VIRGINIA WATER RESOURCES RESEARCH CENTER

Water Quality Assessment Procedures for Virginia: Dissolved Oxygen Assessment of Lakes and Reservoirs

2020 Report of the Academic Advisory Committee for Virginia Department of Environmental Quality

SPECIAL REPORT

VT
VIRGINIA TECH.

VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY BLACKSBURG, VIRGINIA

WATER QUALITY ASSESSMENT PROCEDURES FOR VIRGINIA: DISSOLVED OXYGEN ASSESSMENT OF LAKES AND RESERVOIRS

2020 Report of the Academic Advisory Committee for Virginia Department of Environmental Quality

Written by:

Cayelan C. Carey, Ph.D.
Associate Professor of Freshwater Ecosystem Ecology
Department of Biological Sciences, Virginia Tech

and

Whitney M. Woelmer
Ph.D. Student
Department of Biological Sciences, Virginia Tech

Edited by:
Jane L. Walker
Virginia Water Resources Research Center

Members of the 2020 Academic Advisory Committee to the Virginia Department of Environmental Quality

Stephen H. Schoenholtz, Chair, Virginia Water Resources Research Center / Department of Forest Resources and Environmental Conservation, Virginia Tech

E. Fred Benfield, Department of Biological Sciences, Virginia Tech (retired)

Paul Bukaveckas, Rice Rivers Center / Center for Environmental Studies, Virginia Commonwealth University (VCU)

Cayelan Carey, Department of Biological Sciences, Virginia Tech

Gregory C. Garman, Rice Rivers Center, VCU

Carl Hershner, Department of Biological Sciences / Center for Coastal Resources Management, Virginia Institute of Marine Science, College of William and Mary (retired)

Erin R. Hotchkiss, Department of Biological Sciences, Virginia Tech

Wu-Seng Lung, Department of Civil and Environmental Engineering, University of Virginia

Kevin J. McGuire, Virginia Water Resources Research Center / Department of Forest Resources and Environmental Conservation, Virginia Tech

Daniel McLaughlin, Virginia Water Resources Research Center / Department of Forest Resources and Environmental Conservation, Virginia Tech

Durelle T. Scott IV, Department of Biological Systems Engineering, Virginia Tech

Leonard A. Shabman, Resources for the Future

Eric P. Smith, Department of Statistics, Virginia Tech

Leonard A. Smock, Department of Biology, VCU (retired)

Meredith Steele, School of Plant and Environmental Sciences, Virginia Tech

Kurt Stephenson, Department of Agricultural and Applied Economics, Virginia Tech

Jane L. Walker, Virginia Water Resources Research Center, Virginia Tech

Kang Xia, School of Plant and Environmental Sciences, Virginia Tech

Carl E. Zipper, School of Plant and Environmental Sciences, Virginia Tech (retired)

Water Quality Assessment Procedures for Virginia: Dissolved Oxygen Assessment of Lakes and Reservoirs

Table of Contents

Executive Summary

The Virginia Department of Environmental Quality (DEQ) requested assistance from the Academic Advisory Committee (AAC) in fiscal year 2020 regarding assessment of lakes and reservoirs for dissolved oxygen (DO). This report is a response to the DEQ's paper, "Request for a Scientific Peer Review of DEQ's Assessment Methodology for Lake/Reservoir Dissolved Oxygen" (Appendix A). In response to five specific questions posed by the DEQ, we present the following summarized answers:

(1) Is DEQ's current procedure for DO assessments in lakes/reservoirs sound and scientifically defensible? Does the AAC have any concerns about this procedure besides those already mentioned?

> This report outlines several concerns associated with DEQ's current DO assessment procedure as it relates to being sound and scientifically defensible: (1) it is critical to assess DO in both the epilimnion *and* hypolimnion; (2) spatial monitoring of DO along the reservoir continuum (from upstream to dam) provides valuable information; and (3) temporal averaging should be done on short time scales so as not to mask emerging patterns.

(2) Is the alternative procedure [proposed in DEQ's paper in Appendix A] sound and scientifically defensible? Is it a substantively better approach than the current one? Do its advantages outweigh its disadvantages?

> We have three major concerns with the alternative procedure: (1) it is critical to assess DO in both the epilimnion *and* hypolimnion; (2) spatial DO monitoring and individual assessment of sites along the reservoir continuum (from upstream to dam) provides valuable information, without aggregation throughout the waterbody; and (3) temporal averaging should be done on short time scales so as not to mask emerging patterns occurring within 6-year reporting periods.

(3) Is there another alternative approach that DEQ should consider?

> We propose, in detail, an alternative approach within this report. See pages 13–17 for more information.

(4) Should DEQ restrict DO assessments to vertical profile datasets in lakes/reservoirs or is it appropriate for a surface-only DO dataset to form the basis of an assessment for a lake/reservoir? If the latter, how should surface-only data (or data from shallow profiles) be used in assessments?

> It is critical that a full water column profile be used for DO water quality assessment, so we strongly recommend against using a surface-only DO dataset.

(5) Can the AAC recommend a set of instructions that staff could use 1) to determine whether exceedances at a station are due solely to incomplete fall turnover, 2) to determine to what degree low DO during fall turnover might be exacerbated by pollution, and 3) to downweight fall turnover DO data while still being protective of living resources?

In response to question 5.1, a determination of whether incomplete fall turnover has occurred could be made using a time-series analysis of reservoir and meteorological data, including assessment of the magnitude of thermal stratification, recent air temperature patterns, and differences in DO between the epilimnion and hypolimnion. If water temperature differences between the hypolimnion and epilimnion remain >1°C, it is likely that fall turnover is incomplete or has not yet occurred. This result would necessitate additional sampling events into later autumn to determine if fall turnover was actually incomplete or merely delayed.

In response to question 5.2, low DO conditions during incomplete fall turnover are often related to pollution from nutrients (nitrogen and/or phosphorus) and/or salt. Furthermore, reservoirs often receive a substantial amount of runoff because of their high ratio of watershed area to reservoir surface area, which would strengthen thermal stratification and subsequent hypolimnetic DO depletion. Given that historical nutrient and salt pollution can potentially have long-term (albeit indirect) effects on current DO conditions, a detailed time series of the waterbody's nutrient and salt pollution record is needed to disentangle the relative effects of current pollution on a waterbody.

In response to question 5.3, DO concentrations are important to aquatic life throughout the year, so we do not recommend down-weighting low DO concentrations in the autumn months. Low DO conditions can promote release of nutrients and metals from the sediments into the water column, further exacerbating poor water quality. These conditions should be monitored and reported. Thus, it is critical that DO availability throughout the year be included in lake and reservoir water quality assessments.

Introduction

Background

      The DEQ is responsible for conducting water quality assessments of lakes and reservoirs in Virginia. Through this process, collected data are compared to water quality standards (WQS) to determine if each monitored waterbody supports its designed uses (*e.g.*, swimmable, fishable) or is impaired. The DEQ also develops monitoring and assessment guidance that describes procedures and methods to be used. Updates to the guidance manual are made every two years.

      For fiscal year 2020, the DEQ requested assistance from the AAC concerning the assessment of DO in the Commonwealth's lakes and reservoirs. The DEQ provided a paper entitled "Request for a Scientific Peer Review of DEQ's Assessment Methodology for Lake/Reservoir Dissolved Oxygen" (Appendix A). The paper makes three requests of the AAC:

    (1) review the agency's current process for assessing DO in lakes/reservoirs;
    (2) provide a technical review of an alternative procedure developed by DEQ for assessing DO in lakes/reservoirs that reflects staff recommendations combined with suggestions submitted by a commenter regarding pH criteria assessment;
    (3) advise DEQ on how to handle the assessment of low DO conditions caused by incomplete fall turnover.

The paper concludes with a series of questions posed by DEQ, which are addressed in their entirety in this report.

The importance of dissolved oxygen as a metric of water quality

      Dissolved oxygen concentrations are widely considered the single most important metric of lake and reservoir water quality (Hutchinson 1957, Wetzel 2001). Dissolved oxygen availability is a primary indicator of water quality because DO serves as a "gatekeeper" that controls the rates of many ecological and biogeochemical processes in freshwater ecosystems (Carey *et al.* 2018). For example, low DO concentrations in the bottom waters of lakes and reservoirs promote the release of reduced nutrients and metals from the sediments into the water column, stimulating phytoplankton blooms and degrading water quality overall (Cooke *et al.* 2005). Dissolved oxygen availability also governs habitat quality for freshwater organisms (*e.g.*, fish; Dodds 2002, Kramer 1987). Consequently, monitoring DO concentrations in waterbodies over time is of paramount importance to state and federal agencies and drinking water managers.

      Dissolved oxygen availability may be quantified in two ways: as a mass concentration (generally in milligrams per liter or mg/L) or as percent saturation (%). Percent saturation, which typically ranges from 0% (no oxygen present) to 100% (in equilibrium with atmospheric oxygen concentrations), controls for the differential

solubility of DO at different water temperatures (Figure 1; Wetzel and Likens 1991). Oxygen saturation decreases non-linearly with increasing water temperature; consequently, the concentration of DO associated with 100% saturation at 4°C is 13.09 mg/L, whereas the concentration of DO associated with 100% saturation at 25°C is 8.24 mg/L (Figure 1; Wetzel and Likens 1991). Because of the temperature sensitivity of DO solubility in water, it is important to quantify DO as both a mass concentration and percent saturation when comparing DO conditions across waterbodies (Wetzel 2001).
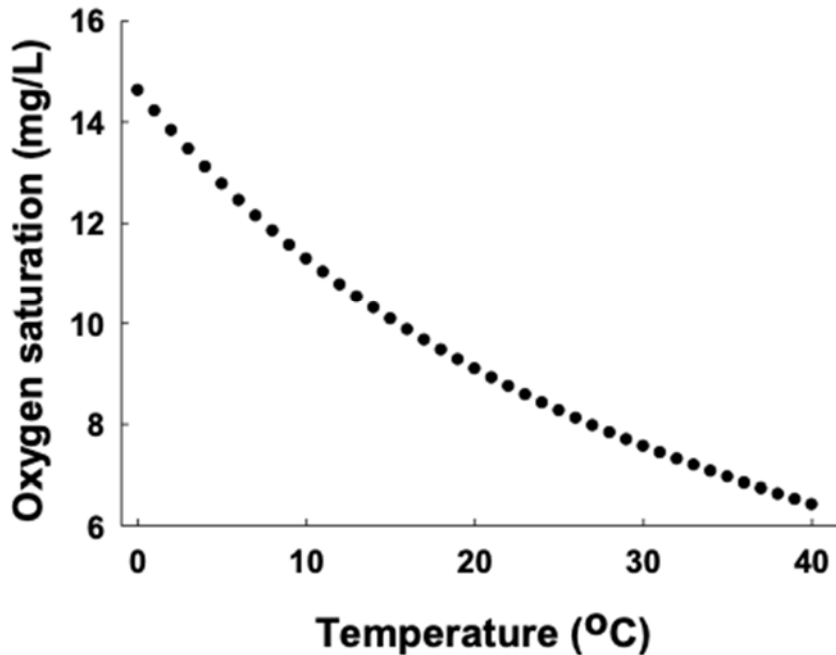
Figure 1. Oxygen saturation decreases with increasing water temperatures. Modified from Wetzel and Likens (1991). The dotted line shows 100% saturation for each water temperature from 0 to 40 degrees Celsius.

Despite the need for temperature correction of DO concentrations, water quality monitoring agencies have historically used mass concentrations in mg/L as their metric of DO availability. The preference for criteria based on mass concentrations may have been related to the lack of reliable methods available in the 1970s and 1980s for measuring percent saturation, in comparison to the well-established method of using Winkler titrations to calculate the concentration of DO (Winkler 1888; see Carpenter [1965] for more information).

Water quality thresholds are still set using mass concentration units (mg/L), even though common handheld DO sensors now provide DO availability in both mass concentration and percent saturation. For example, the U.S. Environmental Protection Agency's (EPA's) National Aquatic Resource Survey (U.S. EPA 2020) considers DO concentrations <3 mg/L to be "of concern" and waters with levels <1 mg/L to be "usually

devoid of life." The EPA's 2012 National Lakes Assessment (U.S. EPA 2017) classifies DO into three classes: "least disturbed" (≥5 mg/L), "moderately disturbed" (3-5 mg/L), and "most disturbed" (≤3 mg/L).

Dissolved oxygen in a stratified waterbody

The concentration of DO in a thermally-stratified lake or reservoir is governed by different factors in the epilimnion (surface layer) than in the hypolimnion (bottom layer; Figure 2; Wetzel 2001). In the epilimnion, oxygen concentrations are primarily governed by: 1) diffusion of oxygen across the air-water interface or entering from inflows; 2) production of oxygen in well-lit waters through photosynthesis by phytoplankton; and 3) consumption of oxygen by respiration of microbes and other organisms (Wetzel 2001). If a waterbody is characterized by high phytoplankton concentrations, there will be high DO concentrations during the day (when photosynthesis rates are at their highest) and low DO at night (when no photosynthesis can occur in the dark, yet respiration is still occurring). During extremely large phytoplankton blooms, high rates of photosynthesis can occasionally result in supersaturation (>100% DO saturation) during the day, until the excess DO in the water diffuses into the atmosphere at night (O'Boyle *et al.* 2013, Welch 1969). Waterbodies with high organic matter concentrations in their surface water will likely exhibit lower DO availability (undersaturation, or <100% saturation) because of greater microbial respiration rates breaking down the organic matter (Jewell 1971, Pereira *et al.* 1994). Oxygen concentrations in surface inflows entering a lake or reservoir could either result in an increase or decrease of oxygen, depending on the DO concentration in the stream inflow in comparison to the lake or reservoir DO concentration. In waterbodies with good water quality (as indicated by the absence of phytoplankton blooms), epilimnetic DO availability should usually be at or near 100% saturation because of oxygen diffusion from the atmosphere into the mixed surface layer (Wetzel 2001).

In contrast, DO availability in the hypolimnion is usually lower than in the epilimnion because of the absence of photosynthesis in the dark bottom waters of lakes and reservoirs (Wetzel 2001). In the hypolimnion, in lieu of groundwater inputs, DO concentrations are primarily governed by: 1) oxygen consumption by respiration of organisms (predominantly microbes) consuming organic matter; and 2) oxygen consumption by reduced chemical compounds (*e.g.*, methane, CH_4; hydrogen sulfide, H_2S; ammonium, NH_4^+; and ferrous iron, Fe^{2+}; Figure 2). The oxygen consumption rates by microbes and chemical compounds are partitioned into biological oxygen demand (BOD) and chemical oxygen demand (COD), respectively (Wang 1980, Wetzel 2001). Lakes or reservoirs with high concentrations of organic matter and reduced chemical compounds in their sediments — *i.e.*, indicators of poor water quality — will generally exhibit hypolimnetic DO availability less than 50% saturation and occasionally hypoxia (low oxygen) or even anoxia (no oxygen). We note that it is common for low-nutrient

(oligotrophic) lakes and reservoirs to exhibit some hypolimnetic oxygen depletion near the sediments during the late summer; however, it is unlikely that this depletion would reach hypoxic or anoxic levels or extend throughout the entire hypolimnetic layer (following Wetzel 2001).
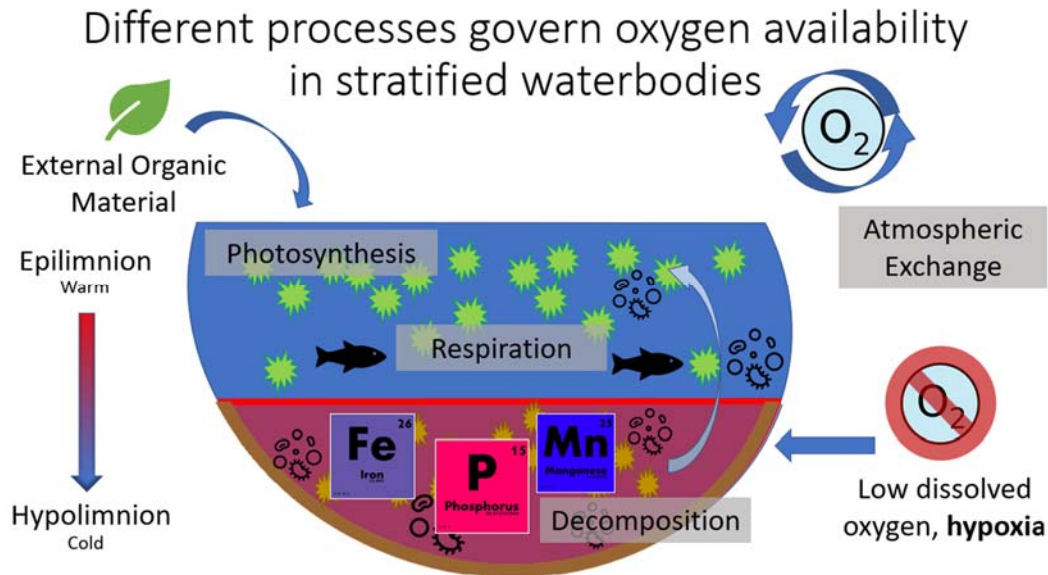
Figure 2. Dissolved oxygen in the epilimnion (surface layer) is governed by different factors than in the hypolimnion (bottom layer) of a thermally-stratified lake or reservoir. In the epilimnion, DO concentrations are primarily governed by: 1) atmospheric exchange and inflows; 2) production of oxygen through photosynthesis; and 3) consumption of oxygen by respiration. In the hypolimnion, DO concentrations are primarily governed by: 1) consumption by respiration during the process of decomposition; and 2) oxygen consumption by reduced chemical compounds.

It is important that a full DO depth profile (*i.e.*, DO at multiple depths from the water's surface to the sediments) is measured when monitoring lakes and reservoirs because epilimnetic and hypolimnetic DO availability reveals different aspects of water quality. In general, we consider DO availability at or near 100% saturation to be indicative of "good" water quality because that water has not been influenced by biological or chemical processes resulting in either supersaturation (>100%) or undersaturation (<100%). In the epilimnion, as described above, supersaturation during daytime would be indicative of phytoplankton blooms, which represent poor water quality because of excess nutrient (nitrogen and/or phosphorus) concentrations that stimulate phytoplankton growth (Schindler 1977, Schindler *et al.* 2008, Smith 1982). In the epilimnion, DO undersaturation would be indicative of high organic matter, which could either occur because of natural sources (*e.g.*, leaching of organic matter

compounds into a waterbody) or anthropogenic sources (*e.g.*, sewage). Anthropogenic inputs into lakes and reservoirs resulting in undersaturated epilimnetic DO concentrations are indicative of poor water quality, which usually occurs at <3 mg/L or <50% saturation and is a common threshold below which many aquatic organisms cannot survive (Davis 1975, Franklin 2014).

Dissolved oxygen availability in lake and reservoir hypolimnia is a key metric of water quality: highly undersaturated or supersaturated DO availability both represent poor water quality status (Hynes 1960, U.S. EPA 2017). Because of the lack of photosynthesis in dark bottom waters, supersaturation resulting from phytoplankton blooms generally cannot occur except in rare cases. During thermally-stratified conditions, diffusion of DO from the epilimnion into the hypolimnion is limited (Wetzel 2001). In waterbodies with poor water quality, DO concentrations will decrease in the hypolimnion soon after the onset of thermal stratification because of high BOD and COD. Hypolimnetic undersaturation is very common in many waterbodies across the U.S., occasionally reaching hypoxia and even anoxia (Dodds *et al.* 2009, Stoddard *et al.* 2016). Hypolimnetic undersaturation is an important indicator of suboptimal water quality conditions because it promotes the release of reduced nutrients and metals from the sediments into the water column.

Global change effects on dissolved oxygen and mixing

As a result of changes in climate and land use, hypolimnetic DO availability is decreasing in lakes and reservoirs globally, posing substantial threats to water quality (Jenny *et al.* 2016). A recent survey of 365 waterbodies indicates that the dominant driver of increased hypolimnetic anoxia (defined as bottom-water DO concentrations <2 mg/L) is increased levels of runoff high in nutrients (nitrogen and/or phosphorus) entering lakes and reservoirs (Jenny *et al.* 2016). Runoff high in nutrients promotes phytoplankton blooms that initially cause DO supersaturation in epilimnetic waters. After the phytoplankton die and sink to the sediments, the increase in organic matter will stimulate hypolimnetic microbial respiration, thereby decreasing hypolimnetic DO concentrations (Müller *et al.* 2012). These low DO conditions will persist throughout the summer stratified period until fall turnover, at which time thermal stratification ceases and the epilimnion and hypolimnion mix (Wetzel 2001).

Fall turnover is caused by decreasing air temperatures in the autumn months, which cool the surface waters until the temperature differential between the epilimnion and hypolimnion becomes zero. Once the temperature and density gradient between the two layers disappears, the epilimnion and hypolimnion are homogenized, and the entire water column will mix (Nürnberg 1984, Wetzel 2001).

Low DO conditions in the hypolimnion may continue throughout the winter months if "incomplete" fall turnover occurs (Effler *et al.* 1986, Nürnberg 1995, Vachon *et al.* 2019). Incomplete fall turnover can occur when air temperatures in the fall do not

decrease sufficiently, so a water temperature and density gradient between the epilimnion and hypolimnion remains. Incomplete fall turnover can also occur in waterbodies that have received an influx of salts and/or other dissolved compounds into the hypolimnion, which intensifies the density differences between the two layers, thereby inhibiting mixing of the two layers (Bubeck and Burton 1989, Judd *et al*. 2005). Because of winter salting and other pollution resulting in greater salt concentrations in many waterbodies (Dugan *et al*. 2017), the likelihood of incomplete fall turnover may increase for some waterbodies.

Incomplete fall turnover has negative effects on water quality because the prolonged thermally-stratified conditions will promote the likelihood of low DO conditions at the sediments. Without the opportunity for mixing with the epilimnion and subsequent infusion of DO, hypolimnetic DO concentrations will continue to decrease until anoxic conditions are reached. Little is known about the prevalence of incomplete fall turnover within Virginia, but it is expected that decreased mixing will likely increase with future climate warming (Gerten and Adrian 2002).

Human-made reservoirs may be particularly susceptible to water quality degradation

Land use and climate change effects on water quality may be exacerbated in human-made reservoirs in comparison to glacially-formed north temperate lakes (Hayes *et al*. 2017). This degradation is likely due to at least three reasons. First, because many human-made reservoirs were built by damming lotic waterbodies and flooding adjacent terrestrial areas, they have large quantities of organic matter, metals, and nutrients in their sediments that are trapped by dams (Kennedy *et al*. 1985, Thornton 1990). Second, reservoirs have significantly higher ratios of watershed area to surface area than naturally-formed lakes (Doubek and Carey 2017), resulting in greater nutrient, metal, and organic matter loads from the landscape. Together, the accumulation of high organic matter concentrations and chemically-reduced compounds will promote high biological and chemical oxygen demand. Third, many reservoirs are located at lower latitudes than glacially-formed temperate lakes, resulting in warmer fall conditions and potentially a higher likelihood of incomplete fall turnover.

Reservoirs are historically under-studied in limnology relative to glacially-formed lakes (Doubek and Carey 2017, Hayes *et al*. 2017) so their sensitivity to global change remains unknown. However, the critical ecosystem services for society that reservoirs provide — *e.g.*, water for drinking, fisheries, irrigation, industry, and recreation (Doubek *et al*. 2019) — necessitate careful monitoring of their water quality, especially of their DO concentrations.

DEQ Questions and Responses

The DEQ requested that the AAC review its current and alternative DO assessment protocols for lakes and reservoirs in Virginia and provide advice concerning assessments with low DO conditions caused by incomplete fall turnover. To assist with this effort, the DEQ provided summaries of these protocols in a document (Appendix A), which we reviewed. The DEQ requested that the review consider the distinguishing marks of strong and weak assessment methodology (Table 1) and address specific questions.

Table 1. Distinguishing marks of strong and weak assessment methodology.

	Strong assessment methodology		Weak assessment methodology
1	Is written clearly and thoroughly enough to be easily understood by the informed public and staff.	1	Is poorly written and leaves out key details; assessment results cannot be replicated just by reading the document. Example: The methodology document says that "samples will be assessed against the criterion" but does not specify the decision rules used to determine WQS non-compliance.
2	Follows the letter of the WQS.	2	Takes considerable creative license with the WQS. Example: The WQS stipulates that a criterion cannot be exceeded more than 10% of the time, but assessment staff are instructed to evaluate the 90th percentile of a dataset against the criterion.
3	Is consistent with how the WQS were developed.	3	Is not consistent with how the WQS were developed. Example: A criterion was developed from monthly arithmetic means of field samples, but staff are instructed to compare annual geometric means of field samples to the criterion.
4	Does not require staff to "use best professional judgment" for key decision points.	4	Is so subjective that assessment decisions differ among staff. Example: Staff are given the discretion to screen out samples taken during "extreme weather events", with no instructions provided on how staff should make this determination.
5	Encourages staff to use "all available and appropriate data."	5	Uses a very narrow definition of "appropriate data", one that excludes potentially useful datasets from the assessment. Example: Staff are instructed to use only agency data, when sister agencies and reputable citizen groups collect similar datasets.
6	Minimizes type II errors as much as possible given available monitoring resources.	6	Allows staff to base assessment on very small and potentially outdated datasets with no safe guards against type II errors. Example: Staff are allowed to decide a water body is meeting a criterion based on a single grab sample.

WQS = water quality standard

(1) Is DEQ's current procedure for DO assessments in lakes/reservoirs sound and scientifically defensible? Does the AAC have any concerns about this procedure besides those already mentioned?

According to the document provided by DEQ, the current procedure is to "pool data by taking all measurements taken in the epilimnion (for Section 187 lakes/reservoirs) or in the water column (for non-Section 187 lakes or when Section 187 lakes/reservoirs are not stratified) over the six-year assessment period.... Vertical temperature profiles enable assessors to determine the bottommost depth of the epilimnion at a station during a monitoring event. Epilimnetic samples are then compared to the minimum DO criterion (4.0 mg/l for most lakes/reservoirs). If more than 10% of these samples are below the criterion, then the lake/reservoir assessment unit will be deemed impaired for DO" (see Appendix A, beginning on page A-5, for more information on the current procedure).

We have several concerns associated with DEQ's current DO assessment procedure (2018) and do not think that it is sound or scientifically defensible for four primary reasons:

First, we recommend taking DO measurements throughout the entire depth profile from the surface to the sediments (thereby capturing both the epilimnion and hypolimnion) at each monitoring site, regardless of thermal stratification conditions. The measurement of DO in the epilimnion is needed to meet the WQS, however, the DO availability in the *hypolimnia* of lakes and reservoirs is a key metric of water quality that also needs to be quantified to understand whole-ecosystem functioning (Gerling *et al*. 2016, Hynes 1960, U.S. EPA 2017, Wetzel 2001). In particular, low DO in the hypolimnion degrades fish and macroinvertebrate habitat and promotes nutrient and metal release from the sediments, which can have subsequent negative effects on water quality long-term (Gerling *et al.* 2016). Thus, a full water column DO depth profile is needed to evaluate water quality.

Second, the current DEQ protocol specifies pooling of data, which means that exceedance rates are calculated from all the epilimnion observations recorded throughout a reservoir, from its inflow tributaries to a dam. We recommend pooling data only within a monitoring station on a given sampling day. It is important to note that water quality varies across the reservoir continuum from the upstream riverine zone, where inflowing tributaries enter the reservoir, to the downstream lacustrine zone at the dam (Thornton 1990). Upstream riverine sites that have been impacted from nutrient and pollutant loading in the watershed sometimes have poorer overall water quality than downstream reservoir sites, where nutrients and organic matter may have already settled out on the sediments (Breitburg *et al.* 2003, Stringfellow *et al.* 2009). Conversely, it may be possible to see decreases in DO when moving from upstream to downstream because of high DO in incoming stream water (Borges *et al.* 2008). Our interpretation of the most defensible approach for monitoring would thus be to pool epilimnetic samples

only within a monitoring station, *not* across the entire epilimnion of the reservoir, as is part of the current DO assessment protocol used by DEQ. Averaging across the entire reservoir masks longitudinal variability in water quality and could prevent the identification of impacted tributaries that bring poor water quality inputs into the reservoir. If there is more than one monitoring site within a spatial unit of the lake/reservoir on a given day, aggregating epilimnetic DO concentrations among the sites within the spatial unit would be acceptable. Ideally, the assessment methodology would allow for the identification of changes in water quality within specific areas of the reservoir.

Third, another concern is the current practice of pooling observations with no temporal averaging over a 6-year assessment period. We support the approach of the 2020 assessment guidance manual, which instructs DEQ staff to calculate monthly medians of chlorophyll and total phosphorus concentrations prior to further analyses. To meet federal and state reporting requirements, DEQ may need to continue performing 6-year assessments, but we recommend that they also consider conducting assessments for shorter timeframes. Although a six-year assessment period provides a larger dataset that represents a larger range of different conditions, pooling observations over a six-year time period could mask water quality changes that would be detected on shorter time scales. For example, in a six-year analysis of DO data in an oligotrophic lake in New Hampshire, Richardson *et al.* (2017) detected changes in water quality that were only possible from comparing year-to-year data. Waiting to assess all six years of data would have prevented the detection of water quality degradation occurring on more rapid time scales.

Fourth, the current DO criterion of 4.0 mg/L may be an inaccurate metric of water quality because it does not consider the differential solubility of DO at different water temperatures. As noted earlier, DO concentrations in mg/L are not as sensitive metrics of DO availability as percent DO saturation, which corrects for differential DO solubility at varying water temperatures (Figure 1). We, therefore, recommend that DEQ begin recording percent DO saturation data along with mg/L data. Consequently, 4.0 mg/L represents a range of different percent saturation levels that encompass both acceptable (>50%) *and* poor (<50%) water quality when using 50% saturation as a threshold for water quality (Figure 3). For example, at 5°C, 4.0 mg/L equals 31% DO saturation, whereas at 35°C, 4.0 mg/L equals 58% saturation (Figure 2). The current DEQ monitoring is focused on epilimnetic conditions in lakes and reservoirs during warm-weather months in Virginia (April 1–October 31), when epilimnetic water temperatures typically range from ~15 to >30°C. Under these conditions, we consider the 4.0 mg/L threshold, which exhibits percent DO saturation above and below 50%, to thus be a less-useful threshold than 3.0 mg/L, which consistently exhibits percent saturation levels below 50% (Figure 3). Moreover, 3.0 mg/L is the current DO threshold

used in the EPA 2012 National Lakes Assessment to identify the most disturbed conditions in the epilimnion of lakes and reservoirs (U.S. EPA 2017).

In summary, the DEQ's 4 mg/L criterion is likely more protective of aquatic life, but it risks classifying some "good" quality water as impaired. In our opinion, using a percent saturation threshold of 50% would be preferred for the DO numeric criterion to enable comparison of DO availability across waterbodies with different temperatures. However, we recognize that it may not be feasible for monitoring agencies to implement this criterion, especially given the EPA's precedent of using a mass concentration threshold. Thus, as the second-best option, we recommend changing the DEQ's DO threshold from 4.0 to 3.0 mg/L to be more inclusive of high temperature conditions that would naturally cause low DO concentrations because of lower DO solubility in warmer water (Figure 3).
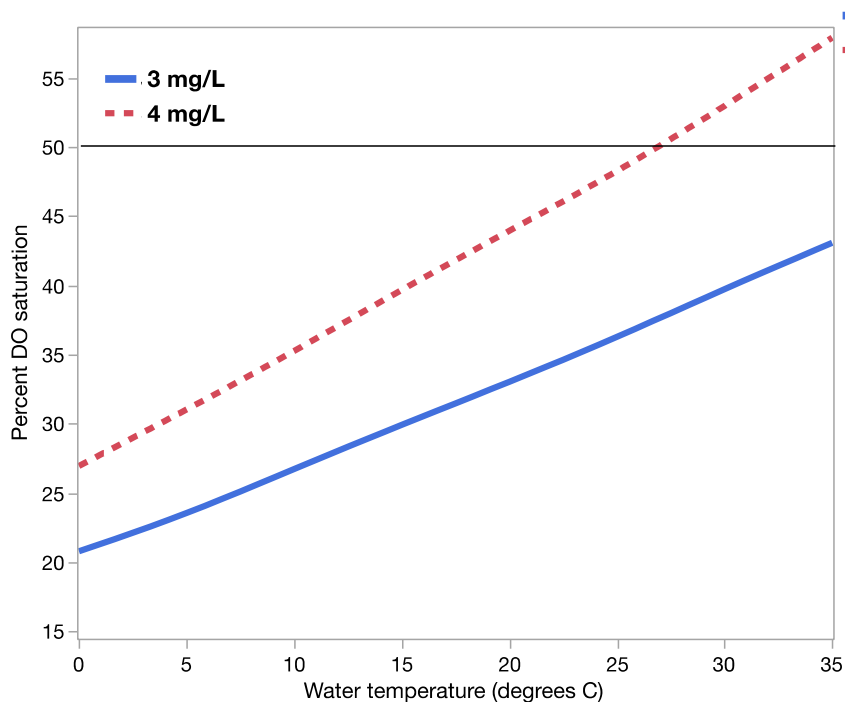
Figure 3. The percent DO saturation corresponding to a given mass concentration (mg/L) varies as a function of temperature. If 50% DO saturation (the black solid line) represented the criterion, the current criterion of 4.0 mg/L could constitute poor water quality (<50% saturation) at cold temperatures or acceptable quality (>50% saturation) at high temperatures. A different value, 3.0 mg/L, would also vary with temperature but would consistently indicate poor water quality, as defined by DO saturation <50%.

(2) Is the alternative procedure sound and scientifically defensible? Is it a substantively better approach than the current one? Do its advantages outweigh its disadvantages?

The DEQ has proposed an alternate procedure for DO monitoring (see Appendix A, pages A-10 to A-12): "One alternative procedure would begin by establishing the location of the thermocline at an individual station (the point where the temperature change between depths is $\geq 1°C$), as is done currently. Epilimnetic or water column DO samples along the vertical profile would then be aggregated via the median. This would represent the average DO at this station, but in a way that does not presume how the samples are distributed or allow outlier samples to have undue influence. The next step would be the aggregation of all 'epilimnon/water column medians' generated at stations visited during the same monitoring run. A median would again be used for the same reasons as above. The end result would be a snapshot of a water body's DO concentration for a particular monitoring run. This snapshot would then be assessed in combination with all the other snapshots taken over the assessment period (see Table 4 [on page A-11] for an illustration of this process). The end result would be an exceedance rate that could be reasonably interpreted as 'Aquatic life had unsuitable habitat X% of the time when the lake/reservoir was monitored.' This exceedance rate would then be evaluated against the 10% rule to determine criteria attainment status for the lake/reservoir assessment unit."

We have several concerns with the alternative procedure provided by DEQ that motivate our proposed new approach, detailed below in the response to question 3. First, we suggest that DEQ monitor both the epilimnion and hypolimnion for DO and record results in both mg/L and percent saturation. Second, as noted previously, spatially aggregating data from all epilimnetic stations precludes the DEQ from detecting site-specific changes in water quality within a lake or reservoir. Third, determining the exceedance rate over a six-year assessment period may mask water quality changes that would be detected on shorter time scales. Given these concerns, we recommend an alternative approach, which we describe in the next section.

(3) Is there another alternative approach that DEQ should consider?

To alleviate the issues noted for both the current assessment approach and the alternative approach presented by DEQ (identified above), we propose a new approach. As noted previously, a percent saturation threshold of 50% would be preferred for the DO numeric criterion to avoid issues related to variable DO solubility at different temperatures. However, we recognize that it may not be feasible for monitoring agencies to implement this change, especially given both the precedent of EPA and DEQ of using a mass concentration threshold. Thus, as the second-best option, we recommend changing the DEQ's DO threshold from 4.0 to 3.0 mg/L. The approach

described below uses a 3.0 mg/L numeric criterion for DO, although it could easily be modified to use a 50% DO saturation threshold.

1. At every current lake/reservoir assessment unit site, measure a full water column temperature and DO depth profile from the subsurface at 0.1 m on every meter interval to the nearest integer meter depth above the sediments (*e.g.*, in 5.2 m deep lake, the water column depth profile would consist of measurements from 0.1 m, 1.0 m, 2.0 m, 3.0 m, 4.0 m, and 5.0 m). Dissolved oxygen should be measured in mg/L and percent saturation.

2. Establish the depth of the thermocline at that site, if it exists, using the operational definition of thermocline (the depth that exhibits the greatest temperature change between depths [at least > 1°C]) to delineate epilimnion and hypolimnion. If there is no temperature change between depths of at least 1°C, then we propose treating the entire water column as one integrated layer for the following steps.

3. For each site individually, determine if the median DO concentration of all of the measurements within the epilimnion on a sampling day is below 3.0 mg/L (see Table 2 as an example). We recommend using this threshold instead of the current DEQ threshold (4.0 mg/L), following the EPA's guidelines for water quality monitoring (U.S. EPA 2017) and because it is generally a more robust metric of water quality (Figure 3).

4. For each site individually, determine if the median DO concentration of all of the measurements within the hypolimnion on a sampling day is below the DO exceedance threshold (see Table 2 as an example). Given the likelihood of DO depletion in the hypolimnion of most reservoirs that is inherent to their construction (see above), we propose using the <3.0 mg/L threshold for hypolimnetic DO exceedances. We recommend this threshold for the reasons described above; it also is the threshold used by the EPA in the 2012 National Lakes Assessment to identify the "most disturbed" waterbodies (U.S. EPA 2017; Figure 3).

5. As noted above, if there is no temperature change between depths of at least 1°C, the entire water column should be treated as one layer. In this case, determine if the median DO concentration of all of the measurements within the water column on a sampling day is below the DO exceedance threshold of 3.0 mg/L.

6. If there is more than one monitoring site within a lake/reservoir assessment spatial unit, aggregate epilimnetic DO concentrations among sites within a spatial unit. Similar to the method described in the alternative procedure by DEQ, the aggregation should be calculated by taking the overall median of the median epilimnetic DO concentrations per site within each lake/reservoir assessment spatial unit for a sampling day.

7. If there is more than one monitoring site within a lake/reservoir assessment spatial unit, aggregate among hypolimnetic DO concentrations among sites within a spatial unit. This aggregation should be calculated by taking the overall median of the median hypolimnetic DO concentrations per site within each lake/reservoir assessment spatial unit for a sampling day.

8. If either the median epilimnion *or* hypolimnion DO concentration of a lake/reservoir assessment spatial unit on a monitoring day is less than the 3.0 mg/L DO threshold, then that spatial unit as a whole is considered to exhibit an exceedance on that day. It would still be considered only one exceedance (not two exceedances) if both layers on the same day had median DO concentrations less than the threshold.

Table 2. A hypothetical temperature and DO depth profile from a monitoring site.*

Depth	Temperature (degrees C)	DO (mg/L)	% DO saturation
0.1	28.0	7.8	100
1.0	27.5	7.7	98
2.0	27.0	7.7	96
3.0	26.5	7.6	95
4.0	18.0	5.8	61
5.0	16.0	5.5	56
6.0	15.0	5.2	52
7.0	14.0	2.3	22
8.0	13.0	1.6	15

*The thermocline is established to be between 3 and 4 m depth, resulting in 4 epilimnetic depths and 5 hypolimnetic depths (delineated by solid black line). The median epilimnetic DO concentration is 7.7 mg/L, which is above the exceedance threshold (<3.0 mg/L). The median hypolimnetic DO concentration is 5.2 mg/L, which is also above the exceedance threshold. As neither the median epilimnetic or hypolimnetic DO concentration is below the exceedance threshold, no exceedances are observed at this site, despite the two lower hypolimnetic depths exhibiting DO concentrations <3.0 mg/L.

9. Calculate the total number of spatial unit-sampling days (*i.e.*, the total number of days in which each spatial unit within a waterbody was sampled) over a maximum two-year (not six-year) rolling-window assessment period. As noted above, pooling data within two years only enables the detection of changes in water quality that are occurring on rapid time scales. For example, if there were three spatial units within one lake that were measured on 7, 4, and 12 sampling days, respectively, over two years, then there would be 23 total spatial unit-sampling days in that two-year assessment period. The exceedance rate is then calculated as the number of times in which any

sampling site exhibited an exceedance over the total number of spatial unit-sampling days (see Figure 4 as an example).

10. A waterbody is considered impaired for DO if the exceedance rate is >10% for the two-year assessment rolling-window period.

11. For calculating descriptive statistics in additional DEQ reports, aggregate data first to calculate month-medians prior to additional analyses *vs.* pooling all data equally within a year.

Spatial Unit A (DEQ)
- 7 sampling days during 2-year period
- 1 day with either an epilimnetic or hypolimnetic DO exceedance

Spatial Unit B (Non-Agency)
- 4 sampling days during 2-year period
- 0 days with either an epilimnetic or hypolimnetic exceedance

Spatial Unit C (DEQ)
- 12 sampling days during 2-year period
- 1 days with either an epilimnetic or hypolimnetic DO exceedance

DO Assessment Summary
7 + 4 + 12 = 23 spatial unit-sampling days
1 + 0 + 1 = 2 days with epilimnetic or hypolimnetic exceedances
Exceedance rate = 2/23 = 8.7%<=Criterion

Figure 4. There are three spatial units within this waterbody, which are measured on 7 days (Unit A), 4 days (Unit B), and 12 days (Unit C), resulting in 23 spatial unit-sampling days total. With 2 exceedances detected, this waterbody is considered not impaired because 2/23 is <10%.

Following the criteria of strong and weak assessment methodology in Table 1, we think this new approach meets most of the categories of a strong methodology: it is written clearly, does not require best judgement, and uses all available data that meet the standards necessary for inclusion (e.g., only complete depth profiles are used, not surface-only measurements). We cannot evaluate the ability of the protocol to minimize type II errors without a long-term comparison of the proposed new protocol to detect water quality exceedances with the previous protocol.

(4) Should DEQ restrict DO assessments to vertical profile datasets in lakes/reservoirs or is it appropriate for a surface-only DO dataset to form the basis of an assessment for a lake/reservoir? If the latter, how should surface-only data (or data from shallow profiles) be used in assessments?

As described previously, it is important that a full water column profile be obtained, so we recommend against using a surface-only DO dataset. Instead, we recommend following the proposed monitoring procedure described in the answer for question 3 (above).

(5) Can the AAC recommend a set of instructions that staff could use 1) to determine whether exceedances at a station are due solely to incomplete fall turnover, 2) to determine to what degree low DO during fall turnover might be exacerbated by pollution, and 3) to downweight fall turnover DO data while still being protective of living resources?

Incomplete fall turnover occurs when surface and bottom waters do not fully mix in the fall. Although most waterbodies in Virginia experience complete fall turnover, there are several factors that may lead to incomplete fall turnover. If surface temperatures remain too warm to homogenize with bottom waters (>1°C difference) or changes in density between the layers from an increase in solutes (*e.g.*, road salt) prevent mixing, it is possible for a waterbody to exhibit incomplete fall turnover. When waterbodies do not completely turn over, oxygen-poor hypolimnetic waters are not able to replenish their oxygen concentrations by mixing with oxygen-rich epilimnetic waters. Incomplete turnover leads to further degradation of hypolimnetic water quality throughout the year (Effler *et al.* 1986, Nürnberg 1995, Vachon *et al.* 2019).

We address the three questions regarding incomplete fall turnover in the following subsections:

1. *Exceedances due solely to incomplete fall turnover.* A determination of whether incomplete fall turnover has occurred could be made using a time series analysis. This approach would assess the magnitude of thermal stratification (*i.e.*,

if a discrete epilimnion and hypolimnion was present, *vs*. isothermal conditions from the surface to the sediments) and how thermal stratification has changed since the previous sampling day within that year at the site. This approach would involve first analyzing patterns of recent air temperature over time since the last sampling event (*e.g.*, have air temperatures decreased such that fall mixing has become possible?) as well as monitoring water temperature and DO depth profiles on multiple sampling days. If water temperature differences between the hypolimnion and epilimnion are still $>1°C$, it is likely that fall turnover is incomplete or has not yet occurred. This result would necessitate additional sampling events into later autumn to determine if fall turnover was incomplete or delayed. It is possible that stratified conditions in the autumn are incorrectly classified as incomplete fall turnover instead of delayed fall turnover if additional sampling does not occur in late autumn and early winter after complete fall turnover has occurred. A waterbody would only technically be classified as having incomplete fall turnover if it continues to exhibit stratified conditions throughout the winter.

It is more difficult to assess if incomplete turnover is occurring when samples are taken only from the epilimnion. Our proposed new sampling protocol should provide an improved assessment of how incomplete fall turnover affects DO concentrations in the water column. For example, with hypolimnetic monitoring, we could observe greater DO concentrations in the hypolimnion during fall turnover (Guarino *et al*. 2005). Ultimately, the new sampling protocol would capture the distribution of DO concentrations between the two water layers, rather than simply showing a decrease in DO in the epilimnion during incomplete turnover (as could occur using the current sampling protocol).

2. *Degree that low DO during fall turnover might be exacerbated by pollution.* Low DO conditions during incomplete fall turnover are often related to pollution from nutrients (nitrogen and/or phosphorus) and/or salt. Given their high ratio of watershed area to reservoir surface area, reservoirs often receive a substantial amount of runoff laden with nutrients and pollution from land use within their watersheds (Doubek and Carey 2017). Furthermore, historical eutrophication of waterbodies was common throughout Virginia (Gerling *et al*. 2016), dating back to the early 20th century when the state was dominated by agricultural land. Because these historical nutrients can remain in the sediments of waterbodies for decades to centuries, hypolimnetic anoxia can promote their release years after the nutrients entered the waterbody (Gerling *et al*. 2016). Thus, hypolimnetic hypoxia caused by nutrient pollution in summer stratified months can be prolonged into the fall when turnover does not fully occur.

Although nutrient pollution promotes low hypolimnetic DO conditions that become prolonged during incomplete fall turnover, salt pollution itself would cause incomplete fall turnover by exacerbating density differences between the epilimnion and hypolimnion. Salinization of freshwaters is an increasingly common problem across the northern hemisphere due to the use of salt in road de-icing (Cañedo-Argüelles *et al.* 2013, Rivett *et al.* 2016). Because increases in solute concentrations alter the density of water, salinization is commonly cited as preventing or lessening mixing in lakes and reservoirs (Bubeck and Burton 1989, Judd 1970, Novotny *et al.* 2008).

In summary, there are several pathways whereby nutrient and salt pollutants can influence DO concentrations. This influence can occur through a recent influx of pollutants (*e.g.*, high-nutrient fertilizer or salt inputs), as well as presence of historical pollutants that can continue to have a negative impact on water quality more than 50 years after introduction to the waterbody (Gerling *et al.* 2016). Given that historical pollution can potentially have long-term effects on current DO conditions, it would be challenging to disentangle the relative effects of current pollution on a waterbody without a detailed time series of the waterbody's pollution record. Analyzing a historical time series to identify changes in nutrient and chloride concentrations in a waterbody exhibiting incomplete turnover would be a useful first step for determining if one form of pollution dominated over the other in driving delayed turnover dynamics.

3. *Downweighting fall turnover DO data and protecting living resources*. Dissolved oxygen concentrations are important to aquatic life throughout the year so we do not recommend down-weighting low DO concentrations in the autumn months, even if they are a result of incomplete fall turnover. These low DO conditions are representative of the conditions that aquatic organisms are experiencing and should therefore be fully included in any analysis of water quality within a waterbody. Furthermore, low DO conditions can promote the release of nutrients and metals from sediments into the water column, further exacerbating poor water quality. Thus, it is critical that DO availability throughout the year be included in lake/reservoir water quality assessments.

References

Borges, P.A.F., Train, S., & Rodriguez, L.C. 2008. Spatial and temporal variation of phytoplankton in two subtropical Brazilian reservoirs. *Hydrobiologia* 607: 63–74. https://doi.org/10.1007/s10750-008-9367-3

Breitburg, D.L., Adamack, A., Rose, K.A., Kolesar, S.E., Decker, M.B., Purcell, J.E., Keister, J.E., & Cowan, J.H. 2003. The pattern and influence of low dissolved oxygen in the Patuxent River, a seasonally hypoxic estuary. *Estuaries* 26: 280–297. https://doi.org/10.1007/BF02695967

Bubeck, R.C., & Burton, R.S. 1989. *Changes in Chloride Concentrations, Mixing Patterns, and Stratification Characteristics of Irondequoit Bay, Monroe County, New York, After Decreased Use of Road-deicing Salts, 1974–1984*. Water-Resources Investigations Report 87-4223. Department of the Interior, U.S. Geological Survey. https://doi.org/10.3133/wri874223

Cañedo-Argüelles, M., Kefford, B.J., Piscart, C., Prat, N., Schäfer, R.B., & Schulz, C.J. 2013. Salinisation of rivers: An urgent ecological issue. *Environmental Pollution* 173: 157–167. https://doi.org/10.1016/j.envpol.2012.10.011

Carey, C.C., Doubek, J.P., McClure, R.P., & Hanson, P.C. 2018. Oxygen dynamics control the burial of organic carbon in a eutrophic reservoir. *Limnology and Oceanography Letters* 3: 293–301.

Carpenter, J.H. 1965. The accuracy of the Winkler method for dissolved oxygen analysis 1. *Limnology and Oceanography* 10: 135–140.

Cooke, G.D., Welch, E.B., Peterson, S., & Nichols, S.A. 2005. *Restoration and Management of Lakes and Reservoirs*. CRC Press.

Davis, J.C. 1975. Minimal dissolved oxygen requirements of aquatic life with emphasis on Canadian species: a review. *J. Fish. Board Can.* 32(12): 2295–2332. doi:10.1139/f75-268.

Dodds, W.K. 2002. *Freshwater Ecology: Concepts and Environmental Applications*. Academic Press. San Diego, California.

Dodds, W.K., Bouska, W.W., Eitzmann, J.L., Pilger, T.J., Pitts, K.L., Riley, A.J., Schloesser, J.T., & Thornbrugh, D.J. 2009. Eutrophication of U.S. freshwaters: Analysis of potential economic damages. *Environ. Sci. Technol.* 43: 12–19. https://doi.org/10.1021/es801217q

Doubek, J.P., & Carey, C.C. 2017. Catchment, morphometric, and water quality characteristics differ between reservoirs and naturally formed lakes on a latitudinal gradient in the conterminous United States. *Inland Waters* 7(2): 171–180. https://doi.org/10.1080/20442041.2017.1293317

Doubek, J.P., Carey, C.C., Lavender, M., Winegardner, A.K., Beaulieu, M., Kelly, P.T., Pollard, A.I., Straile, D., & Stockwell, J.D. 2019. Calanoid copepod zooplankton density is positively associated with water residence time across the continental United States. *PLOS ONE* 14(1): 1–22. https://doi.org/10.1371/journal.pone.0209567

Dugan, H.A., Bartlett, S.L., Burke, S.M., Doubek, J.P., Krivak-Tetley, F.E., Skaff, N.K., Summers, J.C., Farrell, K.J., McCullough, I.M., Morales-Williams, A.M., Roberts, D.C., Ouyang, Z., Scordo, F., Hanson, P.C., & Weathers, K.C. 2017. Salting our freshwater lakes. *Proceedings of the National Academy of Sciences of the United States of America* 114(17): 4453–4458. https://doi.org/10.1073/pnas.1620211114

Effler, S.W., Perkins, M.G., & Brooks, C. 1986. The oxygen resources of the hypolimnion of ionically enriched Onondaga Lake, NY, U.S.A. *Water. Air. Soil Pollut.* 29: 93–108. https://doi.org/10.1007/BF00149332

Franklin, P.A. 2014. Dissolved oxygen criteria for freshwater fish in New Zealand: A revised approach. *New Zealand Journal of Marine and Freshwater Research* 48(1): 112–126. DOI: 10.1080/00288330.2013.827123

Gerling, A.B., Munger, Z.W., Doubek, J.P., Hamre, K.D., Gantzer, P.A., Little, J.C., & Carey, C.C. 2016. Whole-catchment manipulations of internal and external loading reveal the sensitivity of a century-old reservoir to hypoxia. *Ecosystems* 19(3): 555–571. https://doi.org/10.1007/s10021-015-9951-0

Gerten, D. & Adrian, R. 2002. Responses of lake temperatures to diverse North Atlantic Oscillation indices. In: R.G. Wetzel (ed.). *International Association of Theoretical and Applied Limnology Proceedings.* (pp. 1593–1596). *E Schweizerbart'sche Verlagsbuchhandlung.* Stuttgart, Germany.

Guarino, A.W.S., Branco, C.W.C., Diniz, G.P., & Rocha, R.J.S. 2005. Limnological characteristics of an old tropical reservoir (*Ribeirão das Lajes* Reservoir, RJ, Brazil). *Water* 17(2): 129–141.

Hayes, N.M., Deemer, B.R., Corman, J.R., Razavi, N.R., & Strock, K.E. 2017. Key differences between lakes and reservoirs modify climate signals: A case for a new conceptual model. *Limnol. Oceanogr. Lett.* 2: 47–62. https://doi.org/10.1002/lol2.10036

Hutchinson, G.E. 1957. *A Treatise on Limnology (Volume 1 - Geography, Physics and Chemistry).* John Wiley & Sons. 1015 pp.

Hynes, H.B.N. 1960. *The Biology of Polluted Waters.* Liverpool University Press. Liverpool, UK.

Jenny, J.-P., Francus, P., Normandeau, A., Lapointe, F., Perga, M.-E., Ojala, A., Schimmelmann, A., & Zolitschka, B. 2016. Global spread of hypoxia in freshwater

ecosystems during the last three centuries is caused by rising local human pressure. *Glob. Change Biol.* 22: 1481–1489. https://doi.org/10.1111/gcb.13193

Jewell, W.J. 1971. Aquatic weed decay: dissolved oxygen utilization and nitrogen and phosphorus regeneration. *J. Wat. Pollut. Control Fed.* 43: 1457–1467.

Judd, J. 1970. Lake stratification caused by runoff from street deicing. *Water Res.* 4(8): 521–532.

Judd, K.E., Adams, H.E., Bosch, N.S., Kostrzewski, J.M., Scott, C.E., Schultz, B.M., Wang, D.H., & Kling, G.W. 2005. A case history: Effects of mixing regime on nutrient dynamics and community structure in Third Sister Lake, Michigan during late winter and early spring 2003. *Lake and Reservoir Management* 21(3): 316–329.

Kennedy, R.H., Thornton K.W., & Ford D.E. 1985. Characterization of the reservoir ecosystem. In: D. Gunnison (ed.). *Microbial Processes in Reservoirs. Developments in Hydrobiology, Vol 27*. (pp. 27–38). Springer. Dordrecht, Netherlands.

Kramer, D.L. 1987. Dissolved oxygen and fish behavior. *Environ Biol Fish* 18: 81–92. https://doi.org/10.1007/BF00002597

Müller, B., Bryant, L.D., Matzinger, A., & Wüest, A. 2012. Hypolimnetic oxygen depletion in eutrophic lakes. *Environmental Science and Technology* 46(18): 9964–9971. https://doi.org/10.1021/es301422r

Novotny, E.V., Murphy, D., & Stefan, H.G. 2008. Increase of urban lake salinity by road deicing salt. *Science of the Total Environment* 406(1–2): 131–144. https://doi.org/10.1016/j.scitotenv.2008.07.037

Nürnberg, G.K. 1984. The prediction of internal phosphorus load in lakes with anoxic hypolimnia. *Limnol. Oceanogr.* 29: 111–124.

Nürnberg, G.K. 1995. Quantifying anoxia in lakes. *Limnol. Oceanogr.* 40: 1100–1111. https://doi.org/10.4319/lo.1995.40.6.1100

O'Boyle, S., McDermott, G., Noklegaard, T., & Wilkes, R. 2013. A simple index of trophic status in estuaries and coastal bays based on measurements of pH and dissolved oxygen. *Estuaries and Coasts* 36: 158–173. https://doi.org/10.1007/s12237-012-9553-4

Pereira, A., Tassin, B., & Jørgensen, S.E. 1994. A model for decomposition of the drown vegetation in an Amazonian reservoir. *Ecol. Model., State-of-the-Art in Ecological Modelling* P*roceedings of ISEM's 8th International Conference* 75–76: 447–458. https://doi.org/10.1016/0304-3800(94)90039-6

Richardson, D.C., Melles, S.J., Pilla, R.M., Hetherington, A.L., Knoll, L.B., Williamson, C.E., Kraemer, B.M., Jackson, J.R., Long, E.C., Moore, K., Rudstam, L.G. Rusak, J.A., Saros, J.E., Sharma, S., Strock, K.E., Weathers, K.C., & Wigdahl-Perry, C.R. 2017. Transparency, geomorphology and mixing regime explain variability in trends in lake temperature and stratification across Northeastern North America (1975–2014). *Water* (Switzerland) 9(6): 442. https://doi.org/10.3390/w9060442

Rivett, M.O., Cuthbert, M.O., Gamble, R., Connon, L.E., Pearson, A., Shepley, M.G., & Davis, J. 2016. Highway deicing salt dynamic runoff to surface water and subsequent infiltration to groundwater during severe UK winters. *Science of the Total Environment* 565: 324–338. https://doi.org/10.1016/j.scitotenv.2016.04.095

Schindler, D.W. 1977. Evolution of phosphorus limitation in lakes. *Science* 195(4275): 260–262.

Schindler, D.W., Hecky, R.E., Findlay, D.L., Stainton, M.P., Parker, B.R., Paterson, M.J., Beaty, K.G., Lyng, M., & Kasian, S.E.M. 2008. Eutrophication of lakes cannot be controlled by reducing nitrogen input: Results of a 37-year whole-ecosystem experiment. *Proc. Natl. Acad. Sci. U.S.A.* 105 (32): 11254–11258. https://www.pnas.org/content/105/32/11254 (accessed June 2, 2020).

Smith, V.H. 1982. The nitrogen and phosphorous dependence of algal biomass in lakes: An empirical and theoretical analysis. *Limnol. Oceanogr.* 27(6): 1101–1112.

Stoddard, J.L., Van Sickle, J., Herlihy, A.T., Brahney, J., Paulsen, S., Peck, D.V., Mitchell, R., & Pollard, A.I. 2016. Continental-scale increase in lake and stream phosphorus: Are oligotrophic systems disappearing in the United States? *Environmental Science and Technology* 50(7): 3409–3415. https://doi.org/10.1021/acs.est.5b05950

Stringfellow, W., Herr, J., Litton, G., Brunell, M., Borglin, S., Hanlon, J., Chen, C., Graham, J., Burks, R., Dahlgren, R., Kendall, C., Brown, R. & Quinn, N. 2009. Investigation of river eutrophication as part of a low dissolved oxygen total maximum daily load implementation. *Water Science and Technology* 59(1): 9–14. https://doi.org/10.2166/wst.2009.739

Thornton, K.W. 1990. Perspectives on reservoir limnology. In: K.W. Thornton, B.L. Kimmel, & F.E. Payne (eds.). (pp. 1–14). *Reservoir Limnology: Ecological Perspective*. John Wiley & Sons, Inc. New York, U.S.A.

U.S. EPA. 2017. *National Lakes Assessment 2012: Technical Report*. EPA 841-R-16-114. April 2017. U.S. Environmental Protection Agency Office of Wetlands, Oceans and Watersheds Office of Research and Development. Washington, D.C. https://www.epa.gov/sites/production/files/2017-04/documents/nationallakesassessment2012_technicalreport.pdf (accessed: April 21, 2020); data and metadata files available at https://www.epa.gov/national-aquatic-resource-surveys/nla (accessed April 30, 2020).

U.S. EPA. 2020. National Aquatic Resource Surveys. Indicators: Dissolved Oxygen. https://www.epa.gov/national-aquatic-resource-surveys/indicators-dissolved-oxygen (accessed: April 4, 2020).

Vachon, D., Langenegger, T., Donis, D., & McGinnis, D.F. 2019. Influence of water column stratification and mixing patterns on the fate of methane produced in deep sediments of a small eutrophic lake. *Limnology and Oceanography* 64(5): 2114–2128. https://doi.org/10.1002/lno.11172

Wang, W. 1980. Fractionation of sediment oxygen demand. *Water Res.* 14: 603–612. https://doi.org/10.1016/0043-1354(80)90118-9

Welch, E.B. 1969. *Factors Initiating Phytoplankton Blooms and Resulting Effects on Dissolved Oxygen in Duwamish River Estuary, Seattle, Washington.* Water Supply Paper 1873A. U.S., Geological Survey. https://doi.org/10.3133/wsp1873A

Wetzel, R.G. 2001. *Limnology: Lake and River Ecosystems.* 3rd edition. Springer Verlag. New York, NY.

Wetzel, R.G. & Likens, G.E. 1991. *Limnological Analyses.* 2nd edition. Springer-Verlag. New York, NY.

Winkler, L.W. 1888. *Die Bestimmung des im Wasser gelösten Sauerstoffes. Chem. Ber.* 21: 2843–2855.

<u>**Appendix A**</u>

Request for a Scientific Peer Review of DEQ's Assessment Methodology for Lake/Reservoir Dissolved Oxygen

Introduction

States, territories, and authorized tribes must ensure their water quality standards (WQS) are met. As mandated by the Clean Water Act, they must monitor parameters that enable criteria assessment, report assessment information to the public and to the Environmental Protection Agency (EPA), and identify impaired waters (those not meeting WQS) so that appropriate measures can be implemented to restore designated uses (e.g., Total Maximum Daily Load implementation plans). To provide transparency, jurisdictions must describe how they determine whether their WQS are being met. This description is usually not included in the WQS regulation, but instead is typically published separately in a public-facing document devoted solely to implementation procedures. The Virginia Department of Environmental Quality (DEQ) presents its water quality assessment methodology in a document called the Water Quality Assessment Guidance Manual (hereafter referred to as the "guidance manual"). The guidance manual not only shows the public how DEQ determines whether a water body meets its designated uses, but it also serves as an instruction manual for staff so that surface waters across the Commonwealth are assessed consistently and defensibly.

As with anything, assessment methodology can be strong in some areas and weak in others (Table 1). Frequently, weaknesses are detected by new staff members, who tend to read the instructions from a different perspective than more veteran staff members. Sometimes weaknesses are brought to DEQ's attention by informed members of the public, EPA, or the regulated community. Updates are made every two years to the guidance manual to address weaknesses and to ensure that procedures reflect newly adopted or modified WQS and updated guidelines from EPA. DEQ's Integrated Water Quality Assessment Report, which includes the Impaired Waters List, is largely the end result of the data analyses described in the guidance manual.

Table 1. Distinguishing marks of strong and weak assessment methodology

	Strong assessment methodology		Weak assessment methodology
1	Is written clearly and thoroughly enough to be easily understood by the informed public and staff.	1	Is poorly written and leaves out key details; assessment results cannot be replicated just by reading the document. Example: The methodology document says that "samples will be assessed against the criterion" but does not specify the decision rules used to determine WQS non-compliance.
2	Follows the letter of the WQS.	2	Takes considerable creative license with the WQS. Example: The WQS stipulates that a criterion cannot be exceeded more than 10% of the time, but assessment staff are instructed to evaluate the 90th percentile of a dataset against the criterion.
3	Is consistent with how the WQS were developed.	3	Is not consistent with how the WQS were developed. Example: A criterion was developed from monthly arithmetic means of field samples, but staff are instructed to compare annual geometric means of field samples to the criterion.
4	Does not require staff to "use best professional judgment" for key decision points.	4	Is so subjective that assessment decisions differ among staff. Example: Staff are given the discretion to screen out samples taken during "extreme weather events", with no instructions provided on how staff should make this determination.
5	Encourages staff to use "all available and appropriate data."	5	Uses a very narrow definition of "appropriate data", one that excludes potentially useful datasets from the assessment. Example: Staff are instructed to use only agency data, when sister agencies and reputable citizen groups collect similar datasets.
6	Minimizes type II errors as much as possible given available monitoring resources.	6	Allows staff to base assessment on very small and potentially outdated datasets with no safe guards against type II errors. Example: Staff are allowed to decide a water body is meeting a criterion based on a single grab sample.

In 2005, the Academic Advisory Committee (AAC) submitted a report to DEQ entitled "Freshwater Nutrient Criteria" that was used to inform the development of nutrient criteria for a large subset of Virginia lakes/reservoirs studied by the AAC. These water bodies were deemed "significant" and include the Commonwealth's two natural lakes and 122 man-made reservoirs. A significant lake/reservoir is defined as a publicly accessible lake/reservoir that is a

public water supply and/or 100 acres or more in size. Total chlorophyll-a and total phosphorus (TP) criteria became effective for these lakes/reservoirs in 2007 and are listed in 9 VAC 25-260-187 and 9 VAC 25-260-310 of the WQS regulation. Along with providing the agency with recommendations for nutrient criteria, the AAC also advised DEQ on how dissolved oxygen (DO) criteria should be applied to lakes/reservoirs (Appendix C in AAC, 2005). These recommendations were used to support the inclusion of language in 9 VAC25-260-50 stipulating that DO and pH apply only to the epilimnion of a thermally stratified man-made lake or reservoir listed in 9 VAC 25-260-187 (hereafter referred to as Section 187 lakes/reservoirs). Prior to the adoption of this language, DEQ applied DO criteria to the entire water column of both stratified and non-stratified water bodies, which resulted in a number of reservoirs being categorized as impaired when their low DO condition was not due to anthropogenic inputs (e.g., nutrients).

Apart from these recommendations, however, the AAC did not find fault with DEQ's assessment procedures for DO in lakes/reservoirs. The AAC's report describes DEQ's methodology as "sound" and "scientifically defensible". It should be noted that the AAC report does not specifically describe DEQ's DO assessment methodology, but the 2004 DEQ Water Quality Assessment Guidance Manual is cited in the reference section. The 2004 guidance manual instructed DEQ staff to apply DO criteria to lakes/reservoirs in the following manner (bolding added):

Epilimnion:

For each monitoring station, all DO data collected in the epilimnion (delineated using temperature profile or assumed to be the upper 1/3 of the water column) will be **aggregated** and assessed. If the violation rate exceeds 10%, the assessment unit or entire lake/reservoir will be assessed as impaired partially due to one or more pollutants from anthropogenic sources and will be placed in category 5A for TMDL development. If the violation rate is less than 10.5%, assess the hypolimnion.

Hypolimnion:

For each monitoring station, all data collected in the hypolimnion (delineated using temperature profile or assumed to be the lower 2/3 of the water column) will be **aggregated** and assessed. If the violation rate exceeds 10.5%, the lake/reservoir will be assessed as impaired partially due to one or more pollutants. Calculate the Tropic State Indices to determine whether the violations are due to pollutants from anthropogenic sources or natural sources. If the violation rate is less than 10.5%, the assessment unit or lake will be assessed as fully supporting.

Non-stratfied Lakes - Water Column Treated as Homogenous Unit:

If the lake is not stratified (T_t and T_b differential <4ºC) all DO data in the water column will be **aggregated** and assessed. If the violation rate exceeds 10.5%, the assessment unit or entire lake/reservoir will be assessed as impaired partially due to one or more pollutants from anthropogenic sources and will be placed in category 5A for TMDL development. If the violation rate is less than 10.5%, the assessment unit or lake will be assessed as fully supporting.

After 2007, the water quality assessment guidance manual was updated in light of the aforementioned adoptions to the lakes/reservoirs WQS. The 2018 guidance manual instructed DEQ staff to apply DO criteria to lakes/reservoirs in the following manner (bolding in original):

> The 10.5% rule is applicable to assessments for the minimum dissolved oxygen criterion in all assessed lakes and reservoirs for each lake monitoring year. For **§187** lakes/reservoirs, dissolved oxygen samples taken for all months within the lake monitoring year, at all stations within a given lake or reservoir, are assessed only in the epilimnion if the water body is thermally stratified. If not stratified, dissolved oxygen should be assessed throughout the water column. A lake or reservoir is considered stratified if there is a difference of 1ºC /meter. If the differential is < 1ºC /meter, the lake is not considered stratified. Lakes/Reservoirs not listed in **§187 should have all DO samples assessed regardless of thermal stratification determination.** Two or more exceedances and >10.5% exceedance of total samples are required before a water body is listed as impaired for the minimum dissolved oxygen criterion (4 mg/l for most freshwater lakes and reservoirs) under § 62.1-44.19:5 and 7 of the Code of Virginia.

There are some key differences in the 2004 and 2018 instructions besides the treatment of hypolimnion data. First, the 2004 guidance manual instructed staff to apply the "10.5% rule"[1] to "each monitoring station" dataset in a lake/reservoir, while the 2018 guidance directed staff to apply the rule to "all dissolved oxygen samples" within a lake/reservoir. Secondly, the 2004 guidance manual uses the term "aggregate," while that term does not appear in the more recent guidance manual. The AAC members who may have reviewed the lakes/reservoirs section of the 2004 guidance manual may have assumed the term "aggregate" meant that DEQ staff were being instructed to average epilimnetic samples (or water column samples in non-stratified water bodies) by monitoring station. However, DEQ staff interpreted "aggregate" to mean that samples would be assessed collectively, with no averaging. "Aggregate" was likely dropped in subsequent versions of the guidance manual due to the ambiguity of this term.

It is quite possible that the AAC would have still determined that DEQ's DO assessment methodology for lakes/reservoirs was scientifically defensible even if "aggregate" had not been used in the 2004 guidance. But there is enough of a difference between the 2004 and 2018

---

[1] The "10.5% rule" is based on guidelines from the Environmental Protection Agency (EPA, 2002) within the context of conventional parameter 305(b) assessments. It is interchangeable with the "10% rule".

guidance instructions to warrant a new AAC review to ensure that DEQ's procedure is sound and scientifically defensible. This review would also be helpful given DEQ's desire to update the guidance for the [Monitoring and Assessment of Lakes and Reservoirs](#), which provides general guidance for collecting and analyzing lake data. This guidance document was last finalized in 2009. During the spring of 2019, changes were proposed for this document and public comments were submitted to DEQ in response, but the changes were withdrawn due to staff uncertainty. The expertise of the AAC may help resolve this uncertainty.

DEQ requests a review of its current process for assessing DO in lakes/reservoirs to make sure that the AAC still considers it "sound" and "scientifically defensible."   DEQ would also appreciate a technical review of an alternative procedure that reflects staff recommendations combined with the suggestions submitted by a commenter regarding pH criteria assessment.   Additionally, advice is requested on how DEQ should handle the assessment of low DO conditions caused by incomplete fall turnover. The AAC is asked to consider DEQ's assessment methodology and any changes the AAC recommends with respect to the distinguishing marks listed in Table 1.   This paper concludes with questions to assist the AAC with its responses.

Current Process

DEQ staff typically create assessment units for water bodies they judge to be relatively homogeneous—one where surrounding land uses and riverine inputs are expected to be relatively uniform and one where physical dynamics (e.g., thermal stratification) are expected to be similar throughout. While DEQ staff tend to establish riverine assessment units around individual DEQ stations (one ambient water quality monitoring station per assessment unit), it is normal for multiple stations to be located in a single lake/reservoir assessment unit[2]. This is especially true for lakes/reservoirs that are monitored by other data collectors besides DEQ, like citizen groups. The number of non-agency stations is expected to multiply as citizen scientists continue to swell in number. Thus, it is important that there be standardization in the handling of different lake/reservoir datasets.

DEQ staff pool data by taking all measurements taken in the epilimnion (for Section 187 lakes/reservoirs) or in the water column (for non-Section 187 lakes or when Section 187

---

[2] Out of the 137 lake/reservoir assessment units monitored by DEQ over the 2013 to 2018 period, 20% were monitored at two stations and 7% were monitored at three or more stations. When citizen monitoring stations are included, 20% of monitored lake/reservoir assessment units have two stations and 14% have three or more stations. One lake/reservoir assessment unit has nine monitoring stations (five managed by DEQ and four managed by citizens).

lakes/reservoirs are not stratified) over the six-year assessment period (Figure 1). Vertical temperature profiles enable assessors to determine the bottommost depth of the epiliminion at a station during a monitoring event. Epilimnetic samples are then compared to the minimum DO criterion (4.0 mg/l for most lakes/reservoirs). If more than 10% of these samples are below the criterion, then the lake/reservoir assessment unit will be deemed impaired for DO.

Station A (DEQ)
- 7 monitoring events
- 2 DO exceedances
- 31 epilimnetic DO
  measurements over all
  monitoring

Station B (Non-Agency)
- 4 monitoring events
- 0 DO exceedances
- 4 epilimnetic DO
  measurements over
  all monitoring events

Station C (DEQ)
- 7 monitoring events
- 3 DO exceedances
- 32 epilimnetic DO
  measurements over all
  monitoring events

DO Assessment Summary
2+0+3 = 5 DO exceedances
31+4+32 = 67 epilimnetic DO measurements
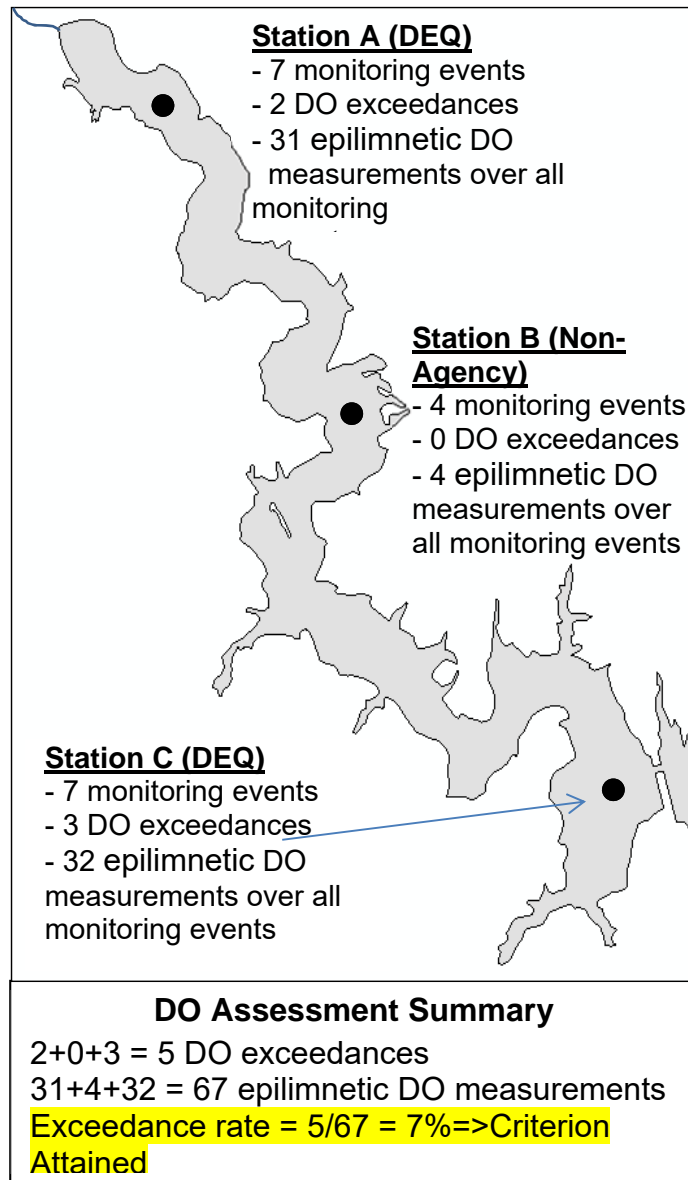Exceedance rate = 5/67 = 7%=>Criterion Attained

Figure 1. Illustration of how DEQ currently assesses DO samples in a hypothetical lake/reservoir assessment unit. Table 2 shows the hypothetical assessment dataset.

Table 2. Current assessment procedure applied to DO (mg/l) samples for the hypothetical stations shown in Figure 1. Underlined measurements are the bottommost epilimnetic measurements. Only the data in the gray box would be used to a calculate an exceedance rate for the assessment unit. Red values are DO criterion exceedance (< 4.0 mg/l).

Station A (monitored by DEQ)							
Depth	4/23/2015	5/26/2015	6/23/2015	7/9/2015	8/25/2015	9/29/2015	10/30/2015
0.3	8.9	7.8	10.0	4.1	8.6	7.6	8.0
1.0	8.8	7.7	9.9	4.0	8.3	7.6	7.9
2.0	8.7	7.5	9.8	4.0	8.0	7.5	7.9
3.0	6.8	7.4	9.8	3.7	5.2	5.0	7.6
4.0	6.7	7.3	9.8	3.7	3.5	4.0	7.5
no. of epilimnetic/water column measurements	5	5	5	5	3	3	5
no. of epilimnetic/water column exceedances	0	0	0	2	0	0	0
Station C (monitored by DEQ)							
Depth	4/23/2015	5/26/2015	6/23/2015	7/9/2015	8/25/2015	9/29/2015	10/30/2015
0.3	9.6	7.9	9.0	4.6	8.5	7.6	7.8
1.0	9.5	7.8	8.9	4.0	8.5	7.5	7.7
2.0	8.8	7.5	8.8	3.6	8.2	7.4	7.6
3.0	7.8	7.5	8.8	3.5	5.0	5.5	7.5
4.0	7.0	6.7	7.8	3.5	4.2	3.9	7.4
no. of epilimnetic/water column measurements	5	5	5	5	3	4	5
no. of epilimnetic/water column exceedances	0	0	0	3	0	0	0
Station B (monitored by non-DEQ data collector)							
Depth	7/6/2015	8/15/2016	9/15/2016	10/15/2016			
0.3	5.0	7.5	6.2	6.6			
1.0							
2.0							
3.0							
no. of epilimnetic/water column measurements	1	1	1	1			
no. of epilemnetic/water column exceedances	1	1	1	1			

total no. of epilimnetic/water column measurements = 67   total no. of epilimnetic/water column exceedances = 5   Exceedance rate = 7%

There have been no objections expressed from the public or EPA about this assessment approach, but questions have been raised internally about whether pooling data in this fashion would be considered a "best practice." One concern is that this approach does not distinguish dependent samples (those collected on the same day, whether at different stations or at the same station within minutes and meters of each other) from independent samples (samples collected a month a part). While EPA guidance does not explicitly restrict the "10% rule" to independent samples and applying it to a collection of dependent samples is certainly a defensible approach, using it to evaluate a mix of dependent and independent samples seems questionable. When temporally independent samples meet the 10% rule, an assessor can conclude that a water body was in compliance with a particular criterion at least 90% of the time a water body was visited. When spatially dependent samples meet the 10% rule, an assessor can conclude that at least 90% of the habitat was in compliance with the criterion. But a dynamic mix of independent and dependent samples does not lend itself to a definitive conclusion about a water body's condition. For instance, if more than 10% of such samples exceed the criterion, it could be that the water body is frequently in non-compliance at one or more depths in the epiliminion. Or it could be that a severe algal bloom resulted in low DO throughout the water column a couple of times during the two years a water body was monitored over the assessment period, and that 100% of the epilimnion was suitable for aquatic life on the rest of the monitored days. Thus, the ratio of the number of exceedances to the number of epilimnetic/water column measurements does not readily communicate whether the aquatic life in a water body is exposed to partially suboptimal habitat frequently or completely suboptimal habitat sporadically. By determining that both of these conditions indicate impairment, the agency is assuming they are equivalent losses of the aquatic life use. Perhaps that is not the case.

The questionable defensibility of this pooling approach was recently brought to staff attention in the context of nutrient criteria assessment in lakes/reservoirs. Until recently, staff has been processing nutrient datasets similar to how they have been processing DO datasets. However, unlike DO criteria, nutrient criteria are almost always paired with basic implementation procedures in the WQS to help ensure that these criteria are assessed in a manner consistent with their derivation. The following language can be found in Section 187 of the WQS (bolding added):

> The 90th percentile of the chlorophyll a data collected at one meter or less within the lacustrine portion of the man-made lake or reservoir between April 1 and October 31 shall not exceed the chlorophyll a criterion for that waterbody in each of the two most recent monitoring years that chlorophyll a data are available. For a waterbody that received algicide treatment, the median of the total phosphorus data collected at one meter or less within the lacustrine portion of the man-made lake or reservoir between April 1 and October 31 shall not exceed the total

phosphorus criterion in each of the two most recent monitoring years that total phosphorus data are available.

Monitoring data used for assessment shall be from sampling location(s) within the lacustrine portion where **observations are evenly distributed over the seven months from April 1 through October 31** and are in locations that are representative, either individually or collectively, of the condition of the man-made lake or reservoir.

This bolded language indicates the importance of the temporal spacing of assessment samples. This is likely because AAC researchers calculated monthly "lake-medians" for the metrics they examined (TP, total nitrogen, chlorophyll-a and total suspended solids) over the April to October period "so as to generate…values that better represent DEQ's lake monitoring schedule" (AAC, 2005). The monthly values were used by the AAC to relate nutrient condition to fisheries status for each water body.   Thus, assessing pooled observations with no temporal averaging is inconsistent with the letter and intent of the WQS (weakness #2 and #3 in Table 1; see Table 3 to see how calculating a 90th percentile on pooled chlorophyll-a samples can lead to a different assessment decision than calculating a 90th percentile on monthly medians). This inconsistency has been addressed in the draft 2020 assessment guidance manual, which instructs staff to calculate monthly medians of TP and chlorophyll-a data before calculating the appropriate descriptive statistics (April-October median and 90th percentile, respectively).

Table 3. Comparison of 90th percentiles calculated on pooled and monthly aggregated chlorophyll-a samples (µg/l) taken at the hypothetical stations shown in Figure 1. The median of the values in the box is used to represent the chlorophyll-a concentration for July.   Red value is an exceedance of the hypothetical criterion (35 µg/l).

Monitoring Date	Station A	Station B	Station C	Monthly Median
4/23/2015	40		25	33
5/26/2015	35		37	36
6/23/2015	35		22	29
7/6/2015		30		
7/9/2015	30		45	30
8/25/2015	10		9	10
9/29/2015	15		10	13
10/30/2015	10		16	13
90th percentile of pooled samples	**39**			
90th percentile of monthly medians	**34**			

This modification to the guidance manual has led to questions over whether DO (as well as pH) data should be treated in a similar fashion as nutrient data for the sake of methodological consistency. If there are serious defensibility issues with the current pooling approach of DO data, then an alternative should be developed and reviewed in time for the next guidance manual update (slated for 2021).

Alternative Approach

One alternative procedure would begin by establishing the location of the thermocline at an individual station (the point where the temperature change between depths is $\geq 1°C$), as is done currently. Epilimnetic or water column DO samples along the vertical profile would then be aggregated via the median. This would represent the average DO at this station, but in a way that does not presume how the samples are distributed or allow outlier samples to have undue influence. The next step would be the aggregation of all "epilimnon/water column medians" generated at stations visited during the same monitoring run. A median would again be used for the same reasons as above. The end result would be a snapshot of a water body's DO concentration for a particular monitoring run. This snapshot would then be assessed in combination with all the other snapshots taken over the assessment period (see Table 4 for an illustration of this process). The end result would be an exceedance rate that could be reasonably interpreted as "Aquatic life had unsuitable habitat X% of the time when the lake/reservoir was monitored." This exceedance rate would then be evaluated against the 10% rule to determine criteria attainment status for the lake/reservoir assessment unit.

One advantage of this approach relative to the current one is that it would ensure that all assessment units are assessed with temporally independent samples rather than a variable mix of spatially and temporally autocorrelated measurements. The assessment summary for a lake/reservoir assessment unit monitored at three stations for two years will be directly comparable to the assessment summary for a lake/reservoir monitored at one station for the same period of time. This approach also prevents a single deep station—which in theory could have many samples to contribute to the denominator of the exceedance rate—from having more influence over the assessment decision than stations that are located in shallower waters. While TP and chlorophyll-a criteria only apply to the lacustrine zone of a Section 187 water body, DO criteria apply throughout its entirety, including the littoral areas and the transitional zone. Using the current approach, these habitats would need to be annexed into their own assessment units in large water bodies to ensure that all parts of the water body are protected. But spatially refined assessment units would not be as important for the alternative approach

Table 4. Alternative assessment procedure applied to DO (mg/l) samples for the hypothetical stations shown in Figure 1. Underlined measurements are the bottommost epilimnetic measurements. Only the data in the gray box would be used to an calculate exceedance rate for the assessment unit. Red values are DO criterion exceedance (< 4.0 mg/l).

	Station A (monitored by DEQ)						
Depth	4/23/2015	5/26/2015	6/23/2015	7/9/2015	8/25/2015	9/29/2015	10/30/2015
0.3	8.9	7.8	10.0	4.1	8.6	7.6	8.0
1.0	8.8	7.7	9.9	4.0	8.3	7.6	7.9
2.0	8.7	7.5	9.8	4.0	8.0	7.5	7.9
3.0	6.8	7.4	9.8	3.7	5.2	5.0	7.6
4.0	6.7	7.3	9.8	3.7	3.5	4.0	7.5
epilimnion/water column DO median	8.7	7.5	9.8	4.0	8.3	7.6	7.9
	Station C (monitored by DEQ)						
Depth	4/23/2015	5/26/2015	6/23/2015	7/9/2015	8/25/2015	9/29/2015	10/30/2015
0.3	9.6	7.9	9.0	4.6	8.5	7.6	7.8
1.0	9.5	7.8	8.9	4.0	8.5	7.5	7.7
2.0	8.8	7.5	8.8	3.6	8.2	7.4	7.6
3.0	7.8	7.5	8.8	3.5	5.0	5.5	7.5
4.0	7.0	6.7	7.8	3.5	4.2	3.9	7.4
epilimnion/water column DO median	8.8	7.5	8.8	3.6	8.5	7.5	7.6
	Stations A & C aggregated by shared date						
	4/23/2015	5/26/2015	6/23/2015	7/9/2015	8/25/2015	9/29/2015	10/30/2015
median of epilimnion/ water column medians	8.8	7.5	9.3	3.8	8.4	7.5	7.8
	Station B (monitored by non-DEQ data collector)						
Depth	7/6/2015	8/15/2016	9/15/2016	10/15/2016			
0.3	5.0	7.5	6.2	6.6			
1.0							
2.0							
3.0							
epilimnion/water column DO median	5.0	7.5	6.2	6.6			

no. of epilimnion/water column DO medians lower than criterion = 1

no. of monitoring events = 11
exceedance rate = 9%

since each station would be weighted equally, regardless of how deep they are. Low DO occurring in a shallow, quiescent cove would not get "swamped out" by all the not-as-low DO samples taken in the well-mixed center of the lake/reservoir, where the water epilimnion/water column is probably deeper. This is especially true if the cove station is sampled on a date when no other samples are taken in the water body, since the sample(s) taken at that station will represent the entire water body on that date.

However, this advantage can also be seen as a disadvantage. It can be argued that deeper stations should be weighted more than shallower stations since they represent more of the three-dimensional habitat. Additionally, a station that is only monitored at the surface (such as one visited by a citizen scientist) would be represented the same in the assessment as a DEQ station that is sampled throughout the entire epilimnion (see the contrast between Station A/ C samples and Station B samples in Table 4). Surface samples may be adequately representative of the epilimnion/water column in well-mixed shallow water bodies, but they may lead to an overestimation of the epilimnion/water column DO concentration in deep water bodies. Thus, in lakes/reservoirs that are monitored by "surface only" data collectors, the alternative approach could lead to more false negatives (i.e., impaired water bodies mistakenly categorized as non-impaired water bodies, see weakness #6 in Table 1). The guidance manual could include a rule that a surface-only DO dataset can only be used in an assessment when an assessor determines they are adequately representative of the epilimnion/water column. But the adoption of such a rule may introduce inconsistency in assessments performed by different staff and result in the exclusion of some non-agency datasets[3] (see weaknesses #4 and #5 in Table 1).

Incomplete Fall Turnover

DEQ Central Office staff recently determined that the majority of lakes/reservoirs were not being sub-segmented into individual assessment units, but instead are being assessed holistically. The guidance manual gives staff the discretion to assess holistically. It is reasonable from a monitoring efficiency standpoint to avoid a highly refined sub-segmentation scheme, since all the sub-segments (or assessment units) would have to be monitored over an assessment period for the entire lake/reservoir to be assessed. Furthermore, it can be difficult

---

[3] Of the 41 most recently active citizen monitoring stations where "assessable" lake/reservoir DO datasets were generated, only five stations are associated with DO datasets that would be categorized as "surface only". But four of them were the sole station in their respective assessment unit. At ten stations, citizen scientists took "bottom only" measurements during the winter and spring while sampling the upper three meters during the rest of the year. Full vertical profiles were consistently generated at most of the remaining stations.

to determine the boundaries of sub-habitats in a large lake/reservoir when bathymetry data are not available. For large Section 187 lakes/reservoirs (such as Lake Anna), the lacustrine zone (where nutrient criteria apply) is typically carved out separately from transitional waters. "Fingers" may also be carved out. But this is typically the extent of sub-segmentation in large lakes/reservoirs[4].

DEQ Central Office staff proposed updating the Monitoring and Assessment Guidance for Lakes and Reservoirs (DEQ, 2009) with language that more strongly recommends sub-segmentation, since staff recognized that vulnerable stations (those prone to DO, pH, and temperature exceedances) may be masked by assessing medium-sized and large lakes/reservoirs holistically. Staff could delineate assessment units around individual stations as they currently do for many riverine segments or they could bracket one assessment unit around multiple stations while creating a separate assessment unit for another station.   It is not expected that sub-segmentation will lead to more impaired waters listings for the majority of the Commonwealth's lakes/reservoirs. But staff have determined that more listings are likely to occur in deep reservoirs (those greater than 15 meters in depth). In such water bodies, non-stratified but hypoxic water columns tend to be encountered as fall turnover begins (see the example in Table 5). Because a thermocline is not detected during these events, the DO criterion applies throughout the water column. When station datasets are pooled across the entire water body and over the entire six-year assessment period (the current procedure), there are generally enough "non-exceedance" DO measurements in the dataset to counterbalance the exceedances that emerge during fall turnover. But when station datasets are assessed individually, stations that are prone to fall turnover-induced hypoxia have a high likelihood of being flagged as impaired.

The obvious solution to this problem is to simply maintain the current practice of assessing deep water bodies holistically and using individual measurements (the current procedure) rather than aggregations of measurements (the alternative procedure) to establish criterion attainment status. However, this seems like a "Band-Aid" solution that does not address the larger problem of potentially masked impairments. While fall turnover-induced low DO is not a problem that can be addressed with a Total Maximum Daily Load, it could still be harmful to aquatic life if there are no oxygenated refuges available. Moreover, it is possible that fall turnover itself could mask low DO caused by anthropogenic influences.   It would be ideal to have a scheme that assessment staff could follow—one that does not rely too much on "best professional judgement"(see weakness #4 in Table 1)—to determine whether exceedances at a station are due solely to incomplete fall turnover. This scheme could also

---

[4] The average size of a lake/reservoir assessment unit (to date) is 460 acres. The largest assessment unit is 30,665 acres and is located in Kerr Reservoir.

Table 5. DO (mg/l) vertical profiles taken at Smith Mountain Lake station 4AROA175.63 for a single monitoring year. Underlined measurements are the bottommost epilimnetic measurements. Red values are exceedances of the minimum DO criterion (4.0 mg/l).

Depth	4/17/2013	5/1/2013	6/4/2013	7/24/2013	9/9/2013	10/8/2013
0.3	11.5	11.4	9.7	10.5	9.6	9.3
1.0	11.8	11.5	10.0	10.9	10.3	9.4
2.0	11.9	11.5	10.2	11.0	10.6	9.4
3.0	11.9	11.5	10.4	10.9	10.2	9.4
4.0	11.7	10.9	10.0	8.6	5.0	9.3
5.0	11.7	10.5	8.0	6.6	2.8	9.2
6.0	11.7	10.1	7.1	5.9	0.6	4.2
7.0	11.8	8.3	6.2	5.0	0.2	3.4
8.0	11.8	8.1	5.6	4.3	0.3	1.9
9.0	11.8	9.0	5.5	4.0	0.5	1.6
10.0	11.4	9.6	5.8	3.1	0.6	1.4
11.0	11.3	9.8	5.9	2.3	0.7	1.3
12.0	11.1	9.9	5.9	2.4	1.0	1.3
13.0	10.9	9.8	6.4	3.4	1.3	1.3
14.0	10.7	9.7	6.6	3.9	1.4	1.2
15.0	10.4	9.7	7.0	4.9	1.5	1.3
16.0	10.3	9.6	7.5	4.6	1.6	1.3
17.0	10.2	9.4	7.8	4.8	1.5	1.3
18.0	10.1	9.4	7.9	5.0	1.4	1.3
19.0	10.0	9.3	8.0	5.0	1.3	1.3
20.0	9.8	9.1	7.9	4.9	1.2	1.3
21.0	9.7	9.0	7.8	4.5	0.9	1.3
22.0	9.5	8.9		4.3	0.8	1.3
23.0	9.4	8.8		4.1	0.7	1.2
24.0	9.3	8.7		3.2	0.6	1.3
25.0	9.2	8.5		3.0	0.3	1.3
26.0	9.0	8.5		2.9	0.1	1.2
27.0	8.8	8.4		2.6	0.0	1.3
28.0	8.6	8.2		2.2	0.0	1.3
29.0	8.4	8.0		1.7	0.0	1.2
30.0	8.2	7.8		1.3	0.0	1.4
31.0	8.1	7.6		1.0	0.0	1.4
32.0	8.0	7.5			0.0	1.4
33.0		7.3			0.0	1.3
34.0		7.1			0.0	1.4
35.0		7.1				2.1
no. of exceedances in epilimnion/water column	0	0	0	0	0	29
no. of measurements in epilimnion/water column	6	7	4	4	4	36
epilimnion/water column DO median	11.8	11.4	10.1	10.9	10.3	1.3

instruct staff on how to downweight fall turnover DO data while still being protective of living resources.

Questions for the AAC

1. Is DEQ's current procedure for DO assessments in lakes/reservoirs sound and scientifically defensible? Does the AAC have any concerns about this procedure besides those already mentioned?

2. Is the alternative procedure sound and scientifically defensible? Is it a substantively better approach than the current one? Do its advantages outweigh its disadvantages?

3. Is there another alternative approach that DEQ should consider?

4. Should DEQ restrict DO assessments to vertical profile datasets in lakes/reservoirs or is it appropriate for a surface-only DO dataset to form the basis of an assessment for a lake/reservoir? If the latter, how should surface-only data (or data from shallow profiles) be used in assessments?

4. Can the AAC recommend a set of instructions that staff could use 1) to determine whether exceedances at a station are due solely to incomplete fall turnover , 2) to determine to what degree low DO during fall turnover might be exacerbated by pollution, and 3) to downweight fall turnover DO data while still being protective of living resources?

Literature Cited

Academic Advisory Committee (2005) January 2005 Report of the Academic Advisory Committee to Virginia Department of Evironmental Quality: Freshwater Nutrient Criteria. Submitted to Division of Water Programs, Virginia DEQ on February 7, 2005. 120 pg.

Virginia Department of Environmental Quality (2009) Monitoring and Assessment Lakes and Reservoirs. Water Guidance Memo No. 09-2005. Richmond, Virginia. 31 pg.

Virginia Department of Environmental Quality (2018) 2018 Water Quality Assessment Guidance Manual. Water Guidance Memo No. GM18-2001. Richmond, Virginia. 110 pg.

U.S Environmental Protection Agency (2002) Consolidated Assessment and Listing Methodology: Toward a Compendium of Best Practices. Office of Wetlands, Oceans, and Watersheds. Washington, DC. 375 pg.