Draft 2024 303(d) Assessment Methods Public Comments

The following are public comments received for the 2024 303(d) Assessment Methodology and the Utah Division of Water Quality's (DWQ) response. Please click the links below to jump to your section of interest.

Daniel Lay

coltenlay@gmail.com

Comment 1A: Identifying Causes of Impairments

Absence or lack of water in perennial systems due to drought, over-allocation of water rights, debris flows/wildfire should be considered a class 4C non-pollutant impairment.

DWQ Response to comment 1A:

As described in the assessment methods (page 12, Table 1 and page 74, Category 4C), DWQ may place a waterbody or parameter-specific impairment in category 4C when DWQ can demonstrate that a beneficial use impairment is driven by pollution and not by a pollutant or pollutant that causes pollution; including use impairments driven by hydrologic modification. DWQ recognizes the potential negative impacts hydrologic modification may have on beneficial use attainment and is currently actively engaged in an internal workgroup focused on improving our ability to identify, quantify, and evaluate the impacts of hydrologic modifications on beneficial use attainment in Utah. No method changes are required for DWQ to place impairments driven by hydrologic modification in category 4C. Stakeholders with information or evidence demonstrating that a beneficial use impairment is related to hydrologic modification may submit that information for review during the 2024 Integrated Report public comment period.

Rob Dubuc, Friends of Great Salt Lake

robd@xmission.com

Comment 2A: Assessments Specific to Lakes and Reservoirs

Specific to Great Salt Lake: The Division should develop a method for assessing GSL salinity.

DWQ Response to comment 2A:

DWQ is conducting an evaluation of the current status and trends of water levels, salinity, habitat availability, and aquatic biota in Gilbert Bay to develop appropriate assessment methods. Without defined numeric criteria, it is challenging to develop and implement clearly defined and repeatable assessment methods for assessing whether Great Salt Lake's Bays' including Gilbert Bay are meeting their beneficial uses.Therefore, new assessment methods for Great Salt Lake, including for salinity, have not been incorporated into the 2024 Integrated Report.

Grant Wilson, Earth Law Center

gwilson@earthlaw.org

Comment 3A: Revising the 303(d) List and Other Categorical Assessments

ELC has advocated for complete and accurate 303(d) lists and 305(b) reports for ten years, including in Utah. As you know, we were pleased that there were several flow-impaired waters listed under Category 4C in Utah's 2022 Integrated Report. We also sincerely appreciate the ongoing dialogue and DWQ's openness to form a workgroup to evaluate potential methodologies to assess non-pollutant pollution impairments. We understand this will not be completed before the 2024 303(d) assessment method public comment period this fall.

While these are all positive developments, we still urge you to include a simple statement allowing for Category 4C listings under a "weight of the evidence" approach in the 2024 Integrated report. Drawing inspiration from the "weight of evidence" approach used for Category 5 waters in California, the methodology could look something like this:

When readily available information strongly indicates the non-attainment of water quality standards due to hydrological modification (i.e., 4C waters), a water segment shall be evaluated to determine whether the weight of evidence demonstrates that a water quality standard is not attained and, therefore, a listing is appropriate under Category 4C.

When making a listing decision based on the situation-specific weight of evidence, the DWQ must justify its recommendation by:
• Providing any data or information including current conditions supporting the decision;
• Describing in fact sheets how the data or information affords a substantial basis in fact from which the decision can be reasonably inferred;
• Demonstrating that the weight of evidence of the data and information indicate that the water quality standard is not attained; and
• Demonstrating that the approach used is scientifically defensible and reproducible.

Additionally, or in the alternative, the DWQ could commit to piloting a "weight of the evidence" approach for Category 4C listings beginning with those rivers that flow into the Great Salt Lake. As you know, the Great Salt Lake is undergoing an imminent ecological crisis in large part due to low inflows from the Jordan, Weber,

and Bear rivers. Listing these rivers for Category 4C hydromodification, assuming such a listing is justified under a "weight of the evidence" approach, would send a strong message that the DWQ is using all available policy tools to help address one of the greatest, existential challenges facing Utah.

DWQ Response to comment 3A:

As described in the assessment methods (page 12, Table 1 and page 74, Category 4C), DWQ may place a waterbody or parameter-specific impairment in category 4C when DWQ can demonstrate that a beneficial use impairment is driven by pollution and not by a pollutant or pollutant that causes pollution; including use impairments driven by hydrologic modification. As the commenter notes, DWQ is currently actively engaged in an internal workgroup focused on improving our ability to identify, quantify, and evaluate the impacts of hydrologic modifications on beneficial use attainment in Utah. The commenter's suggestions have been provided to that workgroup.

No methods changes are required for DWQ to place impairments driven by hydrologic modification in category 4C in the manner recommended by the commenter, and the data and information used to make assessment decisions, including potential 4C determinations are published as part of the Integrated Report and available for public comment. Stakeholders with information or evidence demonstrating that a beneficial use impairment is related to hydrologic modification may submit that information for review during the 2024 Integrated Report public comment period.

Comment 3B: Revising the 303(d) List and Other Categorical Assessments

Finally, I would like to make you aware that there is a movement to recognize the Rights of the Great Salt Lake. This builds from other rights-based movements for waterways across the world--the Atrato River in Colombia, the Whanganui River in New Zealand, various waterways and watersheds in Colorado (recognized by non-binding resolution in the communities of Nederland, Grand Lake, and Ridgway), and the rights of Mar Menor in Spain. The DWQ may wish to consider how the rights of the Great Salt Lake, including mechanisms to give it a voice in governance, could benefit your efforts to protect and restore waterways.

DWQ Response to comment 3B:

Thank you for bringing these efforts to our attention. DWQ is deeply committed to protecting the Great Salt Lake ecosystem and all of Utah's waterways. We look forward to seeing how these efforts progress.

Shera Reems, EPA Region 8

reems.shera@epa.gov
EPA Email with Comments PDF

Comment 4A: Assessments Specific to Rivers, Streams, and Canals, Table 11

How does DWQ communicate to the public the water bodies where early life stage presence has been confirmed?

DWQ Response to comment 4A:

DWQ develops tables that identify species present and associated seasons of early life stages for Utah Pollution Discharge Elimination System (UPDES) permits that desire that specificity when wasteloads are developed. These tables are developed on a case by case basis, and like any information that falls under our GRAMA guidance, can be provided to the public upon request.

Comment 4B: Assessments Specific to Rivers, Streams, and Canals, Table 11

The last paragraph under notes for TDS should be deleted. It appears that this information should only be in the notes field for Sulfate.

DWQ Response to comment 4B:

DWQ appreciates you bringing this to our attention. The paragraph referenced in Table 11 in the TDS notes has been removed from the document.

Comment 4C: High Frequency Assessments for Dissolved Oxygen (DO)

In response to EPA's informal question about the rationale for applying the 10% allowable exceedance frequency to the 7-day average and 30-day average DO criteria to high frequency DO data, UDWQ staff reminded EPA about the following language in R317-2-7: "For water quality assessment purposes, up to 10 percent of the representative samples may exceed the minimum or maximum criteria for dissolved oxygen, pH, E. coli, total dissolved solids, and temperature, including situations where such criteria have been adopted on a site-specific basis."

UDWQ emphasized that the 10% exceedance frequency (aka 10% rule) language is parameter specific. In reviewing the state's assessment method for E. coli (pages 60 to 65), UDWQ only applies the 10% rule to the maximum E. coli criterion. To calculate the 30-day geometric mean, Scenario B of the assessment method suggests a minimum sample size of 5 grab samples. The methodology documents that a waterbody would be considered impaired if more than one geometric mean was exceeded. This method does not appear to require sufficient data to calculate whether 10% of the 30-day geometric means were exceeded. In contrast, for a more comprehensive dataset like high frequency DO, the state applies the 10% rule to the 7-day average and 30-day average and requires sufficient data to evaluate whether those averages are exceeded 10% of the time. Please explain this inconsistent application of the 10% rule cited in Section R317-2-7.

EPA encourages UDWQ to consider revising the high frequency DO assessment method to align with the approach used to calculate the geometric mean for E. coli. Specifically, EPA recommends the state: a) define a less stringent minimum dataset for the calculation of the 7-day and 30-day average DO criterion (e.g., 3 to 5 days to calculate the 7-day average and 10 days to calculate the 30-day average), and b) place waterbodies on the 303(d) list with more than one exceedance of the 7-day or 30-day average.

When assessing for DO, UDWQ should also consider the narrative criteria requirement for water quality conditions not to "…produce undesirable physiological responses in desirable resident fish, or other desirable aquatic life…" (R317-7.2). The 10% rule should not be implemented in a manner that obscures potentially harmful or lethal conditions to aquatic communities.

DWQ Response to comment 4C:

The difference in assessment methods between *E. Coli* and dissolved oxygen is derived from differences in water quality standards between the two parameters. The general 10% exceedance frequency language at R317-2-7.1 b. includes both dissolved oxygen and E. Coli. However, following EPA guidance, Utah's E. Coli criteria includes an additional footnote (R317-2 Table 2.14.1, footnote 7) that specifically applies a 10% exceedance frequency for assessment to the two maximum criteria and excludes the 30-day geomean criteria: "For water quality assessment purposes, up to 10% of representative samples may exceed the 668 per 100 ml criterion (for 1C and 2B waters) and 409 per 100 ml (for 2A waters)." Only the general 10% exceedance frequency language is applied to dissolved oxygen criteria, and national guidance regarding appropriate exceedance frequencies for dissolved oxygen assessment is not currently available. Therefore, DWQ has not changed the exceedance frequency used for high frequency dissolved oxygen methods as recommended by the commenter.

As described in the assessment methods, all waterbodies are subject to applicable narrative standard based water quality assessments such as biological assessments, observations of fish kills, and trophic state evaluations.

Comment 4D: Nutrient Assessments Specific to Headwater Streams (Pages 40 to 41)

During the 2022 Integrated Report cycle, EPA recommended DWQ consider including a column in the table, "Draft Integrated Report: 305(b) and 303(d)" that identifies headwater assessment units for which the numeric nutrient criteria apply. This information would also be helpful in the water quality data files that DWQ shares with the public.

DWQ Response to comment 4D:

All headwater watersheds are georeferenced. As suggested, DWQ will include this column in the referenced table.

Comment 4E: Assessments Specific to Lakes, Reservoirs, and Ponds, Aquatic Life Use Support, Figure 9 (Page 49)

In the second row of this Figure, should there be a “No” after the diamond “>10% of water column exceed criterion?”

DWQ Response to comment 4E:

Thank you for your comment. DWQ has updated the diagram in Figure 9 so that it is properly labeled.

Comment 4F: Temperature and Dissolved Oxygen: Stratified Lakes and Reservoirs, Figure 12 (Page 52)

Based on the information in the Figure, it appears that both a Tier I Not Supporting and Tier I Fully Supporting proceed to a Tier II Assessment. Is this correct?

DWQ Response to comment 4F:

The commenter is correct. The Tier II assessment process is independently applied to all lakes regardless of the results of the Tier I assessment methods.

Comment 4G: Great Salt Lake (Pages 56 to 57)

Previously, UDWQ indicated that the Great Salt Lake Water Quality Strategy components on Aquatic Life Numeric Criteria for Priority Pollutants and Strategic Monitoring and Research were in the process of being updated. Does UDWQ have an updated timeline for the completion of these components?

DWQ Response to comment 4G:

Updating the Great Salt Lake Water Quality Strategy and its core components is ongoing with a targeted completion date in mid-2024.

Comment 4H: Harmful Algal Blooms (Pages 66 to 67)

EPA appreciates DWQ’s work in assessing for Harmful Algal Blooms (HABs) and on updating the assessment method to rely on EPA’s recommended criteria for microcystin and cylindrospermopsin. Please clarify the following:

Does UDWQ intend to apply these methods to benthic blooms similar to those observed in Zion National Park?

DWQ Response to comment 4H:

Thank you for asking for clarification. This assessment method will not be applied to any benthic harmful algal blooms affecting recreational beneficial use at this time. DWQ will be evaluating how to best incorporate appropriate assessment methodology for benthic blooms in the future.

Comment 4I: Harmful Algal Blooms (Pages 66 to 67)

Would UDWQ list waters as impaired for HABs if the waterbody exceeds UDWQ's anatoxin-a warning threshold?

DWQ Response to comment 4I:

Until more information is available on the toxicity of anatoxin-a, DWQ chooses to withhold setting a specific cyanotoxin threshold for anatoxin-a in directly determining beneficial use support. DWQ looks forward to EPA recommending similar thresholds for other common cyanotoxins (including anatoxin-a) in the future.

Anatoxin-a concentrations are, however, considered indirectly in beneficial use assessment. The second independent indicator of beneficial use support is waterbody recreational access or use limitations. Recreational advisories which may also list a waterbody as impaired, are generally issued by a health department when concentrations of anatoxin-a are greater than 15 ug/L.

Ellen Bailey, DWQ

ellenbailey@utah.gov
Email comment PDF

Comment 5A: Harmful Algal Blooms

There is a typo referring to table 16 in the HAB assessment. In the text it refers to table 15.

DWQ Response to comment 5A:

Thank you for pointing out these typos. DWQ has updated all the table references in the Harmful Algal Bloom section to the correct table.

Ashley A. Peck, Wasatch Front Water Quality Council

aapeck@hollandhart.com
Comment Letter PDF

Comment 6A: Harmful Algal Blooms

The Council has significant concerns with this approach because the HAB Guidance allows for the use of cyanobacteria cell density – on its own – to inform public health advisory decisions. This means that an impairment decision could also be based solely on two health advisories having been issued for the water body based on localized cell density sampling that is not representative of the water body as a whole – and notwithstanding whether cyanotoxin concentrations were above recreational guidelines. As EPA has recognized, cell density is not a reliable advisory trigger and is more stringent than the federal approach, in violation of Utah Code Section 19-5-105. Moreover, use of cell density-triggered health advisories as the basis for impairment decisions would impart a wholly subjective standard that likewise violates Utah Code Section 19-5-105.

Under the HAB Guidance, cyanobacterial cell density greater than 100,000 cells/mL on its own could result in a "Warning Advisory" for the waterbody in question. However, relying on cell density is contrary to EPA guidance, which focuses exclusively on concentrations of microcystin and cylindrospermopsin to guide recreational health advisories. Indeed, EPA expressly declined to recommend issuing public health advisories based on cell counts, concluding that "available data are insufficient to develop quantitative recreational values" and that additional research is needed given the inconsistency in epidemiological studies. In other words, allowing cell counts to trigger warning advisories makes the HAB Guidance more stringent than its federal counterpart. This runs afoul of Utah Code Section 19-5-105, under which state standards developed in administering a program under the federal Clean Water Act can be no more stringent than federal standards addressing the same circumstances unless the Water Quality Board makes a written finding, after public comment and hearing, that the corresponding federal standard is not adequate to protect public health and the environment. Here, the Board has made no such finding.

Because the inclusion of cell counts makes the HAB Guidance more stringent than its federal counterpart, DWQ should not tie impairment decisions to health advisories issued under the Guidance based on cell count. Instead, the HAB section of the Draft Assessment Methods should focus on whether microcystin and cylindrospermopsin concentrations have exceeded applicable thresholds. This will allow the final Assessment Methods to avoid being more stringent than, and conflicting with, federal standards. To achieve this, the Council requests that DWQ remove mention of warning advisories, danger advisories, and closures for recreational use from the HAB section of the Draft Assessment Methods or at minimum make both health advisories and cyanobacteria concentration together a trigger for impairment.

Beneficial Use Supported
The beneficial use is fully supported if, over the period of record:
Cyanotoxin concentrations have not been identified above recreational use thresholds (Table 15), AND a Warning Advisory, Danger Advisory~~, or closure has not been issued for recreational access to a waterbody.~~

Beneficial Use Not Supported
The beneficial use is not supported if, in representative samples for recreational uses, in two or more years in the period of record:
Cyanotoxin concentrations above recreational guidelines (Table 15) have been reported in more than three 10-day assessment periods in a recreational season, ~~OR a Warning Advisory, Danger Advisory, or closure~~

~~has been issued for recreational access to a waterbody for two or more 2-week periods in a recreational season.~~

Insufficient Data and Information with Exceedances (IR Category 3)
The waterbody will be placed in the insufficient data category if:
It does not meet either of the Beneficial Use Not Supported criteria (above), but cyanotoxin concentrations exceeded recreational use thresholds (Table 15) in three or fewer 10-day assessment periods in a recreation season, ~~OR a Warning Advisory, Danger Advisory, or closure has been issued for recreational use for less than two 2-week periods.~~ These waterbodies will be prioritized for further sampling and evaluation.

DWQ Response to comment 6A:

Utah Code Section 19-5-105 prohibits the Water Quality Board from making rules that are more stringent than the corresponding federal regulations. The methodology for conducting beneficial use assessments isn't a rule promulgated by the Water Quality Board. Similarly, the EPA guidance referenced by the commenter is not codified in federal regulation. As such, the inclusion of consideration of health advisories in the methodology for conducting beneficial use assessments is neither more nor less stringent than the federal Clean Water Act. Rather, health advisories simply provide an additional line of evidence that is independent from cyanotoxin benchmarks.

DWQ strives to consider all readily available and relevant data in the assessment process. Formal waterbody assessments are not limited to numeric criteria as presumed by the commenter. Health advisories provide an opportunity to incorporate local risk assessment expertise into HAB assessments. Recreational use restrictions provide a direct indicator of whether recreational uses are supported in a waterbody. DWQ works collaboratively with local health departments and stakeholders through our Water Quality Health Advisory Panel to ensure that local health departments have the appropriate resources, scientific background, and technical support to make accurate decisions about health advisories. All assessment decisions are subject to a secondary review and public comments. Final assessment decisions consider the weight of evidence of quantitative data, the quality and robustness of available data, waterbody specific information and expertise, and public comment. Assessment decisions based on health advisory information can be modified in secondary review or following public comment if DWQ finds that the underlying data were inaccurate or if the weight of evidence of other indicators or other waterbody specific information demonstrated that to be appropriate. However, these decisions must be made on a case-by-case basis, carefully considering available information for a specific waterbody. No changes were made in response to this comment.

David Richards, OreoHelix Ecological

oreohelix@icloud.com
Comment Letter PDF

Comment 7A: Biological Assessments, Figure 7

[p. 2] Figure 7 on page 45 has been in the UDWQ IR methods for almost a decade. Its caption states that it is a hypothetical example of O/E, but it is delineated into Desert and Mountain ecoregions. This does not support the description of how the RIVPACS O/E model is used by UDWQ that states that the model is based on fifteen predictor variables which produce many more mostly watershed based expected reference taxa sites throughout the state than just two regions, Desert and Mountain. Figure 7 is obviously outdated and its continued use to illustrate the wonders of the RIVPACS O/E model as a sole metric for assessing biocriteria reaffirms my conclusion that UDWQ has put little effort into improving its understanding, assessment, and management of Utah's rivers and streams using benthic macroinvertebrates. As a result, water quality of Utah's cherished rivers and streams continues to be jeopardized.

…[p. 6] Because this figure continues to be included in IR reports a naïve reader would conclude that UDWQ delineates all of Utah's rivers and streams into only two regions, Desert and Mountain. Again, as I illustrated in Table 2 and discussed E for all mountain streams or all Desert streams would be a constant in this case E = 10 for desert streams, E = 30 for mountain streams. Multiply the denominator in O/E by its constant and all that is left is O, observed number of taxa i.e., Taxa Richness.

DWQ Response to comment 7A:

As stated directly in the caption, Figure 7 is a hypothetical example. The intent is to illustrate that O/E has the same meaning, even though Utah streams vary naturally in local richness. O/E compares the taxa predicted to occur at a stream in the absence of human disturbance against those taxa that were observed at the location. The figure simply indicates that O/E can generate similar scores among streams that naturally differ in biological condition.

E is not a constant and contains site-specific information. The provided table and associated discussion reflects a misunderstanding of how O/E, particularly E, is calculated. RIVPACS model predictions are both site-specific and taxa-specific. Of course one can always multiply a fraction by its numerator to obtain its denominator, but doing so in this case eliminates many of the advantages of using O/E as opposed to using species richness. The calculations used to generate E are what allows these models to parse out natural changes in composition (including richness) from those associated with human-caused stressors.

Briefly, E is calculated by first mathematically identifying groups of reference sites with similar taxonomic composition. For an assessed site, the geospatial predictor variables are used to quantify the probability that an assessed site would fall within <u>each</u> taxonomic group (P_g). The frequency with which each taxon occurs among reference sites within each group is then combined with the assessed site's P_g across all groups to predict the probability of capturing (P_c) each individual taxon observed among all reference sites. E is the sum of those taxa with a 50% or greater chance of being observed at an assessed site based on that site's specific geospatial watershed attributes. Moreover, O is not all taxa observed, but those observed taxa there were expected to be present in the absence of human disturbance (those predicted to occur), which is important because human-stress creates conditions that are advantageous for taxa that never would have otherwise occurred at a site. Richness is unable to make such distinctions.

Details of these calculations are provided in the companion document cited in the methods or the hundreds of peer reviewed scientific papers that have been published on RIVPACS models, many of which have been provided to the commentator in previous Integrated Report cycles.

Comment 7B: Biological Assessments, reliance on O/E as sole indicator of biological integrity

[p. 2] The continued reliance and use of O/E metric by UDWQ as the sole measure of water quality 'biological integrity' is extremely troubling and reflects a poor understanding of biological integrity and the use of bioassessment methods as practiced throughout the world, including other neighboring states and counties, even Salt Lake County. No other water quality agency that I am familiar with relies on one metric, particularly with one such metric that has so many flaws as O/E. In fact, after the millions of taxpayer dollars spent on its development, it is no more informative than the Taxa Richness (number of taxa) metric that is easily calculated and used as the single most important bioassessment metric throughout the U.S…O/E is just one of many metrics.

[p. 6] …I don't agree that using a single taxon richness-based metric, RIVPACS O/E would constitute a robust index of biological integrity. It is only one metric that does not address anything other than richness and apparently does not do an adequate job of that (Richards 2016). There is also no reason to make a 'robust IBI' easily interpretable. Ecological interactions between dozens of organisms and their responses to human caused impairment are anything but easily interpretable. RIVPACS O/E models themselves are not easily interpretable. The data and algorithms used in these models are extremely difficult to obtain and often not available, thus not transparent. Other metrics used by other agencies, such as taxa richness, functional feeding group, etc. are very transparent and easily calculable.

[p. 7]…Although O/E may have an intuitive biological meaning as stated by UDWQ, there are so many assumptions, generalizations, and errors associated with derivation of results that its accuracy in assessing loss of taxa and impairment is highly questionable. There are several other diversity metrics in use throughout the world that are much simpler to derive, provide insights into the causes of impairment, and are much easier to interpret than RIVPACS O/E (Table 1 for example and see Literature Cited). These metrics can easily substitute for O/E or at least supplement it.
For example, richness and evenness are better indicators than O/E for several reasons,
1) they are not confounded with other models (e.g., PRISM, a costly and proprietary model that is not transparent except for those who can afford to pay for its use),
2) they are independently verifiable, and
3) they allow assessment of change at local-scale due to point source impacts.

…[from results of study, p. 31]: There were strong effects of evenness and richness metrics on O/E scores, which apparently often affect biological assessments. Taxa richness obviously effects O/E scores because the O/E model is mostly based on this metric. Evenness directly effects taxa richness in a subsample and consequently directly and indirectly effects O/E scores. These effects need to be accounted for by water quality agencies before assigning an assessment score.

DWQ Response to comment 7B:

DWQ and independent experts reviewed this white paper and provided comments back to the commenter, including citations to numerous peer reviewed scientific papers that refute the claims.

As discussed in previous IR cycles (see pp 70-72 in *2016 Integrated Report Response to Public Comments* and pp 200-211 in *Combined 2018 and 2020 Integrated Report*), many of the opinions expressed in these comments (sometimes verbatim) are incorrect or incomplete. There are numerous peer reviewed scientific papers that have evaluated numerous aspects of O/E models, including: the relationship of O/E to richness and evenness, model complexity, their accuracy in identifying biological impairments, sensitivity to human-caused stressors, and many other topics. Many investigations have shown that O/E models are more accurate and often more sensitive to human caused stress than other biological indicators. As a result, RIVPACS continue to be used by numerous states and other countries around the world. This includes USEPA who uses these models in their national assessment programs. Other scientifically supported biological assessment tools exist, but this does not mean that O/E models are not a scientifically defensible method for identifying biologically degraded streams.

Comment 7C: Biological Assessments, definition of biological integrity

[p. 5] There is an urgent need to understand Clean Water Act biological integrity. Let me present one of the most widely used definitions of biological integrity, as defined by one of the leading experts, James Karr and colleagues.

“Biological integrity refers to the capacity to support and maintain a balanced, integrated, adaptive biological system having the full range of elements (genes, species, assemblages) and processes (mutation, demography, biotic interactions, nutrient and energy dynamics, and metapopulation processes) expected in the natural habitat” … (Angermeier and Karr 1994, Karr and Dudley 1981, Karr et al. 1986).

“Integrity implies an unimpaired condition or the quality or state of being complete or undivided; it implies correspondence with some original condition. Health, on the other hand, implies a flourishing condition, well-being, vitality, or prosperity”. “An ecosystem is healthy when it performs all its vital functions normally and properly; a healthy ecosystem is resilient, able to recover from many stresses; a healthy ecosystem requires minimal outside care” (Karr 1996).

I have heard UDWQ staff try to define biological integrity at several meetings, and they consistently offer the “Readers Digest” condensed version of the above definitions in what I can only interpret as an excuse not to fully engage in the complexities of biological integrity and subsequently not fully protecting Utah’s water quality.

DWQ Response to comment 7C:

DWQ scientists understand Clean Water Act biological integrity and are familiar with these papers, but the information is appreciated. DWQ lacks sufficient information to respond to the comment regarding

presentations at public meetings, but DWQ does sometimes simply present complex topics when presenting to the general public.

Comment 7D: Biological Assessments, O/E Reference Sites

[p. 5] UDWQ also claims that O/E is based on similar reference sites derived from fifteen predictor variables. If this is the case, then the expected number of taxa used in the metric becomes irrelevant…Derivation of expected number of taxa, E is problematic and filled with uncertainty that makes its use highly questionable.

DWQ Response to comment 7D:

Comments related to reference sites and model building were also addressed in previous IR cycles (see pp 70-72 in *2016 Integrated Report Response to Public Comments* and pp 200-211 in *Combined 2018 and 2020 Integrated Report*) along with peer reviewed scientific literature citations in the methods and responses that provides the background on RIVPACS models. In short, the models aren't built with similar reference sites, but with a diversity of sites within the mountain and xeric west to capture natural variability.

Comment 7E: Biological Assessments, loss of predicted taxa

[p. 7] As I have emphasized to UDWQ on numerous occasions, RIVPACS O/E models do not quantify loss of predicted taxa. In the case of UDWQ assessments, O/E quantifies only those taxa that were identified from a single (N = 1) composite sample collected from several types of habitats (including riffles and runs) that can exhibit much variability between the macroinvertebrate assemblages. Samples were also identified in the laboratory using a subsample (typically 600 organisms, with large and rare counts). O/E simply quantifies what was observed in a sample, nothing more. Taxa not identified may have or may not have been lost from the waterbody they just weren't counted because other taxa dominated the sample. UDWQ can only conclude that they simply weren't observed, not lost.

[p. 33]…Many RIVPAC O/E users continue to insist that a reduction in O/E scores reflects the extent to which taxa have become locally extinct due to human activities (UDWQ Integrate Report 2016). This is clearly not the case. The analyses included in this report highlight the fact that subsampling and evenness have significant effects on the number of taxa observed, especially the more uneven a sample and subsample. Taxa weren't lost; they just weren't found. They may not have even decreased in abundance. It is possible that other taxa could have disproportionally increased in abundance for whatever reason and that the 'lost' taxa simply weren't counted. To continue to assume that native taxa have become locally extinct because O/E scores have decreased reflects a gross misinterpretation of RIVPACS O/E models.

DWQ Response to comment 7E:

Many of these comments were also addressed in previous IR cycles (see pp 70-72 in *2016 Integrated Report Response to Public Comments* and pp 200-211 in *Combined 2018 and 2020 Integrated Report*). It is true that O/E does not directly measure the loss of taxa, but it does provide a quantitative site-specific estimate of the relative extent to which species loss has occurred among streams. As previously mentioned, the effect of sample error has been extensively evaluated in peer reviewed primary literature. Sampling error was calculated during model creation and these errors were used to create impairment thresholds. This O/E model

has successfully identified biologically degraded streams, which is its primary function in the IR. It has also successfully been able to document improved biological conditions following stream restoration.

Comment 7F: Biological Assessments, probability of capture >50%

[p. 7] Again, as I have discussed on numerous occasions, probability of captures (Pc's) >50% preclude those very macroinvertebrate taxa that constitute biological integrity in a water body (see definitions of biological integrity provided earlier in this response letter). As an example, waters in the Bonneville Basin and in some other parts of UT have unique mollusk assemblages found nowhere else in the world. Most of Utah's mollusks, including native mussels, clams, and non pulmonate snails do not occur in UT waters at Pc rates > 50%. By relying on RIVPACS O/E > 50% Pc, UDWQ failed to protect the unique mollusk assemblages in UT and apparently was unaware of their declines during the time period when continued molluscan viability may have been protected/ensured. This is a tragic and unjustifiable loss of Utah's unique natural heritage. Reliance on a single metric with > 50% Pc to assess biological integrity also likely is not protecting other rare and uncommon macroinvertebrates (< 50% Pc) that are again by definition, biological integrity.

Calculating 'E' using a probability of capture (Pc) of >50% is extremely problematic and results in a poor assessment of biological integrity. Taxa with Pcs < 50% are likely the most sensitive taxa and the very taxa that respond to impairment more that those with Pc > 50%. The statement that "Using a Pc limit set at greater than 50% typically results in models that are more sensitive and precise, which results in a better ability to detect biological stress" is based on two relatively limited studies that evaluated precision using their own methods, i.e., circular reasoning and these were hardly typical. UDWQ is setting a precedent by using Pc > 50% based on results that are not solidly supported in the literature and not established scientific fact but based on a vague ill-defined term in the two studies, 'sensitivity'.

It appears that UDWQ is more interested in the continued reliance on a single metric (O/E) that had good statistical properties (e.g., more sensitive, and precise) than incorporating other metrics or using a < 50% Pc that may prevent loss of rare, uncommon, and unique taxa and provide greater insights into the types of impairments that Utah waterbodies experience. O/E models may be able to detect large levels of biological stress, but not biological integrity.

…[p. 32] RIVPACS O/E models include a 'probability of capture' (Pc) component. Pc is the probability that a taxon occurs at a reference site and is used in the development of the "E" expected taxa list. To reduce 'noise' in results and to ease interpretation, many users, including UDWQ, use a PC > 50%. That is, the probability of a taxon occurring at a site is estimated to be greater than 50%. The decision to use a Pc > 50% has very strong negative implications for assessing the biological integrity of a river or stream in UT. Many ecologists agree that uncommon and rare taxa should be included in ecological assessments and by including these taxa detection of impacts is improved (Turak and Koop 2003; Nijboer and Schmidt-Kloiber 2004). It is also widely recognized that rare taxa are the first to become extinct due to human disturbance (Leitao 2016). Uncommon and rare taxa have also been shown to disproportionally contribute to ecosystem function and integrity (Leitao 2016). For example, native bivalves are extremely important for maintaining water quality via their filter feeding activity and of much concern for developing NH3 criteria. However, bivalves do not occur in >50% of Utah's reference sites and unionids are likely on the brink of extinction in UT (Richards 2016b). A PC > 50% may easily overlook many, many, taxa that are unique to Utah's rivers and streams including

threatened and endangered species, important ecosystems providers, or simply an unknown number of taxa that occur in < 50% of reference streams. These taxa are the true measure of biological integrity and without which will result in a homogenous, biodiversity -limited condition lacking integrity. These taxa are also the most likely to be most sensitive to impacts because their niche breadth is much narrower that taxa that have Pcs > 50%. There is a well-known saying in ecology; 'rare is common, and common is rare' (Pimm et al. 2014). Modifications to RIVPACS O/E models have allowed researchers and managers in England to monitor rare species and to flag Red Data Book threatened species (http://www.ceh.ac.uk/services/rivpacs-reference-database), however they use much lower Pcs. Utah should consider the same.

DWQ Response to comment 7F:

All of these comments were made, in many cases verbatim, in previous IR cycles (see pp 70-72 in *2016 Integrated Report Response to Public Comments* and pp 200-211 in *Combined 2018 and 2020 Integrated Report*) and DWQ created detailed responses to them. We also sent the comments and responses to other experts in biological assessments to ensure that our responses were valid. They were and remain so. The commenter is encouraged to reread those responses and the peer reviewed scientific literature cited in the responses.

DWQ evaluated a P_c of 0 and 0.5 when Utah's models were initially created. The P_c >0.5 model was more accurate and sensitive. More sensitive models, by definition, are better at identifying biologically impaired sites, which is important because that is the purpose of using the model in the IR. As we previously noted, this is hardly a precedent. There have been a number of studies supporting this.

O/E is not biological integrity but an important aspect of it. Protecting, maintaining and restoring biological integrity is the central goal of the Clean Water Act, but that does not mean that DWQ needs to measure all of the many complex and important aspects of it. Doing so on a statewide basis is simply not possible. Instead, Utah and all states use indicators of biological integrity to evaluate water quality objectives. All numeric criteria work like this and so does the O/E model. As the commenter correctly notes, completely measuring all aspects of biological integrity is simply too complicated to accurately measure, particularly on a statewide basis.

Comment 7G: Biological Assessments, RIVPACS O/E precision and predictive ability

[p. 8] The new O/E model in the draft is claimed to be a less precise predictive model than the previous used by UDWQ. A loss of precision in the updated model should be critically reevaluated. Was this updated model selected because in saves time and money?

As far as I can tell, O/E does not address intermittent stream benthic invertebrate assemblages. It is well known by all lotic ecologists that intermittent streams behave differently than perennial streams and that the benthic invertebrate assemblages differ dramatically. I have assisted the State of Idaho and the State of Arizona in their development of bioassessments for intermittent streams. It behooves UDWQ to realize that

intermittent streams are abundant in Utah and are increasing due to global climate change and that intermittent streams need to have a different bioassessment paradigm and suite of assessment metrics than perennial streams/rivers.

DWQ Response to comment 7G:

These comments were made, in many cases verbatim, in previous IR cycles (see pp 70-72 in *2016 Integrated Report Response to Public Comments* and pp 200-211 in *Combined 2018 and 2020 Integrated Report*) and DWQ created detailed responses to them. As our previous response indicated, the new model was selected because it was more applicable to a more physically and geographically heterogeneous group of streams. In particular, DWQ was interested in incorporating larger rivers. Also, climate change can potentially cause systematic changes in macroinvertebrate composition, even among reference sites, so it is important to update the models periodically to account for these changes.

DWQ has not created assessment methods for intermittent streams and agrees that such methods would be useful. Hopefully resources will be available to expand to intermittent streams in the future.

Comment 7H: Biological Assessments, Figure 8

[p. 9] There is no obvious starting point on the tree. It has two diamonds on the same top level that suggest starting points. However, following the first diamond “Were 3 or more samples collected?” if the answer was yes, leads to a nonsense conclusion, “Beneficial use Not Supported”. UDWQ constantly relies on diagrams to illustrate its methods and rational for decision making and this diagram has been used by UDWQ for several iterations of IRs and IR Methods.

DWQ Response to comment 7H:

DWQ thanks you for your comment. Figure 8 has been adjusted to clarify the starting point of the diagram.

Comment 7I: Biological Assessments, incorporation of 1st and 8th+ order streams and rivers

[p. 9] All aquatic ecologists know that there is a big difference in macroinvertebrate assemblages in typical 1rst order vs. 2nd to 5th streams and between 8th plus rivers and 2nd to 5th order stream (please review the River Continuum Concept by Vannote et al.).

DWQ Response to comment 7I:

This comment and others relating to stream size was made in previous IR cycles (see pp 70-72 in *2016 Integrated Report Response to Public Comments* and pp 200-211 in *Combined 2018 and 2020 Integrated Report*) and DWQ completed detailed responses to them. As the commenter suggests, all aquatic ecologists know of ecological changes that occur along a river continuum, those at DWQ included. This is one of the main advantages of O/E models over other biological assessment methods such as IBIs. O/E scores scale along a continuum of ecological conditions using a variety of site-specific geospatial stream characteristics,

whereas IBIs assume that streams in large, predetermined bins should have similar characteristics. The commenter is encouraged to reread previous responses to comments related to stream heterogeneity in previous IR cycles for additional details.

Comment 7J: Biological Assessments, taxonomic resolution

[p. 9] A coarser taxonomic resolution results in a major loss of valuable information provided by individual taxa when 'rolled up' to higher taxonomic level. It also means that some unique or ecologically valuable taxa may be unaccounted for and lost from the AU without knowledge by UDWQ. For example: combining all species of caddisflies in the genus Rhyacophila at least 5 species or more could be lost without UDWQs knowledge. Or by combining all species of the mayfly genus Baetis, several of the more sensitive species may have been lost. UDWQ is well aware that taxonomic (phylogenetic) similarity has very little predictive power for sensitivity to different types of impairment (Richards 2016, UDWQ 2017).

DWQ Response to comment 7J:

These comments were made, verbatim, in previous IR cycles (see pp 70-72 in *2016 Integrated Report Response to Public Comments* and pp 200-211 in *Combined 2018 and 2020 Integrated Report*) and DWQ created detailed responses to them. The commenter is encouraged to reread previous responses for additional details. In short, DWQ requests that all taxa be identified to the lowest possible taxonomic resolution, but this is not always possible. The taxonomic resolution reflects these practical constraints. Despite taxonomic resolution constraints, this model has been able to accurately identify biologically degraded streams throughout Utah, which is its primary function for IR purposes. DNA barcoding and similar techniques are helping address this issue and hopefully DWQ will be able to incorporate these data in future biological assessment tools. Using traditional taxonomy, using a nationally accredited lab and requesting that all individuals be identified to the lowest possible level of taxonomic resolution seems reasonable.

Comment 7K: Biological Assessments, seasonality effects

[p. 10] Seasonality also affects macroinvertebrate assemblages. Summer season has fewer taxa in larval stages that are needed for taxonomic identification and O/E derivation. Comparing summer collected vs. late autumn to early spring samples increases variability and thus precision and accuracy O/E results (e.g., summer samples likely will have fewer taxa and lower O).

DWQ Response to comment 7K:

DWQ is aware that the composition and abundance of macroinvertebrate assemblages varies seasonally. If the general pattern in this comment was true everywhere, this problem would be easier to address. Utah is a geographically diverse place and the periods of maximum abundance and diversity can vary considerably from region-to-region and year-to-year. This temporal variation undoubtedly contributes to declines in model precision and accuracy. DWQ has tried to account for this variation in the modeling process. Interestingly, all versions prior to the current model included ‘Day-of-Year’ as a predictor variable, but this variable was not a significant source of variation in compositional differences among reference sites for the current model. The

use of PRISM predictor variables likely helps account for some of the systematic difference in temporal variation patterns (see also DWQ response to comment 7L). It is also possible that temporal variation is minimized by other model construction decisions (Pc > 0.5, higher level OTUs). If true, this would explain why these models are more accurate and more sensitive in identifying biologically degraded streams.

Comment 7L: Biological Assessments, reliance on PRISM data

[p. 10] As discussed in earlier comment letters; PRISM models are proprietary black box and as such are not independently verifiable and thus are scientifically invalid. The scientific method requires the possibility of independent validations. PRISM models are not reproducible or transparent, which as we all agree, is what we are striving for.

PRISM models rely on historic data (e.g., most of the climate data metrics in Table 12). As an example, "Watershed maximum of mean 1961-1990 annual number of wet days' was 28-year old past data. Conditions likely have changed substantially in 28 years. Clearly the past has absolutely nothing to do with the macroinvertebrates collected next year. Similarly, the average of multiple years has nothing to do with invertebrate assemblages that are mostly multivoltine or univoltine. Their lives are shaped only by the conditions in the years during which they lived… not over multiyear averages. Variables in Table 12 had nothing to do with environmental conditions during the time when the sampled invertebrates lived. This introduces an unmeasurable and significant error to every Pc calculated and prevents the use of field data, which would be site specific. It may have been useful in developing regional models… but it has no place in continued assessment/monitoring and should never be used as such. Only field measurements should be used when possible.

PRISM data errors are also spatially derived mostly from misuse of regional models to monitor local scale changes. These models will complicate every O/E assessment conducted anywhere that there are natural gradients, introducing error in every local assessment. PRISM data often are not precise, and values can change substantially between small changes in elevation within a watershed and sometimes within a few hundred meters. In addition, PRISM values are model predicted values and subject to error.

DWQ Response to comment 7L:

These comments were made, verbatim, in previous IR cycles (see pp 70-72 in *2016 Integrated Report Response to Public Comments* and pp 200-211 in *Combined 2018 and 2020 Integrated Report*) and DWQ created detailed responses. The commenter is encouraged to reread previous responses for additional details, because as noted in previous responses many of these assertions are incorrect or reflect a misunderstanding of how O/E models are constructed.

PRISM data are not proprietary and are freely available. They have been independently tested and validated. They are used by a very large community of scientists across a wide range of disciplines and are continually updated and corrected. O/E models perform best when the predictor variables describe longer-term, generalized conditions. The predictors are site-specific, but ultimately we are trying to distinguish between

different types (i.e, low elevations vs high elevation, mesic vs. xeric, high- vs low-gradient). Prism data provide excellent reach scale information that strongly correlates with the spatial heterogeneity of stream conditions.

For any given stream, the past is the best predictor of what should be there; long-term data show that community composition is stable. In fact, if bioassessment programs had historical data for all streams, predictive models would be unnecessary. Climate change is causing systematic changes in the composition of macroinvertebrate assemblages, which is one reason DWQ periodically updates the models with new reference site data. DWQ is initiating a new round of reference site collections in 2023.

Comment 7M: Biological Assessments, Implications of evenness on O/E Scores and UDWQ Bioassessments

[from study, p. 32] UDWQ uses a mean O/E score of > 0.76 as 'fully supporting' and in general, a score of < 0.69 as 'not supporting' (UDWQ Integrated Report 2016). If the SEM standardized loadings (coefficients) for the total effects of evenness on O/E scores in Table 9 are reasonable, then that would suggest that a 0.07 decrease in O/E score from 0.76 (fully supporting) to 0.69 (not supporting) would only require a decrease in evenness of about 0.044 (0.037 to 0.053). As discussed in footnote 2, page …, taxa abundances in macroinvertebrate samples are rarely if ever even, and this relatively small change in evenness could easily trigger an assessment from 'fully supporting' to 'not supporting'.

DWQ Response to comment 7M:

Thresholds are derived based on an understanding of model error (which is based on actual field measures) and the specific values represent an attempt to balance type I (false positive) and type II (false negative) errors. This is a common dilemma for any regulatory agency in general and perhaps more so with those using biological data. DWQ has stated in the chapter the cost-benefit of ensuring that type I and II errors are appropriately balanced and are not arbitrarily set. DWQ has been using O/E models for over a decade and they have proven to provide robust assessments of biological use support. Follow-up investigations at impaired sites have almost always revealed one or more human stressors and the index has also been responsive to improving conditions following stream restoration.