Thank you so much for the thoughtful and careful review. We really appreciate it and addressing your comments has greatly improved the manuscript. To summarize, we have made some major changes to the manuscript figures and discussion in response. These include:

- Reframed the discussion around preservation of extreme quantiles throughout the manuscript, including an update to the title of the manuscript itself, to make it clear that high and low quantile changes are preserved but not necessarily threshold-based extremes.
- Added additional figures to illustrate the post-processing we describe in Section 4.3.1. These include time series of precipitation for two cities showing reanalysis, raw, regridded GCM data, and pre- and post-processed bias-adjusted and downscaled data (Figures A3 and A4); global maps for all GCMs showing the number of timesteps swapped for maximum and minimum temperature for the historical period (1960-1980) and future end-of-century period (2080-2100) for SSP3-7.0 (Figures A1 and A2); and global maps for all GCMs showing the number of timesteps clipped for precipitation for the same periods (Figures A5 and A6).
- Significant updates to our discussion of bias adjustment and downscaling in the introduction (Section 1), including preservation of the raw model signals and both advantages and pitfalls of this approach.
- Figures 3 and 4 include all seasons and show the ensemble mean across GCMs. Figure 4 only includes wet days (e.g., precipitation days > 1 mm/day). A version of both figures showing the 99th percentile is in Appendix A (Figure A7 for maximum temperature and Figure A8 for precipitation wet days), and we also show drier day figures (e.g. precipitation days < 10 mm/day) in Appendix A as well (Figures A9 and A10 for the 95th and 99th percentiles, respectively).
- Added ensemble member IDs to both Table 1 and a table we have added to the paper based on other reviewer comments, Table B1, which includes all available GCMs in ESGF and reasons for why some of those GCMs were excluded from the GDPCIR dataset.

Our detailed responses are below in normal text, with reviewer comments in italics. Thank you again for such a helpful review!

Comments on "Global downscaled projections for climate impacts research (GCPCIR): preserving extremes for modeling future climate impacts," by Gergel, Malevich, McCusker, Tenezakis, Delgado, Fish, and Kopp, egusphere-2022-1513.

This manuscript obviously represents a very large amount of important work and is quite commendable for the usefulness of the data produced. Overall I am very positive about the project. I do have some major comments that need to be addressed before publication but that does not diminish the overall quality of the work in my mind.

Thank you for providing a careful review of the manuscript. We have responded to all comments below and discussed how we have incorporated them throughout. Comments are denoted in italics and our responses are in regular font below.

Major comments

1. The manuscript emphasizes preserving extremes -- it's even right there in the title -- but evaluation of the extremes is given only a weak treatment in the manuscript. A better and more complete job of describing the extremes is necessary if, as we see here, the extremes are declared by the authors to be a key component of the project.

*For example, Figures 3 and 4 only show 95th percentile values. That is about 3-4 days per year. The hottest 3-4 days in a year, the wettest 3-4 days, etc. The economic and societal importance of climate extremes are much more apparent at extremes that are less frequent than several times per year. For example, water management and flooding analyses routinely consider 1-in-100 *year* floods or precipitation events. The Pacific Northwest heat wave of 2021 has been estimated at a 1-in-multi-millennium event. Values that are routinely seen several times per year are not near the level that causes the big societal and economic impacts that this manuscript asserts that it is concerned with.*

I request that the authors add to the main text or supplementary figures a comparison of how well their method preserves GCM-predicted trends at more extreme levels. For example, 1-in-10 year, 1-in-20 year, and/or 1-in-50 year extremes would be appropriate, especially for precipitation. Besides the trends, it would be useful to see what the actual values look like. The text describes an issue with some extreme values becoming unrealistic and steps taken to mitigate this. Illustrations and evaluations of these more important and impactful (than 95th ptile) extreme values are needed.

Thank you for these comments. We agree that we were using the word "extreme" when we should have been more explicit about referring to high or low quantiles. We have addressed these issues in a number of different ways that we describe below.

Firstly, we have updated Figures 3 and 4 to show the GCM ensemble mean 95th percentile values for all seasons, and the same figures for the 99th percentile have been added to Appendix A (Figures A7 and A8 for maximum temperature and precipitation, respectively). We have also added figures showing precipitation days < 10mm (Figures A9 and A10), described in the "Line 448" comment response below.

However, we acknowledge that even the 99th percentile values represent the maximum or close to the maximum value for that season, rather than the 1-in-10, 1-in-50, or 1-in-100 year floods or precipitation events. We considered showing these extremes, but since we are not specifically applying a correction to threshold extremes that are not quantile trends, we also do not specifically preserve trends in these extremes discussed in the comment above. Consequently, we did not feel that showing these extremes by, for example, fitting a Gumbel distribution to extreme precipitation values would actually be an effective way of analyzing the performance of the method. We know that we *aren't* preserving threshold extremes like those listed above, since that is a limitation of the QDM method.

In thinking through this, we have also updated some of the language around preservation of extremes, since we do not intend to claim that we are preserving trends in the threshold extremes, but instead trends in the quantiles. Any places throughout the manuscript where this was not clear has been updated. We also made a slight update to the title of the manuscript to make sure this nuance was clear.

To address the request to see what the extreme values look like when they become unrealistic, and the steps taken to mitigate this issue, we have added four new figures to Appendix A, which we discuss in Section 4.3.1. Figure 3 shows time series of reanalysis, raw GCM data, bias-adjusted and downscaled (before post-processing) and bias-adjusted and downscaled (after post-processing) for a single city, Delhi, for a single climate model, MIROC6, with projection data for SSP2-4.5. By post-processing, we mean the "additional post-processing" we describe in Section 4.3.1. Figure 4 shows the same time series for Cairo, Egypt. The infrequent, yet physically unrealistic values occur several times over the full time series, and the "clipping" is apparent from the figures.

To further assess the effects of post-processing across all GCMs that are part of the GDPCIR dataset, we then computed the number of swapped timesteps for maximum temperature, meaning daily timesteps where maximum and minimum temperatures were swapped because minimum temperature exceeded maximum temperature. We computed this over a 21-year period for a historical climatological period outside of the calibration period (1960-1980) and for the end-of-century for SSP2-4.5 (2080-2100). This analysis is shown for the historical period in Figure A1 and for the projection period in Figure A2.

We do the same analysis for precipitation, except in precipitation we apply the "cap" that we describe in Section 4.3.1. The number of "clipped" timesteps is shown in Figures 5 and 6 for the same historical and end-of-century time periods as for maximum temperature. Generally, the number of timesteps post-processed for both variables is low, given that the numbers are computed over daily data in a 21-year climatological period.

2. There is a rich literature on some of the issues addressed here, such as how to bias correct in a way that preserves the original model trends, that is not included in the current manuscript. These works should be appropriately cited since they are concerned with an important part of the submitted manuscript. The way it is now, it gives too much of an impression that the described work exists in a vacuum, which it does not.

For example, Michelangeli et al. 2009 (GRL vol 36 L11708, doi:10.1029/2009GL038401) addressed the issue of how to bias correct a model subject to climate change using the cumulative distribution function transformation (CDF-t) method; H. Li et al. 2010 (JGR vol 115, D10101, doi:10.1029/2009JD012882) described the fundamentals of equidistant quantile mapping some years before the Cannon reference you cite; Pierce et al. 2015 (J. Hydromet v. 16 p. 2421, doi:10.1175/JHM-D-14-0236.1) implemented the quantile trend-preserving bias correction for a large data set as well as comparing it to standard quantile mapping, etc. I'm sure you can find other examples, but overall I think the manuscript as written gives short shrift to the context in which this work was done.

Thanks for this comment and for pointing out the gaps in the introduction. We have restructured the introduction (Section 1), in particular the original paragraphs on bias-adjustment and downscaling, and added a much more thorough discussion of the literature on bias adjustment, including the papers mentioned in this comment and a number of other papers as well. We also include a discussion of preservation of original model trends as well as potential shortcomings of this approach based on this and other reviewer comments.

Minor comments

** Line 98: Please list ensemble members for each model. Not all models have an "r1i1p1f1". Was only one ensemble member per model downscaled? Please state this explicitly.*

Only one ensemble member per model was downscaled due to computational limitations. We have clarified this in Section 2.1 and added the ensemble member for each model in the GDPCIR dataset to Table 1 as well as Table B1.

** Line 323, Figure 2 caption: In the caption please explicitly state which variable is shown.*

We have added variable names to the Figure 2 caption.

** Line 435: For precipitation, is this 95th ptile of all days, or only wet days? Please specify. As noted above, 95th percentile of all days, not even wet days, is not extreme enough to cause substantial societal or economic impacts.*

We have redone Figure 4 so that the 95th percentile shown is *only* showing wet days, or days with precipitation greater than 1 mm. In Appendix A, we also include the same figure for the 99th percentile to illustrate more extreme behavior (Figures A7 and A8). Additionally, Appendix A includes the same two figures for precipitation days less than 10mm (Figures A9 and A10). This is discussed further in our response below to the "Line 448" comment.

** Line 448: The panels in Figure 4 are too small to be useful. Please redraft so that the panels are much larger.*

We have redone Figure 4 and the new panels are larger. Additional changes to Figure 4 are described in the “Line 448” comment below.

** Line 448: It's hard to interpret Figure 4 because some of the areas of concern are in dry areas where the colorbar is just showing a dark blue that is hard to differentiate. It would be useful to add some figures to the supplementary information to address this, for example, an additional set of figures where only regions with p < 10 mm are shown (with their own colorbar). The question I want to answer is whether I should be concerned about the large misses in some locations as illustrated here, or is this confined to regions where there is only very little precipitation anyway. Between the tiny panel size, a colorbar where all values below 12 mm look the same, and only supplying JJA rather than the other seasons as well (which need to be added), I can't answer this important question. Results given for only one model, one scenario, and one season do not inspire general confidence in a large data set.*

We have addressed this comment with major changes to Figure 4 as well as adding additional figures to Appendix A. Firstly, all panels have been enlarged and all seasons are now included. Rather than showing a single GCM as was done previously, we now show the ensemble mean across GCMs. Figure 4 shows the 95th percentile *of wet days*, meaning days with precipitation values greater than 1 mm, whereas previously the figure showed *all* precipitation days. We have added the same figure for the 99th percentile, Figure A8 in Appendix A.

To differentiate trends between wet and dry days, we have added additional analysis showing the same figures for days with precipitation less than 10mm, as suggested. Figures A9 and A10 show the ensemble mean for all seasons for the 95th and 99th percentiles, respectively, for days with precipitation less than 10mm/day. We discuss differences in spatial patterns in these figures in Section 5.1. Generally, in the wet day figures, the largest differences in the bias-adjusted and downscaled data trends vis-à-vis the GCM trends is surrounding the ITCZ (also discussed in Section 5.1). In looking at Figures A9 and A10, we can see that the "large misses" described in the comment are mostly concentrated in drier areas (e.g. Sub-Saharan Africa and the poles), and this behavior is underscored by the fifth column of panels, the difference in change between bias-adjusted and downscaled, where the cause of the difference is the "post" wet day frequency correction (which we apply after downscaling in our pipeline, and thus is not applied to the bias-adjusted, pre-downscaled data).

** Line 573, Figure 8: Why do some of the Y axes say "False"?*

This was a typo and we have corrected it.