Biological Indicators of Oxygen Stress in Marine Water-Breathing Animals

Review #1.

Roman & Levin et al. provide a thorough and authoritative review of indicators of seawater deoxygenation, which I found very useful and interesting. I think it represents a great deal of work and knowledge that readers of many disciplines will be grateful for. All of my comments are minor but the authors may find them useful, especially to improve clarity in several places.

It depends on other nominations but this synthesis manuscript is certainly of wide appeal to the broad geoscience community and discusses how the importance of seawater deoxygenation remains underappreciated in the scientific, political, and management communities, and the general public.

The authors greatly appreciate the time taken to provide this thoughtful and helpful review and believe this manuscript is improved by our incorporation of the suggestions made by the reviewer. Our responses are indicated in bold after the individual comments.

My comments are generally preceded by the line number in the original pdf.

67. Delete 'have'. sounds like the evolution was recent **– Incorporated in revised manuscript**

70. comma after 'temperatures' **– Incorporated in revised manuscript**

93. Maybe a note to caution here that, in general, all indicators can give the wrong impression if they are not thoroughly studied or poorly decided e.g. if the same behavior may be correlated with other impacts. **Good point-thank you. Text added: "Note however that indicators can be misleading if confounding variables are not considered or if the indicator does not have sufficient validation for the level of biological organization considered (e.g. species, population, community or ecosystem)."**

137-142. It may be best here to supplement this important section (especially "there were also clear interactions among stressors in their biotic effects") with a meta-analysis that considered interactions with deoxygenation e.g. Reddin et al. 2020, which is also not as restrictive in deoxygenation threshold i.e. 'hypoxic event' as in Sampaio et al. 2021; Vaquer-Sunyer & Duarte 2011 though note that this paper does not use a conventional meta-analysis approach, which limits the robustness of its conclusions

Reddin, C. J., Nätscher, P. S., Kocsis, Á. T., Pörtner, H. O., & Kiessling, W. (2020). Marine clade sensitivities to climate change conform across timescales. Nature Climate Change, 10(3), 249-253.

Vaquer-Sunyer, R. & Duarte, C. M. Temperature effects on oxygen thresholds for hypoxia in marine benthic organisms. Glob. Change Biol. 17, 1788–1797 (2011).

Excellent suggestion, text added: "Reddin et al. (2020) compared the interactive effects of warming, hypoxia and acidification in causing global marine extinction patterns to modern experimental results of the impacts of these stressors on marine organisms. They found that modern clade responses to these climate-related stressors correlated with the clade genus survival rates over the post-Cambrian Phanerozoic with the interactions of dissolved oxygen and temperate having the strongest negative effects for tropical marine animals."

Figure1. The benthos seems a bit empty (perhaps it has already suffered from hypoxia) of indicators. e.g. as mentioned later the microbial mat, lack of bioturbation, benthic organisms 'snorkelling' the better oxygenated water column. Some of the higher up indicators could show their benthic component. Not clear what the different shades of blue are – simply depth? Purely artistic? **The Figure is being modified to include laminated sediments and benthic organisms 'snorkeling'. The different shades of blue have been eliminated.**

165. Can give a little more information on what constitutes a responsive and less responsive species? Not quite a definition perhaps, but more information would be nice.
Text modified: "The amount of HIF-alpha is expressed at a higher rate, accumulates in the cell and is translocated into the nucleus where it accumulates and is then associated with targeted metabolic responses. The response time of HIF expression varies between the species tested (Alderdice et al., 2021). "

Section 'sensory systems'. The section lacks a summary that the previous section had. A general theme here seems to be that the use of sensory system changes as a deoxygenation indicator is promising but needs much further research before any general application is possible?
Introductory summary paragraph added: "*Sensory Systems:* Low oxygen can impair the sensory systems of marine water-breathing animals. Changes in animal vision, olfaction and perception of sound can be sensitive indicators of oxygen stress which impact population and community ecology. While promising, more research is needed to elucidate the primary and secondary effects of sensory impairment by low oxygen stress on different groups of marine animals to be practical for implementation in monitoring programs."

221. "levels of growth hormone-insulin-like growth factor". I am not familiar with this topic but this description is particularly difficult to understand
Sentence modified: "Hypoxia can also alter the levels of growth hormones thereby negatively affecting growth among individuals relative to normoxic conditions (Hou et al., 2020).

223. Is this sentence needed? Seems obvious. "Cortisol levels increase when oxygen levels are low enough to cause a physiological stress response"
Sentence deleted

245. "understood," rather than "understand"
Corrected

In general, it would improve clarity to use more commas, where appropriate, to show the main sentence from the components that add information but are grammatically unnecessary e.g. the "where appropriate" above. Alternatively, long sentences could be broken into two or more smaller sentences. There are many instances early on in the MS (it seems not so bad later on) but here is one example (L271-4) of both the above points:

e.g. "The laboratory data were then used to develop a model of the endocrine functioning of vitellogenesis of individual fish (Murphy et al., 2009) to examine how the indicators measured as blood and organ concentrations would vary over time and under exposures not replicated in the laboratory."

To

"The laboratory data were then used to develop a model of the endocrine functioning of vitellogenesis of individual fish (Murphy et al., 2009). This allowed examination of how the indicators, measured as blood and organ concentrations, would vary over time and under exposures not replicated in the laboratory."
Thank you, text changed to: "The laboratory data were used to develop a model of the endocrine functioning of vitellogenesis of individual fish (Murphy et al., 2009). This allowed examination of how the indicators measured as blood and organ concentrations, would vary over time and under exposures not replicated in the laboratory. These model results were applied to field data from the northern Gulf of Mexico and the indicators of hypoxia exposure / effects were used to assess hypoxia effects at the population-level (Thomas et al., 2015; Rose et al., 2018a). "

345. "estimate" rather than "ascertain"
Corrected

388. Also consider that spawning adults may have greater sensitivities than other stages (though whether this results in mortality or lower ultimate fertility may determine which section this is more appropriate for)
Sentence added: "Spawning fish with salinity challenges and feeding cessation may also be more sensitive to low oxygen."

413. Three as written form, rather than '3'
Corrected

446-448. This sentence is unclear: "The resting metabolic oxygen demand of the 447 metabolic indices (Deutsch et al., 2020; Penn et al., 2018) occurs at the onset of mortality 448 or anaerobic metabolism (Deutsch et al., 2015)." The demand occurs at the onset of mortality? This doesn't make sense to me.
Yes, the sentence was badly worded and did not add to the description of the indicator so it was deleted

466. DIC should be familiar but doesn't harm to define. **Added: Dissolved Inorganic Carbon**

483-5. Commas are a bit messed up in this sentence.
Sentence changed to: "Both the MI and FMR have the potential to be widely used for direct measurements of metabolic rate in fishes; serve as valuable input data to models; and are important tools to assess fish in a future warmer ocean with less oxygen."

498. "types of soft-bodied fishes, cnidarians (jellies) and ctenophores". Maybe give an example of these fishes. Also, perhaps be more specific with 'cnidarians', since that includes corals, which I don't think would be a good fit here. 'Jellies' is rather informal – can a better/more taxonomic term be found?
Sentence changed to: Examples of macroscopic ligooxyphiles include certain types of soft-bodied fishes (e.g. *Cheruble emmelas* and *Cephaluros ceplalus*), medusa and ctenophores."

504-508. Maybe note that care need taking to account for changing sedimentation or even erosion rates.
Sentence changed to: "The presence of biogenic materials from sentinel/indicator species (e.g. shells, scales, otoliths, environmental DNA (eDNA), bones) in sediment cores could also be used to examine species response to oxygenation through time (considering changing sedimentation and erosion) and to detect hypoxia indicator taxa (e.g. Moffitt et al., 2015)."

536. It would be good to mention that the use of a single indicator taxon may incur error as species niches are not unidimensional, while the use of multiple indicators at once (and multiple threads of evidence in general) should be more robust.
Good point; sentence added: "Note that a single indicator taxon niche is rarely unidimensional such that multiple indictors can provide a more robust effect of deoxygenation."

562-7. Please split this long sentence.
Sentence split: "For example, diel-cycling hypoxia increased acquisition and progression of the pathogen *Perkinsus marinus* (Dermo) infections in oysters (*Crassostrea virginica*; Breitburg et al., 2015). The authors found stronger effects on younger (1 y) oysters and patterns of prevalence and intensity of infections varying with spatial patterns of the frequency and intensity of diel cycling hypoxia."

718. Perhaps add Simpson's 1-D. I am not familiar with the 'ESx' acronym the authors have here. In my mind, rarefaction is a sampling standardization approach, not a diversity metric in itself, but any of these listed metrics could be based on rarified data. E.g. rarefaction curves are usually of species richness values. Also, species dominance and evenness are two sides of the same coin, rather than separate (usually Simpson's D or 1-D).
Simpson's D added: ". Common indices include species richness (*S*), Shannon Wiener (*H'*), Simpson's D, Pielou's *J*, Rarefaction (*ESx*), Hill numbers (^{q}D), and Rank 1 dominance (*R1D*). "

753. Perhaps change “shifts to a dominant that is a hypoxia-tolerant species” to “shifts to dominance by a hypoxia-tolerant species”
Sentence changed to: “Thus, when the community shifts to a dominance by a hypoxia-tolerant species, it can be a good indicator of oxygen stress at the ecosystem level. “

759. Salinity may be another important factor in coastal waters
Salinity added: “. Diversity, evenness and dominance can also be influenced by other factors such as salinity, food availability or contamination, independently or synergistically with deoxygenation (Rozenzweig and Abramsky, 1993; Levin and Gage, 1998; Pilo et al., 2015). “

765. These references for ‘fossil forming biota’ all seem to be about foraminifera. Would be nice to have some macrofauna references e.g. Aberhan, M., & Baumiller, T. K. (2003). Selective extinction among Early Jurassic bivalves: a consequence of anoxia. Geology, 31(12), 1077-1080.
Reference added: “…with the paleo literature replete with examples for fossil forming biota (e.g., Tsujimoto et al., 2008; Aberhan and Baumiller, 2003; Yasuhara et al., 2012; Moffit et al., 2014; 2015; Singh et al., 2015).”

766. I would not say alpha diversity is scaled up to beta diversity as one only needs two sites to obtain a beta diversity estimate (i.e. change between two sites). Correct that gamma diversity is the next scale up from alpha diversity.
Changed as suggested: “Alpha diversity can be scaled up to gamma diversity across gradients or larger geographic scales.”

839. Can be more specific? “than that required to see changes in diversity”. But several diversity metrics use abundance data, so these changes would be visible in those metrics, wouldn’t they? Depends how diversity is defined e.g. if it was species richness, I would agree with the authors.
Thanks- sentence re-written: “At the community level, changes in animal density can occur at a higher oxygen threshold than that required to see changes in species richness or diversity. For demersal fish in the Gulf of California, for instance, DO 3 +/- 1 µmol kg^{-1} was identified as the threshold below which fish density decreased (Gallo et al., 2020), compared to a DO threshold of 7 µmol kg^{-1} for diversity (H’).

959. There is a stray colon and then semi-colons later in the sentence in a way that looks intended but is not the usual way these are used. Please write as a normal sentence
Sentence edited: “Under low-oxygen conditions, tolerant species can become more dominant throughout the food web because their predatory strategies are less affected and/or escape behavior is less impaired relative to other species with which they interact (Breitburg et al., 1997; 1999).

Discussion.

The different aspects of scale that the authors are delineating here are not completely clear to me. Please try to distill a clear summary of each aspect. The first aspect of scaling here appears to be

sampling, i.e. relating the scale aspects of the sample to the wider population that the samples are attempting to capture. Second aspect seems to be level in the biological and ecological hierarchy, which are linked mechanistically and thus scales can be crossed not by statistics but via this mechanistic understanding. The third aspect I am unclear about how it differs. It seems an integration of statistical and mechanistic modelling (aspects 1 and 2)?

Thank you for the excellent suggestion. The Scaling Section has been re-written: "Scaling of indicators is often necessary to enable the observed values of the indicator to be interpreted as representing the state of the system and for results to be expressed on spatial and temporal scales that are ecologically or societally meaningful. Consideration of what types and to what extent scaling is needed is important when selecting an indicator, designing a sampling plan and interpreting and communicating the results of an indicator. Scaling often determines what species and life stages to measure, the specific indicator(s) needed and how to allocate effort to sampling locations and frequency of sampling.

Scaling can employ graphical or statistical analyses to extrapolate the measured conditions of individuals to broader areas than those locations sampled (e.g. sub-regions, basin-wide) or to more generalized timescales than those captured by the data (e.g. month, season, years). This scaling employs the statistical concept of looking for patterns in the data collected at different locations and/or over time and subsequently making key assumptions about how these data reflect broader conditions to infer the population of indicator values.

Another type of scaling analysis is used with indicators to derive a mechanistic understanding of how the indicator logically and causally relates to higher levels of biological organization (population, community or ecosystem). For example, low-DO impaired vision affects detecting prey that determines feeding and growth that affects vulnerability to predator (mortality) and fecundity, which affect population abundance. This integration and scaling across levels of biological organization from the organismal to ecosystem level can be represented in a conceptual diagram (Altieri and Witman, 2006), where low-oxygen stress reduced survivorship and growth of individual mussels and impacted the density and spatial extent of mussel populations. Individuals of a single species could be used to infer the state of the population while observations on multiple species can be leveraged to community (e.g. diversity) and food web levels (e.g. energy pathways). The condition of individuals as indicated by lipid content (e.g. Herbinger et al.,1991) suggests sufficient exposure to low-DO can elicit a response of the bioenergetics and physiology of the individual. Reduced animal condition can be related to the oxygen state of the system and can lead to higher mortality, lowered fecundity and other responses that can be directly related to population, community or food web levels. While values of indicators on subsets of individuals can stand alone to show exposure and responses of individuals, scaling translates indicator observations into potentially more-relevant levels of biological organization and scales of time and space. This mechanistic scaling approach was used by Rose et al. (2018a, b) to examine how reduced growth, increased mortality, and reduced fecundity due to low-DO exposure affected croaker (*Micropogonias undulatus*) population dynamics in the Gulf of Mexico. By using an agent-based model with a 2-D grid that included dynamic DO field, the time-dependent exposures of individuals were simulated and avoidance behavior was projected.

Scaling of indicators can include numerical models which provide a quantitative translation of the indicator into variables that are more relevant to management and

society. Common situations requiring such modeling are when multiple stressors covary and DO effects need to be isolated or when expressing indicators in units explicitly chosen to inform policy (e.g. economic impacts of reduced biodiversity) and management decisions (e.g. fisheries yield). For instance, Franco et al. (2022) scaled low-DO effects to habitat changes of Pacific halibut (*Hippoglossus stenolepis*) in the Northeastern Pacific. They used fisheries-independent data and model predictions from ROMS-BEC of oxygen and a metabolic index was used to map suitable aerobic habitat.

1071. Please break this sentence down e.g. split it. It is unclear. "Modeling involves significant effort beyond the conceptual modeling that is thus done in 1072 scaling of indicators when quantitative links from the indicator to the system state are 1073 needed." Especially "that is thus done"
This sentence was deleted in the revised section in the previous response.

1081. "time and space scales". Temporal and spatial scales
Sentence changed to: "The suite of indicators discussed differ in the temporal and spatial scales of oxygen influence which are reflected in the types of settings where it is most applicable, in possible confounding factors and in the expertise and resources required for application."

Table 1. This table did not display well on the pdf and I am not sure the displayed version is the complete version e.g the edges look like they were cut off. The text in many cells is cut off and the font is difficult to read. Better spread table over two pages than miss information. Acronyms and abbreviations need spelling out in a footnote or somewhere. Column 'Useful/relevant settings': aren't all these indicators to be used in the field? Makes this column dubious in its use. 'Expertise required': why has 'basic' got a question mark? One cell in this column also has a red earmark.
The Table has been simplified, edited and reformatted. See attachment.

1117. What does "non-sublethal" mean?
Corrected to "Sublethal"