Review of: "Disentangling the effects of eutrophication and natural variability on macrobenthic communities across French coastal lagoons" by Jones, Schaal, Boye, Creemers, Derolez, Desroy, Fiandrino, Mouton, Simier, Smith, and Ouisse for PCI Ecology.

Project Summary and Overall Comments:

Changes in the composition and structure of benthic macroinvertebrate communities have been linked to changes in environmental quality, such as eutrophication, hypoxia, and pollution. Thus, benthic macroinvertebrate communities are considered a cornerstone of environmental management and are used as indicators of ecological quality. However, benthic macroinvertebrate communities respond to both anthropogenic and natural stressors, making it crucial to disentangle the effects of natural environmental variability and the effects of anthropogenic stressors. In this study, the authors used environmental survey data form French Mediterranean coastal lagoons to 1) disentangle the effects of anthropogenic eutrophication and natural variability, and 2) understand the links between environmental variables that affect benthic macroinvertebrate communities. Specifically, the authors used various statistical techniques to determine the relationships between natural environmental variability (e.g., lagoon-sea connection, macrophytes, oxygen saturation, salinity), eutrophication (e.g., total nitrogen, chlorophyll *a*, ammonium, phosphorus), macrobenthic community structure, and taxonomy-based indices. Their results suggest that the joint effect of natural variability and eutrophication had the largest impact on macrobenthic community structure and the taxonomy-based indices, followed by either natural variability or eutrophication alone depending on the biotic metrics examined, with each environmental variable combination acting on different aspects of community structure and composition.

Overall, I believe the authors' work is of scientific importance to the fields of ecology and conservation biology. Understanding and disentangling the effects of natural environmental variability and anthropogenic stress on benthic macroinvertebrate structure and composition is vitally important for making sound environmental management decisions. The authors used multiple appropriate statistical approaches to understand the complex and interconnected relationships both between the environmental variables themselves and macrobenthic communities, and the results were presented clearly. Most of my concerns and comments are regarding methodological clarity and therefore repeatability, overinterpretation of results, potential limitations, and comparisons with similar studies. Perhaps my most important comment is the calculation of appropriate reference conditions for M-AMBI as it may impact the results and interpretations.

Again, I wish to re-emphasise the intellectual merit and scientific importance of the authors' work, and I look forward to reading the revised manuscript.

Major Comments: (Specific comments are in the order in which the subject matter appears in the manuscript, with general comments at the end)

Lines 308-309: M-AMBI is not calculated at the replicate level and then averaged at the station level. Only AMBI is calculated at the replicate level and then averaged at the station level, while both species richness and Shannon-Wiener entropy are calculated on the pooled replicates. This distinction is directly stated in the R script created by Sigovini et al. (2013), and is therefore stated in the authors' code "ambi&M-AMBI.R" in Line 46 of the R script "## calculation of AMBI-BC on each replicate (other metrics are calculated on pooled replicates)". If the authors have changed how M-AMBI is calculated by calculating M-AMBI values for each replicate separately and then are taking an average, then they need to state exactly why they have decided to change how M-AMBI is calculated. If not, then please correct Lines 308-309 in the manuscript.

Lines 313-316: The authors defined the reference conditions using the default settings of highest species richness, Shannon-Wiener entropy, and lowest AMBI, irrespective of the WFD physical-chemical status. However, Borja et al. (2012) argues that the inability of M-AMBI to detect the response of benthic macroinvertebrate communities to anthropogenic stressors is often linked to the use of inappropriate methods for setting reference conditions, and they recommend setting reference conditions based on minimally impacted or least disturbed areas. This is especially true for transitional environments like estuaries and lagoons, which typically have lower richness, diversity, and have higher proportions of pollution tolerant taxa than their fully marine counterparts.

Additionally, France has adopted M-AMBI as its national index under the Water Framework Directive and has set its own reference conditions and adjusted the Ecological Quality Status boundaries. As part of the Transitional Waters Mediterranean Geographic Intercalibration Group, France derived reference conditions using minimally impacted sites from lagoons for Thau & Leucate. The final reference conditions chosen for Polyhaline-Euhaline coastal lagoons were species richness = 46, Shannon-Wiener entropy = 4.23, and AMBI = 0.6 (Reizopolou et al., 2018). The High-Good boundary was set at M-AMBI = 0.84, and the Good-Moderate boundary was set at 0.63 (Commission Decision (EU) 2018/229).

I highly recommend the authors re-run their M-AMBI analyses using the reference conditions established by the French government, or create their own using following the same criteria outlined in Reizopolou et al. (2018), to see how their results would change (if they do change), and how those results compare to using the default reference conditions. For example, the authors did not find a significant relationship between M-AMBI and Oxygen saturation, while the Intercalibration Group did find a strong relationship between M-AMBI and Oxygen saturation for France (Reizopolou et al., 2018). Are the different findings a result of different reference conditions, or different statistical tests?

Borja et al. 2012. The importance of setting targets and reference conditions in assessing marine ecosystem quality. *Ecological Indicators*, 12(1), 1-7.

Commission Decision (EU) 2018/229 of 12 February 2018 establishing, pursuant to Directive 2000/60/EC of the European Parliament and of the Council, the values of the Member State monitoring system classifications as a result of the intercalibration exercise and repealing Commission Decision 2013/480/EU (notified under document C(2018) 696) Text with EEA relevance.

Reizopolou et al. 2018. Transitional waters Mediterranean Geographic Intercalibration Group. Benthic invertebrates fauna ecological assessment methods; EUR 29561; Publications Office of the European Union, Luxembourg, ISBN 978-92-79-98373-3, doi:10.2760/625400, JRC114720.

Comment: Disentangling the effects of anthropogenic stressors and natural environmental variables is tricky, especially as stressors and variables can co-vary, even in the absence of anthropogenic impact. For example, Nitrogen and Organic Carbon both naturally vary with sediment grain-size, as does heavy metal concentrations, all of which impact the structure and composition of benthic macroinvertebrate communities. However, it is often impossible to measure every single possible stressor, and therefore there are limitations to our work, and it is vital that we, as researchers, acknowledge those limitations and describe how such limitations may have impacted our findings. There are two types of variables that are not examined by the authors, sediment grain-size and heavy metal concentrations. The authors state in lines 214-217 that they did not include grain-size due to differences in laboratory protocols between lagoons and between years. Therefore, I do expect, nor request, that the authors include grain-size nor heavy metal concentrations in their analysis. I do ask that they address these limitations in the discussion, reiterate why they were not included, and how the exclusion of the aforementioned variables could have impacted their results.

Comment: The authors' work and the manuscript's scientific impact would favour greatly from comparisons with other published work on the sensitivity of biotic indices, such as M-AMBI, to multiple anthropogenic stressors and natural variability. Where do the different works agree? How / why do they differ? Understanding and parsing out the effects of natural variability on biotic indices is paramount to developing robust ecological / biological monitoring programs. I ask the authors to include a section in the discussion comparing their results with previously published work. Below are several studies which aimed to assess and disentangle the effects of natural variability on biotic indices, which can be used as a starting point:

- Basset et al. (2012). Natural variability and reference conditions: setting type-specific classification boundaries for lagoon macroinvertebrates in the Mediterranean and Black Seas. *Hydrobiologia*, 704: 325-345.

- Barbone et al. (2012). Linking classification boundaries to sources of natural variability in transitional waters: A case study of benthic macroinvertebrates. *Ecological Indicators*, 12(1): 105-122.
- Berthelsen et al. (2018). Relationships between biotic indices, multiple stressors and natural variability in New Zealand estuaries. *Ecological Indicators*, 85: 634-643.
- Paul et al. (2023). Evaluating the effectiveness of M-AMBI with other biotic indexes in a temperate estuary. *Marine Pollution Bulletin*, 193: 115194.
- Pelletier and Charpentier. (2023). Assessing relative importance of stressors to the benthic index, M-AMBI: An example from U.S. estuaries. *Marine Pollution Bulletin*, 186: 114456.
- Pollice et al. (2015). Bayesian analysis of three indices for lagoons ecological status evaluation. *Stochastic Environmental Research and Risk Assessment*, 29: 477-485.

Minor Comments: (Specific comments are in the order in which the subject matter appears in the manuscript and code)

Lines 302-318: I am uncertain whether the taxonomy-based indices were calculated using the raw or transformed abundances, given that the abundance was transformed for the macrobenthic community structure analysis. Please clarify in the text.

Lines 331-334: Why was a Pearson correlation of |0.6| chosen as the cutoff for multicollinearity? Cutoff boundaries can be rather subjective and often vary between studies, which makes cross-study comparisons difficult. Please state the reasoning and/or provide citations for the common use of that particular cutoff value.

Lines 337-339: Similar to the previous comment, why was a VIF of 5 chosen as the cutoff? Please state the reasoning and/or provide citations for the common use of that particular cutoff value.

Lines 650-654: The authors claim that their results confirm, along with water renewal and environmental instability, that "primary colonization and/or post-disturbance recolonization of lagoons by marine-originating larvae through dispersal and recruitment" strongly shape lagoon benthic communities. However, because the study was not testing, nor was looking at, larval dispersal and recruitment, the results cannot be said to confirm the importance of colonization / post-colonization of marine larvae in shaping benthic communities. Instead, the authors can suggest, based on other published evidence, that colonization / post-colonization of marine

larvae could help explain the authors' results, or that their results may support such ideas. But, the results of the authors' research cannot be said to confirm something it was not testing.

<u>Lines 836-838:</u> It is unclear how the replacement of sampling by Ekman-Birge grabs with diver-operated sampling would limit the sensitivity of M-AMBI to natural variability, other than the mention reducing habitat heterogeneity. What other reasons for switching sampling techniques would decrease sensitivity? Also, I recommend the authors describe potential disadvantages of switching from Ekman-Birge grabs diver-operated sampling, such as the increased cost of using trained divers and potential selection bias by divers, which would aid a potential reader in weighing the pros and cons of replacing their current sampling methods.

<u>ambi&M-AMBI.R Line 117:</u> As noted in the "Essential amendment to the R script in the Electronic supplementary material of Sigovini M., Keppel E., Tagliapietra D. (2013) M-AMBI revisited: looking inside a widely-used benthic index. Hydrobiologia 717: 41-50", due to changes in the function "factor.scores" in the psych package, Line 117

METRICS.scores <- factor.scores(METRICS.tot, f = METRICS.fa, method = c("components"))$scores

provides incorrect metric scores and should either be removed from the code or a hashtag (#) should be placed in front of Line 117, which ensures that R will not run that line of code. While this correction should not affect the final M-AMBI scores, as the proceeding line of code (Line 118) should overwrite Line 117, hence why I have not marked this as a "Major Comment", it could cause unnecessary confusion for anyone trying to replicate the authors' work. Also, I recommend that the authors re-run their AMBI and M-AMBI calculations after making the correction to ensure that the results do not change.

<u>ambi&M-AMBI.R Lines 123-130:</u> These lines of code are redundant as the metric scores have already been calculated in Lines 116, 118, and 119. This should not affect the results as both methods for calculating the metric scores should give the same outcome, hence why, again, I have not marked this as a "Major Comment". Similar to my previous comment, please either remove Lines 123-130 or place a hashtag in front of each line to reduce the potential for unnecessary confusion and re-run the AMBI and M-AMBI calculations after making the correction to ensure that the results do not change.