

Review of "Evaluation of regional climate models ALARO-0 and REMO2015 at 0.22 resolution over the CORDEX Central Asia domain" by Top et al. 2020

The authors improved the quality of the paper compared to previous versions. In particular, the presented analyses are now more appropriate for an evaluation of models performance. Nonetheless, the quality of the manuscript is still not sufficient to be granted publication. In particular, the description of the results is very poor, full of inconsistencies, not exhaustive and detailed and not consistent throughout the different sections. I invite the authors first to describe each of their figures, extensively providing all the details, and only then summarizing the text. Also, please check that you proceed in the description of the results in the same way for each of the figures. Below you can find a detailed list of some of the main issues of the paper. Be aware that errors are not limited to the mentioned examples, but all the manuscript needs a thorough review. The paper is at a good point and a thorough and patient review should allow for its publication.

- l. 57-58: by whom was the simulation performed?
- l. 61: A new paper by Russo et al. 2020 with COSMO-CLM is available for Central Asia, published on GMD.
- l. 102: comparable with what? you probably mean among each others in a coordinated framework?
- Fig. 1: Specify in the caption that all points with orography higher than 3000 m are set with the same height of 3000m.
- l. 120-122: A regionalization was applied and not the subdomains. Reformulate.
- l. 148-150: line repeated twice. Also, make example of which parameters you refer to.
- l. 153, "to produce an equilibrium for the soil temperature and soil moisture": does not sound good. Reformulate into something like: to let the model reach an equilibrium state for ...
- l. 156-157: it should be "were compared". Check for consistency of verbs tense throughout the text
- l. 161 "since all gridded datasets are characterized by uncertainties": I would avoid such statement, since it is quite obvious. On the other hand I would say that you consider different observational data-sets for the calculation of given metrics, for assessing the reliability of the results.
- Section 2.3: make clear that you use CRU as reference and additional data-sets for assessing reliability of observations over different areas.

- l. 166: was used
- l. 183-185: period a bit confusing, you need to reformulate. Basically you state that Hu et al. found that GPCC is better than CRU and MW for the inner part of Central Asia CORDEX domain. Then you say that precipitation is underestimated in mountainous regions, but are you still referring to the inner part of the domain? Also, you state that GPCC underestimates precipitation in general: but for which region? globally? for Central Asia? do you have a reference?
- l. 199: as for GPCC, where ERAInterim generally overestimates precipitation? references?
- l. 216: by computing the difference between maximum and minimum
- l. 220: specify that you calculate spatial MAE.
- l. 225: over the points of the domain
- l. 225-227: The Taylor Diagrams does not represent, but rather includes calculations of the spatial correlation between model and reference data. In your case you can eventually state that the Taylor diagrams are used to estimate spatial agreement between the climatological means of the different data-sets, by considering different metrics.
- l. 239: maybe it would be nice to indicate the different rows corresponding to annual and different values.
- l. 242-244: I would discuss the results of the table together for annual and seasonal values, after introducing the results of Fig. 2. In any case, you have to provide more details on the results of Table 2 for yearly values, not simply mentioning that they are reported.
- l. 252: Very pronounced biases are also present for REMO, in particular in winter, over the north-eastern part of the domain
- l. 253: the reader should know which areas you are referring to: the map of the subdomains should be introduced directly in the main text and not in the supplementary part.
- l. 255: as you did for ALARO, mention that the bias of REMO exceeds 10C in this case.
- l. 256-258: that's it for spring? please quantify the values of the bias in the two cases, consistently with other seasons. Also, what happens over other regions?
- l. 258-259: as above, provide estimate of biases also in summer
- l. 260: On the contrary

- l. 260-261: Why do you stop providing estimates of the biases? This applies to all sections and figures.
- l. 260-261: what about summer biases of ALARO over Mongolia and North-eastern China?
- l. 265-268: I would move this paragraph before the previous one, at the beginning of the description of Fig. 2.
- Fig. 3: How can you explain very high spatial correlation values for the entire domain when the values of correlation are lower in the case of all subdomains?
- l. 275: how do you define best here? in terms of which metrics? as the points being closer to observations? be more specific. For example for REMO RSV, the model is better in summer than autumn in EEU; RSV is better in summer than Autumn over West Siberia with respect to all metrics; for East Siberia in summer, values of RSV are closer to 1 than in autumn; for WCA, summer correlation is slightly better than in autumn; for the TIB region, both summer and spring results have better correlation than in autumn. Only for ALARO results seem to be better in autumn than for other seasons, given the considered metrics. The description of the Taylor diagrams is not very accurate and precise throughout all the text, making the interpretation of the results very difficult for the reader. I suggest the authors to thoroughly review the comments of the Taylor diagrams, trying in a first essay to write down all possible information and only then summarizing the results. I can understand that this is complicated given the large amount of figures, but this is absolutely necessary given the current state of the manuscript.
- 277-278: not true. What do you consider as "WELL" for normalized standard deviation? For me the good performance of ALARO in EEU in Winter and REMO in winter over WSB are quite arguable. Also, I would not be very convinced about the goodness of the winter results for the ESB domain for ALARO as well as winter results in WCA and autumn results in TIB for both models.
- l. 279-280: how would the limited bias explain the higher SD values? biases could be higher but more homogeneous. Reformulate this period.
- l. 282-285: This part is not exhaustive and accurate and needs to be reformulated: first of all a similar bipolar behaviour is found for different seasons for both models, not only for summer. One good example is SON spatial biases for the entire domain for REMO and the case of both REMO and ALARO in winter. Here it would be opportune to use the fact that you have small mean biases but large MAE for supporting the conclusions on the fact that the biases are the results of compensation effects. Also, poorer performance than what? Given your conclusions about the

performance of ALARO in summer, why your reasoning should work for summer and not for winter, where we also have a very well pronounced bipolar behavior of the bias?

- l. 290: specify that you are discussing spatial variability
- l. 289-293: I would reformulate this part: you basically say that REMO is better than ALARO in simulating spatial variability, except autumn and winter for some domains. Then you say that ALARO better captures spatial patterns except winter and summer over some subdomains. This sounds a bit contradictory.
- Fig. 3: specify in the text that the different Taylor diagrams have different scales.
- l. 300: extremely well? it does not seem so for WCA and ESB; Also performance do not seem very good in November for ALARO over WSB.
- Fig. 4: I suggest you to split the figures in 3 figures, respectively for T2, TMin and TMAX, since at the moment the current figure does not allow to appreciate differences in the different cases. Maybe you can introduce figures for TMIN and TMAX in the supplements.
- l. 326: what happens for the orographically more complex regions? what are the biases?
- l. 327: Specify that the bias of REMO over the Eastern part of the domain reaches 15C in winter.
- l. 329-330: what about all the other parts of the domain in spring and summer?
- l.345-354: Quantify values of bias and MAE.
- l. 352-352: this does not seem the case for autumn and annual values of ALARO and for summer for REMO
- fig.6: I think you are not giving the same importance in the discussion of Fig 6 as for Fig. 3.
- Fig. 6: one general question concerning all Taylor diagrams: do you have any clue why you generally get high values of correlation for the entire domain, when in all subdomains you obtain smaller correlations?
- l. 361: not exact: REMO better also for EUU in summer and winter and in WCA for spring.
- l. 361: specify spatial variability
- l. 366: what happens in the Tibetan plateau? what is the magnitude of these biases?

- l. 365-370: not only here, but also for other figures, fix an order for the discussion of the figures (for example from top to bottom) and follow it throughout the text
- l.367-368: cold bias over northern part of the domain is present in all seasons, except DJF.
- l. 364-371: reformulate and extend all the period. Some parts of the domain are never mentioned. What happens for example in the Himalayas and the Arabian Peninsula? exhaustive description of summer is missing.
- l. 378-379: Specify that in the case of ALARO the bias exceeds 7C.
- l. 376-385: you should extend the part describing the tables, discussing for example biases and MAE, giving indications on the fact that in some case biases are the result of compensation.
- l. 380: not very accurate. In summer REMO is better for the Tibetan plateau in terms of MAE, while ALARO is better in winter for the EEU domain
- l. 381: how can you conclude, from the sentence before where you state that the 2 models are better in autumn, that ALARO simulates TMAX poorly in any season? actually there are 3 cases where ALARO is better than REMO: DJF in EEU and MMA and SON in TIB. Anyhow, in many cases the results of the 2 models are very similar and the MAE is very close to 1. Importantly, you again omit to specify what is good and what is bad, in terms of the given metrics.
- l. 403: it would be better to have a map of the bias of DTR. This could help your discussion that now is too generic based only on the maps of TMIN and TMAX.
- l. 406: the model does not restore its balance, since it is also in equilibrium in winter. Reformulate.
- l. 414-415: Where? In the other sections you started commenting the figures and then the tables. Why you change this now?
- l. 415-420: what about annual values in table 5? what about the spread of observations?
- l. 417-418: please quantify all the biases you mention.
- l. 419: are you sure the Tibetan plateau can be classified as a monsoon region?
- l. 418-420: which figure are you commenting now? If you are discussing Fig. 11, actually it does not seem that the bias is smaller but greater in summer, at least for ALARO.

- l. 423-424: It is not totally true that the largest biases are present over extremely dry areas. One example is Northern India in Summer, presenting a remarkable dry bias despite observations are characterized by highest precipitation values.
- l. 422-423: specify that when you talk about low precipitation you are referring to the observations.
- l. 425: actually over Northern China in REMO the bias exceeds 5mm/month.
- l. 433: "is also present for REMO": why also?
- l. 435-436: which model are you referring to?
- l. 438 and l. 447: greater than -2mm/month
- l. 450-451: how can you claim that the spatial patterns are well represented by ALARO if for each subdomain, despite WCA, the considered metrics present relatively poor values (especially in terms of correlation)?
- l. 449-452: why do you not discuss the observations? how do the different data-sets compare to each others? this is something that you should do when commenting all Taylor diagrams.
- You can extend the discussion of Fig. 10, consistently with the other subsections.
- l. 476-479: Please specify the cases when the bias exceeds the spread of observations. At the moment it seems like the spread of observations is smaller than the bias in almost all the domain and seasons. Is that correct? So your figures would confirm that evinced biases are more inherent to the model than to observations, over almost all the domain in all seasons. One more elegant way (and probably more useful for your goals) to determine those points where the spread of observation exceeds the bias, is by plotting a map of the bias (in the supplements) with a point in correspondence of those points where the spread is larger than the biases.
- l. 482-484: be more precise. It is not clear what you want to express. The observational spread is significantly high over complex-orography regions and not over the entire north-eastern part of the domain.
- l. 486-488: also REMO bias exceeds the spread of the observations over large parts of Mongolia and Northern China.
- l. 489-490: reformulate
- l. 497-498: This is also true for EEU in autumn
- l. 531: Also true over mountainous regions of north-eastern part of the domain.

- l. 536-538: but also in summer, at least for ALARO
- l. 542-543: Reference needed.
- l. 542-542: what are these processes?
- l. 543-544: why should it be? reformulate. Better specify what do you mean by shift in the annual cycle?
- l. 546: how can you state that the bias increases when the snow-covered region expands? Have you directly analyzed snow cover in the two models? If what you affirm is true, this should be a feature of both models. However, winter biases are different in the two cases.
- l. 545-546: are you sure about the warm bias appearing in the North during autumn when the snow appears over this region? Again, did you base these statements on some analysis of simulated snow cover? warm bias is very limited to a very small eastern part of the northern domain in ALARO.
- l. 548-549: I would argue against your conclusions, based on the fact that also REMO is characterized by a very warm bias in winter over the North-eastern part of the domain. I do not think that based on your analyses you can raise strong conclusions on the driver of the bias over the northern part of the domain in both models. For sure I would not state that REMO does not encounter the same problem.
- l. 570: wouldn't Eastern Europe be more consistent with your discussion than Western Russia?
- l. 571: Can you better specify what means acceptable? I would emphasize that the MAE in this case is smaller than the MAE between ERAInterim and CRU.
- l. 575-577: How can you state this without a map of the bias in DTR? can you be sure that for all the points of the domain RCMs produce a smaller DTR in all seasons?