Reply to

2nd Reviewer

Russo, E., Kirchner, I., Pfahl, S., Schaap, M., and Cubasch, U.: Sensitivity studies with the Regional Climate Model COSMO-CLM 5.0 over the CORDEX Central Asia Domain, Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2019-22.

Dear reviewer,

Thank you very much for your effort in reviewing our paper.

Below we go point by point through your technical corrections, presented in italic, detailing how we dealt with your concerns reported in *Bold*.

Thank you.

General Comments

- *I*n order to provide reliable future climate projections, the model should be able to capture the present climate feature realistically. For seeking the optimal setups for regional climate model COSMO-CLM over the CORDEX Central Asia domain, the authors have conducted series of sensitivity simulations for historical periods. With different observation/reanalysis dataset as references, they evaluated the general model performance in capturing the mean climate and variability of temperature, precipitation and daily temperature range and figured out the relative optimal model setups for CORDEX Central Asia domain.

  Though the study is rather regional specific, it is believed to be interesting for the regional climate modelling community. The manuscript is in general well organized. The methods used are reliable and language is good. However, the manuscript suffers from some major problems. The authors will need to address them before the manuscript can be considered for publication in Geoscientific Model Development.

Specific Comments

- *I*t is suggested to reduce to a relative brief introduction about vulnerability of CORDEX Central Asia to the effects of climate changes, say from Page 2 Line 19 to P3 L11. Furthermore, there is a general lack of reviewing studies about model performance evaluation, which are related to the experimental setups, assessment methods and discussion, c.f., Li et al. (2018) and Huang et al. (2015) and so on: Li, D., Yin, B., Feng, J., Dosio, A., Geyer, B., Qi, J., ... Xu, Z. (2018). Present Climate Evaluation and Added Value Analysis of Dynamically Downscaled Simulations of CORDEXEast Asia. Journal of Applied Meteorology and Climatology, 57(10), 2317-2341. Huang, B., Polanski, S., Cubasch, U. (2015). Assessment of precipitation climatology in an ensemble of CORDEX-East Asia regional climate simulations. Climate Research, 64(2), 141-158.

We agree with the referee that the part of the introduction on the vulnerability of Central Asia domain to the effects of climate change should sensibly be reduced, being only secondary to the purposes of the manuscript and making the text hard to read. We will try to summarize this part in a more concise way in the new version of the manuscript. Additionally, the part of the introduction on the state of the art of model performance evaluation and model calibration will be extended, considering, among others, the referee suggested references relevant for the area.

- *T*he authors conducted a series of experiments considering different configurations, which are supposed to be significant for skills of modelling. However, some specific setups, which have been proved to be important in regional climate modelling, have not been considered in the study, such as the technique of spectral nudging (von Storch et al. 2000) and topography. RCM simulation with spectral nudging can add value in reproducing snow water equivalents, coastal winds and some meso-scale phenomena (von Storch et al. 2016), as well as annual mean temperature and precipitation (Tang et al. 2017). The reviewer suggest the authors add one experiment with spectral nudging. In addition, about two additional 25-year long simulations covering 1991-2015, why do not use a period backward, say 1981-2005, so that there are longer spinup time, and same comparison period as other experiments?

We agree with the referee that spectral nudging is a powerful tool in order to add value to several aspects of RCM simulations, as indicated in Von Storch et al. 2016 and Tang et al. 2017. Nevertheless, we think that the use of spectral nudging does not fit well the scopes of our work. In fact, in the paper we want to evaluate general model performance and how it is possible to improve these by using a set of specific physical configurations. Also, we want to determine main model limitations and uncertainties and the possible reasons for them. For doing this, we think that it is of fundamental importance to let the model "free" to develop. We do not think that constraining the model by spectral nudging would be useful in this sense. On top of that, this step is not considered in the main CORDEX-CORE directives and also in the model configuration procedure of the COSMO-CLM community. Concerning the point on why we performed the 25-year long simulations over the period 1991-2015, the response is that we aimed to use for this, the restart file of the reference simulation (01 January 2006). This allowed to save computational time, because otherwise the reference simulation should have been repeated for 25 years, starting at 1981 instead of 1991.

- *T*here are some problems in Figure plottings: a). Figure 1, please plot in lon and lat dimensions rather than in rlon, rlat dimensions; b). Figure 2, it is better to add names of subregions on map rather than using a colorbar; c). Figure 3, the colorbar scheme is rather poor. It is hard to distinguish them on the map. Less and distinguishable colors are suggested to use, with more equal divisions within -10 to 10 and less divisions from (plus minus) 10 to (plus minus) 20.

We modified Fig.1 of the former version of the manuscript as suggested by the referee. We propose now to replace the former figure with Fig.1 of the current document. We also modified Fig. 2 of the former version of the manuscript accordingly to the referee comment. The new figure is shown as Fig. 2 of the current manuscript. We agree with the referee that this new figure

might sensibly help improving the results discussion for different sub-regions. Finally, we also modified Fig.3 and 5 of the former version of the manuscript following the referee suggestion, reducing the number of colorbar breaks. We want to highlight the fact that the new figures, reported here as Fig.3 and Fig. 4, allow now to better discriminate high biases and in particular to notice that, for the case of winter temperatures, these are mainly inherent to the UDEL dataset and that, in general for temperature, biases exceeding 10 °C are only present for a few number of points for areas characterized by particularly complex topography.

- *S*ome descriptions does not reflect the figures or tables. Such as P10 L26, I would not say experiment q in Fig.7 (upper panel) fits to the description; P10 L34, experiment o does not share the use of the setup of j. A thorough revision is needed to catch all these inconsistencies.

Concerning the comment for page 10, line 26, we realize that we were not probably very clear in the description of the figure of seasonal calculated SS. Here we wanted to say that for temperature experiment q has positive values for all seasons, except winter. We will modify the text accordingly. Instead we agree with the referee comment relative to page 10, line 34, and we will try to revise the entire text for similar inconsistencies.

- *I* would not agree the conclusion that The results for the mean climate appear to be independent of the observational dataset used for evaluation and of the boundary data employed to force the simulations. In fact, according to Fig. 3 and Fig. 5, it is clear that skill of simulated mean climate depends on the referred observational dataset. Furthermore, Li et al. JAMC (2018) clearly shows that both observational dataset and boundary forcing have impacts on the skill assessment of simulated mean climate.

Following the referee comment we acknowledge the fact that the highlighted sentence was probably not very clearly expressed. What we wanted to say in this case was that when considering different observational datasets and different boundaries, in our case study, we see that experiment q leads to an univocal positive

improvement of the simulated results, for all variables, in all the cases. Considering this point, we will re-formulate the highlighted part of the text in a clearer way.

- *O*nly whole-region or subregion averaged values for SS or variance ration (Fig. 6 Fig. 8) are not enough. Spatial distribution patterns of these scores are significant for a thorough model quality assessment. I would not suggest to plot every spatial distribution of these scores for each reference dataset, but representative figures are necessary, if not in the manuscript but in the supplementary part.

We agree with the referee. A similar concern was also raised by the other referee. We agree on the fact that analyses on subregions for the climatological means could be very important for the purposes of the paper. For this we now propose to substitute the figure on the SS of the different simulations calculated for single seasons, with Fig.4 of the following document, placing the former in the supplement part of the paper. The new figure shows the SS of the MAE calculated over all the points of each subdomain characterized by similar climatic conditions. This might help to distinguish different biases in different cases, and to determine how and to which degree it is possible to reduce them, through modification in specific physical settings of the model. On the other hand, we think that the analyses of variance are already in their definitive form. In fact, for this we proceeded in the same way as in Gleckler et al. 2008 and also considering Wilks 2006. Basically, the assumption that we follow is that the model, due to its chaotic nature, is not supposed to catch climate variability point by point. For this reason it is better to use regional means when we want to evaluate model variability. We will try to modify the text in order to make this point clearer.

Minor Comments

- *P*6 L8-15: Its better to summarize the data information in a table.

We agree and we will add a table with information for the different observational datasets. Still, we think it is important to also mention these datasets in the text, together with appropriated references.

- *P*7 L6: Tab. 3 not Tab.4, the same for P9 L6 and P12 L14

We agree and will modify the text accordingly.

- *P*7 L7-8 Combine two paragraphs into one

We agree. We will join the two paragraphs accordingly to the referee comment.

- *P*7 L13: Mean Absolute Error to Mean Absolute Error (MAE)

We will modify the text accordingly to the referee comment.

- *P*11 L24-25: It may be only appropriate when you run CCLM driven by similar high quality reanalysis datasets.

Again, here we wanted to show that the model presents the same improvements for experiment q when using NCEP2 and mainly employed ERAInterim reanalysis. We decided to use NCEP2 instead of ERAInterim, cause their resolution is closer to the one of the GCMs ($\sim$ 200 Km) that we aim to use for CORDEX simulations. Despite this more than reasonable choice, we also considered ERAInterim driven simulations in our paper, to show that in the two cases we get almost the same results. Please, find more details concerning this point in the answer to the first referee. We will modify the final version of the manuscript in order to make this point clearer.

- $P12$ L3-19: Please indicate which subpanel of Figure 8 you are descripting in the text.

We agree that the current description of the analysis of the variance is a bit confusing and will try to improve it in the final version of the manuscript, better specifying in each case the considered figure sub-panel, as suggested by the referee.

- $P12$ L26-27: range of absolute differences instead of absolute differences?

We agree. We will modify the text accordingly.

Below we propose some additional bibliography that we will provide in the revised version of the manuscript, if not already present, accordingly to the referee comments.

References

[Li et al. (2018)] Li, D., Yin, B., Feng, J., Dosio, A., Geyer, B., Qi, J., ... Xu, Z., 2018. *Present Climate Evaluation and Added Value Analysis of Dynamically Downscaled Simulations of CORDEXEast Asia*, Journal of Applied Meteorology and Climatology, 57(10), 2317-2341.

[Huang et al. (2015)] , Huang, B., Polanski, S., Cubasch, U., 2015. *Assessment of precipitation climatology in an ensemble of CORDEX-East Asia regional climate simulations.*, Climate Research, 64(2), 141-158.

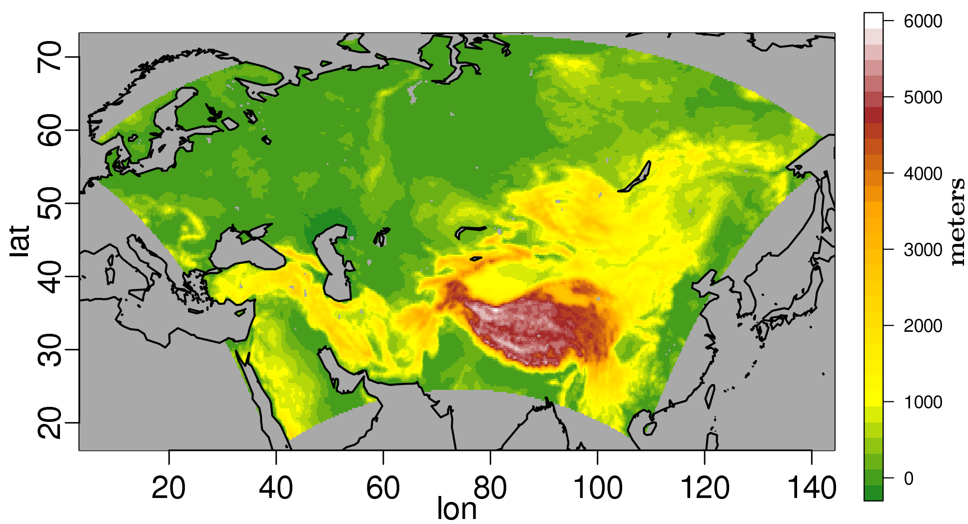

Figure 1: Orography map of the Central Asia simulation domain on a regular grid with a spatial resolution of 0.25 km. Masked in gray are the ocean and the external area of the domain.

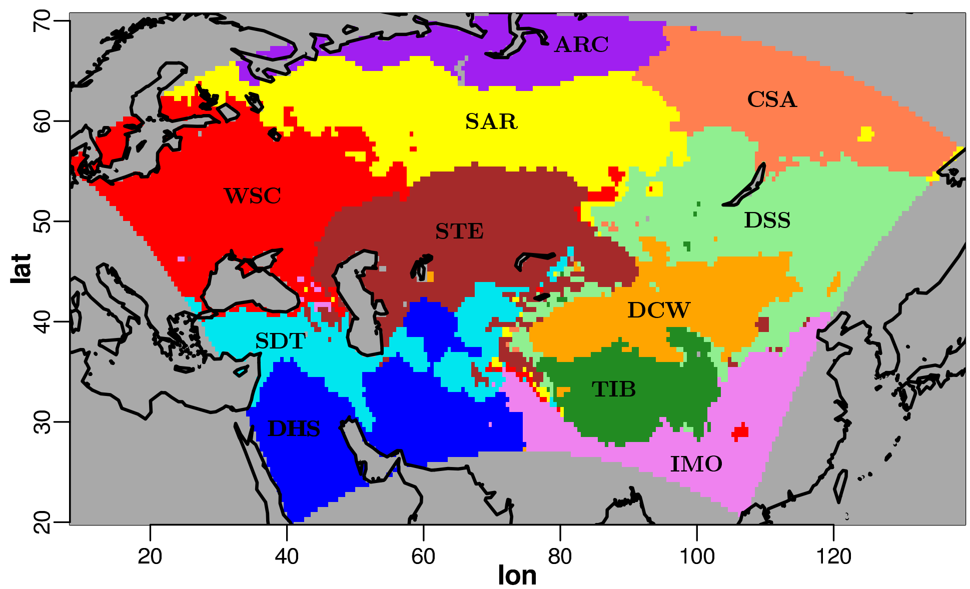

Figure 2: Map of the 11 sub-domains obtained through k-means clustering of the q-normalized monthly climatologies of the three considered variables over the period 1996-2005.

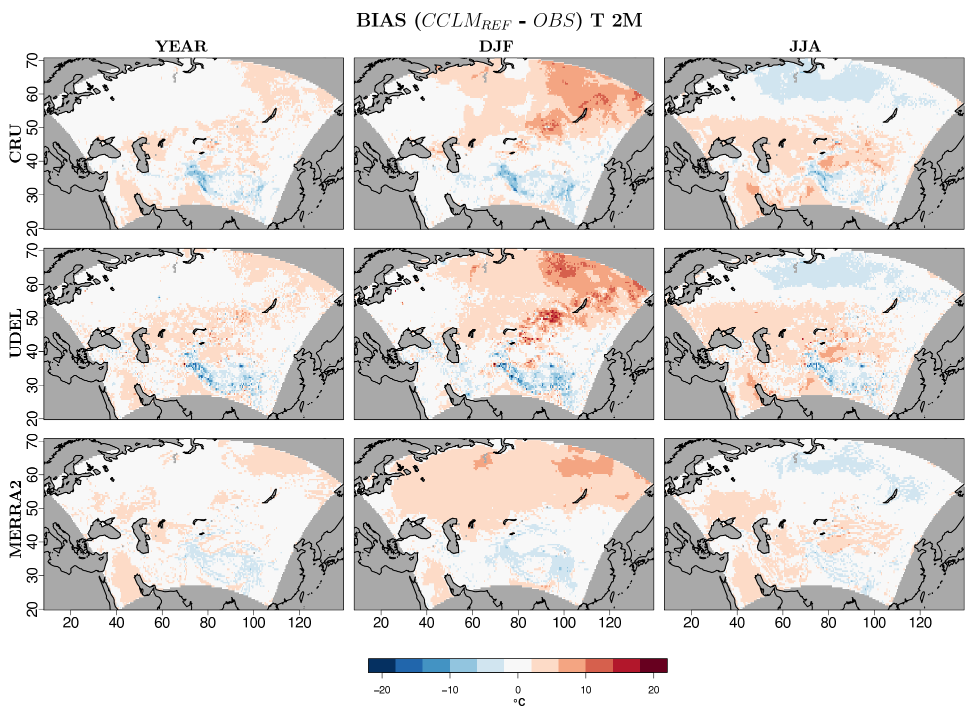

Figure 3: Mean bias of annual mean (*left*), winter mean (*middle*) and summer mean (*right*) near surface temperature (T2M, °C), of the reference COSMO-CLM simulation (a) with respect to 3 considered observational datasets (from top to bottom: CRU, UDEL and MERRA2), for the period 1995-2005.

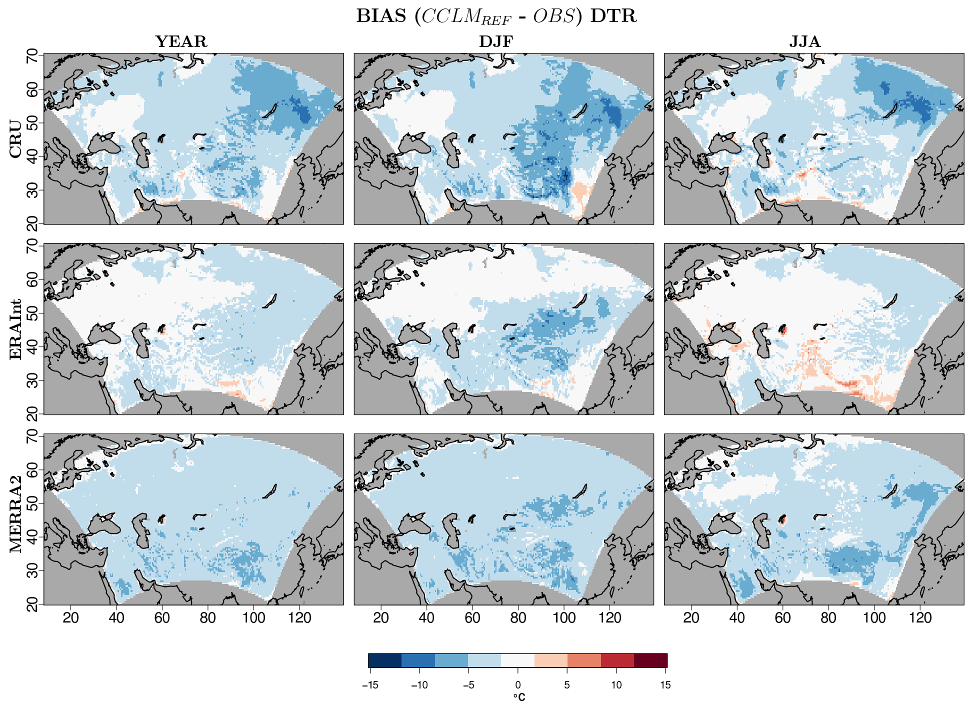

Figure 4: Mean bias of annual mean (*left*), winter mean (*middle*) and summer mean (*right*) Diurnal Temperature Range (DTR, °C), of the reference COSMO-CLM simulation (a) with respect to the 3 considered observational datasets (from top to bottom: CRU, MERRA2 and ERAInterim), for the period 1995-2005.

With kind regards on behalf of the all authors,

Emmanuele Russo