

Geosci. Model Dev. Discuss., referee comment RC1
<https://doi.org/10.5194/gmd-2020-443-RC1>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on gmd-2020-443

Anonymous Referee #1

Referee comment on "COSMO-CLM regional climate simulations in the Coordinated Regional Climate Downscaling Experiment (CORDEX) framework: a review" by Silje Lund Sørland et al., Geosci. Model Dev. Discuss., <https://doi.org/10.5194/gmd-2020-443-RC1>, 2021

Review of the manuscript entitled "COSMO-CLM Regional Climate Simulations in the CORDEX framework: a review" by S. L. Sorland et al.

This work summarizes the contribution of the CLM community to the CORDEX initiative over several standard domains and nested to different GCMs. Results for near surface temperature and precipitation are provided, compared against several global observation-based data sets. It provides very useful information to understand simulation differences and the authors perform a sound analysis considering multi-model and observational uncertainties. I think it meets the criteria to be published, but some extra effort should be made to further clarify the modelling and analysis details to improve the reproducibility of the results and justify some decisions taken. I provide next some specific comments:

1) L.26-27 "For the regional climate projections, it is desired to capture all the ensembles of opportunities" Please, rephrase. The use of ensembles of opportunity is not an aim of regional climate projection, but a necessity arising from a lack of a priori design. It is mostly unavoidable to end up producing ensembles of opportunity, but I wouldn't say there is a desire for them.

2) L.22-37 Please, consider reorganizing the paragraph. Currently you mention "major continental domains" (L.28) before the CORDEX domains are mentioned a few lines later (L.32). EURO-CORDEX is also first mentioned a few lines later without a reference.

3) L.45-49 This paragraph mixes past and current information. It is quite misleading for the reader. It starts with current qualitative ensemble sizes for different CORDEX domains. Then, it poses Europe as the domain with the currently largest ensemble and, then, refers back to the early days of CORDEX, when the African domain was prioritised, as if this were a recent decision to overcome the imbalance. Moreover, ensemble size is revisited in this paragraph without adding any new detail with respect to the figures provided e.g. in L.33 or L.35. A dedicated paragraph on domain ensemble sizes is in order but, please, be precise and provide quantitative information.

4) The reader is commonly referred to other publications to obtain basic details. Please, use references as sources of detailed information, but do include the basics in your manuscript. For example: L.57 "until today only two groups [which groups/models?] were able to conduct all required simulations following the CORDEX-CORE protocol" or L.60 "The COSMO-CLM model has been used for a large set of experiments and run over a wide range [what range exactly?] of resolutions"

5) L.92 What do you mean by "qualified judgment"? Please, state clearly the scope of the study.

6) Given that this work tries to present an overview of the CCLM contribution to CORDEX, I think a clearer description of the model genealogy should be provided. Currently, CLM is presented as "Climate Limited-area Modelling", an international network of scientists aiming to develop community models for regional climate research. If I got it right, COSMO-CLM would then be the adaptation of the COSMO NWP model to climate simulation. Steppeler et al. (2003) is provided as reference for the COSMO model, but this reference does not mention COSMO, but the DWD Lokal Modell (LM). Some renaming seems to have taken place since the early days of CLM in PRUDENCE and ENSEMBLES (L.345). Early references to CLM provide an alternative meaning as the "Climate version of Lokal Modell" (CLM being the model instead of the scientific community) and so does a recent work by Steger and Bucchignani (2020, <http://doi.org/10.3390/atmos11111250>). Please, better clarify the lineage of the model. Particularly, the specific versions used in PRUDENCE and ENSEMBLES (CLM 2.4.6? according to Jaeger et al, 2008, <http://doi.org/10.1127/0941-2948/2008/0301>) should be mentioned. The coupling to the CLM (Community Land Model) in Australia adds some extra confusion.

7) L.151 Is the 5-0-6 or 5-0-16 some kind of semantic versioning system (major-minor-patch)? Some comments in this paragraph seem to imply that model configuration could also be coded in the last number. Can this last number be increased because of a particular "recommended" configuration, without any other change in the model? Are the "recommended" versions mentioned in this paragraph the same as the "default" tuning parameters in Table S1?

8) L.153 What does crCLIM stand for?

9) L.153-160 Is COSMO-crCLIM still endorsed by the CLM community? From the description given, it apparently branched off CCLM4 and followed an independent development. Will these developments be incorporated back to CCLM6 (this could be added to the outlook paragraphs at the end of the paper)? Is COSMO-crCLIM adopting new CLM developments?

10) L.160 Is CORDEX-CORE used as a synonym for simulating at 0.22 deg. resolution? In principle, CORDEX-CORE requires simulating for most domains. It is not just a matter of resolution. Also, in table S1, a CORDEX-CORE framework is stated along with spectral nudging in EAS-22. Would the CCLM CORDEX-CORE contribution mix different model versions (COSMO-crCLIM, COSMO-CLM5-0-x) and nudging settings depending on the domain?

11) L.182 Was this reduction of the standard vertical levels (40 to 35) done for computational efficiency? It seems odd to increase the top of the atmosphere and reduce the amount of vertical levels.

12) L.222 "African setup" was previously defined as "Tropical setup" (L.185). Or does this version include also the other developments for Africa (L.187-190)?

13) L.226 There was only a minor bug fix from version 5-0-2 to 5-0-9? As mentioned above, the reasons to increase this last number in the version specification should be clarified, so the user of the data knows whether different versions can be compared. Apparently, not only model version, but many other subtle changes were applied (Table S1). I think the clear identification of these differences is one of the main outcomes of this paper. With so many small changes, the attribution of the different results to a specific change is problematic, though.

14) Indicate in the caption of Table S3 the meaning of the parenthesis in WAS-44. Is it that no evaluation run is available?

15) L.252 "allowing for a fair comparison ..." It is not really fair: despite being global, the amount and quality of the background observations used by these data sets greatly differ across domains (e.g. USA or Europe compared to Africa).

16) L.282 How wide is the relaxation zone?

17) L.293 "The GCMs listed below [...] represent a wide spread of climate changes over Europe, or because they are part of the CORDEX-CORE framework or external projects" Which models were selected for each reason? Please, be precise. This GCM selection excluded some CCLM simulations from the study (L.279).

18) L.309 "bias is masked (shown in white on maps) when being smaller than the observational range" Please, clarify the exact procedure for reproducibility purposes. Bias w.r.t. the mean of the observations was masked when the value of the model lies between $\min(\text{obs})$ and $\max(\text{obs})$? From the sentence above, it seems that you are computing the observational range $R(\text{obs}) = \max(\text{obs}) - \min(\text{obs})$ and masking if $\text{abs}(\text{bias}) < R(\text{obs})$, i.e: if $\text{mean}(\text{obs}) - R(\text{obs}) < \text{model} < \text{mean}(\text{obs}) + R(\text{obs})$

19) L.331 I would avoid the word "transferability" when model configuration is changed across domains. Transferability experiments are the opposite of your approach: 'perform simulations with all modeling parameters and parameterizations held constant over a specific period on several prescribed domains representing different climatic regions' (Takle et al., 2007). But you use different configurations for each domain and advocate (L.605) for the re-tuning of the model for the target domain.

20) L.334 Table S3. This summary (mean bias) compensates positive and negative bias regions. It is not a performance metric anymore. A value of zero can be achieved by wild, opposite biases. See e.g. WAS-22 T2m JJA. It also penalizes biases of the same sign (EUR44-CCLM4 T2m JJA: 1K) more than opposite biases (EUR44-CCLM5 T2m JJA: 0.38K). It is also not very useful to compare across regions (e.g. the wild opposite biases in AFR-44 T2m JJA score 0.27K). I would suggest using the mean absolute bias or a quadratic mean if you'd like an extra penalty for large biases. Are values masked out in Figure 2 included in the mean bias? It would be helpful to add this measure to the panels of this figure to avoid going back and forth between Fig. 2/3 and Table S3. The table is still OK to summarize other seasons. A background color in the table cells according to the value of the score would also be helpful to easily unveil bias patterns.

21) L.367-369 This might change when using a score that does not compensate opposite biases

22) L.428-433 Remind the spectral nudging in this paragraph outlining simulation differences

23) L.467 "During the winter season, there is a warm bias over Northwest India and a cold bias ..." Add "and the Ethiopian highlands" after NW India.

24) L.472 The spatial variability depends on the domain, and so does the ability of the models and observations to reproduce it. It is misleading to mix in a single Taylor diagram the scores for different regions. The information in Figure 4 is duplicated in Figures 5-8 (L.501). I would suggest keeping just Figures 5-8 with an extra effort to make the ERA-Interim values more outstanding. Also, properly zoomed Taylor diagrams should go to the main manuscript. Current zoomed versions (Figs. S17-20) in the supplementary material do not have proper axes and span different areas of the Taylor diagram. Thus, they can hardly be compared to each other.

25) Figure 7 shows a good example of my comment above. It seems that observational uncertainty (the spread of symbols corresponding to observational data sets) for DJF precipitation in Africa is smaller than in Europe. This goes against intuition, considering the poor observational coverage of these data sets over Africa. Likely, this is the result of large-scale precipitation gradients in Africa, a domain covering a tropical precipitation belt along with subtropical desert regions. This spatial structure is easily captured by any observational data set (correlation ~ 0.99) or model simulation (correlation 0.90-0.95). Therefore, different models or observations for a given domain can be compared in a single Taylor diagram. However, the comparison across domains (as in Figure 4) is more tricky.

26) L.530-535 "the choice of the driving data has a bigger influence on the performance" (also L.632) It is not clear what performance you are referring to. RCM performance can only be assessed with "perfect" boundary conditions (i.e. reanalysis, ERA-Interim in this case). The authors discuss GCM-driven simulations on an equal footing with evaluation simulations (L.541-542). GCM-driven simulations incorporate errors from the GCM and RCM. In order to disentangle both error sources, GCM-driven simulations should be compared to evaluation, reanalysis-driven simulations, not to observations. All GCM-driven simulations should "perform" worse than evaluation simulations. Otherwise, an error compensation would be occurring between the GCM and the RCM (this is not desirable). Also, added value can be discussed with GCM-driven simulations. The authors mix the added value with the garbage-in/garbage-out problem (L.561-563). A simple added value measure could easily be incorporated in the Figures by adding the biases and/or Taylor diagram points corresponding to the driving GCM output (e.g. as the crosses used for Raw ERA-Interim). Given the simulations considered in the study, a clearer discussion of RCM performance, GCM/RCM errors, garbage-in/out and added value should be provided.

27) L.544 "the bias patterns and model performance are" Rephrase. Bias is also a model performance measure

28) L.545-547 "We have shown [...] that the model also has to be re-tuned to obtain a model configuration that is optimal for the domain" Well, strictly, you have only shown re-tuned results for each domain. The need for that has been left to previous references. Also, L.635 "The results from this large COSMO-CLM model ensemble indicate that an RCM-modeler can do a lot when it comes to improve the model performance" This is not derived from your results. You can start the sentence at "An RCM-modeler ..."

29) L.633-635 Please, rephrase for better readability. Split into two simpler sentences?

30) L.647 Please, expand or clarify what ICON is.

31) Small typos:
L.354 micorphysics

L.373 remove "As" at the beginning of the sentence
L.387 remove parenthesis around (Panitz et al, 2014)
L.522 more close -> closer
L.595 2018 -> 2020
L.679 publically -> publicly