This version of the manuscript is much improved from the original. In particular, the figures have been greatly improved, and the addition of the Huntley et al. (2023) simulations brings the manuscript closer to its stated goals of investigating the impact of the downscaling of climate-model output on the comparison of that output with pollen-based reconstructions. I still think, however, that it is wrong to describe the Beyer et al. (2020a) data as "model simulations" (line 129) or "model output[s]" (line 149), because it was already downscaled/interpolated, and several steps removed from the actual Singarayer and Valdes (2010) etc. data. I fear that a naïve reader might think that there is therefore no benefit to the downscaling of actual model output when an easy-to-use data set can be found in an R package.

> We thank to reviewer for acknowledging the improvements made to our paper.
>
> Regarding our reference to 'model simulations' and 'model output', we standardised our terminology in response to this reviewer (Reviewer 2 RC5) who themselves suggested "model output" and "proxy reconstructions" were the correct terms to differentiate between the two sources of climatic data we are comparing. We have however added a comment on Line 578 to address the final point about the differences in the models utilised: "We note that whilst the latter [Beyer et al. 2020a] is not a direct output from a GCM, it is easily accessible for consumers (rather than producers) of model data, includes more sophisticated initial downscaling that takes advantage of a few runs of a high resolution GCM, and is likely to be used by others in the future as a starting point for further delta-downscaling." We have also edited Table 1 to further highlight the difference between the two model datasets.

Is there any benefit to using "real" (Huntley et al., 2023) model output as opposed to previously downscaled data? I agree that there may not be a statistically significant difference in performance, but Fig. 3 hints at a practical difference between the two (i.e the RMSEs for the Huntley et al.-based results appear lower than those for the Beyer et al.-based results).

> We are unsure which figure the reviewer is referring to here, as Figure 3 demonstrates absolute bias values rather than RMSE (with no clear differences between Beyer et al. (2020a) and the HadCM3 GCM output in any of the figures). We note however that we included normalised RMSE (NRMSE) values for all of the comparisons, which is a standardised metric of coherence, and discuss the differences in NRMSE between the different models throughout the text (e.g. Lines 369-371: "In West North America, NRMSE is higher in the HadCM3 model outputs compared to that from Beyer et al. (2020a), with no differences between resolutions in the latter (when debiased using WC) and a slight improvement in coherence due to downscaling in the former"). However, there are no consistent 'practical' differences between the Beyer et al. (2020a) and the Huntley et al. (2023) models.

I think the authors did a good job of responding to reviewers' comments. However, I think three of their responses to my comments should also appear in the text. These are:

"AC1: We thank the reviewer for suggesting that we include a comparison with directly downscaled HadCM3 outputs. We have done so, using a model time series from Huntley et al (2022), which is an updated version of that used to generate Beyer et al. (2020a). The conclusions of our paper do not change… (p.14-15)"

> We have added this sentence to Lines 119-122: "We applied a new suite of functions in the *pastclim* R package (Leonardi *et al.,* 2023) for delta-downscaling model output, including model-data comparisons with directly downscaled HadCM3 outputs from Huntley et al (2023), which is an updated version of that used to generate Beyer et al. (2020a), as well as the model time series from Beyer et al. (2020a)."

"AC3: Based on this, we do not believe that the take home message is 'why bother' but that careful consideration should be required to determine when downscaling is important, given that coherence between proxy records and model outputs does not change significantly. We understand that the reviewer is 'disappointed'… (p. 18)" (And I should point out that I was disappointed by the paper, not the results.)

> We have added this sentence on Lines 536-539: "However, we stress that our take home message is not 'why bother', but that careful consideration should be required to determine when downscaling is important, given that coherence between proxy records and model outputs does not change significantly."

"AC20: We thank the reviewer for this suggestion. We have added the HadCM3 GCM from Huntley et al. (2022) to our analysis and find highly similar results with that of the Beyer et al. (2020a) output. Pertinently, we also find no statistically significant differences in coherence with proxy records between the HadCM3 GCM model output at 30-min and at 5-min resolution. We have retained Beyer et al (2020a) since it is an easily accessible product that includes more sophisticated initial downscaling that takes advantage of a few runs of a high resolution GCM, and it is likely to be used by others in the future (particularly consumers of climatic models) as a starting point for further delta-downscaling. (p. 22-23)"

This last comment exposes the authors' affinity for the data sets in the R package. They may be easy to use, but I don't think "easy" and "optimal" are the same thing.

> We have added this sentence, and a caveat, to Lines 576-581: "Overall, we present a streamlined pipeline for delta-downscaling climate model time series within the *pastclim* R package (Leonardi et al. 2023), and we have presented testing of downscaling using both HadCM3 model output (Huntley et al. 2023) and the product of Beyer et al (2020a) directly available within the package. We note that whilst the latter is not a direct output from a GCM, it is easily accessible for consumers (rather than producers) of model data, includes more

sophisticated initial downscaling that takes advantage of a few runs of a high resolution GCM, and is likely to be used by others in the future as a starting point for further delta-downscaling."

Throughout the text, Huntley et al. (2022) should read Huntley et al. (2023).

We have made this correction throughout.