Thank you 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:

- Discussion of our choice of ERA5 for a reference dataset (added to Section 2.2)
- 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.
- Added an additional section to the paper, Section 4.4, which describes the dataset in further detail, and a section to the supplemental material, Appendix C, which describes the downscaling pipeline specifications.
- Throughout the paper, we have carefully reviewed the terminology of when to use "model" versus "GCM". We only use "model" when referring to methodology that could be applied to both regional and global climate models.
- To address the question of how comprehensive the GDPCIR dataset is, we added a table to the supplemental material, Table B1, which includes all available GCMs in ESGF and reasons why certain GCMs were not included in the GDPCIR dataset.
- Figures 3 and 4 now include all seasons and show the ensemble mean across GCMs. Figure 4 only includes wet days (e.g., precipitation days > 1 mm/day). The corresponding figure for the 99th percentile is in Appendix A (Figure A8), and we also show drier day figures for precipitation in Appendix A as well (Figures A9 and A10). We also include the 99th percentile for maximum temperature (Figure A7).
- Figure 5 analysis has been corrected and the historical time period over which statistics are computed has been extended, and the updated version of Figure 5 reflects these changes.
- Figure 6 now includes all GCMs and we have updated the calculation of bias adjustment and trend preservation error based on reviewer comments.

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

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

This work presents a comprehensive assessment of a new global dataset based on bias adjustment and downscaling of CMIP6, which could be of great interest for the impacts community. The paper is well-written and methodological details are meticulously explained. Still I recommend to address some points before the manuscript is considered for publication.

Overall comments

ERA5 data are used as reference to bias-correct climate model output. Since the focus of the dataset is the impacts community and ERA5 is a reanalysis dataset, the appropriateness of ERA5 for impact studies should be discussed. Higher resolution than other observational-based products is an advantage, but was it evaluated compared to real observations (by the authors or in other works which could be cited)? Also note that bias adjustment methods which preserve *trends for all quantiles (such as QDM) rely to a greater extent on the reference dataset used for calibration, thus presenting a larger sensitivity to the observations used for calibration (especially for precipitation) as shown by Casanueva et al. 2021. This issue related to uncertainties due to the reference dataset should be also discussed.*

Thank you for pointing out this gap in the manuscript. We have added an additional paragraph to Section 2.2 that discusses the appropriateness of ERA5 for impact studies and cites a number of recent articles which compare ERA5 to observations as well as other commonly used reanalysis products (e.g. NCEP CFSR, JRA-55, MERRA-2, and ERA5's predecessor, ERA-Interim). All of the studies we found that compare ERA5 to observations found that ERA5 was the "best-performing" dataset after actual observations, in other words, the best performing reanalysis dataset in comparison to observations. We note the biases these studies found, which were most prevalent in the tropics. We discuss these biases separately for temperature and precipitation given that most studies focus on one or the other. Additionally, we updated Section 3.1 to include a brief discussion of the sensitivity to the choice of reference dataset in QDM and how our application of QDM herein also inherently includes these regional biases found in ERA5.

About the bias correction and downscaling methodology, preservation of the raw model signals is not always the preferred approach (e.g. if the model does not represent basic processes right or has biases in large-scale processes, its local-scale trends should not be trusted either, see Fig. 5 in Maraun et al. 2017). In this sense, some recent works show that bias adjustment can lead to more realistic climate change signals (e.g. by bringing the GCM closer to more realistic regional climate models, Casanueva et al. 2019) and to more plausible threshold-based indices e.g. in insular regions (Iturbide et al. 2022). Also, despite QDM preserves trends in all quantiles, they are not preserved for some moderate/extreme indices based on absolute threshold exceedances. In this sense, it is supported the use of simple and parsimonious bias adjustment methods (e.g. Iturbide et al. 2022). Did authors try other simpler bias adjustment methods together with the downscaling methodology? It is recommended to include some discussion along these points.

This is a great point, and we have added discussion of this to the introduction. It is indeed the case that though QDM preserves trends in all quantiles, they are not preserved for some moderate and extreme indices, and we explore this in the results section in Section 5.2. We did not specifically evaluate simpler bias-adjustment methods in conjunction with the QPLAD method. Previously the authors had extensive experience working with the CMIP5 NASA-NEX-GDDP dataset that exhibited issues with extremes. That formed part of the motivation for finding a method that would perform much better at the tails.

Regarding the potential use of these data by the impact community, many impacts require other variables beyond temperatures and precipitation, which other initiatives as ISIMIP provide (Lange 2019). This limitation should be mentioned.

This comment is addressed by the text additions requested in the L65 comment. We have also updated this section in response to other reviewer comments.

Regarding the GDPCIR dataset, it is written that it "is publicly hosted on the Microsoft Planetary Computer". Does it mean that is is publicly available? Does one need to have an account and/or pay for this service? If that is the case their use by the impacts community is not very straightforward. Why using this system instead of other completely free services, more aligned to the research community? Also, are the files following CF standards for metadata? All this is relevant and should be explained.

Thank you for these questions. We have updated the "publicly hosted" text to mention that it is publicly available, free, and hosted on the Microsoft Planetary Computer. We added an additional section, Section 4.4, which describes the final dataset in more detail, including its size (23 TB), metadata and data type, as well as how to access it. We have also mentioned in the same section how we attempted to follow CF standards whenever possible but do not strictly enforce them. The dataset is hosted on the Microsoft Planetary Computer both because we received support for the project from Microsoft AI for Earth and because we wanted the dataset to be publicly available alongside other widely used geospatial datasets.

Regarding the pipeline, some indication of the running time and resources would be nice to have.

We have added a section to supplemental, Appendix C, on pipeline specifications along with a reference to it in the new Section 4.4.

Specific comments

L9 Here it seems that they use these two methods as independent ones, but later I found that QDM is used for BA and QPLAD for downscaling one after the other, thus, they are used together. Please clarify this in the abstract.

We adjusted the wording in the abstract (Lines 7-10 in the original manuscript) to clarify that QDM is used for bias adjustment and QPLAD used for downscaling.

L34-37 Downscaling and adjustment are a bit mixed in these lines. BA is a mere correction and can be applied even with no resolution mismatch between model and observations. Statistical downscaling comprises also other statistical approaches which actually transfer large-scale information into the local scale (see Maraun and Widmann 2018, Gutierrez et al. 2019), some of which quite sophisticated (Baño-Medina et al. 2021). BA includes an implicit downscaling step if the resolution of the reference data is higher than that in the model, but it is not purely a statistical downscaling method.

Thank you for pointing this out. We have made significant updates to the discussion of bias adjustment and downscaling in the introduction (Section 1) based on this comment and other reviewer comments as well. As part of these updates, we have adjusted the wording so as not to be misleading or convolve the meaning of bias adjustment and downscaling.

L38, 51, 156, 158, 165, 170, 220, 221. Better to use "model" than "GCM" in these descriptive lines since BA can be applied to any (global or regional) climate models.

Thank you for this suggestion. We have updated the terminology to "model" instead of "GCM" in all of those instances except for L51, as that bit was removed from the manuscript, and the last two suggestions in 220 and 221, as there we specifically do want to refer to global climate models rather than regional climate models, which is also consistent with the reference to the Dai 2006 paper.

L41, 150 What do you mean by "standard"? The variety of QM is huge, better to add whether empirical or parametric or a reference to guide the reader a bit more.

We have removed this terminology throughout the manuscript and instead been explicit in discussing different types of QM methods. Thank you for pointing this out.

L53, 54 In Lange (2019) the final resolution of 0.5º is due to the reference data used, not to the BA-SD methodology itself (at the line seems to suggest). I mean, too coarsely resolved BA-projections are due to limited availability of high resolution global datasets (before ERA5, 0.5º was the best one could get on global scale), regardless of the BA or SD applied. It is true that this is a limitation for the use of such projections in impact studies, but the sentence could be better phrased. Note also that 1) CMIP6 models corrected with the BA-SD by Lange 2019 are used in a large intercomparison project devoted to impacts research: ISIMIP), and 2) large resolution mismatch between model and observations is also problematic (Maraun 2013) thus, in general, high-resolution reference dataset is not the solution as long as model data are still too coarse. Also in line 54 "this effect …. dampens or amplifies...", I do not see how coarse resolution dampens or amplifies trends in the tails, the authors may be referring to the BA or SD methodology not resolution per se.

Thank you for raising these issues. Those sentences were not clearly worded. Indeed, we have extensively updated the discussion of downscaling in general and the ISIMIP dataset specifically in this section based on other reviewer comments. Thus, the text to which this comment applied no longer exists.

L59 "Several CMIP6 downscaling datasets", I think the authors mean "Several CMIP6 downscaled datasets".

We have updated this line to "Several CMIP6 downscaled datasets…"

L62 "has made a CMIP6 dataset available", I think the authors mean "has made a bias-adjusted CMIP6 dataset available", or similar, because CMIP6 is available from other public sources (ESGF).

We have updated to "...has made a bias-adjusted CMIP6 dataset available for commercial applications"

L65 It is worth to highlight that ISIMIP provides a large number of variables, not limited to temperatures and precipitation, since many impact studies require them, and this is an important advantage.

Thank you. We added a list of variables that are covered by the ISIMIP CMIP6 dataset, and also added the Lange 2021 ISIMIP3b reference (the correct one for this dataset). We also added to the text that while the additional variables are a key feature to note for ISIMIP, when these additional meteorological variables are not available, a met disaggregator may be used, such as MetSIM. We added the relevant citation, Bennett et al 2020, for this addition.

L69 "and no longer widely used", I think "is" is missing.

Added "is" to both clauses in this line.

L79 Where? Make reference to the appropriate section.

We have added links to the project codebase in Github and Zenodo archive in this line.

L97 I think brackets are missing for the reference.

We fixed this reference (added parentheses).

L99 How many ESGF models where missed? Is the selected subset retaining the ensemble spread of all CMIP6?

To address this question, we added a table in the supplemental material, Table B1, which lists all CMIP6 models and which ones were included in the GDPCIR dataset. For models that were excluded, we list reasons why. This should help to clarify exactly which models we weren't able to include and what the rationale was for each. The selected subset does retain the ensemble spread of CMIP6, and we discuss this by referencing Meehl et al 2020 (which contains ECS values for CMIP6 models) and mention how we include low ECS models, high ECS models, and many in the middle of the range (added to Section 2.1, Table B1 added to Appendix B and referenced in Section 2.1).

L119 Why the 5 extra days in 360-day calendars where filled with the average of adjacent days instead of leaving them as NaN? Averaging could be fine for temperature, but what about precipitation? I checked Pierce et al. 2014 and did not find this filling procedure.

The filling procedure we described was used in the LOCA dataset (described in Pierce et al 2014), however we realize now that neither the LOCA paper nor its supplemental information includes this technical detail. This technical information is on the LOCA website: https://loca.ucsd.edu/loca-calendar/ as well as in a LOCA report (https://gdo-dcp.ucllnl.org/downscaled_cmip_projections/techmemo/Downscaled_Climate_Projections_Addendum_Sept2016.pdf). We have added a citation for the LOCA website, where the filling method is described in the LOCA calendar section on the website. We have clarified this in the manuscript. We chose to fill the days rather than leaving them as NaNs because it is difficult for many impacts modeling applications to have NaNs for any days of the year, as daily temperature and precipitation data is needed. Additionally, we wanted to provide consistency across models in our final downscaled dataset, without having NaNs in some models for some timesteps and not others.

Table 1. I suggest to write the information on the SSPs in a more easy-reading way, with four columns and X denoting the SSPs for each model (in rows). Please also add simulation run.

We have updated Table 1 as suggested and added ensemble members to the table. Please see also Table B1 in Appendix B for a full list of models in the CMIP6 model inventory, including models that are not part of the GDPCIR dataset.

L178 Please clarify what is meant with "Traditional downscaling methods". As said before, typical statistical downscaling builds empirical relationships between large-scale variables (predictors) and local scale predictands (e.g. by means of a linear model), which is not a difference or ratio.

We have taken out the "traditional" terminology and instead are now more specific in discussing bias-adjustment and downscaling methodologies throughout the manuscript.

L187 Which one is the coarse resolution? This information needs to be included in this section, now is found for the first time in line 222. I was wondering which method was used for interpolation and found this information in Sect. 4.1. I would suggest to refer here to Sect. 4.1 with further details.

To clarify, we have added that the coarse resolution is 1-degree in our study. We have also added a reference to Section 4.3 to guide the reader for further detail on our implementation of QDM, which includes a discussion of the interpolation methods used.

L189 Please mention here QDM explicitly, maybe in brackets, to guide the reader.

We have added an explicit mention of QDM.

L222 "adjustment to both", I think the authors mean that the GCM wet-day frequency was adjusted to the observed counterpart, otherwise both are adjusted to what? Also, although less common, it is also a problem when the models are much drier than observation. For this Themeßl et al. 2012 introduced the frequency adaptation. Is there any way you account for such dry biases?

We apologize for the confusing language. We did mean that we applied a wet day frequency correction to both the reanalysis and the GCM data, following Cannon et al. (2015). We detail this further in Section 3.3 where we state that all values at the 1-degree bias-adjustment grid that are less than a specified threshold of 1 mm/day, which was also the threshold used in Hempel et al (2013), are replaced by a nonzero uniform random value between 0.5 to 1, non-inclusive. We also adjusted this text to be more clear.

We do not apply a frequency adaptation as described in Themeßl et al. (2012).

L252 Why did not the authors use conservative for temperatures as well?

Thank you for this question. We do not use conservative remapping for temperatures (as we do for precipitation) because we believe that bilinear interpolation is better for a smoothly varying variable like temperature. We have added an explanation to Section 4.1 of why we used bilinear for temperature and conservative area for precipitation, as well as the pitfalls of any method for precipitation. We also have added a citation for the Rajulapati et al (2021) paper which describes these regridding pitfalls in further detail.

L257-258 "using the regridding method described above" and "using the same regridding methods as in the GCM output" seem to be something different but they refer to the same, right? Please rephrase.

These are the same methods, that is correct. We have updated the text in Section 4.1 in describing regridding of ERA5 to make it more explicit that the methods are the same. The wording for this was confusing before, thank you for pointing it out.

L267 "100 equally spaced quantiles" Do the authors work with 100 percentiles, then? In line 190 it was said that the number of quantiles is equal to the number of timesteps (20x31), as one would expect as Cannon's QDM works with all quantiles (default option), if I am right. Please clarify.

In the referenced Line 190, that description where "the number of quantiles is equal to the number of timesteps" refers to the implementation of QPLAD, not QDM. In Cannon's QDM, 100 quantiles are also used. We have added clarification that these implementations differ between QDM and QPLAD to this paragraph in Section 3.2.

L292 Shouldn't be "percentiles" instead of "quantiles"?

Thank you for the careful read. We adjusted this terminology to "percentiles" and also adjusted the text a bit to clarify that the nearest quantile adjustment factor is applied.

L294 Please explain how the method deals with new extremes, i.e. quantiles of the future period which were not reached in the reference dataset (see different extrapolations in Themeßl et al. 2012).

We have added a description of how the QDM method deals with new extremes to Section 4.2 and discussed the extrapolation methods described in the Themeßl et al. 2012 paper (the method we use is the "QMv1" method evaluated in that study).

L315-316 Is the description of the 2b panel right? It seems to represent the adjustment factor per day of the year and quantile (for the 0.25º gridbox over Miami), I do not see how spatial analogs are shown. The term "spatial analogs" is quite confusing, since for Miami the downscaled value is obtained through the adjustment function calculated between the 1º gridbox and the 0.25º gridbox over Miami, right? As far as I understood the other nearby 0.25º gridboxes do not affect downscaling for Miami, thus "spatial analogs" should be better phrased or clarified. Also how is the analog for each quantile selected from the 620 possible analogs? The mean? Randomly?

Great catch, there was a mistake in L315-316 in describing panel 2b. It is indeed showing the adjustment factors for all quantiles and days of the year. We have updated that text to make it more clear what is being shown in panel 2b.

This interpretation of downscaling for Miami is correct. Nearby 0.25 degree grid cells do not affect downscaling for Miami, however they are related by the coarse 1 degree grid cell that encloses them. In other words, "spatial analogs" refers to an analogous day in the reference (of the 620 days for that quantile and for that day of year) that has a spatial pattern of 0.25 degree grid cells associated with the 1 degree grid cell that encompasses them. Thus "spatial" comes from the sixteen 0.25 degree grid cells while "analog" comes from the fact that the adjustment factors are actual days in the reference time series. We added some language to the text to clarify what is meant by "spatial analogs".

The analog for each quantile is selected from the 620 possible analogs by selecting the nearest quantile to the quantile assigned to that day during QDM bias adjustment. In other words, a day of the year assigned a QDM quantile of 0.25 will get the closest QPLAD quantile to 0.25 from the 620 available ones for that day of the year. We have further updated the discussion of spatial analogs to elucidate this part of the method. Additionally, we added a brief justification of why we chose Miami, Florida for illustrating the method.

L338 I guess this unrealistic values come from the adjustment factors by QDM (please add it e.g. in brackets) since adjustment factors are also applied in the downscaling step.

Yes, we meant that the unrealistic values come from the QDM adjustment factors. We have specified that they come from QDM in this sentence.

L342 "that this" I think one should be removed, otherwise I do not understand the sentence.

We have deleted "that" and added "behavior" to clarify what "this" is referring to.

L414-415 Not sure what this sentence means. Of course results should be consistent with the described methodology. Moreover, is the reference to Fig.2 right?

This sentence was intended to clarify that the bias-adjusted model results shown in Figure 3 were post-processed in the same way as our downscaled model data, which we don't implement as part of the pipeline but instead only for the analysis in Figure 3. The figure reference was incorrect, it should have been 3b. We have updated the figure reference and also clarified the wording so that this distinction in application of post-processing between bias-adjusted and downscaled data is clear, and to make it clear that we apply the same post-processing to bias-adjusted data shown here, but outside of the pipeline.

L418 I was wondering here which data were used in Fig.3b, bias-corrected or downscaled? From this comment about more extremes in higher resolution, it seemed to be downscaled, but then I saw the reference to Fig.A2. Since this confusion comes from time to time, please try to be very clear, e.g. add "after QDM" if you are discussing both but want to refer to bias-corrected only. Also "biascorrected -model" in the figure title should rather be "biascorrected – raw model".

We apologize for any confusion here and have updated the terminology to avoid this. The figure title now reflects the usage of "GCM" rather than "model" and thus it is "bias-adjusted - raw GCM" when referring to the QDM bias-adjusted data. Any reference to "downscaled" data necessarily means that bias-adjustment was also applied, but we are clear about this now throughout the text. Thank you for these suggestions.

L427 Was this behaviour also present in other GCMs? Do the same conclusions about accentuating the Artic amplification hold for other GCMs? Are changes similar? How robust is this result? I would suggest to show this result (Figs. 3 and 4) for the multi-model ensemble median/mean in the supplementary material.

We do see this behavior in other GCMs in terms of Arctic amplification for some seasons, but the strength of the signal depends on the GCM (we investigated this extensively to determine how robust the signal was). To be more comprehensive, we updated figures 3 and 4 to show the ensemble mean across all GCMs included in the GDPCIR dataset and now include all seasons. Additionally, to show more extreme days, we have included results for the 99th percentile for both figures in Appendix A (Figure A7 for maximum temperature and Figure A8 for precipitation). For Figure 4, we now include only days with > 1mm of precipitation to avoid extreme dry days. Additionally, in supplementary materials, we include a version of Figure 4 showing only days < 10mm for the 95^{th} and 99^{th} percentiles (Figures A9 and A10, respectively).

L439 "and" is missing before the second ratio.

We have added in the missing "and" before the second mention of ratio.

Fig.4 caption and titles. Please refer to "raw model" instead of "model".

We updated the language to include the "raw" qualifier. Also, throughout the manuscript, we have updated our usage of the terms "GCM" and "model". Where we discuss QDM and QPLAD methods, we use the term "model" since the methods can be applied to either GCMs or RCMs. When discussing our implementation or results, we use the term "GCM" to denote our application.

L472 In fact, summer days, tropical nights, and annual wet days depend on thresholds. Why is the opposite mentioned? It is precisely in these indices where one can find a fair evaluation of QDM, since it preserves trends in quantiles.

We have updated this section to clarify what we had intended - we did not intend to imply that summer days, tropical nights and annual wet days did not depend on thresholds, but that they were less "extreme" than impacts metrics like days over 95F or precipitation totals above a higher threshold.

L481 Null hypothesis should be rejected if p-value<0.05 and not rejected if p-value >0.05, thus it should be >. Do you use K-S for distributions of annual indices, thus 10, 15, 20 years only? Aren't they too few data to fit distributions?

This was an error, thank you for catching it. The text should have read that the null hypothesis is not rejected if the p-value is > 0.05. We have updated this both in our analysis as well as the text, thus we also recomputed the K-S tests with this update. Additionally, we collapsed the calibration and validation periods into one longer historical period (described in Section 5.2.1) so that the difference in trend during the historical period would not impact the results. Consequently, for the annual indices, 30 data points are used for precipitation and 35 for temperature, and for the seasonal metrics, 4x those numbers (aggregated seasonally on an annual basis). We acknowledge that 30 data points minimum is recommended for Kolmogorov-Smirnov significance testing, thus our results may be impacted by having the minimum number of data points recommended.

Table 2. I do not see the rationale behind the order of the indices within the table. They could go from mean, to moderate and extreme or ordered by input variable. What's the interest of days above 90ºF? To my knowledge that is not an ETCDII index. Also, please use (or add in a new column) the ETCDDI nomenclature in the table, e.g. tropical nights for "tn_days_above". Otherwise, what is the index column representing?

Thank you for this suggestion. We have updated Table 2 so that the order of indices is more intuitive - it is now ordered by input variable. It is true that days over 90F is not an ETCCDI index, but it is a metric that is widely used in impacts modeling, thus we have included it here, along with other metrics that are often used in impacts modeling (e.g. seasonal maximum and minimum temp and precipitation, days over 90, etc). We have also added a new column (the "name" column) which includes the ETCCDI nomenclature for the ETCCDI indices. We also removed the surface variable column as this information is already contained in the "description" column.

Sect. 5.2.1 Why showing results for Miami? GCMs do not represent correctly coastlines, especially those with coarse resolution. Is it a land gridbox in all models original resolution? Do you apply any land-sea mask?

This is a good point. We have updated the analysis in this section to include all cities, both inland and coastal. However, because of the inveterate issues with coastlines in GCMs, we show the inland cities in the main text (still Figure 5) and coastal cities in Figure A11 for completeness.

The land gridbox is not at the GCMs' original resolution; for GCMs, the resolution for the analysis in Figure 5 is at the 1-degree bias-adjustment grid, the resolution for downscaled data is at the 0.25-degree grid, and the resolution for reanalysis is at the native N320 regular Gaussian grid. We do not apply any land-sea mask.

L511 Is here bias adjustment referring also to downscaling?

Yes, here this refers to data after both bias adjustment and downscaling have been applied. We have added clarification to section 5.2.2 that explicitly mentions this.

L517 Please mention that multiplicative factors are used for precipitation.

Thank you for this suggestion. We have updated the description of the equations to reflect the updates we made to the analysis that we have described in the following (L520 comment) response. We have noted that equations 6 and 7 are differences for maximum and minimum temperature as well as precipitation. This text update is in Section 5.2.2.

L520 The error in Eq. 7 is calculated as the difference (in absolute value) between the climate change signal in bias-adjusted data and raw data. Climate change signals are usually calculated over 20-yr periods or more, so I do not see how the error is calculated on annual basis. Furthermore, the error in Eq. 6 between bias-adjusted model and reference should not be calculated annually, because climate model simulations do not have a day-to-day nor year-to-year correspondence with observations, thus the error should be calculated using 20-year periods.

As discussed in our response to the next comment, we updated this calculation. Previously it was computed on a seasonal, daily basis, which as noted is not correct. We updated the error in Eq. 7 to be computed over daily 21-year climatologies, where the median error is the median over all days in the year, with each day composed of a 21-year climatology with a 31-day moving window.

Fig.6 It would be convenient to display the first panel with aspect ratio of 1:1. Also, it would help the interpretation of the results to have information about the error (Eq. 6) or bias in the raw simulation and bias-adjusted (only QDM). One idea could be to have another raw of panels with the scatter plots for these quantities. About these results, it is a bit of an issue that for temperatures the error in present climate of bias-adjusted data (X axis) is larger than the modification of the change signal (Y axis). Is then the modification justified? This should be at least discussed.

Thank you for these suggestions and helpful questions. Because Figure 6 initially only included one GCM for all cities, we decided to update the analysis to include all GCMs. We also updated the error calculations as described in the above comment and in the text. Figure 6 now shows boxplots with the range of error across all GCMs included in the GDPCIR dataset. We also updated our computation of median errors so that it is now performed over the daily climatologies. With the updated analysis of bias-adjustment and trend preservation error, the error in the modification of the change signal is now generally lower across GCMs, with the exception of a few coastal cities. We describe this behavior in the text in Section 5.2.2 and why the level of modification is justified, especially for coastal areas in which the method is doing more work, so to speak, in correcting the change signal from the GCM.

L546 Please mention that these regions are considered for the cities in the previous section. This information was first found in Fig. 7 caption.

This detail is now added to the introductory paragraph of this section (e.g. the first paragraph in Section 5.2).

L558 What is meant by Gaussian interpolation?

Thank you for this question, this term was a typo, the interpolation is not Gaussian. The native ERA5 grid is a Gaussian grid, which does not apply here. We have updated the text to remove the incorrect reference to this in Section 5.2.3.

Figure 8. Correct "False" in Y-axis.

We have corrected this.

Figure A1: is it first mentioned in the conclusions? Then why is not A2 the first one?

We have reordered and added a number of figures to Appendix A, which contains all supplemental figures. We also moved Figure A1 to a new Appendix C section because we felt that the text description of the figure warranted its own section.