Interactive comment on "On the linearity of the temperature response in Holocene: the spatial and temporal dependence" by Lingfeng Wan et al.

Anonymous Referee #3

Received and published: 26 February 2019

The linearity of the externally forced temperature evolution during the Holocene is investigated using climate model simulations forced by the total or by individual external forcing factors. In particular, it is tested whether the total forced Holocene temperature variability is a superposition/sum of the individual externally forced temperature responses. Moreover the linearity of the forced temperature response is tested on different spatial and temporal scales. The addressed topic is interesting and important.

Major comments:

- please revise the method section. Sometimes it is not clear what was done and why it was done. Please see specific comments below.

Reply: Thank you for your comments. We have revised the method section substantially, adding much more details and clarifications on the model setup, data processing, bootstrap method, et al. see replies to referee #1 and #2.

- the discussion should be more extensive, in particular the limitations of the study (please see the following remarks)

Reply: Thank you for your comments. The final section has been written, with much more complete discussion on the limitation of the study here.

- only a single simulation for each forcing is available. Therefore, a correct definition of external and internal variability is not possible. The internal variability likely differs between the individual simulations and the internal variability is likely not constant during the individual simulations. By summing up the four individual simulations it is not certain that the internal variability cancels out. Moreover, the internal variability might depend on the time and spatial scale. In addition, the ALL-forcing experiment still includes the internal variability. Please make this more clear in the text and discuss.

- an ensemble of Holocene simulations with that model is not available. Therefore, although incorrect, because the internal variability might depend on the forcing, it might be useful to get an estimate of the internal variability of the different time and spatial scales from a long control simulation with the same model.

Reply: Yes. Section 1 and 4 have been written to clarify this issue. Also, see reply to reviewer #1 on the general questions.

- I am wondering if it makes sense to investigate the shorter time and also partly the regional scales if only one ensemble member is available. The signal to noise ratio on the shorter time and regional scales might require a larger ensemble size to make a robust statement? Using a control simulation - please see previous point - an estimate of the signal to noise rate might be possible.

Reply: Agreed. This is only a rough estimation. See the revised section 1 and 4 on the limitations.

- I am wondering if the following definition is useful: "Since our study above shows that the linear response is largely valid for orbital and millennial variability, but not for centennial and decadal variability, we define the variance of the orbital and millennial variability crudely as the linear signals, while define the variance of the sum of the centennial and decadal variability, which is dominated by internal variability, as the linear noise." Please comment.

Reply: Given the single realization we have, there is no precise way of separating signal and noise. In this particular case, since all the four forcing factors are at orbital and millennial time scales, the forced signal should be in these long time scales, and the noise should be at shorter time scales, if linear response is assumed (which is largely confirmed). So, this gives a rational to for our crude estimation of signal and noise. If, for example, we discuss volcanic forcing and solar variability, this separation of signal and noise is no longer effect and an ensemble is necessary. This has been discussed now in the revised section 1 and 4.

- Laepple and Huybers (2014) have shown that "a multiproxy estimate of sea surface temperature variability that is consistent between proxy types and with instrumental estimates but strongly diverges from climate model simulations toward longer timescales. At millennial timescales, model-data discrepancies reach two orders of magnitude in the tropics, indicating substantial problems with models or proxies". Please discuss the implications in the context of the findings

Reply: A good comment. Our conclusion is valid only for this model. If the model internal variability is indeed so much lower than in the observation, the implication of this study to the real world will be limited. This point is added now in section 4. It is an interesting issue to be explored in the future.

- please describe the filtering method in more detail. It is not clear to me what kind of polynomial was used for the LOESS. Moreover, it is not clear whether the authors used several iteration to get more 'robust' estimates. More important, what is the influence of the LOESS-filtering method on the result, in particular on the linearity of the response.

Reply: The filtering is discussed in more detail now. LOESS is used here only as one low pass filter. We use this to be consistent with Marsicek et al (2018). (Our original motivation is to interpret the millennial variability found in Marsicek et al). Marsicek et al also verified the locally weighted regression (Loess) by generalized additive model (GAMM) fit. We test

some of our results simply using running mean and the results remain qualitatively similar.

- please describe the method - used to compute the significance of the correlation – in more detail. If I understand the authors correctly, an AR1 process is only fitted to the ALL-forcing simulation on the different time scales. The Monte-Carlo method is then used to produce an ensemble (PDF) of fitted curves. Then the correlations between the fitted curves and the ALL forcing run are computed and the 95% confidence level is determined afterwards. If I understood the authors correctly, I am wondering if this method is sufficient. I would think that an AR1 process has to be fitted to the ALL forcing run and the superposition (sum of the response of the four individual simulations). Then two ensembles - one for the ALL forcing and one ensemble for the superposition – have to be computed using the Monte-Carlo method. The correlations between these two ensembles have to be used to determine the confidence level. Please make also more clear why you choose the AR1 as a benchmark and how robust the parameter of the AR1 process is, in particular for the orbital time scale.

Reply: We think that the randomization on ALL should be sufficient. This is because the key here is to use randomization to destroy the serial relation between ALL and sum. This can be done by randomize either ALL or sum, or both of them. Indeed, we have tested both cases, randomizing one or both time series and confirmed they are the same.

The reviewer is correct in that, strictly speaking, the AR(1) coefficient should be different for each region and should be used for the test of significance. Here, we used the global mean as a common test, mainly for simplicity. Most importantly, our focus here is on the linear response features over the globe, between different regions. Therefore, a common test makes it easy for comparison among different spatial scales and regions. For example, if regional tests are performed, it will be impossible to plot the significance test on the summery figure of Fig.3 and Fig.4 for comparison of different spatial scales. Similarly, it will be hard to compare the value as well as the significance among different regions and spatial scales in Fig.5 and 6. In addition, the global mean AR(1) is meant as a crude representation of most AR(1)'s for different regions. Indeed, except for the orbital scale, the global mean AR(1) is larger than most of the regional AR(1) so that the global mean AR(1) serves as a stricter test. At the orbital scale, the global mean AR(1) is about the middle of the regional AR(1)'s. Finally, we did emphasize that, if one's focus is on a specific region, the regional AR(1) should be used for re-evaluation of the significance. These points are now discussed explicitly in section 2.2.

- it is not clear to me why the authors did not do a spectral analysis of the runs like e.g. wavelet analysis, power spectrum, cross power spectrum ...

Reply: Our study is a first preliminary study. Our interests here is mainly on the linear responses on slow time evolution at the orbital and millennial scales. Given only 11,000 years, it is difficult to derive spectral details with high significance. Nevertheless, we agree it will be interesting to explore the spectral features in the future.

- why was the analysis based on the model grid and not on climate modes using e.g. EOF analysis?

Reply: Fixed region is more practical for using model to interpret the real world proxy. Our original motivation is to interpret the regional climate response over North America and Europe as discussed in Marsicek et al (2018). For overall climate response in the model, it is a good idea to perform this in the EOF space.

Minor comments:

- please be more precise (whole text): please rewrite sentences like 'the linear response is strong' => the response is almost linear; the response is similar to that of a linear system

Reply: Thank you for your comments. We have attempted to clarify these terminologies.

- whole text: I would prefer: forcings => forcing factors

Reply: Done!

- page 3, line 8-9: Please rewrite the sentence

Reply: We have deleted it here and explained the data processing in much more details later in 2.2 as follows:

"To the time scale, we decompose a full 11,000-yr annual temperature time series (from 11 ka to 0 ka) in 100-yr bins (a total of 110 data bins, or points, each representing a 100-yr mean) into three components. The three components are to represent the variability of, roughly, orbital, millennial and centennial timescales. Following Marsicek et al. (2018), we derive the orbital and millennial variability using a low-pass filter called the locally weighted regression fits (Loess fits) (Cleveland, 1979). First, the orbital variability is derived by applying a 6500-yr Loess fit low-pass filter onto the temperature time series, and therefore contains the trend and the slow evolution longer than ~6500 years. Second, we apply a 2500-yr Loess fit low-pass filter onto the temperature time series; then, we derive the millennial variability using this 2500-yr low-pass data subtracting the 6500-yr low-pass data. Finally, centennial variability is derived as the difference between the 100-yr binned temperature time series and the 2500-yr low-pass time series. In addition, we also derive a decadal variability time series. First, we compile the 10-yr bin time series from the original 11,000-yr annual time series (of a total of 1,100 data points, each representing a 10-yr mean). Second, we apply a 100-yr running mean low-pass filter on the time series of the 10-yr binned data. Finally, decadal variability is derived by using the 10-yr binned time series minus its 100-yr running mean time series."