Water Quality Assessments Program

2022 Integrated Report Assessment Methodology

Draft: Dissolved Oxygen Delisting Methodology Development

Oct. 8, 2020

Contents

Figures

Executive Summary

For the 2018/2020 Integrated Report (IR), the IR improvement team established a new delisting methodology to address the 2012 report's "lack of specific detail for delisting" and to create a process that is transparent and consistent for stakeholders. The 2018/2020 Integrated Report adopted the binomial test as a statistical approach that mirrors the listing methodology. This delisting methodology is inadequate for dissolved oxygen delistings. This white paper outlines the inadequacies of the delisting process as applied to dissolved oxygen, presents analysis on different considerations made during development of a dissolved oxygen specific delisting procedure, and outlines open questions needing external stakeholder engagement.

Overall:

- The uniqueness of the Dissolved Oxygen (DO) standard and resulting assessment methodology makes the statistical binomial test inappropriate for delisting.
- Grab sampling programs have a high probability of incorrectly identifying impaired assessment units as non-impaired.
- Continuously collected dissolved oxygen data provides a more robust picture of a waterbody's dissolved oxygen status at the Assessment Unit level.
- Assessment staff at DEQ propose the following dissolved oxygen delisting methodology:

    A. Full critical period (or spawning option)
    1. Continuous metrics analysis results in a category 2 designation of attaining criteria *and*
    2. Dataset must include a minimum 3 years of data that have 80% of the critical period (July 1 through September 30) represented and criteria are met in each of the 3 years.

    Or

    B. Short term probe deployments
    1. Dataset includes minimum 3 years of data that contains at least 5 full days of continuous dissolved oxygen per critical period month per year *and*
    2. <10% (binomial) of samples are below criteria as described in the Delisting Waterbodies Section of the Integrated Report Assessment methodology.

- To allow for monitoring partners time to adjust sampling programs, DEQ is proposing a temporary delisting methodology based on grab samples for the 2022 and 2024 IR cycles only:
    A. Dataset includes 3 years of data that contain at least 2 results for each critical period month.
    B. There are no excursions of any applicable criteria

Background

Non-Applicability of Current Delisting Process

For the 2018/2020 Integrated Report (IR), the IR improvement team established a new delisting methodology to address the 2012 report's "lack of specific detail for delisting" and to create a process that is transparent and consistent for stakeholders[1]. The 2018/2020 Integrated Report adopted the binomial test as a statistical approach that mirrors the listing methodology. The binomial process adopted in the 2018/2020 Integrated Report states that waters will be removed from the 303(d) list if the number of sample excursions of the criteria supports a 90% confidence level that the waterbody exceeds criteria less than 10% of the time. See Appendix A: Critical Values for Delisting Conventional Pollutants for the conventional parameter delisting methodology from the 2018/2020 Assessment Methodology.

This delisting methodology is inadequate for dissolved oxygen delistings. EPA's document proposing the binomial test for listing identifies three assumptions that need to be true for the binomial model to be appropriate[2]:

1. The response can have only two outcomes (e.g. attainment and exceedance);
2. The underlying probability of exceedance, p, remains constant from sample to sample; and
3. Samples are obtained through an independent random sampling design.

Dissolved Oxygen datasets violate two of the three assumptions used in the statistical model. First, assumption 2 requires that the waterbody's "exceedance probabilities at the monitoring site not change seasonally."[2] The probability of a dissolved oxygen excursion varies at both a seasonal and diurnal scale. Secondly, data generated through continuous probes are not an independent random sampling design.

In addition, the multiple metrics identified in the standard (30 day mean, 7 day minimum, absolute minimum, etc.) confound the binomial test. Simultaneous application of multiple "criteria" make application of the binomial test difficult. These multi day metrics also incorporate a statistical base inherently, making application of the binomial test based on discrete distributions inappropriate.

The EPA guidance document specifically calls out Dissolved Oxygen as problematic for the binomial test approach, and recommends that "when seasonality is observed in the data, a statistician experienced in the analyses of seasonal time series should be consulted to determine if an alternative approach or model should be employed."[2]
Additionally, the Integrated Report Assessment methodology[3] does not apply the binomial test to listing "where a statistical threshold or proportion used to determine attainment is already clearly expressed in Oregon's water quality standards"[4] such as continuously monitored

---

[1] Integrated Report Improvements- Delisting Methodology White Paper. 5/11/2018.
https://www.oregon.gov/deq/FilterDocs/iri-delisting.pdf
[2] Consolidated Assessment and Listing Methodology- Toward a Compendium of Best Practices. U.S. Environmental Protection Agency Office of Wetlands, Oceans, and Watersheds. July 2002. Pg. 62.
https://www.epa.gov/sites/production/files/2015-09/documents/consolidated_assessment_and_listing_methodology_calm.pdf
[3] Methodology for Oregon's 2018 Water Quality Report and List of Water Quality Limited Waters, Oregon DEQ, 4/2020. https://www.oregon.gov/deq/wq/Documents/irMethodologyF1820.pdf
[4] Integrated Reporting Improvements- Statistical Methods for Listing and Assessment of Large and Long Term Data Sets White Paper, Oregon DEQ, 5/11/2018. https://www.oregon.gov/deq/FilterDocs/ir-statmethods.pdf

dissolved oxygen. The dissolved oxygen assessment does not use the binomial test for listing purposes.

Delisting considerations

The requirements to delist parameters from the 303(d) list should be higher than the requirements to add to the list.[5] This ensures that a waterbody is not delisted prematurely or without solid supporting evidence. A delisting represents that a previously identified issue is no longer applicable for a given waterbody. As waterbody protections are removed (permit limitations, waste load allocations, TMDL developments, etc.) the agency should ensure that the delisting is appropriate. In addition, delisting an actually impaired assessment unit may result in an inconsistent back and forth effect, where the assessment unit moves in and out of the 303(d) list depending on when sampling occurred. This scenario is problematic for regulators and the regulated community as it offers no assurance from one cycle to the next.

Dissolved oxygen exhibits seasonal and diurnal patterns that influence concentrations. To determine attainment status of the waterbody, DEQ needs to be confident that the dataset adequately captures the periods of criteria excursions.
There are three aspects of a minimum data requirement that would need to be developed to minimize the risk of delisting a waterbody that is still impaired:

- Sampling frequency
- Time of year, i.e. critical period
- Number of years

Sampling Frequency

Overview of Current Assessment Methodology

The Dissolved Oxygen Standard provides for two different assessment methodologies, depending on what type of data is available[6]. See Appendix B: Dissolved Oxygen standard for the standard's language.
Using the Cold Water criteria as an example, dissolved oxygen may not be below 8.0 mg/L as an absolute minimum. However, "at the discretion of the Department, when the Department determines that adequate information exists, the dissolved oxygen may not fall below 8.0 mg/L as a 30-day mean minimum [**30-D –** 30 day average of daily mean], 6.5 mg/L as a seven-day minimum mean [**7-Mi –** 7 day average of daily minimum], and may not fall below 6.0 mg/L as an absolute minimum."[6]

By providing a more complete picture of instream dissolved oxygen conditions, instances of dissolved oxygen lower than 8.0 mg/L may be allowed, down to an absolute minimum of 6.0 mg/L, and still be protective of fish and aquatic life. Note that in the Assessment Methodology, in order for an assessment unit to meet category 2 (attaining) conditions, the assessment unit has to meet all three of the 30-D, 7-Mi, and alternative absolute minimum criteria simultaneously. The Dissolved Oxygen Implementation Procedures defines the requirements for sufficient information needed to evaluate using the continuous data metrics.

---

[5] Listing and Delisting Procedures: Binomial and Large Data Set Assessment Options, Slide 51, 12/06/2017, https://www.oregon.gov/deq/FilterDocs/irimp-listingdelisting.pdf
[6] OAR 340-041-0016

To evaluate compliance with a year-round criteria using continuous data, DEQ considers July 1 – September 30 to be the critical period for monitoring attainment. Enough continuous data to calculate at least 45 consecutive daily mean values must be available within the critical period. This ensures that multiple 30-day rolling averages can be calculated. At a minimum, this requirement provides at least 15, 30-day rolling averages and 39 rolling 7-day averages. This covers at least 50% of the days within the critical period.
To evaluate compliance with the spawning criteria using continuous data, daily means must be available for 45 or more days within the critical period defined by the applicable spawning date range for the waterbody. This ensures that multiple 7-day rolling averages can be calculated.[7]

The flow charts shown in Figure 11 and Figure 12 in Appendix C: Assessment Methodology describe how this is implemented in the Integrated Report Assessment.

Are grab samples sufficient to identify delistings?

To determine if grab samples are sufficient to identify waters in which previously identified problems are no longer issues, DEQ developed a two-part question:

1. Are grab samples effective in capturing diel variations in DO?
    a. Are a majority of criteria excursions during a time-period in which grab samples are least likely to occur (earlier than 5am or so)?
2. Are relatively infrequent grab samples sufficient to say, with confidence, a previously identified impairment is no longer affecting beneficial uses?
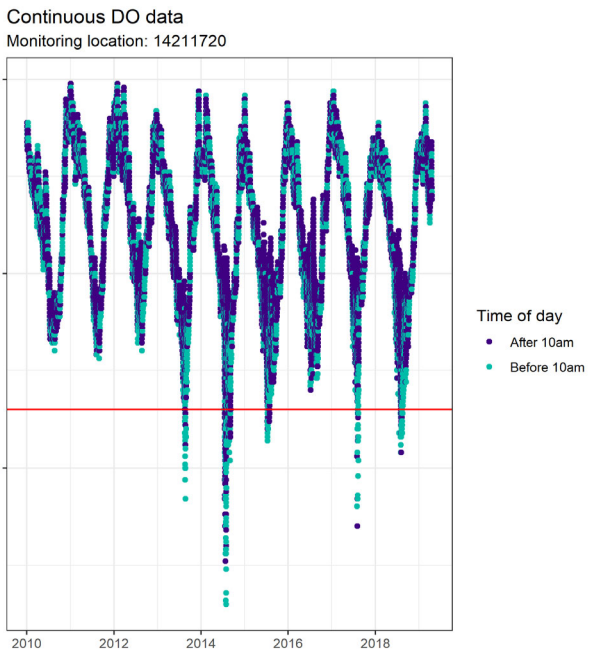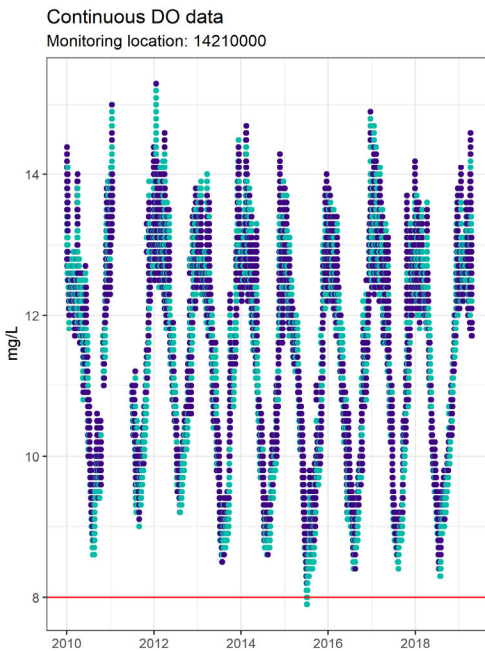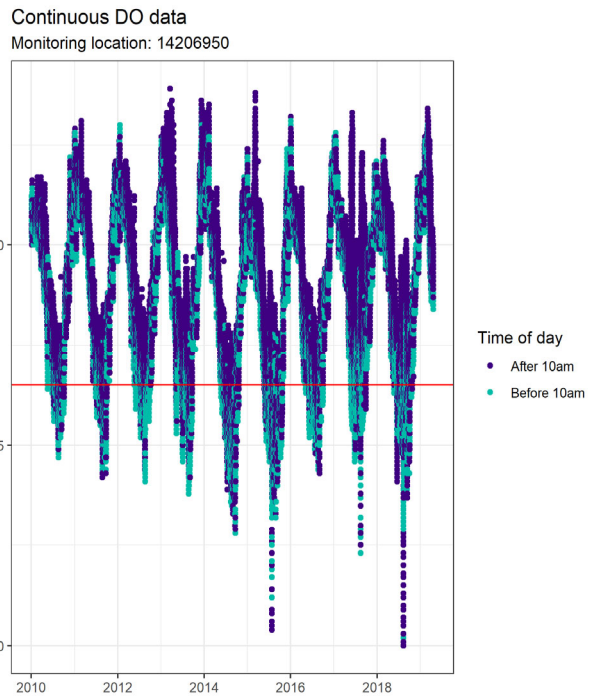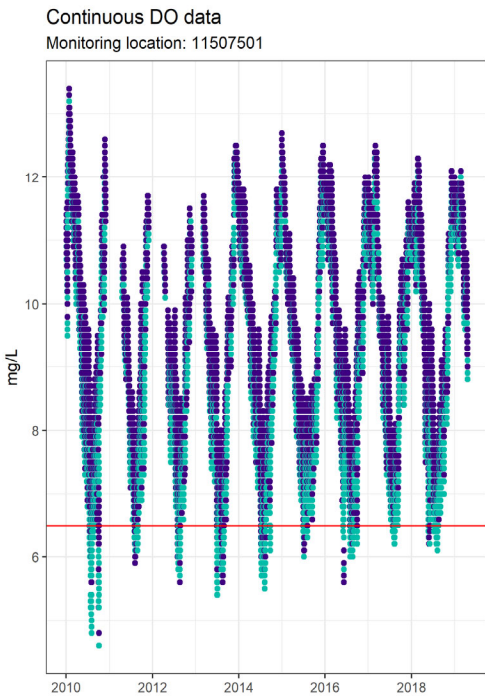
Figure 1. Selected graphs of USGS continuous Dissolved Oxygen Data

The dataset used in this analysis is USGS continuous DO results from 2010-01-01 through 2019-04-21 from 60 monitoring locations. DEQ used USGS data for the ease of obtaining both raw data from the USGS National Water Information System (NWIS) and summary statistics from Oregon's Ambient Water Quality Monitoring System (AWQMS). In an effort to save time, additional quality checks were not performed on the raw data, and USGS quality control codes were not factored into the analysis; all data was used as if it was validated. The analysis was performed for comparative purposes, and not actual assessments. Data that had not been

checked for quality assurance were sufficient, and the analysis was completed for year round criteria.

1. Time of day analysis

For each USGS monitoring location, the graph shows all available data. The data was color coded depending on time of day. Samples between midnight and 10 a.m. are colored light blue, and samples between 10 a.m. and midnight are dark purple. Figure 1 shows a selection of this data.

At a glance, the graphs for these locations appear to show that excursions typically happen over multiple days and there are few short intraday low DO events. There are a handful of instances where a monitoring location only had criteria excursions in the midnight to10 a.m. timeframe. This indicates that grab samples may be sufficient to capture excursions, if the sampling is frequent enough to capture low DO events.

2. Does grab samples accurately describe waterbody Dissolved Oxygen conditions?

The next step was to ascertain whether grab samples accurately describe waterbody conditions. If a dataset consisting of purely grab samples indicate a category 2 condition, would that assessment agree with an assessment conclusion based on continuous data? This analysis assumes that an assessment conclusion using the continuous data metrics (30-D, 7-Mi, and alternative absolute minimum) is the 'true' condition of the waterbody. Continuous datasets capture results at both high frequency (~15 minute intervals) and for 24 hours a day over the span of the deployment. These datasets comprise the most complete picture of DO available. The first step was to assess the USGS monitoring sites using the 2018/2020 year round continuous methodology on a site-by-site basis. Following the 2018 Integrated Report Assessment methodology, the monitoring location was assigned an impaired designation if any of the following were true:

1. ≥ 2 30-D (30 day average of daily mean) below the 30-D criteria
2. ≥ 2 7-Mi (7 day average of daily minimum) below the 7-Mi criteria
3. ≥ 2 daily minimum values are below alternative absolute minimum criteria

This gave us the 'true' condition of the waterbody. The data were analyzed on a site-by-site basis, as opposed to an assessment unit, in order to account for potential differences within an assessment unit. This analysis identified 46 USGS locations that have a 'true' impaired condition.

To identify if a grab sampling program would accurately describe the 'true' condition of the waterbody, The datasets were narrowed to an extended critical period of June 1 – September 30, and to working hours (08:00 – 17:00). A sampling program was simulated by randomly sampling 1 result per month for the 4 month critical period, for each year available in the dataset (mean number of years available = 6). This produced a simulated grab sample dataset of n = 4 * (number of available years) per monitoring location. The simulated dataset was then processed through the instantaneous DO methodology. The monitoring location was assigned an impaired designation if >10% of results (using the binomial) were below the absolute minimum value (the shaded cells in Table 6. OAR 340-041-0002 Table 21). The impairment status identified by the simulated grab sampling program was compared to the results of the continuous assessment methodology. A successful assessment was determined if the two different methodologies agreed. This simulated process was repeated 5,000 times at each monitoring location, and the percent of simulated runs that produced the "correct" result was calculated.

To tease out a time of day component to this, the simulations were repeated with both morning only samples (08:00-12:00) and afternoon only samples (13:00-17:00). Results of the simulations are in Table 1.

Sampling program	Metric	Percent of monitoring locations that meet metric
Critical Period Monthly- 08:00-17:00	>= 90% of simulations produce the correct result	57%
Critical Period Monthly- 08:00-17:00	>= 80% ...	57%
Critical Period Monthly- 08:00-17:00	> 50% ...	61%
Critical Period Monthly- 08:00-17:00	< 50% ...	39%
Critical Period Monthly- 08:00-17:00	0% ...	26%
Critical Period Monthly Morning, 08:00 - 12:00	>= 90% of simulations produce the correct result	57%
Critical Period Monthly Morning, 08:00 - 12:00	>= 80% ...	57%
Critical Period Monthly Morning, 08:00 - 12:00	> 50% ...	61%
Critical Period Monthly Morning, 08:00 - 12:00	< 50% ...	39%
Critical Period Monthly Morning, 08:00 - 12:00	0% ...	26%
Critical Period Monthly Afternoon, 12:00-17:00	>= 90% of simulations produce the correct result	48%
Critical Period Monthly Afternoon, 12:00-17:00	>= 80% ...	50%
Critical Period Monthly Afternoon, 12:00-17:00	> 50% ...	59%
Critical Period Monthly Afternoon, 12:00-17:00	< 50% ...	41%
Critical Period Monthly Afternoon, 12:00-17:00	0% ...	33%
Critical Period Weekly- 08:00-17:00	>= 90% of simulations produce the correct result	63%
Critical Period Weekly- 08:00-17:00	>= 80% ...	63%
Critical Period Weekly- 08:00-17:00	> 50% ...	67%
Critical Period Weekly- 08:00-17:00	< 50% ...	33%
Critical Period Weekly- 08:00-17:00	0% ...	22%

Table 1. Grab sample simulation results

Table 2 shows the average accuracy of each simulated sampling program's ability to detect impairments.

Simulated Sampling Program	Average Accuracy of 5000 Simulations
08:00 - 17:00 critical period monthly grab sampling	60%
08:00 - 12:00 critical period monthly grab sampling	62%
03:00 - 17:00 critical period monthly grab sampling	56%
08:00 - 17:00 critical period weekly grab sampling	67%

Table 2. Average simulated grab sampling program accuracy

For all three monthly sampling scenarios (mixed day, morning sampling, afternoon sampling) approximately 40% of the locations received the "correct" conclusion on less than 50% of the simulated sampling runs. The range of the locations receiving the "correct" conclusion on greater than 90% of the simulations ranges from 48-57%. This indicates monthly grab sampling cannot reliably detect impairments in approximately 40-50% of these USGS sampling locations.

Moving the sampling frequency to weekly during the critical period slightly improves the ability to correctly identify impairments. More than 90% of simulations derived the correct impaired designations at 63% of the monitoring locations.

These numbers do not inspire the confidence that a monthly grab sampling program is of sufficient frequency to accurately describe if a previously identified DO issue has been resolved.

Short-term continuous probe deployment

DEQ staff expressed concern over the requirement for continuous data collection and the costs incurred for deployment of continuous monitoring probes. Staff noted that by mandating long-term continuous data to delist, there would need to be a change in what data DEQ recommends Designated Management Agencies collect, and what data are requested in watershed management plans.

To address this concern, an exploration of shorter duration continuous probe deployments (i.e. 3-day, 4-day, 5-day) showed it would be sufficient to accurately categorize the waterbody condition. Short-term deployments may be easier datasets for monitoring partners to collect. In addition, short-term deployments increase the number of locations that can be monitored in a year if given the same amount of hardware.

To test the efficacy of these types of deployments, the simulations were repeated with short-term deployments. Again, the deployments were narrowed to the extended critical period (June 1 through September 30) for the entire available dataset for each monitoring location (mean number of years available = 6). A monthly short-term probe deployment was simulated by picking a random day for each of the four months, and then calculated n number of consecutive

daily minimums where n = number of complete days of each deployment. This dataset was processed through the Integrated Report instantaneous DO methodology. Following the methodology and DO Implementation guidance, the daily minimum results were compared to instantaneous criteria (the shaded values in Table 6. OAR 340-041-0002 Table 21).

Table 3 shows the results of this simulation. Note that a 2-day deployment refers to 2-full days. For example, 2-day deployment could be installed on Monday and retrieved on Thursday.

Sampling program	Metric	Percent of monitoring locations that meet metric
2 day deployment monthly	>= 90% of simulations produce the correct result	74%
2 day deployment monthly	>= 80% ...	76%
2 day deployment monthly	> 50% ...	85%
2 day deployment monthly	< 50% ...	15%
2 day deployment monthly	0% ...	2%
3 day deployment monthly	>= 90% of simulations produce the correct result	76%
3 day deployment monthly	>= 80% ...	78%
3 day deployment monthly	> 50% ...	89%
3 day deployment monthly	< 50% ...	11%
3 day deployment monthly	0% ...	0%
4 day deployment monthly	>= 90% of simulations produce the correct result	78%
4 day deployment monthly	>= 80% ...	83%
4 day deployment monthly	> 50% ...	89%
4 day deployment monthly	< 50% ...	11%
4 day deployment monthly	0% ...	0%
5 day deployment monthly	>= 90% of simulations produce the correct result	80%
5 day deployment monthly	>= 80% ...	85%
5 day deployment monthly	> 50% ...	89%
5 day deployment monthly	< 50% ...	11%
5 day deployment monthly	0% ...	0%

Table 3. Short term continuous probe deployment simulation results

Table 4 shows the average accuracy of each simulated short-term probe deployment's ability to detect impairments.

Simulated Sampling Program	Average Accuracy of 5000 Simulations
2 full day continuous probe deployment per critical period month	84%
3 full day continuous probe deployment per critical period month	86%
4 full day continuous probe deployment per critical period month	87%
5 full day continuous probe deployment per critical period month	89%

Table 4. Average simulated short-term continuous probe sampling program accuracy

The four and five complete day deployments both saw approximately 80% of locations receiving the "correct" conclusion on greater than 80% of the simulations. All short-term deployment scenarios perform significantly better than pure grab sample programs. The binomial process used for delisting other parameters assumes a 90% confidence level. Five full-day continuous probe deployments approach this confidence level.

Location of sites used in simulations
Unfortunately, spatial heterogeneity in these impaired USGS sites was limited. Generally, the sites are grouped near either Klamath Falls or the northern Willamette Valley (Figure 2). This may skew the simulation results, as the locations likely do not represent the full range of dissolved oxygen conditions statewide.

Figure 2. Location of USGS sites identified as Impaired by continuous data analysis and used in simulated model runs
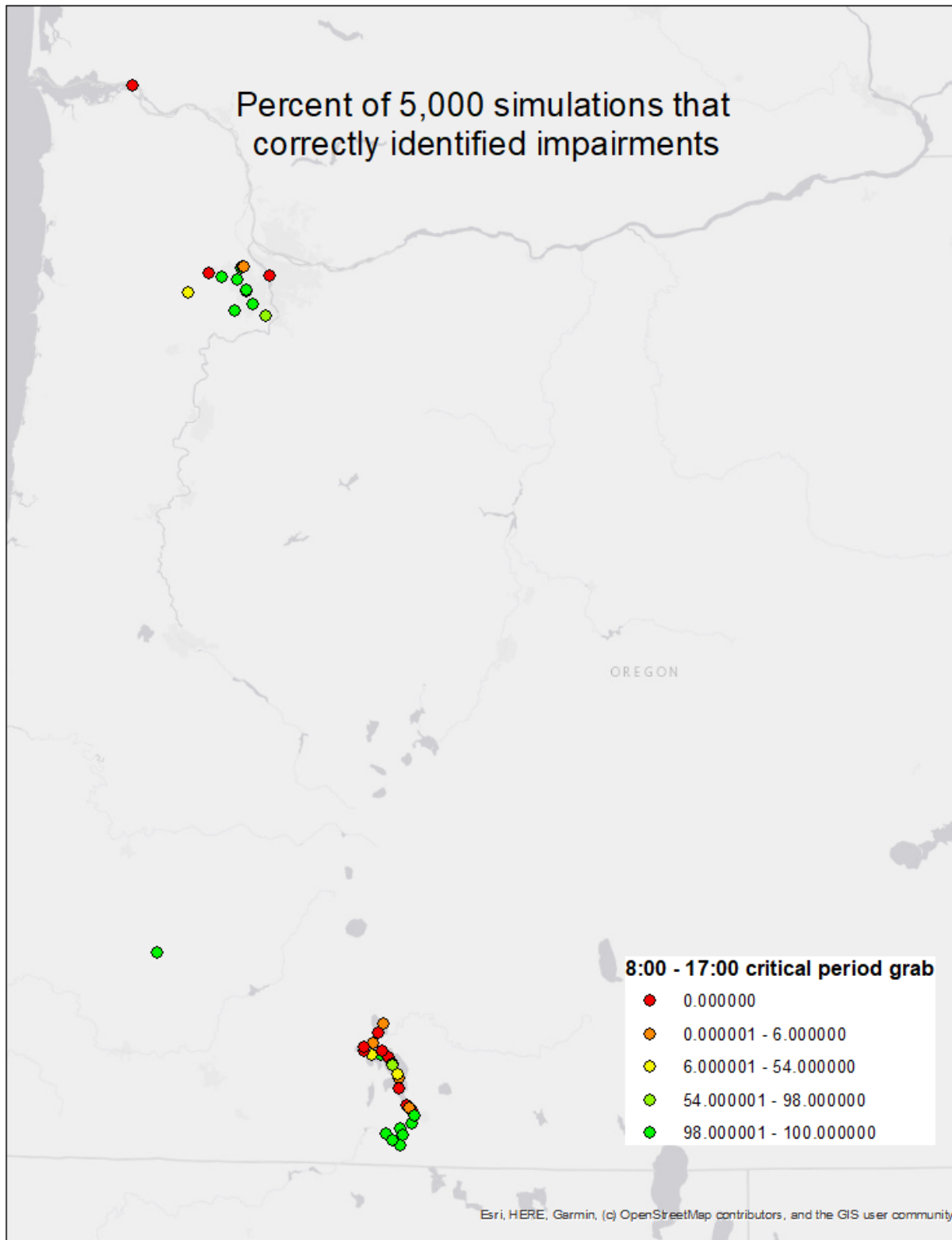
Figure 2. 8:00-17:00 critical period grab sample ability to detect impairment

Percent of 5,000 simulations that correctly identified impairments

5 full day deployment
- 1.000000 - 8.000000
- 8.000001 - 44.000000
- 44.000001 - 83.000000
- 83.000001 - 94.000000
- 94.000001 - 100.000000

Esri, HERE, Garmin, (c) OpenStreetMap contributors, and the GIS user community
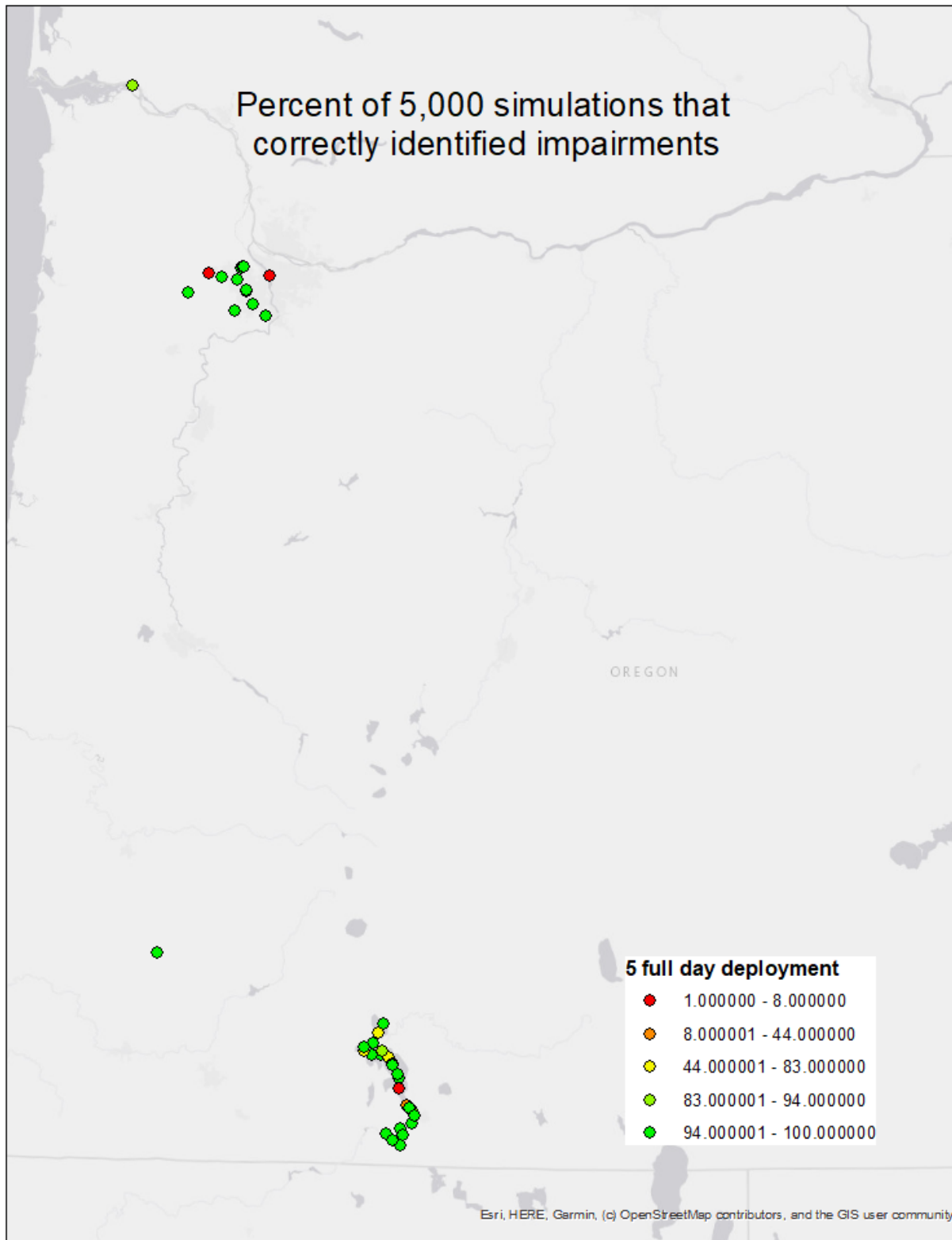
Figure 3. Monthly critical period 5 full-day continuous probe deployment ability to detect impairment

A grab sampling program performed significantly worse in Northern Upper Klamath Lake than it did in the Northern Willamette Valley. The cause of these differences was not analyzed.

Conclusion

Due to the inability of monthly sampling programs to identify impairment conditions with confidence, DEQ recommends a continuous data requirement to delist waterbodies for dissolved oxygen. The error rates of the short-term deployments are significantly lower and may provide a path to delisting without requiring full critical period continuous datasets. By utilizing monthly 5-day continuous deployments throughout the critical period, DEQ approaches the 90% confidence level needed to delist most other parameters. However, there may be a need to include a grab sampling delisting methodology for a limited time to allow monitoring partners time to adjust monitoring programs.

Time of Year - Critical Period

Year-round assessment of Oregon's Dissolved Oxygen Water Quality Standard: Interpretation and Application Procedures identifies July 1 – Sept. 30 as the "Critical Period" for dissolved oxygen[7]. To assign a waterbody an "attaining" (category 2) or an "impaired (category 5) status, an assessment unit's dataset needs to contain a minimum amount of data within this critical period. The minimum critical period requirements are:

- 15 "30 day averages" (30-D) for continuous data methodology.
- 5 samples (either grab, or daily minimum) for instantaneous data methodology.

These data minimums were developed based on the implementation guide to ensure the assessment captures periods of excursions. Figure 4 shows the day of year of all the freshwater dissolved oxygen year-round excursions. The blue critical period bars represent the beginning and end of the critical period. The area between the red "90th Percentile" bars represents the time-period that contains 90% of all dissolved oxygen excursions (June 20 – October 28). To account for the considerable additional data collected during the summer months, Figure 5 shows the percentage of results that are excursions. This analysis shows that to reliably detect impairment, the dataset needs to include a reasonably complete picture of the critical period to ensure excursions are not being missed. In addition, to assesses attainment of criteria, the critical period likely needs to either be expanded or shifted several weeks later, but that is an action outside of the scope of this project and will be considered when there are updates to the DO listing methodology.

---

[7] Oregon's Dissolved Oxygen Water Quality Standard: Interpretation and Application Procedures, Oregon DEQ, 9/2019.

Dissolved Oxygen Criteria Excursions by Day of Year

Date of year-round criteria excursion for all Monitoring Locations in 2018/2020 data window
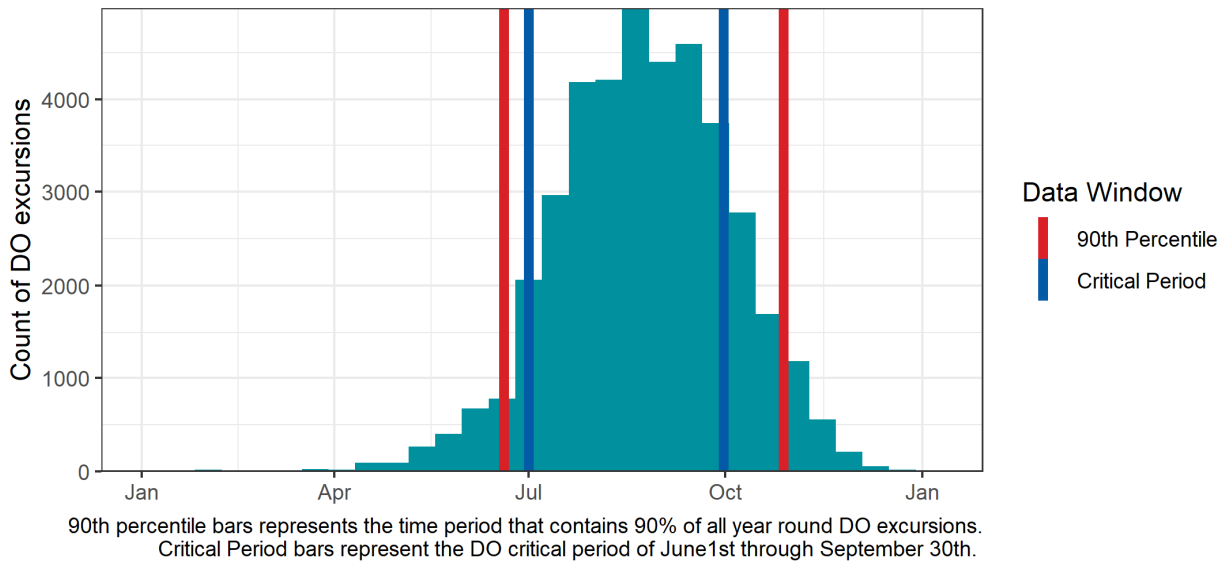
90th percentile bars represents the time period that contains 90% of all year round DO excursions.
Critical Period bars represent the DO critical period of June1st through September 30th.

Figure 4. Dissolved Oxygen Excursions by Day of Year

Dissolved Oxygen Criteria Excursions by Day of Year

Percent of samples below criteria for all Monitoring Locations in 2018/2020 data window
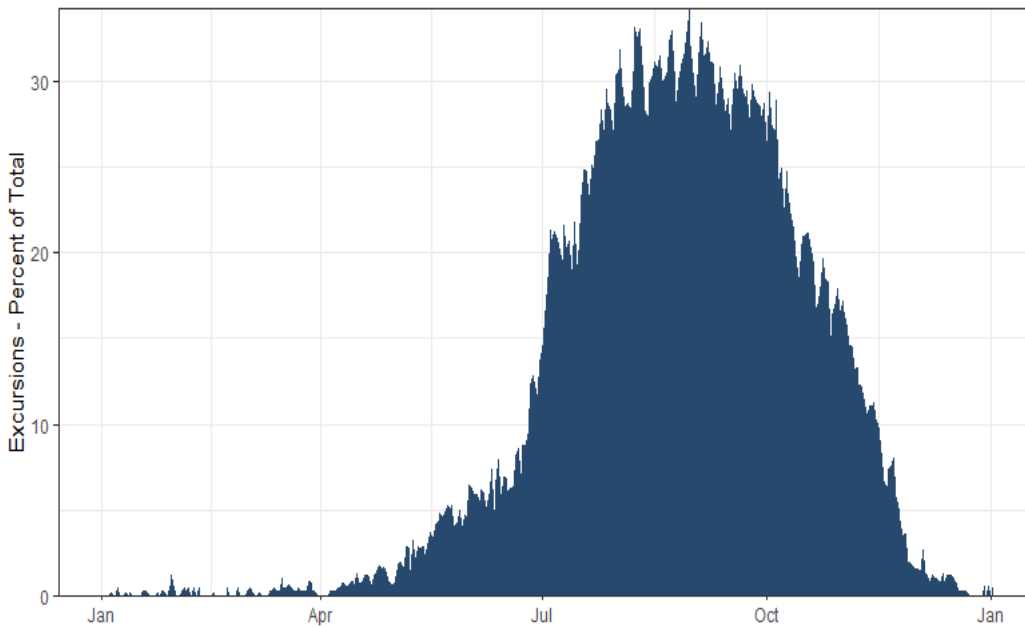
Figure 5. Percent of samples below DO criteria

Spawning Season Critical Period

When assessing against the spawning criteria, a different critical period needs to be identified. Figure 6 shows the Dissolved Oxygen data that was assessed against the spawning criteria in the 2018/2020 Integrated Report. As Oregon has multiple spawn periods throughout the state, and these spawning periods have different time scales, Figure 7 shows the number of spawning period samples and excursions as a percent of the spawning season. The very first day of the

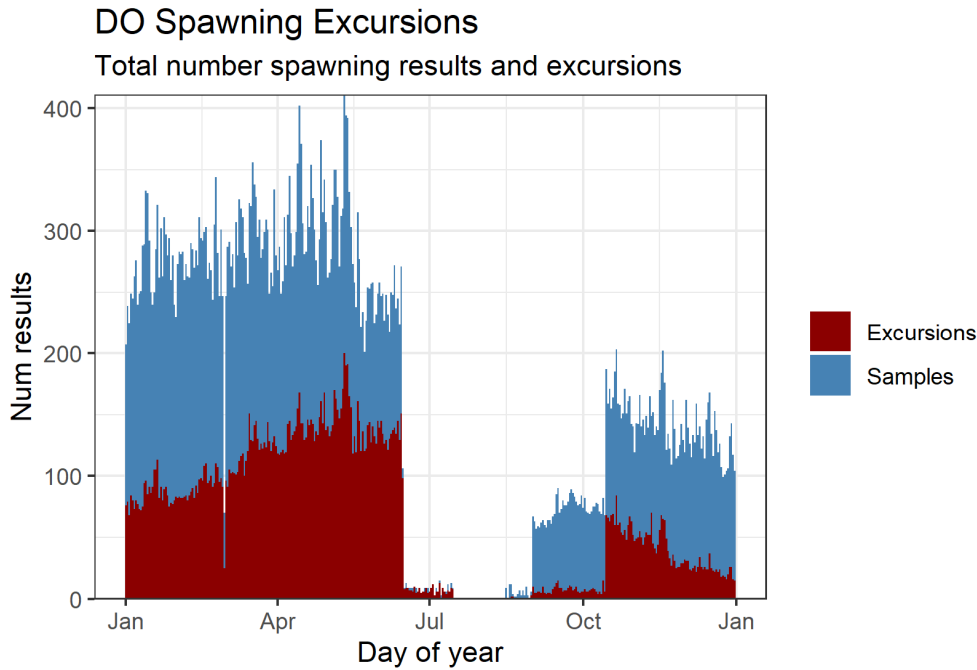spawning season is represented as 0% and the very last day of the spawning season is represented as 100%.
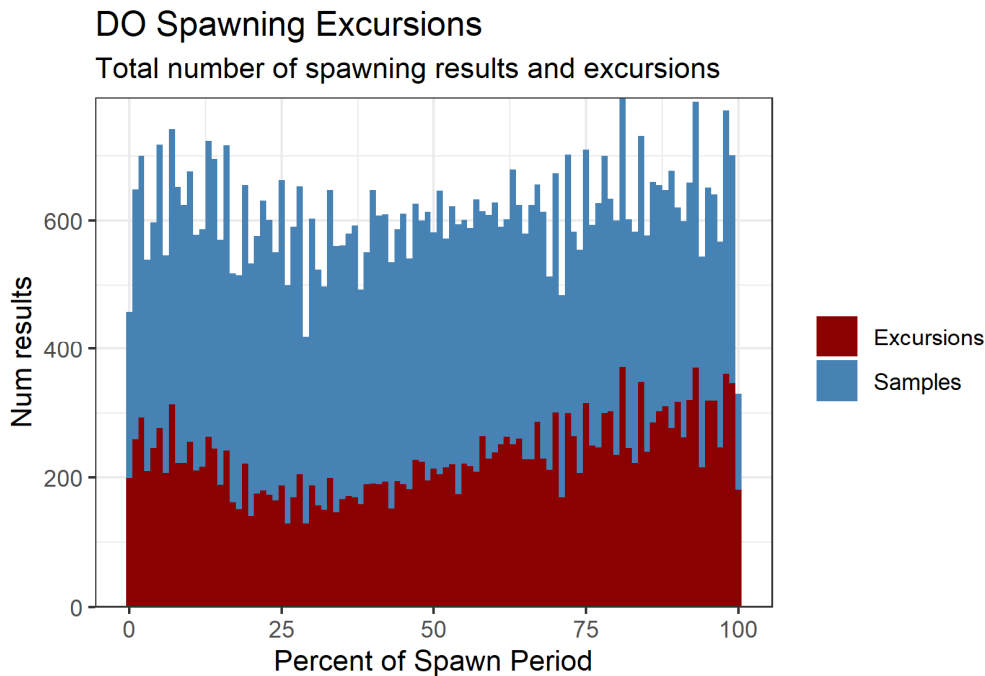

Figure 6. Spawning data and excursions

Figure 7. Spawning data and excursions by percent spawning period
While there does seem to be an increase in the percent of samples that are an excursion of the spawning criteria (11 mg/L) in the summer months, winter month samples also frequently show concentrations below criteria.

There is some variation in the pattern for different spawning periods, but the overall pattern indicates that the spawning critical period should spawn the entire spawning period. This would make an interesting topic for DEQ staff and stakeholder discussions, as there may be difficulty to deploy continuous probes during winter months.

Number of Years

Climatic variables heavily factor into the probability of a waterbody to experience dissolved oxygen criteria excursions. Waterbodies are more likely to experience periods of low dissolved oxygen during hot, dry, and low flow conditions. Conversely, during cool, wet, and high flow periods, waterbodies are less likely to experience low dissolved oxygen conditions. Datasets must include data from the range of these expected conditions to accurately identify dissolved oxygen attainment. Figure 8 illustrates the length of data at monitoring locations used in the 2018/2020 Integrated Report. Nearly 70% of all monitoring locations had 3 years of data or less. Over 30% of all locations had 3 or more years of data.
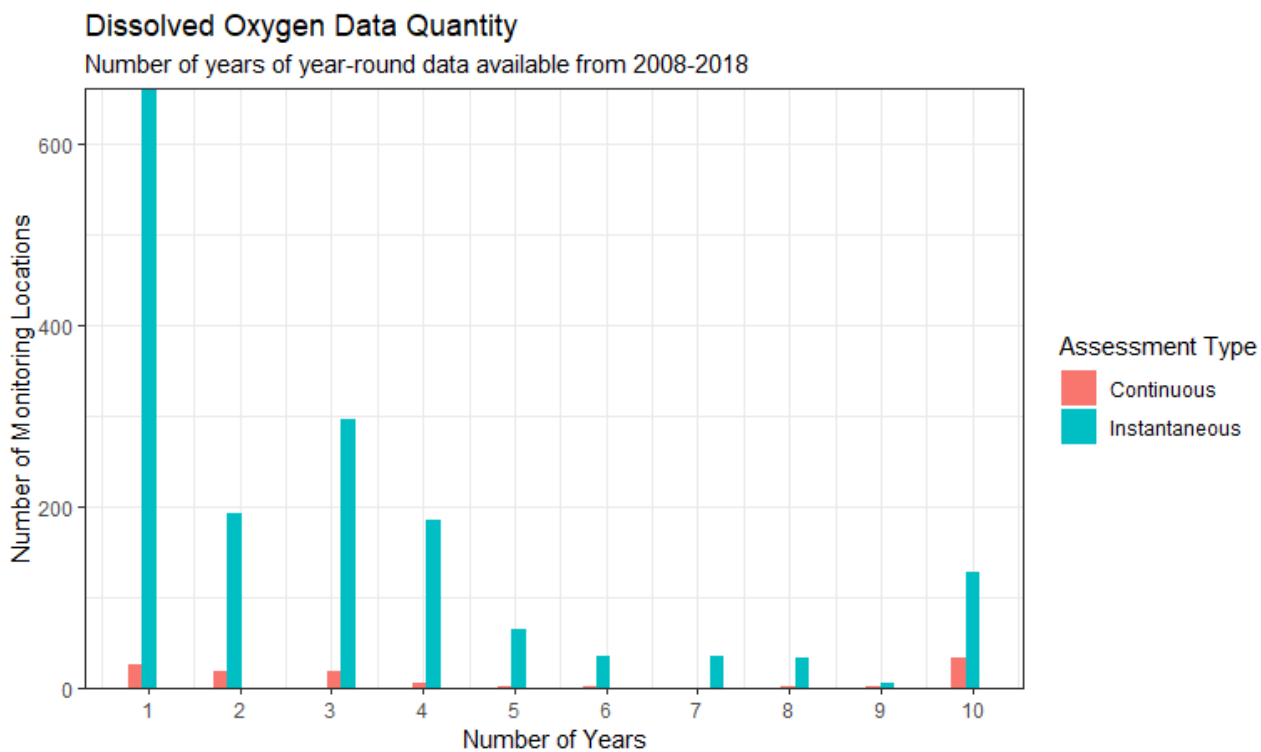

Figure 8. Length of Dissolved Oxygen datasets

To test the effects of differing number of years on short-term deployments to detect impairments, DEQ repeated the monthly critical period 5 day continuous probe simulations described in Table 3 and Table 4.

For this analysis, data were limited to USGS locations with 6 or more years of data. Each simulation limited the dataset to bins of 1-5 years. The years were selected at random from the full time period available at each dataset. DEQ narrowed the dataset to the extended critical period (June 1 through September 30) for the reduced dataset for each monitoring location.

Simulations of a monthly short-term probe deployment were done by picking a random day for each of the four months, and then calculated 5 consecutive daily minimums. This dataset was then processed through the Integrated Report instantaneous DO methodology. Following the methodology and DO Implementation guidance, daily minimum results were compared to

instantaneous criteria (the shaded values OAR 340-041-0002 Table 21). Then, each simulation was repeated 5,000 times and the average ability to detect impairments was calculated.

Number of years	Average Accuracy over 5,000 simulations
1	76 %
2	85 %
3	87 %
4	88 %
5	90 %

Table 5.Average accuracy of 5-day deployment simulations on year-limited datasets

Considerations for climatic variability should be taken into account when selecting a minimum data requirement for delisting. Using a shorter time period raises the risk of improperly excluding low DO conditions that may have been present when the assessment unit was initially listed as impaired.

Two options for setting minimum data requirements have been discussed.

1. Setting a minimum number of years.
2. Consider measured climate variables to ensure data period represents a range of climate conditions. (e.g. compare measured years to 30 year normal)

Setting a minimum number of years can be easier to assess, and gives the agency the ability to communicate clear data expectations to partners. However, there can be no assurance that the data is collected only during a dry-cycle instead of a wet-cycle, where we expect DO concentrations to be relatively high. Connecting minimum data requirements to measured climate variables ensures a full range of climatic conditions are represented, but increases assessment complexity, and makes long-term data collection effort planning difficult. Setting minimum years of data collection is another area where conversations can be initiated to determine acceptable error rates.

Proposed Dissolved Oxygen Delisting Methodology

Initial concept of DO Delisting Option would be:

A. Full critical period (or spawning option)
    1. Continuous metrics analysis results in a category 2 designation of attaining criteria *and*
    2. Dataset must include minimum 3 years of data that have 80% of the critical period (July 1– September 30) in each year represented.
<div align="center">Or</div>
B. Short term probe deployments
    1. Dataset includes minimum 3 years of data that contains at least 5 full days of continuous dissolved oxygen per critical period month per year *and*
    2. <10% (binomial) of daily minimums are below criteria as described in the Delisting Waterbodies Section of the Integrated Report Assessment methodology.

In addition, for the 2022 Integrated Report cycle only, DEQ is proposing a third method for delisting using grab samples. This method is included in the 2022 and 2024 cycles to allow monitoring partners to adjust to continuous sampling methodologies.

C. Grab samples
   a. Dataset includes 3 years of data that contain at least 2 results for each critical period month.
   b. There are no excursions of any applicable criteria

If conditions described in A, B, or C are met, then the assessment unit may be delisted. For year round analysis, Option A would require a continuous dissolved oxygen Probe to be deployed from July 1 through September 30 for three years in the IR data window. Option B would require continuous Dissolved Oxygen Probes to be deployed for 5 full days in July, August, and September for 3 years in the IR data window. To get 5 full days of data, the probe installation would likely be 7 day deployments.

For delisting of spawning listings, the critical period will make up the spawning period. Internal DEQ discussions identified a desire to develop a delisting mechanism that included an alternative to long-term full summer probe deployments. Given the modeled grab sampling was unable to reliably detect impairments, 5 full-day deployments per month during the critical period was added. Modeled deployments show that this can approach a 90% success rate that limits improperly delisting impaired waterbodies.

Recognizing that 3 years of 80% critical period coverage with a continuous probe is a challenge to meet, DEQ analyzed how many monitoring locations assessed in the 2018/2020 Integrated Report would meet the data requirements. Figure 9 and Figure 10 show how many locations meet various data quantity metrics give the 3-year and 80% metric. In the 2018/2020 dataset, 50 monitoring locations meet this metric. There is a low level of sensitivity to these numbers in the dataset. Setting the critical period coverage from 80% down to 60% made only 6 additional sites available for delisting analysis. Keeping the 80%, but reducing the required years to 2 made 62 locations available for delisting analysis. DEQ developed a tool to visualize these changes. The tool can be found online: https://travispritchard.shinyapps.io/DO_delisting_app/

Analysis was not performed for how many monitoring locations would be available for delisting under the 5 full day monthly probe deployment during the critical period scenario; however, it is assumed that more monitoring groups would start collecting continuous data in waterbodies identified as impaired.

Figure 9. Data quantity by monitoring location as compared to a continuous data requirement of 3 years of 80% critical period
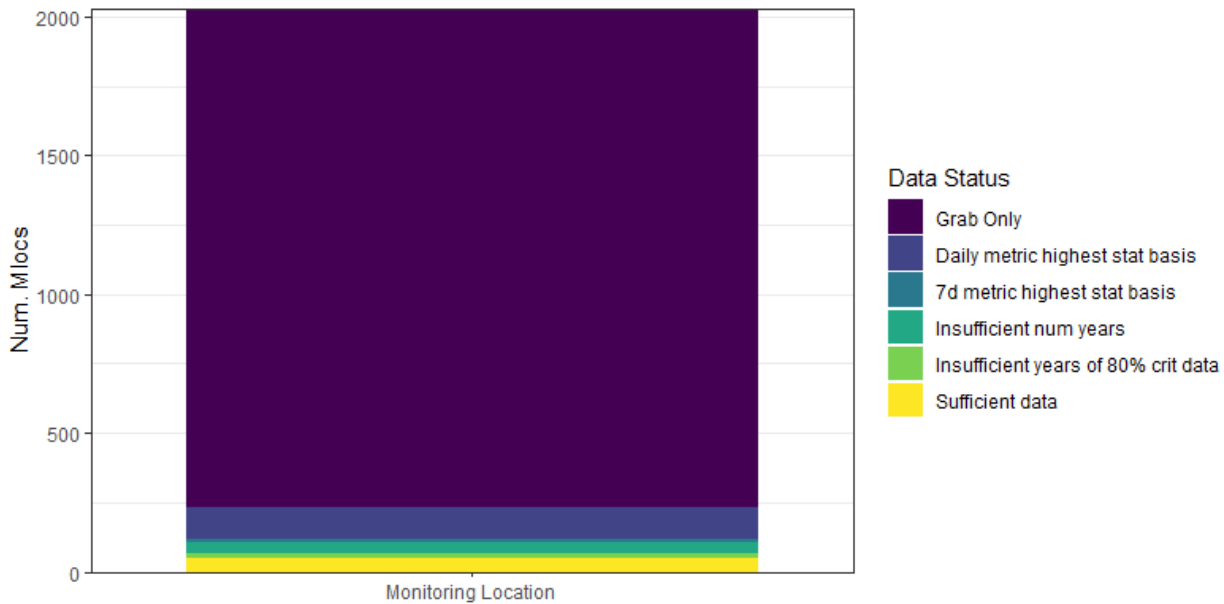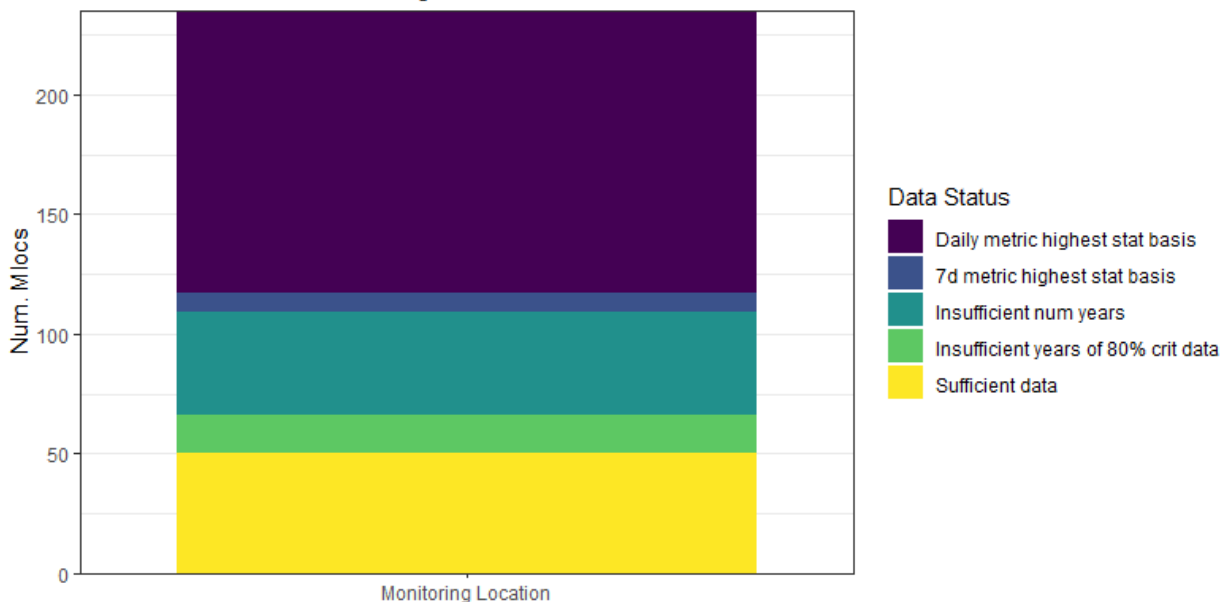

Figure 10. Similar to Figure 9. Data quantity by monitoring location as compared to a continuous data requirement of 3 years of 80% critical period, but without grab samples

This delisting methodology will give DEQ assurance that listed waters will not be improperly removed from the impaired waters list. However, monitoring plans will have to be modified if removing dissolved oxygen impairments is a goal of a water quality, monitoring program.

Conclusion

The analysis strongly suggests grab samples do not provide sufficient confidence to assert that a previously identified impairment has been resolved. The proposed DO delisting methodology ensures that assessment units that are removed from the 303(d) list for dissolved oxygen are truly protective of aquatic life.

This proposal includes a temporary methodology for delisting based on grab samples. This methodology, while not as robust as continuous monitoring, provides a level of assurance while allowing data gatherers time to adjust monitoring programs. To continue exploring this proposal,

Stakeholder engagement
DEQ is interested in input from stakeholders about the following:

1. What should the minimum number of years be for critical period sampling?
    a. This DEQ proposal has 3 years; future Integrated Reports will likely have 4-5 year data windows.
        i. Is 3 years enough? Too much?
        ii. Should the number of years be a percentage of data window?
        iii. Should the number of years be tied to climate variability and climate triggers?
2. What is the critical period for spawning listings?
    a. The entire spawning period? Shoulders? A percentage?
3. Given the low level of confidence of grab samples to accurately identify non-impaired conditions, should the temporary grab method be included?
    a. What is the appropriate balance between protectiveness of Oregon's waterways and ability to use the data we currently have?
        i. Not including grab samples may inappropriately keep some assessment units on the 303(d) list. Is this tradeoff appropriate?
4. What should the minimum number of critical period grab samples be to include in a delisting procedure?

Appendix A: Critical Values for Delisting Conventional Pollutants

Null Hypothesis: Actual exceedance proportion is >10%
Alternate hypothesis: Actual exceedance proportion is ≤10%
Minimum confidence level is 90%
A minimum sample size of 15 is required.

Maximum number of sample excursions to delist as impaired for conventional pollutants

Sample Size	Delist if excursions ≤
15	1
16-18	2
19-25	3
26-32	4
33-40	5
41-47	6
48-55	7
56-63	8
64-71	9
72-79	10
80-88	11
89-96	12
97-104	13
105-113	14
114-121	15
122-130	16
131-138	17
139-147	18
148-156	19
157-164	20
165-173	21
174-182	22
183-191	23
192-199	24
≥200	See generalized delisting method for formula to calculate the number of excursions

Generalized Binomial Delisting Procedure

For sample sizes greater than 200, calculate α from the left tail probability of the cumulative binomial distribution:

α = 1 – Excel® Function BINOMDIST (k_a-1, n, p_1, TRUE)

Where n = the number of samples,

K_a = maximum number of measured exceedances to determine a waterbody is attaining, and should be removed from the 303(d) list, and

p_1 = unacceptable exceedance proportion.

BINOMDIST( ) is an Excel software function that returns cumulative left tail binomial probabilities.

The number of excursions required to delist is the value of k_a, where the initial value of ka=1 for n=10. k_a is incrementally increased by 1, until 1-α ≤ 0.90.

Appendix B: Dissolved Oxygen standard

340-041-0016 Dissolved Oxygen

Dissolved oxygen (DO): No wastes may be discharged and no activities may be conducted that, either alone, or in combination with other wastes or activities, will cause violation of the following standards: The changes adopted by the Commission on January 11, 1996, become effective July 1, 1996. Until that time, the requirements of this rule that were in effect on January 10, 1996, apply:

(1) For water bodies identified as active spawning areas in the places and times indicated on the following Tables and Figures set out in OAR 340-041-0101 to 340-041-0340: Tables 101B, 121B, and 190B; and Figures 130B, 151B, 160B, 170B, 180A, 201A, 220B, 230B, 260A, 271B, 286B, 300B, 310B, 320B, and 340B, (as well as any active spawning area used by resident trout species), the following criteria apply during the applicable spawning through fry emergence periods set forth in the tables and figures and, where resident trout spawning occurs, during the time trout spawning through fry emergence occurs:

(a) The dissolved oxygen may not be less than 11.0 mg/l. However, if the minimum intergravel dissolved oxygen, measured as a spatial median, is 8.0 mg/l or greater, then the DO criterion is 9.0 mg/l;

(b) Where conditions of barometric pressure, altitude, and temperature preclude attainment of the 11.0 mg/l or 9.0 mg/l criteria, dissolved oxygen levels must not be less than 95 percent of saturation;

(c) The spatial median intergravel dissolved oxygen concentration must not fall below 8.0 mg/l.

(2) For water bodies identified by the Department as providing cold-water aquatic life, the dissolved oxygen may not be less than 8.0 mg/l as an absolute minimum. Where conditions of barometric pressure, altitude, and temperature preclude attainment of the 8.0 mg/l, dissolved oxygen may not be less than 90 percent of saturation. At the discretion of the Department, when the Department determines that adequate information exists, the dissolved oxygen may not fall below 8.0 mg/l as a 30-day mean minimum, 6.5 mg/l as a seven-day minimum mean, and may not fall below 6.0 mg/l as an absolute minimum (Table 21);

(3) For water bodies identified by the Department as providing cool-water aquatic life, the dissolved oxygen may not be less than 6.5 mg/l as an absolute minimum. At the discretion of the Department, when the Department determines that adequate information exists, the dissolved oxygen may not fall below 6.5 mg/l as a 30-day mean minimum, 5.0 mg/l as a seven-day minimum mean, and may not fall below 4.0 mg/l as an absolute minimum (Table 21);

(4) For water bodies identified by the Department as providing warm-water aquatic life, the dissolved oxygen may not be less than 5.5 mg/l as an absolute minimum. At the discretion of the Department, when the Department determines that adequate information exists, the dissolved oxygen may not fall below 5.5 mg/l as a 30-day mean minimum, and may not fall below 4.0 mg/l as an absolute minimum (Table 21);

(5) For estuarine water, the dissolved oxygen concentrations may not be less than 6.5 mg/l (for coastal water bodies);

OAR-340-041-0002 Table 21
Dissolved Oxygen and Intergravel Dissolved Oxygen Criteria

DO Standard	Concentration and Period[1] (All Units are mg/L)				Use/Level of Protection
	30-D	7-D	7- Mi	Min	
Salmonid Spawning		11.0[2,3]		9.0[3] / IGDO: 8.04	Principal use of salmonid spawning and incubation of embryos until emergence from the gravels. Low risk of impairment to cold-water aquatic life, other native fish and invertebrates.
Cold Water	8.0[5]		6.5	6.0	Principally cold-water aquatic life. Salmon, trout, cold-water invertebrates, and other native cold-water species exist throughout all or most of the year. Juvenile anadromous salmonids may rear throughout the year. No measurable risk level for these communities.
Cool Water	6.5		5.0	4.0	Mixed native cool-water aquatic life, such as sculpins, smelt, and lampreys. Waterbodies includes estuaries. Salmonids and other cold-water biota may be present during part or all of the year but do not form a dominant component of the community structure. No measurable risk to cool-water species, slight risk to cold-water species present.
Warm Water	5.5			4.0	Waterbodies whose aquatic life beneficial uses are characterized by introduced, or native, warm-water species.
Marine / No Risk	No Change from Background				The only DO criterion that provides no additional risks is "no change from background". Waterbodies accorded this level of protection include marine waters and waters in Wilderness areas.

Table 6. OAR 340-041-0002 Table 21

340-041-0006 Definitions

(15) "Daily Mean" for dissolved oxygen means the numeric average of an adequate number of data to describe the variation in dissolved oxygen concentration throughout a day, including daily maximums and minimums. For calculating the mean, concentrations in excess of 100 percent of saturation are valued at the saturation concentration.

(27) "Intergravel Dissolved Oxygen" (IGDO) means the concentration of oxygen measured in the water within the stream bed gravels. Measurements should be taken within a limited time period before emergence of fry.

(38) "Minimum" (Min) for dissolved oxygen means the minimum recorded concentration including seasonal and diurnal minimums.

(39) "Monthly (30-D) Mean Minimum" for dissolved oxygen means the minimum of the 30 consecutive-day floating averages of the calculated daily mean dissolved oxygen concentration.

(59) "Spatial Median" means the value that falls in the middle of a data set of multiple intergravel dissolved oxygen (IGDO) measurements taken within a spawning area. Half the samples should be greater than and half the samples should be less than the spatial median.

(73) "Weekly (7-D) Mean Minimum" for dissolved oxygen means the minimum of the seven consecutive-day floating average of the calculated daily mean dissolved oxygen concentration.

(74) "Weekly (7-Mi) Minimum Mean" for dissolved oxygen means the minimum of the seven consecutive-day floating average of the daily minimum concentration. For application of the criteria, this value is the reference for diurnal minimums.

(22) "Estuarine Waters" means all mixed fresh and oceanic waters in estuaries or bays from the point of oceanic water intrusion inland to a line connecting the outermost points of the headlands or protective jetties.

(34) "Marine Waters" means all oceanic, offshore waters outside of estuaries or bays and within the territorial limits of the State of Oregon.

Appendix C: Assessment Methodology

Determining Applicable Criteria

The application of the various dissolved oxygen criteria is based on designated fish use as described in the tables and figures in OAR-340-041-016 (1). For convenience, the interpretation of this information is detailed in the Dissolved Oxygen Standard Implementation Guidance and depicted for reference in the DEQ water quality standards mapping web tool (in development).

Time Period

Spawning Time-Period: The spawning criteria shall be applied for places and times indicated, in the tables and figures referenced in OAR-340-041-0016 (1), as having active salmon and steelhead spawning, or any additional assumed spawning by resident trout species. Listed status of waterbodies in violation of the spawning criteria is in effect only during the applicable spawning date range for the waterbody.

Year-round: The year-round dissolved oxygen criteria apply year round. For some locations, a more stringent spawning criteria may apply in addition to the year round criterion for part of the year. Listed status of waterbodies in violation of the year-round criteria are in effect year-round.

Critical Period: The critical period for assessing compliance with the year-round dissolved oxygen standard is the summer period July 1 – September 30, when seasonal trends in dissolved oxygen are expected to be near annual minimums.

Assignment of Assessment Category

Instantaneous Dissolved Oxygen Concentration

The D.O. criteria metrics are absolute minimum D.O. concentrations referenced in OAR-340-041-016 (1)(a) –(6) (Table 16). These criteria are also depicted in grey boxes on OAR-340-041-0006, Table 21 (see Table 15, above).

Instantaneous Minimum Dissolved Oxygen Criteria to Protect Aquatic Life, Table 7.

Dissolved Oxygen Standard	Salmonid Spawning	Cold Water	Cool Water	Warm Water	Estuary	Marine
D.O. Criteria (mg/L)	11.0*	8.0	6.5	5.0	6.5	No change from background
% Saturation Allowance	Not less than 95 % saturation	Not less than 90 % saturation	—	—	—	—
IGDO Criterion (mg/L)	8.0	—	—	—	—	—

*Shall be 9.0 mg/L if data shows the IGDO criterion of 8.0 mg/L is also attained.

Category 5: Water Quality Limited, TMDL Needed (303(d) List)
Where greater than 10% of the samples within the IR data window collected on separate days for the time-period of interest (spawning or year-round critical period) are less than the appropriate criterion according to the exact binomial test and are also less than the percent saturation allowance.

Category 4: Water Quality Limited, TMDL Not Needed
TMDLs needed to attain applicable water quality standards have been approved (Category 4A), other pollution control requirements are expected to address the pollutant and result in the attainment of water quality standards (Category 4B), or impairment is not caused by a pollutant (Category 4C).

Category 3: Insufficient Data
Fewer than 5 samples within the IR data window collected on separate days for the time-period of interest (spawning or year-round critical period) with no sample less than the appropriate criterion, *and* all samples less than the appropriate criterion are also less than the percent saturation allowance.

Category 3B: Insufficient Data: Exceedances
Fewer than 5 samples within the IR data window collected on separate days for the time-period of interest (spawning or year-round critical period); where at least one sample is less than the appropriate criterion *and* is also less than the percent saturation allowance.

Category 2: Attaining
Less than or equal to 10% of samples within the IR data window in the time-period of interest (spawning or non-spawning) are less than the appropriate criterion according to the exact binomial test *and* are also less than the corresponding percent saturation allowance.

Continuous Time Series Dissolved Oxygen Concentration

The Department shall apply the Monthly (30-D) Mean Minimum, Weekly (seven-day) Minimum Mean, and alternate absolute minimum, when it determines sufficient continuously monitored data is available.

For calculating daily means and minimums, measurements from at least 22 hours in each day must be available. Sufficient data will include, but may not be limited to, at least 29 daily mean values for calculating a 30-day average, and at least 6 daily mean values for calculating a seven-day average.

To assess the year-round criteria using continuous data, at least 15 instances of the 30-D metric data must be collected during the year-round critical period (July 1 – September 30) within the integrated report data window. To assess the spawning criteria using continuous data, 15 instances of the 7-D metric must be collected during the spawning period within the integrated report data window.

In the absence of sufficient continuous monitoring of dissolved oxygen, attainment of the dissolved oxygen criterion shall be assessed as instantaneous or "grab" measurements. The daily minimum dissolved oxygen concentration shall be used as the "grab" sample unit.

Sites having insufficient data to be assessed as continuous data will be assessed according to the instantaneous criteria in the previous section. Where multiple samples are collected on the same day, the minimum DO concentration will be used in the assessment.

For the details of the following procedures please see Figure 5 and Figure 6.

Category 5: Water Quality Limited, TMDL Needed (303(d) List)

Where the Department concludes that sufficient continuously monitored data has been collected, it shall assign waterbodies to Category 5 if *any* of the following criteria are exceeded:

Year-Round

- Two or more of the 30-D consecutive rolling averages of the daily mean of dissolved oxygen concentration *and* for those water bodies classified as cold water, the corresponding 30-day average of daily mean percent saturation is less than the applicable criterion.
- Two or more of the 7-Mi consecutive rolling average of the daily minimum concentration of dissolved oxygen is less than the applicable criterion.
- If both of the year round (30-D or 7-Mi) metrics are attained, two or more of the daily minimum concentration of dissolved oxygen is less than the Min. alternate minimum criteria (Min) (Table 15).

Spawning

- Two or more of the 7-D consecutive rolling average of the daily mean of dissolved oxygen concentration *and* the corresponding 7-day average of daily mean percent saturation is less than the applicable criterion, or 9.0 mg/L if data shows the IGDO criterion is also attained.
- If the year round 7-D metric is attained, two or more of the daily minimum concentration of dissolved oxygen is less than the Min. alternate minimum criteria (Min) (Table 15).

Category 2: Attaining
Where the Department concludes that sufficient continuously monitored data has been collected, it shall assign waterbodies to Category 2 if all of the following metrics are attained:

Year-Round

- No more than one of the 30-D consecutive rolling averages of the daily mean of dissolved oxygen concentration *and* for those water bodies classified as cold water, the corresponding 30-day average of daily mean percent saturation is less than the applicable criterion.
- No more than one of the 7-Mi consecutive rolling average of the daily minimum concentration of dissolved oxygen is less than the applicable criterion.
- If both the year round (30-D and 7-Mi) are attained, no more than one of the daily minimum concentration of dissolved oxygen is less than the Min. alternate minimum criteria.

Spawning

- No more than one of the 7-D consecutive rolling average of the daily mean of dissolved oxygen concentration *and* the corresponding 7-day average of daily percent saturation is less than the applicable criterion.
- If the year round 7-D metric is attained, ≤ 1 of the daily minimum concentration of dissolved oxygen is less than the Min. alternate minimum criteria (Min) (Table 15).

A. Calculating Percent Saturation
For evaluation of instantaneous or "grab" samples, the percent saturation corresponding to each sample of dissolved oxygen concentration shall be evaluated when applicable criteria are exceeded to determine if conditions of barometric pressure, altitude, and temperature preclude attainment of the standard.

For evaluation of continuous metrics, the corresponding 30-D (cold water year-round criteria) or 7-D (spawning criteria) percent saturation metrics shall be evaluated when applicable criteria are exceeded to determine if conditions of barometric pressure, altitude, and temperature preclude attainment of the standard.

Direct field instrument measurements of percent saturation are preferred and shall be used if available. However, if corresponding percent saturation data is unavailable, and corresponding water temperature data is available, the value can be calculated using Equation 2 (below)[8]. When the dissolved oxygen saturation is measured in excess of 100 percent, the saturation value used shall be limited to 100 percent for the calculation of metrics. If percent saturation is unavailable or cannot be calculated, DEQ shall apply the applicable spawning and cold-water criteria.

---

[8] Pelletier and Chapra. 2008. Qual2Kw theory and documentation (version 5.1), *Washington Department of Ecology*, Olympia, WA.

Equation 1

$$DO_{Theo} = e\wedge\left[-139.34411 + \frac{1.575701\times10^5}{T} - \frac{6.642308\times10^7}{T^2} + \frac{1.243800\times10^{10}}{T^3} - \frac{8.621949\times10^{11}}{T^4}\right]$$
$$* \left(1 - (0.0001148 * Site\_elvm)\right)$$

Where e = a constant, the base of the natural logarithm (≈2.71828)

T = Temperature in Kelvin

Site_elvm = Site elevation in meters (recorded field value or derived from a Digital Elevation Model)

Equation 2

$$PS = 100 * \frac{DO_{Meas}}{DO_{Theo}}$$

Where PS = Percent saturation dissolved oxygen

DO_{Meas} = Measured Dissolved Oxygen in mg/L

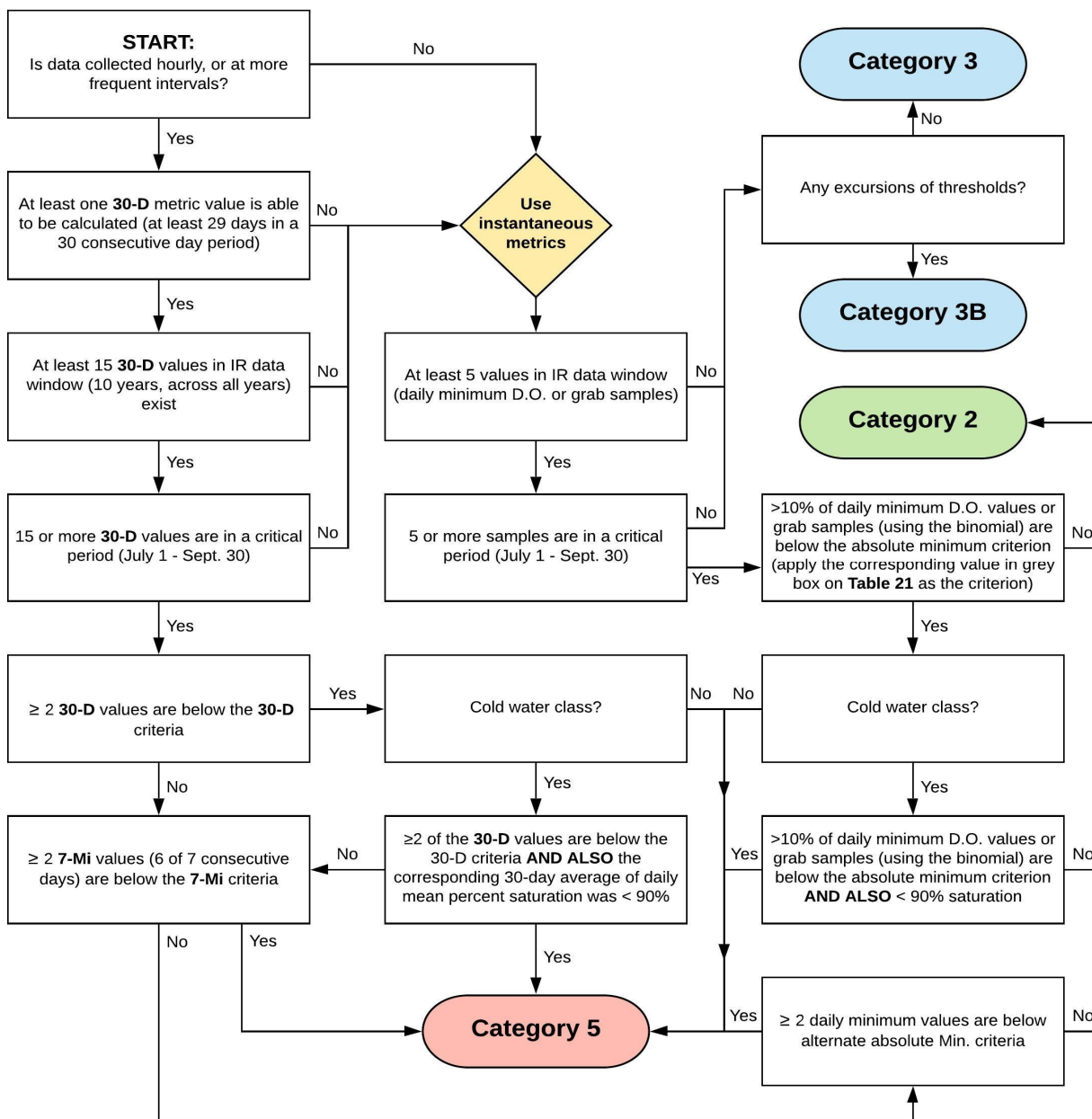DO_{Theo} = Theoretical Dissolved Oxygen in mg/L

Figure 11. The decision tree for assessment of the dissolved oxygen year-round criteria
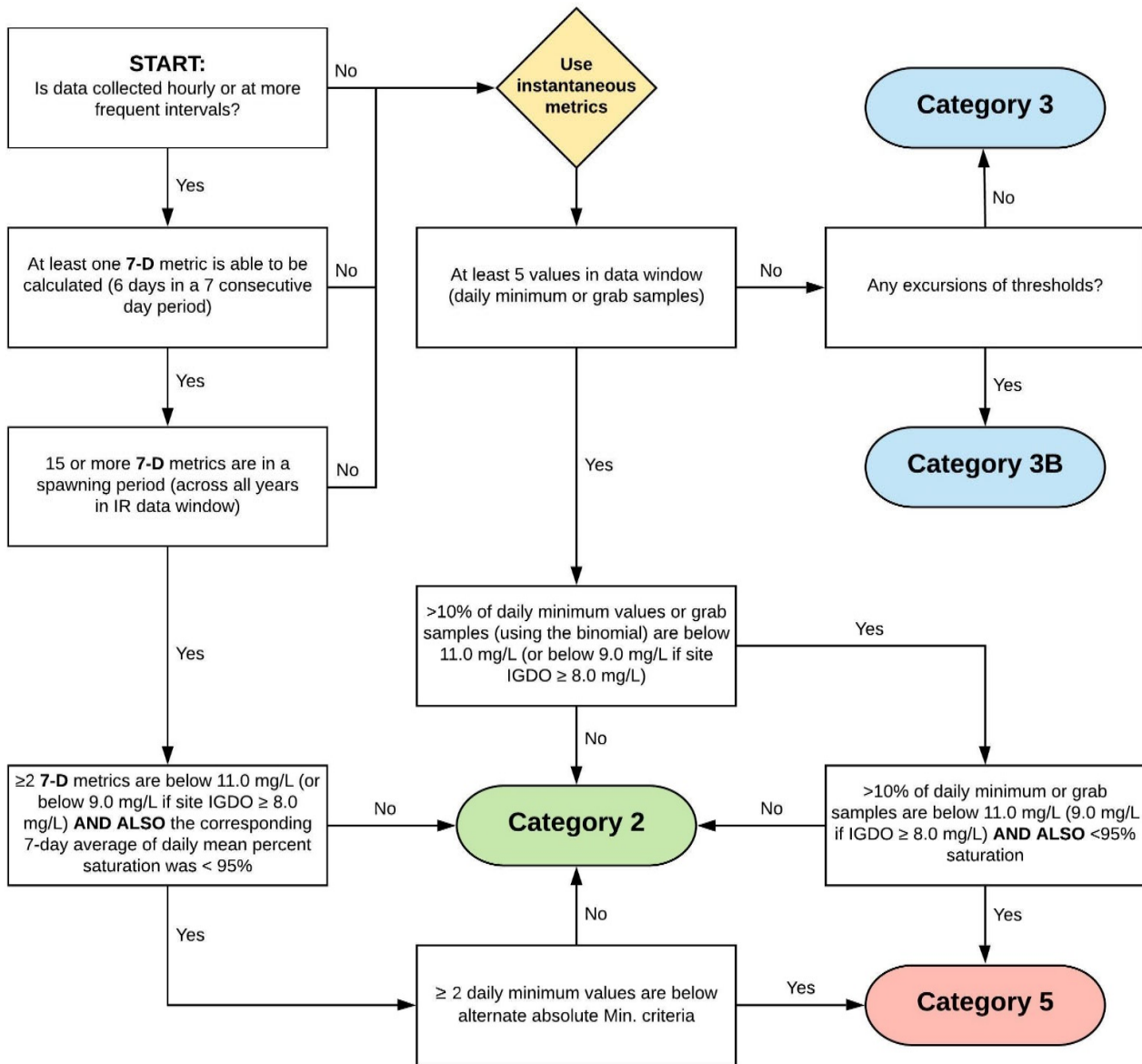
Figure 12. The decision tree for assessment of the dissolved oxygen spawning criteria

Alternative formats

DEQ can provide documents in an alternate format or in a language other than English upon request. Call DEQ at 800-452-4011 or email deqinfo@deq.state.or.us.