Reply to the anonymous reviewer 1

We would like to thank the reviewer for his/her comments. The remarks have definitely improved the manuscript. Below you will find a point-by-point reply, with the review given in **black** and our reply given in **blue**. We hope that we have answered all questions sufficiently. The line numbers we included in our answers refer to the marked changes document.

Major comments

My major concern about this paper is that it is at least partially based on a comparison between apples and pears. Runoff from the CLIMBER-2 model is compared to sediment records representing a combination of runoff and dust deposition (Ti/Al record of site ODP 967) and dust deposition (ODP 659). Since dust emissions are a strongly non-linear function of ground cover, wind and soil moisture, comparing runoff with dust deposition is hardly justified. They are of course related, during dry periods you expect less runoff and more dust, but their relation is probably far from linear. A direct comparison of runoff with the Ti/Al record is partly justified, because the Ti/Al record is expected to be also a proxy for runoff. This could also be the reason for the higher correlation of CLIMBER-2 runoff with the Ti/Al record compared to the correlation with the dust record at ODP 659. In principle the CLIMBER-2 model output probably includes all variables needed to diagnose the dust emission flux using e.g. the simple model described in Bauer & Ganopolski, 2010. This would allow a more straightforward comparison between model and the sedimentary records presented in the paper.

Non-linearity can clearly be seen when comparing peaks of Dust and Runoff. We thank the reviewer for putting this forward and have added this to the discussion of the results. We will discuss the non-linearity nature of the coupling in more detail, including referring to the Bauer & Ganopolski 2010 paper (thank you for the reference). We do not have the possibility to run the dust model ourselves, because not all CLIMBER-2 output is available to us at this stage. Added changes are:
at lines 232-238: "Although correlation is high for some time intervals, there is a non-linear behaviour between runoff, which results from precipitation and evaporation, and dust peaks from Site 967 and the Ti/Al record from Site 659. High runoff peaks do not always correspond to high dust or Ti/Al signatures in the records. We have illustrated this by comparing the high and low peaks of Ti/Al and the dust records with the corresponding peaks of the runoff records (Figure 8c,d). For both the high (orange) and low (blue) peaks of Ti/Al (Figure 8c) a clear trend is visible. On the contrary, for the dust record (Figure 8d), the high peaks (red) show a more linear trend compared to the low peaks (blue). Nonetheless, correlation coefficients are moderate too low for all comparisons."

And lines 287-289: "The lower correlation with the dust could be expected since dust emissions are a strongly non-linear function of ground cover, wind and soil moisture (e.g. Bauer and Ganopolski, 2010)."

Because of its important effect on both the water cycle and dust emissions, I'm missing a description of what happens to the vegetation over the Sahara and Sahel in the model over the simulation period and how that could have affected runoff and dust and therefore the comparison with the sediment records.

Yes, we understand. This was included in a first version of the manuscript, but the most outstanding result was the link between runoff and the sediment records. We of course agree that it is important to mention. We have included a new figure (now figure 3), which shows the vegetation fraction, either grass, trees or desert, over the two grid boxes. A short discussion is included in Section 3 (line 137 - 141) and the vegetation fractions are now included in the discussion of the results (line 290 - 293):
"The vegetation shows a high correlation for trees and desert with Ti/AL for the Sahel region (-0.734 and 0.781, respectively), and grass and desert for the Sahara region (-0.783 and 0.783). On the contrary, the correlations between vegetation coverage over the two regions is generally poor compared to the dust of Site 659, illustrating the strong non-linear relationship between vegetation and dust outside of the African continent."

A discussion of uncertainties in the forcings is missing. There are for example large uncertainties in the atmospheric CO2 concentration. Proxy reconstructions show a large uncertainty, particularly in the amplitude of 'glacial-interglacial' CO2 variability. The paper by Stap et al. 2016, just to name a model-based reconstruction where two of the authors of this paper are co-authors, shows a very different CO2 trajectory across the Pliocene-Pleistocene transition than that used in the simulations presented in this paper. I'm not saying that CLIMBER-2 should be re-run with all these alternative forcings, but a critical discussion of the possible impact that the choice of a particular forcing could have on the results presented in the paper is needed.

Yes, we agree. There are a couple of CO_2 reconstructions in the current literature that show a different behaviour. The same hold for proxy reconstructions, albeit these are not continuous records. The simulations with the Climber-2 run have been run prior to the work by Stap et al. (as the reviewer pointed out we were involved in both studies) and both come from different methodologies. We have added an additional paragraph in the discussion on this, also referring to a recent paper (still in discussion) which nicely shows a comparison of different CO_2 reconstructions (Figure 6 in Berends et al., CPD, 2020; doi: 10.5194/cp-2020-52). Added text (line 275 - 282):
" The climatic variability in the model is largely determined by changes in the forcing records; NH and Antarctic ice sheets, atmospheric greenhouse gas forcing of CO_2, and orbital variations (Laskar et al., 2004). Both timing and magnitude of the forcing will have an impact on the changes shown by the model. The ice-sheet forcing imposed here is based on a 3-D ice-sheet model constraint by the LR04 benthic d^{18}O stack by Lisiecki and Raymo (2005), which also determined the age scale of the CO_2 reconstructions. Ice-sheet changes can be different, but are constraint by the locations. On the other hand, CO_2 is much less constraint since proxy data over this time period are sparse, and model-based reconstructions can be quite different (see for example Figure 6 in Berends et al., 2020). Particularly, the reconstruction from Stap et al. (2016) shows a much larger amplitude in CO_2, whereas that from Willeit et al. (2019) employs a smaller amplitude over our time period."

Minor comments

lines 27-29: What is meant by 'completely'? There are plenty of other studies that could be cited here, showing that, at least if CO2 is low enough, orbital variations are enough to get pronounced glacial cycles: e.g. Abe-Ouchi et al., 2013 and Ganopolski & Calov 2011.

We meant that it can induce ice sheet growth, but not large ice sheet as seen during glacial maxima. This is the same as for example Ganopolski & Calov (2011) showed, for which large ice sheets do occur but only with low enough CO2 concentrations. We have revised the wording to (lines 28-30):
" Although radiative forcing of orbital variations is too small to force the world into or out of a glacial state with significant ice sheets, they are key to initiate ice-sheet growth and to pace glaciations (e.g. Bintanja and Van de Wal, 2008; Ganopolski and Calov, 2011)."

line 88: what does 'quality' mean here?

This mean how healthy the layer is, i.e. water content and thus prone to grow vegetation on it. We revised the sentence to (line 96-97):
"On the contrary, runoff and precipitation also depends on the water content and amount of vegetation that grows on the upper soil layer. "

lines 104-106: sentence is unclear

The sentence is revised to (line 116-117):
"When forcing records are kept constant we use the present-day ice sheet and a pre-industrial level of 280 ppm for CO_2."

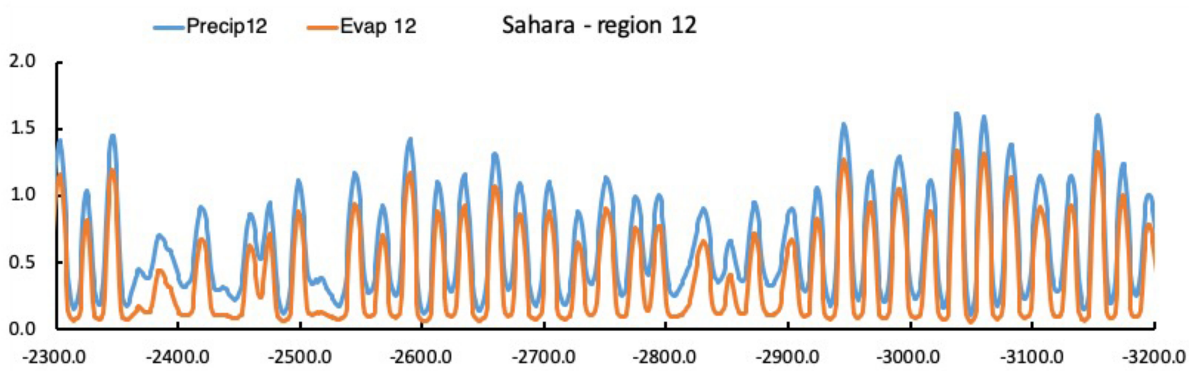
lines 126-127: how has the tuning been done? Moreover, that the LR04 stack has almost no precession for the early Pleistocene.

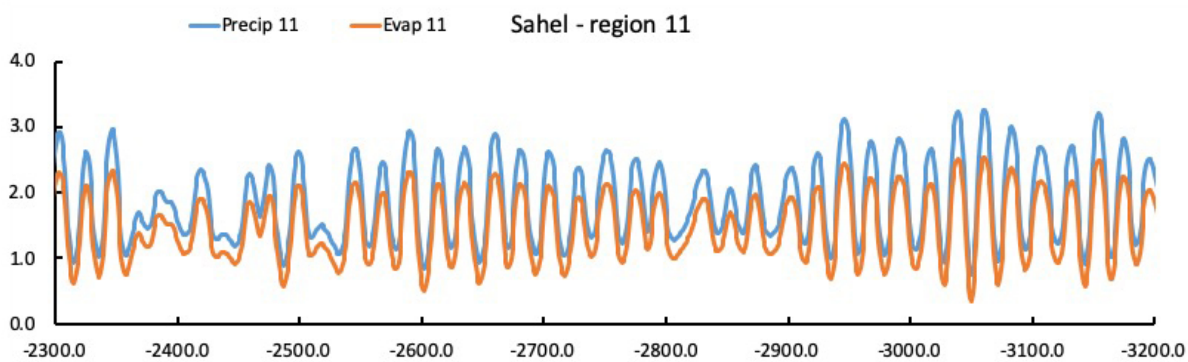
As mentioned in Wang et al., 2010 (where the same data is used), the data for the last 2.6 Myr follow the same age scale, which is actually included in the LR04 stack (for benthic d18O). Between 5.2 and 2.6 Myr ago the data is retuned to minima in the Laskar et al. (2004) 65N insolation curve. This has been added to the text (line 138 - 140).

lines 136-139: Would be interesting to see the time series for precipitation, evaporation and runoff for the two grid cells. Also, what is happening to vegetation in these grid cells? Could it be that the increase in evaporation is related to an expansion of vegetation in the Sahara grid cell? If vegetation is growing over the Sahara I guess that more water should be available to evaporate because roots have access to deeper soil layers...?

We have added an additional figure (now Figure 3) that shows the vegetation fractions over time from 3.2 to 2.3 Myr ago. Figures of precipitation and evaporation are shown below. We have revised the text (line 155 - 162) as follows:
"For the Sahel region the runoff is strengthening following the African summer monsoon, driven by an increase in NH insolation. The increase in precipitation causes an increase in

trees, which replaces desert and grass (Figure 3b). This enhances evaporation, which is stronger than for grassland, causing the peaks in runoff. In contrast, the runoff values of grid box 12 (Sahara Desert) do not increase by the strengthened monsoon, but show peaks of low runoff during precession maxima. Although precipitation is enhanced during the summer monsoon when the air from the Atlantic Ocean reaches land, higher temperatures provide more room for water to evaporate, in combination with an increase of grass cover (Figure 3a) In the case of grid box 12, this additional precipitation is therefore compensated by an increase in evaporation. Also, during precession maxima precipitation is reduced and vegetation disappears, which leads to a strong decrease of evaporation and minima in the runoff."





lines 142-143: is this possibly related to changes in Atlantic meridional heat transport and subsequent changes in the position of the ITCZ when NH ice sheets start to grow and decay?

There is a link there, because increased precipitation is related to a northward shift of the ITCZ, but this holds for precession as well. With the data we have from the model we cannot fully investigate the link with heat transport and obliquity. The figure below shows the obliquity frequency (filtered at 0.0245 ± 0.003) of the four climatic forcing runs: orbit only (O: orange), orbit + CO2 (OG: green), orbit + ice sheets (OI: blue) and orbit + CO2 + ice sheets (OIG: red). The obliquity strength (i.e. amplitude) is mostly equal prior to 3.0 Myr ago for all 4 runs, when ice-sheet are included (OI: blue and OIG: red), there is a clear strengthening of the obliquity frequency in the Runoff. We have added a note on this in the paragraph that follows (line 176 - 177):
" Also, the power of the obliquity frequency of runoff is increased in the OI and OIG relative to the O and OG simulations."

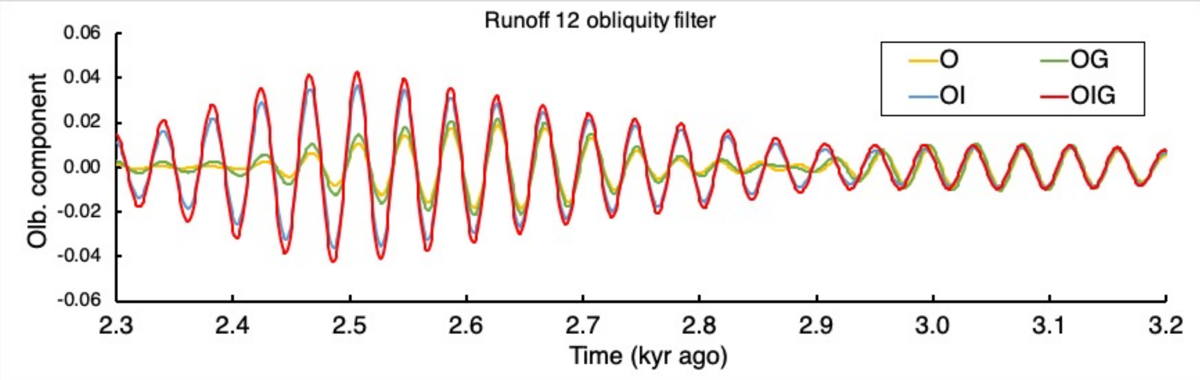


Fig. 3: Please mention in the caption that the y-axis for precession is reversed in 3a. It took me a while to figure out that maxima where actually minima.

This is now Figure 4, this has been added, the same for other figures where it was missing.

lines 149-152: I have read this sentence 10 times, but still do not understand what it means.

We understand, the sentence is changed to (line 172 - 176):
"This shift can be attributed to the imposed lag in the tuning of the LR04 benthic d18O data when calibrating the depth-age scale. The time lag between obliquity (41-kyr) and its related frequency component in the LR04 stack is gradually increased from 3 kyr prior to 3 Myr ago towards 5-6 kyr up to 1.2 Myr ago. This follows from an anticipated slower response time of the growth of larger Pleistocene ice sheets (Lisiecki and Raymo, 2005)."

line 155: 're-tuned age model of Sites 659'. It is not a spectrum of the age model, but of the dust record, right?

Yes correct, sentence is changed to (line 180):
".. and the dust record on the re-tuned age model of Site 659."

line 254: 'we correlation combined': rewrite
Changed to:
"Following, we correlated the combined runoff output of grid box 11 and 12 with the Ti/Al record."

line 257: 'which representing' -> representing
Agreed, removed 'which'

line 259: 'that indicating' -> indicating
Agreed, removed 'that'

Fig. 9d: and how are lags and leads represented? Please add a legend with arrow directions to clarify. Color scale is missing in d.

Yes, the colour scale is missing, but not needed for the interpretation, with orange-yellow within black lines as significant and strong power. Arrow direction is also explained in the caption. We have changed the caption accordingly.

References

Stap, L. B., de Boer, B., Ziegler, M., Bintanja, R., Lourens, L. J. and van de Wal, R. S. W.: CO2 over the past 5 million years: Continuous simulation and new d11B-based proxy data, Earth Planet. Sci. Lett., 439(April), 1–10, doi:10.1016/j.epsl.2016.01.022, 2016.

Bauer, E. and Ganopolski, A.: Aeolian dust modeling over the past four glacial cycles with CLIMBER-2, Glob. Planet. Change, 74(2), 49–60, doi:10.1016/j.gloplacha.2010.07.009, 2010.

Abe-Ouchi, A., Saito, F., Kawamura, K., Raymo, M. E., Okuno, J., Takahashi, K. and Blatter, H.: Insolation driven 100,000-year glacial cycles and hysteresis of ice-sheet volume., Nature, 500(7461), 190–3, doi:10.1038/nature12374, 2013.

Ganopolski, A. and Calov, R.: The role of orbital forcing, carbon dioxide and regolith in 100 kyr glacial cycles, Clim. Past, 7(4), 1415–1425, doi:10.5194/cp-7-1415-2011, 2011.