Author response to the review of Anonymous Referee #3

Referee 3, thank you for reviewing our manuscript. We are delighted to take all of your comments into account to improve the manuscript. Our answers on all questions, suggestions and remarks can be found on the next pages. Firstly, we summarize the major changes we will make to the revised version of the manuscript based on the comments of the different reviewers:

- *We will include an analysis of the annual cycle over the subdomains as defined by the IPCC6 report (Iturbide et al., 2020) which are situated within the CAS-CORDEX domain. The results, both for the RCMs and the gridded datasets, for the mean temperature and precipitation are given in Fig. A1 and A2.*
- *We will approach the differences between the gridded datasets in a different way. The spread between the gridded datasets (Fig. A3) will be used as an estimate of the uncertainty.*
- *We will improve the discussion section by describing which model features can explain the significant biases that were obtained over certain regions.*
- *We will include some additional recently published scientific papers in our revised manuscript e.g. Harris et al. 2020; Wang et al. 2020; Zhu et al. 2020.*

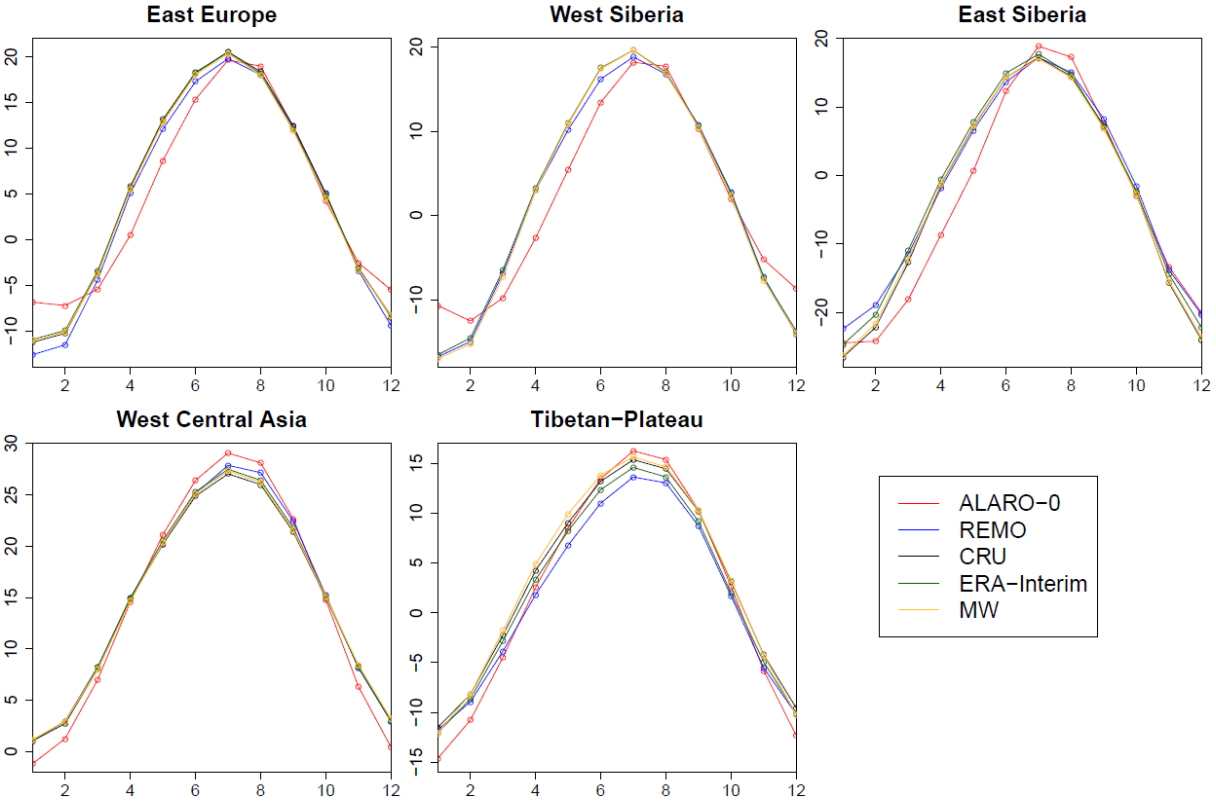

Fig. A1: Annual cycle of the mean temperature (°C) over different subdomains.

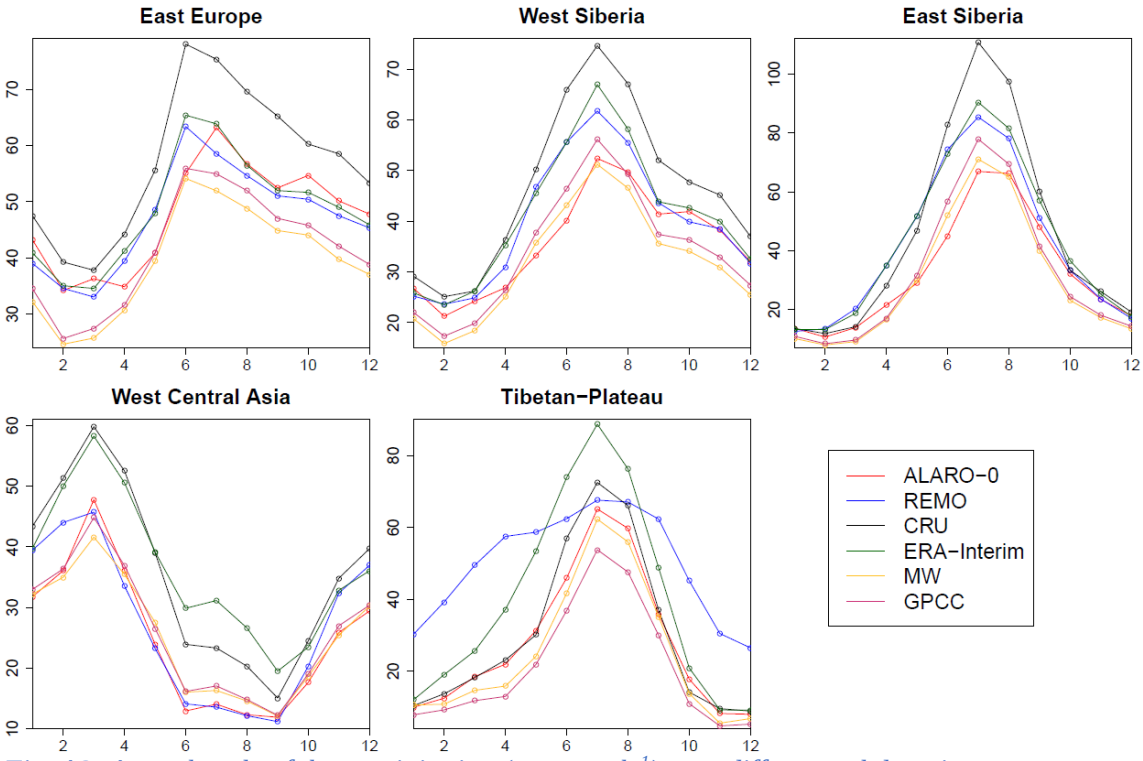

Fig. A2: Annual cycle of the precipitation (mm month^{-1}) over different subdomains.

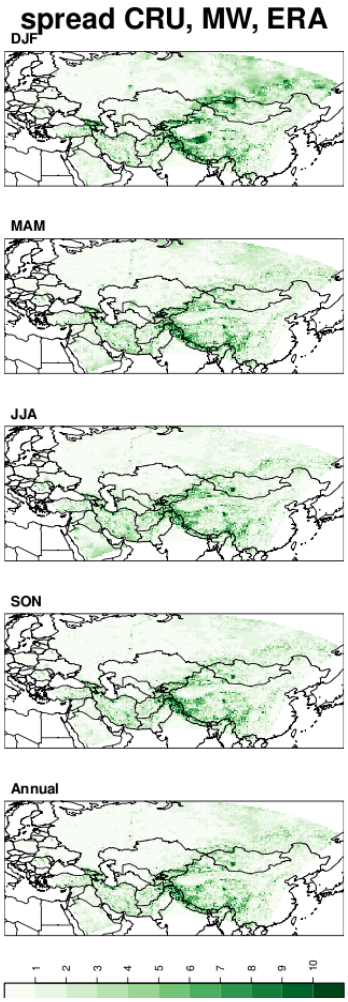

Fig. A3: Spread in mean temperature between the gridded datasets CRU, MW and ERA-Interim.

This paper describes the results of two models (REMO and ALARO-0) simulations over CORDEX Central Asia domain. Authors compared simulated temperature and precipitation climatology and concluded that both the models are capable to reproduce CAS climate. Reading the paper I had an impression it is a kind of technical report but not a scientific manuscript suitable for GMD. I do not see any science by describing how large biases in models are without any reasonable explanation where they come from. Authors took models which were tuned for Europe, implemented them for CAS, obtained huge biases and concluded: "That's it." Therefore I would recommend the manuscript for publication only in case it will be substantially revised.

Major points

1. Analysis (but not referring to other models results) of model biases is required. Where they come from? Is it large scale atmospheric circulation or local processes, e.g. atmosphere – land heat/moisture exchange? In this sense it would be interesting to look in mean sea level pressure (MSLP) biases. For example, the warm temperature DJF bias as well as huge overestimation of DJF precipitation in REMO could be because of underestimation of Siberian High.

We will improve our discussion section taking this comment into account. We are currently investigating possible causes that could explain the obtained biases (e.g. cloud cover, snow cover) and we will include our findings in the revised version of the manuscript.

2. The models show quite a substantial differences in biases. Considering the eastern part of CAS it is clearly seen that in cold seasons REMO simulates 2m temperature much better then ALRO. Furthermore ALRO results with almost 10K bias over quarter of the domain are inacceptable. The opposite is seen for precipitation which is simulated by ALRO better. Based on these results authors can take heat (moisture) fluxes as well as heat (moisture) transports from both the models (assuming that "better" model reproduces better fluxes (transports)) and try to analyze which of them leads to produce mentioned above biases.

We will improve our discussion section by trying to explain the obtained biases.

3. For better understanding I would also recommend to analyze the climatological annual cycle of some quantities, like temperature, precipitation and heat fluxes at least for the eastern part of the domain (from Mongolia to the east), where the biases are really large. For such a big domain with a plenty of climatological zones Taylor diagrams are more a kind of speculation. E.g. in case the climatological temperature varies from +30C in the South to -30C in the North spatial correlation will be high with any kind of model.

We agree with this remark. To gain insight into the model's performance and limitations we will include in the revised version an analysis of the annual cycles based on monthly means for five subdomains. However, we still find it valuable to do the evaluation (and make the Taylor diagrams) over the complete CAS-CORDEX domain since this region is set as a standard domain. Many papers use currently different subdomains over Central Asia and due to the small differences in the definition of these domains they applied the results cannot be equally compared. Standard regions such as the CORDEX and IPCC regions avoid this problem, that is why we will keep the scores over the complete domain in our manuscript.

4.Authors should have a more deeper look into previous studies done with the same models. In particular ones were done with REMO. Since REMO existence (more then 20 years) there are many papers with REMO simulation results over regions partially included in CAS, e.g. whole the northern part: Niederdrenk, 2013 (PhD), Niederdrenk et al., 2016 (Clim. Dyn.), Sein et al., 2014 (Tellus); south-eastern part: Xu et al, 2018 (Clim. Dyn.).

We took these papers into account and will refer to some of them in our updated text.

5.Authors claim that some of the biases come from the ERA-Interim forcing. That is quite an ambitious conclusion, in particular for Siberian continental climate. This conclusion has to be proven with some additional simulations. It is not a big deal to take a lateral boundary conditions from some of the global climate model, to simulate ca. 10 years and to look if the large scale biases are similar or not. I think with available computer recourses it should be just 3-4 working days.

Indeed we cannot claim that the biases are due to the ERA-Interim forcing without investigating this feature. We removed the text parts where we are claiming this.

Minor points

L. 23: I do not think that with large scale 8-10K 2m temperature biases and more then 100% precipitation biases over quarter of the model area both models reproduce climate "reasonably well".

For the precipitation we get sometimes more than 100% due to the very low amounts as discussed in the text. For example, if there is 1 mm of precipitation and the models estimate 2 mm monthly precipitation, the relative precipitation bias is huge. Therefore, we added the absolute differences as well in the supplementary material. Additionally, there is the spread between the gridded datasets. From the newly created annual cycles it can be seen that the RCMs are mostly within the spread of the gridded datasets.

L.24-25: It has to be done in this work, but not postponed to the unclear future

This would make the paper too long.

L.35: Even being a not an expert in CORDEX and even for CORDEX domains mentioned by authors, I know much more works based on multi-model regional simulations. E.g. Africa: Paxian et al. (JGR-Atmos, 2016); Mediterranean: Damaraki et al. (Clim.Dyn, 2019), Gaertner et al. (Clim. Dyn, 2017), Soto-Navarra et al. (2020, Clim.Dyn).

Since there are quite some publications about multi-model regional simulations we made a selection, discussing all of them is not in the aim of this paper that handles about CAS-CORDEX where there are no multi-model regional simulations available. Including all of the other domains would make the paper too long but we will add some of these references.

L.61: "Absence of reliable observational data sets". Over China and Russia? Maybe 20 years ago "yes" (describing CRU data authors site work from 1999), but at the present time it sounds at least strange.

We agree, Harris et al. (2014) is indeed better to refer to for the current information about CRU and we will add as well the Harris et al. (2020) reference which was published after we submitted our manuscript. We included the 1999 reference since this one describes the strategy and methodology of CRU.

2. Methods. See above (L.35) Central America: Cabos et al. (2019, Clim. Dyn.), Southeast Asia: Zhu et al. (2020, TAC), Arctic: Akperov et al. (2019, Global and Planetary Change; 2018, JGR)

We will at least refer to Zhu et al. (2020) in our updated paper.

L.94: I would remove word "sea". In a middle school I have learned that Black, Caspian Red and Baltic seas are seas, but it is hard to say that they are barely covered with CAS domain.

We agree, the Black Sea, Caspian Red Sea and Baltic Sea are seas in the CAS-CORDEX domain. We removed "sea" and replaced it with "open ocean" since we wanted to stress that the domain mainly exists out of landmass.

L.96: Before claiming it, authors should "google" a word "HighResMIP". In the framework of this project there are many global climate model simulating climate on 25 km resolution, i.e. the same resolution as authors use for their regional simulations.

We added the reference of Haarsma et al. (2016) with information about HighResMIP to the text.

L.106 and in other places: I would suggest to use not "coupled zone", but "sponge zone". Forcing a regional model with reanalysis has nothing to do with coupling.

To overcome confusion we will use "relaxation zone".

L.129: But what about dynamical core itself? Please explain at least in the way it is done for ALRO above, i.e. special discretization, advection (e.g. in ALRO it is based on semi-Lagrangian algorithm and what about REMO?)

See table S1 in the supplementary materials where these specifications are mentioned. We opted not to mention all of them in the text because of the readability and to keep the text as concise as possible.

L.137-138: What about upper boundary? Which height does it have? 10hPa? 50hPa?

The upper boundary of ERA-Interim configures for 60 levels in the vertical, with the top level at 0.1 hPa (https://www.ecmwf.int/en/elibrary/8174-era-interim-archive-version-20).

L.202: As far as I know almost all the atmospheric models (including REMO and ALADIN) provide direct output of Tmax and Tmin which are obtained every model time step. Why not to use them directly?

This is correct, Tmax and Tmin were used directly from the model output of REMO and ALARO-0. We reformulated our text to avoid confusion.

3. Results: As I mentioned in "major points", not only seasonal means but also climatological annual cycle for the quantities averaged over different areas has to be included.

We agree, we have added the annual cycles.

L.229: Exceeded. How much does it exceeded? On the plot I can only see that it is larger then 10K.

It depends on the subregion or the location. In winter the maximum bias obtained for REMO and ALARO-0 at one particular point is respectively 16.8 °C and 19.2 °C when compared to CRU.

L.234: What has Scandinavia to do with Mongolia? They have completely different climate. In the same way REMO group can write: Paxian et al. (2016) showed a strong precipitation bias over Guinea in Africa. Maybe that is also a reason of REMO prcip. bias over East Siberia?

We agree and we will add additional information.

L238: Actually the strongest cold bias over Europe in REMO is at Spring. It is not visible in most of the papers, because mainly they show DJF and JJA only.

Yes, that is true. We included all seasons to report our results as honestly as possible.

L.360 (Fig.8) Relative difference in mm/month? I think it should be in (%)

Indeed, we corrected this.

To all the figures with biases: For the biases I would avoid linear color bar and extend it for larger values. E.g. for the temperature something like: 0,1,2,3,5,7.5,10,12.5,15 and for precip. (%) 0,10,20,30,50,75,100,125,150,200

We will reduce the classes of the color scales in order to improve the readability of the figures and we will use a non-linear color bar as suggested.

L.405: What the Czech Republic has to do with Central Asia? Do they have similar climate? I have here the same claim as at L.234. Authors should provide arguments which has something to do with CAS and not speculations like: we have warm bias in Mongolia, because in French Polynesia is to rainy.

We agree.

L.414: I would not say that up to 10K large scale temperature bias is something which is VERY well

Biases over 10 °C are mainly found over the regions where the reference datasets are less reliable (see spread reference datasets in the newly created maps). We agree that we should formulate this differently e.g. the results are within the range of uncertainty of the used gridded datasets. Additionally, for some parameters significant biases are present over parts of the domain for some seasons and cannot be explained by the uncertainty in the gridded data. For example, the ALARO-0 RCM has a large positive temperature bias in winter over the northern part of the domain. The REMO model has difficulties in reproducing the observed precipitation patterns over the orography of Central-Asia. We agree that the biases observed in this study should be kept in mind when presenting future projections. We find it therefore important to publish an exhaustive evaluation study. In this evaluation study we saw that the main patterns are modelled correctly and therefore we concluded that we can move on towards climate projections. We will add to our conclusion that these large biases should be kept in mind when looking to the future projections. Additionally, to deal with the biases in impact studies, several bias adjustment methods have been tested within the AFTER project and the most suitable method will be applied before simulations for impact studies are done with these climate data. It is not in the scope of this evaluation study to explain the details about bias adjustments and impact modelling but to avoid misunderstandings we will add that bias adjustment is one of the possibilities when mentioning that the RCMs can be used for future projections.

L.423: "..assigned to this forcing". As it was mentioned above (Major points), before speculating about it, please do some simulations with different forcing.

We agree, we cannot claim that the biases are due to the ERA-Interim forcing without investigating this feature. We removed the text parts where we are claiming this.

L.433: "Ozturk et al. . . ., but they did not explain it." And? If Ozturk did not explain it, it is over? Why don't you try to explain it in your manuscript.

We will improve our discussion section by trying to explain the obtained biases.

L.428, 448, etc. New et al. (1999). You discuss present climate and present observational data set citing a work from 1999? There is a quite a big difference between the number of observations before 1999 and now.

Indeed there is a difference between the number of observations in the beginning of our evaluation period (1980) and the end (2017). New et al. (1999) is rather describing general features about gridded datasets, that is why we mentioned this reference. We agree that it is better to refer to more recent and concrete papers for the CRU dataset. Recently a new paper for the CRU data was published (Harris et al., 2020) and we updated our text, taking this paper into account.

Fig. 11: I think should be MW, but not WM. As well as (%), but not mm/month

Indeed, we corrected this.

Conclusion: In the scientific sense conclusion is very poor simply describing how large model biases are only. The only one "explanation" of their origin is "models are good, but observations are bed", based on results obtained more then 20 years ago, in 1999. I would suggest to authors to bring more "scientific analysis" into the manuscript considering comments written above. Maybe it will bring the paper from "technical report" to "scientific manuscript".