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

Oliver Bothe (Referee)
ol.bothe@gmail.com
Received and published: 22 February 2019

General Comments

Wan et al. use the TraCE-21ka simulations in their manuscript "On the linearity of the temperature response in Holocene: the spatial and temporal dependence" to provide an initial assessment of the linearity of the climate response to various assumed forcings. They study this separately for a number of spatial and temporal scales. The idea behind the manuscript can result in a valuable contribution to our understanding of past and future climate changes, to assessing paleo-simulations, and to studying paleo-observational records. I do not have real major concerns but I think various clarifications and additional discussions are necessary before the manuscript could be accepted for publication. These clarifications should be re-evaluated by a round of revisions and therefore I, nevertheless, recommend major revisions.

Specific Comments

1: Could the authors please discuss more clearly, why they think that the assumptions on potential linearity hold (see also the major comments by anonymous referee 1). This discussion could also include, how the specific setup of the TraCE simulations hampers or supports the approach. It may help to include in these discussions a priori knowledge/references on forced and internal variability across temporal and spatial scales.

Reply: This point has been discussed in much more detail in the revision. Section 1 and section 4 are written. More clarifications are also given now. See the reply to referee #1, general comment (1).

2: Similarly, I think it is necessary to discuss, at least shortly, how the simulations implement the various forcings and how this may influence the results.

Reply: Thank you for your comments. We have added substantially more details on the simulation in subsection 2.1. We also added a paragraph here on the experimental design and its usefulness for linear response study. "It should be noted that the linear response can't be assessed if the individual forcing experiments are performed with the forcing superimposed one-by-one. In this approach, the four external forcing is added sequentially, for example, first the ice sheet, second the ice sheet plus orbital forcing, third the ice sheet, orbital and GHGs, and finally, applying all four forcing of ice sheet, orbital, GHGs and

melting water. In this experimental design, the full forcing response is by default the response of the sum response after adding the four forcing factors together, and therefore can't be used to test the linear response. It should also be noted that our four individual forcing experiments, although in principle feasible for assessing linear response, are not designed optimally for the study of Holocene climate. This is because, except for the variable forcing, all the other three forcing factors is fixed at the 19ka condition. As such, the mean state is perturbed from the glacial state, not a Holocene state. This may have contributed to some unknown deterioration on the linear response discussed later. Nevertheless, we believe, our major conclusion should hold approximately. This is because, partly, the response is indeed almost linear for orbital and millennial variability as will be shown later."

3: Could the authors please discuss, why correlation coefficients and the linear index are appropriate measures of the linearity of the responses as studied. Could they please also clarify, which information is added by the linear index and how to interpret the index in this context.

Reply: see Reply to specific comment 5 to referee #1. The correlation represents the similarity of the ALL and SUM, but can't evaluate the absolute magnitude of the two responses. Even if two time series is perfectly correlated, their magnitudes can differ by an arbitrary constant. The linear error is to reflect the magnitude of the relative error between the ALL and SUM. More clarifications are added in the text in section 2.2 on this. "However, the correlation does not address the magnitude of the response. Even if S_t and T_t has a perfect correlation r=1, the two time series can still differ by any constant factor in their magnitudes. Therefore, we will also use the linear error index L_e to evaluate the magnitude of the linear response."

4: Is the linearity assumption even valid for time scales where internal model processes are known to dominate. That is, we can be quite certain a priori that the decadal scale will be dominated by internal climate over the last 11k years.

Reply: With our approach of single realization, the linear response assumption will fail if the internal variability is dominant. Then, a large ensemble is needed to suppress internal variability. This is discussed in detail in the reply to referee #1, general comment (1).

5: Could the authors please stress that their conclusions really only hold for the specific setup of the TraCE simulations used.

Reply: Thank you for the reminder. Yes, we have been specifically clear about this in a sentence and discussion in section 4. "The result here represents the first such assessment and is carried out for a single variable (surface temperature) in

a single model (CCSM3). It should therefore be kept in mind that the assessment could differ for different variables, in different models, for different periods and for different sets of forcing factors. For example,…"

6: More generally, I think the manuscript is missing a dedicated and thorough discussion-section.

Reply: Thank you for your comments. The last section has now been rewritten and substantial discussions are added. Section 1 is also written to address many potential ambiguities.

7: Page 1 Line 25ff (P1L25): I do not think the authors show this causality conclusively. Anyway, the SNR plots do not show or are even intended to show, according to the manuscript, why linearity is consistent between millennial and orbital scales, but why linearity is strong in these regions. That is, in my understanding, the manuscript does not support this sentence in this form.

Reply: We think the referee is referring to this sentence: "On the millennial scale, the linear response is still strong in the NH over many regions, albeit weaker than on the orbital scale". The reviewer is correct in that the paper is not to show millennial and orbital are consistent, it is only to show the linear response in different regions for each time scale separately. The comparison is only made by the comparison between the two in Fig.5, instead of SNR in Fig.6. Therefore, we think this statement is valid.

8: P2L22ff: Could the authors please be more specific how this answer can benefit our understanding?

Reply: The linear response is the base for attributing the response to each forcing factors. Indeed, if a regional response is far away from a linear response (that is the sum response is far away from the total response), the attribution of forcing factor becomes not very useful, because there will be large amount of total response that can't be attributed to any forcing. Indeed, our original motivation for this work was to understand a specific climate response in a particular region (Europe + North America) at millennial time scale. But, we realized that we didn't even know if this response is a good linear response to the four individual forcing in the first place. The work is then expanded systematically to all the regions and at all timescales. We have added more explanation on this point in the new section 1: introduction and section 4: summary and discussion.

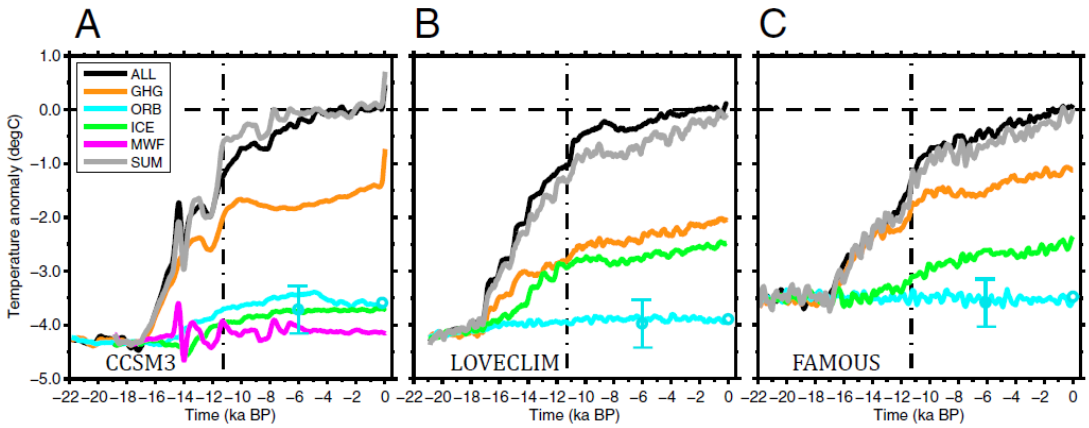
9: P3L3: I do not think the results by Shakun and colleagues or Marsicek and colleagues allow statements about how reasonable CCSM3's climate sensitivity is.

Reply: We have relaxed the sentence to "suggesting a potentially reasonable

climate sensitivity in CCSM3, at global and continental scales." We think this is a reasonable statement.

10: P3L7ff: I think the last part of this sentence requires a reference.

Reply: Thank you for your comments. The reference is Liu et al., 2014, Figure 2A. (We have changed the citation more specifically as: "Figure 2A of Liu et al., 2014"). From this figure we can see in the Holocene period have small change by CO_2 than in the deglacial which warming response is dominated by the response to CO_2. Since the ice sheet retreating is also an important warming effect in the deglaciation in LOVECLIM and FAMOUS, we have also added this factor in the sentence "as it removes the deglacial warming response that is dominated by the response to increased CO_2 and ice sheet retreat".



11: P3L8: I am not sure, I completely understand how the authors perform their binning.

Reply: Thank you for your comments. The binning simply means we use the 100 year mean data as one data point. So, the 11,000 years of data is binned into 110 data points, each representing the average of 100-year. Partly, this binning is to be consistent with Marsicek, et al (2018). We have made some clarification on this and sometimes used "mean" to replace "binning".

12: I find much of the method description on page 4 unclear. For example: I am not fully clear how the authors produce their various time series.

Reply: We apologize for the ambiguity. We have rewritten this part. Basically, the Loess fit is a low-pass filter. The filtered data is therefore described as low-pass filtered data. This paragraph of deriving the various time series have been written.

13: P4L19: Is an AR1 process appropriate or could other models be more appropriate?

Reply: We use the AR(1), instead of white noise, is because we test against the reduced degree of freedom in the low pass the data.

14: P4L23: Could the authors please give more details on their bootstrap. From my point of view, this description does not allow for reproduction of the significance tests.

Reply: Thank you for your comments. Take the 100-yr binned data for the Holocene for example. The ALL run global mean temperature time serial have 110 points of data, each representing a 100-yr bin. For one realization, the order of the data is swapped randomly. Then, the sum is used to compare this realization once to derive one Le. Since the randomly swapped realization is not related to the sum response, one should expect a large error Le. Here, we perform the random realizations for 1,000,000 times. This gives us 1,000,000 values of Le, forming the PDF of Le values. The minimum 5% level is then used as the 95% confidence level. One can reproduce the results using the matlab function bootstrap. An introduction on bootstrap is given in the reference Efron, 1979 or Wikipedia (https://en.wikipedia.org/wiki/Bootstrapping_(statistics)). These detailed explanations are added in the revision.

15: a) I understand that the author's interest is only in the linearity of the response. However, considering Figure 1, I think, the manuscript will benefit if the authors also discuss the visual discrepancy between the SUM and the ALL series. A difference of about 1K between both series between 11K BP and 3K BP is a relevant feature. Indeed, even if the response is largely linear according to the correlation, the potentially smaller nonlinearity appears to be more important here.
b) To add on this, I wonder whether the setup of the simulations really allow for the analyses? However, I am not familiar enough with the setup of the TraCE "single" forcing simulations.

Reply: Thanks for this good point. Indeed, the assessment of linear response depends on the time period and is a "global" measure here. It does not exclude the discrepancy at some times when the nonlinearity or internal variability can be large. Some comments on this point has been added in the discussion on Fig.1. "It should be noted that, the goodness of the linear response is based on the entire period and is meant for the response of the time scale to be studied. Therefore, even for a good linear response at long time scales, the sum response may still differ from the total response significant at some particular time. For example, for the orbital scale response in Fig.1b, even though the linear response is good according to the correlation and L_e, there is a 1^{o}C difference between the sum and total at 11ka and 3ka. Therefore, for the orbital scale response, the linear response mainly refers to the trend-like slow response comparable with orbital scale, instead of response features of shorter time scales." The dependence of the linear response assumption on time period is also discussed in

section 4.

More descriptions of the TraCE-21ka simulation setup have been added. The relevance for the assessment of linear response has been discussed in the reply to question 2.

16: P5L28: Could the authors discuss this complexity in more detail, please?

Reply: Thank you for your comments. What we mean is that the goodness of linear response depends on the region and time scale. This sentence has been changed to: "This suggests that the goodness of the linear response depends on both the region and time scale. This further highlight the need to study the linear response at regional scales."

17: P6L5: As far as I can see the authors do not discuss the reason, at least not in depth.

Reply: Thank you for your comments. The reason is discussed in section 3.3. For orbital variability (Fig.5a1-a3), the linear response is strong in most regions in the NH across all three spatial scales, with the correlation coefficients above 0.8. In the SH, the linear response is also strong over the continents, but is poor over the ocean. This leads to the significantly reduced linear response in the SH as discussed in Fig.3a-4a. Since there are more continents in NH than SH and generally speaking, the linear response in continents strong than oceans, the linear response in the NH is better than SH. This is only an explanation from one angle, certainly not a full explanation. We have modified this sentence to: "Part of the reason of the stronger linear response in the NH than over SH will be discussed later"

18: P6L8: Could the authors please be specific, why this should be treated with caution.

Reply: Usually, the linear response becomes better for larger spatial scale, because the large spatial average suppresses noise (internal variability). This case is the opposite. A note is added on this.

19: Considering the centennial time scale: the authors diagnose that the linear response on centennial scales is poor. However, at least in the correlations in Figure 3, there appear to be many regions where linearity still is of modest importance. That is while I agree with the assessment that there is "no strong" linearity, the authors appear to dismiss linearity on centennial scales to easily on page 6.

Reply: Our statement of "no strong" linearity is derived the statement on the centennial variability: "The median linear response on the centennial timescale

in either hemisphere across spatial scales (*f>3*, Fig.3c and Fig.4c) is no longer significant, with few correlation coefficients larger than 0.3 and contributing less than 10% of the variance." We are not very clear what the referee is inferring here.

20: P6L32: The description appears to exclude the continent of Australia.

Reply: Thank you for the careful observation. A note is added on Australia.

21: P7L1: a) Could the authors please discuss later on, why the response over the continents should be different from the oceans on these very long time scales. b) Could they please also discuss what the strong internal variability over southern oceans implies for reconstruction efforts.

Reply: Good questions! We don't know the reason. A comment is added. We plan to further explore this in the future.

22: Could the remaining ice sheets and the last freshwater forcing implementations influence the results generally and specifically the poor linearity over North America?

Reply: Again, this is a good question. We plan to further explore the physical mechanism of the linearity response in the future.

23: P7L18: It suggests so for this set of simulations.

Reply: We added "in this model". A general note on the dependence of our results to model, time period, climate variable, et al, is also added in section 4.

24: P7L29: The authors write at a number of instances "North America" but the results differ notably within North America if I interpret the visualizations correctly.

Reply: Thank you for the careful observation. We have changed North America to Canada.

25: P7L31ff, P8L11ff, Figure 7: I am not sure whether these parts add anything to the other analyses.

Reply: This figure is meant to give some intuition of the scatter of the relationship between correlation and SNR.

26: P8L6: I appear to be unable to see the poor SNR over southern North America.

Reply: We clarified it now as "the North America continent outside the central North America".

27: Figure 1: Are the linear errors really the same in panels (a) and (b)?

Reply: Thank you for your comments. Yes, they are same. I have check it. The Le of Fig.1a is 0.632 and Fig.1b is 0.626. So they have different in the third decimal. But in this paper we only keep two decimal.

Technical Comments

T1: If I understand it correctly, the manuscript will receive language editing by Copernicus if it is accepted. Nevertheless, I think it will help further reviews if the authors check the language everywhere for clarity and grammatical correctness.

Reply: Thank you for your comments. We have gone through the manuscript carefully several times.

T2: Some of the Figures (particulary Figures 3 to 6) are not publication ready. While I assume that Copernicus is going to assist the authors in this if the manuscript is accepted, it probably would shorten the time between submission and final publication if the authors improve on the Figure-quality for the next round of reviews already.

Reply: We have improve the figures.

T3: Could the authors please check that all Figure captions are correct. I was not sure.

Reply: We have checked it again.

T4: P4L20: I think "valid" is not the correct expression, here.

Reply: We didn’t find valid in P4L20, but in P7L20. We have change the wording.

T5: P8L2: Could the authors please skip the exclamation mark.

Reply: Thank you for your comments. All right, fixed it.

T6: Acknowledgements: I think the authors have to acknowledge the repository or the persons which/who produced and provided the data. (I assume this was the Climate Data Gateway at NCAR.)

Reply: Thank you for your comments. All right, fixed it.