Point-by-point Reply to Referee #2

Comment published online on the 4th of January, 2022

This work evaluates the performance of the Weather Research & Forecasting (WRF) model version 3.8.1 at convection-permitting scale over southern Peru using different configurations combining microphysics, cumulus, longwave radiations, and planetary boundary layer physic schemes. For this purpose, different comparisons for two years, 2008 and 2012, were performed between the WRF outputs and the observations of both weather stations and precipitation gridded products.

The topic addressed in this study is relevant, being the results here found of high value for the climate scientific community. The manuscript is well written and structured, with an appropriate discussion of the results, showing clear and concise conclusions. Therefore, I recommend this study for publication in the Geoscientific Model Development (GMD) journal after minor revisions. My comments are as follows:

Thank you for taking your time to read our manuscript in detail and for your positive and constructive comments.

Major comments:

1. The main analysis was based on 2008 as the precipitation over Madre de Dios was more or less standard (L437-438) in that year. In addition, the year 2012 was selected to test the model performance for wetter conditions. In this regard, as analysis focused on domain 2, how were 2008 and 2012 for the whole d02 domain? Were they also "standard" and "wet"? To clarify this point, it could be good to represent the climatology of domain 2 in Figure 2.

We use PISCO as the main observational data set to perform that analysis. However, PISCO doesn't cover the entire domain 2 as stated in the manuscript. Thus, we decided to include the climatology for Madre de Dios in the paper instead of the one for the entire domain 2.

Figure R2.1 shows the analysis for the part of domain 2 included in PISCO (without the hatched area of Fig. 1). Even if a new domain is considered and a new precipitation mean is obtained for the same period of time (1981-2010), the annual precipitation anomalies (Fig. R2.1a) show that the year 2008 is a standard year in terms of precipitation, and that the year 2012 is also a wet year considering a larger domain. Additionally, Fig. R2.1b shows that the seasonal cycle is rather similar to the one presented in Fig. 2 in the manuscript, which means that well differentiated rainy and dry seasons are observable in both years compared to the 30-year climatology.

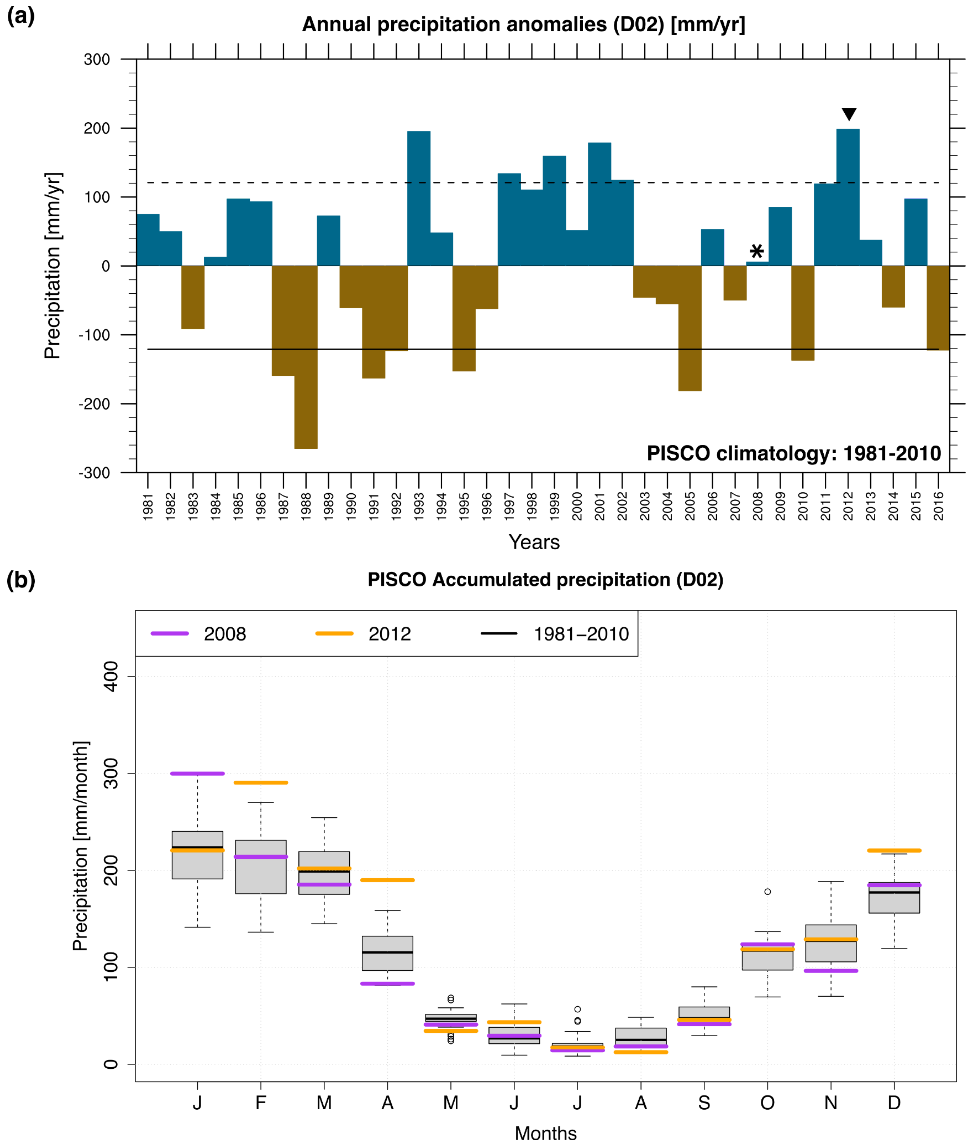

Figure R2.1: (a) Annual precipitation anomalies (in millimetres per year), where the asterisks denotes our main study year 2008 and the triangle the wet year 2012, and (b) monthly accumulated values of precipitation (in millimetres per month) for years 2008 and 2012 (purple and orange horizontal lines, respectively) compared to the climatology (1981–2010, in grey, using a box and whisker plot).

2. For me, one of the most interesting parts of this study is the one related to section 3.4 (seasonal cycle over the northeastern flatland). For these analyses, a comparison could be made with observational values (for example, ERA5) to see which combination of parameterizations are closer to the "reality"? On the other hand, and just as a curiosity, why did the authors select IMERG and CHIRPS (and not the other two, i.e., ERA5 and TRMM) to compare with PISCO?

This is something suggested also by referee 1. Hence, we replace the precipitation panels and add IMERG for the daily cycle (Fig. R2.2) and IMERG and observations for the seasonal cycle (Fig. R2.3). Fig R2.2 shows that Micro13, South America and Kenya are able to capture relatively well the precipitation of the first half of the day, but they all miss the peak in precipitation in the afternoon. Conversely, NoCumulus is able to capture the amount of precipitation during the afternoon correctly. Also the seasonal cycle shows a rather good alignment between IMERG and the observations and the

Micro13 run. We will replace these panels in the corresponding figures in the new version of the manuscript. We will not include ERA5, as it is not an observation-based data but a modeled reanalysis product, which is the driver of WRF and hence, it is not fully independent.

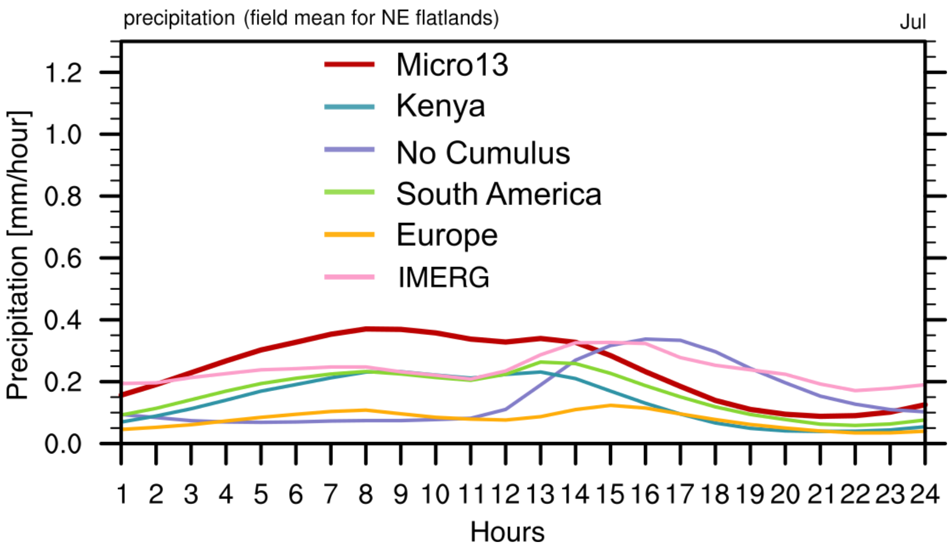

Figure R2.2: Monthly mean daily cycle for July of a field mean over the northeastern flatlands for precipitation (mm) including also IMERG (pink line).

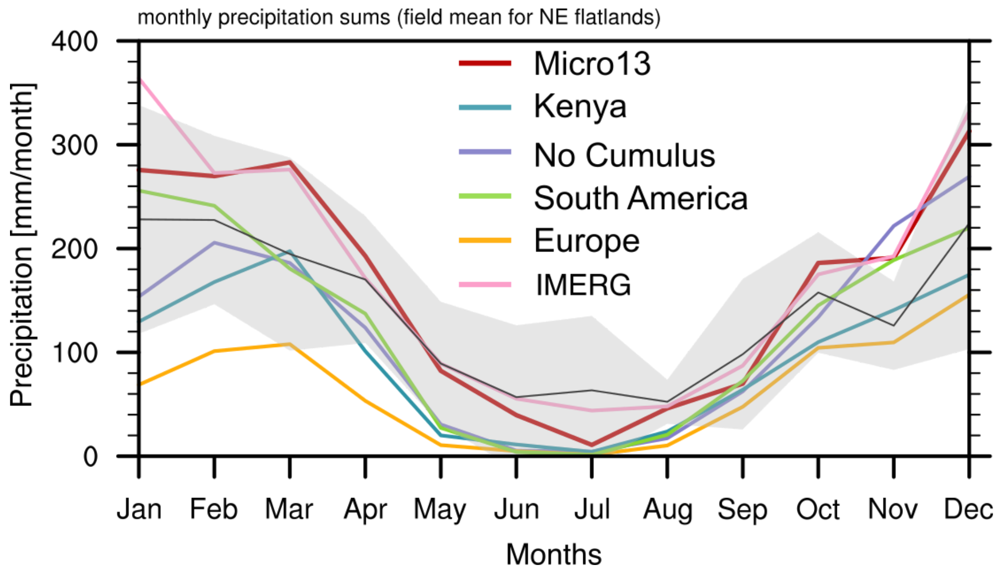

Figure R2.3: Seasonal cycle of a field mean over the northeastern flatlands for (a) monthly precipitation sums (mm/month) including also IMERG (pink line). The black line indicates the average of the station data and the gray shaded area indicates plus and minus one standard deviation.

The reason why we generally focus on IMERG and CHIRPS is that these are the most advanced gridded observational data sets, including a variety of station data as well. TRMM is the predecessor of IMERG and hence not only the temporal but also the spatial resolution is better in IMERG, also the quality of the dataset itself must be assumed to be better in IMERG. ERA5 cannot be considered as a gridded observational dataset, as it is a reanalysis product, which is based on model simulations. Due to the fact that ERA5 is the driving data set of the WRF runs, we prefer to compare the output of WRF to somewhat more independent data sets, i.e., CHIRPS and IMERG.

3. Although the analyses proposed here at the monthly scale provide very valuable information on the characterization of the model and the different parameterization schemes, the use of the daily scale could provide additional information on how the model represents extreme values at a high spatial resolution, and which scheme is better in this regard for the different regions.

As stated in the manuscript, the temporal analysis at finer temporal resolutions such as 15-days, 10-days, pentads or daily was also carried out for our analysis, and the same results are observed for the different intervals. In general, the RMSE increases and the temporal correlations decrease as we increase the temporal resolution. The statement in the manuscript is based on Fig. R2.4 below, that is neither shown in the manuscript nor in the supplementary material. Hence, we will add this figure for daily temporal resolution to the supplementary material.

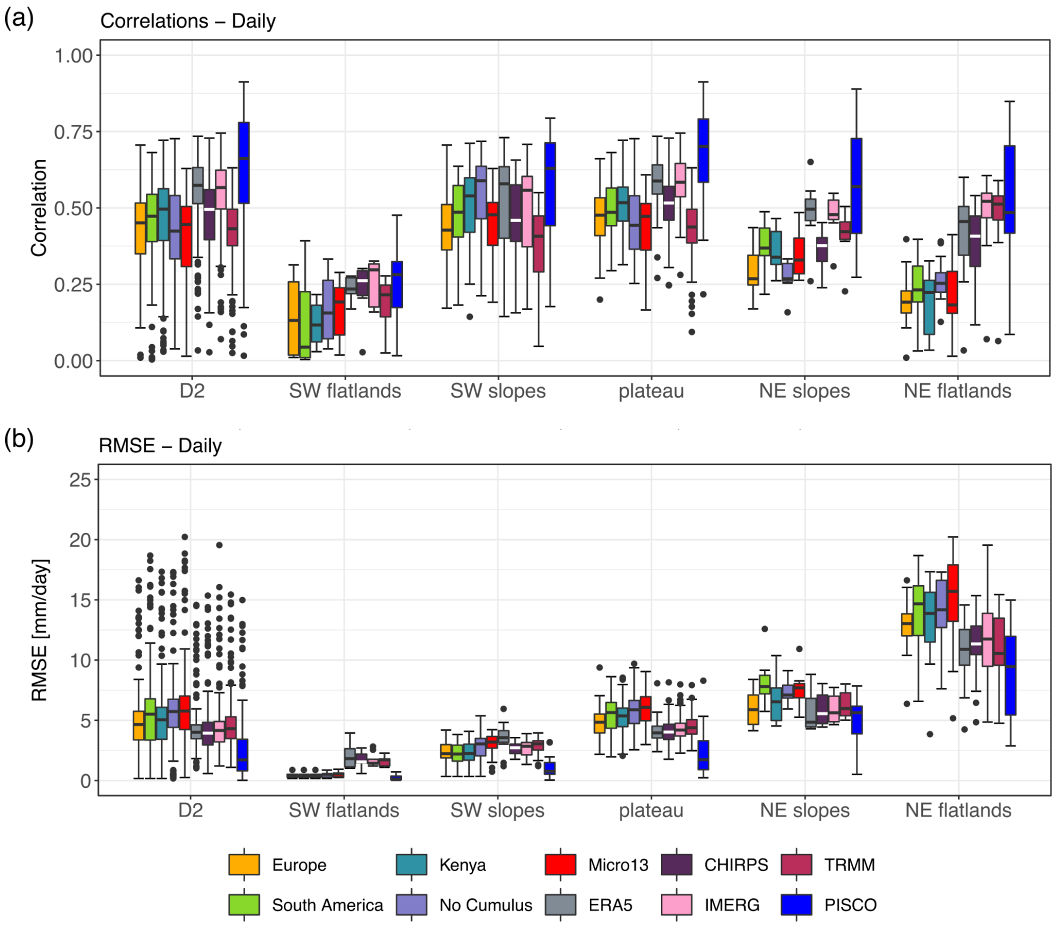

Figure R2.4: (a) The temporal correlation and (b) root-mean-square error (RMSE) between the annual cycle for the year 2008 of measured and simulated daily precipitation sums at the nearest grid point to the station's location shown for the different parameterization options and gridded observational datasets. The whiskers extend to the value that is no more than 1.5 times the inter-quartile range away from the box. The values outside this range are defined as outliers and are plotted with dots.

For the evaluation of the representation of extreme values over each region, we plan to perform a short analysis on daily extreme precipitation values exceeding the 90th percentile.

4. If I understood correctly, the results of domain 3 were not used for the evaluation of the model performance. I agree with the authors to focus the evaluation on domain d02, since, as mentioned by the authors in L428-430, there are not enough observations for a comparison at the finest spatial resolution (i.e., 1 km). However, due to the high computational cost required to run the model for an additional domain at 1-km, it might be good to better specify what was the final purpose of using a third domain, covering the department of Madre de Dios, in such a high spatial resolution.

As stated in the manuscript, once the best configuration of the model has been determined, different climate simulations can be carried out to evaluate the effect of global warming or to investigate the interactions between the land and the atmosphere. In our case, we plan to investigate the effect of global warming over Madre de Dios, a region of Peru that can be considered a biodiversity hotspot and where the ecosystem provides everything that people need there (e.g., raw materials, fresh water, climate regulation, etc). At the same time, some threats are affecting the region such as illegal logging, deforestation or gold mining. These activities sustain to some extent the economy of the region, but at the same time they jeopardize the sustainable development of the region. New high resolution simulations over Madre de Dios will provide some insight about how the region is expected to change under climate conditions, and to infer the effect of those changes on the activities carried out in this biodiversity hotspot. With this in mind, it was also important to include the third domain in the tests, as some of the test runs include a two-way nesting configuration, which means that the result of the innermost domain influences the larger domains and vice versa.

We will include some new lines about this in the revised version manuscript to clarify the final purpose of the highly demanding third domain of the simulations.

Minor comments:

- L121 "...so a spin-up period of two months is enough to balance the fluxes between the atmosphere and soil in WRF": Here, maybe, I would not affirm that a 2-month spin-up period is enough since, as the authors concluded, a longer spin-up period is probably needed for the simulations (L315-318 and L467-469).

As we point out in the paper, we have performed a test with a longer spin-up period and we cannot see a systematic improvement of precipitation sums in the seasonal cycle, so a 2-month spin-up period should still be sufficient.

- In section 2.1., please specify the number of vertical levels used in WRF, the top of the model, and the time-steps applied in simulation.

All the sensitivity simulations include 49 vertical eta levels until the model top at 50 hPa, and the adaptive time step was employed while running the simulations. No nudging was applied to the input data. These details will be added to section 2.1 as suggested also by referee 1.

- Please provide information about the time resolution of the weather station data in the main text.

Thank you for pointing out this missing information. We will add it to the new manuscript:

“The weather station data from Peru are provided by the Servicio Nacional de Meteorología e Hidrología (SENAMHI) del Perú, the data from Bolivia by the SENAMHI Bolivia, and the data from Brazil by the Instituto Nacional de Meteorologia (INMET). These data are available with a daily temporal resolution.”

- L211: Did the authors check other interpolation methods? Please justify why bilinear interpolation was used instead of others (e.g., nearest neighbor).

We have not considered other interpolation methods. The bilinear interpolation is widely used in the literature, and that is the method that we usually follow. We think that the results are insensitive to the interpolation method selected, but we will redo some of the analysis performing also the nearest-neighbour interpolation to assess the sensitivity of the results to the chosen interpolation method.

- L327 "In the NE flatlands, the pattern correlation is rather good compared to the temporal correlation": Here, it could be good to remember that the comparison is between the results from Figure 5 and Figure 3.

We will add the reference to these two figures in the new version of the manuscript, as suggested by the referee.

- L346-347 "However, No Cumulus shows a general excess of precipitation in the whole domain" and L378 "... except for the last which overestimates the amount": It is hard for me to see that the No Cumulus combination generates more precipitation, in general, for the entire domain than others parameterization schemes (e.g., Micro13). In this regard, it could be good to indicate the mean value of the accumulated monthly precipitation for the entire domain, or even the mean for the different five regions.

We will perform this analysis for each of the runs and different regions, and provide the numbers in form of a table, which will either be shown in the manuscript or in the supplementary material.

- L356-359 "Except for the Micro13 parameterization option, most of the simulations...in the NE flatlands compared to the other parameterization options": Were the authors referring to the results obtained from the transect? Please, clarify this point.

Yes, we refer to the transect as stated in line 350. We will clarify that in the new version of the manuscript.

- L363-364 "For the plateau, the simulations agree with PISCO on the rather dry conditions, except for Micro13 and CHIRPS that show wetter patterns": Here, I would suggest removing the information for

CHIRPS as the comparison seems to be between simulations and PISCO. Otherwise, I would change the sentence to express it in another way.

As noted by the referee, the comparison is between simulations and PISCO. We will reformulate this sentence in the new manuscript to:

"For the plateau, the simulations agree with PISCO on the rather dry conditions, except for Micro13 that shows wetter patterns. These wetter conditions are also represented by CHIRPS."

Figures

- Figure 3a-e: I would suggest changing the color of the box for CHIRPS. Here, the median is sometimes hard to see.

We agree that the median is hard to see in the boxes for CHIRPS. However, as we had a hard time to select well distinguishable colors (all the gridded observational data sets, except for PISCO, share the same color family), we will color the median in white and leave the color as it is.

- Figure 3f-l: I would suggest changing the colors of the lines bordering the different regions. It is sometimes difficult to differentiate between the borders of the regions (i.e., plateau, SW slopes, SW flatlands, NE slopes, and NE flatlands) and the borders of Madre de Dios. Also, if not necessary, it could be good to remove the black lines in all the maps. Do these lines represent the borders between countries?

As identified by the referee, the black lines in the maps represent the country borders. We will consider removing those lines from the plots, and include only the border of Madre de Dios together with the lines bordering the five regions. The last will then be colored in grey.

- Figure 4b: I would suggest changing the range on the y-axis for this case to better show the box and whisker plot.

In Fig. 4, all the box and whisker plots share the same range in the y-axis. This facilitated finding the regions with the largest and lowest RMSEs. We have tested this for panel 4d (we think the referee refers to Fig. 4d instead of 4b), but as it does not give more information to the reader, we would like to keep it as it is in the current version of the manuscript and to have the same range for all the y-axes.

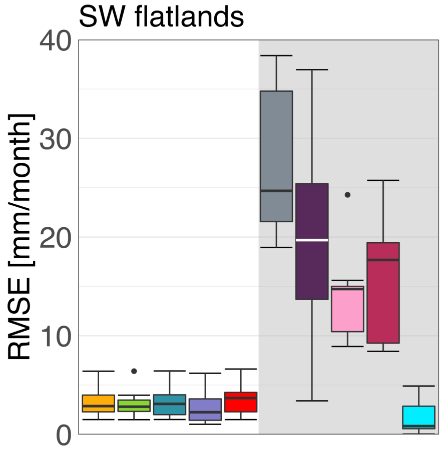

Figure R2.5: As Fig. 4d in the manuscript, but with a smaller y-axis range.

- Figures 6 and 7: I would suggest adding the borders between the five regions (as in Figures 3 and 4) in order to better follow the discussion of the results.

As for Fig. 3 and 4 we will remove the country borders and add instead the lines bordering the five different regions.