**Handling editor: Holger Kreft**

The reviewers raise a number of points that you need to address before we can consider publication. Most of their comments circle around clarifying concepts, improving parts of the writing, and finding a more balanced discussion. Reviewer #2 is concerned about the reproducibility of the study - a point I also think is extremely important.

>>> Thank you for the consideration of our revised manuscript. We have taken into account all the comments of the reviewers. Specifically, we have:

1. added a new Table 1 defining the metrics used in this review,
2. modified Fig. 1 (fully), 2b and 2c to clarify the concepts of temporal scales and to make them clearer. Also, we have changed the color palette of Fig. 3 and added a new supplementary Fig. 1,
3. added the rules used for selection of literature in order to make our study more reproducible (with some examples),
4. clarified the definition of temporal grain and why there is a confusion about temporal grain in the scientific literature (with the help of the new Fig. 1),
5. clearly listed the guidelines for authors to share the temporal features of their studies without ambiguity,
6. better emphasized the limitations of the specific selection of studies considered.
7. corrected typos.

All the differences between the original version and the new one are highlighted in blue on the new manuscript.

**Reviewer #1**  
The authors made a big piece of work with this review, but I would like to encourage them to clarify some points and the presentation of the results and to go further in their recommendations. I do not think this last objective (L42) is totally fulfilled in the manuscript.

>>> Reviewer 1 refers to the guidelines that we provide for authors to deliver the temporal features without ambiguity. We do agree that those guidelines were not clearly listed in the first version of the manuscript. Together with a fully revised Fig. 1 that we refer to, we provide a clear list of steps to remove the ambiguity about temporal features (L319):

*“If one wants to study the temporal scaling of biodiversity trends* (*e.g.* Thompson et al., 2002)*, a clear definition and description of all the temporal features of both the study design and the analysis needs to be considered (Fig. 1C, 1D). Thus, future studies should provide the following features: 1) the temporal grain of the sampling (i.e. black dots in Fig. 1D), 2) the temporal grain of the analysis (red boxes in Fig. 1C), 3) the number of samples used to form the temporal grain of the analysis (i.e. the number of black dots in a red box in Fig. 1D), 4) how these samples have been processed to obtain the temporal grain of the analysis (summed/averaged/modelled) 5) how these samples are clumped or spread in time, 6) the temporal lag of the analysis (Fig. 1C), and 7) the temporal lag of the sampling (Fig. 1D). We argue that these guidelines allow for a complete representation of temporal structure, as they differentiate the temporal features of the sampling from those of the analysis (i.e. difference between Fig. 1C and Fig. 1D). This would allow a better assessment of the impact of the temporal features on biodiversity and its trends in the future. These recommendations also apply to spatial features, which are more often considered and simpler to display with maps.”*

First, the context and explanations about temporal scaling could be more developed. The theory of temporal scale effects is mentioned without explaining the implications when calculating trends beyond the good practices of providing the grain (L274-276). What are the implications of having a different temporal grain from the sampling to the metric? What are the implications of 2.5 min versus 10min? The selection of the grain of the sampling depends on what is measured (e.g. Bonthoux & Balent 2012 Journal of Ornithology https://doi.org/10.1007/s10336-011-0766-2). Should it be discussed?

>>> We do agree that the concept of temporal grain and its impact on biodiversity trends was the most challenging to clearly define. To help clarify concepts of temporal scaling, we revised Fig. 1.

Also, we thank Reviewer 1 for notifying us about the study of Bonthoux & Balent (2012) that informed us about Thompson et al. (2002), a rare article exploring this topic (*“CVs (i.e. count variations) did not appear to be affected by changes in count duration.”,* Thompson et al., 2002). Thus, we added this latter reference in the Discussion L319. However, current scientific literature still lacks clear messages about the impact of the temporal grain on biodiversity trends, especially due to a lack of consistency in “temporal grain” definition. It is the latter message that we wanted to highlight in our manuscript (*Issues of temporal grain* paragraph in Discussion and Fig. 1). From a personal point of view, given that spatial grain does have an impact on biodiversity trends, the effect of temporal grain is also expected, but empirical assessments of this assertion are lacking in the literature.

Here, you focused on diversity trends based on occurrences, but among these studies, many measured their trends using data sampled to assess the relative abundances of birds.

(the following answer is the same as for reviewer 2)

>>> We agree with reviewer 2 that not considering abundance-based metric is missing part of the picture, and we do emphasize on this point in the introduction L101:

*“We focus our review on incidence-based metrics (e.g. species richness) as they aim at assessing biodiversity of entire metacommunities and we do not consider abundance-based metrics as they often focus on a limited set of species’ population (e.g. multi-species indicators; Gregory & van Strien, 2010).”*

And L289 by adding:

*“[...] but we stress that those abundance-based metrics, which are often found decreasing, are beyond the scope of our study”*

Including the abundance-based metrics in our review would be problematic and we tried to make it clear in the introduction. First, multi-species indicators (MSI, *e.g.* forest/farmland bird indicator) are focusing on a restricted set of species and ecosystems. Second, some abundance-based metrics are too specialized on a specific topic (*e.g.* community temperature index, L103). Finally, considering abundance-based metrics (*e.g.* multi-species indicators, LPI…) would massively increase the number of studies, leading to a too broad focus of the manuscript and blurring the take-home messages. We do think that focusing only on abundance-based metrics could be the content in the future of an entire review (as specified L290).

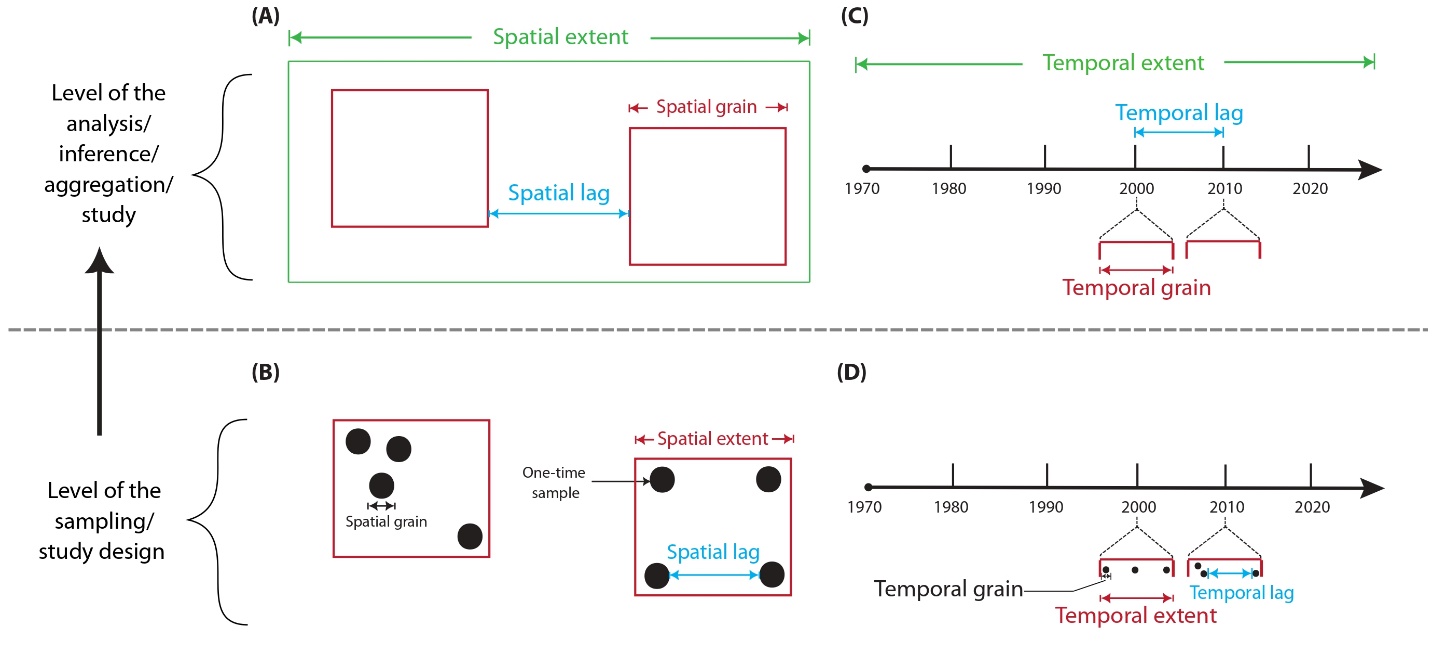
Also, I do not understand how you extracted or measured the temporal grain of the studies. I think I would not be able to reproduce the calculations from the description and examples in the methods section.  
In the example of Wretenberg et al. (2010) you provided L279-280, when I read the paper, I think that I would say that the temporal grain is provided: 10 min. In Monnet et al., even if the change in protocol from 5 to 10 min can make the calculations difficult, I have no idea how you got to 0.09 hour as a temporal grain.

>>> We have clarified (L162, L199) that temporal grains are provided in decimal hours (it is now corrected in the column names of Table 1). These temporal grains are given for the smallest unit of area at which the biodiversity metric is computed. For instance, in Monnet et al. (2014), the indicated temporal grain is 5 minutes: 5/60 = 0.08333 rounded to 0.08 h. Other example: articles using the North American Breeding Birds Survey are summarizing the data at the scale of the routes. Each route is divided into 50 census points surveyed for 3 minutes each. Thus: (3\*50)/60 = 2.5 h (those examples are added to the revised manuscript L167-173).

Concerning Wretenberg et al. (2010), the temporal grain was indeed specified and we thank the reviewer for pointing it out (to quote Wretenberg *et al.*, 2010: “*The observer noted all bird species seen and heard during five minutes.*”). We have added this value in the table of references, and we now cite Davey et al. (2012) to illustrate the point that temporal grain is sometimes not provided clearly (L302).

Overall, despite the examples and the figure, I didn't find the explanation of the temporal grain of the metric very clear.

>>> We have clarified this in new Fig 1, in which we clearly differentiate between the spatio-temporal features (*i.e.* grain, extent and lag) of the analysis Fig. 1a, 1c (above the dotted grey line) and the spatio-temporal grain of the samples Fig. 1b, 1d (below the dotted grey line). Also, we modified the caption of Fig. 1 to emphasize this difference. Finally, we also note that the lack of clarity and consensus concerning the concept of temporal grain in the reviewed scientific literature is an important point of our manuscript.



***Fig. 1 (in colours):*** *Illustration of the concepts of spatial (A, B) and temporal (C, D) features used in this review. Even though they are named in the same way, spatio-temporal grains, extent and, lags are different according to whether one is referring to the analysis conducted (i.e. above the dotted grey line: A, C) or to the study design (i.e. below the dotted grey line: B, D).*

L289 What if the authors keep the maximum between the two samples?

>>> Concerning species richness, if authors keep the maximum number of species observed out of two samples, the spatial grain specified should be the one at which the maximum value of the metric has been registered.

L291: Fig 1b?

>>> Corrected

L363 "this needs to be addressed' I think you should go a step further and provide clear guidelines regarding the good practices on reporting the temporal grain of the metrics and of the sampling plan, with examples of good practices from the literature if they exist.

>>>We now provide these guidelines on L322-332 (or see above first reply to reviewer 1). Also, for the reader to understand that we are referring to this part, we modified the sentence as follows: “[…] *this can be addressed by following our guidelines for specifying the full temporal features of a study”*.

2) I also have comments and suggestions to improve the figures:  
  
Figure 1b: I understand the definitions, but I found the figure did not help to understand them.  
I would rather represent the temporal grain as a zoom from a point in the timeline. Are the y-axis and the continuous variable needed in the figure?

>>> These are several very good points to make Fig 1b clearer. We use the recommendations of Reviewer 1 to modify entirely the Fig. 1. With the revised Fig. 1, we emphasize the fact that even though spatio-temporal features are named the same (*e.g.* spatial grain, temporal grain…), they do not indicate the same process if one is referring to the study design (Fig. 1B, 1D) or to the inference (Fig. 1A, 1C).

Figure 2: Even if the legend explains it, the varying size of birds used for aesthetic reasons is very confusing. I would suggest changing it. Are they too many European-wide studies to add them to the map?

>>> We have unified the bird sizes for each map and added the European-wide studies on the map as follows:

|  |  |
| --- | --- |
| **(A)** | **(B)** |

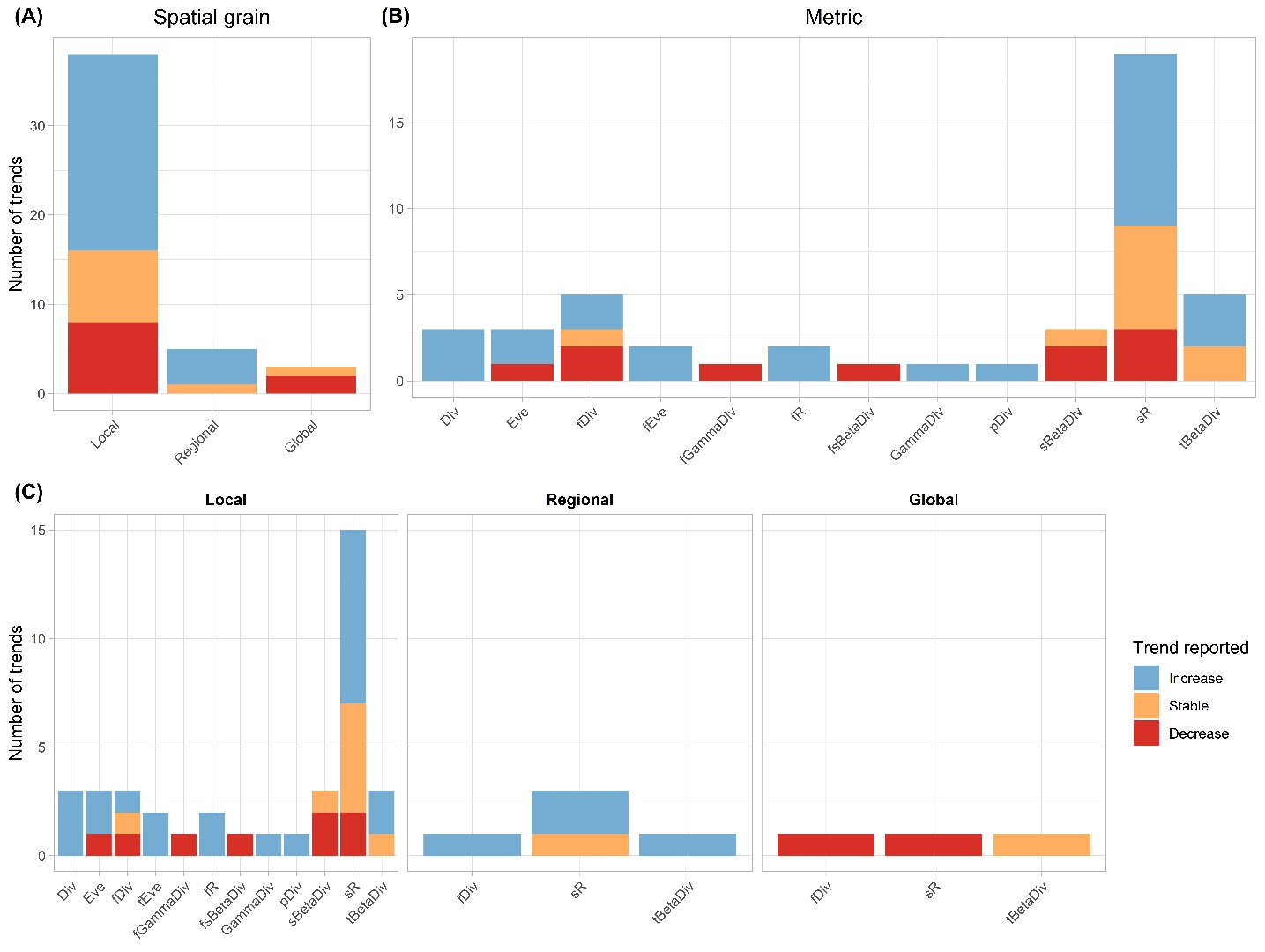
We have also added “*Birds indicate countries where the studies were conducted*” in the figure caption.

Figure 3: In the text, panel c in Figure 3 is used to compare only tBetadiv, sR and Fdiv between local, regional, and global scales. I would consider panel c by a barplot with only these three indices grouped by spatial extents.

>>> Concerning panel c, we understand that displaying the overall number of metrics is a lot of information. However, we do think that removing it would eliminate one key information, that is: in the scientific literature, a higher number of biodiversity facets are found for local grains over regional and worldwide grains.

Also, I think this graph is not colour-blind friendly.

>>> Originally, we used the “turbo” palette from the `viridis` package which is supposed to be color-blind friendly. However, in order to avoid any ambiguity, we have now changed the color palette to "*RdYlBu*" from the RColorBrewer package as follows:



3) The authors resume the trends into the three categories: increase, decrease, and stable. It must have been not easy in some papers. From there, could you also provide some recommendations of good practices when reporting results on temporal trends in scientific papers?

>>> This is an interesting suggestion. I think reporting both the significance and effect size of a temporal trend should be done when possible. However, I did not find any location in the manuscript where these recommendations could be put as it may be out of our scope. However, in the context of our literature review, we think that we can provide warnings about the temporal lag vs the temporal trend reported. Indeed, an increase of biodiversity with a temporal lag equal to the temporal extent, as considered in our manuscript (*i.e.* only 2 points in time: the first and the last) can be the results of a combination of increases and decreases, which are often not reported in the final text. Thus, we added text L189-192.

4) Finally, a so appealing title and this big work of review call for many expectations that cannot be all adressed. I understand that some questions might have been discarded by the authors when delimiting their work of review. Still, I think some are in the scope of the question of spatio-temporal scaling and a couple of them should be at least mentioned or discussed if not addressed. Among them, I would like to know what the authors thought about the implications on the analyses of the following points:  
- The effect of diachronic versus continuous data to calculate

>>> We do not fully understand what the word “diachronic” refers to in this context. If it refers to the variation of temporal lags (*e.g.* 5 years temporal lag vs. yearly surveys), we would reply that long temporal lags does not indicate all the fluctuations of biodiversity (as explained just above). If this is not the case, we would be thankful if the reviewer could give more explanations about this comment.

- The implications of the trade-off between long extent (L351) and huge time lags

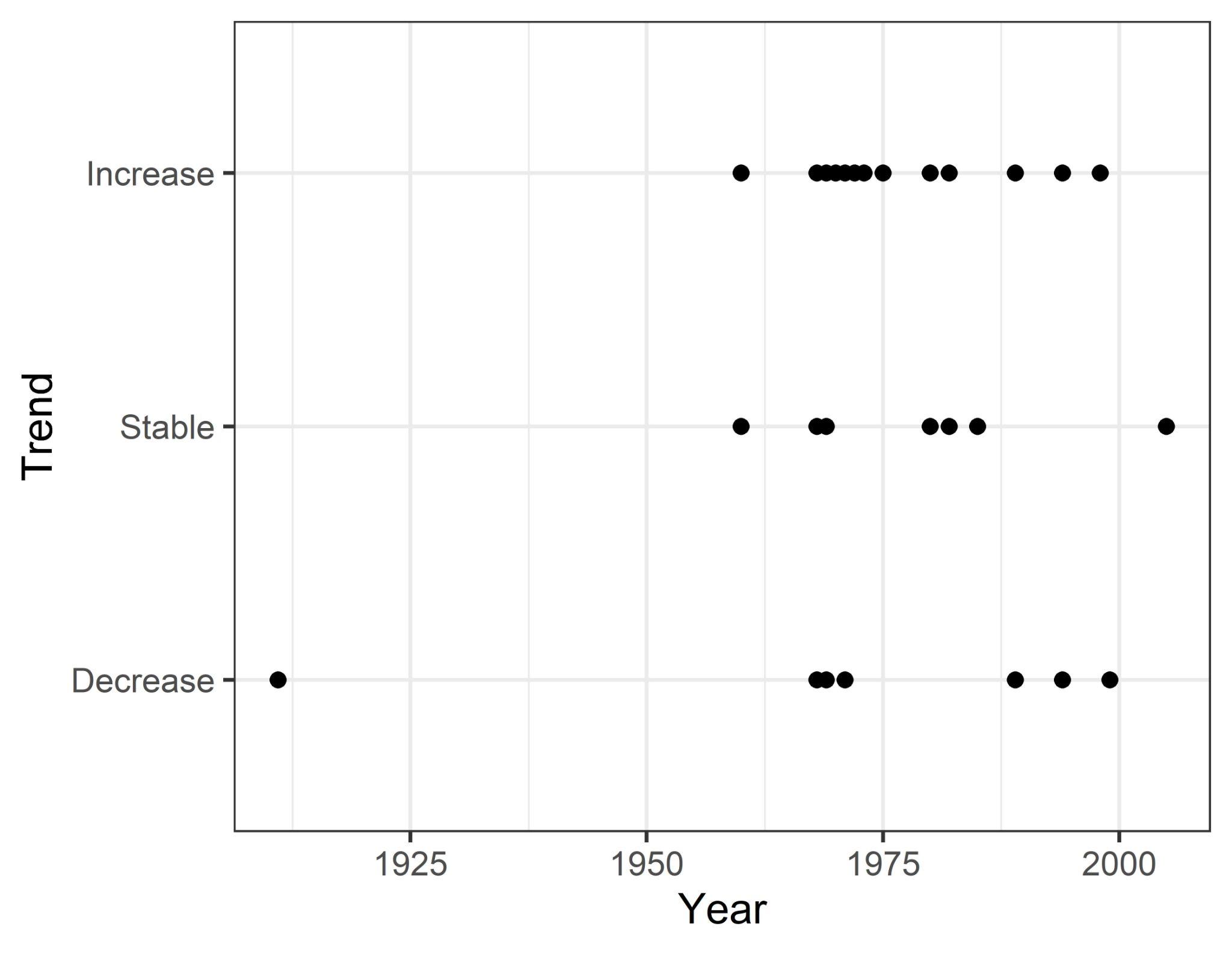
>>> This is an interesting topic. I think the trade-off between long extent and large temporal lags must be considered according to the scope of the study. If the main question is to assess and address the historical changes of biodiversity, long time series (even with large temporal lags) should be considered (added L387). On the other hand, short temporal lags should be favored when assessing, for instance, the influence of recent policy changes or specific conservation actions.

- The implications of the different groups of birds that the authors may have considered in their studies (all species, common birds, some guilds…) to calculate the trends and compare them

>>> We tried as much as we could to select studies which were not focusing on a specific family/guild, but which were considering all bird species at once (see reply to reviewer 2 about the studies selection steps).

- Did you notice an effect of the studies' starting point on the trends (from the studies starting in the 1960s vs in the 2000s)?

>>> Good point, we have checked this (see below the new supplementary Figure 1). A visual check shows there is no link between the starting year and the trend. We added this information L216



- Did the studies performed with the same data and same grain find the same trends?

>>> Yes, studies using the same data, temporal extent and spatial grain reported the same trends (*e.g* Jarzyna & Jetz, 2017, 2018; La Sorte et al., 2009; La Sorte & Boecklen, 2005; they all report an increase of species richness at local scale in North America). That’s why we’ve decided to account for this pseudo-replication, and to always consider only one of the contributions in Fig. 3. The list of all of these similar studies is in Table 1.

Line-by-line comment:  
I noticed throughout the manuscript many sentences that are not very informative:  
- L59-61 ("critical", "we know", "affects macroecological patterns", "great interest")

>>> Clarified  
- L71

>>> Reworded  
- L339 Which kind of study is Fraixedas'?

>>> Clarified by specifying it is a literature review.  
L149 Does the threshold of 50 x50 km² discriminate well the categories local or regional, or are some studies close to this threshold?

>>> Studies that are closest to the 50x50 = 2,500 km² grain threshold are Chase et al. (2019, with 11,000 km² grain) and Chase et al. (2019, 800 km² grain). There is also one study for which the grain is = 2,500 km²: Keller et al. (2020). Thus, except for the latter, the 50x50 threshold seems like a good one.

L228: The first sentence of the discussion seems to show that the main result of your review (*i.e.* negative trends on a global scale and none on smaller scales) is already known. You may need to explain the specificities of your contribution on this point compared to the existing literature.

>>> The contrast between local vs global trends was hypothesized for the first time by Sax & Gaines (2003). However, there was no empirical assessment of this hypothesis before Vellend et al. (2013) who made a meta-analysis on plant biodiversity. Other papers followed (*e.g.* Chase et al., 2019; Dornelas et al., 2014) but they were always limited by the spatial extent, the resolution of the data and were studying a variety of taxa. This idea was also well summarized in an opinion paper by Vaidyanathan (2021). However, our review is the first to put together studies assessing bird biodiversity trends using spatial replications along with their heterogeneous spatio-temporal features, study designs and modeling methods. Thus, the main contribution is to show that this hypothesis (*i.e.* contrast between local vs global trends) is underpinned by the current scientific literature. We summarized this L250-255:

“*The mismatch between global vs. local trends of diversity was hypothesised for the first time by Sax & Gaines (2003) and empirically demonstrated afterwards (Chase et al., 2019; Dornelas et al., 2014; Vellend et al., 2013). However, our review is the first to confirm this hypothesis based on published results, putting together studies assessing bird biodiversity trends using spatial replications along with their spatio-temporal features, study designs, and modelling methods.*”

L305 on how many?

>>> Clarified. It now reads as *“[…] (we found 24 studies out of the 244 references resulting from the queries made in Web of Science).”*

L309-311 reference missing

>>> Added Rosenberg et al. (2017)

L317-320 and so, what did they bring?

>>> We now clarify that L352:

*“[…] who show mismatching trends through spatial grains”*

330-333 Explain what developments in EBBA and GBIF make "hope to see trends with spatial replicates at national grain soon"?

>>> Clarified with: *“[…] by providing data at large spatial extent, […]”*

L364 unclear - what are your recommendations in modelling?

>>> We recommend using cross-scale models, which can predict biodiversity, and its trends, through a continuum of spatio-temporal grains. In order to give more details to the reader, we added “*[…] by cross-scale models*” and we refer to Keil & Chase (2019).

**Reviewer #2:**

1) Defining biodiversity and related metrics. While there is a general agreement according to which "global biodiversity is undoubtedly decreasing" (L228), there is a debate in the scientific community regarding the nature of the biodiversity crisis and the need to rely on multiple metrics. Even the term "decline" is challenging to define when simultaneous metrics are considered. So overall what is meant by biodiversity? Is it only species richness? Is it all species-based metrics? Or all existing metrics? Giving a formal definition of the 12 retained metrics (lines 117-121) in a glossary or in a supplementary item would be a good start as well as a broader approach to define biodiversity.

>>> We are considering many metrics and a decrease of one can be reflected by an increase of another (e.g. increase of species diversity and decrease of functional diversity at local scale in Schipper et al., 2016). Concerning the sentence cited by reviewer 2 (*"global biodiversity is undoubtedly decreasing"* L228), we are referring to the most common incidence-based metric: species richness, which is decreasing at global scale (Barnosky et al., 2011; Diamond, 1989; Sax & Gaines, 2003; Smith et al., 1993…). We have clarified this on lines L236 and L56. Otherwise, when only the word *“biodiversity”* is used without specifying any metric (for easier reading), then it refers to all the metrics.

Importantly, we agree that a table with definitions of the metrics would be helpful. Thus, we have added a new Table 1:

***Table 1:*** *definition of the metrics used in this review*

|  |  |
| --- | --- |
| **Metric name** | **Definition** |
| *Species richness (sR)* | Number of species. |
| *Functional richness (fR)* | Number of morphological, physiological, and/or ecological traits spread across species. |
| *Evenness (Eve)* | Similarity of the relative abundances between species. The closer are abundances of species to each other, the higher the evenness. Examples are Pielou’s or Simpson’s evenness indices. |
| *Functional evenness (fEve)* | Evenness applied to functional richness. |
| *Diversity (Div)* | Diversity metrics which combine both richness and evenness. Examples are the Simpson’s and Shannon’s indices. |
| *Functional diversity (fDiv)* | Diversity metrics which combine functional richness and functional evenness. |
| *Phylogenetic diversity (pDiv)* | Diversity of evolutionary lineages, often measured as the sum of branch lengths of a phylogenetic tree. Example is Faith’s phylogenetic diversity (Faith, 2006). |
| *Temporal beta-diversity (tBetaDiv)* | A measure of dissimilarity of species composition between two time periods at a single location. Example is the temporal Jaccard index. |
| *Spatial beta-diversity (sBetaDiv)* | A measure of dissimilarity of species composition between two locations in space at a single temporal snapshot. Example is spatial Jaccard index. |
| *Functional spatial beta-diversity (fsBetaDiv)* | Spatial beta-diversity applied to composition of functional traits in communities. |
| *Gamma diversity (GammaDiv)* | Total diversity of all communities in a larger region. |
| *Functional gamma diversity (fGammaDiv)* | Total functional diversity of a larger region. |

2) Global vs local scale. What is meant by "global biodiversity" also needs clarification. Some analyses consider local biodiversity but include it in meta-analyses whose objective is to characterise trends on a global scale [1,2]. Can a collection of local studies scale up to a global scale?

>>> We considered trend assessment as global (*i.e.* spatial grain = global) when studies assessed the trends for all the birds on Earth (*i.e.* **not** a cumulation of local studies with a global spatial extent). However, assessing biodiversity trends in this way is challenging, hence the very few numbers of global trends found in the scientific literature. The only manuscript claiming to assess global decline is Jarzyna and Jetz (2018) who used the work by Szabo et al. (2012). The latter used a set of extinction lists in order to assess global biodiversity decline since 1500.

Concerning ref. [1,2] cited by the reviewer 2, those are typical examples of what we emphasize in our manuscript: for both references, the **spatial extent** is global, while the **spatial grain** is local. Thus, we considered them as assessments of local biodiversity change.

3) The selection process for the papers considered would be difficult to reproduce. What and why some papers are excluded would be useful in a supplementary material.  
How the 156 references are processed to select the 24 final papers? This is key to the understanding of the corpus and to potential future replication.

>>> Even though these steps were scattered along the first version of our manuscript, we added a summary of the following steps L137-148:

1. Studies considering the temporal trends of diversity metrics (*i.e.* species/functional/phylogenetic richness and their variations). Rejecting metrics which are focusing on the abundance of only one or a few sets of species (*e.g.* all the multispecies indicators, *i.e.* MSI).
2. Studiesassessing and summarizing the temporal trend over **spatial replicates.** This was the most restrictive condition.
3. Studies which were as heterogeneous in the spatial replicates as possible, in order to not have temporal trends influenced by one specific driver. For instance, we rejected studies which were assessing the temporal trend after a perturbation (*e.g.* impact of logging in Hill & Hamer, 2004; Roels et al., 2019 before/after tree planting; Sirami & Monadjem, 2012 who studied the impact of shrub encroachment; Xu et al., 2018 who studied the impact of urbanization…). Also, we rejected studies which were assessing temporal trends for a specific type of ecosystem (*e.g.* Latta et al., 2011 who focused on Ecuadorian Andes; Scarton, 2017 who focused on the lagoon of Venice)

4) The limits of community metrics. It's surprising that pure and simple abundance metrics (population size) are not considered. We understand the choice of focusing on species richness and related metrics. But by ignoring multi-species average trends in abundance, part of the story is missed. The trends in LPI type of metrics or those targeting birds in particular (Indicators produced by European or National Bird Census) are essential to capture biodiversity responses to global changes and missed by community metrics. This should be more emphasized.  
For instance, L228 again, "there is still no evidence of this negative trend at local and regional scales" is somewhat misleading. First, isn't it twisted to expect local trends to reflect the global ones? Indeed, I do not expect the global trend to "propagate" locally but rather the opposite. Second, the query adopted in this paper does not capture the local evidences of loss in habitats and populations or species massively documented. And this is a methodological decision, not a lack of evidences. I take one example at random following a rapid search on google scholar about individual species extinctions: Ford et al. 2009. Extinction debt or habitat change? - Ongoing losses of woodland birds in north-eastern New South Wales, Australia. Biological Conservation. This is a case (among tons) of evidences for a regional extinction of individual species. It's of course filtered out in the query conducted in the paper because it's not using species richness. But isn't it contributing to biodiversity dynamics? Isn't it a possible evidence that global changes have local consequences ? This is mentioned briefly L270-272. But too briefly and without the due attention.  
I suggest that this issue to be also proposed as a necessary angle to adopt when approaching biodiversity trends: not focusing on aggregated metrics, whatever the scale and resolution issues, but also on other aspects (population sizes, local extirpations of individual species, habitat destruction...). The authors seem aware of this, but I think it should be made clearer to the reader.

(the following answer is the same as for reviewer 1)

>>> We agree with reviewer 2 that not considering abundance-based metric is missing part of the picture, and we do emphasize on this point in the introduction L101:

*“We focus our review on incidence-based metrics (e.g. species richness) as they aim at assessing biodiversity of entire metacommunities and we do not consider abundance-based metrics as they often focus on a limited set of species’ population (e.g. multi-species indicators; Gregory & van Strien, 2010).”*

And L289 by adding:

*“[...] but we stress that those abundance-based metrics, which are often found decreasing, are beyond the scope of our study”*

Including those abundance-based metrics in our review would be problematic and we tried to make it clear in the introduction. First, multi-species indicators are focusing on a restricted set of species and ecosystems. Second, some abundance-based metrics are too specialized on a specific topic (*e.g.* community temperature index). Finally, considering abundance-based metrics (*e.g.* multi-species indicators, LPI…) would increase massively the number of studies, leading to a too broad focus of the manuscript and blurring the take-home messages. We do think that focusing only on abundance-based metrics could be the content in the future of an entire review (as specified L290).

5) The discussion about the meaning and implications of any given trend needs more cautious. For instance, take a metric of beta-diversity reflecting homogenization. An increase means less diversity and is possibly interpreted as a deterioration from conservation perspectives. It can result from an increase in local species richness. So simply counting increases or decreases (as proposed in the section "overall trends" p 10) is hazardous. At least some kind of preferable direction for each metric should be agreed on I suppose. After all, 20 increases + 1 decrease is similar than 1 increase and 20 decreases if one ignores the increase/decrease of what? Why an increase "per se" should be better? I missed introduction/discussion about the lack of meaning of any trend if no objective or expectation is associated to the trend. Again a table with the definition and meaning of each metric would be useful.

>>> We do agree that, in the first version of the manuscript, we sometimes used the word “biodiversity” without specifying the metric we were referring to (especially in the discussion and introduction). Thus, we modified the manuscript, so the correct metric(s) are mentioned (modified L105, L236, L245, L280, L275, L376). Concerning the section “overall trend”, this is a descriptive paragraph in which we are referring to all the metrics together, thus we do not discuss the meaning of the trends. Also, Table 1 of definitions added (see above).

6) The study challenges the spatial/temporal resolution of the ~60 studies collected. But the resolution of this approach is also a potential problem. For instance, in figure 3c when the conclusion for metrics x grain rests on very few numbers (3 papers for 3 metrics at global scale), it's unclear what can really be concluded. One point is not really considered is that we still miss data for such a review. With 300 studies analyzed instead of 50, maybe the conclusions would be different? So in other words, could regional and local scale declines in biodiversity simply still be under sampled? I think that concluding anything general from Fig. 3c is really risky.

>>> Good point! Indeed, the lack of data is a key argument of this manuscript. We dedicate an entire paragraph L366-387 to it in the discussion (**Lack of spatio-temporal coverage**). Our take on this is that the global decline in species richness is currently mainly happening in the global south, for which we do not have spatially replicated data at local grains. Here, all the studies assessing temporal trends at local grain are mainly from either Europe or North America where most of the declines happened before data collections. We now discuss this on lines L375-380 as follows:

*“We have reasons to think that the dominant increase of diversity (e.g. species richness, Shannon index…) reported in the reviewed studies applies only on the north hemisphere, especially due to temperature increase that impact positively species richness through species’ range shifts (Walther et al., 2002) and because most of the extinctions in Europe and North America might have happened before data collection.”*

Other comments:  
L105-113. Another justification to work on birds is that they cover a large spectrum of strategies, diet, habitat etc… They are not all similarly responsive to a given disturbance (natural or anthropogenic)…So this variability offers the possibility to formulate and test specific predictions.

>>> This is a very good point that we now discuss on lines L121-123 as follows:

*“Finally, they represent a large spectrum of functional traits (e.g. diets, morphology, ecology), habitats, and responses to perturbations, and are thus suitable for tests of macroecological theories.”*

L131. With these replications…You mean with Keller compared to Barnagaud ? "More general and reliable"… than what scenario? Please explain to avoid confusion.

>>> Clarified: “*By assessing trends over spatial replicates, the trend reported at one spatial grain is more general and statistically reliable.*”

Fig. 2. Why 24 studies ? I cannot find this number above (Why not 20 as in L180 ? I am confused)

>>> Corrected L199 to 24.

L251. This is really hazardous to equate the "quality of biodiversity" to ecosystem services unless you want to endorse a hard anthropocentric and instrumental lens. Some services need very few species but many individuals (eg game species for food), others would need very specific functions but not superior any "quality"…So I would not use the ecosystem services concept here, it's really not needed. That losing beta-diversity is bad for the ecology of species and systems because it is a loss of potential interactions is I think a much stronger justification.

>>> Fair point. Replaced by: “*This local homogenization can be seen as a threat for ecosystems, as the replacing species do not necessarily provide the same ecosystem functions as the replaced ones*”

L255. Is increasing functional diversity always wanted? Some habitats harbor poor functional diversity. Increasing "per se", the number of species or functions is meaningless from a conservation perspective. It is really context dependent I think. Is maximizing functional diversity everywhere a desirable goal? It would be meaningless and lead to maximum functional homogenization. That some ecosystems have less diversity (whatever the metric used) seem perfectly fine. What matters is their trajectory, the cause of the changes…I suggest more cautious is used when a drop/increase in diversity is taken as a problem/improvement per se.

>>> It is indeed a tricky topic. The benefits of the direction of a temporal trend, whatever the metric, depends a lot on the location in space and time at which it is assessed. We try as much as we can throughout our manuscript to not give any value judgments when talking from a general point of view (*i.e.* talking about several metrics, several locations in space and time…). See for example L207-216.

L259-269 and L335-352. Very good points. Some of this could I think come earlier and introduce the necessary nuances to interpret the results (see concern expressed above regarding the difficulty to interpret a local increase in species richness or function as a synonym of no biodiversity decline or as a good sign for the "quality" of biodiversity).

>>> Thank you. We agree that these are two important points of our review. However, introducing this information earlier in the manuscript (*e.g.* introduction) would need introductions of the concepts and thus would increase the number of words in a manuscript that is already quite long. We think that having these points in the discussion is better for the reader.

L275. Explain what is the "species-time" relationship.

>>> Explained (now L297).

L273. Issues of temporal grain and extent. It's unclear what introduces uncertainty (sampling error) and what can produce real and more problematic bias (estimating false negative/positive trends because of this discrepancy in temporal grain). Moreover, I felt that more attention should be given to the variability in temporal "extent". Is it better to rely on 2 time series of 10 years or 1 of 20? Most of the time, only the extent of the longest time series is displayed by a given study, even though it does not reflect the overall abundance and quality of the data. For instance, in Dornelas et al., 2014 ([1] below), not all time series extend from 1960 to 2000, only the longest ones. So what is the extent in this case? It might be interesting to consider the impact of length by looking at the time series one by one for example, or by considering the average length of the time series in a given study. Several solutions can be discussed, but I think it is important to raise this point as well, because otherwise the temporal extent of the data is overestimated.

>>> First, we updated Fig. 1 entirely in order to better explain the spatio-temporal features. We do agree that we should warn the reader that the temporal extents retained in our table are the maximum for each reference. We added this information in Fig. 2 caption as the temporal extents are all displayed on it. Second, as pointed out in the first version of the manuscript L169, we warn about the fact that the temporal trend retained is the one from the first and last points in time and that it does not represent all the fluctuations of the dynamic. However, we emphasize this point by adding text L189-193.

L329. Why national standardization per se is a problem here? Because it's different among countries? I guess standardization would be a good thing for comparing trends among countries, right?

>>> Our phrasing was unclear: we meant that the standardization is made at the national scale (which indeed is a good thing) but that those standardization criteria are different for each nation, making the inter-national comparisons challenging. We now phrase it L359-362.

Fig. 2. Varying the size of the bird for aesthetic reason does not work. The norm is that different sizes in a symbol represent different sample sizes. Maybe use different shape of different species for aesthetic reason and given true information related to the spatial grain or size on that figure.

>>> The size of the birds has been unified for each map as follows:

|  |  |
| --- | --- |
| **(A)** | **(B)**  Diagram, map  Description automatically generated |

[1] Dornelas M, Gotelli NJ, McGill B, Shimadzu H, Moyes F, Sievers C, et al. Assemblage Time Series Reveal Biodiversity Change but Not Systematic Loss. Science 2014;344:296-9. https://doi.org/10.1126/science.1248484.  
[2] Vellend M, Baeten L, Myers-Smith IH, Elmendorf SC, Beauséjour R, Brown CD, et al. Global meta-analysis reveals no net change in local-scale plant biodiversity over time. Proc Natl Acad Sci 2013;110:19456-9. https://doi.org/10.1073/pnas.1312779110.

***References***

Barnosky, A. D., Matzke, N., Tomiya, S., Wogan, G. O. U., Swartz, B., Quental, T. B., Marshall, C., McGuire, J. L., Lindsey, E. L., Maguire, K. C., Mersey, B., & Ferrer, E. A. (2011). Has the Earth’s sixth mass extinction already arrived? *Nature*, *471*(7336), Article 7336. https://doi.org/10.1038/nature09678

Chase, J. M., McGill, B. J., Thompson, P. L., Antão, L. H., Bates, A. E., Blowes, S. A., Dornelas, M., Gonzalez, A., Magurran, A. E., Supp, S. R., Winter, M., Bjorkman, A. D., Bruelheide, H., Byrnes, J. E. K., Cabral, J. S., Elahi, R., Gomez, C., Guzman, H. M., Isbell, F., … O’Connor, M. (2019). Species richness change across spatial scales. *Oikos*, *128*(8), 1079–1091. https://doi.org/10.1111/oik.05968

Davey, C. M., Chamberlain, D. E., Newson, S. E., Noble, D. G., & Johnston, A. (2012). Rise of the generalists: Evidence for climate driven homogenization in avian communities. *Global Ecology and Biogeography*, *21*(5), 568–578. https://doi.org/10.1111/j.1466-8238.2011.00693.x

Diamond, J. M. (1989). The present, past and future of human-caused extinctions. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *325*(1228), 469–476; discussion 476-477. https://doi.org/10.1098/rstb.1989.0100

Dornelas, M., Gotelli, N. J., McGill, B., Shimadzu, H., Moyes, F., Sievers, C., & Magurran, A. E. (2014). Assemblage Time Series Reveal Biodiversity Change but Not Systematic Loss. *Science*, *344*(6181), 296–299. https://doi.org/10.1126/science.1248484

Faith, D. P. (2006). The Role of the Phylogenetic Diversity Measure, PD, in Bio-Informatics: Getting the Definition Right. *Evolutionary Bioinformatics*, *2*, 117693430600200000. https://doi.org/10.1177/117693430600200008

Gregory, R. D., & van Strien, A. (2010). Wild bird indicators: Using composite population trends of birds as measures of environmental health. *Ornithological Science*, *9*(1), 3–22. https://doi.org/10.2326/osj.9.3

Hill, J. K., & Hamer, K. C. (2004). Determining impacts of habitat modification on diversity of tropical forest fauna: The importance of spatial scale. *Journal of Applied Ecology*, *41*(4), 744–754. https://doi.org/10.1111/j.0021-8901.2004.00926.x

Jarzyna, M. A., & Jetz, W. (2017). A near half-century of temporal change in different facets of avian diversity. *Global Change Biology*, *23*(8), 2999–3011. https://doi.org/10.1111/gcb.13571

Jarzyna, M. A., & Jetz, W. (2018). Taxonomic and functional diversity change is scale dependent. *Nature Communications*, *9*(1), 2565. https://doi.org/10.1038/s41467-018-04889-z

Keil, P., & Chase, J. M. (2019). Global patterns and drivers of tree diversity integrated across a continuum of spatial grains. *Nature Ecology & Evolution*, *3*(3), 390–399. https://doi.org/10.1038/s41559-019-0799-0

Keller, V., Herrando, S., Voříšek, P., Franch, M., Kipson, M., Milanesi, P., Martí, D., Anton, M., Klvaňová, A., Kalyakin, M. V., Bauer, H.-G., & Foppen, R. P. B. (2020). *European Breeding Bird Atlas 2: Distribution, Abundance and Change*. Lynx Edicions. https://www.ebba2.info/results/

La Sorte, F. A., & Boecklen, W. J. (2005). Changes in the diversity structure of avian assemblages in North America. *Global Ecology and Biogeography*, *14*(4), 367–378. https://doi.org/10.1111/j.1466-822X.2005.00160.x

La Sorte, F. A., Lee, T. M., Wilman, H., & Jetz, W. (2009). Disparities between observed and predicted impacts of climate change on winter bird assemblages. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1670), 3167–3174. https://doi.org/10.1098/rspb.2009.0162

Latta, S. C., Tinoco, B. A., Astudillo, P. X., & Graham, C. H. (2011). Patterns and Magnitude of Temporal Change in Avian Communities in the Ecuadorian Andes. *The Condor*, *113*(1), 24–40. https://doi.org/10.1525/cond.2011.090252

Monnet, A.-C., Jiguet, F., Meynard, C. N., Mouillot, D., Mouquet, N., Thuiller, W., & Devictor, V. (2014). Asynchrony of taxonomic, functional and phylogenetic diversity in birds. *Global Ecology and Biogeography*, *23*(7), 780–788. https://doi.org/10.1111/geb.12179

Roels, S., Hannay, M., & Lindell, C. (2019). Recovery of bird activity and species richness in an early-stage tropical forest restoration. *Avian Conservation and Ecology*, *14*(1). https://doi.org/10.5751/ACE-01330-140109

Rosenberg, K. V., Blancher, P. J., Stanton, J. C., & Panjabi, A. O. (2017). Use of North American Breeding Bird Survey data in avian conservation assessments. *The Condor*, *119*(3), 594–606. https://doi.org/10.1650/CONDOR-17-57.1

Sax, D. F., & Gaines, S. D. (2003). Species diversity: From global decreases to local increases. *Trends in Ecology & Evolution*, *18*(11), 561–566. https://doi.org/10.1016/S0169-5347(03)00224-6

Scarton, F. (2017). Long-term trend of the waterbird community breeding in a heavily man-modified coastal lagoon: The case of the important bird area “Lagoon of Venice.” *Journal of Coastal Conservation*, *21*(1), 35–45. https://doi.org/10.1007/s11852-016-0470-8

Schipper, A. M., Belmaker, J., Miranda, M. D. de, Navarro, L. M., Böhning-Gaese, K., Costello, M. J., Dornelas, M., Foppen, R., Hortal, J., Huijbregts, M. A. J., Martín-López, B., Pettorelli, N., Queiroz, C., Rossberg, A. G., Santini, L., Schiffers, K., Steinmann, Z. J. N., Visconti, P., Rondinini, C., & Pereira, H. M. (2016). Contrasting changes in the abundance and diversity of North American bird assemblages from 1971 to 2010. *Global Change Biology*, *22*(12), 3948–3959. https://doi.org/10.1111/gcb.13292

Sirami, C., & Monadjem, A. (2012). Changes in bird communities in Swaziland savannas between 1998 and 2008 owing to shrub encroachment. *Diversity and Distributions*, *18*(4), 390–400. https://doi.org/10.1111/j.1472-4642.2011.00810.x

Smith, F. D. M., May, R. M., Pellew, R., Johnson, T. H., & Walter, K. S. (1993). Estimating extinction rates. *Nature*, *364*(6437), Article 6437. https://doi.org/10.1038/364494b0

Szabo, J. K., Khwaja, N., Garnett, S. T., & Butchart, S. H. M. (2012). Global Patterns and Drivers of Avian Extinctions at the Species and Subspecies Level. *PLOS ONE*, *7*(10), e47080. https://doi.org/10.1371/journal.pone.0047080

Thompson, F. R., Burhans, D. E., & Root, B. (2002). Effects of Point Count Protocol on Bird Abundance and Variability Estimates and Power to Detect Population Trends. *Journal of Field Ornithology*, *73*(2), 141–150. https://www.jstor.org/stable/4131137

Vaidyanathan, G. (2021). The world’s species are playing musical chairs: How will it end? *Nature*, *596*(7870), 22–25. https://doi.org/10.1038/d41586-021-02088-3

Vellend, M., Baeten, L., Myers-Smith, I. H., Elmendorf, S. C., Beausejour, R., Brown, C. D., De Frenne, P., Verheyen, K., & Wipf, S. (2013). Global meta-analysis reveals no net change in local-scale plant biodiversity over time. *Proceedings of the National Academy of Sciences*, *110*(48), 19456–19459. https://doi.org/10.1073/pnas.1312779110

Walther, G.-R., Post, E., Convey, P., Menzel, A., Parmesan, C., Beebee, T. J. C., Fromentin, J.-M., Hoegh-Guldberg, O., & Bairlein, F. (2002). Ecological responses to recent climate change. *Nature*, *416*(6879), Article 6879. https://doi.org/10.1038/416389a

Wretenberg, J., Pärt, T., & Berg, Å. (2010). Changes in local species richness of farmland birds in relation to land-use changes and landscape structure. *Biological Conservation*, *143*(2), 375–381. https://doi.org/10.1016/j.biocon.2009.11.001

Xu, X., Xie, Y., Qi, K., Luo, Z., & Wang, X. (2018). Detecting the response of bird communities and biodiversity to habitat loss and fragmentation due to urbanization. *Science of The Total Environment*, *624*, 1561–1576. https://doi.org/10.1016/j.scitotenv.2017.12.143