



EGUsphere, referee comment RC2
<https://doi.org/10.5194/egusphere-2022-489-RC2>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on egusphere-2022-489

Louis François (Referee)

Referee comment on "Refining data-data and data-model biome comparisons using the Earth Movers' Distance (EMD)" by Manuel Chevalier et al., EGU sphere,
<https://doi.org/10.5194/egusphere-2022-489-RC2>, 2022

Review of "Refining data-data and data-model biome comparisons using the Earth Movers' Distance (EMD)" by Chevalier et al.

- Does the paper address relevant scientific questions within the scope of CP?

Yes, this contribution looks promising as a new method to analyse pollen data for the past and compare them with vegetation model reconstructions. It is thus directly connected with key aspects of paleoclimate reconstruction and, hence, falls fully within the scope of CP.

- Does the paper present novel concepts, ideas, tools, or data?

Yes. To my knowledge, the application of the EMD to the analysis of pollen data is completely new. This concept is quite interesting.

- Are substantial conclusions reached?

Yes. The authors prove the applicability of the EMD for analysing/comparing pollen data/vegetation model reconstructions. In particular, they show that the EMD has the advantage of conserving more information from the original pollen data. Reconstructions look to be more stable through time compared to classical methods based on the biome with highest affinity score, because these latter methods do not offer a continuous measure of vegetation state (i.e., they are discrete classes).

- Are the scientific methods and assumptions valid and clearly outlined?

Yes, the main method used, the EMD, is well explained and references are provided. Other methods are statistical analyses, which relatively well explained.

- Are the results sufficient to support the interpretations and conclusions?

Yes.

- Is the description of experiments and calculations sufficiently complete and precise to allow their reproduction by fellow scientists (traceability of results)?

Yes, probably. The authors have even developed a R package that certainly helps for such a reproduction of their results, as well for analysis of other pollen data.

- Do the authors give proper credit to related work and clearly indicate their own new/original contribution?

Yes.

- Does the title clearly reflect the contents of the paper?

Yes.

- Does the abstract provide a concise and complete summary?

Yes.

- Is the overall presentation well structured and clear?

Yes, it is very well-organised and generally clear.

- Is the language fluent and precise?

I am not native English-speaker, but language looks fine to me.

- Are mathematical formulae, symbols, abbreviations, and units correctly defined and used?

Yes.

- Should any parts of the paper (text, formulae, figures, tables) be clarified, reduced, combined, or eliminated?

The paper has a reasonable length. All parts look necessary. Description of Test 2 (lines 201-216) could be improved. It is difficult to read, especially the method used to establish the second EMD distribution.

- Are the number and quality of references appropriate?

Yes.

- Is the amount and quality of supplementary material appropriate?

Yes.

Comments

In their paper, Chevalier et al. establish a new method based on the EMD to analyse pollen data (change in space and time) or compare them to model vegetation reconstructions. This method is novel, quite interesting and promise to be of broad applicability. The paper is well structured and very well written. It can be published after very minor revision. I have only a few remarks and suggestions that the authors could consider in their revision:

- The concept of biome is integrative, i.e., it is used to represent (within classes) the overall vegetation present at a given location. So, only one biome should exist at a given location. Thus, the words "dominant biome" should be avoided, and replaced by something like "biome with the highest affinity score" (with the pollen data).
- The authors claim that the biomes are discrete quantities, and for that reason, the methods based on the biome with highest affinity score is presented as less robust than the use of the EMD, which is more continuous. However, with the EMD, the authors use the biome concept and their affinity scores. So, some of the "discontinuities" associated to the discrete definition of biomes still remain, especially when mega-biomes are used as done here. Actually, the EMD method developed here could equivalently be applied using plant functional types (PFTs) and PFT scores rather than biomes/biome scores. For instance, Henrot et al. (2017) (Palaeogeogr . Palaeoclim. Palaeoecol. 467, 95-119, 2017) compared model reconstructions with vegetation data at the level of PFT, i.e., using PFT scores. This allows to keep more information from the original pollen data,

that are provided at the genus level. In this case, biome maps are just created to illustrate/capture vegetation distribution in a single map.

- The EMD method allows to measure a (continuous) distance in the multidimensional space phase with the scores of the different biomes. You can thus show for instance (as in Figure 6) how the distance to a present-day biome has varied in the past. However, the biome phase space is multidimensional and the distance does not tell you in which direction you move. Are you moving towards more forests (and if yes towards which type of forests) or towards more grasslands or deserts? This information is quite important to characterize past vegetation. So, the EMD alone is not sufficient to reconstruct precisely past vegetations. It must be combined with a measure of directions in the phase space. In the method presented here, this role is played by the change in the biome with highest affinity score (so the classical method). But could it be improved to achieve a continuous method to evaluate such directions?

- I have not found many typos. Only on line 436, "toa" does not make sense. I guess the authors want to say: "... as a metric to support ..."

The authors may want to discuss topics (2) and (3) in the discussion section.