The comments of the Editor are in black, and our responses are in blue.

We really appreciate the positive feedback provided by the Editor on our revised manuscript. We have implemented most of the suggested changes, and we detail our choices a little more below.

The authors have carried out a very substantial revision of the manuscript. The new version now covers related literature and is much more balanced concerning the advantages and disadvantages of the different approaches and metrics for analysing the differences between data- or model-based reconstructions of, in particular, past vegetation reconstructions. The authors also include new analyses based on the suggestion from one reviewer.

The presented method is novel and will likely be adopted and further developed by colleagues working in the field. I have some further suggestions and questions for improving the manuscript. The necessary revisions would be minor.

Major comment:

It would be nice to finish the abstract with a sentence on what has been learned by applying EMD to the previously published study on past simulated and reconstructed biomes for Europe. Just stating that the method has been applied does not demonstrate any new insights. The scientific/knowledge advance of the example application should also be very clear in the discussion.
We have included another sentence to highlight some of the benefits of using the EMD, namely that 1) the EMD allows to reconstruct more gradual biome changes while unary estimates look more flickering and 2) we were able to determine which comparisons were statistically significant from those that were not, while accounting for the multidimensionality of the underlying biome affinity score distributions.

Minor comments and questions:
1.) Is “unary” vegetation estimates really the right term? The term is, at least, not usual, and how the unary numeral system applies to past vegetation reconstructions might not be clear for all readers.
We have looked for other terms since the beginning and settled for “unary” because it was the one that ticked the most boxes. We acknowledge this term is unusual but we believe that our definition of the meaning of this word works in the context of this paper and will enable interested readers to follow.

2.) Abstract: Better “when minor variations in pollen percentages occur” instead of “when minor variations in pollen percentages change”?
We agree that this sentence could have been more clear. However, we replaced “change” by “modify” as we believe it is less ambiguous in this context.

3.) Abstract, “to quantify the mismatch between vegetation distributions available in various

formats": Can it be made clearer what is meant by "vegetation distributions"? Maybe reconstructions of "biome distribution and vegetation composition"? "Available in different formats" sounds like a technicality. Necessary here?
Rephrased as follow: "To overcome this limitation, we propose using the Earth Movers' Distance (EMD) to quantify the mismatch between vegetation distributions (e.g. between distributions of affinity scores)"

4.) Please, check carefully how literature is cited. In the introduction, Allen et al (2020) is cited for determining the global accuracy between data sets, but Allen et al. (2020) mainly addressed the magnitude of past and future simulated biomes shifts, not some kind of accuracy between the model and data, at least not quantitatively. Furthermore, a term like "correspondence" would be more suitable in the sentence than "accuracy".
We replaced accuracy with similarity to better account for the diversity of studies involved in our references. We also split the references to better reflect which reference refers to what.

"(Prentice et al., 1996, 2000). The transformed pollen data can thus be directly compared with model simulations of the same period (e.g. Cao et al., 2019; Prentice et al., 1998) or other pollen data of different periods (e.g. Allen et al., 2020) and the "agreeing" and "disagreeing" pairings (i.e. binary assessments of the compared biome or PFT estimates with the highest affinity score) are counted to determine the global similarity of the compared datasets."

5.) When Allen et al. (2010) is cited the first time, I would make clear that the conical correlation analysis was based on NPP "per PFT" (whereby modelled PFTs were matched with pollen PFTs). This important detail is not clear from the formulation so far. In other words I would add "per PFT" or so.
Yes and no. Based on Allen et al. (2010), the analysis was performed at the biome level (what we wrote). It just happened that the biomes were derived from NPP per PFT. We included that information as follow: "Pollen-based biome scores have also been directly compared to biome scores estimated from the net primary productivity per PFT produced by LPJ-GUESS using canonical correlation analyses (Allen et al., 2010).".