Reply to Paulo Borges
Sources of confusion in global biodiversity trends

General comments from Paulo Borges

The organization and structure of the manuscript are logical and coherent, making it easy to follow the authors' line of reasoning. The writing style is clear, concise, and technically sound, with appropriate use of terminology.

This manuscript underwent evaluation by two reviewers who provided significant suggestions for improvement. A agree with both reviewers that there is the need of some additional work.

Please perform the additional analyses suggested by both referees.

Dear Paulo Borges,

First of all, we would like to thank you and both reviewers for your time to assess our manuscript and your constructive comments. We appreciate that you felt it is a timely contribution and that you enjoyed the reading, which confirms the positive reactions we had after posting the preprint online.
We've addressed all the concerns raised in the revised version. We now acknowledge that clarifications were needed especially regarding the aim of the review and the choices we made for the reviewing process and the categorization of the conclusions. We provide below a point-by-point response to the comments.

Just a note: we would like to make sure that our aim and methodology is not misinterpreted. It appeared to us that the reviewers understood our work as a formal systematic review. However, we did not perform (and did not aim to perform) a formal systematic review as defined in the *Cochrane Handbook for Systematic Reviews of Interventions*, 2023. Systematic reviews follow a specific methodology and aim which differ from our approach. In particular, contrary to systematic review we are not trying to generate a quantitative estimate of biodiversity change from a meta-analysis. Rather, we perform a literature review (as routinely found in most journals) to identify potential sources of confusions among the global biodiversity changes assessments. This is now clearly explained in the revision, that is available on the preprint server (https://doi.org/10.32942/X29W3H).

That being specified, we hope the revised manuscript will be found suitable for publication.

On behalf of all authors,
Maelys Boennec

Reviewer #1:

In their manuscript, Boennec et al investigate the potential sources of variability in global biodiversity trend estimations. Since identifying these sources, or controlling them in future estimates, is key for understanding trends underpinning conservation actions and policy decisions, the importance of this work is undeniable. The manuscript is well-written and logically structured and it also scrutinises ample information to estimate factors potentially distorting biodiversity trend estimates.

We thank the reviewer for these positive comments.

Major comments

(1) *Only Web of Science was used to gather related literature. This not only raises the problem of an incomplete dataset, which the Authors discussed, but it also can lead to substantial biases and a non-replicable data collection. The Cochrane handbook for systematic reviews of interventions (Higgins and Cochrane Collaboration 2020) clearly advices for querying multiple databases to avoid biases and Pozsgai et al. (2021) reported that particularly Web of Science is unsuitable for reproducible searches because its result set (even if only Core Collection is used) depends on the subscription of the institution the search was initiated from. This, to a certain level, can be mitigated by reporting the exact date and institutional background of the search(es). However, other databases (such as Scopus), from this respect, may be more suitable for similar purposes.*

(R1.1) We fully acknowledge the limitations of our search strategy, and we have carefully considered the suggestions made to enhance the rigor and comprehensiveness of our study. As suggested, we now have specified the date and institution background in the manuscript (line 130-131).

Regarding the database selection, we acknowledge the importance of utilizing multiple databases for a comprehensive literature review. In our study, we focused on Web of Science (WoS) due to its well-established reputation and comprehensiveness.

As suggested, we launched an equivalent query on Scopus, using the following request:

TITLE (biodiversity OR population OR communit* OR indicator OR natur* OR richness OR species OR (biological AND diversity) OR abundance OR assemblage OR *flora OR *fauna) AND TITLE (trend OR dynamic OR (time AND series) OR declin* OR loss OR extinct* OR increas* OR gain OR coloni* OR change OR fluctuat* OR trajectory) AND ABS (temporal OR time) AND ABS (analys* OR model* OR stud* OR quantifi*) AND NOT TITLE ((human AND population) OR (urban AND population)) AND (TITLE (global OR worldwide) OR ABS (global OR worldwide)) AND PUBYEAR < 2022 AND (LIMIT-TO (SUBJAREA,"AGRI") OR LIMIT-TO (SUBJAREA,"ENVI")).*

For comparison, we queried papers published before 2022. While we found 2,008 matches in Web of Science, we found 9,094 matches in Scopus. This difference might seem huge but it has been previously reported (Chadegani et al. 2013) as Scopus is based on a wider range of databases and non-indexed journals including a lot of grey literature. Still, Scopus did not recover all the papers we identified with WoS. Only 1,727 papers were common to both WoS and Scopus (i.e. 86 % of the WoS corpus). Therefore, by using only Web of Science, we may not have captured all possible sources matching a query performed in Scopus, but we also were able to retrieve 281 in WoS but not is Scopus. Moreover, WoS is anyway covering all journals relevant to the field in ecology, macroecology or conservation, which are similar in Scopus and WoS (the top twenty journals in Ecology listed by the *Observatory of International Research* in particular are covered by WoS and

Scopus). In that sense, it is most likely that relevant studies to our review will be mostly found in those topic journals. But above all, our goal is not to be quantitatively exhaustive (we did not conduct a systematic review) but to cover a representative set of the datasets, methods and approaches used in global biodiversity studies. In fact, a systematic review and comprehensive meta-analysis would have been beyond the scope of this study. Additionnally, working with a too large database from the beginning, mostly dominated by non-relevant studies would have led to non-reproducible results.

In the revision, we acknowledge the potential for undiscovered literature in other databases as we have not conducted a systematic review, and argue that our conclusions are most probably robust as they are based on most key papers in the field (lines 377-378; 381-382).

> **(2)** *Although the authors presented numerical summaries, a formal analysis is badly missing. Claims, such as "The time span did influence the conclusions" in line 290 are not statistically tested, we only see the differences between numbers – whether these differences are significant or not cannot be told. However, I believe, machine learning (e.g. GBM or random forest) and/or factor mining approaches would be fully suitable for this type of analysis.*

(R1.2) We agree that statistical tests were missing, even though we were not expecting any strong statistical argument regarding the number of papers considered for each single test. However, the specific tests proposed by the reviewer are not the most appropriate in our opinion, as we only have qualitative variables, and also because we lack the power to perform the advanced statistical analysis proposed. The most appropriate analysis in our opinion was to perform chi square tests in order to test whether the conclusions drawn depended on either (1) the assessment approach, (2) the dataset used, (3) the number of taxonomic groups (split in qualitative categories), (4) the time span (split in qualitative categories), (5) the ecological level, and (6) either the methods or the identity of the studied drivers. We performed these tests both for the trends and the drivers assessments.

> **(3)** *When evaluating trends, splitting the trends into decreasing, increasing and stagnating, and mixed is too coarse – mixed can mean too many different trends. I suggest splitting the mixed to three further categories as "mostly decreasing", "mostly declining", and "equal". Moreover, merging stable and increasing trends may also be confusing.*

(R1.3) We acknowledge that the definitions and thresholds we used were lacking clarity and we clarify them now in our revision. Yet, note that our aim was to classify whether the conclusion of a paper falls into identifying a global decline or having mixed trends. These are the two most common conclusions for the global biodiversity crisis at the moment. What is considered as "mixed trend" in our paper is what the authors of the papers we reviewed concluded themselves. It is not our interpreted classification. Therefore, dividing the « mixed » results into further categories, and having for instance a « mostly declining trends » category would mean to reinterpret the conclusions made by the reviewed studies. We didn't want to do that but instead we think it's important to specifically demonstrate that some studies provide results that do not allow any clear directional conclusions.

We agree though that the thresholds we have used to consider if the results were falling into each of the categories were not clear enough. Papers classified as « declining trends » were those concluding declines for higher than 50 % of the species or populations considered. We double-checked our conclusions by going through again all the papers classified as « mixed trends ». We classified the papers as « mixed » when studies presented as many increasing as decreasing trends and/or a majority of no trends (as expressed line 175). For instance, Daskalova et al. (2020) identify

15 % of decline, 18% of increase and 67 % of no net changes. Finally, we only merged the papers concluding « no effect of the drivers » with the papers concluding « positive effects of the drivers » together as they represent only three papers. Stable (if understood here as “mixed”) and increasing biodiversity trends were not merged at all.
All information we extracted along with the source papers is still available on the metadata supplement file. In this file we also include our categorization along with the detailed conclusions as made by each paper.

Minor comments

(4) *Figure 2C repeats info from the text*

(R1.4) We deleted figure 2C.

(5) *Alluvial plots look nice but, in my opinion, Figure 4 is too busy for this kind of visualisation. An online available interactive version of the same plot, or rethinking what and how is shown in the figure, might help though.*

(R1.5) We completely agree that alluvial plots are not that easy to read, but this figure is the most important summary of the analysis and needs to be presented in all its complexity. We however changed the panel disposition to help the reading of it.

(6) *Numbers of used papers in some parts of the analysis do not add up for me: if 91 papers were scrutinised why the sum of papers is Table 1. is 54? Also, how the numbers (27 and 32) for Figure 4 and Figure 5 were calculated?*

(R1.6) There are 91 papers in total but 19 are methodological papers, which makes it only 72 papers total with metadata regarding the conclusions, datasets etc. However, among these 72 papers, 44 concern trends and 34 concern drivers (with 6 concerning both trends and drivers, as explained in the methods section of the manuscript, lines 168-170). 20 are reviews, 4 are reports from the grey literature and 48 are papers analysing directly empirical datasets.

Overall, figures are changing depending on the analyses that are conduct because not all papers are eligible to all types of analyses. For instance, methodological papers were used for discussion only, as mentioned in the manuscript (line 163). Table 1 refers to the category of papers based on empirical datasets (n=48). Some of them are using both BioTIME and the Living Planet Database, which is why the sum is 54 in Table 1.

Figure 4 rely on papers that consider trends and that are based on empirical datasets. This concerns n=32 papers (Fig. 4B). However there were papers for which the information regarding the temporal extent of the datasets was unavailable. We were thus able to construct Figure 4A only with papers having this information, ended up with 27 papers for this particular analysis.

(7) *Typos and unclear wording:*

a. line 10: comma after “remains”
b. line 11: not clear what is “being of social..”
c. line 73: should be clarified what pressures
d. line 76: climate change is also anthropogenic
e. line 146: summarises instead of “summaries”
f. line 208 “declined” the drivers?

(R1.7) Thank you for pointing out the typo. All of these are adressed in the revision.

Reviewer #2:

This manuscript intends to conduct a literature review of studies documenting the temporal dynamics of global biodiversity. The authors found that reviews and meta-analyses, along with the use of global indicators, are more likely to conclude that trends are declining and provide advice on future directions for research. This is a timely publication in view of many contradicting studies being released, namely on the insect extinction narrative (although not limited to it in any way). I do have a few suggestions for improvement as in many parts the text or analyses were not clear to me and in my view still require further thought and work.

We thank the reviewer for these positive comments and for his suggestions which have helped to improve the quality of our manuscript.

Major comments

(1) *It is not clear if there were thresholds applied and what they were to the different response categories. Most studies have at least some trends diverging from the general picture, how were these considered, in decline or mixed/factor-dependent results? Wonder if a continuous measure would not be more appropriate for the analyses, like % of declining trends among all species in a study.*

(R2.1) We acknowledge that the thresholds we used were fuzzy and lacking clarity. However, we decided from the very beginning not to use percentages as we are not claiming to perform any meta-analyses with the aim at producing quantitative estimates. Our review is conceptual with the aim to discuss major sources of confusions in global biodiversity change studies and the corresponding debate.

Yet, the thresholds we have used to consider if the results were falling into each of the categories might not clear enough. This issue, also raised by reviewer 1 (see R1.3), is now addressed in the revision. Papers classified as « declining trends » were those concluding declines for higher than 50 % of the species or populations considered. We double-checked our conclusions by going through again all the papers classified as « mixed trends ». We classified the papers as « mixed » when studies presented as many increasing as decreasing trends and/or a majority of no trends (as expressed line 175). For instance, Daskalova et al. (2020) identify 15 % of decline, 18% of increase and 67 % of no net changes.

The « factor-dependent trends » were distinguished from the « mixed » category as these relate to papers we had actually difficulty to conclude about. For instance, Blowes et al. (2019) report that « the direction and magnitude of change differs across geographic regions » making difficult to conclude if the results are « global decline » or « mixed trends », as they do not claim that there are as much increases as decreases nor a majority of no trend, but rather that some locations present mostly declining trends whereas others show mostly increasing trends and others a mix of them all. In such cases, we think it was important to respect the message claimed by the authors and for such papers we did not find it adequate to put them in the « mixed trends » category. We admit that this factor-dependent trends category might still seem too coarse, but we want to emphasize that this category only includes six papers when considering the trends and nine papers when considering the drivers. As such, we do not believe that managing to highlight clear conclusions from them would substantially change our claims.

(2) *I am missing a formal analysis of the results. It is often the case that conclusions are drawn from percentage of studies reaching some trend with no analysis of the data. I think at least a statistical model where the response variable is decline vs others (or the %declining as mentioned above) and the independent variables are the characteristics of the study is needed.*

(R2.2) This issue, also raised by reviewer 1 (see R1.2), is now addressed as follows in the revision: We agree that statistical tests were missing, even though we were not expecting any strong statistical argument regarding the number of papers considered for each single test. However, the specific tests proposed by the reviewer are not the most appropriate in our opinion, as we only have qualitative variables, and also because we lack the power to perform the advanced statistical analysis proposed. The most appropriate analysis in our opinion was to perform chi square tests in order to test whether the conclusions drawn depended on either (1) the assessment approach, (2) the dataset used, (3) the number of taxonomic groups (split in qualitative categories), (4) the time span (split in qualitative categories), (5) the ecological level, and (6) either the methods or the identity of the studied drivers. We performed these tests both for the trends and the drivers assessments.

Minor comments:

(3) *Everything mentioned in this study is about taxonomic diversity in one way or the other. It would be important to also mention functional, phylogenetic and eventually network diversity. Did the authors find any studies covering the trends in these facets of diversity? If not, maybe it is worth mentioning in the discussion why that is the case.*

(R2.3) Indeed, we did not questioned functional, phylogenetic nor network diversity as we focus on the debate that we highlight on the introduction. We thus excluded such papers as considering the « wrong type of data » while performing the screening process. This is now specified in the manuscript (lines 149-152).

(4) *Ln 60 - Regarding the use of the term biodiversity, two suggestions to add regarding many sources of confusion: https://doi.org/10.1146/annurev-environ-120120-054300 and https://doi.org/10.1016/j.cub.2022.12.003*

(R2.4) These are now cited.

(5) *Ln 168 – Wonder if and how it is possible to study drivers without trends as they are drivers of trends? Maybe some clarification here is needed?*

(R2.5) We agree. This is now clarified in the revised version (lines 166-169).

(6) *Table 1 is not clear. Is "Contant" "Constant"? Or "Content"? Maybe better naming is needed in any case. In Taxa, how were the numbers calculated? For example, 8 in total means 8 large groups? How were the large groups identified?*

(R2.6) Thank you for pointing out the typo. We replaced it by « content », which refers to the content of the databases. We agree that the information regarding the classification of the taxonomic groups was missing. This was specified in the methods section (line 193-195).

(7) *Ln 306 – I would not say it is surprising. Species loss is always preceded by population decline, so trends of population decline must always be stronger than trends of extinction.*

(R2.7) We agree. The word « surprinsingly » has been deleted.

(8) *Ln 482 – No mention is given to the recent Kunming-Montreal agreement and targets.*

(R2.8) Indeed, the manuscript was submitted before the agreement. We've now updated the manuscript accordingly (lines 519-521).

References

Chadegani, A. A., Salehi, H., Yunus, M. M., Farhadi, H., Fooladi, M., Farhadi, M., & Ebrahim, N. A. (2013). A Comparison between Two Main Academic Literature Collections : Web of Science and Scopus Databases. *Asian Social Science*, *9*(5), Article 5. https://doi.org/10.5539/ass.v9n5p18

Daskalova, G. N., Myers-Smith, I. H., & Godlee, J. L. (2020). Rare and common vertebrates span a wide spectrum of population trends. *Nature Communications*, *11*(1), Article 1. https://doi.org/10.1038/s41467-020-17779-0

Blowes, S. A., Supp, S. R., Antão, L. H., Bates, A., Bruelheide, H., Chase, J. M., Moyes, F., Magurran, A., McGill, B., Myers-Smith, I. H., Winter, M., Bjorkman, A. D., Bowler, D. E., Byrnes, J. E. K., Gonzalez, A., Hines, J., Isbell, F., Jones, H. P., Navarro, L. M., … Dornelas, M. (2019). The geography of biodiversity change in marine and terrestrial assemblages. *Science*, *366*(6463), 339-345. https://doi.org/10.1126/science.aaw1620