Dear editor,

Thank you for the valuable feedback on our manuscript. We are pleased that you and the reviewer found interest and worth in our manuscript. We addressed each comment and found that the manuscript is indeed clearer now, and we hope to have it approved after your consideration. We have added more citations, context and examples, as well as revised the data retrieval code and added more details in the Methods section. We detail specific changes below (line numbers refer to the clean, revised PDF file).

> *"The reviewer pointed out that some background is lacking in the introduction. Citing Gotelli, Graves & Rahbek 2010 PNAS in the intro could be useful."*
>
> *"[…] relevant prior literature must be referenced in your introduction to provide the necessary scientific basis for understanding. So, please provide more literature references in the first paragraph of your introduction."*
>
> *Lines 7-9: Please provide a short description for each example (taxonomic errors, geographic inaccuracy, or sampling biases).*
>
> *Lines 9-11: Please cite studies that have done this before! If there are no previous studies, it is important to highlight the novelty of your study.*

We have added more context in the introduction section, but we've found it hard to expand the first paragraph. We believe that this particular paragraph needs to be more succinct, at the same time that it provides a complete picture of the paper, framed in a storytelling structure. Then, on the second paragraph, we start providing literature background and many examples, to help the reader understand where our paper stands in the scientific context. The references suggested were indeed helpful; we have included Gotelli *et al*. (2010) in the introduction (L. 28 & 33), Hurlbert

and White (2005) (L. 14, 50 & 66), Hurlbert and Jetz (2007) (L. 14 & 50), Ficetola *et al*. (2014) (L. 15 & 66); we have also added Isaac, Mallet, and Mace (2004) (L.86), Rondinini *et al*. (2006) (L. 86), and Windsor *et al*. (2022) (L. 311 & 325). Finally, we have fixed the IUCN reference, as pointed out by the reviewer in the annotated manuscript (L. 46), and added citation regarding all software infrastructure used and availability of our scripts (L. 158-165). Along with these references, we have included discussions about the quality and accuracy of range maps (L. 46-58), examples of data biases (L. 7-11), and we addressed the uniqueness of our study compared to previous research (L. 13-18). We hope that this structure is clarifying and appealing to the readers.

> *[…] I would add the need to provide more details on how you cleaned GBIF data before actually using it. You also need to provide many more details on data retrieval, which date range?*

Regarding the GBIF data cleaning, retrieval and the validation steps, we added a more detailed description of our process (L. 144-165). Data retrieving is detailed in the source code provided on our repository (code/09-GBIF_data.jl). We used the list of species from the network dataset and retrieved any occurrence within our bounding box (between longitudes -20.0 and 55.0 degrees; and latitudes -35.0 and 40.0 degrees). We changed two things regarding geographical information during data retrieval:

1. We had originally added an argument to capture only the data that had information about the continent location equals to “AFRICA”. However, this constrained our search by thousands of occurrences, and added a filtering layer that could be interpreted as “data

cleaning". We removed that because we don't perform any kind of data cleaning with the IUCN dataset.

2. We added, however, a constraint to exclude occurrences falling within a 4°x4° degrees square around the "Null Island" (see code/09-GBIF_data.jl#L42 on our repository). Rather than a data cleaning step, we see this as a geographic filter complementing the bounding box we used to retrieve data from GBIF and to limit IUCN range maps. This is important because the existence of coordinates is a fundamental part of our analysis, and a (00,00) value can be interpreted as a null coordinate.

Our results changed with that (we provide updated numbers in lines 201-237, and Figure 04), but remained conceptually the same. Mainly, with the inclusion of occurrences of *Canis aureus* out of Africa, we now have 56.2% of its presence pixels within its original IUCN range map (in contrast with the previous number, 0%). However, this proportion after we consider its preys' ranges is still null. Therefore, we still suspect of a taxonomic confusion in this case, since many of the occurrences still appear to refer to other species of *Canis* based on location and due to the mismatch with the preys. The fact that the results remained conceptually the same regardless of the number of occurrences retrieved from GBIF can be interpreted as an evidence of the robustness of our method.

Some taxonomic investigation needed to be done due to inconsistencies between GBIF and IUCN nomenclatures which would block our analysis completely (in the case of *Taurotragus oryx*). However, in the case of *Canis aureus*, as discussed in the paper, we could not track which occurrences or which maps corresponded to which species, which we think is a great illustration of our point. We didn't perform any further data cleaning procedures because this could hinder

the goals of the paper, which was to investigate the characteristics of the data available for research. As we needed to compare two sets of raw data, and data cleaning protocols are dependent on the use objectives, we chose to avoid data manipulation both for GBIF and IUCN datasets.

> *Lines 124-131: I have a doubt about your data validation. First, you mentioned that you compiled point observation data from GBIF. Then you converted that data in the presence or absence of the focal taxon. How did you transform these data in the presence or absence once the GBIF provides just occurrence points (i.e. presence data)? This was not well explained. In general, absence data are not provided because the non-detection of a species in a location does not mean a true absence. This could just be a detection issue. Thus, it is needed to calculate the pseudo-absence data. Please explain this better in your manuscript.*
>
> *[…] the section on geographical data validation needs improvement, by explaining how you produced absence data from presence-only data base.*

For the validation step, we understand our phrasing was confusing, and we have rewritten this section to clarify our method (L. 151-153). Rather than using absence data, what we did was to transform the point data into raster files, and then restrict our range maps to the locations where we had GBIF occurrences. Therefore, the pixels where there are no GBIF occurrences are not considered true absence of a species, but absence of **data** about the occurrence of that species.

Finally, we have addressed all the other minor comments throughout the text as detailed below:

- *Lines 68-70: This mismatch cannot just be a result of the overestimation of the predator's range but it could be the result of misestimating both the predator's and prey's range. Please rephrase the sentence.*

  We've added citations in the Methods section to support our argument that a mismatch between ranges can be originated from different sources (L. 68-70).

- *L. 214-217 can be suppressed.*

  Agreed. We have suppressed lines 214-217.

- *PeerJ uses a structured abstract. I highly recommend authors to adhere to it when resubmitting the revised version.*

  The abstract is now restructured according to PeerJ guidelines.

- *Avoid citing tables and figures in the Discussion.*

  References to tables and figures in the Discussion section were suppressed, with the exception of Figure 05 (L. 305), which is an example of our results to clarify an important point of the discussion.

- *Lines 21-23: Are the cited literature examples of models that do not take ecological interactions into account? Use e.g. before.*

  The cited literature are not examples of models that do not take ecological interactions into account, but studies that have demonstrated how ecological interactions might be responsible for shaping a species' distribution. We've rewritten the sentence to clarify that.

- *Lines 68-70: This mismatch cannot just be a result of the overestimation of the predator's range but it could be the result of misestimating both the predator's and prey's range. Please rephrase the sentence.*

  Lines 84-89: we have rewritten the sentence to include the consideration misestimation of both species in a trophic interaction. We have also added citations and examples.

- We have addressed the grammatical suggestions throughout the text, except for the legend of Figure 1. We thought that the suggested phrasing adds a little bit of confusion to the comprehension of the figure.

We hope our revised manuscript is now more appropriate for publishing at PeerJ. Reiterating our appreciation for yours and the reviewer's thoughtful comments, we are available to answer any further questions.

Best regards,

The Authors