Reply to

Anonymous Referee

Russo, E., Soerland, S.L., Kirchner, I., Schaap, M., Raible, C.C. and Cubasch, U.:

Exploring the Parameters Space of the Regional Climate Model COSMO-CLM 5.0 for the CORDEX Central Asia Domain, Geosci. Model Dev. Discuss.,

https://doi.org/10.5194/gmd-2020-196.

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

- *Page 4 line 9: Panitz et al. (2014) describes an evaluation simulation forced by ERA-Interim, not a future projection study; cite Dosio et al. (2015) and or Dosio and Panitz (2016) instead.*

  - Dosio, A. and H.-J. Panitz (2016): Climate change projections for CORDEX-Africa with COSMO-CLM regional climate model and differences with the driving global climate models. - *Dosio et al. (2015): Dynamical downscaling of CMIP5 global circulation models over CORDEX-Africa with COSMO-CLM: evaluation over the present climate and analysis of the added value. Clim Dyn 44, 26372661 (2015).*

  - Dosio, A. and H.-J. Panitz (2016): Climate change projections for CORDEX-Africa with COSMO-CLM regional climate model and differences with the driving global climate models.

  We will correct the previous reference taking into account the new ones suggested by the referee.

- *Page 5, line 21: Zhang et al (2004) cited, but reference is missing*

  We will introduce the missing reference in the reference list.

- *Pages 5 and 6, section 2.3: which spatial resolution did you use for the comparisons between model data and observations? I assume 0.5°. Please, mention it and say why you chose the specific spatial resolution and how you did the remapping.*

We conducted our analyses considering a spatial resolution of 0.5°for the comparison between model data and observations. Prior to the calculation of the considered metrics the model data were remapped onto the grid of the CRU dataset. For temperature we used a linear remapping, while for precipitation and cloud cover a conservative interpolation approach was employed. Following the comment of the reviewer we realized that such information is missing in the manuscript and we will provide it in section 2.3 of the new manuscript, where the analysis methods and metrics are discussed.

- *Page 8, line 9: any idea why PS is lower for PRE than for T2M and CLCT? Just indicating this fact is not very satisfying.*

  The value of PS is particularly low for precipitation because of higher biases with respect to the values of the uncertainties in this case. On the other hand, biases for T2M and CLCT are more in the range of the corresponding uncertainties. This is evident from Fig. 9 and is discussed in section 3.3, where we analyze the role of different uncertainties on the computation of the considered metrics. In the new version of the manuscript we will explain at the end of line 9 in page 8 that more analysis on this point will be introduced in the following sections. Also, we will try to extend the discussion in section 3.3. concerning the role of different uncertainties on the considered metrics.

- *Page 8, line 10: must be Tab. 2, not Tab. 3*

  Will be corrected accordingly.

- *Page 8, section 3.1: altogether, 9 parameters have been selected, which are recommended to conduct the objective calibration procedure following Bellprat et al(2012). These 9 parameters are the 7 most sensitive parameters that show largest variation in PS, and in addition, two further, namely uc1 and soilhyd, which have been selected from the interpretation of PS dependency on each variable. Why not also rat lam and tur len being characterized, like uc1 and soilhyd, as parameters with particularly small variations in PS calculated for single variables ... (see Page 8,line 5). To my opinion, especially the tur len values $\geq$ 500 m are too high, and the smaller value shows slight improvements for CLCT and PRE. Baldauf et al. (2011) also demonstrated the sen-*

sitivity of results of NWP to the values of tur len with improvements using smaller values, even smaller than the lower limit of 100 m used here. I recommend considering at least also tur len in a subsequent objective calibration study.

Despite acknowledging the importance of additional parameters as suggested by the reviewer, our a priori decision was to select a maximum of 2 parameters for each of the model physical scheme. This choice was deliberately made for keeping the "costs" of a possible calibration procedure limited. Following the reviewer comment, we will make clearer in the new version of the manuscript the reasons for our decision. At the same time, we will try to highlight the fact that other parameters such as tur_len play an important role. Concerning the parameter rat sea, instead, we do not agree with the referee on the fact that it plays such an important role for the region.

- *Page 9, line 13: must be* $c_lnd, notc_land; deletethea$

  Will be corrected.

- *Page 9, line 15: for example here the authors assign the model bias, here with respect to T2M, to structural problems in the model formulation. But whats about the quality/reliability of observations in such sub-regions like those representing Siberia? I would expect at least a short paragraph in the manuscript discussing this aspect. I cite: As models are frequently tuned on the basis of observational data, misguided model development can easily result from not taking into account observational uncertainties. For example, tuning models to observations in regions where the mean model bias strongly depends on the selected observational data set (e.g. in Norway) can deteriorate the model performance. These are the first two sentences of the Conclusions from a publication of Prein and Gobiet (2017) that perfectly describes the impacts of uncertainties in observations on regional climate analysis.*

  The considered metrics take already into account different sources of uncertainties in their definition, among which the one related to the use of different observational data-sets and the one related to the interannual variability of the reference observational data-set. For the case of T2M over Siberia (here SAR, CSA, DSS of Fig. 2 of the manuscript) Russo

et al. 2019 showed that the model presents a remarkable warm bias in winter over Siberia. Despite the fact that over some point this large bias is associated with high uncertainty in observational data-sets, the comparison against different observations confirmed its sign and pattern: there is surely some problem for the model over this region. The issue would eventually be how to accurately assess the magnitude of this bias over the entire points of the region. The fact that all the perturbed parameters do not show significant improvements in simulated temperatures over the area (SS derived from PI), and a consequent reduction of the bias, is indicative of the fact that the model is likely missing or not accurately reproducing processes important for the region. It has to be acknowledge though that one possible reason for the high values of PI and its small variations when perturbing parameter values is that these changes are dumped by higher values of the uncertainties with respect to the bias over the region. Therefore for a better interpretation of our results we decided to include a section in the paper, discussing the role of uncertainties on the calculated metrics. From Fig. 9 of the former version of the manuscript it is possible to see that the role of the different uncertainties compared to the bias is relatively small for the 3 subdomains of Western Siberia, for almost all the months. This supports the idea that for the region, the model bias does not change significantly when changing parameter values. There is an underlying reason for the evinced biases that could possibly be reconducted to model formulation. As already said before, we acknowledge the fact that the discussion on the effect of the different uncertainty sources as presented in the former version of the manuscript could sensibly be improved and extended. We will try to do so in the new version of the manuscript. Additionally, considering the 2nd reviewer comment, we realized that in our previous conclusions we did not give enough weight to the fact that the evinced model sensitivity might change when changing the model setup, for example changing the size of the domain or the model resolution. We will try also to consider this point in the new version of the manuscript, when referring to errors in the model formulation.

- *Page 11, section 3.4: I assume that the PS analysis has been performed for T2M, PRE, and CLCT together. This is not mentioned in the text.*

  Following the referee comment we realized that we did not specify how the PS analysis is conducted in section 3.4. We will specify it in the new version of the manuscript.

- *Page 11, line 8: please explain why you only used the parameters e surf, rlam heat, rat sea, and entr sc for the transferability study. I would have expected that you would have considered also qi0, uc1, fac rootdp2. With e surf and qi0 you then would have considered the two parameters that you identified as those with the largest effect on model performance, as you state in your Conclusions. Furthermore, rlam heat, rat sea, entr sc, qi0, uc1, and fac rootdp2 are those parameters that had been considered by Bellprat et al (2012) in their objective calibration study. This would, perhaps, give the opportunity for some comparative discussions on the results achieved for corresponding parameters.*

  We actually selected a priori 2 parameters for which the model seems to be particularly sensitive over the Central Asia domain and 2 for which it is not. In our opinion, seeing that some parameters that are not sensitive in one case are sensitive in the other, is already sufficient for supporting the hypothesis that calibration analyses should be performed when changing the domain of study. In this sense, according to the evinced results, we do not think it is necessary to perform further tests.

Comments Figures

- *Figure 3: please, indicate in the caption that the red marker represent the PS values for the default values of the tested parameters (see also Table 2)*

  will be corrected in the new version of the manuscript.

- *Figures 5, 6, and 7:It would be of advantage for the reader to group the experiments carried out in this study according to the physical processes the respective parameters are assigned to (as you did in in Table 2). It would be much easier for the reader to follow the discussions in the text also in the figures Example: on page 8, line 32, the authors describe, for T2M, changes in model performance over the Tibetian*

Plateau due to value variations of the surface parameters e surf and pat len. In Fig. 5 the reader finds the results for pat len in the upper part, those for e surf nearly at the end. This makes it hard to synchronize a discussion/interpretation in the text with the corresponding visualization in the figure.

We will try to sort the different experiments of Fig. 5-7 as suggested by the reviewer in the new version of the manuscript. On the other hand, concerning the referee comment on the discussion, we previously discussed the figures focusing on different regions and we would like to use the same approach in the new version of the manuscript. However, we will acknowledge the referee comment, reviewing the text for making the discussion more synchronized where necessary.