Reply on RC2

João António Martins Careto et al.

Author comment on "Added value of EURO-CORDEX high-resolution downscaling over the Iberian Peninsula revisited – Part 2: Max and min temperature" by João António Martins Careto et al., Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2021-208-AC2, 2021

General comments to Referee #2

AC: We are very grateful for your kind and positive comments and suggestions. We appreciate all of them. We sincerely think that your revision allowed an overall improvement of the manuscript

RC2: Part II of this study applies the distribution added value (DAV) to the validation of temperatures from EURO-CORDEX downscalings and their respective parent global datasets.

This manuscript shares its methodology with Part I, in which the same assessment is performed for precipitation. The work done in this study contains original research and fits very well into the scope of GMD.

As with Part I, I found especially the Introduction well written. It does a great job at setting the stage for the following presentation of the data and results. I find, however, a certain mismatch between the nine pages of meticulously presented results and the merely 21 lines of conclusions at the end.

Specific comments:

RC2: As with Part I, I found the use of the term "relevant" and "significant" to be very problematic and misleading. I would highly recommend to refrain from the use of "significant" unless in the context of an objectively calculated measure of significance.

This affects the complete discussion of the results, the conclusion, and the abstract. Especially worrying and ambigious uses are "not as significant" (by what standard?) and even "somewhat significant".

The description of the results is often unclear and difficult to follow. I suggest to carefully proofread the whole manuscript regarding the use of English language.

AC: As in Part I, the term "significant" is used to depict larger values and not derived from some statistical significance analysis. We thank the reviewer for raising these issues and decided to change terms such as "significance", "strong" or "weak" by "higher", "lower", "smaller", "large", "notable", "noteworthy" or "noticeable".

RC2: line 38: Please rephrase, it is unclear what is meant with "these uncertainties are usually derived"

AC: We thank the reviewer for the suggestion and removed the word "uncertainties" since we were referring to the GCM shortcomings.

RC2: I suggest two rephrase instances of "weaker/stronger gains".

AC: We thank the reviewer for the suggestion and as before we checked the whole manuscript and changed the words weaker/stronger to larger, higher, lower or smaller.

RC2: line 74: "spatial representation and correlation" of what?

AC: We decided to remove this sentence from the manuscript.

RC2: line 82: Please explain what is meant with "different driving mechanisms in comparison to observations".

AC: We decided to remove the whole sentence from the text.

RC2: line 182: unclear what is meant with "reveal closer proximity".

AC: Here we mean that the PDFs almost overlap with the observational one. To avoid confusion for the readers we changed to "…overlap more reasonably than…".

RC2: line 184: I do not understand the use of "namely" here.

AC: We decided to change "namely" to "where" and "with" to "reveal" in the same sentence.

RC2: line 204: What is a small bias in this context? What magnitude are we talking about? And do you mean a small ratio of standard deviations, or a ratio close to 1?

AC: The PDFs for temperature closely follow a Normal Distribution, thus changes in the mean and standard deviations between models and observations will have impact on the DAVs. However, we did not compute these differences as it falls outside the scope of this work and relied instead in previous studies that used the exact same observational

dataset to evaluate the EURO-CODEX models.

RC2: line 211: Why does the data assimilation of temperature observations in ERA-Interim make it so that it is diffuclt to achieve added value only for TASMIN, but not for TASMAX? Without further explanation, I find this argument speculative.

AC: On this sentence "Another reason could be owed to the incorporation of temperature observations into the ERA-Interim" we do not specify if we are referring to either TASMAX or TASMIN. In fact, the assimilation of temperature into this reanalysis makes it extremely difficult for models to reveal added value and this is applied for both temperatures. Still the issues around 0 $^{\circ}$C occurs for the TASMIN and not the TASMAX. To avoid confusions, we changed the text to:

"The results here are a reflection of the smaller differences between Era-Interim and the IGD due to the incorporation of temperature observations into ERA-Interim (Prömmel et al., 2010), together with the overestimation of the temperature bins around 0 $^{\circ}$C for the RCMs (Fig S1 right side)."

And removed line 211.

RC2: line 216: Please rephrase "the panorama is not so different".

AC: We thank the reviewer for this suggestion and changed "panorama is" to "values are".

RC2: line 219: Please rephrase "contradicts the results".

AC: We thank the reviewer for the suggestion and change "contradicts the results" to "reveals a different result".

RC2: line 226: unclear: "allied with improvements"

AC: We decided to change to "together with improvements".

RC2: line 229: unclear: "close performance to the driving ERA-Interim"

AC: Here we mean that the scores of both high and low resolutions are very close. Still to avoid confusion by the reader we changed the statement to: "…for autumn reveal neutral DAVs representing a similar performance to the driving ERA-Interim in terms of the Perkins skill score."

RC2: line 261: unclear: "strong or most points"

AC: We decided to remove "strong or". The phrase changes to: "The overall results correlate well with those shown in Fig. 2a i.e., models with most points with added value

will inevitably reveal gains at the Iberian Peninsula scale and vice-versa."

RC2: line 275-277: unclear

AC: We added some examples to the sentenced and changed accordingly to "At the seasonal scale, RCMs such as CCLM, CNRM63 DMI for summer or CNRM53 for winter, which reveals detrimental effects in locations from the interior also show noteworthy losses in Fig. 2b."

RC2: line 289: unclear "lower values are derived from the losses"

AC: In this case we are comparing the regional and the spatial figures. We added " …in the regional DAVs…" after "These lower values…"

RC2: line 318: unclear "no connection is found for each downscaled GCM"

AC: We decided to change the phrase to "The overall results for RCMs driven by the same GCM are similar for most cases, while no connection is found between models forced by different GCMs."

RC2: line 323: not sure about the use of "yet" here. Is there really a contrast between the two sentences?

AC: Yes, there is a contrast because in this sentence "In fact, the CNRM GCM in McSweeney et al. (2015) reveal a good performance, together with NOAA, MOHC and MPI." We are referring to the good performance of these GCMs, which means that they could possibly reveal higher Perkins skill score value, thus it will be more difficult for models to display an added value. However, the EURO-CORDEX RCMs are still able to substantially improve the driving information.

RC2: line 349: unclear

AC: This phrase is an introduction for the following sentences, still we changed to: "From the PDFs in Fig S4, one can infer the large gains found for TASMIN and for TASMAX in Fig. 4.".

RC2: I find the last section to be very short and somewhat lacking. The whole section is written in one paragraph of 21 lines and seems, to be honest, incomplete.

I recommend to revisit this section, provide a thorough summary, present the main conclusions, and discuss their relevance in context to other studies and to the field of regional climate modelling as a whole.

AC: We thank the reviewer for the suggestions. We revised the whole section, and we

kindly ask the reviewer to read down below the changes done to the document in order check if we fulfilled the request:

"In this work a Distribution Added Value metric proposed by Soares and Cardoso (2018) is used to gauge the added value of higher resolution simulations confronting each low-resolution driving GCM or Era-Interim, with the IGD observations over the Iberian Peninsula as a baseline. The DAVs were applied to the EURO-CORDEX Hindcast (1989-2008) and Historical (1971-2005) simulations for the variables maximum and minimum temperatures. Overall, the high-resolution models are able to provide some added value, particularly for TASMAX. On the other end, minimum temperature reveals some difficulties in obtaining added value for the Hindcast simulations, mainly when the whole PDF is considered and also for the extremes in the Historical simulations, partly owed to problems associated with the snow-albedo-atmosphere feedback, derived from uncertainties related to snow cover, depth, and melt (García□Díez et al., 2015; Minder et al., 2016; Terzago et al., 2017). These uncertainties substantially affect the PDFs from the RCMs around 0 ^{o}C, by overestimating the frequency of events in comparison with the observations. This poor representation has a more significant impact on the Hindcast simulations, where the ERA-Interim reanalysis does not reveal these kinds of issues owed to the incorporation of temperature observations (Prömmel et al. 2010). Nevertheless, the finer details from the downscaling allow a more spatial variability of temperature, which in the end could result into added value. This fact is particularly relevant for coastal regions, where the improvements of the boundary between ocean and land result in larger gains in comparison to points located in the interior. For the maximum and minimum temperature extremes, the results are more limited. Still for the TASMAX extremes winter and summer reveals an evident added value, contrasting with the neutral or negative DAVs for spring and autumn. As for the TASMIN extremes, depending on the season, the DAVs depend more on the season. For instance, the 10th percentile for winter is too low, not incorporating the problems around 0 ^{o}C, thus revealing some added value. The opposite occurs for spring and autumn which revealed more neutral values. As for summer the temperatures are too high for snow in most part of the territory, yet the models still revealed losses.

For the Historical, no connection is found between each GCM downscaling group, while the results for multiple RCMs forced by the same GCM reveal a similar range. Each driving simulation has its own resolution and performance, which could impact the added value from the higher resolution as in Careto et al (2021) for precipitation. However, the interpolation with the orographic correction with a constant lapse rate might in the end have a more significant impact in the performance of the individual GCMs and RCMs. Moreover, other factors such as the parametrizations of smaller scale processes and the representation of feedback systems can have a major impact in how a variable represent reality. Another factor that may play a major role is related to how well do GCMs represent storm tracks. If a GCM is not able to properly represent storm tracks, then the downscaling RCMs will inherent these issues.

Similar to the Hindcast simulations, the gains of the Historical spatial DAVs are also more focused in coastal regions. Models which reveal more substantial gains and neutral or slightly positive DAVs in the interior, tend to have higher values at the Iberian Peninsula scale. Although, care must be taken when comparing the spatial and regional DAVs as both follow slightly different approaches. Still both methods correlate well, primarily for the whole PDF. In the case of the extremes, the different thresholds for each individual point makes it difficult for a direct comparison.

A second methodology was also implemented, following the secondary results in Careto et al (2021). In this case, all data is interpolated with an orographic correction to the resolution from the observations. The interpolation from the high-resolution observations to each of the low-resolution model grids degrades the observational PDF. However, by

downscaling the driving low-resolution with an orographic correction considering a constant lapse rate, unrealistic values can be generated, not only due to the interpolation, but also derived from the higher resolution orography. This method substantially improves the representation of temperature at the cost of not considering land-atmosphere feedbacks, thus resulting in a larger uncertainty. In fact, for this case, the low-resolution scores are improved, resulting in an overall lower DAV.

The higher resolution EURO-CORDEX models do reveal some added value for temperature. In many situations the RCMs can improve the lower-resolution signal, all relative to the observational dataset considered. For instance, in the Historical simulations, the high-resolution mitigate the annual and seasonal biases from the GCMs, due to a better representation of the observable PDF. On the other hand, the problems found for the RCMs, particularly for the Hindcast simulations do affect the DAV values negatively, with these models revealing losses. Nevertheless, the higher resolution, together with a finer representation of the local topography, improved coastal resolution and consequently land-sea contrasts, and also improved local feedback systems which ultimately leads to an add value.

There are very interesting results in this manuscript, and I am looking forward to a revised version.

Technical comments:

RC2: Throughout the manuscript: The unit "km" is written as "Km", often with a space missing in front of "Km".

AC: We thank the reviewer for pointing out the issue and changed accordingly.

RC2: Please consider to use either British English (kilometre, analysed, normalised) or American

English (~ized, ~zation), instead of mixing the two.

AC: We thank the reviewer for the suggestions and checked the whole document.

RC2: The term "convection permitting" is wrongfully written as "convective permitting".

AC: We thank the reviewer for the suggestion and decided to change these terms.

RC2: Please remove commas before in citations like "Lhotka, (2018)"

AC: Corrected.

RC2: line 17: Please consider using "development" instead of "appearance"

AC: We thank the reviewer for the suggestion and changed to "development".

RC2: line 18: "spamming" -> maybe supposed to be "spanning" ?

AC: Yes, Corrected.

RC2: line 34: There is something off with this sentence.

AC: Corrected. We changed this sentence to "Over the years, the increasing public attention and motivated by the Intergovernmental Panel on Climate Change (IPCC), tools for assessing past, present and future climate conditions were developed, the so-called Global Climate Models (GCM)." and removed the following one.

RC2: line 46: dremove the "." after reference.

AC: Corrected.

RC2: line 66: affects -> affect

AC: Corrected.

RC2: line 72: rephrase "over for most"

AC: Corrected. Removed "for".

RC2: line 85: overperform -> outperform ?

AC: Corrected.

RC2: line 89: Please rephrase "while underestimate for lower altitudes".

AC: We decided to rephrase the paragraph to: "A common issue across all RCMs is related to the snow-albedo-temperature feedback, through a misrepresentation of the surface energy balance (García☐Díez et al., 2015; Minder et al., 2016; Terzago et al., 2017). Uncertainties in the snow depth, melt and cover could have a potential impact on surface air temperatures around 0 ^{o}C. Biases in the albedo representation leads to positive feedback, thus enhancing too cold temperatures during winter."

RC2: line 94: fix capitalisation in "probability Density Functions"

AC: Corrected.

RC2: line 301: "Although" does not fit to the rest of the sentence. And "spam" -> "span"?

AC: Corrected, changed to "Despite…" and yes it is supposed to be "span".

RC2: line 360: rephrase "superior"

AC: Corrected from "…superior to…" to "…above…"

RC2: line 362 and 366: "had" -> "have"

AC: Corrected.