Review of “Sensitivity studies with the Regional Climate Model COSMO-CLM 5.0 over the CORDEX Central Asia Domain” by Russo et al.

General Comments:

In 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:

(1) It 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 CORDEX—East 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.

(2) The 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?

(3) There 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 (±) 10 to (±) 20.

(4) Some 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.

(5) 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.

(6) A temperature bias exceeding 15ºC is too large, especially when the simulations are driven by high-quality reanalysis dataset rather than by GCMs generally used by CORDEX community. A fully discussion on this issue is necessary.

(7) Only 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.

Minor Comments:

(1) P6 L8-15: It's better to summarize the data information in a table.
(2) P7 L6: Tab. 3 not Tab.4, the same for P9 L6 and P12 L14
(3) P7 L7-8 Combine two paragraphs into one
(4) P7 L13: 'Mean Absolute Error' to 'Mean Absolute Error (MAE)'
(5) P11 L24-25: It may be only appropriate when you run CCLM driven by similar high quality reanalysis datasets.
(6) P12 L3-19: Please indicate which subpanel of Figure 8 you are descripting in the text.
(7) P12 L26-27: range of absolute differences instead of absolute differences?