Thank you for the opportunity to review for Climate of the Past the manuscript entitled “The climate and vegetation of Europe, North Africa and the Middle East during the Last Glacial Maximum (21,000 years BP) based on pollen data” by Davis B. and coauthors.

The Last Glacial Maximum or LGM is a key past period selected since last decades to provide proxies-model comparisons in the framework of PMIP and following projects. This paper follows different papers which aimed to reconstruct quantitatively the climate conditions in Europe during the LGM from pollen data (eg Peyron et al., 1998, Tarasov et al. 2000, Guiot et al., 1999, Jost et al., 2005; Wu et al., 2007; Bartlein et al., 2011, Cleator et al., 2020). All these papers are based on the same fossil pollen dataset, but on different methods : (1) PFT method with ANN calibration (Peyron et al., 1998; Tarasov et al., 2000, Jost et al., 2005, Bartlein et al., 2011), and (2) methods which take into account the effect of the CO2 on plants as Inverse Modelling (Guiot et al., 1999; Wu et al. 2007) or recent algorithms developed by Cleator et al. (2020) and used in Pini et al., (2022) and Wei et al., (2021).

The objective of this paper is to update these studies and to propose a new synthesis on the climate in Europe during the LGM based on the modern analogue method (MAT). The positive point of this paper is that it’s based on new pollen datasets. It is based on the MPD dataset (8000 modern pollen spectra instead of 800 to 1500 for previous studies), and it also proposes a new fossil dataset (63 instead of 18), which allows to strongly increase the spatial resolution of the results and to better understand the climatic patterns during the LGM from proxies data.

For these two reasons, I think that the paper presents interesting findings in terms of results to be published in Climate of the Past ; however, I also think that it cannot be published in its current version.

My first point concerns the choice of the method to reconstruct LGM climate changes. You have selected the MAT: why? It’s a key point because previous studies have evidenced that it’s not easy to find reliable modern analogues for the LGM vegetation, and that assuming past CO2 equivalent to modern one may induce biases in climate reconstruction (Guiot et al., 2009; Prentice et al., 2017). So why do you use the MAT on your data instead of the IM or the Cleator method which are the only methods to take into account the CO2 changes? I like the MAT for Holocene but I think that the MAT, as used here for LGM, is not appropriate for several reasons.

The first one is that the vegetation of the LGM is mostly steppe, and there is, with the MAT, a possible confusion between warm and cold steppes, which can lead to a bias to the reconstruction of too warm climate conditions for the LGM. There is a method that take into account this bias by distinguishing warm and cold steppes (Tarasov et al 1998, JQS) , but this is not what was used here. Here, the MAT is applied directly to the PFT scores of the undifferentiated steppe biome. I think that the fact of not differentiating the steppes can lead to an important bias in the results obtained in this study. You should add a figure (with the basic statistical tests R2, RMSE) plotting the climate parameters estimated/observed for the modern samples of the steppe biome and see if we have no deviation. We should also add a figure (supp mat ?) with the location of the modern analogues chosen for each of the fossil spectra classified in steppe.

The second one is that CO2 is not really taken into account in this paper. You compare the results obtained here with the MAT with the already published results of Wu et al (2007) based on IM developed by Guiot et al. 2000 (Guiot et al., 2009). I consider that it is insufficient because the datasets used (modern and fossil) are different and therefore hardly comparable. It would be necessary to compare your results with the recent results of Cleator et al (2020, values are available in supplementary mat). The solution that I recommend is to apply the Inverse modelling developed by Guiot on your new datasets presented here, or the algorithm developed by Cleator et al (2020), cf in Pini et al (2022).

If this is not possible, one of the solutions would be to apply to your data a multi-methods approach - WA-PLS or machine learning methods (Random Forest, Boosted Regression Trees)- as often applied now (Salonen et al.,2014; Brewer et al., 2008; Peyron et al., 2013; Robles et al. 2022 ...) to be sure of your results.

Once concerns are addressed, I feel the manuscript will be much closer to being an outstanding contribution to knowledge in this time period, and a key paper to validate model outputs.

Minor points :

Abstract, line 28 “Previous pollen-based climate reconstructions based on MAT show…”: which ones? The MAT has not been often used to reconstruct LGM climate, and the PFTs method cf Peyron et al., Tarasov et al and others references IS not a variant of the MAT, so correct it.

Introduction

-lines 40 to 52 : more references are needed

-lines 62-34: “the pollen-based reconstructions that show the greatest disagreement with climate models have themselves been criticized for not considering the possible effect of low atmospheric CO2 on the physiological relationship between plants and climate (Ramstein et al., 2007)”. The significant bias of CO2 in climate reconstructions for glacial periods must be further explained here, as well as the developed methods that take it into account: inverse modelling by Guiot et al., 2000, 2009; the recent algorithm of Prentice et al., 2017 and Cleator et al., 2020.

-line 65 Inverse modelling, the ref is missing; please add Guiot et al, 2000 (Guiot, J., et al Inverse vegetation modeling by Monte Carlo samgpling to reconstruct paleoclimate under changed precipitation seasonality and CO2 conditions: application to glacial climate in Mediterranean region, Ecol. Model., 1, 119–140, 2000.) and Guiot et al 2009.

-line 100-103 the chronology of the LGM needs to be further explained here as the LGM time window is very close to the Heinrich stadial 1 (17.7 ka) and 2 (23.7 ka).

-line 126: other proxies: which ones? Speleothems?

Methods

-line 177 “more recent studies”: which ones?,

“although the exact record (EPD site #Entity) “: ???

-lines 178-180 “ We estimate that we have excluded 16 of the 17 European sites used by Binney et al. (2017) , 5 of the 6 European sites used by Allen et al. (2010), 28 of the 33 sites used by Cao et al. (2019) and 27 of the 71 sites used by Kaplan et al. (2016)”. So finally, how large is your dataset? How many marine cores ? How is the spatial coverage of these new sites?

-line 194 « The count of *Larix* was amplified by a factor of 10 due to its low pollen representation (Binney et al., 2017)”: why only *Larix*? Other taxa are under or over represented: how do you manage that?

-line 213“ we did not apply this additional procedure and present only the merged steppe biome”: I disagree with that (see my major point) because a possible confusion between warm and cold steppes can lead to a bias in the climate reconstruction to too warm climate conditions for the LGM.

-line 220 “to match fossil samples with modern calibration pollen samples”: the MAT is an assemblage approach which require no statistical calibration, so correct it (the modern pollen samples dataset is not a calibration dataset as it’s the case for the WAPLS for example).

-line 221-223 “This is a similar approach to that used by Peyron et al. (1998) and Jost et al. (2005) who also applied pollen PFT scores to reconstruct LGM climate from pollen data, but who used a neural network technique which is a variant of the standard MAT (Chevalier et al., 2020)”. I disagree with that, there is a confusion here in the principle of each method. The Artificial neural networks used by peyron et al and others studies IS NOT a variant of the MAT. It’s a method close to machine-learning methods, with a real calibration dataset and not easy to check because similar to a black box; in contrast the MAT is very simple, based on an dissimilarity calculation. The only common point is that both methods use PFTs scores to overcome problems associated with the lack of modern analogue but that is all.

-line 242 “The size and distribution of the modern training set in climate and vegetation space is important”: yes, I strongly agree with that, the role of the modern dataset is a key one see papers of Turner et al., 2021; Salonen et al; Dugerdil et al., 2021 for example. I think that the differences in the different climate reconstructions evidenced here are mainly due to the size of the modern dataset.

-line 259 “It was therefore decided not to apply this filter”, so how to you take into account the autocorrelation in your data?

-line 263 A part on the climate parameters reconstructed here is lacking, as statistical tests to be sure that these climate parameters are not autocorrelated; how is calculated the error bars?

-lines 267-272: refs are missing;

-line 312 "Similarly, quantitative climate methods have been applied to individual marine pollen records (Combourieu Nebout et al., 2009; Fletcher et al., 2010)": some key references are missing, as the MF Sanchez Goni team.

-line 331 "In this study we have taken the closest point on land as the modern climate for the calculation of anomalies": better to take a regional temperature range

-lines 337-347 "we did not adjust the pollen assemblage for the over-representation of Pinus in the marine pollen samples" This poses the problem of *Pinus* transport over very long distances in open environments as the LGM vegetation; this is particularly true for marine cores but it is also true for some terrestrial sites. So the question of excluding or keeping *Pinus* needs to be more investigated and tested may be on a site-by-site basis.

-line 363 "such as [site #3] and [site #58]" ; better to give the name of the sites

-lines 377-380 "The main arboreal biomes found at the LGM include Taiga (TAIG), Cool Mixed Forest (COMX), Cool Conifer Forest (COCO) and Xerophytic Scrub (XERO), with just a single occurrence of Cold Mixed Forest (COMX) and Warm Mixed Forest (WAMX). We do not record any Temperate Deciduous Forest (TEDE), Tundra (TUND) or Desert (DESE) biomes at any site at the LGM." Could you explain more the location of the differents biome patterns?

- in the text, many taxa are not in italic: please correct it

- lines 441-443 "The first test was to compare our MAT results with previous pollen-climate reconstructions based on the same LGM sites but using different methods. These previous reconstructions include the neural-network methodology of Peyron et al. (1998) and Jost et al. (2005)". I don't agree, it's not a validation test: not the same method, not the same surface datasets, so we cannot really compare the results. Moreover, the LGM spectra used in previous studies and here are probably not the same, that too can bias the results. OK for me in the discussion but not in this part as a validation test. Same for Wu et al, 2007.

-lines 443-444 "the neural-network methodology of Peyron et al. (1998) and Jost et al. (2005) which we call MAT-NN, as well as the Inverse Modelling approach by Wu et al. (2007) which we call INV." First, the neural networks methodology of peyron et al. is NOT a MAT method, so you cannot call it MAT-NN, it's a non-sense. Second, could you use the name of the method given in the reference papers? Please check, I guess it's the PFT method for Peyron et al and I.M. for Wu et al. which are correct.

-line 472 "We compare the chironomid record with our MAT reconstruction…": you don't compare the chironomid record, you compare the temperature inferred from the chironomid record, please correct it

Samartin et al (2016) not Samaratin et al (2016)

-lines 510-512 « The second consequence of lower seas levels is that terrestrial pollen sites were located further from the moderating effect of the ocean than they are today, resulting in a localised modification of the climate experienced by the site irrespective of regional or global changes.": a ref is lacking

-lines 531-538: “In terms of regional climate, the major ice sheets would have provided significant barriers to westerly atmospheric circulation, or even north-south circulation in the case of the Alps and Pyrenees. As well as representing a physical obstruction, the thermodynamic response of the atmosphere to these high, cold obstructions would have been to encourage the formation of areas of semi-permanent high pressure, similar to those found today for instance over the Greenland ice sheet. In addition, the Laurentide ice sheet located over North America would have generated downstream effects over Europe. These physical and thermodynamic effects would have affected the direction of storm tracks, as well as more local climatic effects commonly associated with ice sheets such as strong katabatic winds.”: refs are lacking

-Line 563: “despite arboreal pollen forming 70-80% of the pollen assemblage”: a significant part of the arboreal pollen is due to *Pinus* which is clearly overestimated in LGM pollen assemblages due to long distance transport in open areas as during the LGM.

-lines 615-616: “expected, areas of forest reconstruct similar or increased precipitation compared to today, and areas of steppe indicate deceased precipitation (see next section).” The CO2 effect on climate reconstruction (see recent papers by Cleator et al. and Prentice et al) is not discussed, please add a part on this point.

-line 618 correct “archaezoological”

-line 669 PMIP = Paleoclimate Modelling Intercomparison Project, not “Palaeo-model Intercomparison Project”, correct it; many key refs on PMIP project are missing: Jost et al., 2005; Tarasov et al …

-line 372: “suffer from the same problems of dating control, unclear provenance and a potentially limited taxa assemblages.” I don’t agree with that, you kept a lot of them for your study.

- line 677: “and the Neural Networks method which is a version of MAT (MAT-NN) “: the method developed by Peyron et al and Tarasov et al is named the PFT method and IS NOT a version of the MAT. It’s a method based on Artificial neural networks close to machine-learning methods, with a real calibration dataset and similar to a black box; both methods use PFTs scores to overcome problems associated with the lack of modern analogue but that is all.

-lines 678-690: see my major concern; I think that the fact of not differentiating the steppes can lead to the warm temperatures reconstructed here with the MAT; please check.

-line 721: diatom not Diatom

-line 730 check “Hughes et al (Hughes et al., 2006)”

-line 755 “19.1 ºC” or -19.1 ºC ?

-line 763 “This compares with -7.2 ºC for our 63 pollen sites”: not sure it makes sense to calculate the mean for 63 sites given the regional climate patterns

-lines 778-784: Good to add a comparsion with the brGDDTs temperature record from Padul (Rodrigo-Gámiz et al., 2022).

-line 806: I think a part on the comparison of these results with LGM model outputs is lacking.

-lines 856-857 “Nevertheless, one of the most consistent signals in our dataset is for an increase in summer precipitation over many areas of Southern Europe and the Mediterranean”. In south Spain, the reconstructed biomes is steppe or xerophytic, with a lot of *Artemisia* and chenopodiaceae: these taxa are characteristic of dry environments (semi-desert), so how do you explain the wetter than today conditions reconstructed?

-check your reference list : Allen et al., 2008 a and b, two refs for Peyron et al 1998 ..

I realize the authors may find my comments difficult to approach, but I sincerely hope they accept them as well-intentioned guidance. It should not be difficult to address them. Once concerns are addressed, I feel the manuscript will be much closer to being an outstanding contribution to knowledge in this time period.