GMD-2021-294 Model evaluation paper

Answers to RC1:

We thank the reviewer for the suggestions and we answered (in blue) to each single comment (in black), focusing on the best solution to satisfy RC1 and RC2 comments (both available on the online discussion).

The paper evaluates the impact of an increase of horizontal resolution on the representation of extreme events in CMIP5 and HighResMIP type models. I think that this could become a valid contribution, but I would like to encourage the authors to extend the evaluation and to provide additional information and plots as outlined in the comments below.

As already stated in the introduction we tried to match as much as possible RC1 suggestions in the new version of the manuscript.

Major:

- The use of "high" resolution for 1 degree and "very high" resolution for ¼ degree is misleading. Why not use "standard" and "high" resolution. Otherwise, you would need to come up with "extremely" high and "ultra" high for the next resolution upgrades.

  In the new version of the manuscript we use the terms "standard" and "high" instead of "high" and "very-high".

- To my best knowledge, ERA5 does not assimilate precipitation and precipitation is only available as a diagnostic of 18h forecasts that are used for data assimilation. This is not a big problem, but could you please specify in the text where the precipitation field is coming from? Also, the grid and resolution of ERA5 should be specified (I think it is not ¼ degree).

  In the new version of the manuscript we do not rely on ERA5 precipitation for model evaluation. Also following the Reviewer #2 comment, we decided to use a new observational dataset (in addition to the already involved CHIRPS dataset) instead of ERA5. The MSWEP (Beck et al. 2019) dataset is a global precipitation product with a 3-hourly 0.1° resolution available at a 3-hourly temporal resolution, covering the period from 1979 to the near present. The dataset takes advantage of the complementary strengths of gauge-, satellite-, and reanalysis-based data to provide reliable precipitation estimates over the globe. With this dataset we compute seasonal averages and both daily and 6-hourly percentiles to evaluate model results (same as done based on ERA5 in the previous version of the paper, but over a shorter period [1981-2014).

  The ERA5 resolution is now made explicit in section 2.2.

  References:

  Beck, H. E., Wood, E. F., Pan, M., Fisher, C. K., Miralles, D. G., van Dijk, A. I. J. M., McVicar, T. R., & Adler, R. F.. MSWEP V2 Global 3-Hourly 0.1° Precipitation: Methodology and Quantitative Assessment, Bulletin of the American Meteorological Society, 100(3), 473-500, 2019.

- I would recommend comparing against two observation datasets for precipitation. This would allow to judge how much (or little) we actually know about mean and extreme precipitation fields.
  As introduced in the previous comment we added a new precipitation dataset (MSWEP, Beck et al. 2019) for the evaluation of the model ability in representing averages, extreme and intense events at the 6-hourly and daily time frequency.

- The evaluation of extreme events is interesting. But please also discuss the mean temperature and precipitation fields in more detail and include plots of the mean fields and biases. It is difficult to judge the quality of the representation of extreme events if the quality of the mean field representation is unclear. In particular, as you are referring to average representations for temperature and precipitation in the summary.
  Temperature and precipitation averages, commented in the manuscript, are shown in supplemental figures S1-S6, now defined following the reviewer suggestion to "put the reference and the bias fields into the same figure" keeping separate "6h/24h and DJF/JJA" fields (see the last RC1 major comment). Be aware that, following the reviewer #2 suggestion, we had to change the color schemes used, to present the data based on the IPPC visual style guide.

- Why are you focussing so much on the 99th Please also add plots and discussion of the 90th percentile.
  We added maps of the 90th percentiles for temperature (figures S7-S10) and precipitation, computed with MSWEP (daily S11-S12 and 6-hourly S13-S14) and CHIRPS (only at the daily frequency, S15-S16). These figures are now considered when commenting intense events in sections 3 and 4.
  Old figure S4 is now S17.

- There should be more discussion on the impact of a resolution upgrade on temperature and precipitation predictions that was observed by other modelling groups.
  Regarding the dependence of the extreme temperature representation on the horizontal resolution, the relative summary section has been modified as:
  "*It is well known that the representation of precipitation extreme indices is more dependent on the horizontal resolution than what we expect for temperature extreme indices (Wei et al. 2019). Anyway, on average, the highest resolution CMCC model (VHR) is better than the lower resolution model (HR) in representing average intense (90p) and extreme (99p) events of temperature both in terms of patterns and magnitude. This is true for daily and 6-hourly based statistics. Also VHR results are quite in agreement with CMIP6 multi-member average of daily intense and extreme temperature indices (Scoccimarro and Navarra, 2021)*"

  Regarding the dependence of the extreme precipitation representation on the horizontal resolution, the relative summary section has been modified as:
  "Regarding the precipitation distribution, the VHR model performs better in representing averages and intense events, but more pronounced biases appear in VHR compared to HR when focusing on extreme events, with a more evident degradation

in the daily statistics compared to the 6-hourly. This latter result reduces the confidence we usually attribute to the highest horizontal resolution in modelling extreme precipitation, and is consistent with single model analysis based on CAM5.1 atmospheric model (Wehner et al. 2014) suggesting a positive bias over most of the globe in the representation of extreme events at ¼ degree horizontal resolution, and also with multi-model recent findings (Bador et al. 2020) suggesting that highest resolution models tend to produce more pronounced extremes than lower resolution ones. . In addition many of them show lower skill in representing observed patterns, both in terms of intensity and spatial distribution, at higher resolution compared to their corresponding lower resolution version."

Added references:
Scoccimarro E., Navarra A.: Precipitation and temperature extremes in a changing climate. Chapter 2 in "Hydrometeorological Extreme Events and Public Health" Wiley book, 320 pages. ISBN: 978-1-119-25930-5.

Wehner MF et al.: The effect of horizontal resolution on simulation quality in the Community Atmospheric Model, CAM5.1. J. Adv. Model. Earth Syst. 6, 980–997. doi:10.1002/2013MS000276, 2014.

Wei, L. X., X. G. Xin, C. Xiao, et al.: Performance of BCC-CSM Models with Different Horizontal Resolutions in Simulating Extreme Climate Events in China. *J. Meteor. Res.*, 33(4): 720-733. doi: 10.1007/s13351-019-8159-1, 2019.

There should also be a discussion how the move to storm-resolving resolutions could change the situation.
We added the following sentence to the summary section:
"In principle, horizontal resolution increases should improve the representation of extreme storms such as tropical cyclones (Scoccimarro et al. 2020) and for this reason also the representation of the associated short term extreme precipitation should improve, but this is not the case for the model object of this study, and it is also confirmed by recent studies on the same topic (Wehner et al. 2021)."

Furthermore, I would like to know how the parametrisation schemes change when resolution is increased from 1 degree to ¼ degree.
No change was applied, to follow the PRIMAVERA (EU project) protocol, as now specified in section 2.1 on *numerical experiment description*.

- It is very hard to compare the fields in the figures at the moment. Please put the model fields, the reference (ERA or observations) and the bias fields into the same figure. You can separate 6h/24h and DJF/JJA. At the moment, a lot of flipping through the pages is required to compare the fields.
  This was our first choice when starting to collect and present our results, but then we moved to the "shorten" version you have seen, for readability. Anyway, to satisfy the reviewer request, in the new version of the manuscript we went back and all the figures are presented following the suggestion to put the model fields, the reference and the bias fields into the same figure, separating 6h/24h and DJF/JJA. Adding also

the 90th percentile maps we now have 10 figures in the main manuscript and 17 supplemental figures, where S7-S16 refer to the 90th percentile.

Minor:

l17: "for the definition of the extreme condition" Please re-word

Rewritten as: "For a more detailed evaluation we use both 6-hourly and daily time series, to compute indices representative of intense and extreme conditions."

l21: "for average precipitation"

Sorry, I don't understand this point.

l26: "lost opportunities" What does this mean?

We removed this part of the sentence that in the new version is:

"An extreme climate event can have an impact on human activities, either as direct and indirect damages and, unfortunately also as loss of human life."

l29: "GCM simulations" -> Simulations of GCMs"

Done.

l178-179: This should be re-worded

Rewritten as:

"*This result suggests that a higher horizontal resolution is not sufficient to improve the representation of extreme temperature events at the highest time frequency considered. Consequently, the worsening of model biases in high frequency (6-hourly) temperature statistics derives from deficiencies of the current version of model components and parameterizations in representing high-frequency processes."*

Figure 1: This may be an ignorant question, but I guess the 99th percentile could also be for negative temperature values. Whether you are looking into hot or cold temperatures should be specified somewhere.

Not sure to understand the point. This is the 99th percentile computed over temperature time series that can contain negative values too. With this said it can be that, at least at high latitudes, the 99th percentile is still negative, despite sitting on the right tail of the temperature distribution.

Figure 5: Please use [mm/d] and not [%].

Done.