Editor's comments:

The preprint "Using a large-scale biodiversity monitoring dataset to test the effectiveness of protected areas in conserving North-American breeding birds" has been reviewed by two experts in the field. They agreed on the value of the research presented, and the soundness of the overall approach: the manuscript addresses an important question of conservation ecology "Are protected areas effective?", and brings new insights into how to answer that question, specifically by accounting for environmental variation across sites to allow better quantification of effects. Both reviewers outlined a number of concerns, mostly around adequate presentation of the research, and some methodological aspects, and I would encourage to take them into consideration to improve the manuscript. Particularly, I agree with the reviewers that some terms could be made more precise (e.g. "protection effectiveness" being different from a "good designation", and the choice of considering only native species made more explicit in the terminology). The relevance of species richness as a metric for change is worth discussing, as is detectability (though I agree with the reviewer that a simple mention might be enough). Overall more information on the methods (sample size, raw data distribution and motivation for winsorization, motivation for having two phylogenetic models) need to be included. Detailed suggestions can be found in the review, about a number of aspects including writing style and references, which I won't comment on here but seem like good options to improve the manuscript.

We are sending here a new version of our article following the first round of the review process. We are very grateful to the recommender and to both reviewers for their great and very constructive comments on the earlier version of the manuscript. As you will see in the detailed answers, we have applied almost every comment, which helped a lot to improve the manuscript in terms of analyses and writing. There was only two suggestions that we did not incorporate: compare our results with eBird data (which would be highly interesting but requires a lot of work) and changing scale to full routes (we answered in detail to this comment). We hope that you will find the new version of the manuscript convincing and are looking forward to read your next comments.

Reviewer 1 comments (Willson Gaul):

Cazalis et al. assess the effectiveness of protected areas for preserving birds, using both assemblage-level and individual species-level analysis. The main contribution of this paper seems to me to be the explicit consideration of land cover type when comparing protected and unprotected areas, such that protected areas are compared to unprotected areas of the same land cover type. At the individual species level, the study evaluates the different effects of protected areas on species with different habitat requirements and human tolerance, highlighting which species are likely to be most benefited by protected areas. Overall I find the study interesting and useful, especially because it evaluates whether protected areas are more successful for some habitat types and some types of species – this seems like it could be of direct interest to applied managers. I also find this study interesting in the context of other recent overviews of biodiversity change that, like this study, have found little change in assemblage-level metrics but greater change in species-level metrics [1][2][3]. I think this paper would benefit from more explicit discussion of other studies that have found little assemblage-level change, and what implications that has for using assemblage-level metrics (particularly species richness) to measure biodiversity change.

I have one concern (discussed below) about the authors' decision to exclude non-native species from analysis, and I have a concern about "winsorizing" the data in the statistical methods (also discussed below). Finally, I have listed some minor comments about word choice and sentence structure; these minor sentence-level issues are a slight barrier to smooth reading, but generally did not interfere with my ability to understand the article.

The introduction and/or discussion could benefit from addressing other studies that have found small changes or no changes in species richness over time and in response to disturbance [1][2][3]. This study compares species richness of protected and unprotected sites, with the assumption that species richness would decline more in unprotected than in protected sites because of local extinctions and population declines in non-protected areas. This would lead to an observed pattern of lower species richness in unprotected areas. However, Supp & Ernest [3] reviewed published studies and found that species richness did not respond strongly to disturbance. They suggested that "community-level measures are poor indicators of change..." [3]. Supp & Ernest [3] did find stronger species-level responses to disturbance, which is in line with the results of this study. I think it would be useful to put this study and its results into the context of broader biodiversity trends, and especially to mention other studies that, like this study, have found that differences in individual species metrics (e.g. individual species abundance changes, species turnover) are larger than differences in species richness [1][2][3].

Thank you for this very good suggestion. We amended the text in the discussion to add this point and the references: "In one sense, this is not surprising, as several large-scale studies found that assemblage metrics – particularly species richness – are relatively resilient to disturbance through species substitution (Dornelas et al. 2014; Supp and Ernest, 2014)". L432-435. However, we did not wish to include this in the introduction as we still expect richness and abundance to drop if disturbance keeps increasing.

If I understand correctly, this study excluded non-native species from analyses (line 210). It may be reasonable to limit the study to native species if the goal is to evaluate the effectiveness of PAs in preventing local or regional extinctions and population declines. However, excluding non-native species does not make sense if the objective is to study the effectiveness of PAs in conserving ecological communities, because communities are changed by the addition of non-native species. The abstract of this study states that the study investigated "protected areas effectiveness in conserving bird assemblages" (line 22) and this is also mentioned in the discussion (line 418). In fact, I think this study investigates protected areas' effectiveness in conserving *native bird species*, even if the assemblages in which those native species live have changed due to the addition of non-native species. Excluding non-native species could be masking important changes in the ecological community. I think it would be helpful for the authors to clarify whether this study is aiming to study native species conservation or entire community assemblages. If the aim is to study community assemblages (e.g. line 22), then I think all species detected should be included in the assemblage analyses. If non-native species are excluded from analyses, then I would find it clearer if the terms "species richness" and "abundance" were changed changed to "native species richness" and "abundance of native species" throughout the text. This will clarify exactly what aspects of the community are being studied. In particular, claims like that made in line 488 that "...we measured PAs effectiveness as the difference in abundance or richness between protected and unprotected sites" are not quite correct if non-native species have been excluded. The actual difference in species richness or abundance between sites is unknown in this study, and may differ from the difference between native species richness or native species abundance.

Non-native species are not of main interest for us, as we do not expect conservation to have a positive impact on them (often the opposite). But we do not want this exclusion to bring questions about the robustness of our results. And indeed, if non-native species replace or remove native species, their exclusion could affect our results. Therefore, we ran the analyses (assemblage and species level) with and without these species, which showed consistent results. Results without these 7 species are still presented in the main text, the new results (with these species) are presented in appendix. We clarified this in the methods: “We also excluded seven non-native species, as they are not the focus of conservation efforts. The main dataset we analysed included 400 species in total. To test if removing non-native species can bias analyses (e.g, because they replace native species) we also ran analyses including these species (results presented in Appendix S6).” L.212-216

I am concerned about the winsorizing of abundance values for species (line 217). If I understand the methods correctly, individual species abundance is analyzed using GAMs with a Poisson error distribution. Did the authors check whether the data fit distributional assumptions before or after winsorizing, e.g. by looking at QQ plots or histograms of residuals [4]? In general, it is preferable to model the original data values using an appropriate error distribution rather than transforming or changing extreme values. If the authors still decide to winsorize the data, then they should justify this by reporting whether the winsorized data fit a Poisson error distribution better than the un-transformed data did. I do not know how much the choice of whether to winsorize the data affects the results, but in general I think it would be preferable to model these data using their true values and an appropriate error distribution, rather than changing those values to the value of the chosen quantiles.

We removed winsorization of species abundances and changed Poisson distribution for negative binomial, which allowed a better fit. We also removed winsorization of estimates PAEFor (which was used to remove aberrant estimates) and instead included only species with more than 150 observations. Some species between 100 and 150 observations have so few PAs in their Kernell that they ended up having extreme estimates. We clarified this L.309-311.

Minor comments For spelling or word change suggestions, I have put the proposed new word in **bold**.

Line 17-19: “which whereby...” is odd wording. Remove the word “which”?

We followed the recommendation L.17

Line 27-28: Wrong sentence structure. Perhaps delete “are the one avoiding human activities” so that the sentence reads “At the species level, we found that species that avoid human activities tend to be favoured by protected areas.”

This paragraph changed because of a small change in the results. L26-32.

84-89: I think the comparison to population trend-based methods can be shortened. A criticism of trend-based methods is not a major development of this study. The example (starting at “Hence...” and ending at “...20 times more important” can be removed.

We followed the recommendation L.84

96-137: These 2 paragraphs are the most informative and important part of the introduction. The choice of counter-factuals and how that influences the interpretation of measures of PA effectiveness seems to be a major theme of this study. These 2 paragraphs are a well-written, concise overview of the issues.

132: Re-cite the meta-analyses here? They were originally cited in line 98, but I ended up scrolling up through the text looking for those citations.

We followed the recommendation L.128

184: "effectiveness **in** selecting the most interesting sites..."

This paragraph has changed due to other comments L183-188

185: "effectiveness in creat**ing** ..."

This paragraph has changed due to other comments L183-188

202: "presence **of** PAs ..."

This paragraph has changed due to other comments L.202-204

204: "starting stop of each route" is confusing. Maybe better to say "**first stop** of each route".

We followed the recommendation L.206

210: I find the phrase "non-native" species more familiar than "non-indigenous species" but this is probably just a matter of convention. The authors could change it or leave it as it currently is.

We followed the recommendation across the whole manuscript

216: Consider at least mentioning detectability and the difference between abundance and detections. Unless they have already corrected for detection, what the authors did here was sum detections, not abundances. Using this metric for analysis assumes that detectability was constant across routes and protected and unprotected areas. This assumption should be mentioned.

We have now added a few sentences about detection, saying that detection is particularly expected to vary between vegetation structure types, which is controlled for in our models: "We acknowledge that these values correspond only to detected birds rather than true abundances. Detection is known to vary between habitats, depending on vegetation structure (Pacifici et al., 2008). This could lead to a difference of detection probability (and thus of perceived abundance) between protected and unprotected sites if vegetation structure differs; controlling for vegetation structure in our analyses reduces this bias." L.232-237.

220: Include which software was used (e.g. QGIS, ArcGIS, R).

We followed the recommendation L.265 and L271.

228: Include a reference for the quoted IUCN definition of a protected area.

We followed the recommendation L.250-251.

244: Regarding "effectiveness can vary with protection level", has the introduction already cited studies showing this? Perhaps provide a reference here.

We added two references, already cited in the introduction L.269

250: "(**May** to June)"

Actually it was March (we mistakenly wrote Mars). L.276

262: “... not analysed because **they were** too scarce.”

We followed the recommendation L.289

Fig 3: The negative relationship between human-affinity and PAEfor is hard to see on the scatter plot. Drawing the regression line from the LM model reported in Table 1 might help make the relationship more visible.

Drawing only one line was misleading as the line was not really on the top of the points because of the effect of habitat preference of species. Therefore, we plotted a line for forest species (taking deciduous forest species as they are the most represented in our sample) and a line for non-forest species (taking semi-open species for the same reason). With the new analyses, this relationship is not strictly significant anymore (P=0.087), thus we weakened conclusions on this relationship in the text but we would like to let the figure and the results to keep the idea that it might be an effect of PAs even if we cannot strongly prove it here.

390-505: Overall, the Discussion is interesting and well written.

422: The claim that forest PAs have “more forest-typical bird assemblages” would be stronger if all species (including non-native species) were included in analysis.

As stated earlier, results are consistent and are now both presented in the article. See new Appendix S5 and L212-216 or L.394-397 in the main text.

452: “changed” should be “**changes**”

We followed the recommendation L.488

507-508: This sentence seems weaker than the rest of the Discussion. It does not add anything new.

We removed this sentence L.520.

Reviewer 2 comments (Anonymous):

Cazalis et al. investigate the effectiveness of protected areas on birds in North America using a large-scale citizen science monitoring program. I read this manuscript with interest. The authors address a relevant question related to the effectiveness of protected areas in conserving North American breeding bird assemblages, and for this purpose they use a relevant approach that differs from what has been used mostly in previous assessment studies. Although I am confident about the relevance of this manuscript, I think there are some issues to sort out, and clarifications are needed throughout the manuscript. I also have several methodological concerns regarding this study, and I think this may be addressed before being further considered for publication.

It seems to me that the proportion of sites under protection is relatively low compared to the proportion of sites under no protection status, and even lower when considering sites according to main habitat (main vegetation structure). I would expect this type of bias in sampled sites to obviously lead to a lack of power, and potentially of significancy, in the analysis of PA effectiveness. Also, I am wondering why the authors did not consider the major habitat type based on the proportion of land cover classes within buffer surrounding the entire routes? it would increase the sample size of PA sites. If the authors were worried about the representativeness of a single major habitat across the entire BBS routes, they could have considered using the shapefile provided with BBS data '*bbsrte_2012_alb*' (which corresponds to polylines) to split entire BBS routes into segments (or sections, to use the same terminology as in the manuscript) and then assigned major vegetation structure and PA proportion to each of these segments, instead of using only small sections of BBS routes. See my detailed comments below.

We can answer this comment quite precisely as we had the same thoughts when we started this study. At first, we considered bird observations at the scale of the full route and we calculated landscape variables (including protection) in a buffer of 500 m around the full route. We quickly thought that they were not of good size to correlate landscape variables to assemblages. Indeed, we thought that summing bird abundances that were distant of 50 km did not make sense. More specifically, we had some routes for which a protected forest covered half of the route and we were trying to correlate these landscape characteristics with counts of birds living in open habitats detected on the other half of the route. Shifting to routes' beginning allowed us to reduce land use heterogeneity within routes compared with full routes. Moreover, the proportion of protected routes in full routes was not higher than the proportion of protected routes when we considered route beginning only (4.4 % of full routes are protected by more than 50 %, 5.3 % of the route beginning). Therefore, we thought we needed more local bird counts and thought about splitting the routes in segments of 5 points using BBS polylines as suggested by the reviewer. Unfortunately, an email exchange with Keith Pardieck (US coordinator of BBS) made us realise it was impossible. The problem is that the route polylines the reviewer refers to are a working document, aiming to give directions to observers, but do not represent the actual route done by observers. When an observer goes for the first time on a route, she will travel along this route and stop every half mile to set a point. But if she cannot stop before a mile, every other point will be moved away (to quote Keith: "Generally speaking, when a route is first established and a safe pull-off can't be found at 0.5 mile, then observer may extend distance to find a safe spot. The next stop should be placed 0.5 mile from that extended location."). Therefore, the length of these polylines do not reflects the length of the real route. Then, splitting the routes into segments would have lead in numerous cases to link bird counts to landscapes variables of places distant by several kilometres. In the end, the only precise GPS point we have is the starting point, this is why we only used the first segment of the route.

The mismatch between our sample size and Wood's comes from different definitions of protected areas. We used the World Database on Protected Areas (the reference when talking about PAs) while Wood et al. 2014 use a much wider definition of PAs - because they have a different objective – from the PADUS database (Protected Areas Database of the US). Particularly, they incorporate areas categories as GAP 3, which they define as "GAP 3 lands accounted for 72% of the area of BBS routes within public lands, and have permanent protection from conversion of natural land cover, but are subject to resource extraction. GAP 3 lands include most National Forest lands, where many private inholdings are located." Considering these areas as PAs lead to consider, for instance, Nevada State as protected by 77 % (4 % GAP1, 9 % GAP2, 59 % GAP3, 5 % GAP4) while the WDPA considers 15 % as protected.

To conclude on this answer, yes our sample size in terms of PAs is relatively low, especially for shrub and herbaceous routes. But it would not have been higher if we have considered full routes and is not due to a drastic and biased selection of routes. It is rather due to the fact that PAs are rather scarce and biased against herbaceous areas as we emphasised in Appendix S4. We think that we did the best we could with the BBS dataset on this specific issue.

Although I get the purpose of using a phylogenetic regression - especially in the context of this study - I think this should be clearly justified (e.g. to account for potential phylogenetic relatedness among species). Also, I am wondering why the authors implemented two different types of phylogenetic models (Brownian and Lambda). I did not find any information in the ms. that could justify the use of multiple phylogenetic models… Justification is needed. Finally, as recommended by Ives & Garland (2010), PLR model coefficients were estimated using bootstrap. see my detailed comments below.

Indeed, it was not relevant to use two different phylogenetic models and we therefore removed the Lambda model as suggested, making changes across the whole manuscript. We also changed the way confidence intervals were calculated, using bootstrap, which did not affect the results. We added this information L.345-346.

DETAILED COMMENTS

L78: compare to what? inside vs. outside PAs I guess… Please specify

Indeed, it was comparison of protected and unprotected sites. We clarified this point in the main text L.79

L132: "all three meta-analyses": which one? you change paragraph, so it would help to cite the studies again, or specify something like "discussed above", to help the reader.

We added the reference to clarify the sentence L.128.

L159: "positive effect of PAs" - what kind of effects? provide examples. Given the studies cited, it might be relevant to specify those are mostly designation effects…

We detailed the results of these studies L.152-160.

L184: "effectiveness to select the most interesting sites" - it does seem to me that the designation effect that this sentence refers to, is mentioned for the first time here or at least not clearly stated/defined in the Introduction. This should be better defined in the Intro. Indeed, the expected differences between protected and non-protected sites may result from either the 'designation effect' (the initial state of biodiversity) or from the 'protection effect' (protection efficiency per se). On the first case, it will be assess through a comparison of the bird assemblages (or biodiversity) inside and outside the protected sites, and on the latter case, it will require the comparison of the state of the response variable before and after the designation of PAs… Given the purpose of this study, I think this would be worth clarifying this in the Introduction.

We clarified this in the introduction "These measures combine two types of effects: the effectiveness at selecting as PAs sites of higher-than-average conservation interest (i.e. differences that existed at the time of PA designation); and effectiveness at maintaining species richness and abundance within existing PAs (i.e. differences established subsequently to PA designation)." L.87-90.

L187: I didn't find any mention about the data from first-year observers. How were BBS route-year data from first-year observer treated? Same for data collected during poor weather. According to Kendall et al. (1996), and following other studies using NA BBS data (e.g.

Huang et al. 2014; Wood et al. 2015), records surveyed by first year observers should be removed to minimize observer bias.

Indeed, we had not initially removed these observations. It is now done and specified in the text L.218-221.

L191: “25-mile routes”. I suggest add or replace by the correspondence in kilometers.

We followed the recommendation across the whole manuscript.

L195: Given the landscape characteristics across North America (often one dominant land cover type over broad extents) I am wondering whether this represent a high proportion of routes, i.e. those with multiple land cover types.

As stated in the first answer, full routes were more heterogeneous than route beginnings in terms of land use, and did not represent PAs better.

L213: where does this threshold of 5 years come from? did you perform sensitivity analyses to test this? or was it an arbitrary value?

It is an arbitrary value, chosen to include a maximum of routes (few routes monitored each year) and at the same time have a good sampling quality (if only one year was considered, assemblage would have very low richness). It is now clarified in the text L.228-230.

L241-242: what is the sample size? how many sites classified as within PAs and how many as outside PAs? Looking at the map presented in Appendix S1, it seems there are very few site located in protected areas overall… what’s the proportion of these within each habitat type? This disequilibrium in the number of BBS routes considered as “protected” vs. “unprotected” may affect the analyses.
Also I am wondering what would be the proportion of BBS routes with PAs cover > 50% if the routes were considered in their entirety… The authors may want to have a look at the paper from Wood et al. (2014), as the methodology to define protected sites differ and as such, the ratio protected/unprotected sites does not seem to match (at least the number of protected routes seems way larger in their study).

As stated in the long comment, we acknowledge that sample size in terms of PAs is relatively low but we think that it is inherent in the dataset. About the sample size numbers, they were scattered in the paper and thus not easy to find, we are sorry about that. We now added them in the Landscape data paragraph L.285-287.

L256: I have some concern regarding the temporal mismatch between the bird data and the land cover data. How this delay of ca. 10-15 years may not affect the results? I would suggest using the MODIS-based GLCC instead, for a better match in the temporal extents of the data sets…

Thank you for this very relevant comment. We changed for the 2011 version of land cover as it is the central year of our sampling period. We did not use the MODIS land cover but instead the ESA CCI land cover (http://maps.elie.ucl.ac.be/CCI/viewer/index.php) that has a better resolution. We have applied the modifications in the manuscript L.281-290.

L270-271: argument and reference towards the biases related to altitude and productivity were already presented in the intro, no need for this here.

We removed these sentences L.297.

L291: “fully” vs. “unprotected route” - what does it mean exactly? 100% vs. 0%? Please clarify.

We are sorry for this unclear sentence and we now have clarified it in the text L.319-320. Indeed, we predicted values for 0% protected routes and 100% protected routes.

L305: Were all the BBS routes considered to calculate this human-affinity index, or only the subset of the routes’ sections further considered for the analyses? in the latter case, this may bias the estimation of how sensitive species might be to human disturbance if all survey routes are not gathered to provide such estimate (?), and as such this may not be totally representative of the habitat preferences or sensitivity of each species to human disturbances… This should be at least discussed.

All routes were used in order to avoid biases. We clarified this in the text L.335-338.

L308: What is the purpose of testing these two types of phylogenetic models? Would the Brownian motion model not be enough to test for an effect of the phylogenetic relatedness among species on the response to PA?… Please explain

As suggested, we kept only the Brownian motion model.

L310: what kind of trees? (e.g. primary backbone tree of Hackett) Please specify.

We are sorry for this forgetting. We have specified in the text that we used the Hackett trees of Jetz’s phylogeny L.342.

L311: why choosing to use this specific method (maximum clade credibility) to summarise phylogenetic trees, instead of another common method such as using a consensus tree ? Both methods can be criticized, and depending on the case, they may end on very similar results. However, depending on the number of trees selected, and then depending on the shape of the posterior density, you may not sample the highest probability topology… A consensus tree could have a higher credibility than any of the trees in your selection. Do you have any thought, additional information about this?

This method was used because edge lengths, which are not given by consensus trees, are needed in phylolm models.

L314: if neither a table nor a plot are presented as results, I think estimate coefficients would be appreciated…

We have added the coefficients in the text L.354-360.

L350-L353: this corresponds to result interpretation and should be more relevant I think in the Discussion… To counteract this lack of significance, I think it would be worth considering the entire BBS routes, as I mentioned above already. This lack of significance (and power regarding the analyses) could in somehow raise questions about the robustness - and the relevance - of the results presented here.

As this explanation was repeated in the discussion, we completely removed these sentences L.388.

L423-424: and why is this so? what’s the difference between these two phylogenetic models that could explain results differ… this echoed what I am saying above, and ask for more details about these two models and why they are both considered in this study.

As answered before, we only kept the Brownian motion model.

L427-431: I don't think this is needed here. It is all "theoretical" (would not require the exact same model), and not even feasible/done through this study. Answering this type of question would rather require to consider analysing a metric such as the Phylogenetic Diversity (e.g. Zupan et al 2014; Thuiller et al. 2015). I would suggest removing this section, but I leave it up to the authors…

As suggested, we removed this part L.467.

L432-433: in other words, the species difference in PAEFor relies on phylogenetic relatedness among species

We changed the sentence for "This suggests that much of the effect attributed to habitat preferences under the linear model relies on phylogenetic relatedness among species, which is not surprising as bird habitat preferences and phylogeny are correlated." L.467

L502-505: maybe considering eBird data for example could be useful, even as a joint analyses of both data sets…

We agree this would be very interesting, but would represent substantial additional work, which is beyond the scope of this paper. But we would be very happy to participate if the reviewer intends to work on this question!

WRITING

I would suggest maybe asking for an external English proofread, or at least a very careful reading of the manuscript. According to me, some pieces of the text could definitely be made simpler and easier to read. Need to be consistent in the use of present or past tense throughout the manuscript, and even when referring to the study.

Indeed, they were several mistakes in the previous version of the manuscript due to a lack of attention in the last checks, we are sorry about it. This revised version have been read with high attention, looking for English and phrasing mistakes.

L16: I would suggest rephrasing as "using neighbouring sites to protected areas as "counterfactuals"

We followed the recommendation L.15-16.

L27-28: remove "area the one avoiding human activities"

This paragraph has changed due to a change in the results.

L80: "populations inside these areas" instead of "in them"

We followed the recommendation L.80.

L102: "act as"

This sentence has changed because of other comments.

L104: please rephrase: "to the former if it had not been protected"

We followed the recommendation L.100.

L120-123: need to rephrase. I suggest "Indeed, ignoring this could lead to comparing sites that are not expected to have similar biodiversity regardless of their protection (e. g. protected grasslands or unprotected forests), without neglecting the effect of protected areas by preventing habitat changes (e. g. deforestation or urbanization)."

We rephrased but changed the second part of the sentence, as we felt that the suggestion of Reviewer 2 changed the meaning of the sentence. "On the one hand, not controlling for habitat can lead to comparing sites that are not expected to have similar biodiversity regardless of their protection status (e.g. protected grasslands vs unprotected forests). On the other hand, controlling for habitat type can result in an overlooking of the effects that PAs have on biodiversity by preventing habitat changes (e.g. deforestation or urbanisation)." L116-120

L182: "we use the term"

We followed the recommendation L.183.

L185: "in creating" or "to create"

This paragraph has been rephrased due to other comments L.183-188.

L469: "compared" - past tense

We followed the recommendation L.504. More broadly, we payed attention in our last reading to past use.

L482-484: this sentence is not clear to me… "in relation to pairwise comparisons"??

We changed the sentence to clarify "A main advantage of using large biodiversity monitoring datasets (such as bird monitoring schemes) rather than pairwise comparisons is thus the possibility of applying a well-defined and repeatable control." L.516-518.

L486: I suggest rephrasing: "our results emphasize that it is impossible to clearly measure the effectiveness of PAs in conserving species diversity without defining precisely what is expected of them"

We followed the recommendation L.520-521.

L500: "in North America"

We followed the recommendation L.534.

Appendix S1: "herbaceous" in the legend

We followed the recommendation.