

Response to the comments of Referee #2

We would like to thank the reviewer for his/her positive evaluation of our work and for the constructive comments. Our responses to his/her comments are presented below in blue italic fonts while the comments themselves are in normal font.

This manuscript presents an analysis of the performance of climate model simulations of the last millennium. I don't really have much to say in the way of suggestions and recommend it for publication in CP. The analysis is extensive and seems appropriate, the paper is not earth-shattering in conclusions, but it is certainly a useful analysis of the current state of climate model performance (as far as can be deduced from recent proxy-based reconstructions). It is a lengthy paper but as the model-data comparison is extensive, this is not entirely avoidable. However, the English in some places is a little hard work, and it could do with some editing and rewriting in parts.

As also suggested by the first referee, we will edit and rewrite some parts of the text for the revised version to make it clearer and easier to read. A part of the material will also be suppressed or moved to the supplement to reduce the length.

One area that particularly stands out is the description of volcanic response in 6.1, where it is not always clear which two quantities are being compared for many instances of "larger" and "smaller" etc. In addition to the overall magnitude of forcing, I would expect that the regional distribution of the forcing could be somewhat uncertain here (and thus lead to regional discrepancies between models and reconstructions) but this does not seem to be mentioned in the text.

In the revised version, we will ensure that when two quantities are compared, it is clear which one is larger or smaller. As the magnitude of the forcing is indeed very uncertain at regional scale, we will remove it from the plots (Fig. 8; Figs S11-S13) and discuss the uncertainties in this forcing in the revised version of the manuscript.

I'd also prefer more direct language that does not skate around the issue of model data disagreements where they exist. Throughout the paper, discrepancies are often attributed to "uncertainties" when in fact uncertainty, if correctly accounted for (which the methods used can potentially do), should not in itself give rise to significant discrepancies. The problem is surely with errors that lie outside the range of estimated (or tested) uncertainties. The authors could do worse than globally search for "uncertain" and ask themselves whether it would not be clearer to talk frankly about errors.

The paper also seems to take a rather rosy view of the reconstructions. I realise that the authors did not set out to assess the reconstructions, but taking them as a ground truth (albeit with their stated uncertainties) seems potentially misleading. Given the wide range of results reported for NH temperatures e.g. in the IPCC AR4 "spaghetti plot" of Fig 6.10 (which shows persistent disagreements of order 0.5C even after heavy smoothing), it seems optimistic to expect reconstruction accuracy to be reliable on regional scales. Where models agree reasonably with each other (but not with the reconstruction) and forcing uncertainties are not considered to be large, gross errors in reconstructions cannot be excluded.

We agree with the reviewer that it is important to make the distinction between uncertainties and discrepancies. We also agree that we cannot exclude gross errors in reconstruction that are larger than the given uncertainties. Nevertheless, it is impossible to state for sure that a difference between reconstructions and model results is due to errors in model physics, in the forcing or in the reconstructions. This is why in that case we prefer to mention the disagreement and then discuss its

possible causes. The fact that the quality of reconstructions is a potential source of discrepancy between model results and reconstructions (in particular in the Southern Hemisphere because of the lower number of records compared to the Northern Hemisphere) is mentioned several times in the submitted manuscript, starting in the abstract. Nevertheless, we will check again carefully in the revised version, as suggested, that in all sections we are clear enough about those points, stating it explicitly each time when a difference between reconstructions and model results cannot be attributed to uncertainties alone and discussing all the potential sources of disagreement, including inadequate or underestimated uncertainties in the reconstructions. To better elucidate the scope of the uncertainty estimation for reconstructions, we will be more explicit about which sources of uncertainty were accounted for in each regional reconstruction.