Environment; The Ohio Manufacturers' Association; Counsel, Mississippi River Collaborative et al.; and California Water Boards.

Comment synopsis: EPA received comments questioning the statistical modeling approach it took to develop its recommendations. Commenters suggested that the approach is oversimplified conceptually without sufficient support for the models' assumptions, while others commented the approach is overly complex statistically. Some commenters stated that the statistical approach precludes comparisons of the models' accuracy to other technical approaches, while other commenters felt EPA should have provided an explanation as to why an empirical stressor-response approach was pursued in lieu of alternative technical approaches. Some comments pointed to the historical literature on Great Lakes water quality research as one way EPA can further support its recommendations.

Some comments were critical of EPA's selection of assessment endpoints for aquatic life, recreation, and human health, asserting that the protection of water resources conferred by nutrient criteria should be directed at structural and functional ecosystem services associated with resiliency and sustainability, rather than controlling individual nutrient pollutants to protect ecosystem health.

Other comments EPA received focused on the geospatial and temporal scales of the models. Some comments asserted that the constrained geographic and temporal representation in the NLA data limit EPA's models in terms of accurately quantifying the stressor-response relationships and characterizing the effects. Other comments stated that the models ignore geographic differences in lake water quality, including site-specific nutrient limitation conditions, large-scale watershed factors such as land use and land cover, or confounding factors within lakes that can affect the stressor-response relationship. Related to the geographic issues, some commenters expressed concern about how EPA classified lakes. As a remedy, one commenter suggested that EPA develop a hierarchy of models based on the 14 Level III-Aggregate ecoregions that occur within each state.

One feature of EPA's recommendations is the ability of states to incorporate state-specific data into the models when developing numeric nutrient criteria. However, some commenters expressed concern that such an accommodation may not be utilized if states do not possess the requisite water quality data. In a broader sense, some commenters contended that EPA's recommendations as a whole may not be implemented, regardless of the flexibilities inherent in them, because nutrient pollution management, in their view, should be holistic and based on watershed management of all nutrient pollution sources rather than narrowly managed through costly water quality-based controls.

Finally, some commenters asserted that EPA's models cannot be used to infer cause-effect relationships between nutrient stressors and responses in the same way that EPA infers cause-effect relationships with toxic pollutants, citing differences in the modes of action and pathways of effects between toxic and nutrient pollutants as the basis for this view. Another comment claimed that EPA has a statutory and regulatory requirement to establish cause-effect links (rather than correlations) for nutrient pollutants, citing CWA sections 303 and 304, as well as the regulations at 40 C.F.R. 131.11.

Response: The implementing regulations of the CWA establish that state water quality criteria must be based on science and be scientifically defensible (40 C.F.R. 131.11(a)). These expectations apply to EPA as well, including in the exercising of its discretionary authority under CWA Section 303(c)(4), which authorizes EPA to promulgate water quality standards under certain conditions, and CWA Section 304(a), which authorizes EPA to publish, and "from time to time thereafter revise," criteria and information that support the achievement of the goals expressed in CWA Section 101(a).

Consistent with this requirement of scientific defensibility, as well as the specific requirements under CWA Section 304(a) that describe the different types of scientific information, EPA published draft recommendations that inform the protection of aquatic life, recreation, and human health in lakes and reservoirs from nutrient pollution and its effects. Throughout the drafting process, EPA relied upon the judgment of its scientists to identify high-quality data and the appropriate modeling approach through which accurate, predictive relationships could be produced. EPA identified a highly credible, national-scale source of water quality data from the NLA program, offering a unique opportunity to explore and develop predictive models that could serve as recommendations under CWA Section 304(a). EPA considered a variety of modeling approaches that best fit the scales of the data and offered the most rigorous approach from which it could reliably draw inferences about nutrient pollutants and specific responses to them. EPA concluded that an empirical, hierarchical Bayesian modeling approach was the most rigorous approach for analyzing the water quality data. EPA provided detailed descriptions of the data and the modeling approach in its criterion document and the associated peer-reviewed journal articles, which EPA provided to the public during the

comment period. In addition, EPA provided to the public all of the data and the model codes that were used to develop the draft recommendations.

EPA also relied upon the scientific literature to support its decisions throughout the drafting process. In its criterion document and the associated peer-reviewed journal articles, EPA documented the scientific basis for the following core elements of the draft recommendations: the modeling approach, the conceptual models, the configuration of the models, and the interpretations of the model outputs. Another important element of EPA's draft recommendation was the evidentiary basis establishing the relationships between nutrient pollutants and their effects on state designated uses, an element that CWA Section 304(a) expresses should be associated with 304(a) criteria and information. Again, EPA pointed to a long record in the scientific literature in its criterion document and the associated peer-reviewed journal articles that established the cause-effect relationship between nutrient pollutants and their myriad effects. From this record, EPA was able to identify and select sensitive assessment endpoints that it could model using the NLA data to develop numeric criteria that are protective of the three designated uses. EPA characterized these results for each of the models it developed and provided model visualizations that aid in interpreting the model results.

EPA recognized that the data and the technical approaches used to model them would generate uncertainty around the predictions, which is a feature common to all data and modeling techniques. Unlike deterministic numerical models, however, posterior model evaluations such as sensitivity analysis are not applicable to hierarchical Bayesian models because the model simulations draw from a distribution of data rather than from single values, which can greatly affect model outputs if they are varied in a deterministic model. Nonetheless, EPA documented the sources of uncertainty in its models in its criterion document and the associated peer-reviewed journal articles. EPA went further in communicating model uncertainty by including it as a model parameter that can be specified in its model visualizations. In this way, EPA provided flexibility to states and authorized tribes with regard to this important risk management decision that states and authorized tribes can consider and specify when deriving numeric nutrient criteria.

In choosing to model data from the NLA, EPA also understood that the resulting model predictions could potentially be made more accurate if additional data were included in the models. To accommodate this, EPA configured the model codes so that more data, including site-specific data, could be simulated. EPA illustrated this model flexibility by conducting a case study, which is described in Appendix A of the criterion document. By crafting flexibility in the models to simulate additional data, EPA did not intend to create additional monitoring and data generation requirements for states and authorized tribes. Rather, EPA intended to create capacity in the models to generate more accurate relationships if states and authorized tribes chose to investigate them further.

Finally, EPA sought to find regulatory balance in its recommendations, offering constructive information that states and authorized tribes can use to reliably derive numeric nutrient criteria for lakes and reservoirs while at the same time avoiding prescriptive, “one-size fits all” recommendations.

Category 1.7 General Comments: Range of Allowable Values in Sliders

Commenters: Wyoming Department of Environmental Quality; Clean Water Action et al.; and R.T. Angelo and J. DeLashmit.

Comment synopsis: Some commenters suggested that EPA provide more narrow credible intervals in its recommendations as a way of constraining the range of candidate criteria that might be based on the uncertainty boundaries. Other commenters suggested that EPA provide wider credible intervals as a way to increase the representativeness of the nutrient–Chl *a* models (TP and TN-dissolved inorganic nitrogen [DIN]).

Response: EPA used an empirical, hierarchical statistical modeling approach to quantify the stressor-response relationships. As part of that approach, EPA estimated the uncertainty associated with the modeled relationships and expressed it as the credible interval around the mean relationship between the modeled parameters. To further aid understanding of model uncertainty, EPA developed visualizations of the certainty level or credible interval (R Shiny apps), providing a slider that allows different levels of certainty to be selected and explored by users. In a general sense, the range of uncertainty associated with each of the models is fixed, driven by the underlying data and coefficients in the models. Within that range of uncertainty, states can quantitatively specify the uncertainty by selecting different certainty levels (or, credible intervals). In EPA’s view, selecting the credible interval is a risk management decision that can be determined by states and authorized tribes and then can be transparently communicated with their citizens.

Category 1.8 General Comments: Clean Water Act Section 304(a) Regulation

Commenters: G. Hess; Missouri Department of Natural Resources; Wyoming Department of Environmental Quality; Association of State Drinking Water Administrators; Riverkeeper, Inc.; Virginia Association of Municipal Wastewater Agencies; Louisiana Department of Environmental Quality; American Farm Bureau Federation; Center for Biological Diversity; American Water Works Association; Water Environment Association of Texas and the Texas Association of Clean Water Agencies; Central Valley Clean Water Association; Kansas Department of Health and Environment; National Association of Clean Water Agencies; New Jersey Department of Environmental Protection; Florida Department of Environmental Protection; Public Employees for Environmental Responsibility; Tulane Institute on Water Resources Law and Policy; Colorado Department of Public Health and Environment; North

Carolina Department of Environmental Quality; California Water Boards; and Counsel, Mississippi River Collaborative et al.

Comment synopsis: EPA received several comments on CWA Section 304(a) authorities. Comments included requests that EPA clarify the purpose of the recommendations under Section 304(a) and suggestions that EPA provide more detailed reasoning for replacing rather than building upon its previous 304(a) ecoregional-based nutrient criteria for lakes and reservoirs published in 2000 and 2001. Other comments included requests that EPA publish its recommendations as technical guidance under CWA Section 304(a), or as technical tools outside of EPA's CWA authorities, rather than as recommended criteria under Section 304(a). One comment suggested that EPA take an opposite approach and derive numeric values from its models, publishing these specific values as recommended 304(a) numeric nutrient criteria. Some commenters urged EPA not to use its recommendations to replace its previous 304(a) ecoregional-based criteria recommendations for lakes and reservoirs, while others asked that EPA clarify whether they are still scientifically valid.

Some comments stated that the recommendations are non-binding on states and, as a consequence, the recommendations will have little influence in protecting the nation's surface waters from nutrient pollution. Some commenters suggested that EPA imposed an enforceable timeline for numeric nutrient criteria adoption. One comment cited EPA's lack of consultation with other federal agencies in considering threatened and endangered species as part of its recommendations.

As an extension to the comments received regarding Section 304(a) recommendations, some commenters asked that EPA include in its recommendations information that can guide a state in choosing alternative approaches to the recommendations themselves. Some commenters expanded on this topic, requesting that EPA provide specific information on how states can comply with the requirements at 40 C.F.R. 131.11(b)(iii), which calls on states to use "other scientifically defensible methods" in establishing criteria in the event that states decide not to follow the recommendations. Other commenters requested specific information that states can use to comply with 40 C.F.R. 131.20(a), specifically information that justifies a state's discretion not to adopt numeric nutrient criteria based on EPA's Section 304(a) recommendations at the time it submits the results of its triennial review to EPA.

Finally, EPA received comments urging it to create a regulatory structure that supports state adoption of its Section 304(a) recommendations and to provide regulatory support to states that aids the implementation of any numeric nutrient criteria developed and adopted from the recommendations, including regulatory measures that enforce their implementation.

Response: EPA's recommendations were developed under CWA Section 304(a), which calls on EPA to publish, and "from time to time thereafter revise," criteria and information that support

the achievement of the goals expressed in Section 101(a). Consistent with Section 304(a), EPA sought the latest science as the basis for drafting recommendations, drawing upon a well-known and highly credible source of water quality data and using analytical techniques appropriate for modeling and drawing inferences from the data. EPA also understood from 20 years of feedback and engagement with states that empirical connections between nutrient pollutants and their effects would have strong support from states, stakeholders, and the public. These factors, including the plain language of Section 304(a), contributed to EPA's view that its recommendations serve as a revision to and replacement of the ecoregional nutrient criteria it published previously in 12 separate documents under Section 304(a) in 2000 and 2001.

EPA does not consider the technical approach used in its recommendations to conflict in any way with previous scientifically defensible approaches it has recommended under Section 304(a). States may use these criteria recommendations separate from or in conjunction with other scientifically defensible approaches to derive numeric nutrient criteria for lakes and reservoirs. Additionally, states may use EPA's recommendations as an independent, yet complementary line of evidence that corroborates other lines of evidence from which they derive numeric nutrient criteria for lakes and reservoirs.

EPA sought to find regulatory balance in its recommendations, offering constructive information that states can use to reliably derive numeric nutrient criteria for lakes and reservoirs while at the same time avoiding prescriptive, "one-size fits all" recommendations. As such, EPA refrained from recommending specific values from the models. Instead, EPA chose to craft flexibilities in the recommendations in which states can customize the models with state-specific data to generate candidate criteria to reflect the environmental conditions of their lakes and reservoirs. Furthermore, EPA continues to provide technical support to states and authorized tribes for deriving numeric nutrient criteria through its N-STEPS program. By providing model visualizations, EPA sought to provide an operating tool that would facilitate state regulatory decision making and adoption of numeric nutrient criteria. EPA also refrained from offering prescriptions on what constitutes "other scientifically defensible methods" or the conditions under which a state would not use its recommendations, choosing instead to let states articulate the scientific and technical rigor associated with their preferred data and methods.

Category 1.9 General Comments: General Compliments

Commenters: Virginia Department of Environmental Quality; Association of State Drinking Water Administrators; Ecosystem Consulting Service, Inc; Upper Neuse River Basin Association; Tip of the Mitt Watershed Council; Virginia Association of Municipal Wastewater Agencies; Montana Department of Environmental Quality; Alliance for the Great Lakes; North Carolina Water Quality Association; City of Springfield, Missouri; Metropolitan St. Louis Sewer District;

Missouri Department of Natural Resources; Arkansas Department of Energy and Environment; Federal Water Quality Coalition; American Fisheries Society et al.; Iowa Environmental Council; Oregon Department of Environmental Quality; Association of Metropolitan Water Agencies; Water Environment Association of Texas and the Texas Association of Clean Water Agencies; Water Quality, National Council for Air and Stream Improvement Inc.; Idaho Department of Environmental Quality; Wisconsin Department of Natural Resources; Kansas Department of Health and Environment; National Association of Clean Water Agencies; New Jersey Department of Environmental Protection; Florida Department of Environmental Protection; Iowa Department of Natural Resources; NEIWPC; Oregon Lakes Association; Tulane Institute on Water Resources Law and Policy; Missouri Public Utility Alliance; Colorado Department of Public Health and Environment; and California Water Boards.

Comment synopsis: summary

EPA received numerous comments expressing support for its recommendations, citing the scientific rigor of the recommendations, as well as the regulatory benefits associated with the technical approach. For example, commenters found that the overall risk-based approach is scientifically sound, documentation of the technical approaches provided transparency, the data are high quality and come from credible sources (EPA's NLA), and the conceptual models and the ecological principles that structure the stressor-response models are well-founded in the scientific literature. Commenters also expressed appreciation to EPA for being responsive to requests to extend the comment period for the draft recommendations.

Others offered specific remarks on the stressor-response approach, noting that the recommendations offer states an alternative approach to the ecoregional-based nutrient criteria for lakes and reservoirs that EPA published under CWA Section 304(a) in 2000 and 2001 that were based on a reference condition approach. Other comments elaborated, stating a preference for the recommendations because they explicitly link nutrient pollution stressors—nitrogen, phosphorus, and algal-derived organic carbon (i.e., Chl *a*)—to adverse effects on designated uses, which many states expressed as a critical component in communicating the importance of nutrient criteria to stakeholders and the public.

Several commenters remarked that EPA's statistical modeling approach was innovative and welcomed it as a departure from some of EPA's more conventional modeling approaches that have been used in the past to develop national recommendations under CWA Section 304(a). Additionally, EPA received widespread praise from state water quality agencies for developing the model visualizations (R Shiny apps) as part of its recommendations. Many found the visualizations intuitive to operate and foresee them being beneficial for decision making because they provide clarity on the complex interactions represented in the models. Other commenters found EPA's recommendations innovative because the models can be customized to reflect more local-scale environmental and water quality conditions. Many state water

quality commenters observed that such accommodations for state-specific data will be beneficial to them in developing numeric nutrient criteria.

Overall, commenters, including state water quality agencies, interpreted EPA's recommendations as a sign that it continues to recognize nutrient pollution as a serious threat to the quality of the nation's surface waters, affirming its commitment to achieving the CWA's overarching goals expressed in Section 101(a) (the "integrity" goal) and in supporting states' CWA authorities under Section 303 (water quality standards). Many state water quality agencies praised EPA for including specific protections of surface drinking water sources from nutrient pollution and its effects (e.g., cyanobacteria and microcystin) in its recommendations.

Response: EPA appreciates all the constructive comments it received and recognizes the significant investment of time by commenters to offer specific feedback on the draft recommendations.

Category 2 Problem Formulation

Category 2.1 Problem Formulation: Management Goals

Commenters: American Water Works Association.

Comment synopsis: Commenters approved of EPA's consideration of source water protection in drafting the recommended rules and that addressing water quality upstream of drinking water facilities is critical to protecting human health and minimizing treatment costs. Comments emphasized that EPA programs should emphasize protecting source water and enhance such efforts when possible.

Response: EPA thanks the commenter for these comments and wholeheartedly agrees, which is why drinking water endpoints were considered a critical component in problem formulation and why recommended values were developed to protect this use.

Category 2.2 Problem Formulation: Assessment Endpoints in General

Commenters: Wyoming Department of Environmental Quality, G. Hess, North Carolina Water Quality Association, Water Environment Association of Texas and the Texas Association of Clean Water Agencies, Wisconsin Department of Natural Resources, National Association of Clean Water Agencies, New Jersey Department of Environmental Protection, and California Water Boards.

Comment synopsis: Some commenters expressed concern about the specific endpoints selected; comments suggested that zooplankton:phytoplankton (Z:P) biomass was hard to interpret and was potentially affected by grazing, that dissolved oxygen was too simple and

data too sparse, and that microcystin per Chl was too variable and unrelated to nutrients. Other commenters provided general comments that the assessment endpoints were too narrow and that EPA should consider including recreational values for swimming and aesthetics, fishery metrics, and clarity/turbidity. Other comments suggested that EPA was assuming that phytoplankton was dominated by harmful algal blooms (HABs) and nuisance algae, and that EPA should consider historical use of lakes in setting assessment endpoint targets.

Response: The recommended criteria document lays out a scientific basis for the incorporation of various assessment endpoints replete with detailed risk hypotheses linking the selected variables to nutrient pollution. Commenters did not address the scientific arguments that justified the selection of the endpoints. EPA acknowledges that relationships of these endpoint variables to drivers is variable and has provided defensible methods to account for and reduce much of this variability. Zooplankton are grazed by fishes which is why this endpoint was selected, since it represents the efficiency with which primary producer carbon can be transmitted to secondary consumers. Dissolved oxygen data are among the most commonly collected and available data, and EPA found the data density for this variable to be greater than others; thus, states and authorized tribes should have little difficulty in collecting and applying this endpoint. Lastly, EPA acknowledges that microcystin production for a given level of Chl varies as evidenced by the figures shown in the criterion document; however, there is clear and widespread evidence in these data and the literature (detailed in the recommendation) that the risk of toxin production and concentration increase with Chl concentration. This variability is, in part, why EPA is providing flexibility to states and authorized tribes to make risk management decisions that reflects their risk tolerance in light of this variability.

EPA acknowledges that there is a larger population of potential assessment endpoints available to states and authorized tribes, but few that are available at the national scale required for this effort. EPA agrees with the commenter regarding the difficult of applying, for example, a single recreational value for swimming or aesthetics nationally given regional and even subregional differences in user perception. EPA did indeed pursue fishery metrics—these are captured in the aquatic life use assessment endpoints chosen. Lastly, turbidity/clarity were not expressly considered in these recommendations because of the difficulty in estimating turbidity from algal versus non-algal sources; however, nothing in these recommendations restricts states from developing relationships between turbidity and Chl on their own where they feel turbidity/clarity criteria would help in reducing impacts of nutrient pollution on designated uses. They can use the provided Chl-TN-TP models to set nitrogen and phosphorus targets to meet turbidity-based Chl targets.

Figure 1 in the recommended criteria document does not suggest that phytoplankton is comprised of only HABs or nuisance algae, only that those elements are components of phytoplankton and it would be improper to include them elsewhere.

Lastly, with regards to accounting for the historical use of lakes in setting endpoints, states are responsible for designating the uses of their waters, and they set these in a manner consistent with the CWA. States and authorized tribes can consider the historical use of lakes when setting these designated uses. In these recommended criteria, EPA has recommended assessment endpoint targets that ensure protection of three potential designated uses: aquatic life protection, recreation, and drinking water source protection.

Category 2.3 Problem Formulation: Assessment Endpoints in General

Commenters: Wyoming Department of Environmental Quality, Massachusetts Department of Environmental Protection, Arizona Department of Environmental Quality, National Wildlife Federation, North Carolina Water Quality Association, American Water Works Association, Oregon Department of Environmental Quality, Association of Metropolitan Water Agencies, Florida Department of Environmental Protection, Oregon Lakes Association, and Colorado Department of Public Health and Environment.

Comment synopsis: Several commenters expressed concern that additional cyanotoxins (e.g., anatoxins, cylindrospermopsin) were not considered for use as drinking water endpoint measures in addition to microcystin, arguing for regional variation in bloom species and toxin production, as well as the importance of these other toxins to protecting drinking water. Alternatively, there were comments that EPA should use microcystin directly rather than TN, TP, and Chl *a*, and that EPA should consider not including the drinking water criterion value of 0.3 micrograms per liter ($\mu\text{g/L}$) microcystin as a target because effective treatment technologies exist. Commenters also questioned whether or not these approaches apply to other drinking water standards. EPA also received comments regarding the use of alternative human health assessment endpoint measures including nitrate, manganese, clogging/filtration concerns, and disinfection byproducts.

Response: EPA acknowledges and even addressed the importance and significant threat imposed by other toxins in the recommended criteria. However, to develop nationally recommended criteria using a stressor-response modeling approach, EPA was limited by the availability of national scale datasets; the most defensible dataset available was the NLA microcystin dataset. Under 40 CFR part 131.11(b), states can pursue the development of criteria for other toxins using scientifically defensible methods.

Measuring microcystin directly may be helpful for the relevant assessment endpoints; however, it is still not routinely done across the United States. Additionally, the development of TN and TP targets are still needed for effective nutrient management to protect the management goals represented by those endpoints.

With regards to not applying drinking water microcystin targets for modeling purposes, EPA received as many comments noting that some states apply drinking water criteria to source water conditions and others that appreciated and supported the flexibility to select the state appropriate threshold. To support this wide range of needs, EPA will continue to provide states and tribes the flexibility to select a range of microcystin endpoints. In addition, the methods and decisions used here apply for this numeric nutrient criteria effort only.

Finally, with regards to other endpoints, for some of these (e.g., nitrate), national CWA Section 304(a) criteria recommendations already exist. For the others, again, national scale datasets to develop criteria for those measures were not available at the time of analysis. But also, again, under 40 CFR part 131.11(b), states can pursue the development of criteria for other measures using scientifically defensible methods.

Category 2.4 Problem Formulation: Aquatic Life Use Endpoints

Commenters: Wyoming Department of Environmental Quality; Massachusetts Department of Environmental Protection; R.T. Angelo and J. DeLashmit; Clean Water Action et al.; Iowa League of Cities; National Wildlife Federation; Center for Biological Diversity; City of Springfield, Missouri; Metropolitan St. Louis Sewer District; Missouri Department of Natural Resources; Idaho Department of Environmental Quality; National Association of Clean Water Agencies; Nevada Division of Environmental Protection; North Carolina Water Quality Association; and Florida Department of Environmental Protection.

Comment synopsis: Some commenters expressed specific concerns about the aquatic life use endpoints. For the Z:P biomass measure, these included concerns about its linkage to aquatic life, its appropriateness given a lack of state zooplankton data, observed increases in fishery production beyond a ratio of 0, that most lakes are managed for sports fisheries, and that its use might lead to oligotrophication. For the hypoxia measure, these included clarification on which fish were being used as endpoints, whether Z:P is more important than the hypoxia endpoint, how to handle lakes where mixing occurs before the end of the growing season, whether EPA considered other biota susceptible to hypoxia, and what to do if a state has no dimictic lakes. EPA also received comments asking why they did not consider other aquatic life use endpoints, including vulnerable species, zooplankton alone, other invertebrates, and changes in biological community structure. Lastly, EPA was asked about sublethal effects of microcystin on aquatic life.

Response: The recommended criteria document lays out a scientific argument for the incorporation of various assessment endpoints replete with detailed risk hypotheses linking the selected variables to nutrient pollution. Section 304a of the CWA requires that EPA develop criteria “...*accurately reflecting the latest scientific knowledge...on the kind and extent of all identifiable effects on health and welfare including, but not limited to, plankton, fish, shellfish,*

wildlife, plant life, shorelines, beaches, esthetics (sic), and recreation...and on the effects of pollutants on biological community diversity, productivity, and stability, including information on the factors affecting rates of eutrophication." The technical document states that zooplankton are grazed by fishes, which is why this endpoint was selected because it represents the efficiency with which primary producer carbon can be transmitted to secondary consumers. It describes the importance of this phenomenon to the lake food web, inclusive of, but not limited to, sports fishes. Moreover, zooplankton, phytoplankton, and their consumers are all aquatic life per se, the totality of which is considered within the purview of protection under the CWA. Disturbance to the productivity of lakes and stability of lakes reflected in this ratio is within the requirements of Section 304(a). The condition of apex predator populations may not be a de facto indication of ecosystem stability as many such fisheries can require active management. EPA acknowledges that there may be increases in fishery production at Chl levels above those where zooplankton biomass no longer increases in response to phytoplankton biomass, such systems may result from increased zooplankton productivity:biomass ratios, alternative food sources, or fishery management activities (e.g., harvest), among others. This reality does not remove the fact that this ratio is linked to aquatic life use as described in the criterion document. Lastly, the risk of oligotrophication does not apply to this endpoint because states may select ratios that reflect a range of trophic states. Finally, states are able to adjust the ratio to reflect desired secondary production goals for the lake, including those that are beyond close coupling of Z:P biomass.

Dissolved oxygen data are among the most commonly collected and available data; EPA found the data density for this variable to be greater than others, therefore, states should have little difficulty in collecting and applying this endpoint. EPA detailed the fisheries (cool-water and coldwater) that are the focus of this endpoint. EPA does not consider this or the Z:P biomass ratio endpoint more or less important in developing criteria; both are important and should be considered, where applicable, by states in exploring criteria development that are catered to individual lake designated uses. With regards to lakes where mixing occurs before the end of the growing season, EPA stated that these endpoints apply to the growing season; it is, therefore, appropriate to use surface Chl samples from throughout the growing season. EPA did consider other biota linked to hypoxia but lacked sufficient data on a national scale to develop such models. Lastly, states without seasonally stratified lakes or which do not otherwise meet the parameters for applying the hypoxia model should not apply that model to develop Chl *a* endpoints.

EPA did consider alternative aquatic life use endpoints but found there were insufficient data on a national scale to develop any of the alternative aquatic life use endpoints commenters mentioned, including for fishes, amphibians, vulnerable taxa, or biological community structure. NLA did not collect fishes or amphibians or vulnerable taxa and did not develop assemblage-based multimetric or other indices for zooplankton or phytoplankton that could be used. Similarly, EPA lacked sufficient data to develop national scale stressor-response models

using microcystin and sublethal invertebrate responses; moreover, lethal or sublethal effect criteria for such toxins would benefit from toxicologically based endpoint development approaches.

Category 2.5 Problem Formulation: Assessment Endpoints for Recreational Criteria

Commenters: Massachusetts Department of Environmental Protection; R.T. Angelo and J. DeLashmit; Wyoming Department of Environmental Quality; Montana Department of Environmental Quality; City of Springfield, Missouri; and Metropolitan St. Louis Sewer District.

Comment synopsis: Some commenters expressed concerns about recreational criteria assessment endpoints. Some expressed concern that microcystin endpoints might not protect all potential human health effects and questioned whether effects from dermal contact are less sensitive to ingestion; some also suggested that EPA did not explain why the microcystin endpoint protects against skin reactions. Other comments suggested the microcystin endpoint was not necessary because EPA already has recommended 304(a) criteria for microcystin, and thus, nutrient criteria to protect against it were not needed. Another commenter recommended adding a transparency target to protect recreation and still another requested cyanobacteria-to-Chl models so that other recreational endpoints might be developed based on other cyanobacterial density levels used for recreational assessments.

Response: The recommended criteria document lays out the scientific argument for the incorporation of various assessment endpoints replete with detailed risk hypotheses linking the selected variables to nutrient pollution. One commenter suggested EPA did not provide an explanation for why microcystin is protective against skin reactions. EPA did not comment on the protectiveness of the microcystin endpoint for skin reactions specifically; EPA's 2019 recommended recreational criteria for microcystin details the appropriateness of microcystin levels for protecting recreation, based on incidental ingestion, and is used as a basis for adopting microcystin here. With regards to the sufficiency of microcystin to protect designated uses, 40 CFR 131.11 directs that criteria be developed based on sound scientific rationale and contain sufficient parameters and constituents to protect the designated use. EPA acknowledges that microcystin is readily measurable, but it is not routinely measured by states. Combined with the variability in the temporal and spatial expression of microcystin to nutrients and, thus, difficulty in capturing the expressed risk, additional parameters and constituents like Chl *a* and nutrients are justified to ensure protection of recreation from the effects of nutrient pollution.

Transparency that could affect recreational safety is influenced by a variety of factors and EPA acknowledges that algal biomass is one of those. However, the lack of a readily available dataset and approach for dissecting the variable influences of these factors and isolating Chl across the range of lake types nationally was an impediment. Also, the unavailability of national

transparency targets related to swimmer safety and, as noted by other comments, the variable nature of user transparency preference across states were impediments to selecting transparency as a target. Many states have already adopted water clarity or turbidity criteria to protect aquatic life uses; the states may wish to evaluate whether those criteria also protect recreational uses.

EPA did not provide cyanobacterial density/biovolume – Chl stressor-response models because neither of these cyanobacterial attributes were considered for use as assessment endpoints for this effort. Users interested in these modeled relationships within the recreational endpoint model should contact EPA or consider constructing such models using the NLA dataset for themselves.

Category 3 Analysis

Category 3.1 Analysis: National Lakes Assessment Data

Commenters: Virginia Department of Environmental Quality; Upper Neuse River Basin Association; Massachusetts Department of Environmental Protection; Louisiana Department of Environmental Quality; Wyoming Department of Environmental Quality; Footprints in the Water, LLC; Iowa League of Cities; North Carolina Water Quality Association; Missouri Department of Natural Resources; Arkansas Department of Energy and Environment; Federal Water Quality Coalition; Texas Commission on Environmental Quality; Oregon Department of Environmental Quality; Water Environment Association of Texas and the Texas Association of Clean Water Agencies; Idaho Department of Environmental Quality; Wisconsin Department of Natural Resources; Kansas Department of Health and Environment; National Association of Clean Water Agencies; New Jersey Department of Environmental Protection; Iowa Department of Natural Resources; Florida Department of Environmental Protection; Tulane Institute on Water Resources Law and Policy; Colorado Department of Public Health and Environment; and North Carolina Department of Environmental Quality.

Comment synopsis: Commenters expressed concern about the use of NLA data for deriving nutrient criteria. These concerns included the large spatial extent of the data, the limited number of repeat samples collected from individual lakes, the possibility that drought or flooding affected samples collected in particular years and in particular locations, the degree to which a single sample in each lake represented the range of possible conditions in a lake, and other related issues. Commenters also expressed concern about the lack of references to quality assurance project plans for the NLA.

Response: The national models estimate relationships between parameters at the level of individual samples. For example, the model for microcystin estimates the relationship between Chl *a* concentration and microcystin in a water sample. The intent of these models is not to

characterize conditions within individual lakes, which is an exercise that is not necessary to recommend national numeric nutrient criteria. Instead, the models provide insight into the likely range of a measurement of interest (e.g., microcystin) given a different measurement (e.g., Chl a). The NLA data are ideally suited for these models because they provide samples collected from a wide variety of lakes that span the country. When the modeling unit is taken into account, the relevance of many factors weakens. For example, drought conditions may lower water levels and increase the concentrations of solutes in a lake. However, in terms of the unit of analysis, a sample collected during a drought merely contributes another combination of parameters in a water sample, expanding the range of conditions and diversity of samples in the modeled data. A similar argument applies to particular, unique lakes highlighted by commenters.

References to NLA quality assurance project plans have been added to the criterion document.

Category 3.2 Analysis: Depth in the NLA

Commenters: Louisiana Department of Environmental Quality, New Jersey Department of Environmental Protection, and Iowa Department of Natural Resources.

Comment synopsis: Commenters asked about the measurement of depth in the NLA and how that measurement translates to monitoring. Commenters also asked how criterion models could be applied to lakes with complex bathymetry and a variety of depths.

Response: Depth in the NLA is measured at the deepest point in lakes and at the midpoint of reservoirs. These depth measurements can be viewed as an approximation to the true maximum depth of the lake. Practitioners measuring lake depth to use in the models can employ similar methods as used by the NLA or other approaches (e.g., bathymetric maps) to identify the deepest point of a lake. Because depths can vary depending on time of year and vary with location in the lake, EPA recommends that practitioners consider the effects of variations in depth on final criteria and weigh those variations when selecting the final criterion. For example, small variations in depth may not appreciably change the magnitude of derived criteria. With regard to the question about the applicability of the models to lakes with complex bathymetry, EPA notes that the influence of depth on some models (e.g., TP-Chl, zooplankton, hypoxia) has been taken into account in a way that is consistent with the mechanistic influence of depth on the model results at the sample location. So, in lakes with complex bathymetry, the use of water depth at the sample location likely yields the most accurate predictions.

Category 3.3 Analysis: Other Data Sources

Commenters: Nebraska Department of Environment and Energy; R.T. Angelo and J. DeLashmit; Wyoming Department of Environmental Quality; City of Springfield, Missouri; Metropolitan St.

Louis Sewer District; Water Environment Association of Texas and the Texas Association of Clean Water Agencies; Iowa Department of Natural Resources; Maryland Department of the Environment; NEIWPC; and North Carolina Department of Environmental Quality.

Comment synopsis: Commenters identified several data sets that they suggested would help EPA refine criterion models. Another commenter suggested that the Missouri state data used to refine national models in the included case study was not representative of state conditions.

Response: The models rely on simultaneous measurement of several parameters, and datasets identified by commenters generally did not have the required measurements to refine the national model. Datasets in which measurements are incomplete can, in some cases, be used to refine the models (see example in Appendix A). However, for fitting the national criterion model, EPA focused on the complete data sets available in the NLA. The 2017 NLA data was not publicly available at the time of this work, but after it is released and becomes available, national criterion models could be updated using these data; however, it is not anticipated that the addition of the 2017 NLA data would alter the models in an appreciable manner.

With regard to the representativeness of state data used to refine the national models in the case studies, note that the unit of modeling is based on the individual sample, and the intent of the models is to estimate relationships between different parameters. Therefore, the important aspect of representativeness in this application is whether the full range of possible conditions is observed in the dataset. When state data are used to refine the national model, state data usually represent a narrower range of conditions than the national model, and the inclusion of the national model ensures that the full range of conditions is represented. Hence, the range of conditions available in the state data (and its representativeness) is much less important.

Category 3.4 Analysis: Field Collection Methodologies

Commenters: Louisiana Department of Environmental Quality, Wyoming Department of Environmental Quality, Iowa League of Cities, Water Environment Association of Texas and the Texas Association of Clean Water Agencies, and Nevada Division of Environmental Protection.

Comment synopsis: Commenters expressed concern that field collection protocols used by the NLA were not the same as commonly used protocols in state monitoring programs. Examples of these differences included depth-integrated versus grab samples, mesh size for zooplankton samples, and quantification techniques for microcystin. Commenters were also concerned about the effects of different bottles used during each NLA survey. Finally, commenters noted that the intake location for water supply lakes differs from the surface samples collected by NLA.

Response: The NLA used a depth-integrated sampler to collect water samples; therefore, the criterion model measurements of different water quality parameters (e.g., Chl *a*, TP, and TN) also reflect depth-integrated values. In contrast, many state programs rely on grab samples collected at a single depth. EPA acknowledges that some differences in water quality measurements is possible when comparing depth-integrated samples with grab samples. However, because the unit of measurement is the sample, maintaining internal consistency can resolve many of these issues. For example, a unit of Chl *a* is associated with a particular probability of exceeding a certain microcystin concentration, and this relationship will be applicable whether the unit of Chl *a* is estimated from a depth-integrated sample or from a grab sample.

Measurements of zooplankton biomass would likely vary depending on the mesh size used when sampling. Within the NLA data, consistent methods were used, but if state zooplankton data are used to refine these models, then conducting additional work to examine the comparability of state and NLA data may be advisable.

Similarly, NLA measurements of microcystin concentration are internally consistent, but when combining national and state data, any differences in lab protocol used in the state data may need to be evaluated.

Effects of different types of sample bottles were examined in the NLA and no significant difference was observed in microcystin concentration based on bottle type.

The effects of intake location are relevant when deriving criteria protective of drinking water uses because differences in Chl *a* and microcystin concentrations are expected between surface and deep water locations. However, as the modelling unit is the sample, we expect that the relationship between Chl *a* and microcystin to be consistent whether samples are collected from deep waters or shallow waters; therefore, criterion derived for Chl *a* can be effectively applied to deep water samples.

Category 3.5 Analysis: Zooplankton Model – General Comments

Commenters: Iowa League of Cities; Wisconsin Department of Natural Resources; New Jersey Department of Environmental Protection; Florida Department of Environmental Protection; North Carolina Department of Environmental Quality; and Counsel, Mississippi River Collaborative et al.

Comment synopsis: Commenters asked for clarification on how measurement errors were modeled, how seasonal mean phytoplankton biovolume was estimated, how the ratio of Z to P was converted to a ratio between zooplankton and Chl *a*, and suggested that additional peer review was needed. Other commenters asked for clarification regarding the influence of direct

measurements of phytoplankton biovolume in the model relative to other measurements (zooplankton biomass, Chl *a*). Other commenters suggested that the criterion models were flawed because of the potential for very high candidate Chl *a* criteria, and they suggested that EPA impose an upper limit of Chl *a* = 25 µg/L. Commenters also suggested that a lag in the zooplankton response relative to changes in phytoplankton would be expected. Commenters also observed that the proportionality constant, *b*, that was included in the criterion document was not in the R scripts, and another commenter requested that a background screen be added to the Shiny app for the zooplankton model.

Response: As described in the criterion document, a lumped estimate of the variability in direct phytoplankton biovolume measurements is estimated in the model. This variance includes contributions from temporal, within-lake spatial, and measurement variability. In the Bayesian network models, this variance is incorporated into the model such that estimates of seasonal mean phytoplankton biovolume are represented as distributions of possible values, rather than as single measured quantities.

The models compute an estimate of the ratio between zooplankton and phytoplankton biomass not zooplankton and Chl *a* as suggested by the commenter. In the model, direct measurements of phytoplankton biovolume provide a second, independent estimate of phytoplankton biovolume, and both it and Chl *a* contribute to the estimated phytoplankton biovolume for each sample. The correct R script for the zooplankton model has been provided with the criterion document. The zooplankton model has been reviewed both as a separate article in the peer-reviewed literature (see references in the criterion document) and as an external review of the criterion document.

The suggestion that EPA impose an upper limit of Chl *a* criterion values is predicated on the assumption that exceeding a specific threshold for Chl *a* is known to affect designated uses. This threshold may exist, but the commenter provided no evidence in support of the proposed threshold of 25 µg/L. The models in the current criterion document are intended to provide empirical support for Chl *a* thresholds corresponding to specific endpoints that then correspond to protection of specific designated uses.

The comment that a lag in the zooplankton response would be expected relative to changes in phytoplankton applies to models in which one is attempting to model temporal changes in food web structure. In contrast, the criterion model seeks only to represent average patterns in the relationship between zooplankton and phytoplankton biomasses across many lakes; it does not attempt to resolve temporal changes. Consideration of temporal changes would be an interesting exercise if data that is well-resolved in time were available.

The comment that the proportionality constant, b , is missing from the script is correct, and the text in the document has been corrected to reflect this change. A *Background* tab has also been added to the zooplankton Shiny app.

Category 3.6 Analysis: Stressor-Response Approach – General

Commenters: Footprints in the Water, LLC; Federal Water Quality Coalition; Minnesota Environmental Science and Economic Review Board; Wisconsin Department of Natural Resources; National Association of Clean Water Agencies; and Florida Department of Environmental Protection.

Comment synopsis: Commenters suggested that the statistical approach used to estimate relationships was useful for exploring patterns, but not for deriving criteria. Other commenters asked for further justification for distributions that were assumed for different measurements and parameters. Commenters also referred to past suggestions from the Science Advisory Board review of draft guidance on stressor-response approach that suggested the use of tiered weight of evidence approach, reminded EPA that correlation of variables does not imply causation, asked whether confounding variables were considered in the analysis, and finally, suggested that the statistical methods required further peer review.

Response: The statistical method used to estimate criterion models is based on well-established statistical modelling techniques. The criterion document has undergone extensive peer review both on its own and in independent journal articles (as referenced in the criterion document). These statistical models are well-suited for deriving criteria because they accurately predict the range of possible responses, given the values of different predictors, and as such, provide information that is directly relevant to environmental decision making. The models do not use the correlation of variables to infer causal relationships. Instead, as recommended by EPA's 2010 Stressor-Response guidance, the models are used to estimate known causal relationships among different variables. Covariates that were identified as strongly influential through exploratory analysis (i.e., possible confounding variables) are included in the models as needed for different variables (e.g., depth in TP-Chl model). These exploratory analyses are described in the referenced journal articles.

Tiered weight-of-evidence approaches are another scientifically defensible approach that states or authorized tribes may utilize for deriving nutrient criteria. The stressor-response techniques used in these nationally recommended criteria does not preclude the use of these other approaches by states or authorized tribes when adopting nutrient criteria into their water quality standards.

The specification of the distributions of different variables (e.g., normal, log-normal) followed standard practice for statistical modeling. Environmental measurements are often log-normally

distributed, and the fit of this distribution was confirmed with quantile-quantile plots. In cases in which distributions other than log-normal or normal were assumed (e.g., negative binomial), extensive analysis was conducted to identify the most appropriate distribution (see independent journal articles referenced in the criterion document for examples).

Category 3.7 Analysis: Zooplankton Model – Effect of Depth and Other Covariates

Commenters: Wyoming Department of Environmental Quality, Wisconsin Department of Natural Resources, and Iowa Department of Natural Resources.

Comment synopsis: Commenters asked for clarification as to why depth was used as a classifying variable in the zooplankton model, and they asked whether other variables were considered. Commenters also asked how depth should be used in assessment when samples may be collected from lake locations with different depths. Commenters also questioned the selection of the three classes of depth used in the analysis.

Response: Depth was selected as a classifying variable due to its strong effect on structuring lake biological communities. For zooplankton, lake depth can determine whether benthic zooplankton are included when samples are collected from the water column. Initial exploratory analysis of the NLA data also indicated that lake depth, after Chl *a* concentration, was a strong predictor of zooplankton abundance (see associated journal article on zooplankton referenced in the criterion document). The difference in relationships estimated for lakes in each depth class reflects the strong effects of lake depth on zooplankton biomass. Other variables were also considered in exploratory analysis, but the magnitude of variability in the zooplankton biomass measurement limited the degree to which additional classification variables could be considered. Because the NLA randomly selected lakes to sample, the effects of other possible covariates (e.g., presence of top-down feeding pressure from fish, macrophytes in shallow lakes, regional differences in phytoplankton or zooplankton composition) likely exerted a random effect on the estimate relationships, which would increase residual variability, but would not alter the mean relationship. The stratification status of lakes may also provide useful information, including the Wisconsin-specific predictor of stratification status, but as noted above, a limited capacity for classifying variables was available with the zooplankton model.

The effect of depth was taken into account in the model by dividing the data set into three depth classes with similar numbers of samples. The selection of three classes represents a compromise between the need to maintain a sufficient number of samples within each class to estimate the mean relationships and the need to best represent the continuous effect of changes in depth. A greater number of classes yielded too few samples in each group, while a smaller number of classes did not fully account for the effects of depth. For these national

criteria, EPA opted for model simplicity; states and authorized tribes can explore different depth classes using local data.

Criterion models are based on samples collected at the deepest point of lakes and in the middle of reservoirs, and estimated relationships reflect these sample locations. Zooplankton assemblage structure may vary with location in the lake, especially when comparing littoral zone samples with samples collected in the middle of the lake. States and authorized tribes may wish to further study the effect of sample location with state and local zooplankton data sets.

Category 3.8 Analysis: Zooplankton Model – Data Availability

Commenters: Massachusetts Department of Environmental Protection, Upper Neuse River Basin Association, and Louisiana Department of Environmental Quality.

Comment synopsis: Commenters noted that zooplankton measurements are rarely collected by states during routine monitoring, which limits the opportunities for refining models using state-specific data.

Response: EPA recognizes that zooplankton monitoring is rarely conducted by states, but notes that the zooplankton model and the other models all can be used with just the national data to derive defensible nutrient criteria that protect aquatic life designated uses. Refinement of the models with state data is optional, and the availability of this option hinges on the availability of state monitoring data. States that wish to further explore refining the zooplankton model by collecting these data are welcome to do so. EPA is ready to assist states and authorized tribes in developing monitoring plans for zooplankton as needed.

Category 3.9 Analysis: Zooplankton Model – Threshold Selection

Commenters: R.T. Angelo and J. DeLashmit, Wyoming Department of Environmental Quality, Montana Department of Environmental Quality, Arkansas Department of Energy and Environment, Wisconsin Department of Natural Resources, New Jersey Department of Environmental Protection, Florida Department of Environmental Protection, and NEIWPC.

Comment synopsis: Commenters asked for more guidance regarding the management threshold for the ratio between $\log(\text{zooplankton biomass})$ [$\log(Z)$] and $\log(\text{phytoplankton biomass})$ [$\log(P)$].

Response: The slope between $\log(Z)$ and $\log(P)$ provides an indication of how the degree to which zooplankton and phytoplankton biomasses are coupled. A loss of this coupling indicates that food web connectivity is weak and aquatic life is not fully supported in the lake. Hence, when the slope between $\log(Z)$ and $\log(P)$ is zero, aquatic life is not protected. Lower percentiles of the posterior distribution of possible mean criterion values (i.e., higher values of

the credible interval) can be used to account for model uncertainty when setting criteria about this threshold; slightly higher thresholds can be selected to provide a margin of safety before the loss of designated use protection.

As pointed out by commenters and in the text of the criterion document, oligotrophic lakes have higher values of $\log(Z)/\log(P)$. Different threshold for these types of lakes may be possible, and these thresholds could be demonstrated to be protective of the existing conditions in the lake. Traditional approaches based on quantifying these conditions in a reference set of lakes may be appropriate in these applications to ensure designated use protection.

Category 3.10 Analysis: Deepwater Hypoxia – General Comments

Commenters: Ecosystem Consulting Service, Inc; Louisiana Department of Environmental Quality; Wyoming Department of Environmental Quality; Iowa League of Cities; City of Springfield, Missouri; Metropolitan St. Louis Sewer District; Oregon Department of Environmental Quality; Idaho Department of Environmental Quality; Wisconsin Department of Natural Resources; National Association of Clean Water Agencies; Florida Department of Environmental Protection; NEIWPC; and Counsel, Mississippi River Collaborative et al.

Comment synopsis: Commenters expressed concern that Chl *a* thresholds calculated from the criterion model did not match with previously established thresholds for particular lakes. Some commenters asked how model assumptions regarding initial dissolved oxygen concentration, first day of stratification, and the model for dissolved oxygen < 2 milligrams per liter (mg/L) were validated. With regard to the use of dissolved organic carbon (DOC) in the model, commenters noted that DOC originates from autochthonous production in the lake as well as from watershed loading, and they also noted that increased DOC and lake browning has been observed, suggesting that separate DOC criteria were warranted. Commenters also suggested that the term “depth below the thermocline” was unclear, and that DOC and depth-below-thermocline data were not commonly available. Commenters also requested that dissolved oxygen requirements be provided for different fish and different life stages. Commenters also asked whether the models accounted for high elevation lakes and posed questions regarding the application of the criterion model, including how to measure the depth below the thermocline and what a refuge of zero meters indicated. Commenters also suggested using mechanistic models to derive criteria. Commenters asked about how one can interpret combinations of conditions in which the R Shiny apps do not yield criterion values. Commenters suggested that the assumption of a continuous loss of dissolved oxygen from the deep waters over the course of the summer was a worst-case scenario, and they wondered whether this was an appropriate basis for criteria.

Response: The degree of comparability between Chl *a* criterion calculated with the current criterion model and with existing criteria depends on a variety of factors, including the methods

used in deriving the existing criteria. Criteria would be expected to be similar in cases in which similar approaches were used. Otherwise, interpreting differences in criteria computed from different methods can be very difficult. When selecting a final criterion value to protect aquatic life, states and authorized tribes should select the criteria that is protective of the most sensitive endpoint for which data exist. Comparisons of candidate criteria with conditions in individual lakes that are known to harbor cool- or cold-water fish suffer from the same issue—the underlying basis of the criteria may differ from a simple assessment of observed conditions in the lakes of interest. When sufficient data from an individual lake is available, updating the model with lake-specific data may address some of the comparability issues.

The hypoxia model makes use of two assumptions: (1) that the first day of stratification can be estimated from mean annual temperature and (2) that the initial DO concentration in deep water can be estimated from the minimum annual air temperature. Both of these assumptions are consistent with previously published analyses (see citations in Yuan and Jones, 2020). The values of two parameters are also closely related in the statistical model, and errors in one parameter can be compensated by adjusting the value of the other. More specifically, in the present model, initial dissolved oxygen concentration in each lake is specified as a single value based on minimum air temperature. The first day of stratification, in contrast, is estimated as a parameter estimated from observed data. Hence small variations in the initial dissolved oxygen concentration in a lake around the value predicted by the minimum air temperature can be taken into account by adjusting the date of stratification. That is, the initial dissolved oxygen concentration and first day of stratification together determine the y-intercept of the straight line fit to the observed data, but because only the first day of stratification is fit to the data, a robust estimate of the fit can be calculated. Validation of the initial date of stratification by comparing to other studies is useful, but note that the main intent of this parameter is to estimate the correct y-intercept of the dissolved oxygen depletion curve; this y-intercept is a function of both initial dissolved oxygen concentration and first day of stratification. Also, as noted in Yuan and Jones (2020), the estimated first day of stratification in the statistical model is consistent with the day of the year at which transport of dissolved oxygen from the surface layers of the lake is slowed, which is a different definition of the first day of stratification than is commonly used.

The model approach of representing dissolved oxygen measurements < 2 mg/L as censored data is discussed extensively in Yuan and Jones (2020); therefore, this discussion is not repeated in the criterion document. In short, the approach allows EPA to accurately model the linear decrease in dissolved oxygen concentrations while accounting for the fact that the relationships becomes non-linear as dissolved oxygen approaches zero. Nonlinear models can be fit to this relationship but they do not lend themselves as readily to the hierarchical structure that is used to estimate model coefficients within each lake.

EPA recognizes that DOC originates from both watershed and within-lake sources and does not distinguish between the two when fitting the hypoxia model. The browning of lakes has been noted in the peer-reviewed literature, but recommending DOC criteria is beyond the scope of the current effort. Ranges of DOC provided in the R Shiny apps are not intended to indicate recommended values, but instead simply reflect the range of DOC values observed in the NLA dataset. When monitoring data for DOC and/or depth below the thermocline are not available, EPA suggests using other data sets to infer likely distributions of values for these parameters to use in the models.

EPA acknowledges that the term “depth below thermocline” may be confusing and has added it to a glossary of terms for the criterion document.

Dissolved oxygen requirements for different fish species and life stages are detailed in other publications; providing that information is beyond the scope of the current work. The intent of the R Shiny apps is to provide the users with the flexibility to account for different fishes’ dissolved oxygen requirements unique to the systems of interest.

High-elevation lakes are taken into the model via the specification of elevation in the hypoxia model. EPA recommends that states wishing to measure depth below the thermocline follow a similar protocol as used by the NLA to measure vertical profiles of temperature. In lakes with depths that vary spatially, EPA recommends that depth-below-the-thermocline be measured with an approach similar to that used in the NLA. Similarly, in lakes in which the thermocline deepens over the summer (and therefore reduces the value of the depth below the thermocline), EPA recommends that the seasonal mean depth-below-the-thermocline be used in the model, as this statistic best represents the process by which lake depth affects the rate of dissolved oxygen decrease in deep waters.

In the R Shiny app, the minimum refuge depth of zero meters provides the logical lower bound to the possible refuge size. The availability of this selection does not imply that a selection of a refuge depth of zero meters protects aquatic life. Criteria derived using the national criterion models, as with any criteria, require submission to EPA under CWA Section 303(c), whereupon EPA will review the submission for scientific defensibility and for protection of the applicable designated use(s).

As for mechanistic models, the current criterion models do not preclude the use of other scientifically defensible approaches for states and authorized tribes to use to derive criteria that are protective of their designated use(s). Mechanistic models are informative, but not feasible for application over the large spatial area and the many different lakes included in this criterion model. Metalimnetic maxima in dissolved oxygen are incorporated in the measured values of depth-averaged dissolved oxygen in the model; therefore, the existence of such a maximum does not influence the applicability of the model. However, as with any criteria, when

conditions for a particular water body can be shown to be unique, site-specific criteria can be derived. EPA's goal with this national-level effort is to develop nationally recommended criteria that will protect the vast majority of lakes from the effects of nutrient pollution without being too over- or under-protective.

The modeled scenario in which dissolved oxygen in the deeper waters of the lake decrease linearly over the course of the summer is applicable to stably stratified lakes. In other lakes, this scenario might be viewed as a "worst-case" scenario, in which stratification is stable throughout the summer. The potential impacts on aquatic life are severe (i.e., fish kills) when refugia are not available, and considering worst-case scenarios is consistent with the need for criteria to be protective of designated uses. Furthermore, other flexibilities in specifying parameters (e.g., refugia size) allow states to incorporate their own risk management decisions in calculating criteria based on this scenario.

In cases in which the R Shiny app does not return Chl *a* criteria, the combination of parameters yields conditions in which either cool water refugia are always present or cool water refugia are never present. In either case, states and tribes can use this information to guide management decisions for those systems.

Category 3.11 Analysis: Zooplankton Model – Applicability to Other Systems

Commenters: Wasatch Front Water Quality Council and Wyoming Department of Environmental Quality.

Comment synopsis: Commenters asked whether a regional version of the zooplankton model was available, and they questioned whether the model was appropriate for certain types of systems (e.g., large shallow lakes).

Response: The criterion model is fit to data collected from lakes samples by the NLA; therefore, the estimated relationship is applicable to this population of lakes. Application of the model to types of lakes that were not sampled by the NLA (e.g., stock ponds) may be possible, but EPA recommends that states and authorized tribes conduct further study to determine whether it yields protective result for these systems. Similarly, other types of lakes that can be demonstrated as differing substantially from the mean trends estimated in the criterion model may be candidates for site-specific criteria.

Category 3.12 Analysis: Deepwater Hypoxia – Other Measures

Commenters: Ecosystem Consulting Service, Inc; Montana Department of Environmental Quality; Iowa League of Cities; and Wisconsin Department of Natural Resources.

Comment synopsis: Commenters suggested that other measures of oxygen depletion or hypoxia may be more appropriate. Examples of other measures included areal hypolimnetic oxygen demand and anaerobic respiration products (e.g., sulfides). Commenters also suggested that other dissolved oxygen threshold for other fish life stages be considered.

Response: Other measures of hypoxia are informative, but EPA chose to model depth-averaged dissolved oxygen because of its direct relationship to known thresholds for fish and because of its relationship to eutrophication status of a lake.

As described in other responses, the specification of the targeted dissolved oxygen in the R Shiny app provides users with an opportunity to tailor nutrient criteria to fish species and life stages that exist in lakes of interest. In the calculation of the depth-averaged dissolved oxygen from the specific dissolved oxygen threshold, the entire refugia has a dissolved oxygen concentration that is at least as high at the dissolved oxygen threshold.

Category 3.13 Analysis: Deepwater Hypoxia – Applicability

Commenters: Iowa League of Cities; Alliance for the Great Lakes; City of Springfield, Missouri; Metropolitan St. Louis Sewer District; Missouri Department of Natural Resources; Arkansas Department of Energy and Environment; Florida Department of Environmental Protection; and Missouri Public Utility Alliance.

Comment synopsis: Commenters asked for clarification on the applicability of the hypoxia endpoints. Specific questions included whether the endpoint is applicable to non-dimictic lakes with only warm water fish, whether the endpoint is applicable to states that only enforce the dissolved oxygen criteria in epilimnetic waters, whether the endpoint is applicable only to states that explicitly designated waters as cool- or cold-water fish habitat, and whether the endpoint is applicable to lakes that stratify weakly or discontinuously. Commenters also asked for clarification regarding the definition of dimictic lakes and asked whether the initial selection of lakes for analysis excluded known, dimictic lakes. Commenters also noted that some warm water fish have optimal temperature ranges similar to cool water fish, and they asked whether the endpoint should apply in these circumstances. Commenters also suggest that the relationship between dissolved oxygen and depth below the thermocline was only applicable to certain types of lakes.

Response: The criterion model for the hypoxia endpoint was fit using data from an initial screen of NLA lakes to identify those that were likely to be dimictic. This previously published screen was based on a statistical analysis as described in the cited reference in the criterion document, and, as with any statistical model, it is possible that the screen was conservative and did not include a few lakes that are dimictic. However, the addition of a few samples is unlikely to substantially alter the mean effects estimated by the model.

As stated in the criterion document, the underlying mechanism of the hypoxia endpoint is predicated on the assumption that the temperature sensitivities of certain cool- and cold-water fish force them to seek deeper water refugia during the summer. Temperature requirements of certain warm water fish may be similar, and states and tribes can extend the criterion model to apply to those fish where appropriate. In states in which fish that are sensitive to warmer temperature are not found (e.g., some southern states), this endpoint may not be applicable. That is, the fact that the R Shiny app allows the computation of criteria in certain locations does not indicate that application of the criterion in that location is expected. Application of the hypoxia endpoint is not limited to states that have specifically designated cool- and cold-water habitat use, as cool- and cold-water fish exist in states without such refinements to their designated uses and represent aquatic life in those states.

The underlying mechanism of the hypoxia model provides insights into how the criterion model can be extended to systems that stratify weakly or discontinuously. In such cases, an evaluation of the length of time that stratification persists and the associated depletion of oxygen during this time may inform questions of whether the criteria applies for such systems.

EPA acknowledges that some states only monitor for compliance with the dissolved oxygen criteria in lake epilimnions. These existing criteria should not affect the application of nutrient criteria that include dissolved oxygen in the causal pathway linking increased nutrients to effects on aquatic life use. That is, specifying criterion values for every step in a causal pathway linking increased nutrients to effects of designated uses is not, in general, necessary.

EPA acknowledges that the definitions of dimictic lakes in the criterion document are unclear. EPA has clarified this text to refer to seasonally stratified lakes where appropriate.

Commenters notes that the relationship between dissolved oxygen and depth below the thermocline was not applicable to certain lakes. Here, the commenter misunderstands the model, as the model expresses a relationship between *depth-averaged* dissolved oxygen and depth below the thermocline among different lakes—a relationship that is very different from a relationship between the depth and dissolved oxygen measured at that depth. This latter case characterizes the shape of the depth–dissolved oxygen profile, whereas the criterion model expresses the relationship between depth-averaged dissolved oxygen and depth below the thermocline among lakes.

Category 3.14 Analysis: Deepwater Hypoxia – Temperature Thresholds

Commenters: Wyoming Department of Environmental Quality; City of Springfield, Missouri; Metropolitan St. Louis Sewer District; and Nevada Division of Environmental Protection.

Comment synopsis: Commenters suggested that the critical temperature provided in the criterion document were conservative. Other commenters requested additional information regarding critical temperatures for different fish species and different life stages.

Response: The information provided in the document regarding potential critical fish temperatures was intended only to illustrate the types of resources that can be used to determine critical temperatures for relevant fish species for a state or authorized tribe. The slider for temperature in the R Shiny app provides the means for customizing the criterion calculation to fish species in lakes of interest and to different fish life stages; a state or authorized tribe can consult with game and fish agencies and other references to select this temperature to develop criterion values that are protective of fish species in their specific lakes. The slider for critical temperature was revised to accommodate a broader range of temperatures.

Category 3.15 Analysis: Microcystin Model – General

Commenters: NEIWPC, Tulane Institute on Water Resources Law and Policy, and North Carolina Department of Environmental Quality.

Comment synopsis: Commenters asked how the allowable exceedance frequency might be computed from field data using either continuous or regularly sampled data. Also, some commenters were uncertain as to how the functional form for Chl *a* versus microcystin was specified.

Response: The exceedance frequency specified in the R Shiny app is computed from the fitted distribution of microcystin concentrations, given a Chl concentration. Directly estimating this exceedance frequency for a particular lake from field measurements would not be necessary to derive a criterion value. However, if states or authorized tribes were interested in estimating this quantity directly from field data, they could collect temporally intensive (e.g., daily) measurements of Chl *a* and microcystin. As described in detail in the criterion document, the functional form for Chl *a* vs. microcystin is based on the combination of several distinct relationships: (1) the linear relationship between Chl and total phytoplankton biovolume, (2) the piecewise linear relationship between cyanobacterial biovolume and microcystin concentration, and (3) the quadratic relationship between Chl and relative cyanobacterial biovolume. The upper edge of the distribution is estimated as an upper percentile (e.g., 90th) of the distribution of possible microcystin concentrations.

Category 3.16 Analysis: Deepwater Hypoxia – Stratification

Commenters: Arizona Department of Environmental Quality; Ecosystem Consulting Service, Inc; R.T. Angelo and J. DeLashmit; Wyoming Department of Environmental Quality; Montana

Department of Environmental Quality; Iowa League of Cities; City of Springfield, Missouri; Metropolitan St. Louis Sewer District; Wisconsin Department of Natural Resources; New Jersey Department of Environmental Protection; and Florida Department of Environmental Protection.

Comment synopsis: Commenters suggested that EPA use relative thermal resistance (RTRM) to measure stratification rather than a simple temperature gradient, and they suggested that the geometry ratio is not appropriate for dendritic reservoirs. Commenters also asked about the effects of other factors on stratification, including controlled releases from reservoirs, wind, and variations in stratification strength away from the midpoint of reservoirs. Commenters also noted that the model for the first day of stratification described in Demers and Kalff (1993) accounted for less variability when mean annual temperatures exceeded 10 degrees Celsius.

Commenters also highlighted the availability of predictions of lake temperature profiles in other resources (Winslow et al. 2017) and asked EPA to clarify temperature requirements in the metalimnion versus epilimnion. Commenters also suggest that other measures of habitat availability may be more appropriate.

Response: EPA acknowledges that RTRM provides an alternate measure of stratification strength. However, a broadly accepted threshold of 1 °C/m was available for identifying stratified conditions using a simple temperature gradient. Furthermore, the temperature gradient and the geometry ratio were used only to identify a subset of profiles that were used to fit the model; they are not used directly when calculating the criterion values. The applicability of the geometry ratio to dendritic reservoirs may be weak, but the subsequent data screen based on the temperature profile would then identify stratified profiles for analysis from these systems. Other factors, such as controlled releases and wind events, can influence stratification strength, which in turn can potentially alter the pattern of dissolved oxygen depletion over the summer. In these cases, EPA reminds states and tribes that the criterion models can be updated with site-specific information and then used to derive criteria that apply specifically to particular reservoirs and lakes. Other scientifically defensible approaches can also be used. In lakes with complex shapes, states and authorized tribes may want to derive criteria for different discrete segments of the reservoir.

Commenter observations about the increased variability of the Demers and Kalff (1993) relationship above 10 degrees Celsius are correct, but this increase in variability was attributed by Demers and Kalff to monomictic reservoirs. Because the current models are fit using data from dimictic reservoirs, these issues may be less important. Furthermore, the parameters estimated by Demers and Kalff were used in the current model only as prior distributions, which were then adjusted by empirical data, diluting the direct influence of these initial estimates on the final models. Finally, as noted in Yuan and Jones (2020), the initial stratification date estimated in these models is more strongly linked to the reduction of oxygen

transport from surface layers to deeper waters, which may account for differences with a stratification date based solely on temperature profiles.

The data on temperature profiles in Midwest reservoirs is extensive, but not matched with corresponding data on lake water chemistry; therefore, at this time it cannot be incorporated into the present models.

The hypoxia model is based on release of temperature restrictions in the epilimnion. EPA assumed that, in general, metalimnion temperatures are less than epilimnion temperatures. Other measures of available habitat in the deep lake water may have other advantages, but the simplicity of refuge depth provides for a broad applicability of the approach and the incorporation of a greater variety of data sets.

Category 3.17 Analysis: Microcystin Model – Other Cyanotoxins

Commenters: Florida Department of Environmental Protection.

Comment synopsis: Commenters noted that other cyanotoxins (e.g., cylindrospermopsin, anatoxin, saxitoxin) may be important.

Response: EPA acknowledges the importance of other cyanotoxins, but at the present time, microcystin data are the only data available to develop national models, as noted in the criterion document. Furthermore, microcystin is one of the most commonly observed cyanotoxins, and so, managing nutrients to control for microcystin can be an effective initial step to controlling all cyanotoxins. At this time, EPA has published recommended health advisory values and swimming advisory values for cylindrospermopsin that are protective of drinking water sources and recreational designated uses, respectively. If a state or authorized tribe has sufficient monitoring data, then it may be possible to derive nutrient criteria to protect against the formation of cylindrospermopsin; however, it would only be necessary if it is demonstrated that the formation of cylindrospermopsin is more sensitive to nutrients than the formation of microcystin. At this time, EPA has not published protective values for anatoxin and saxitoxin, so if a state or authorized tribe wanted to derive nutrient criteria to protect against their formation, an analysis of what it means to be protective of the designated use(s) would be needed in addition to collecting monitoring data to derive nutrient criteria.

Category 3.18 Analysis: Microcystin Model – Detection Limit

Commenters: Texas Commission on Environmental Quality.

Comment synopsis: Large numbers of below-detection-limit measurements impede the ability to develop useful models.

Response: Censored data were considered in depth during preliminary analyses. After this work, as described in the criterion document, EPA determined that a negative binomial distribution accurately represented the distribution of microcystin concentrations, including values that were below detection limits, and the model of microcystin using this distribution accurately accounted for variations in observed data. Additional analysis supporting this approach is described in detail in a supporting reference (Yuan and Pollard 2017).

Category 3.19 Analysis: Microcystin Model – Threshold Selection

Commenters: Wyoming Department of Environmental Quality; R.T. Angelo and J. DeLashmit; Iowa League of Cities; Alliance for the Great Lakes; City of Springfield, Missouri; Metropolitan St. Louis Sewer District; Missouri Department of Natural Resources; Iowa Environmental Council; Wisconsin Department of Natural Resources; National Association of Clean Water Agencies; New Jersey Department of Environmental Protection; Florida Department of Environmental Protection; and Counsel, Mississippi River Collaborative et al.

Comment synopsis: With regard to the microcystin threshold for drinking water source, different commenters suggested either that the adjustments for drinking water treatment should not be allowed and the drinking water microcystin threshold applied in source waters (in the case of a malfunction in the treatment plant), or that the microcystin threshold should never be applied in source waters. Commenters also asked for more explicit recommendations regarding the effects of treatment on microcystin concentrations. Commenters also noted that Chl *a* criteria calculated from the model were less than Chl *a* concentrations in certain drinking water sources and lower than existing criteria in some states, and they concluded that the proposed criterion models were too conservative. With regard to the microcystin threshold for recreation, commenters asked that the criterion document text be clarified to indicate that the 8 µg/L threshold applied for incidental ingestion in children and asked that the R Shiny app highlight this threshold as the recommended value. For both the drinking water and recreational microcystin thresholds, some commenters raised issues with how those thresholds were derived. Others suggested that the criterion document text be edited to emphasize that the same model applies to both recreation and drinking water source protection, and that EPA allow values higher than 8 µg/L for the microcystin threshold. Commenters also observed that a wide range of Chl *a* criterion can be derived based on management choices of exceedance frequency and credible interval—and therefore the Chl *a* criteria based on recreation should be used in support of aquatic life criteria. Finally, commenters suggested that the criteria models account for duration and frequency components specified in the drinking water health advisory and recreational criteria.

Response: Several studies have quantified a range of possible effects of conventional drinking water treatment technologies on microcystin concentration. In reviewing these studies, EPA determined that, as noted in the criterion document, the variety of effects precluded EPA from

identifying a single value for the effect of drinking water treatment. The references included in the criterion document provide a representative sample of the types of studies that have quantified the treatment effect, and EPA suggests that states refer to these and other studies in determining how to adjust the microcystin threshold. Some water quality treatment systems also may not be effective in removing microcystin, and EPA suggests that states consider this possibility when making risk management decisions. Comparisons between the present criterion models and existing state criteria are difficult because of the differences in the methods used to derive the criteria. Similarly, observing that a particular lake has Chl *a* concentrations exceeding a certain threshold while having no episodes of elevated microcystin does not inform discussions of the overall validity of the model. Data from a single lake can be incomplete or anomalous relative to the overall mean trends. In lakes with long data records, adjusting the current model using lake-specific data is also possible.

Editorial comments to the text regarding the description of the recreational and drinking water models have been considered and incorporated where appropriate.

Comments regarding the analysis to derive health advisory and recreational thresholds for microcystin are not within the scope of the current effort. These thresholds have undergone their own public comment process; EPA suggests that commenters refer to those documents for further details regarding these questions.

Microcystin threshold values higher than the recreational threshold are not allowed on the R Shiny app. As other commenters have noted, the exceedance frequency and credible intervals provide the mean for risk management decisions to be incorporated in the criterion derivation. The magnitude of the threshold concentration for microcystin is not viewed as a risk management decision. Similarly, the wide range of possible Chl *a* concentrations for a given microcystin threshold highlights the importance of public participation in the risk management decisions used to derive criteria, and the R Shiny app provides an important tool for communicating the meaning and effect of these decisions to stakeholders.

Frequency and duration components for microcystin recommended by the health advisory and by the recreational criteria can be incorporated by using the exceedance frequency parameter, and, if necessary, additional analysis as specified in the appendix of the criterion document.

[Category 3.20 Analysis: Microcystin Model – Other Covariates and Measures](#)

Commenters: Ecosystem Consulting Service, Inc; Massachusetts Department of Environmental Protection; Iowa League of Cities; Water Environment Association of Texas and the Texas Association of Clean Water Agencies; Idaho Department of Environmental Quality; Wisconsin Department of Natural Resources; New Jersey Department of Environmental Protection; Iowa

Department of Natural Resources; Florida Department of Environmental Protection; Counsel, Mississippi River Collaborative et al.; and California Water Boards.

Comment synopsis: Commenters suggested that other factors might influence concentrations of microcystin, including weather, lake depth, turbidity, presence and expression of microcystin producing genes, TN:TP, and stratification. Other commenters observed that phycocyanin is more closely associated with cyanobacteria biovolume and microcystin concentration. Commenters also asked how spatial variability within a lake affects criterion values. Commenters also suggested that the relative abundance of cyanobacteria is not always correlated with Chl and that all cyanobacteria do not produce microcystin.

Response: EPA acknowledges that many other factors can influence microcystin concentration, but notes that these factors usually do not affect the relationship between cyanobacterial biovolume and microcystin. For example, some studies have observed that high microcystin concentrations are more likely to be observed in stably stratified lakes. One explanation for this trend is that stably stratified systems are well-suited for cyanobacteria, and higher cyanobacterial biovolumes are expected under these conditions. Then, higher concentrations of microcystin are possible. In this case, the stable stratification affects the overall biovolume of cyanobacteria, but it does not affect the relationship between cyanobacterial biovolume and microcystin, which is the focus of the criterion model. A similar logic applies to other covariates such as lake depth and weather. Focusing on the relationships represented in the criterion model (i.e., Chl -> phytoplankton biovolume, Chl -> relative biovolume of cyanobacteria, biovolume of cyanobacteria -> microcystin) helps identify specific factors that may influence the model.

The comment that some cyanobacteria do not produce microcystin is correct. The use of cyanobacterial biovolume as a predictor variable in the criterion model acknowledges the fact that scientific understanding is still incomplete regarding which cyanobacterial species produce microcystin. Indeed, the potential for producing microcystin varies within a given species, and predicting the conditions under which a species produces microcystin is the subject of active research. Using the biovolume of all cyanobacteria eliminates the bias that would be introduced by the current incomplete understanding of which species produce microcystin. Furthermore, use of the biovolume of all cyanobacteria retains enough data to permit a robust estimation of the statistical relationship. This topic is addressed in detail in an article in an external journal (Yuan and Pollard 2019).

The increased measurement of phycocyanin during routine monitoring is encouraging, but at the time of this analysis, national data on phycocyanin was not available.

Cyanobacterial biovolume and microcystin concentration are often patchily distributed across a lake, but data were not available to characterize the magnitude of this variability. This

patchiness can be taken into account as another factor for risk management, and, as data accumulates, spatial variability across a lake may be quantified in the future.

Category 3.21 Analysis: Phosphorus, Chlorophyll Model – General Comments

Commenters: Montana Department of Environmental Quality, Iowa League of Cities, Idaho Department of Environmental Quality, Wisconsin Department of Natural Resources, and New Jersey Department of Environmental Protection.

Comment synopsis: Commenters expressed concern about the possibility that credible intervals selected for different models would propagate error through the models. Commenters also noted the variability in TP concentration about a given Chl *a* concentration and expressed concern about the accuracy of the resulting TP criteria. Commenters also asked for more information regarding the application of limiting and ambient criterion values.

Response: One of the advantages of the Bayesian networks used in the criterion models is that they propagate error among different modeled relationships. For example, in the model relating Chl *a* to microcystin, uncertainty in estimating cyanobacteria biovolume from Chl *a* is propagated to the relationship estimating microcystin concentration from cyanobacteria biovolume because all relationships in the network are estimated simultaneously. Specifying a credible interval on which to base the Chl *a* criterion then accounts for uncertainty in the estimated network of relationships. When computing a TP criterion corresponding to a Chl *a* threshold, uncertainty in the network of relationships linking TP to Chl *a* is taken into account by specifying a second credible interval. Because the function of Chl *a* in both of these networks is that of a predictor variable (i.e., plotted on the x-axis of bivariate plots), specifying two separate credible intervals is appropriate. Residual error in both cases is associated with the response variable (microcystin concentration in the first model and TP in the second model). Hence, application of separate credible intervals on each of the models does not compound errors.

As noted in the criterion document and in associated journal article, the TP-Chl *a* models provides accurate predictions of TP, with root mean square (RMS) error for \log_e transformed TP of 0.52. Variability in TP that appears in bivariate plots between Chl *a* and TP does not account for variations in non-phytoplankton sediment among samples. Mean and variance estimates of non-phytoplankton sediment within an ecoregion can provide an accurate range of estimates for this parameter, which can then be translated to accurate estimates of TP.

Limiting criterion values for TP can be thought of as the minimum possible TP concentration for a given Chl *a* concentration. These values may be useful for setting load limits, but further analysis would be needed. The ambient criterion values predict TP concentrations that are most likely to be observed, and therefore, are comparable to monitored values.

Category 3.22 Analysis: Phosphorus, Chlorophyll Model – Other Covariates

Commenters: Iowa League of Cities, Missouri Department of Natural Resources, Texas Commission on Environmental Quality, and Florida Department of Environmental Protection.

Comment synopsis: Commenters suggested many other covariates that might influence the TP-Chl relationships, including color, alkalinity, residence times, lake productivity, dissolved phosphorus, and stratification status. Commenters also suggested that internal loading of phosphorus needed to be considered in the model.

Response: The TP-Chl model expresses the relationship between TP and Chl observed within a water sample; viewing the model with respect to this sampling unit can help identify covariates that are likely to influence the relationship. For example, internal loading of phosphorus influences the total amount of phosphorus that is introduced to a lake per unit time, and within a water sample, internal loading may add to dissolved phosphorus compartment of TP. In the criterion model, this contribution is taken into account via the term for dissolved phosphorus. Other factors, such as residence time, influence how much of a particular load of phosphorus is taken up by phytoplankton, but at the unit of a water sample, residence time effects have already been actualized. That is, in the water sample, the effects of residence time have already occurred in terms of changes in the relative proportions of dissolved phosphorus and phytoplankton-bound phosphorus. The underlying relationship between phosphorus bound within phytoplankton and Chl concentration does not change when residence time changes, nor does it change due to variations in the values of other covariates such as stratification status or color.

The approach for identifying relevant covariates to the TP-Chl model followed EPA's 2010 stressor-response guidance in that a conceptual model identifying the different components of TP was specified, and variables that informed quantification of different nodes in the conceptual model were then included in the model.

Category 3.23 Analysis: Nitrogen, Chlorophyll Model – General Comments

Commenters: Wyoming Department of Environmental Quality, Iowa League of Cities, Missouri Department of Natural Resources, Idaho Department of Environmental Quality, Wisconsin Department of Natural Resources, and Florida Department of Environmental Protection.

Comment synopsis: Commenters requested more detail regarding the difference between TN and TN-dissolved inorganic nitrogen (DIN), particularly with regard to the resulting criterion values, and asked for clarification regarding a numeric value for “near-zero” DIN. Commenters also asked about the difference between the ambient and limiting criterion values. Commenters asked about the joint effects of TN and TP, suggesting specifically that EPA refer to

papers by Dolman, consider the potential confounding effect of TP on the estimated TN relationship, and examine the utility of examining the ratio between TN and TP in predicting cyanobacterial dominance. Commenters also suggested that the residual variability of TN was too large in bivariate plots.

Response: More explanation has been added to the criterion document regarding the difference between TN and TN-DIN. As with recommended TP criteria, limiting criterion values for TN can be thought of as the minimum possible TN concentration for a given Chl a concentration. These values may be useful for setting load limits, but further analysis would be needed. The ambient criterion values reflect TN concentrations that are most likely to be observed and are therefore comparable to monitoring data.

Papers by Dolman and coauthors describe an alternate approach for modeling relationships between Chl a , TN, and TP, focusing on Chl a as the response variable. As described in the criterion document, the relationships between Chl a and TN and between Chl a and TP in the present analysis focuses on modeling the components of TN and TP in the water column, and one of these components is nitrogen and phosphorus bound within phytoplankton. As such, the model approach is very different from that proposed by Dolman, so comparisons are difficult. The potential confounding effect highlighted by some commenters is based on the observation that TN and TP are strongly correlated in monitoring data. Based on the current analysis, this correlation is expected because both nitrogen and phosphorus are present within phytoplankton. Hence, changes in phytoplankton biovolume yield concurrent changes in measured TN and TP and give rise to the correlation between the two measurements. Viewed in this way, the correlation between the two parameters does not confound estimated relationships. When variations in TN and TP are viewed in the context of this conceptual model, the utility of examining the ratio between TN and TP is also limited, because much of the variations in this ratio may be driven by factors that are not related to phytoplankton, such as the amount of inorganic sediment (for TP) and the amount of dissolved organic nitrogen (DON) (for TN). Some potential for examining the differences in nitrogen:phosphorus stoichiometry among major phytoplankton groups may be possible, but this work is not within the scope of the current effort.

As noted in the criterion document, the TN-Chl a models provides accurate predictions of TN, with RMS error for \log_e transformed TN of 0.37. Variability that appears in bivariate plots between Chl a and TN does not account for variations in DON among samples; therefore, the residual variability of TN visually may appear to be large.

Category 3.24 Analysis: Model Statistics

Commenters: Wyoming Department of Environmental Quality; Massachusetts Department of Environmental Protection; North Carolina Water Quality Association; Oregon Department of

Environmental Quality; Water Environment Association of Texas and the Texas Association of Clean Water Agencies; Water Quality, National Council for Air and Stream Improvement, Inc.; Wisconsin Department of Natural Resources; National Association of Clean Water Agencies; Iowa Department of Natural Resources; Florida Department of Environmental Protection; and Colorado Department of Public Health and Environment.

Comment synopsis: Commenters requested more information regarding the accuracy of the criterion models, including a definition of RMS error, RMS values by ecoregion, a table comparing RMS among different models, and a discussion of the accuracy of models fit to ecoregions with small numbers of samples. Commenters also expressed concern regarding the residual variability and uncertainty in estimates of mean relationships, suggesting that high uncertainty and residual variability rendered models unusable and required that safety factors be added to models, and that basing criterion values on a low percentile of the credible interval (e.g., 25th percentile) implies that the remaining 75 percent of sites are overprotected. Commenters also suggested that EPA consider other diagnostic tests and statistics to characterize model performance, including sensitivity analysis, comparison to lake-specific data, comparison to existing state criteria, examination of Type I and Type II error, and tests of statistical significance. Commenters also suggested that EPA consider other broad issues regarding model performance, suggesting that models are too “black box,” that the relationship between nutrient concentration and Chl *a* should be a straight line, that different model assumptions need validating and ground truthing, and that Bayesian machine learning is not an appropriate statistical technique to use for this application. Commenters also wondered why bivariate relationships shown in the criterion document are stronger than observed in their own data, and they suggest that EPA consider threshold responses.

Response: EPA notes that RMS is a commonly used measure of model prediction accuracy, quantifying the average difference between predicted and observed values. Hence, RMS provides an ideal statistic for characterizing model performance. The units of RMS values are the same as the units of observed response values, and, due to differences in the responses among models, comparisons of RMS among models would not be meaningful. Therefore, a table of RMS values, as requested by the commenter, has not been added. The TN-Chl *a* and TP-Chl *a* models estimate different model coefficients for different ecoregions, and as described in the document, ecoregion-specific model coefficients are estimated in a hierarchical model structure, such that estimates for ecoregions with small numbers of samples can “borrow” statistical strength from other ecoregions, improving the overall performance of the model. Additional data collection with some ecoregions or states can further improve the performance of the model.

Ecological measurements are variable due to temporal variability and measurement errors, and the response variables modeled in the criterion document manifest this variability. Hence, residual variability is expected in the criterion models due to these factors, plus the potential

effects of unmodeled predictors. By quantifying model uncertainty, EPA has provided the means to account for this uncertainty when calculating criterion values. Indeed, the transparent display of the effects of model uncertainty can enhance efforts to communicate to stakeholders regarding the effects of management decisions to address uncertainty. Safety factors, if needed, can be added, but the estimates of credible interval provide the means for directly accounting for model uncertainty. The comment suggesting that basing criterion values on the lower bound of the credible interval results in over-protection reflects a lack of understanding of the meaning of the credible interval. The credible interval expresses uncertainty regarding the position of the estimated mean relationship and provides no information regarding the proportion of sites that the criterion value applies to. That is, a criterion based on the 25th percentile of the posterior distribution of possible mean values does not imply that 75 percent of sites are overprotected. Instead, the selection of this percentile expresses the uncertainty of the estimate of the mean relationship between stressor and response.

The RMS statistics provided in the criterion document quantify the predictive performance of the model, a key statistic for characterizing model accuracy. Other statistics recommended by commenters are only tangentially related to model performance. For example, significance tests are used in controlled studies to test whether the application of a treatment yields a change in the response that is greater than expected due to random variability. In the present case, known relationships among variables are modeled; therefore, significance tests provide no additional information. That is, the intent of the criterion models is not to test whether or not elevated nutrients have an effect on the response that is greater than expected due to random variability. Similarly, sensitivity tests are typically applied to mechanistic models that include many user-specified parameters, and the tests reveal which of the parameters have the greatest effect on model predictions. In the present models, the number of parameters is far less than used in mechanistic models, and these parameters quantify effects that have been identified through exploratory analysis to be strong. Hence, sensitivity tests would be uninformative. Type I and Type II errors are most commonly discussed in terms of medical diagnostic tests, in which a diagnostic test provides information regarding the likelihood of a certain illness, and the presence or absence of the illness can be independently confirmed. In the present analysis, the criterion models are used to establish values that protect against impairment from excess nutrients rather than to diagnose a condition, and hence, the relevance of Type I and Type II errors is not apparent. Finally, a comparison of criterion model results to existing criteria is difficult, given that different methods were used to derive the different criteria. When data for comparing derived criteria to performance in individual lakes are available, updating the existing models with lake-specific data would be recommended.

EPA does not agree with comments that suggest that the models are too “black box.” All model equations are explicitly described in the criterion document, and bivariate plots are shown of all observed data with modelled relationships. Comments that Bayesian machine learning are not appropriate are inaccurate because Bayesian network models are not the same as Bayesian

machine learning. Commenters that suggested that the relationship between Chl and nutrients should be a straight line agree with the model formulation, as the relationship the models assume that the relationships is a straight line in log-log space.

With regard to threshold responses, the Vitense et al. (2017) article recommended by one commenter discusses changes between alternate stable states in lakes and requires further work to apply at the broad spatial scale of the current analysis.

As indicated in the text, the observed data shown in Figure 21 has been binned and averaged, which may explain the stronger bivariate relationships noted by one commenter.

Category 4 Characterization

Category 4.1 Characterization: Incorporating State Data – Need More Guidance

Commenters: Wyoming Department of Environmental Quality; Association of State Drinking Water Administrators; Iowa League of Cities; Idaho Power Company; Water Quality, National Council for Air and Stream Improvement, Inc.; Idaho Department of Environmental Quality; Wisconsin Department of Natural Resources; Florida Department of Environmental Protection; Nevada Division of Environmental Protection; and Oregon Lakes Association.

Comment synopsis: Commenters asked that EPA clarify when incorporation of state data is recommended, asked that EPA provide a more user-friendly way to incorporate data, asked that more detailed guidance be provided regarding the process for incorporating state data, asked for details regarding mechanisms for requesting EPA assistance in using state data, and asked for more case studies. Commenters also asked for details regarding additional assumptions associated with combined national-state models. Commenters asked whether national models and national-state models have precedence over existing site-specific models. With regard to the Iowa microcystin case study, one commenter had a specific question regarding the Chl *a* thresholds derived in the study.

Response: Details regarding the use of state data in conjunction with the national models have been added to the criterion document, as well as relevant questions and answers. Assumptions inherent to combining state and national data are also described in the criterion document. State data sets differ tremendously in the variables that were measured, the lab and field techniques, and the structure of the database in which they are stored. For these reasons, providing user-friendly tools for incorporating state data is not possible. Instead, EPA is ready to provide assistance to states and authorized tribes that are interested in incorporating their data in the national models. As these case studies are complete, they will be documented and made available.

As with any CWA Section 304(a) criteria, national criterion models and associated criteria are recommendations, and states can derive criteria with other scientifically defensible methods that are protective of their designated use(s). Ongoing site-specific work to derive criteria can use information and insights from the national criteria to enhance site-specific derivation efforts.

One commenter noted the difference in microcystin concentrations associated with Chl *a* in Figure 35 and Figure 36 of the Iowa microcystin case study. Differences in the predicted microcystin concentrations are expected because Figure 35 plots mean microcystin concentrations, whereas Figure 36 plots instantaneous, grab sample concentrations.

Category 4.2 Characterization: Incorporating State Data – General Comments

Commenters: Ecosystem Consulting Service, Inc; Wisconsin Department of Natural Resources; and Oregon Lakes Association.

Comment synopsis: Commenters were concerned about the coarseness of the ecoregions used in some of the criterion models and recommended the use of regional data in deriving criteria.

Response: The commenter notes differences between two neighboring lakes in Oregon that are distinct in terms of nutrient loads and trophic status. EPA notes that many aspects of lake location and setting are taken into account by the hypoxia model, and that ecoregions only influence particular terms in the TN-Chl *a* and TP-Chl *a* model. In cases in which a lake is particularly unique relative to other lakes in its vicinity, state and authorized tribes may want to consider site-specific criteria.

Category 4.3 Characterization: Limitations and Assumptions

Commenters: Nevada Division of Environmental Protection, Iowa Department of Natural Resources, and Florida Department of Environmental Protection.

Comment synopsis: Commenters expressed concern that the draft criterion models did not account for naturally eutrophic systems. Commenters also asked whether variability in model coefficients at smaller spatial scales (e.g., Level IV ecoregions) was examined.

Response: The criterion models focus on support of designated uses and do not attempt to characterize natural expectations for different lakes and reservoirs. Naturally eutrophic waters, where endpoint values differ from those specified as attaining designated uses, would be potential candidates for site-specific criteria.

The density of NLA samples did not support consideration of spatial variability at a resolution below that of the Level III ecoregions. The effects of these finer grain ecoregion specifications

can be considered if appropriate monitoring data are available at the state level. The methods for estimating coefficients for different Level III ecoregions is described in detail in the criterion document, and this hierarchical approach is well-suited for examining even finer scale classifications.

Category 4.4 Characterization: Duration and Frequency

Commenters: Wyoming Department of Environmental Quality; Association of State Drinking Water Administrators; Ecosystem Consulting Service, Inc; Massachusetts Department of Environmental Protection; Virginia Association of Municipal Wastewater Agencies; Louisiana Department of Environmental Quality; Iowa League of Cities; North Carolina Water Quality Association; City of Springfield, Missouri; Metropolitan St. Louis Sewer District; Oregon Department of Environmental Quality; Idaho Department of Environmental Quality; Wisconsin Department of Natural Resources; National Association of Clean Water Agencies; New Jersey Department of Environmental Protection; Florida Department of Environmental Protection; Iowa Department of Natural Resources; and Colorado Department of Public Health and Environment.

Comment synopsis: Commenters asked for clarification on the use of seasonal mean values for duration component and the specification of a not-to-exceed value for the frequency component of the recommended criteria. Commenters suggested that the use of seasonal means does not account for impacts that occur on shorter time scales (e.g., a temporary loss of deep water refugia). Commenters also asked how NLA data collected during the summer could be used to infer year-round conditions and how variations in growing season length in different locations were considered. Finally, commenters had specific questions regarding the use of arithmetic versus geometric means and regarding how multiple samples within individual lakes were handled in the analysis.

Response: The recommended duration of Chl a , TN, and TP criteria is a seasonal geometric mean value. As described in the criterion document, this recommendation arises from the sampling design of the NLA, which collected samples during the summer, and from an understanding of the timescale at which the effects of increased nutrients are manifested. States are encouraged to examine the appropriateness of this recommendation in their regions, as the length of the growing season can vary strongly depending on location. Nutrient concentrations in southern latitudes may be best characterized by longer sampling windows, whereas in northern latitudes, shorter sampling windows may suffice. The duration of the Chl a , TN, and TP criteria is distinct from that of the effects, which vary by endpoint. For example, the microcystin model links criterion values to the probability of an exceedance of the specified threshold; therefore, the duration associated with microcystin concentration is very short (e.g., a single day). Durations for the other endpoints differ as well.

The criterion document recommends the use of the frequency of exceedance of the Chl a , TN, and TP criteria with estimates of within-lake variability to calculate operational criteria (Appendix D in the criterion document). EPA is aware that interannual sampling variability can be high, as is sampling variability associated with estimating seasonal means from a small number of samples in any single lake. To that end, EPA recommends that states conduct additional analysis of their monitoring data to estimate interannual and sampling variability and incorporate those values into their adopted criteria. The appendix to the criterion document provides details on this process, and the text has been revised to re-emphasize this point.

A geometric mean is the appropriate statistic to characterize the central tendency of a log-normally distributed distribution. Log-normally distributed values are frequently observed for environmental variables such as Chl a , TP, and TP. However, in certain locations, the temporal variability of these measurements may be less variable and can be modeled with a normal distribution. In these cases, arithmetic means are appropriate as well.

Multiple samples were collected in a subset of lakes in the NLA. Each of these samples was included separately in the statistical analysis, and these repeat measurements were used to characterize the within-lake variability of each measurement.

Category 5 Implementation

Category 5.1 Implementation: Management, Most Sensitive Use, and Other Issues

Commenters: Footprints in the Water, LLC; Alliance for the Great Lakes; North Carolina Water Quality Association; City of Springfield, Missouri; Metropolitan St. Louis Sewer District; Missouri Department of Natural Resources; Federal Water Quality Coalition; Association of Missouri Cleanwater Agencies; American Fisheries Society et al.; American Water Works Association; Oregon Department of Environmental Quality; Idaho Department of Environmental Quality; New Jersey Department of Environmental Protection; Iowa Department of Natural Resources, Maryland Department of the Environment; NEIWPC; Coalition of Greater Minnesota Cities; Missouri Public Utility Alliance; Colorado Department of Public Health and Environment; California Water Boards; Wyoming Department of Environmental Quality; Virginia Association of Municipal Wastewater Agencies; Mississippi River Collaborative et al.; and Minnesota Environmental Science and Economic Review Board.

Comment synopsis: summary

EPA received numerous comments urging it to publish guidance on how states and authorized tribes can implement numeric nutrient criteria that are developed using its CWA Section 304(a) recommendations and subsequently adopted under Section 303(c). Specifically, the comments expressed the need for clarification on how the Section 304(a) recommendations affect and

interact with other state CWA authorities, as well as more detailed procedures that can guide states' implementation of the CWA regulations (Table 1).

Table 1. CWA regulations referred to by commenters.

Examples of issues raised	CWA program	CWA section	Federal regulation(s)
<ul style="list-style-type: none"> • Protection of all designated uses, not just the most sensitive use • Antidegradation policies • Variance procedures • Triennial reviews 	Water quality standards	303(c)	<ul style="list-style-type: none"> • 40 CFR § 131.11 • 40 CFR § 131.12 • 40 CFR § 131.14 • 40 CFR § 131.20
<ul style="list-style-type: none"> • Monitoring methods 	Water quality monitoring	303(d) 305(b)	<ul style="list-style-type: none"> • 40 CFR § 130.4
<ul style="list-style-type: none"> • Data requirements for listing and delisting decisions 	Water quality assessments	303(d) 305(b)	<ul style="list-style-type: none"> • 40 CFR § 130.8
<ul style="list-style-type: none"> • Pollutant source identification and tracking • Model development 	TMDLs	303(d)	<ul style="list-style-type: none"> • 40 CFR § 130.7
<ul style="list-style-type: none"> • Averaging period calculations 	National Pollutant Discharge Elimination System (NPDES)	402	<ul style="list-style-type: none"> • 40 CFR § 122
<ul style="list-style-type: none"> • Statewide nutrient pollution reduction actions/assistance 	Nonpoint source pollution	319(h)	Not applicable
<ul style="list-style-type: none"> • Statewide nutrient pollution reduction actions/assistance 	Water quality assistance grants	106	Not applicable

In addition to the statutory and regulatory guidance sought by commenters, EPA also received comments asking for guidance that (1) supports communicating the CWA Section 304(a) recommendations to stakeholders (e.g., using the recommendations as a way of communicating the risks that nutrient pollution poses to designated uses), and (2) describes the financial costs to point source dischargers in implementing and complying with nutrient water quality standards based on the 304(a) recommendations. Commenters also asked that implementation guidance be published either before the end of the comment period or before the publication of the final recommendations.

Response: EPA recognizes that states and authorized tribes have statutory and regulatory commitments under the CWA to implement water quality standards they adopt under Section 303(c). As such, EPA will publish and request comments on draft implementation guidance for these 304(a) nutrient criteria recommendations in the near future.

Category 6 Supporting State Criteria

Category 6.1 Supporting State Criteria: Candidate Criteria

Commenters: Wyoming Department of Environmental Quality, Massachusetts Department of Environmental Protection, North Carolina Water Quality Association, Oregon Department of Environmental Quality, Idaho Department of Environmental Quality, Wisconsin Department of Natural Resources, Kansas Department of Health and Environment, National Association of Clean Water Agencies, New Jersey Department of Environmental Protection, Maryland Department of the Environment, NEIWPC, and North Carolina Department of Environmental Quality.

Comment synopsis: Commenters expressed various concerns regarding the output from the models. This included the modeling tools, especially the level of information or guidance provided for those tools, such as incorporation of state data, reconciling the criteria generated by the tools with state existing criteria (e.g., durations and frequencies), and the life cycle of these values, variability in the value generated and balancing errors, the issue of competing uses or appropriateness of uses for which criteria are developed, and the variables that are produced as candidate criteria.

Response: Several commenters asked for expanded guidance, training, and implementation resources to help guide the selection of input values and application of the R Shiny app models. This included how to select input values and depth classes, adopt values for groups of sites, apply and update with state data, and print results or make notes on each run. EPA is preparing these additional resources in response to these requests, some of which will be included with the final document as a frequently-asked-questions section and others that will be made available in the form of future direct support via webinar, workshop, or direct analytical help. EPA acknowledges that meeting every user need is difficult and that not every technical resource need is even known. Interested states are encouraged to contact EPA if interested in additional support, especially for the incorporation of state data. There were a few related comments asking about sample collection to increase model precision and which temporal duration to use. Model precision is a state policy decision, so there is no single decision as it depends on each state's or tribal user's desired certainty. The temporal duration of the criteria was stated in the document as the growing season. Lastly, one commenter asked about the lack of code availability. The code was made available as supplemental material on regulations.gov

and is still available for interested parties. Please contact EPA if there is difficulty in accessing that code.

Other commenters asked about how to reconcile model outputs with existing state criteria, including duration and frequency and the life cycles of these values and whether or not they might change over time. The CWA puts the responsibility for adopting standards on states, which are approved by EPA. States should establish criteria for these standards using CWA Section 304(a) guidance, 304(a) guidance modified to reflect site specific conditions, or other scientifically defensible methods (40 CFR §131.11). In addition, the CWA requires from time to time that EPA update 304(a) criteria with the latest scientific knowledge. States with EPA-approved numeric nutrient criteria for lakes or reservoirs that meet these requirements should consult with EPA regarding the protectiveness of their criteria through the regular water quality standards triennial review process—this would include duration and frequency elements. Consistent with the regulation, states and authorized tribes have the flexibility to use 304(a) guidance modified to reflect site specific conditions or other scientifically defensible methods to establish criteria that are protective of designated uses.

With regards to variability in the values generated by the different models, specifically whether different input parameters result in different values or whether different inputs could be used by opposing stakeholders to justify different values, the same inputs produce the same model output. In addition, input conditions have to be justified scientifically. Most input conditions require simple measurements (e.g., DOC concentration or latitude and longitude) and are only marginally deliberative. Exceedance frequency and certainty level are deliberative, but no more so than policy decisions that states and authorized tribes make for a variety of applications; these should be reconciled among stakeholders using the approaches that states and authorized tribes already use in such instances.

Other commenters asked about how to weigh error risks across model output with a concern that the values produced may be “over-protective.” Over-protection and balancing assessment error are not water quality standards concepts. 40 CFR 131.11 (a) requires that states adopt those water quality criteria that protect the designated use. Such criteria must be based on sound scientific rationale and must contain sufficient parameters or constituents to protect the designated use. For waters with multiple use designations, the criteria shall support the most sensitive use.

Similarly, some commenters asked about the issue of competing uses, the appropriateness of some uses for which criteria are being developed, and whether historic uses should be considered. States and authorized tribes are responsible for adopting water quality standards (40 CFR 131.4) and, thus, the state or authorized tribe sets the uses pursuant to the requirements of the CWA and regulations (e.g., with regards to existing uses). Nothing in the

recommended criteria changes this. Moreover, CFR 131.11(a) requires that for waters with multiple use designations, the criteria support the most sensitive use.

Category 6.2 Supporting State Criteria: Training

Commenters: Association of State Drinking Water Administrators; Idaho Department of Environmental Quality; New Jersey Department of Environmental Protection; Iowa Department of Natural Resources; Florida Department of Environmental Protection; NEIWPC; Colorado Department of Public Health and Environment; California Water Boards; Tip of the Mitt Watershed Council; Texas Commission on Environmental Quality; American Fisheries Society et al.; American Water Works Association; and Water Quality, National Council for Air and Stream Improvement, Inc.

Comment synopsis: EPA received several comments requesting that it provide information that inform, guide, and justify certain model parameters, including the selection of credible intervals, slopes for the zooplankton model, and alternative numeric values for a chosen assessment endpoint (e.g., microcystin concentration). Other comments expressed an interest in EPA extending hands-on training and conducting technical outreach with state water quality staff working on nutrient criteria development as a way to build state technical capacity with the models, provide communication tools on the models, and guide integration of state water quality data into the models. In addition, some comments requested that EPA consult with state water quality agencies, providing information that can guide additional state sample collection and data generation at the scales appropriate for inclusion in EPA's models. Other comments raised concerns that the criterion document did not include sufficient technical documentation in terms of how to apply the model, with some comments requesting that EPA provide more detail on the sequential workflow associated with applying the model (e.g., more detailed descriptions of how to use the model visualizations to generate candidate numeric nutrient criteria), as well as a tool that allows states to compare candidate numeric nutrient criteria as a function of different inputs to the same model.

Response: EPA illustrated through its case study (Appendix A) an example of how state-specific data can be used to increase the accuracy of the models. Additionally, EPA created a platform through its model visualizations (R Shiny apps) as a way for states to explore and apply the model in developing candidate numeric nutrient criteria.

In response to the comments, EPA has developed additional information that illustrates a more detailed sequential workflow when applying the model, including points in the criteria development process where states can make appropriate risk management decisions (see Appendix E). Additionally, EPA has revised the model visualizations to clarify the terms and improve the aesthetics and navigation of the visualizations. EPA has also included additional content to facilitate interpretation of the model results.

EPA considers the publication of the final recommendations as an important first step in supporting development of numeric nutrient criteria for lakes and reservoir by states and authorized tribes to adopt into their water quality standards. EPA is committed to providing the necessary technical support and outreach sought by states, as it has for the past 15 years, under the N-STEPS program. The N-STEPS program will be a key mechanism through which EPA can consult, partner, and collaborate with states and authorized tribes on applying all aspects of these recommendations for numeric nutrient criteria development. States and authorized tribes seeking EPA technical support in applying its recommendations through the N-STEPS program should contact their [EPA Regional Nutrient Criteria Coordinator](#).

Category 6.3 Supporting State Criteria: R Scripts

Commenters: Wyoming Department of Environmental Quality, Virginia Department of Environmental Quality, Association of State Drinking Water Administrators, Louisiana Department of Environmental Quality, Missouri Department of Natural Resources, Oregon Department of Environmental Quality, Kansas Department of Health and Environment, Iowa Department of Natural Resources, Colorado Department of Public Health and Environment, and North Carolina Department of Environmental Quality.

Comment synopsis: Commenters provided a number of requests related the R scripts, including providing the R code, or making it easier to access; providing the R Shiny code; providing a data dictionary that detailed the R packages needed, describing abbreviations; not using random seeds to make results consistent each time; making it easier to include state data; making the code generally easier and potentially like desktop spreadsheets; and then adding minor edits. Commenters also commented on the R Shiny app models, including providing a glossary; providing a table of models for which waterbodies they apply to and the required inputs; clarifying the input data needs; providing units for each input; providing duration and frequency of each output; and clarifying the source of the data.

Response: The R code script was readily available in the docket as evidenced by the many commenters who ran and commented on the code; it will continue to be available. EPA can provide technical assistance to anyone having difficult accessing the code. The R code makes it clear which packages are needed, but EPA plans to provide updated support information with the final recommendations that clarifies much the information requested regarding packages and the other details requested vis-à-vis a data dictionary/glossary. The models run using random seeds that produce comparable results each time and they should not be fixed to provide the same results each time implying a determinism absent in the underlying models. EPA reaffirms that this code should be run by people with sufficient statistical expertise and facility with R coding. EPA can provide technical assistance in incorporating state or local data for those states and authorized tribes interested in doing so and who may be constrained by

available expertise. However, it is not possible to replicate this modeling in desktop spreadsheet software.

EPA is updating the R Shiny apps to provide many of the improvements users requested to better understand and select input conditions, understand and run the models, and interpret the output.

Category 6.4 Supporting State Criteria: Supporting Information for the Document

Commenters: Wyoming Department of Environmental Quality; Louisiana Department of Environmental Quality; Montana Department of Environmental Quality; North Carolina Water Quality Association; City of Springfield, Missouri; Metropolitan St. Louis Sewer District; Missouri Department of Natural Resources; and American Water Works Association.

Comment synopsis: Some commenters requested additional supporting information on the modeling. Specifically, some commenters requested a primer on Bayesian network models. Others requested the criteria be published as technical guidance rather than recommended criteria, arguing the selection of endpoints and criteria is a state-led process. Lastly, a commenter requested that EPA demonstrate the defensibility of the model relationships including, for example, the relationship between zooplankton and phytoplankton biomass, and air temperature and day of stratification, among many others.

Response: There are many excellent and well-developed textbooks that explain Bayesian modeling approaches much more thoroughly and appropriately than EPA could describe in this document, which applies this type of modeling approach for recommended criteria development. Moreover, explaining the approach requires that the reader already have training in statistical modeling. EPA encourages users to seek out the support of appropriately trained experts, including at EPA, to support efforts to incorporate their own data into these models as they would with any advanced model application (e.g., water quality models, toxicological testing, etc.).

In this criteria document, EPA has provided comprehensive technical support along with the recommended criteria values available through R Shiny apps for Chl *a*, TN, and TP to protect three different designated uses from the adverse effects of nutrient pollution. EPA acknowledges that criteria can be derived using a variety of assessment endpoints that have been shown to be sensitive to nutrient pollution; the commenter did not argue with the reasoning behind EPA's selection of sensitive assessment endpoints that have data available nationally, but noted that they would have selected different ones. Under 40 CFR part 131.11(b), states may choose to modify these criteria based on site-specific conditions or adopt alternative criteria using other scientifically defensible methods, including additional or alternate assessment endpoints.

Finally, EPA has demonstrated, either directly or by reference, the defensibility of all the relationships posited in its models. Users are referred to the extensive scientific record cited, including the scientifically peer reviewed manuscripts by EPA scientists involved in developing these recommended criteria, which provide even additional detail on the demonstrated defensibility for these relationships.

Category 6.5 Supporting State Criteria: Supporting Information for the R Shiny Apps

Commenters: Federal Water Quality Coalition and Colorado Department of Public Health and Environment.

Comment synopsis: Some commenters expressed a desire for greater detail on some of the R Shiny app windows, specifically background information for zooplankton and hypoxia models, guidance on units, as well as edits to the text to improve interpretation. There were also questions about critical temperature for taxa and the potential for input conditions to produce conditions outside the range for a state. Lastly, there was a comment regarding providing details on sensitivities, error likelihood, and uncertainty.

Response: EPA is updating the R Shiny app code to provide background information for the zooplankton page and to clarify language regarding units and interpretation. With regards to critical temperature, states and authorized tribes should use temperatures consistent with scientifically defensible guidance from regional fishery experts. EPA provided detailed citations about the scientifically defensible sources for the critical temperature values EPA selected as examples; state fishery expertise can and should be relied upon to provide more state-specific targets, as long as they are scientifically defensible. EPA notes that the chronic standard is more appropriate for the data structure of these models rather than the acute value. Lastly, it is possible that the combinations of depth, depth below thermocline, DOC, refugia depth, and dissolved oxygen threshold produce model results outside the experience of the models. In that case, a state should consider not using the endpoint model or updating the model with state-specific data that encompasses this population of lake conditions to expand the experience of the model.

Lastly, the comment regarding the models' sensitivities, error, and uncertainty likely reflects a misunderstanding regarding the types of models used. Such terms are typically applied to deterministic, mechanistic water quality models where input sensitivities and error likelihoods are important because such models are not probabilistic and true uncertainty, in the statistical sense, cannot be calculated. The models used here are inherently probabilistic; input parameters are sampled from statistical distributions based on observed error and uncertainty, and they produce output with estimated uncertainty. Moreover, the user can choose to select the level of uncertainty or confidence through the credible interval selection—something not generally reproducible in mechanistic models. Lastly, the estimates used were not intentional

or conservative, but demonstrative. Again, the user can select the credible interval base on risk tolerance and this is not set by the Agency a priori.

Category 6.6 Supporting State Criteria: Output Magnitudes

Commenters: City of Springfield, Missouri; Metropolitan St. Louis Sewer District; and Minnesota Environmental Science and Economic Review Board.

Comment synopsis: Some commenters expressed concern that the output magnitudes generated by the models were overly conservative or excessively over-protective. Specific examples from Missouri reservoirs were provided that indicate the hypoxia models would result in lower Chl criteria than site-specific criteria developed and approved. Another comment suggested using single grab samples and the most restrictive credible intervals yields overly protective endpoints for the “vast majority of lakes.” They further state that lakes that have “never experienced microcystins are presumed to have high levels.”

Response: EPA does not agree that the models generate values that are over-protective or under-protective—these are concepts that are value based. The models generate values that protect the assessment endpoints outlined in the risk hypotheses. States and authorized tribes have the flexibility to set values for credible intervals that adjust their risk tolerance and thus to the level of protection sought. EPA can also not comment on the “vast majority” of lakes since the commenters did not provide specific analytical results as to the extent of lakes presumably affected by the array of decisions feeding into any model. The example from Missouri some commenters provided suggested the values set are overly protective and noted that several reservoirs in the state support cool water fishes at higher Chl levels than the model produces. However, comparisons of criteria to observations at single lakes may not be informative because the criteria are based on long-term average relationships and do not account for inter-annual variability that may be represented in an individual lake dataset. In addition, the Jones et al. (2011) paper cited by the commenters (Jones, J. R., M. F. Knowlton, D. V. Obrecht, and J. L. Graham. 2011. Temperature and oxygen in Missouri reservoirs. *Lake and Reservoir Management* 27:173–182.) itself indicates in Figures 9 and 10 that the study reservoirs cannot meet the optimal cool-water target conditions at Chl concentrations above approximately 0.5 mg/L, consistent with the EPA models.

With regards to microcystin-based Chl targets, the recommended criteria document lays out a scientific argument for the incorporation of various assessment endpoints replete with detailed risk hypotheses linking the selected variables to nutrient pollution. Commenters did not address the scientific arguments that justified the selection of the endpoints. EPA acknowledges that the relationships of these endpoint variables to drivers is variable. EPA has provided defensible methods to account for and reduce much of this variability. EPA acknowledges that microcystin production for a given level of Chl varies as evidenced by the figures shown;

however, there is clear and widespread evidence in these data and the literature (detailed in the recommendation) that the risk of toxin production and concentration increases with Chl concentration. This variability is, in part, why states have the flexibility to set their own credible intervals that reflects their risk tolerance in light of this variability.

Category 6.7 Supporting State Criteria: Applicability Given Within-Lake Variability

Commenters: Virginia Department of Environmental Quality, Massachusetts Department of Environmental Protection, and Wyoming Department of Environmental Quality.

Comment synopsis: Some commenters expressed concern about the effects of intra-lake variability in specific predictors on the models. Specifically identified by several commenters was the effect of variations in lake depth, which depth ought to be used, and how spatial variability in depth ought to be considered in applications of the models by states and authorized tribes.

Response: The NLA study sampled lakes across a wide range of conditions, including those that capture a wide variety of the variability found within individual large lakes. Thus, much of the variability expected within single large lakes is incorporated into the models. Users applying these models should use predictor values derived in a way that is consistent with the way the NLA predictors in the models were derived—namely using NLA methods. For the zooplankton endpoint, for example, depth classes were based on NLA index site depth (the depth at the deepest point up to 50 meters in natural lakes or else the midpoint of the lake and the middle of reservoirs). For hypoxia, the only depth measures are the depth below the thermocline and depth of summer refugia; both of these are set by the user and are uninfluenced by within lake variability. The microcystin endpoint does not depend on depth. Lastly, only the TP model depends on depth, and that model once again used NLA-derived index site depth values, which is the value at the deepest point in the lake up to 50 meters in natural lakes or else the midpoint of the lake and the middle of reservoirs.

Category 6.8 Supporting State Criteria: Applicability – National to Site-Specific

Commenters: Wasatch Front Water Quality Council, Massachusetts Department of Environmental Protection, Upper Neuse River Basin Association, Federal Water Quality Coalition, Texas Commission on Environmental Quality, Idaho Department of Environmental Quality, Iowa Department of Natural Resources, Florida Department of Environmental Protection, and California Water Boards.

Comment synopsis: Commenters suggested that criteria should account for site-specific conditions and that a single sample did not sufficiently characterize individual lakes. Other commenters asked for further guidance on the application of criterion models applied at

different levels of aggregation. That is, some models specified candidate criteria for groups of sites (e.g., depth ranges for the zooplankton model) and other models specified criteria using lake-specific information (e.g., latitude/longitude for the hypoxia model). Commenters also suggested that EPA develop ecoregional-specific criteria to account for natural differences among Level III ecoregions, whereas some commenters suggested that variations at spatial scales that are smaller than Level III ecoregions need to be taken into account. Similarly, commenters suggested that grouping data by state political boundaries (as specified in the Iowa microcystin case study) did not sufficiently account for spatial variability.

Response: The criterion models account for site-specificity to different degrees because of differences in the mechanisms represented by the criterion models. In the case of the hypoxia model, lake water temperature is a critical parameter because it determines the timing of spring stratification and the release of water temperature constraints in the surface layer for cool- and cold-water fish. Because lake temperature varies strongly depending on lake elevation and geographic position, including site-specific information helped EPA develop a more accurate criterion model. Conversely, the model linking microcystin to Chl *a* concentration represents the relationship between phytoplankton biomass and the likelihood of different microcystin concentration within a water sample, and this relationship is relatively stable among different locations. Hence, site-specific information is less important to include in this model and in other models that address groups of lakes.

The recommended national criterion models do not preclude the adoption of site-specific criteria in lakes that have been studied intensively. The national models may also inform these site-specific efforts by providing a broader spatial context for lake specific criteria.

Comments noting that a single sample from each lake is insufficient to characterize conditions in that lake misunderstand the intent of the stressor-response approach. Stressor-response models are used to characterize the relationship between two variables (e.g., Chl *a* and microcystin), and this relationship is used to identify levels for one of the variables (e.g., Chl *a*) that is associated with a desired level for the second variable (e.g., microcystin concentration). The model is not intended to characterize conditions in a single lake; therefore, there is no need for multiple samples from every lake.

The criterion models incorporate ecoregional variations where they were found to account for substantial amounts of variability in the responses (i.e., the TN and TP models). In other models, ecoregional differences were not strong relative to the effects of other predictor variables. Variations at smaller spatial scales than Level III ecoregions may be important in certain regions, and incorporating these variations by examining more spatially intensive state monitoring data sets is possible, given the hierarchical structure of the criterion models. Similarly, as noted in the Iowa microcystin case study, other approaches for grouping data beyond using state boundaries are possible using the same hierarchical model structure.

Category 6.9 Supporting State Criteria: Applicability – Unsampled Lake Types

Commenters: Louisiana Department of Environmental Quality; National Wildlife Federation; Alliance for the Great Lakes; Florida Department of Environmental Protection; and Counsel, Mississippi River Collaborative et al.

Comment synopsis: Commenters asked for additional guidance for the applicability of the criterion models to coastal lakes, especially with regard to the salinity gradient. Commenters also requested clarification regarding the applicability of the criterion models to the Great Lakes, and the effect of the national criterion models to ongoing efforts to improve water quality in the Great Lakes. Commenters also requested clarification for the reasons for the wide range of Chl *a* criteria for different lake characteristics specified in each criterion model.

Response: The criterion models were developed using NLA data. Relationships estimated in the national criterion models may be informative when interpreting data collected from systems other than represented in the NLA data, and further evaluation of the applicability of these models is warranted. Some models estimate relationships between measurements at the level of individual samples (e.g., TN-TP-Chl *a* and Chl *a*-microcystin models), and therefore, the relationships estimated in these models are more likely to be applicable in lakes that were not included in the sampled population.

As stated in the criterion document, these criterion recommendations do not affect existing water quality standards and TMDLs that have been approved by EPA.

The wide range of Chl *a* criteria associated with different types of lakes is a direct consequence of the differences in how lakes manifest the effects of increases phytoplankton biomass. By comparing candidate Chl *a* criteria for different designated uses, states and authorized tribes can ensure that criterion values protect the most sensitive use.

Category 6.10 Supporting State Criteria: Constraints (Data)

Commenters: Tip of the Mitt Watershed Council, Wyoming Department of Environmental Quality, Virginia Department of Environmental Quality, Arkansas Department of Energy and Environment, Federal Water Quality Coalition, American Fisheries Society et al., and Wisconsin Department of Natural Resources.

Comment synopsis: Commenters noted that the lack of local monitoring data for certain parameter (e.g., DOC, microcystin, zooplankton biomass) impeded their ability to evaluate the national criterion models.

Response: Data for variables that characterize a lake when running the criterion models (e.g., DOC) are necessary, and when these data are not available for particular lakes, states and

authorized tribes may opt to use other data to characterize the range of possible values for their waters. For example, DOC data are available from the NLA, and examining DOC data in a particular state or ecoregion can provide insights into the range of possible concentrations. Values from this range can then be used to inform criteria derivation. EPA notes, however, that the importance of DOC in determining lake responses to nutrients may provide impetus to incorporate these measurements in routine monitoring.

Response variable data (e.g., microcystin and zooplankton) also may not be collected during routine monitoring. For these data, EPA notes that the national criterion models provide applicable nutrient criteria without the need for refinement with state monitoring data. Hence, state monitoring data for these variables is not necessary when using the national models to derive criteria. However, EPA also observes that data for these parameters provide direct insights into the condition of lakes within a state—insights that can guide efforts to manage these waters.

Category 6.11 Supporting State Criteria: Alternative Methods

Commenters: Virginia Department of Environmental Quality; Riverkeeper, Inc.; City of Springfield, Missouri; Metropolitan St. Louis Sewer District; and Federal Water Quality Coalition.

Comment synopsis: Commenters asked that the criterion document emphasize that other methods can be used to derive criteria. Commenters also asked for examples of other criterion derivation methods and clarification of what would constitute a scientifically defensible method.

Response: As already stated in the draft and in the final criterion document, the recommended criterion models do not preclude the use of other scientifically defensible methods for deriving numeric nutrient criteria. These methods, including mechanistic models and reference condition approaches, are described in detail in previously published EPA guidance (<https://www.epa.gov/nutrient-policy-data/criteria-development-guidance>). The broad coverage of this topic in existing EPA guidance can also provide insights into what constitutes scientifically defensible approaches for numeric nutrient criteria derivation.

Category 6.12 Supporting State Criteria: Derivation Efforts – Sampling Designs

Commenters: Wyoming Department of Environmental Quality and Iowa Department of Natural Resources.

Comment synopsis: Some commenters expressed concerns that the national models relied heavily on modeled relationships and assumptions from NLA, rather than on datasets collected under a nutrient criteria development study, and that this introduced additional uncertainty.

Further comments requested guidance on the minimum sample sizes, where samples should be taken, and how data should be summarized (e.g., mean, median) before entering it into the model.

Response: The recommended criteria do not preclude states and authorized tribes from pursuing their own independent scientifically defensible nutrient criteria development studies to develop specific numeric lake criteria. The model construct provides robust estimates for any single lake (i.e., reasonable certainty) and allows users to adjust the credible interval to account for their risk tolerance given the model uncertainty. Moreover, the option exists for states and authorized tribes to update model coefficients with state-specific data to further address concerns over uncertainty related to regionally sufficient relevant data.

In regards to the appropriateness of data to use, for any state effort to collect data to update the models, data gathered consistent with the underlying NLA sample design is recommended. These methods can be found at <https://www.epa.gov/national-aquatic-resource-surveys>. Very few model inputs require summarization of data. The obvious one is DOC; growing season geometric mean concentrations, consistent with the description in the technical document, is recommended. Most all other inputs are discrete values not requiring summarization (e.g., latitude, longitude, elevation, critical temperature, depth below the thermocline, refugium depth, dissolved oxygen threshold, target microcystin concentration, allowable exceedance frequency, maximum lake depth, or ecoregion).

Category 6.13 Supporting State Criteria: Combined Criteria

Commenters: Virginia Department of Environmental Quality, Virginia Association of Municipal Wastewater Agencies, Missouri Department of Natural Resources, Federal Water Quality Coalition, National Association of Clean Water Agencies, Missouri Public Utility Alliance, and California Water Boards.

Comment synopsis: EPA received a few comments requesting that states have the flexibility to derive and adopt only Chl *a* criteria using the models rather than numeric criteria for nitrogen and phosphorus, citing EPA's [Guiding Principles on an Optional Approach for Developing and Implementing a Numeric Nutrient Criterion that Integrates Causal and Response Parameters](#) published in 2013 (hereafter *Guiding Principles*). Specifically, the comments suggested that the adoption of model-based numeric criteria for nitrogen and phosphorus as independently-applicable water quality criteria should be dependent on the response variables in the models (e.g., dissolved oxygen, microcystin) or be adopted only when a broader watershed context of nutrient pollution management indicates that independently applicable water quality criteria are warranted. Along this line, other comments elaborated further, stating that model-based numeric criteria for nitrogen and phosphorus should be considered within a weight-of-evidence context that relies upon symptom-specific measures of nutrient pollution (e.g., dissolved

oxygen depletion, poor water clarity, measures of fish health) to confirm nutrient pollution as the cause (i.e., combined criteria). Some comments raised concerns that adopting model-based numeric criteria for nitrogen and phosphorus could constrain implementation due to the high cost and the risk of false positive impairment decisions in water quality assessments.

Response: EPA recognizes that uncertainty associated with models used to develop numeric water quality criteria for nutrient pollutants—nitrogen and phosphorus—presents a challenge for states and authorized tribes in deriving and implementing an independently applicable water quality criterion. To aid state decision making, EPA published information in 2013 (i.e., *Guiding Principles*) that described the analytical context, environmental conditions, and regulatory requirements under which states might structure numeric nutrient criteria for the causative pollutants (nitrogen and phosphorus) to be applied dependently, activating as determinant for CWA assessments only when the response variables are exceeded.

States with robust monitoring programs may use their discretion in deciding whether the analytical context and environmental conditions outlined in EPA's 2013 *Guiding Principles* apply to its use of EPA's recommendations, and whether their resulting combined numeric nutrient criteria comply with the regulations described therein. Should states choose to derive combined numeric nutrient criteria using EPA's recommendations, EPA encourages them to document their rationale for dependently applicable numeric nutrient criteria consistent with the implementing regulations described in its 2013 *Guiding Principles*.

In light of its *Guiding Principles* for combined criteria, EPA recommends that states and authorized tribes adopt independently applicable numeric nutrient criteria derived from its recommendations for lakes and reservoirs to ensure protection of their designated uses. First, EPA has identified environmental effects (i.e., assessment endpoints) that are sensitive indicators of water quality degradation and threats to designated uses. Reliance upon those effects, in the first instance as dictated under combined criteria, may undermine a state's intentions to manage its water quality in a preventative manner. Second, EPA's models illustrate, quantitatively and visually, predictable effects as both nitrogen and phosphorus concentrations increase. Although model uncertainty exists, such uncertainty can be accounted for should a state choose to derive independently applicable numeric nutrient criteria, obviating the reliance upon a combined numeric nutrient criteria. Finally, EPA published its *Guiding Principles* after consulting with states and in response to their technical challenges of deriving numeric nutrient criteria in rivers and streams, which when modeled using stressor-response approaches, sometimes exhibit a high degree of uncertainty in selecting a precise nutrient stressor threshold. Fashioning lake and reservoir numeric nutrient criteria, derived using EPA's criteria recommendations, as combined criteria, may be unnecessary because these criteria recommendations are based on strong stressor-response relationships. Additionally, EPA also cautions states and authorized tribes from using a combined criterion approach for lakes and reservoirs because excess nutrient loading that is allowed to increase to the point

where one observes oxygen-depleted conditions or the presence of microcystin in harmful amounts could lead to long-lasting changes in lake water quality that may be difficult to reverse (see Duarte et al. 2007; Scheffer 2004).

Category 6.14 Supporting State Criteria: Existing Criteria

Commenters: Arizona Department of Environmental Quality, Wyoming Department of Environmental Quality, Virginia Department of Environmental Quality, Massachusetts Department of Environmental Protection, South Carolina Water Quality Association, Association of Missouri Cleanwater Agencies, and Minnesota Environmental Science and Economic Review Board.

Comment synopsis: EPA received several comments requesting that states should have discretion in developing numeric nutrient criteria using EPA's recommendations in light of a state's existing nutrient criteria. Other comments urged that EPA's CWA Section 304(a) recommendations should not supplant a state's preferred technical approaches for developing numeric nutrient criteria for lakes and reservoirs, nor should they be used as justification by EPA to criticize technical approaches states used in the past to adopt nutrient criteria that EPA approved. In those cases, commenters suggested that states should be permitted to retain prior state-adopted and EPA-approved nutrient criteria. Finally, some commenters urged EPA not to use its recommendations to force realignment of a state's regulatory priorities in developing nutrient criteria, nor should they be used to alter a state's broader nutrient pollution management strategy.

Response: EPA's criteria recommendations, which are issued under CWA Section 304(a), do not conflict with past scientifically defensible approaches that states and authorized tribes utilized to develop numeric nutrient criteria for lakes and reservoirs, current approaches states and authorized tribes are exploring, or prospective approaches states and authorized tribes will consider as a means to develop numeric nutrient criteria for lakes and reservoirs. Furthermore, the implementing regulations (40 C.F.R. 131.11(b)) allow states and authorized tribes to exercise their discretion in following EPA's recommendations or to choose other scientifically defensible methods when developing water quality criteria. Pursuant to Section 304(a), EPA's recommendations are non-binding and therefore cannot supplant a state's existing numeric nutrient criteria for lakes and reservoirs. Although states must comply with 40 C.F.R. 131.20(a) with respect to EPA's recommendations, the recommendations themselves do not compel states to alter their regulatory priorities or broader nutrient pollution management strategy. In fact, EPA considers its recommendations as a complement to a state's existing water quality standards priorities—enhancing and accelerating regulatory activities that may already exist.

Category 7 Editorial Comments

Category 7.1 Editorial Comments: General

Commenters: Wasatch Front Water Quality Council, Louisiana Department of Environmental Quality, Wyoming Department of Environmental Quality, Montana Department of Environmental Quality, National Wildlife Federation, Iowa League of Cities, Wisconsin Department of Natural Resources, Nevada Division of Environmental Protection, and Maryland Department of the Environment.

Comment synopsis: Several commenters identified typographical errors and made editorial suggestions to clarify the document.

Response: EPA thanks the commenters for identifying errors in the document. EPA reviewed the editorial suggestions and edited the criterion document as appropriate.