In our point-by-point author response below, referee comments are in italics, our responses are in regular font, and line numbers refer to the latest version of the revised manuscript and are denoted in yellow and comma-separated. The authors would like to express our gratitude to the topical editor and to the anonymous reviewers who carefully reviewed the initial manuscript submission as well as the revised version.

Response to Referee 2 comments on revised manuscript

Comments to "Global downscaled projections for climate impacts research (GDPCIR): preserving extremes for modeling future climate impacts", by Gergel et al.

The authors did a great job and the manuscript has substantially improved in the revised version. Thank you for all clarifications and new additions. I have only some additional minor comments.

Thank you so much for the thoughtful review of our revised manuscript. Your insights have greatly improved the manuscript and we are really grateful for your feedback. We have made the following changes described below.

L38 At this stage more basic references to bias adjustment methods should be given instead of works about multi-variate BC methods (i.e. François et al., 2020b) which have not been even mentioned yet and are not considered in this work. I would recommend to replace that citation by Maraun and Widmann 2018, Räty et al 2014 and references threrein. Also note that the references François et al 2020a and 2020b seem to refer to the same work so, please, correct the references list and its mention throughout the manuscript.

We have replaced the reference to Francois et al., 2020 with Maraun and Widmann 2018 and Raty et al. 2014. Additionally we have fixed the Francois et al 2020 reference throughout the manuscript (introduction and methods, see line numbers below).

L26-27, L38, L169 (and any other Francois et al 2020 reference occurrences)

L38-39 "or methods that use deep learning neural networks (Baño-Medina et al., 2021)" The cited work is not dealing with bias correction, but with statistical downscaling using large-scale predictors which is a different approach for statistical downscaling. Thus, it does not apply to mention it here. Please remove this part of the sentence.

We have removed this part of the sentence from the manuscript.

L38

L40 Mention also empirical quantile mapping (Déqué 2007) after Li et al. 2010, which in my view is the most widely used QM method and implementation.

We have added Déqué 2007 after Li et al. 2010, and also added Déqué 2007 to the references.

L40, L749-752

L41-42 I would write "VALUE Cost Action experiment" or "VALUE experiment" instead of "VALUE study".

We have updated this to “VALUE experiment” as suggested.

L41

L44-45 Please add "parametric" before "quantile mapping approach", to differentiate from the previous methods.

We have added “parametric” before “quantile mapping approach” to make this clear.

L44

L50 I would add to the reference "and references therein".

We have added this clause and also added “e.g.,” before the reference to accommodate this addition.

L50-51

L55 Please add "such as threshold-based indices" after "for climate extremes".

We have added the clause “such as threshold-based indices” after “for climate extremes”. There was also a space missing in this line which has been fixed.

L58

L56 I do not think that Casanueva et al. 2019 generally support the use of trend-preserving methods. In fact, they found the modification of the signals in GCMs (by empirical quantile mapping) to produce, in specific cases, more realistic signals than those of the original raw GCMs. Thus I would remove the citation there.

Thank you for catching this, we have removed the Casanueva et al. 2019 citation in this line.

Lines 50-61 could be better organized. For instance, it would make more sense to mention Lehner et al. 2023 and Casanueva et al. 2020 in the same sentence since they are very much related. Also "an additional question..." is worth to mention but it is a bit lost there. I see two messages in the paragraph which are a bit mixed in the current version: 1) general question about preservation or not of the climate change signals (also the mean, not only about quantiles) since trend-preserving corrections are a sensible choice when a climate model simulates a credible climate change signal (Maraun 2016) but could be questioned otherwise, 2) if preservation is desired, QDM is one of the best performing methods.

Thank you for suggesting this. We have reorganized the paragraph to reflect the two take-away messages you outlined. We moved up the Lehner at al. 2023 reference (and related sentence) to follow the first mention of the QDM method. We also moved up the discussion of the Qian 2021 paper. We also mention that in spite of the uncertainty in the future climate signal, one of our goals was explicitly to preserve trends in moderate to extreme climate indices, and the Casanueva et al 2020 study found that QDM in particular did perform better in preserving trends, even though trend-preserving methods do not *necessarily* perform better for threshold-based indices.

L50-62

L108 This is the first time SSP is mentioned and the abbreviation is defined a few lines later.

Thank you for catching this. We have replaced “SSP” with “ScenarioMIP experiment”.

L110

L119 I think it should be Table 1 instead of A1.

Thank you for catching that, indeed it should be Table 1, we have corrected this.

L121

L252-259 I guess the described approach refers to the "pre" wet day frequency adjustment. If so, mention the second "post" adjustment afterwards or state more clearly that this paragraph refers to the "pre" adjustment.

That is correct. We have added a sentence after this mentioning that after downscaling, all values below the specified threshold (1 mm/day) are replaced by 0 mm/day. In the first paragraph of Section 3.3, we mention that this is the “post” WDF adjustment.

L261-262

L357 (and Sec. 4.3 in general) "The analog day for that quantile is 1.5º". Being the analog day a temperature does not make much sense to me (the analog day should be another day), please rephrase. I am sorry for insisting but still do not get why adjustment factors are called analogs, at least some times. I think that analog day is not exactly the adjustment factor, but the day corresponding to the closest quantile within the 620 days used for calibration, isn't it? If so, I would rather refer to them in the text as spatial adjustment factors as they are in the figures and equations. Also, as authors corroborated, downscaled data for Miami (Fig.2) depends only on the adjusment factor of the 0.25x0.25 over Miami, so the adjective "spatial" is a bit misleading. I do not fully get panel c showing "all possible adjustment factors for 15 August", what are the bars for each quantile showing? Please clarify the caption description.

Thank you for these comments and our apologies for any confusion here. What we meant by "The analog day for that quantile is -1.5" should have read "The analog-based adjustment factor for that day is -1.5", but since that is a bit confusing we have rephrased it to "The spatial adjustment factor for that quantile is -1.5". For consistency, we also updated the caption for panel a to use the term adjustment factors. For panel c, we added clarification that by "all possible adjustment factors" we mean "corresponding to all quantiles".

That is correct that the "analog day" corresponds to the closest quantile within the 620 days used for calibration. Each quantile refers to a given day, or "analog" of the reference training period, and then the adjustment factors represent the difference (or ratio) in empirical quantiles of the reference data at coarse and fine resolution. The reason we use the term "analog" is because each of the day values does correspond to an actual day in the coarse and fine resolution calibration data. In Figure 2, the "bars" in panel c show that the adjustment factors are not perfectly smooth given that the values from which they are computed in coarse and fine calibration data for each quantile refer to actual days.

L358, L360, L366, L367

L586-588 So the next question would be, at which resolution are errors in Eq. 6 and 7 calculated? at the final 0.25x0.25º or the coarse 1x1º or the GCM original resolution? Or do authors just took the gridbox over each city for each dataset regardless of the resolution? The latter would not be entire fair due to the different representativeness of each grid box. I think that the fairest would be a comparison on the 1x1º grid (i.e. upscaling the results at high resolution) since departures from the raw data are analyzed and some added value of the high-resolution could still be present after upscaling. Same question for Sect. 5.2.3: at which resolution are raw and downscaled datasets compared? Please clarify.

We do take the gridbox over each city for each dataset, which is what we mean by “at the pixel level” in that section. We acknowledge that this is not entirely fair both due to the differing resolutions but also due to the artifacts of regridding, and we discuss this in Section 5.2.2. However, alternatively, if we had upscaled the downscaled data to a 1-degree grid and upsampled the climate model data to a 1-degree grid, these steps would also have had undesired effects (for example, loss of information in the case of upscaling the downscaled data). We explored various options for this analysis in depth, and given all of the issues discussed herein and in the manuscript decided that we felt this was the most faithful method to the resolutions of the original data. But we acknowledge that other decisions in computing the errors could have been made that would have likely impacted the results shown.

L588 Not sure why ERA5 is mentioned here, since these analyses compared raw vs bias corrected plus downscaled data.

We mention ERA-5 here because in describing the methods we used for Equations 6 and 7, in Equation 6 x refers to historical daily climatological reference data, which is from ERA-5, so it in fact relevant here.

I believe the confusion stems from the fact that ERA-5 is not used in Equation 7, but our description of the method and how we depart from the Lange 2019 method includes a discussion of both Equations 6 and 7.

L593 "variables" can be removed.

We have removed “variables” and added “maximum and” as we intended to mean both temperature variables (e.g. maximum and minimum temperature).

L598-599

Sect. 5.2.2 In my view the quantities of Eqs. 6 and 7 should not be called "errors", especially the one in Eq. 6 which represents the difference between raw and BA plus downscaled data in the annual cycle, since it should imply an improvement by construction. I would rather use the term "differences" or "effect of BA plus downscaling" and "effect in trend preservation". Check also line 651.

Thank you for this suggestion; the authors discussed this terminology extensively when we were making method decisions on this analysis. We have kept the term but we added a description of what we mean by “error” that discusses what you mention above. Additionally, we have updated the terminology earlier in the manuscript before we introduce explicitly what we mean by the term “error”. At the beginning of Section 5.2, we replaced the term “error” with “modification”.

L526, L573-575

Sect. 5.2.2 and Sect. 5.2.3 have almost the same title, consider to change 5.2.3 to something a bit more specific.

Thank you for this feedback - we have renamed section 5.2.3 to “Relative trend preservation of ETCCDI indicators aggregated over selected regions”.

L608

L627 and caption of Fig. 7 "change in period average" is commonly denoted as climate change signal. Same in lines 636-637 and caption of Fig.8 and 9.

Thank you for this suggestion; we have replaced the “change in period average” terminology with “climate change signal” in all instances mentioned above except in the Figure 9 caption, where we feel that this terminology would be a bit awkward and “change in period average” for threshold counts makes more sense as a descriptor for the figure.

L633, Figure 7 caption, L641, Figure 8 caption

L644 Consider to add that this is in line with other studies such as Casanueva et al. 2020, since the lack of preservation of derived indices signal has been already reported in other studies.

Thank you for this suggestion. We have added a clause that this result is in line with Casanueva et al. 2020 and also Dosio 2016 (and added the latter article to the reference list).

L647

Fig. 5. Consider to plot the bars for the raw GCMs with the thin blue frame as in Fig. A11, because the grey rectangle is sometimes hard to see.

Thank you for this suggestion; however, we have decided to keep Figure 5 as is since the contrast between the bias-adjusted and downscaled GCM (dark blue) and raw GCM overlain (light blue/grey) is discernible even for the few cases where for a given city and GCM the number of Kolmogorov-Smirnov tests passed is the same for both raw and bias-adjusted and downscaled data.

Table B1 What does GC in "GC CMIP6" mean?

The “GC” in Table 1 refers to Google Cloud, e.g. the Google Cloud CMIP6 collection. We have clarified this when referring to Table 1 in the text by adding a “(GC)” after our mention of the Google Cloud CMIP6 collection.

L120

Please mention either in the methodology or around line 70 (where ISIMIP is mentioned) that, within ISIMIP3, bias adjustment is developed at coarse resolution and subsequent stochastic statistical downscaling (based on The MBCn -multivariate quantile mapping bias adjustment method- algorithm by Cannon 2017) to a finer resolution of 0.5° is the final product. So the procedure is similar to the QDM plus QPLAD procedure.

Thank you for this suggestion. We have updated the description of the ISIMIP dataset to mention that the latest version of the ISIMIP3BASD methodology uses this approach (which is the method used in the CMIP6 dataset we describe). We also reworked the description of ISIMIP a bit, moving the Lange 2019 reference to the end of the sentence and adding the Cannon et al 2018 reference (our citation information has 2018 versus 2017 as the date of publication for that article).

L69-72

References

Cannon, A. J.: Multivariate quantile mapping bias correction: an N-dimensional probability density function transform for climate model simulations of multiple variables, Clim. Dynam., 50, 31–49, https://doi.org/10.1007/s00382-017-3580-6, 2017

Déqué, M. (2007) Frequency of precipitation and temperature extremes over France in an anthropogenic scenario: model results and statistical correction according to observed values. Global and Planetary Change, 57, 16–26. https://doi.org/10.1016/j.gloplacha.2006.11.030.

Dosio, A. (2016) Projections of climate change indices of temperature and precipitation from an ensemble of bias-adjusted high-resolution EURO-CORDEX regional climate models. Journal of Geophysical Research—Atmospheres, 121, 5488–5511. https://doi.org/10.1002/2015JD024411.

Maraun D (2016) Bias correcting climate change simulations—a critical review. Current Climate Change Reports 2(4):211–220. https://doi.org/10.1007/s40641-016-0050-x

Maraun, D. and Widmann, M. (2018) Statistical Downscaling and Bias Correction for Climate Research. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781107588783.

Räty, O., Räisänen, J. and Ylhäisi, J.S. (2014) Evaluation of delta change and bias correction methods for future daily precipitation: intermodel cross-validation using ENSEMBLES simulations. Climate Dynamics, 42, 2287–2303. https://doi.org/10.1007/s00382-014-2130-8.