Dear Editor,

I am submitting this rebuttal in response to the reviewers' comments on our manuscript titled "Spatial and Temporal Representation of Marine Fish Occurrences," which is available online.

We are grateful for the opportunity to address the observations raised by the reviewers regarding our contribution. We have diligently incorporated their feedback, and it has undoubtedly contributed to enhancing the quality and robustness of our manuscript.

Sincerely,

Horacio Samaniego

Universidad Austral de Chile

horacio@ecoinformatica.cl

Reviewer #1: The authors have made changes to the MS, however many of these are cursory in nature. There remain many issues with the work that need to be corrected, some of which I did not identify in my previous reading of the MS.
The English has not been improved and is not of publication standard. I list some of the issues below but these are a small subset. The authors might consider a professional editing service if they are unable to secure help otherwise.
Below I provide my comments on the author responses, followed by general and specific comments on this version of the MS.

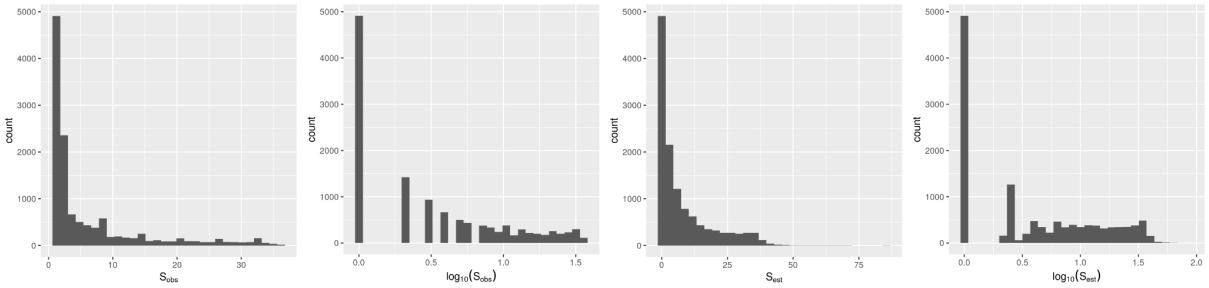
== Responses ==

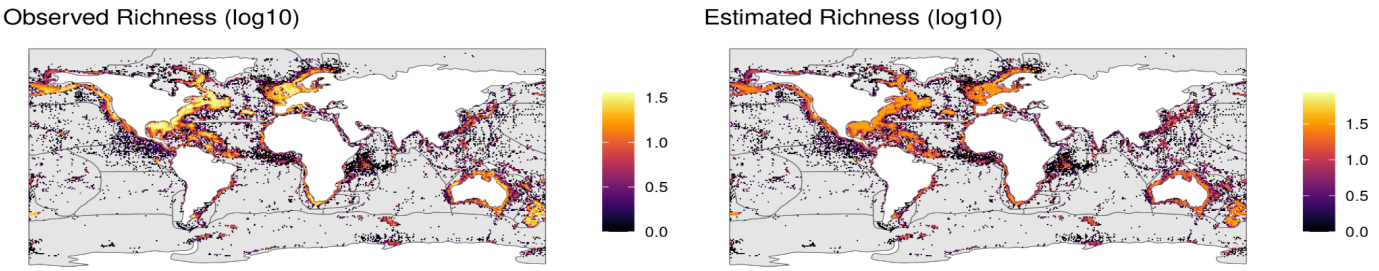
1. The SRI index is still problematic and the authors have not addressed my concerns. Any ratio will suffer from the small denominator problem. 2/4 has almost the same bias as 1/2. Consider instead a bivariate colour scheme to show both observed and expected richness, possibly log scaled so more species-rich cells do not dominate the shading.

We appreciate the comment.
We have created a new figure in the appendix that illustrates the spatial distribution of the Species Richness Index (SRI) alongside the original observed species richness and estimated species richness values utilized in SRI calculations. This figure is intended to provide readers with a clear understanding of the disparities in magnitude between observed and estimated species richness (S).

Nonetheless, we maintain our classification, as we believe it offers a more accurate representation of the relative knowledge of species occurrence numbers and contributes to maintaining the manuscript's focus. It is important to underscore that while the ratios of 12/24th and ½ do indeed signify a considerable disparity in the number of missing species, the primary objective of this study is not to provide an exhaustive enumeration of species absent from the database records. Rather, our aim is to furnish a macro-level description of the most substantial differences between observed and estimated species richness, facilitated by the Species Richness Index (SRI). We value the suggestion to include the new Figure A.1, as it will enable readers to gauge the extent of this difference should they wish to explore it further.

Lastly, we regret to admit that we are encountering some difficulty in comprehending the final suggestion concerning the modification of the color scheme. The mapping of observed and estimated richness is accomplished through the Species Richness Index (SRI), which represents the ratio of these two richness values. We attempted the use of log10(S), but regrettably, it did not result in an improvement in the shading. To provide further clarity, we have included histograms and log maps for your review and assessment.





2. Taking an average of the richness estimators itself has issues. What is the variability of the different estimators? Is it skewed? Why not use a single richness estimation metric that provides confidence intervals?

The literature has shown evidence that richness estimators differ in their performance depending on the type of data, sampling design and scale of analysis (see Foggo *et al.,* 2003, for instance). Additionally, most of the studies analyzed by Walther *et al.,* (2005) show that nonparametric estimators (e.g. Chao and Jackknife estimators) perform better than other methods. Considering this, we have chosen to follow what was done in Mora *et al.,* (2008), i.e. generating accumulation curves through different methods and averaging these results to obtain a "smoothed" curve (based on the review of biostatistical methods in Gotelli & Conwell, 2001).

We have chosen this procedure as it integrates a wide variety of information coming from these different estimators.

References:

Foggo, A., Attrill, M. J., Frost, M. T., & Rowden, A. A. (2003). Estimating marine species richness: an evaluation of six extrapolative techniques. *Marine Ecology Progress Series*, *248*, 15-26. DOI: 10.3354/meps248015

Gotelli, N. J., & Colwell, R. K. (2001). Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology letters*, *4*(4), 379-391. DOI:10.1046/j.1461-0248.2001.00230.x

Walther, B. A., & Moore, J. L. (2005). The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. *Ecography*, *28*(6), 815-829. DOI: 10.1111/j.2005.0906-7590.04112.x

3. Remove the formula for the mean. It is entirely redundant in a publication such as this.
The equation has now been removed.

4. The argument for the classes remains unconvincing. These boundaries are entirely arbitrary and potentially misleading.

Although these classes may appear to be arbitrarily assigned, they offer a clear and uncomplicated means of categorizing SRI. We selected these values after a thorough evaluation of the SRI distribution, believing that they provide a balanced approach for qualitatively assessing the representativeness of database records. We appreciate your comment, as it highlighted the need for clarification, which was not explicitly addressed in earlier versions of the manuscript. Consequently, we have now included (i) an explanation of the methods used to define these classes and (ii) have included the histogram in the appendix to further illustrate this.

5. In regards to the responses to reviewer 2, I agree that cartograms are unnecessary.

== General comments ==

6. GBIF includes OBIS.
https://www.gbif.org/network/2b7c7b4f-4d4f-40d3-94de-c28b6fa054a6
Is direct use of OBIS therefore necessary? Use of both could lead to double counting of records, although hopefully the removal of duplicate coordinates will avoid this. If there are additional records in OBIS due to delayed updates or other such issues then please quantify this.
Indeed OBIS is among the most widely used sources of marine life data in the GBIF[1]. However, since both platforms include different data sources and methodologies, as well as important differences in temporal and spatial scales at which data were collected, several studies suggest

[1] https://obis.org/usecases/*(

examining both platforms with care (Ziska *et al.,* 2020). Recent studies have also shown unequal contributions between the two repositories, with very low percentages of shared data (Moudrý & Devillers, 2020; Chollett & Robertson, 2020).
While we only work with unique data and remove any duplicates, we do not quantify the difference in data between the two platforms as this has not been the focus of this study. We have therefore placed any reference to this information in the supplementary material.

References:

Moudrý, V., & Devillers, R. (2020). Quality and usability challenges of global marine biodiversity databases: An example for marine mammal data. *Ecological Informatics*, *56*, 101051. DOI: https://doi.org/10.1016/j.ecoinf.2020.101051

Chollett, I., & Robertson, D. R. (2020). Comparing biodiversity databases: Greater Caribbean reef fishes as a case study. *Fish and Fisheries*, *21*(6), 1195-1212. DOI: https://doi.org/10.1111/faf.12497

7. An 80 cm interval for the analysis of fish sizes is extremely large, to the point where it is perhaps larger than the median fish size. Is this size correct? Did the authors actually use 80 mm? If 80 cm is the unit used for the analyses then they need to be redone.
Thanks for the acute observation. This is much appreciated.
We rechecked the analysis and figure 5 was reformatted.The analysis is in cm units indeed. Bodysize is normally distributed after a log10 transformation. We now report the size interval for the range between the 1st and 3rd quartile. This is all included in section 3.5.1 of the Results.

8. The authors refer to the hexagonal lattice as being of one degree resolution. However, this does not make sense when using an equal area coordinate system. It is better to use 10,000 km^2 as this is a reasonable approximation of the area of a one degree square cell at the equator that is also easily understood. If the authors have used an approximation of a one degree hexagon at the equator to determine the size then this needs to be explicitly stated. Alternately look into the H3 system. https://h3geo.org/
Thanks for this observation we have indeed performed the analysis gridding the globe in 10^5 km^2 and have now added the cellsize units (square-kilometers) to avoid confusion. It seems that it was not entirely clear that the cartography on which we worked (bioregions of Costello et al., 2017) was projected to cylindrical equal area. This uses the WGS84 coordinate system making each cell equal area at a global scale. This is now explicit in the text.

9. Why use the last 10% of the accumulation curves? Why not use the same period of years across all regions so the values are comparable? Although given the data were subset to 1980 and later, why not just state the number of years?
We used the last 4 years of data input (i.e. 2016-2020) which corresponds to the last 10% of the species accumulation curve that was constructed for each bioregion. We have included this description in the Methodology, section 2.2.2.

10. Hypotheses are given in the introduction but there are no associated statistical tests. What are provided are largely summary statistics. I suggest that these be reframed as research questions to avoid the term "hypothesis".
Thanks for your comment. We corrected it.

== Specific comments ==

11. P1, L20. "This may have dreaded and detrimental effect". The English expression is incorrect. And "dreaded" is not the ideal word choice here.
Thanks for your comment. We corrected it.

12. P2, L3. "overlapped on the marine bioregions worldwide."
We have now rephrased this paragraph.

13. P2, L30. "discovered revealing" -> "discovered, revealing" (otherwise it reads that they have been discovered to be revealing).
Off course! Thanks very much for your comment. We corrected it.

14. P3, L2. "considers" -> "include".
Thanks for your comment. We corrected it.

15. P3, L6. The argument in this paragraph is not well presented. It needs to be rewritten.
Thanks for the opportunity to sharpen the wording of this paragraph. We have rewritten this.

16. P3, L6. "species richness is often used to represents".
Done, we have taken care of this in the previous comments

17. P3, L13. The issue with "hands-on" I raised in my previous review has not been corrected. Replace the term as it is inappropriate.
Thanks for your comment. We corrected it.

18. P6, L18. "wrangling" is conversational. It is better to refer to "manipulation".
Thanks for your comment. We corrected it.

19. P7, L6. It is clearer to refer to a cylindrical equal area projection using the WGS1984 datum.The cited reference is not appropriate as it is not the reference definition, merely an analysis where this projection was also used.
And which of https://epsg.io/53034 and https://epsg.io/54034 was used?
Please also note that this projection is not recommended for global extent analyses so it should not be used in any case.
https://pro.arcgis.com/en/pro-app/latest/help/mapping/properties/cylindrical-equal-area.htm
A better coordinate system to use is Mollweide or similar. https://epsg.io/53009
We have now replaced the reference with the specific EPSG code employed in this study https://epsg.io/54034
It's important to emphasize that the Mollweide projection is well-suited for representing global data, but like any projection, it introduces undesired distortions, particularly at higher latitudes (for more information, see: https://desktop.arcgis.com/es/arcmap/latest/map/projections/mollweide.htm).

20. P7, L24. Sentence is awkwardly phrased.
Thanks for your comment. We rewrote the paragraph.

21. S2.2.2. This is not clearly described.
Thanks for your comment. We rewrote the paragraph.

22. P9, L1. An 80cm interval is extremely large. Is this correct? Should it instead be 80 mm? If 80 cm is the unit used for the analyses then this needs to be redone.
Redone. Please see previous point 7. The size is in cm and this info is now in the text. We have redonde the analysis. In order to avoid the use of arbitrary intervals, we have changed these plots to histograms using approximately 30 bins.

23. Fig 1. Why have the values been normalised? Why not show the actual values in legends? There is space if the figure is reorganised. (The panels are too small at the moment in any case).
The primary objective of Figure 1 is to visually depict the marine bioregions to the readers, as originally outlined by Costello et al. (2017). Furthermore, our intention is to offer visual reinforcement for the data already expounded in Table 1. While Table 1 contains the precise species diversity and richness values for each bioregion, Figure 1 serves a schematic purpose, providing an illustrative representation of this data. These are the reasons to provide normalized values.

24. P11. "scale resolution" and "resolution scales" -> just refer to resolution.
Thanks for your comment. We corrected it.

25. P11, L13. This edit is incorrect. Please revert it.
Thanks for your comment. We corrected it.

26. Fig 2. Remove the hexagon outlines. They obscure the patterns for cells containing data.

Thanks, we have now removed the hexagon outlines.

27. Fig 3. Y-axis label is misspelt.
Thanks for your comment. We corrected it.

28. Fig 5. Remove the linear correlation coefficient. Use a non-linear function if one is to be fitted since these are log-linear relationships. The caption should not refer to the dotted line since it has been removed.
Thanks for your comment. We corrected it.

29. P18, L4. "We used hexagonal grids that fit the geographic reality of marine ecosystems". This is not a good explanation. Hexagons are a tessellation that better fit the spheroid than do square grid cells.
Thanks for your comment. We rewrote the sentence to make it clearer to the reader.

30. P19. This text is largely meaningless: "Regarding the units of analysis, here we estimate the species richness at the grid level in order to obtain more uniform results on the distribution of occurrences and avoid overestimating the SRI for marine bioregions"
We have rewritten this paragraph retaining only essential information. Thanks.

31. P19, L13. "we recommend using grids that really allow observing macro-ecological patterns". This is hardly a novel recommendation given it is basic common sense. It is also poorly phrased.
Removed.

32. P19, L19. "GBIG"
Thanks for your comment. We corrected it.

33. P19, L22. Does this value correct for OBIS records that are also in the GBIF database?
Yes it does.

34. P21, L1. sub-soil?
We have removed such term and rephrase this paragraph.

35. P20, L7. Diametric?
Thanks for your comment. We removed this misplaced word.

Reviewer #2: All of my suggestions were strictly taken into account. Also the analysis I suggest as a test were seriously performed. I agree with the cartograms point.

We sincerely appreciate your valuable feedback. We have diligently endeavored to incorporate all of the provided suggestions and are confident that these revisions have enhanced the overall quality and strength of the manuscript.

36. Is table 1 needed inside the paper or could it be put as an appendix?
We consider it imperative to maintain this table as we refer to it consistently throughout the progression of the paper. Its removal would necessitate substantial modifications to Figure 1.