GMD-2021-294 Model evaluation paper

Answers to RC2:

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 RC2 and RC1 comments (both available on the online discussion).

Major comments

1. The introduction does not provide sufficient context to the study. In particular, the authors should expand it to include a discussion on prior studies which have looked at the impact of resolution on the ability of GCMs in simulating extreme events.

We thank the reviewer for this suggestion and in the new version of the manuscript we added the following part to the introduction, also supporting the discussion of the results presented in section 5:

"Regarding the extreme temperature representation, based on data at the daily frequency, it has been shown that GCMs tend to have warm bias over most land areas (Li et al., 2021) and the horizontal resolution plays a minor role with respect to the one played in the extreme precipitation representation (Kharin et al. 2013; Wei et al., 2019). Typically the warm extremes are computed based on maximum daily temperature, but in this work we want to verify the potential improvements induced by the increased resolution in the representation of extreme temperature events defined at two different time frequency (daily and 6-hourly). For this reason we investigate the distribution of daily and 6-houry average temperature, instead of maximum daily temperature (Scoccimarro and Navarra, 2020).

Regarding the extreme precipitation representation, Based on simulations from single GCM, some improvement in skill at higher resolution for some measures of extreme precipitation over certain regions of the globe have been found in the past (Wehner et al. 2014, Kopparla et al. 2013) and only recently, multi-model assessment on this topic have been done, confirming that increasing the horizontal resolution to ¼ of degree (the highest adopted by the model object of this study), the magnitude of simulated daily (Bador et al. 2020) and sub-daily precipitation (Wehner et al. 2021) extremes is increased. On the other hand this is not associated to a systematic improvement in the simulation of precipitation extremes when compared to observations and, quantitatively, at the global scale, the intensification of precipitation extremes at increased resolution varies substantially from model to model (Bador et al. 2020). Also, for grid point GCMs (as opposed to spectral GCMs), the fraction of land precipitation increases, largely due to better resolved orography (Vannière et al., 2019; Terai et al., 2018; Demory et al., 2014)."

Added References:

-Demory, M.-E., Vidale, P. L., Roberts, M. J., Berrisford, P., Strachan, J., Schiemann, R., and Mizielinski, M. S.: The role of horizontal resolution in simulating drivers of the global hydrological cycle, Clim. Dynam., 42, 2201–2225, https://doi.org/10.1007/s00382-013-1924-4, 2014.

-Kopparla, P., Fischer, E. M., Hannay, C., & Knutti, R.: Improved simulation of extreme precipitation in a high-resolution atmosphere model. Geophysical Research Letters, 40, 5803–5808. https://doi.org/10.1002/2013GL057866. 2013.

- Kharin, V.V., Zwiers, F.W., Zhang, X. *et al.* Changes in temperature and precipitation extremes in the CMIP5 ensemble. *Climatic Change* **119,** 345–357. https://doi.org/10.1007/s10584-013-0705-8. 2013.
- Li, C., Zwiers, F., Zhang, X., Li, G., Sun, Y., & Wehner, M.. Changes in Annual Extremes of Daily Temperature and Precipitation in CMIP6 Models, *Journal of Climate*, *34*(9), 3441-3460. 2021.
- 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. 2021
-Terai, C. R., Caldwell, P. M., Klein, S. A. Tang, Q., and Branstetter, M. L.: The atmospheric hydrologic cycle in the ACME v0.3 model, Clim. Dynam., 50, 3251–3279, https://doi.org/10.1007/s00382-017-3803-x, 2018.
-Vanniere, B., Vidale, P. L., Demory, M.-E., Schiemann, R., Roberts, M. J., Roberts, C. D., Matsueda, M., Terray, L., Koenigk, T., and Senan, R.: Multi-model evaluation of the sensitivity of the global energy budget and hydrological cycle to resolution, Clim. Dynam., 52, 6817–6846, https://doi.org/10.1007/s00382-018-4547-y, 2019
-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.

Field Code Changed

2. While I agree with using two products to evaluate the model's precipitation, I would have to disagree with the use of ERA5 for that purpose. Precipitation is not assimilated in reanalyses and is thus a product of the model used to create it. Although ERA is a superior product to its predecessor, there are many known issues with ERA5 precipitation. See for example:
Rivoire, P., Martius, O., & Naveau, P. (2021). A comparison of moderate and extreme ERA-5 daily precipitation with two observational data sets. Earth and Space Science, 8, e2020EA001633. https://doi.org/10.1029/2020EA001633
Hu, G., Franzke, C. L. E. (2020). Evaluation of daily precipitation extremes in reanalysis and gridded observation based data sets over Germany. Geophysical Research Letters, 47, e2020GL089624. https://doi.org/10.1029/2020GL089624
Crossett et al. (2020) Evaluation of Daily Precipitation from the ERA5 Global Reanalysis against GHCN Observations in the Northeastern United States. Climate, 8, 148; doi:10.3390/cli8120148
It would thus be better to use another observational product to evaluate the model.
In the new version of the manuscript we do not rely on ERA5 precipitation for model evaluation. Also following the Reviewer #1 comment on the same topic, 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 three suggested references have been added to the text to justify the choice to do not use the ERA5 precipitation for comparison. This is the sentence added to the text:

"*Since there are many known issues with ERA5 precipitation (Rivoire et al., 2021; Hu et al., 2020; Crosset et al. 2020), for the evaluation of the model performance in representing the precipitation distribution we build on MSWEP version 2 observational data set (Beck et al. 2019): The Multi-Source Weighted-Ensemble Precipitation (MSWEP) global precipitation dataset is available at a 3-hourly temporal resolution, covering the period from 1979 to the near present, with an horizontal resolution of 0.1 degrees. The dataset takes advantage of the complementary strengths of gauge-, satellite-, and reanalysis-based data to provide reliable precipitation estimates over the globe*."

Added 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.

-Rivoire, P., Martius, O., & Naveau, P. (2021). A comparison of moderate and extreme ERA-5 daily precipitation with two observational data sets. Earth and Space Science, 8, e2020EA001633. https://doi.org/10.1029/2020EA001633

-Hu, G., Franzke, C. L. E. (2020). Evaluation of daily precipitation extremes in reanalysis and gridded observation based data sets over Germany. Geophysical Research Letters, 47, e2020GL089624. https://doi.org/10.1029/2020GL089624

-Crossett et al. (2020) Evaluation of Daily Precipitation from the ERA5 Global Reanalysis against GHCN Observations in the Northeastern United States. Climate, 8, 148; doi:10.3390/cli8120148

3. I am not sure that daily mean 2m temperature is the best variable to evaluate extreme temperatures in the model. Why not use tasmin and tasmax? And Figure 3 could be sent to the supplementary material as it is hardly discussed.

We agree that looking at the distribution of daily tasmax is different from looking at the distribution (and relative tails) of average daily temperature, but the approach used in the current manuscript (also used in Scoccimarro and Navarra, 2021) has some advantages such as the fact that tasmax parameter depends on the model time step length (different in the different versions of the model) while average temperature (daily or 6-hourly) is independent from the model time step. Also, the usage of values averaged over a period (daily or 6-hourly) instead of tasmax gives an information more exhaustive from the human health perspective: e.g. a few minutes (model time step) with 42ºC might be less problematic for the human body than 6 hours at 38 ºC.

In addition, since tasmax is defined only at the daily frequency (this is true for all of the CMIP5 and CMIP6 model output available on ESGF), it is impossible to compare the model horizontal resolution role in representing daily and 6-hourly statistics.

Last but not least, we recently retrieved CMCC-HR4 and CMCC-VHR4 tasmax and tasmin fields from the ESGF repository because we found a bug on both these daily datasets.

4. The color schemes used to present the data makes it difficult to understand the results. For one, it saturates very quickly. For example, on Fig. 1, it is nearly impossible to distinguish values between -6 and -20 (when it is printed on paper). And also, there are similar colors on both sides of the 0 point (e.g. green on Fig 3.). It made reading through the precipitation subsection particularly difficult, as I couldn't get a good sense of the size of the biases that were being shown. My suggestion would be to refer to the IPCC visual style guide:
https://www.ipcc.ch/site/assets/uploads/2019/04/IPCC-visual-style-guide.pdf
Following this suggestion, In the new version of the manuscript color schemes have been defined following the IPCC visual style guide. We also had to modify the structure of the figures to put the model fields, the reference and the bias fields into the same figure, dividing 6h/24h and DJF/JJA to follow reviewer #1 request.

5. I think the manuscript would benefit from an attempt at explaining some of the results that are presented. The authors described the convection scheme in Section 2.1, because "it is worthwhile to mention for our discussion on precipitation biases", but the convection scheme is never referred to when the results are presented. Does it explain the differences between results obtained with 6-hourly data and daily mean? And if so, how? Could dry biases in the model play a role in the extreme of near surface temperature? Was the impact of resolution on extreme temperature and precipitation evaluated by other groups using the CAM model? Are the results consistent with those found here? Furtheremore, Vaniere et al. (2019) has shown a significant impact of resolution on precipitation over mountainous areas in HighResMIP models. Are the results presented here consistent with that study and others that have looked into this issue previously? Given that this is a single model study, it is difficult to evaluate if the results are model dependent. Expanding the discussion would help in that regard.
Vaniere et al. (2019) Multi-model evaluation of the sensitivity of the global energy budget and hydrological cycle to resolution. Climate Dynamics, 52, 6817–6846
We extended the description of the convection scheme in the standard version of CAM4 comparing it to the one adopted by CAM5:
"*In other words the deep convection scheme is triggered based on a minimum positive threshold of CAPE, same as in the standard resolution of the CAM5 model (Wang and Zhang, 2013)."*
And this is supporting the added text in the discussion:
"*The high-resolution version of the model generates excess extreme precipitation in the wet, warm regions, or seasons, consistently with findings based on experiments carried out with the CAM5 atmospheric model at the same resolutions (Wehner et al, 2014), highlighting once again the importance of an extensive model tuning at the high resolution*".
The differences between extreme precipitation biases in 6h and daily data moving from standard to high resolution is not that evident, thus we didn't link this to the description of the precipitation parameterization.
Regarding the role of dry biases we assume that this comment is related the 99p bias since the average precipitation (S3 and S4 figures in the new version of the manuscript) tends to show a wet bias. A first investigation on the role of such dry bias in modulating extreme near surface temperature, does not suggest a systematic relationship: the bias of the 99p of precipitation during summer (Figure 6) is dry for the low resolution model but wet for the

high resolution model over part of the Maritime Continent and South America, while the bias in 99p of near surface temperature is positive for both models, over the same regions (Figure 4). During winter, for both models, the most pronounced positive biases in 99p of temperature (Figure 3) are over regions where the bias in the 99p of precipitation is negligible (Northern Hemisphere north of 70°N) and over South America where the bias in the 99p of precipitation is positive in the standard resolution version and negative in the high resolution version (Figure 5).

The work by Vanniere et al. (2019) has been now mentioned in the introduction (see the answer to your first comment) and a comment on the CMCC-CM2 model results within the Venniere et al. analysis is provided in the conclusion as:

"In addition it is important to note that moving from the standard to the high resolution of CMCC-CM2, the model behaves as most of the models participating to the HighResMIP project with the tendency to an increased fraction of land precipitation in the highest resolution, same as for the fraction of land precipitation caused by moisture convergence increased with resolution (Venniere et al. 2019). Also, in CMCC-CM2 model, the orographic precipitation captures most of the change of precipitation due to resolution, consistently with most of HighResMIP models (Venniere et al. 2019)".

Added references:

- -Vanniere, B., Vidale, P. L., Demory, M.-E., Schiemann, R., Roberts, M. J., Roberts, C. D., Matsueda, M., Terray, L., Koenigk, T., and Senan, R.: Multi-model evaluation of the sensitivity of the global energy budget and hydrological cycle to resolution, Clim. Dynam., 52, 6817–6846, https://doi.org/10.1007/s00382-018-4547-y, 2019
- Wang X., Zhang M.: An analysis of parameterization interactions and sensitivity of single-column model simulations to convection schemes in CAM4 and CAM5. Journal of Geoph. Research Atm. https://doi.org/10.1002/jgrd.50690. 2013.
- 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.

Minor comments

CMCC-CM2-HR and CMCC-CM2-VHR might be the name of the models, but it is strange to refer to a model with a resolution of 1 deg (for the atmospheric component) as high resolution. It might be easier for the reader to simply refer to the two configurations as standard resolution (1 deg) and high resolution (0.25 deg). To be clear, I am not suggesting changing the name of the models, but simply to use the terms standard resolution and high resolution (or something along those lines) when referring to MCCCM2-HR and MCC-CM2-VHR.

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

The authors should mention the name of the experiment from which the data are taken. Vaniere et al. (2019) noted different responses in terms of the impact of resolution on precipitation between grid point and spectral models. As such, the type of atmospheric model should be highlighted and the authors should mention whether their results are consistent with that prior study.

We added a sentence in 2.1 section to indicate the experiment from which the data are taken: *"In the current analysis we investigate the hist-1950 HighResMIP experiment as described in section 2.3.".*
Also we made explicit the grid point configuration in section 2.1:
"*The CMCC general circulation has been developed in several configurations (Cherchi et al. 2019). The model uses as atmospheric component the CAM Atmospheric component (CAM4, Neale et al. 2010) in its grid point configuration*"
A comparison of CMCC-CM2 model results to other HigResMIP results, based on Venniere et al. analysis, is now part of the discussion (see the answer to your last major comment).

p.1, line 26: "A climate variation can have an impact on human activities...". I am not sure what the authors mean by "climate variations: in this context, but this phrasing is a bit odd. I would suggest rewriting.
Rewritten as:
"*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*."

p. 1, line 27: "Extreme climate events are involved in the vast majority of the most severe episodes." The most severe episodes of what?
The sentence has been removed also in accord with the rephrasing of the previous one.

p. 2, line 32: "was designed to understand the role of the horizontal resolution." The role of horizontal resolution on what?
Rewritten as:
" *.. was designed to understand the role of the horizontal resolution in improved process representation in all components of the climate system*"

p.2, line 33: "based on two versions of the GCM"
Done.

p. 2, line 34: " differing only in their atmospheric horizontal resolution"
Done.

p.2 line 41: "However, such analyses has employed rather low frequency data…"
I am not sure what analyses the authors are referring to (or what they mean by low and high frequency), but many studies have used daily or sub-daily data to look at extremes in climate models. See for example:
Wehner M, Lee J, Risser M, Ullrich P, Gleckler P, Collins WD. 2021 Evaluation of extreme sub-daily precipitation in high-resolution global climate model simulations. Phil.Trans.R.Soc.A379: 20190545. https://doi.org/10.1098/rsta.2019.0545
Wehner MF et al. 2014. 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.
And references therein.
The sentence has been modified as;

"However, most of the analyses on extreme events employ rather low frequency data (tipically daily), and short-duration high-intensity precipitation events can easily escape detection if high-frequency data are not used (Meredith et al. 2020, Scoccimarro et al. 2015)."

The two mentioned references are part of the manuscript and helped to improve the manuscript based on the answer to the major comments of the reviewer, see for instance new line 70:

"..and only recently, multi-model assessment on this topic have been done, confirming that increasing the horizontal resolution to ¼ of degree (the highest adopted by the model object of this study), the magnitude of simulated daily (Bador et al. 2020) and sub-daily precipitation (Wehner et al. 2021) extremes is increased "

and new line 507:

"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 analysis on the same topic (Wehner et al., 2021)."

p.2 line 62: "The two models object of this study..   degree in VHR."
I would recommend moving this sentence to the previous paragraph when the authors discuss the atmospheric component of the model.

Done.

p.4, line 115: "Also, the positive HR DJF bias over eastern Europe is more than halved in VHR". To me, it seems like it disappears, but it might be due to the colorbar.

This is now more clear based on the new color scheme proposed (see new Figure 1 lower panels).

p.4, line 118: "The positive extreme temperature bias between 30N and 60N shown by the HR model during JJA is partially reduced in VHR." This seems to happen mostly over Europe and Asia, not so much North America.

Rewritten as:

"The positive extreme temperature bias between 30°N and 60°N shown by the HR model during JJA (Figure 2 lower left panel) is partially reduced in VHR especially over Europe and Asia."

p.4, line 119: "the 5 to 7C positive JJA bias over the western coast of South America in HR results haved in HR". That might be the case, but it is really hard to see in the figure. Also, some words seem to be missing in that sentence.

I'm not sure about the source of the aforementioned phrase, because it seems different from what I see in the submitted manuscript, but anyway, the sentence in the new version of the manuscript is:

"Similarly, the 5 to 7°C positive JJA bias over the western coast of South America in HR, results halved in VHR"

We think that the new color scheme is more appropriate (see new figure 2)

p.5, line 129: "the model extreme precipitation is compared to…"

Done.

p.5, lines 155-164. I have to confess I didn't quite understand that explanation.

This part has been partially rephrased as follow:

"*The worsening of the extreme precipitation bias moving from the HR to the VHR model along the tropics, especially in the Southern Hemisphere during JJA, is also associated to a deterioration of the representation of the fraction of precipitation associated to extreme events with respect to the total precipitation: this metric is obtained accumulating the water of all the events more intense than the 99p, and normalizing it by the total amount of precipitation in the considered period (season by season). Figure S17 shows that both models reasonably well capture this metric in both seasons compared to MSWEP, but the VHR model tends to overestimate such amount over the southern Hemisphere, except for the Australian domain. In particular, the strong positive bias of DJF average precipitation over Australia (up to 4 mm/d, Figure S3, lower panels) can't be attributed to the positive (higher than 15 mm/d, Figure 5 lower panels) bias found for extreme events, but must be associated to a right shift of the remaining part of the precipitation distribution, more pronounced for the non-extreme events as partially confirmed by the positive bias in the 90p metric over the same season (Figure S11)*"

p.6, line 188: Replace PRIMAVERA by HighResMIP

Done.