“Disentangling the effects of eutrophication and natural variability on macrobenthic communities across French coastal lagoons”

Please find below the comment from Referees **(in bold, that we have numbered**) and our answers (in blue). Citations from new version of the manuscript are *in italic*.

Reviewer #1

Thank you for the opportunity to review the preprint, "Disentangling the effects of eutrophication and natural variability on macrobenthic communities across French coastal lagoons". This paper used a broad dataset from 29 lagoon systems located along the French Mediterranean coast to identify the environmental drivers of macroinvertebrate communities, based on a number of key features including the level of connection to the sea, water column and sediment characteristics and local macrophyte community. Further, this paper aimed to identify the contriubtion of anthropogenic drivers, linked to eutrophication and oxygen levels, to the structure and distribution of the macrofauna assemblages, in order to better understand the response of various indicators (multivarite and univariate indices) to natural or anthropogenic stressors, given the highly dynamic nature of these systems.

This paper was a pleasure to read and to review, and was one of the very few papers that I have reviewed lately that doesn't require major revision. I must commend the authors on their effort and the quality and standard of this work, and their attention to detail, particulalry with regard to the statistical robustness of this paper. It is refreshing to see (and review) a piece of work that has given proper thought and consideration to the data processing and analysis, and statistical testing required to address the aims and objectives of the paper, which I find is worryingly neglected in many of the other papers I review. I enjoyed reading this paper, I found it interesting and relevant and will be of value to the scientific community.

I include here just a few minor comments and suggestion provided in more detail below.

Introduction:

Question 1.1. Is there a specific definition for lagoon (like those in this context)? If so, it might be good to brielfy include it in the beginning of the introduction somewhere, particularly for people that are not used to working with these systems. For example, I am more familiar with estuarine systems, and I'm not sure how they differ or compare with one another.

We have added some additional information.

“Coastal lagoons are inland, shallow and transitional water bodies, usually oriented parallel to the coast, separated from the sea by a barrier, connected to the sea by one or more restricted inlets which remain open at least intermittently (Kjerfve, 1994). These ecosystems represent 13% of the world’s coastline (Barnes, 1980) and provide essential ecosystem services, especially through nutrient regulation and sequestration as well as climate regulation (Kermagoret et al., 2019; Levin et al., 2001).” [Lines 53 - 57]

Question 1.2. Line 69 - is it meant to be "lagoons"? (or lagoonal systems?)

We have reworded to be clearer.

“Consequently, lagoon benthic communities, just like estuarine ones, generally present low species richness and high abundances of tolerant species...” [Lines 68 - 72]

Question 1.3. Lines 71/72 - Perhaps consider briefly explaining what the "transitional waters quality paradox" is, which would help readers that are not very familiar with this work.

We have added some additional information

"These paradoxes highlight the problems faced by transitional waters concerning the development of methodologies permitting the evaluation of their ecological quality status, because it results from both anthropogenic pressures and natural characteristics." [Lines 72 - 74]

Question 1.4. Lines 122/123 - I'm not really used to the use of the word compartment(s) in the way that the authors have used it in the introduction, but that's just me. Perhaps just try to be conscious of making it clear as to what types of compartments you are referring to. Like here, my immediate reaction was "what compartments/what are you referring to here), but a simple fix in this line would be to move the content in brackets (water column, sediments...) to immediately after the word "compartments".

We have reworded to be clearer and we have moved the bracket section just after the word "compartment".

"In this study, we make use of the full potential of large-scale environmental surveys of the various physical and biological compartments (water column, sediments, macrophytes and benthic macroinvertebrates) of French Mediterranean coastal lagoons..." [Lines 123 - 128]

Methods:

Question 1.5. In figure 1, this is the first place where you mention the groupings of the stations into salinity types and group membership. It might be nice to briefly introduce this (and you don't have to include the details that you provide later on) in the Study sites section. I also didn't notice any details about how the sites were classified into the different salinity groups (not sure if I missed this).

The French Mediterranean lagoons have been classified according to their median salinity and the Venice classification (Anonymous, 1958). All results are detailed in Le Fur et al., 2018. We have therefore added theses references to the legend.

Anonymous, 1958. The Venice System for the classification of marine waters according to salinity. Limnology and Oceanography 3: 346–347.

"Map of the 41 stations sampled in the 29 lagoons located along the French Mediterranean coast (Gulf of Lion and Corsica). Each station's salinity type (according to their median salinity and Venice classification (Anonymous, 1958) in Le Fur et al. (2018)) and ..." [Lines 161 - 165]

Question 1.6. Lines 167-169: "and to large inter-lagoon..." - Is this meant to be "two"? or are there some words missing here? This part of the sentence is a bit confusing and difficult to follow, perhaps consider re-wording it.

We have reworded to be clearer:

"The different hydro-morphological characteristics of the studied lagoons lead to a great diversity of salinity regimes with 8 oligo- or meso-haline lagoons and 21 poly- or eu-haline lagoons (according to Le Fur et al., (2018), Figure 1) and to large discrepancies between lagoons in terms of annual variability of temperature, salinity and oxygen saturation." [Lines 166 - 169]

Discussion:

Question 1.7. Lines 650-654 "Overall, our results..." - I'm not sure if your tests and results actually showed the second part of this (primary colonization and/or post-disturbance recolonization of lagoons by marine-originating larvae through dispersal and recruitmen). If you think it did then you might have to elaborate on this, to show how you came to this conclusion.

Clearly not. We have reworded to be clearer.

"Overall and based on other published evidence, our results could be explained by (i) water renewal from marine origin, (ii) primary colonization and/or post-disturbance recolonization of lagoons by marine-originating larvae through dispersal and recruitment and (iii) environmental (in)stability, which are known to strongly shape lagoon benthic communities (Basset et al., 2006; Pérez-Ruzafa et al., 2007b; Pérez-Ruzafa and Marcos-Diego, 1992), as proposed in the confinement theory (Guelorget and Perthuisot, 1983)." [Lines 649 - 654]

Question 1.8. Line 656 - which communities are you referring to here? Do you mean all macrobenthic assemblages overall?

We are referring to our results. We have reworded to be clearer.

"Water temperature was another important driver of lagoon macrobenthic communities, SR and macrofauna density in our study. Temperature had a negative effect on these communities, probably directly through thermal stress, and indirectly through its effect on oxygen dynamics." [Lines 655 - 657]

Question 1.9. Paragraph from line 688. There is a bit of disconnect in this paragraph. What is the point you are trying to make with regards to the stations in La Palme? And then how does this part relate to the next part where you seem to be summarizing a key finding?

Our results for benthic communities on the La Palme lagoon (LPN and LPS stations) are poorly explained by environmental variables considered in the manuscript (see Figure 4). This could be the result of the strong variations in salinity recorded throughout the year, which we have not been considered here. We have movedthis paragraph after modifications to link it to the previous and the following paragraphs.

"The environmental variables considered in the manuscript do not unfortunately explain the communities described at the LPN and LPS stations. These two macrofauna stations sampled were actually classified in a group generally associated to small eutrophic lagoons and mainly composed of stations sampled in lagoons evaluated by the WFD as being in a poor or in a bad physico-chemical status even through is considered in a good physico-chemical status according to Chla concentration. This could be the result of the strong variations in salinity recorded throughout the year, which we have not been considered here. In this lagoon, marked annual salinity variations are observed (between 3.6 and 76.8 in Wilke and Boutiere (2000)), variations indetectable with the available summer salinity data. Consequently, the eight environmental variables selected to explain the overall macrofauna communities were not able to explain the one characterizing this lagoon. This result also reinforces the importance of natural environmental instability probably has on structuring benthic macrofauna communities." *[Lines 693 - 703]*

Question 1.10. Lines 751-753 - "as previously done for freshwater and marine organisms (Alonso and Camargo, 2006; Boardman et al., 753 2004; Camargo et al., 2005)" - it might be useful to briefly elaborate on what these studies actually showed or found, in this context.

We have added some additional information

"Overall, experimental work on a few well-selected species and populations sampled in lagoons presenting different levels of natural variability would greatly improve our understanding of the effect of eutrophication on lagoon benthic invertebrates via inorganic nitrogenous compound toxicity. Even trough marine animals appear to be less sensitive to nitrate than freshwater animals, nitrate toxicity to benthic invertebrate increases with increasing nitrate concentrations and exposure times (Alonso and Camargo, 2006; Boardman et al., 2004; Camargo et al., 2005) Nitrate toxicity may also increase with decreasing body size, water salinity, and environmental adaptation. In lagoon environments, where organisms are often smaller in size, this could also be of importance." [Lines 754 - 761]

Question 1.11. Line 755: What are the low and intermediate abundance taxa you are referring to here? And how did you classify these groups? I might have missed this, but I don't remember seeing details about this, and these groups feature in the paper. Something to consider since I couldn't really relate to this as I read this sentence.

We have added some additional information in the brackets.

"Furthermore, mean [Chla] appears to have a strong negative effect on macrobenthic communities by impacting the diversity of low and intermediate abundance taxa (i.e. SR and H' are diversity indices that give more weight respectively, to low abundance/rare and intermediate abundance taxa)..." [Lines 762 - 764]

Question 1.12. Line 761 - What exactly do you mean by this decimeter scale?

This is a mistake, we wanted to indicate "meter scale". We have modified the text.

"Despite not being the appropriate index to investigate beta diversity, the standard deviation of SR at the intra-station scale was significantly correlated negatively to mean [Chla], which seems to point towards a homogenization of benthic communities at the meter scale (i.e. a few meters separate each replicate of a given station) linked to eutrophication." [Lines 767 - 771]

Question 1.13. Line 798 - In this paragraph, is it worth discussing the possible influence of the reference values used in the calculations of this index? What do other studies use as reference values? Are there no national or regional standards and thresholds that could be used for the parameters used in this study? How could this possibly affect the index and the perceived response of the system?

We have added the reference values in the Methods section and complementary information in the Discussion section. This question is related to Question 2.2 of Reviewer 2 (see our answers below).

"We computed M-AMBI using the R function available in Sigovini et al. (2013) and we defined the good reference conditions as the highest values of SR and H' (46 and 3.73, respectively) and lowest value of AMBI (0.31) measured in the dataset, irrespective of the WFD physico-chemical status described above." [Lines 311 - 314]

"The M-AMBI (Muxika et al., 2007) is currently used to evaluate the WFD ecological state of poly- and eu-haline French Mediterranean lagoons based on soft sediment benthic invertebrate, with reference conditions using minimally impacted sites (Reizopolou et al., 2018)." [Lines 818 - 821]

"First, as recommended by Borja et al. (2012) who pointed that the inability of this index to detect anthropogenic responses is often linked to the use of inappropriate reference conditions, the currently used poly- and eu-haline vs oligo- and meso-haline categories (Provost et al., 2012; Reizopolou et al., 2018) could be replaced by one based on the three lagoon-sea connection levels that strongly influence the M-

AMBI or on another typology like the choked vs restricted typology (Barbone et al., 2012; Basset et al., 2013), with the test of new reference lagoons (Reizopolou et al., 2018).” *[Lines 853 – 859]*

Question 1.14. Lines 904-907 "Finally a functional approach..." - But didn't the previous sentence just contradict this recommendation? (regarding the limtations of using traits due to the plasticity) - perhaps this just needs to be re-worded.

We have reworded to be clearer.

“Finally, a functional approach based on well-informed biological traits and biomass data could help overcome the identified taxonomic limitations and build a more holistic framework linking natural variability and eutrophication of coastal lagoons to macrobenthic communities but the intra-specific plasticity for many biological traits makes this approach questionable today.” [Lines 932 – 936]

Abstract:

Question 1.15. I'm just confused by this sentence of the abstract: "Conversely, AMBI was the only tested index that uniquely responded to eutrophication variables, which nonetheless explained less than a third of its variability" - But I thought the results from this paper showed that AMBI was not very responsive, and didn't show any significant relationship with [Chla], among other variables? And AMBI (only M-AMBI) is not included in Figure 5 which represents the variance partitioning.

This sentence is indeed ambiguous. AMBI was only correlated (using multiple linear models) to NOx (one of the eutrophication variables) but not to chlorophyll *a* concentration, the common eutrophication indicator used in our study. We have therefore removed this unessential sentence of the abstract.

Reviewer #2

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)

Question 2.1. 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.

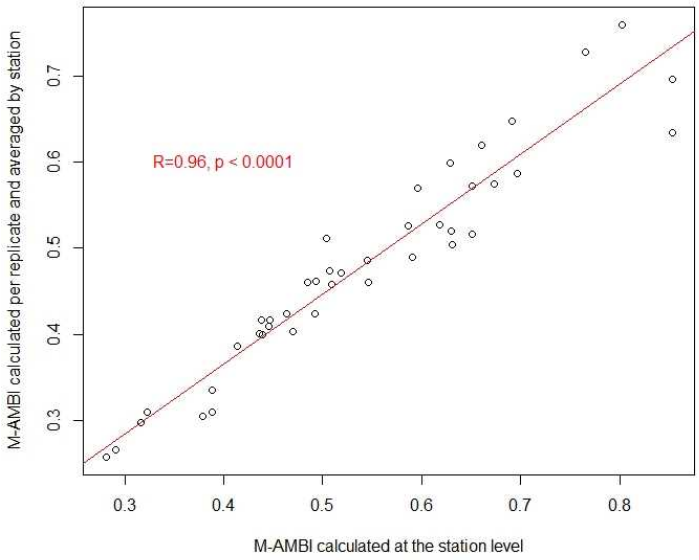
We used the exact script created by Sigovini et al. (2013) considering each replicate as a sample and without specifying the st= parameter to pool replicates by station. Additional indices were also computed at the replicate level. The mean and standard deviation were then computed for both indices.

It is true that we are therefore not perfectly calculate the M-AMBI as recommended in the WFD. This enabled us to compare directly our results with results from others countries such as been done by Intercomparison Group mentioned in Question 2.2. In our case, this means considering each replicate as an independent station and thus providing a mean value for species richness, Shannon index and AMBI, together with a variability indicator, as shown in Figure 4.

The comparison of the results obtained by calculating the M-AMBI at station level or by averaging the replicates reveals a small difference (on average 0.06) which follows exactly the same trend. We have illustrated this link in the figure below, where the Pearson correlation coefficient is 0.96 ($p<0.0001$). This does not alter our hypotheses and conclusions. We have therefore decide to retain our analyses. The sentence has been removed and replaced in the Methods section to be true.

"To characterize the macrobenthic community structure, we calculated the commonly used M-AMBI (Multivariate AZTI Marine Biotic Index), the three indices composing the M-AMBI - species richness (SR), Shannon diversity index (H') and AMBI (Muxika et al., 2007) - plus the inverse of Simpson's index (N2) and the total macrofauna density. The AMBI is an index based on the classification of benthic species into ecological groups, from the most sensitive to disturbances (i.e. organic matter enrichment) to the most opportunistic ones (Borja et al., 2000)." [Lines 303 - 308]

"All indices were computed at the replicate level ; the mean and standard deviation (SD) values were then calculated for each station." [Lines 314 - 315]



Question 2.2. 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.

This question is related to Question 1.13 of Reviewer 1. Reference conditions are highly influential parameter in assessing marine ecosystem quality. In our paper, we defined the reference conditions using the default settings of highest species richness (SR=46), Shannon-Wiener entropy (H'=3.73), and lowest AMBI (AMBI=0.31), irrespective of the WFD physical-chemical status. We had previously run the M-AMBI analyses using the reference conditions established by (i) the French government (SR=46, H'=4.23 and AMBI=0.60) and using two different reference sites (ii) Thau lagoon (SR=27, H'=3.73 and AMBI=0.31) or (iii) Leucate lagoon (SR=46, H'=3.66 and AMBI=2.11). We have compared these 3 options with the default reference conditions used in the paper and M-AMBI values were highly correlated (between 0.99 and 1).

We have added the reference values used in the Methods section and complementary information in the Discussion section:

"We computed M-AMBI using the R function available in Sigovini et al. (2013) and we defined the good reference conditions as the highest values of SR and H' (46 and 3.73, respectively) and lowest value of AMBI (0.31) measured in the dataset, irrespective of the WFD physico-chemical status described above." [Lines 311 - 314]

"The M-AMBI (Muxika et al., 2007) is currently used to evaluate the WFD ecological state of poly- and eu-haline French Mediterranean lagoons based on soft sediment benthic invertebrate, with reference conditions using minimally impacted sites (Reizopolou et al., 2018)." [Lines 818 - 821]

"First, as recommended by Borja et al. (2012) who pointed that the inability of this index to detect anthropogenic responses is often linked to the use of inappropriate reference conditions, the currently used poly- and eu-haline vs oligo- and meso-haline categories (Provost et al., 2012; Reizopolou et al., 2018) could be replaced by one based on the three lagoon-sea connection levels that strongly influence the M-AMBI or on another typology like the choked vs restricted typology (Barbone et al., 2012; Basset et al., 2013), with the test of new reference lagoons (Reizopolou et al., 2018)." [Lines 853 – 859]

Question 2.3. 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.

Sediment grain-size, heavy metal concentrations or hypoxia are indeed key parameters in the distribution of benthic communities of soft bottom. Unfortunately, these data are not available on the 29 coastal lagoons studied. We already mentioned a few elements and we have addressed new limitations in the Discussion.

" Disentangling the effects of anthropogenic stressors and natural environmental variables remains nevertheless challenging, especially as stressors and variables can co-vary, even in the absence of

anthropogenic impact. Moreover, it is often impossible to measure effects of every single possible stressor on macrobenthic communites. In our study, grain-size could not be included to differences in laboratory protocols between lagoons and between years and heavy metal contaminations were not available at the same temporal and spatial scale (one measurement per lagoon every ten years). It should be relevant to re-run the analyses when such data would be available at the macrofauna sampling sites to test the effect of this supplementary anthropogenic driver to check if it would be discriminating benthic communities." [Lines 806 – 814]

"Indeed, sediment contaminants like biotic interactions are likely to affect benthic macrofauna at fine spatial scales (ca. 1 km) that are more relevant to low mobility macroinvertebrates (Berthelsen et al., 2018; Brauko et al., 2020, 2015; Carvalho et al., 2006)." [Lines 881 – 883]

"More generally, it appears key to increase our understanding of the effects of other disturbances like hypoxia and heavy metals on lagoon invertebrates and to reevaluate the small-scale distribution of benthic organisms in French coastal lagoons, more than 30 years after the historical studies by Amanieu et al. (1977), Guelorget et al. (1994) and Guelorget and Michel (1979)." [Lines 886 – 889]

Question 2.4. 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.

Understanding and parsing out the effects of natural variability and anthropogenic pressures on biotic indices is paramount to developing robust ecological monitoring programs. We could indeed include a section focusing uniquely on the exhaustive comparison of our results with the literature. Instead, we have chosen to compare each of these points throughout the discussion, citing several of the references proposed here. This has enabled us to focus on the underlying mechanisms and less on the problems associated with indicator development, which is not the focus of this work. In response to this comment, we have included recent references in this new version to give the reader as much information as possible on the subject.

"This index has previously been shown to be sensitive to pressures related to eutrophication, like concentration in chlorophyll a and total nitrogen (Derolez et al., 2014) but it is also known to be sensitive to natural environmental variability (e.g. water salinity) and lagoon hydro-morphology (e.g. surface and depth) (Barbone et al., 2012)." [Lines 821 – 824]

"Our study reinforce the difficulty of M-AMBI to clearly identify natural disturbance but also the high sensitivity of this multimetric index to anthropogenic pressures such as eutrophication or metal contaminations and pesticides in transitional waters (Paul et al., 2023; Pelletier and Charpentier, 2023; Pollice et al., 2015). The strong joint effect is probably related, as previously discussed, to the same interconnections between lagoon-sea connection that buffers eutrophication, temperature that promotes phytoplankton blooms and eutrophication that homogenizes benthic habitats." [Lines 847 – 853]

"First, as recommended by Borja et al. (2012) who pointed that the inability of this index to detect anthropogenic responses is often linked to the use of inappropriate reference conditions, the currently used poly- and eu-haline vs oligo- and meso-haline categories (Provost et al., 2012; Reizopolou et al., 2018) could be replaced by one based on the three lagoon-sea connection levels that strongly influence the M-AMBI or on another typology like the choked vs restricted typology (Barbone et al., 2012; Basset et al., 2013), with the test of new reference lagoons (Reizopolou et al., 2018)." [Lines 853 – 859]

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

Question 2.5. 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.

We have added some additional information on the calculation of taxonomic indices.

" The raw (i.e. untransformed) macrofauna abundances were used to calculated all these indices." [Lines 308 - 309]

Question 2.6. 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.

In our dataset, a Pearson correlation of |0.6| corresponds to a p-value < 5. 10^{-5} and was chosen as a reasonable compromise for removing the most correlated variables from the following analyses.

"To limit multicollinearity and avoid artificially inflating the fit of linear models, we reduced the number of environmental variables (see section Environmental variables) for all future analyses by considering Pearson correlations < |0.6|, corresponding to a p-value < 5. 10-5 and leading to 19 remaining environmental variables." [Lines 331 - 334]

Question 2.7. 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.

The variance inflation factor is a measure to analyse the magnitude of multicollinearity of model terms. A VIF less than 5 indicates a low correlation of that predictor with other predictors. A value between 5 and 10 indicates a moderate correlation, while VIF values larger than 10 are a sign for high, not tolerable correlation of model predictors (James et al. 2013). We thus removed variables with a VIF >5. We added the reference in the manuscript to justify our choice.

"We computed the variance inflation factor (VIF, 'vif.cca' function) to check for multicollinearity and sequentially removed variables with a VIF > 5 (James et al., 2013)" [Lines 337 - 339]

James, G., Witten, D., Hastie, T., and Tibshirani, R. (eds.). (2013). An introduction to statistical learning: with applications in R. New York: Springer.

Question 2.8. 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.

Clearly not. We have reworded to be clearer.

"Overall and based on other published evidence, our results could be explained by (i) water renewal from marine origin, (ii) primary colonization and/or post-disturbance recolonization of lagoons by marine-originating larvae through dispersal and recruitment..." [Lines 649 - 654]

Question 2.9. Lines 836-838: It is unclear how the replacement of sampling by Ekman-Birge grabs with diveroperated 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.

We have reworded and added some information directly in the text to be clearer.

"Secondly, the replacement of a sampling done with an Ekman-Birge grab by a diver-operated sampling would limit heterogeneity between replicats by (i) verifying that only unvegetated sediments are being collected and (ii) describing the presence and type of macrophyte near the sampling site (e.g. 10 meter radius). This could represent a bias in many homogeneous environments where random selection of the sampling zone is quite possible but macrophytes are often heterogeneously distributed in coastal lagoon and areas of unvegetated sediments are rare. Selection and description of the sampling area by divers could therefore be relevant in these shallow environments even if this often means additional human and financial costs." [Lines 859 - 866]

Question 2.10. ambi&M-AMBI.R Line 117: As noted in the "Essential amendment to the R script in theElectronic 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.

This correction has been done

Question 2.11. ambi&M-AMBI.R Lines 123-130: 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.

This correction has been done

Reviewer #3

This manuscript attempts to use macrobenthic invertebrates and environmental variables to disentangle natural and anthropogenic drivers of coastal lagoons across 29 systems in the Mediterranean region of France. General results include that lagoon hydro-morphology (connection to sea and lagoon surface which alter salinity and temperature), as well as habitat diversity (macrophyte morphotypes) being the primary drivers of macrofauna distribution. Anthropogenic stressors such as eutrophication resulting in hypoxia are superimposed on this natural instability altering species richness and diversity. Using these findings, the sensitivity of M-AMBI was assessed and it was found to be more sensitive to eutrophication than natural variability. Subsequent suggestions were made to make M-AMBI more applicable to coastal lagoons.

While the results are interesting, there are some concerns regarding inconsistencies and gaps in the data collated. They include but are not limited to:

Question 3.1. - Line 191-194: Each estuary was only sampled once, with three replicates per station. Some larger estuaries had southern and northern stations. In addition, sampling occurred 3 years apart for some systems e.g. 2006 and 2009. This raises some concerns as this is sufficient time for significant changes to have occurred in macrobenthic communities as a result of other influences such as natural climatic variability.

Our dataset did not allow us to discuss the temporal variability of benthic community structure, which could indeed differ over a 3-year period as result of natural or anthropogenic variabilities. The strength of our dataset lies in the fact that it considers a wide environmental gradient along 21 coastal lagoons for which we have biological data combined with environmental data over the same period. As specified in the manuscript, benthic community structure and environmental conditions are characterized only once, using the same method, in 2006 or 2009 depending on the station. We therefore focused our analyses on the link between benthic community structure and the lagoon environment. We have reworded to be clearer.

"The sampling was carried out once a time in 2006 or 2009 (depending on the station) during the "Réseau de Suivi Lagunaire" and WFD campaigns." [Lines 150 - 151]

Question 3.2. - Line 209-214: It is uncertain whether the estimation of the Organic Matter Content for some sites is correct as these areas are at times between 285m and 707m from the actual sampling sites in a habitat known for its environmental variability.

This problem concerns only 5 stations (on 41 stations in total). To fill in these gaps, we have chosen to allocate the spatially closest data (or mean on several stations) to the monitoring station. This solution may not be perfect, but it ensures that these stations are not excluded from the analysis. This method is currently

applied in many large-scale spatial datasets where geospatial statistics are used to estimate data usually not available at the same spatial resolution as benthic macrofauna.

Question 3.3. - Line 214-217: Sediment particle size has been omitted from the analysis due to changing laboratory protocols. This is a concern as grain size is widely agreed to be a primary driver of the macrobenthos.

This question is related to Question 2.3 of Reviewer 2 (see our answers above). We were unable to include grain size in our analysis even though it is indeed a key parameter in the distribution of benthic communities of soft bottom, in general. We have grain size data but they are unfortunately not uniform across the whole dataset, in particular for the estimation of the percentage of mud. The analysis of data in each of the French lagoons as part of a habitat mapping project (Miramont et al., 2023) illustrates that the grain size is relatively homogeneous in these environments (mud percentage between 10 and 50%). We nevertheless preferred not to include this parameter to avoid adding a degree of variability in the distribution of benthic communities that would have been difficult to explain.

That being noted, this study is intended to be a widescale analysis to identify overarching trends. In these cases, it is extremely difficult to account for all discrepancies, and the authors have clearly described their study and been transparent with the data limitations, as such, I feel that the manuscript has some very interesting findings and should be published as is.

Thank you for yours questions and this remark.

References cited/added

Anonymous, 1958. The Venice System for the classification of marine waters according to salinity. Limnology and Oceanography 3: 346–347.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (eds.). (2013). An introduction to statistical learning: with applications in R. New York: Springer.

Miramont Arthur, Menu Marion, Vaz Sandrine, Ouisse Vincent (2023). Carte du Substrat des lagunes méditerranéennes françaises listées au titre des masses d'eau de transition de la Directive Cadre sur Eau : critère 2 de la typologie d'habitats lagunaires du projet CHAMILA. IFREMER. https://doi.org/10.12770/b887bc5c-0114-4f13-8624-4c47c095abbe

Paul, J., Nestlerode, J.A., Jarvis, B.M., 2023. Evaluating the effectiveness of M-AMBI with other biotic indexes in a temperate estuary. Marine Pollution Bulletin, 193: 115194. https://doi.org/10.1016/j.marpolbul.2023.115194

Pelletier, M.C., Charpentier, M., 2023. Assessing relative importance of stressors to the benthic index, M-AMBI: An example from U.S. estuaries. Marine Pollution Bulletin, 186:114456. https://doi.org/10.1016/j.marpolbul.2022.114456

Pollice, A., Arima, S., Lasinio G.J., Basset, A., Rosati, I., 2015. Bayesian analysis of three indices for lagoons ecological status evaluation. Stochastic Environmental Research and Risk Assessment, 29: 477-485. DOI: 10.1007/s00477-014-0885-4