Title: Sensitivity of precipitation in the highlinds and lowlands of Peru to physics parameterizations options in WRF3.8.1

Authors: Santos J. González-Rojí, Martina Messmer, Cristoph C. Raible, and Thomas F. Stocker

Reference: gmd-2021-307

Journal: Geoscientific Model Development

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:

Major comments:

1. The main analysis was based on 2008 as the precipitation over Madre the 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.

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?

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.

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.

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).
- 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.
- Please provide information about the time resolution of the weather station data in the main text.
- L211: Did the authors check other interpolation methods? Please justify why bilinear interpolation was used instead of others (e.g., nearest neighbor).
- 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.
- 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.
- 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.
- 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.

<u>Figures</u>

- Figure 3a-e: I would suggest changing the color of the box for CHIRPS. Here, the median is sometimes hard to see.
- 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?
- Figure 4b: I would suggest changing the range on the y-axis for this case to better show the box and whisker plot.
- 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.