Reply to anonymous reviewer comments, 24.2.2014, concerning the manuscript of Rehfeld, Trachsel, Telford & Laepple in discussion for Climate of the Past (doi:10.5194/cp-2016-13)

04/04/16

Summary

We would like to thank the reviewer for his/her detailed comments which will help to improve the clarity and quality of the manuscript.

The reviewer mainly comments on limitations of our study related to the coarse representation of the simulated vegetation and the simulated climate. These are valid points which we are aware of. We will ensure that they are more properly represented in the revised manuscript. However, as we demonstrate below, the reviewer's points are overstated and the limitations brought up by the reviewer do not affect our main conclusions.

Reviewer's comments are given in grey. Emphasis in *italics* was added to highlight main points.

Point 1: Number of taxa used in the study

> **Reviewer's comments:**
> The authors use a vegetation model and climate model to simulate the process of reconstructing climate from pollen data, and in turn to assess the ability of pollen-based methods to accurately reconstruct seasonal Holocene climate change.
> This is a interesting and novel approach, and although similar virtual experiments have been conducted with other proxies, this is the first time that I know of where it has been applied to pollen. Pollen-based climate reconstructions have been widely used in data-model comparisons, and large discrepancies have been found between these reconstructions and climate model simulations during the Holocene, particularly in terms of seasonality. Investigation of potential seasonal bias in pollen-based reconstructions is therefore of particular interest and importance.
> *The study is generally well written and presented, but has a number of critical issues that I do not think can be easily resolved. The most obvious of these is the unrealistically low number of virtual taxa, or in this case PFTs, used in the transfer-function.* To some extent the authors themselves acknowledge this (lines 550-554) “Furthermore, the methods we have tested are limited by the low number of plant functional types, as large-scale PFT-based pollen reconstructions use roughly 2-3 times the number of PFTs (as e.g. in Davis et al., 2003; Mauri et al., 2014).” *This is actually an underestimate, since of the 8 PFT’s used by Rehfeld et al, 3 are tropical (tropical evergreen trees, tropical deciduous trees and C4 grasses), leaving just 5 PFT’s for the extratropics such as Europe.* Davis et al. 2003 and Mauri et al. 2014 use 22 PFT’s for Europe, which is more than 4 times the number used by the authors in their study. *Taxa based pollen-climate transfer functions commonly use upwards of 50-60 taxa.* These numbers are important because the individual behavior of the PFTs/Taxa and their climatic tolerances constitute the degrees of freedom necessary to reconstruct multiple climatic variables, and particularly those that may show close co-

variance as cited as a potential problem by the authors.

As we state in the manuscript, the number of Plant Functional Types (PFTs) in our model study is lower than what is generally used in a real-world large-scale reconstruction exercise (eg. in Davis et al., 2003, Mauri et al., 2014, Mauri et al., 2015). We use 9 PFTs (one of which is representing bare soil, or desert fraction), whereas e.g. Mauri et al. (2014) use 22 (two of which are virtual). In a given region, the number of contributing PFTs is lower, as some PFTs only appear in some region, thus leading the reviewer to conclude that the difference between our modeled and the real world PFTs is up to factor of four.

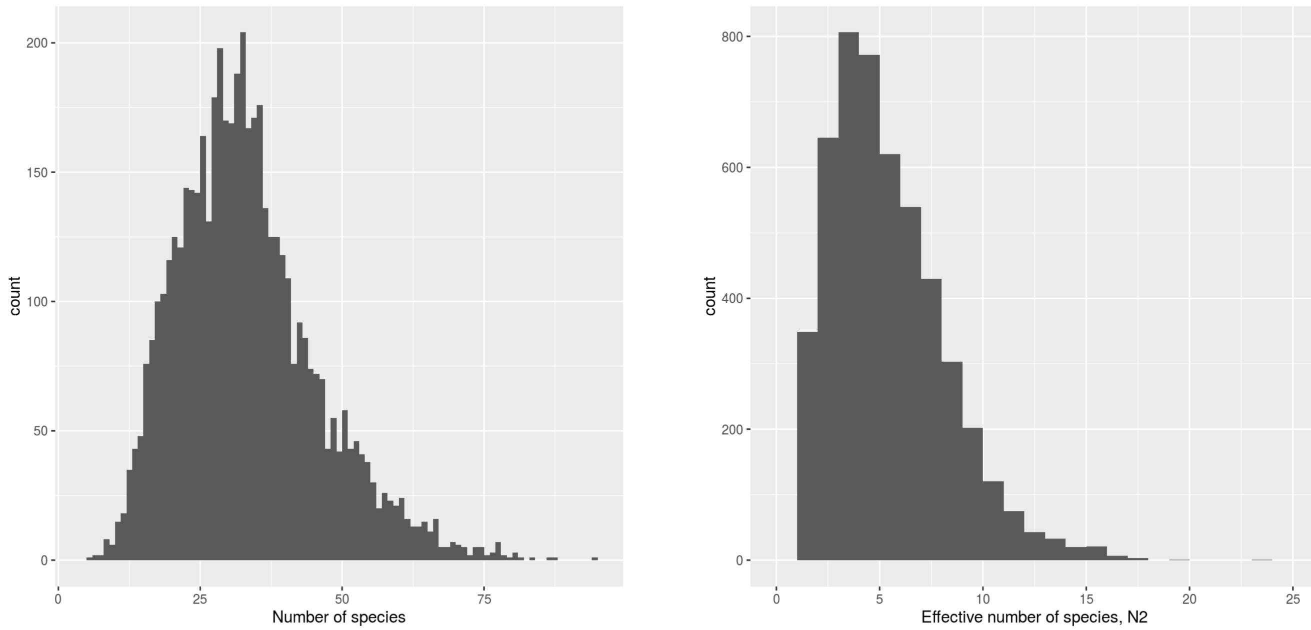
However, what is ultimately relevant for the calibration and reconstruction efforts is the information contributed by the PFTs or taxa; in other words how many of them actually contribute to the pollen (or PFT) diagram in a relevant way. This strongly differs from the number of PFTs/ the number of taxa.

The effective number of species can be quantified by the Hill's number N2, which is an entropy-based measure for the vegetation diversity (Hill, 1973). An analysis of 4990 sites in the European modern pollen database, which formed the basis of Mauri et al. (2014, 2015) and many other large-scale pollen-based reconstructions shows that the median effective number of species (N2) is 4.9, much lower than the median palynological richness of 31 taxa (Fig R1 below) . If the pollen data of the European Pollen Database were assigned to PFTs we would expect the N2 for the PFTs to be even lower.

In our study we only use sites (grid points) with a Hill's number larger than 2 (p.4, l. 19). Of the 458 grid points we use, the median N2 for the fossil data (shown in Fig. 1d in the manuscript) is 2.9, and for the modern calibration data it is 2.7. Therefore, although the number of PFTs is much lower in the model, the diversity and effective number of species is not much lower than that in actual pollen-climate reconstructions.

In the revised version of the manuscript, we will discuss the differing number of taxa as well as the difference in the effective vegetation diversity.

Figure R1: Distribution of the number of species (left) and effective number of species (N2) in modern samples in the European Pollen Database.



Point 2: Winter vs. summer temperature reconstructability

> Furthermore, the PFT's used in the study by Rehfeld et al. have extremely broad climatic tolerances (deciduous trees, evergreen trees, grass..) that can be expected to have little diagnostic power. No pollen-climate transfer function should or would be based on such a low number of taxa/PFT's with such broad climatic sensitivity, and it is therefore disingenuous of the authors to compare their own over-simplified approach with the approach used in actual pollen-climate reconstructions. For instance the authors infer that because they were unable to reliably reconstruct winter temperatures, this should also be a problem for actual pollen-climate reconstructions. In reality, the problem with winter temperatures is just as likely to be a result of the authors over-simplified experimental design and the use of a limited number of PFT's with limited winter temperature sensitivity.

In our study we have followed the standard workflow for pollen-based reconstructions. The reconstructability, or non-reconstructability, of climate variables is often inferred from the transfer function r2 and the RMSEP in cross-validation (e.g. in Mauri et al., 2014, Frechette et al., 2008). These test diagnostics are based on the modern calibration data alone. As we show in Fig. 7 in the manuscript, the transfer function estimated r2 for winter temperatures is similar to that for other temperature variables. Therefore, the transfer function diagnostics suggest, that winter temperature is reconstructible. We agree with the reviewer that the true reason for low *actual* reconstructability of winter temperatures may well be that winter temperatures have little influence on the modeled vegetation (which might be realistic in at least in some regions of the world, such as Siberia).

However, even if the winter sensitivity of the model vegetation were unrealistically low, this would only strengthen our conclusion that transfer function diagnostics based on modern calibration data alone are not sufficient to characterize reconstructability.

Point 3: Resolution of the calibration climate dataset

> This problem is likely to be compounded by the use of climate data for calibration from a climate model with low spatial resolution, and where the spatial variability of climate is highly smoothed compared to the real world. On the one hand this reduces the variance of climate and vegetation in the training set and on the other, it greatly increases the propensity for spatial auto-correlation that the authors also highlight as a problem in their study.

We of course agree that the spatial climate fields from our climate model simulations have a much lower resolution than for example the 0.5 minute resolution interpolated instrumental dataset (Hijmans et al. 2005) used in Mauri et al. (2015).
However, the resolution in itself is neither the determining factor for the variance of the climate explanatory variable, nor for the spatial autocorrelation. Most importantly, the covariance structure between the different climate variables (e.g. summer and winter temperatures), is not directly a function of the resolution. Given the large-scale structure of spatial climate and especially temperature variations (e.g. Hansen, & Lebedeff, 1987) we do not expect a strong influence of the resolution, except in some areas where elevational gradients, not well represented in the coarse model topography.

This is demonstrated in the Figure R2 below, which compares the distribution and relation of the gridpoint winter and summer temperatures in the 0.5 minute resolution temperature field (Hijmans et al. 2005) with the gridpoint temperatures from our low resolution ECHAM5-MPIOM simulation in Europe (30-60N, 0-30E) including several mountain ranges. The 61 land grid-points from the climate model cover most of the phase-space spanned by the 13 million grid points of the 0.5min resolution field, except high-altitude regions represented by the lower-left tail. The model field further shows a similar correlation between the seasons. Thus we see no reason to expect that the low resolution would bias our results towards less skill in reconstructing multiple variables. We will include a detailed discussion in the revised manuscript.

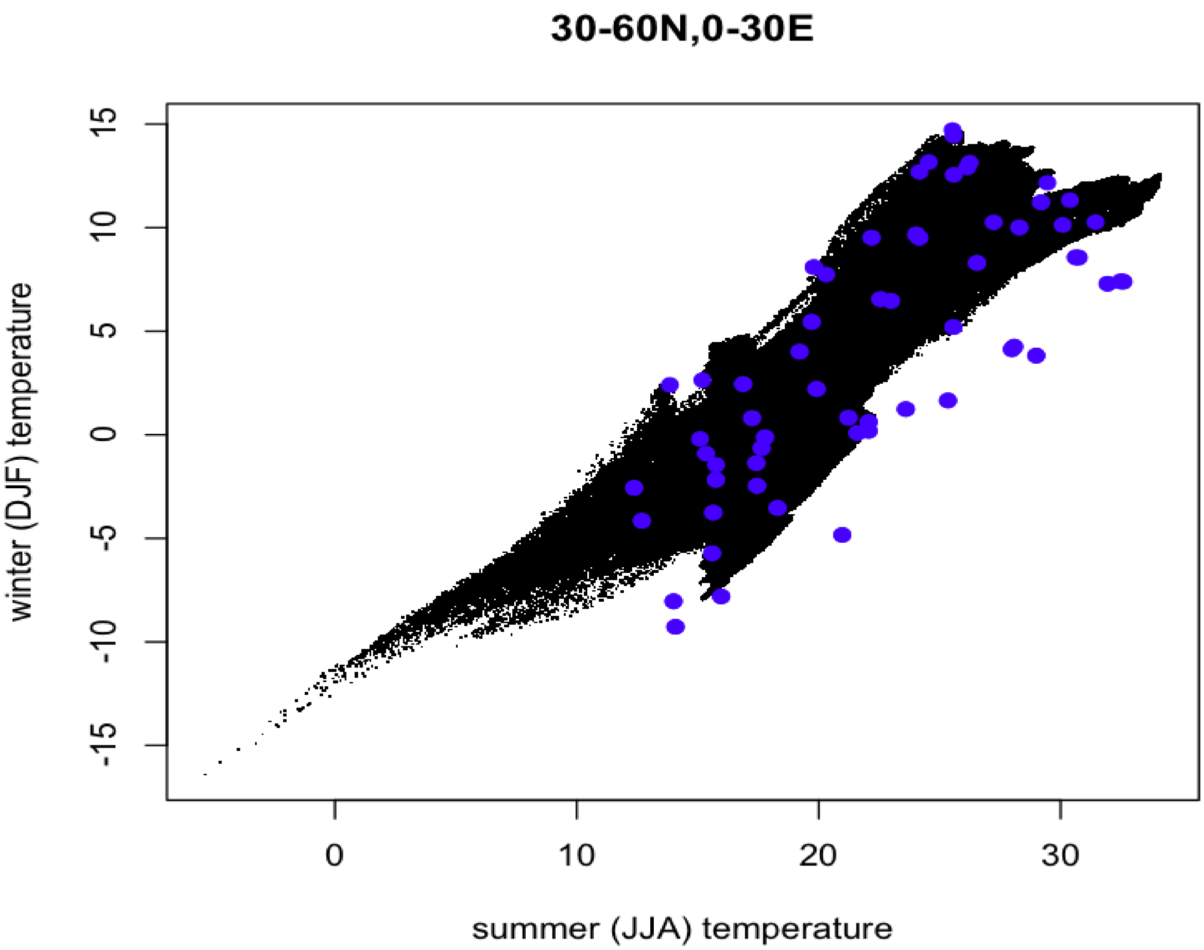


Figure R2: Distribution and relation of gridpoint winter and summer temperatures in the 0.5'-resolution temperature field of (Hijmans et al., 2005) and the gridpoint temperatures of the ECHAM5-MPIOM simulation (Fischer & Jungclaus, 2011) used in this study.

Spatial autocorrelation is often not considered in papers reconstructing climate from pollen (e.g. in Bartlein et al., 2011), and will tend to be a larger problem in the densely sampled pollen databases than in our low resolution data, as each pollen site has many geographically close neighbors which can be used as an analogue in the modern analogue technique.

Point 4: Simplification

> *Whilst some simplification should be expected in a 'virtual' study like this, it is important not to over-simplify to the point where the study itself is so far removed from any actual application that the results are not comparable.* The problem here is that the authors consistently conflate their results with those from actual pollen-climate reconstructions (as in the title), and therefore are at risk of presenting a fallacious argument that the average reader who is not so familiar with the topic will likely interpret at face value.

We fully agree that the complexity of the vegetation representation in the model as well as the simulated climate evolution are a strong simplification of the reality. Therefore, results on the Holocene evolution of specific PFT's, the actual spatial pattern of PFT's, or the reconstructability of a certain climate variable in a certain region should not be directly translated to the real-world.

On the other hand, conclusions about reconstruction methods and the relation of spatial calibration and downcore reconstruction only require a consistent dataset of climate and vegetation parameters in space and time and do not depend on details of the climate evolution or vegetation response, as long as the dataset is realistic enough that we can apply the real world reconstruction workflow. The major factor shaping these results is that the modern spatial relationships between climate variables is different from the changes in the relationships over time, which is a robust feature related to the transient insolation forcing.

In the revision, we will check in detail again if all our statements are either independent from the model-world specifics , or are clearly marked that they just apply to the model world. Furthermore, we will emphasize the limitations of our study further by extending paragraph 4.1, and by highlighting their impact in the conclusion paragraphs. We do, however, find the title of our manuscript appropriate, as it clearly expresses that we work with pollen-related methods in an ideal model world.

Point 5: Repeating the study with a more elaborate vegetation model (LPJGuess)

> The subject of the paper is nevertheless interesting, and one that would otherwise be worthy of publication. I would therefore encourage the authors to collaborate with someone who has more experience in pollen-climate modeling, and to use a vegetation model such as LPJGUESS which can simulate a greater number of PFT's/Taxa so that the analysis can be more comparable with how pollen-climate transfer functions are actually applied.

As we have shown above, our analysis is comparable to actual pollen-climate transfer functions, as the effective number of species present in the model data is within the range of the numbers observed for the European Pollen Database, which formed the basis e.g. for (Mauri et al., 2014; Mauri et al., 2015).

We agree that using a more complex vegetation model, such as LPJGUESS, would be worthwhile in a future study as this would not only increase the number of PFT's but also include a more realistic interaction between climate and vegetation. This would allow to also test other properies of the reconstruction e.g. effects on the time-scale dependent variability, or potential time-lags. We will thus include this proposal in the outlook of our study.

However, we expect that the main result of this manuscript, the limitations of spatial modern calibrations, would only be strengthened. This is because a more realistic climate-vegetation response will even differ more in time versus space, if for example the modern vegetation is not yet in equilibrium with the modern climate state.