Reply to

1st 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

- *T*he level of the manuscript is quite poor; especially the Introduction is too "educational" and seems more similar to a technical tutorial for PhD students than to a scientific paper for experts. Many concepts are explained in details, but they are well known by the scientific community working with RCMs, and could be replaced by proper references. On the other side, a detailed synthesis of the state of the art is completely missing, especially for what concerns previous sensitivity studies performed with CCLM or other regional models. There are specific works by Bellprat et al. (2011, 2012) or Bucchignani et al. (2016) that have been referenced, but were not mentioned. In this view, I suggest the following papers to be analyzed and referenced:

  Avgoustoglou E., Voudouri A., Khain P., Grazzini F., Bettems J.M.: Design and Evaluation of Sensitivity Tests of COSMO model over the Mediterranean area. Perspectives on Atmospheric Sciences, 1:49-55.

  Voudouri A., Khain P., Carmona I., Avgoustoglou E., Kaufmann P., Grazzini F., Bettems J.M. (2018): Optimization of high resolution COSMO model performance over Switzerland and Northern Italy, Atmospheric Research, 213:70-85.

We agree with the referee and think that the introduction could

be improved and made more easily readable. Also following the suggestion of the second referee, sensibly shortening the part of the introduction on the description of vulnerability of Central Asia to the effects of climate change should help in this sense. At the same time, we will try to expand the part of the introduction on the state of the art of model performance evaluation and model calibration. For this we will consider the suggested references, together with additional ones. We will additionally try to express concepts well known in the community in a shorter form, when possible.

- *M*y major concern is however related with robustness and significance of the results. A key aspect is that none of the simulations represents correctly the observed fields used as reference. Consequently, there is no value in analyzing the relative performance of the simulations, simply because all do in a terrible bad way. A temperature bias larger than 15 °C or a precipitation bias larger than 200 % is not acceptable.

  A key question is about the reason of this shortcoming. Is it due to errors in the reference CCLM model configuration? a key parameter is certainly the time step adopted (dt), whose value is not specified in the manuscript, but it needs particular care. Alternatively, is it due to errors in the boundary conditions? The authors decided to use NCEP2, but I personally would prefer ERA-Interim, which are characterized by higher resolution, so reducing the resolution jump (other critical aspect).

The concerns raised by the referee are very important. Despite we think that we partly tried to take care of them already in the manuscript, at least partially in the experiments design, we recognize that we were not very careful in the treatment of some points, not clearly specifying them in the text. Trying to answer to the referee comment, in a first place we want to highlight the fact that the paper, given the large amount of tests conducted, is very important in order to understand whether the evinced biases are characteristic of the model itself and if they can be reduced by properly configuring the model and to which degree. For this, in our experiments we tried to be careful in as many points as possible as to isolate different sources of uncertainties. For this purpose, we performed additional simulations driven by

ERA-Interim to test the effect of different boundaries. The results of these simulations are not particularly different than the ones driven by NCEP2, presenting a very similar bias with the considered observations, for all variables (Fig. 1). We realized that we did not carefully discussed this point in the paper and we will better highlight it in the new version of the manuscript. With this additional simulations we aimed to show that the given biases are not due to the effect of considered boundary data. Or at least that, in our case, the use of mainly employed higher resolution reanalyses such as ERAInterim as boundaries, does not affect significantly model performances. Additionally, the improvement of the results when using NCEP2 or ERAInterim, with the finally proposed configuration (experiment q), is also very similar in the two cases (Table 4 of the original manuscript). We want to claim the fact that we decided, very consciously, to use NCEP2 instead of normally employed ERAInterim, for a clear reason: to try to reproduce the resolution jump (mentioned by the referee) that we will have using GCMs for the CORDEX-CORE simulations. In fact, the plan within the COSMO-CLM community is to use, for the CORDEX-CORE simulations, 3 GCMs: MPI-ESM, HadGEM and NorESM. Their resolution is respectively around 210km, 210x140km and 270x210 km. In this sense, also considering the results of the mentioned simulations, we think that our choice of using NCEP2 instead of ERAInterim is more than justifiable. We realized that we were not accurate enough on this point in the text. In the final version of the manuscript we will include information on the three models and (as supplementary material) the picture of the bias of the ERA-Interim driven simulation, together with related information. Other conducted simulations that we performed and that could help in isolate the reasons for the model bias are the two additional simulations that we performed with diffetrent soil layers and the CCLM multi-layer snow model (Fig. 2). These results are important because they show that it is not possible to improve model performance in terms of winter temperature over Siberia following previously suggested hypotheses (the snow model produces even warmer biases over the area in winter), and the given bias is very likely due to the model formulation itself. This is very important because it highlights the necessity to put more efforts in the improvement of the simulation of snow processes and permafrost in COSMO-CLM. We

will provide the new figures of the bias of winter temperatures for these simulations in the supplementary part of the final version of the paper. The only model uncertainty factor that we did not consider in our former version of the manuscript is the different time step. For tackling this issue, following the referee comment, we now conducted a new simulation with a different timestep of 120s instead of 150s. The biases against observations calculated for this new simulation are reported in Fig. 3 (in this case only CRU is used for T2M and DTR, while GPCC is considered for PRE). As you can see, using a different time-step the results do not change significantly. This confirms even more that the given biases are characteristic of the model and do not depend upon the referee suggested sources of uncertainty. We will add this figure as supplementary material to the new version of the paper, and will better discuss it in the final manuscript.

Beside these considerations, the most important thing that must be considered in order to properly address the referee comment concerning the validity and significance of the results is the comparison and proper discussion of the different observational datasets. It is certain that the evinced model biases are quite remarkable in some cases, in particular during winter for temperature. For precipitation, the biases exceed 100% over large parts of the domain, but not 200% as suggested by the referee. This is normal for many RCM studies, especially for the Tibetan Plateau, as already indicated in the former version of the manuscript. Biases of 100% in precipitation are evident in most of the CORDEX simulations, for different domains and models (Kotlarski et al. 2014, Russo & Cubasch 2016, Park2013, Martynov2013). In any case, these biases, supported by the already mentioned analyses, certainly confirm the importance of trying to investigate different model configurations in order to improve model performance over the area and to determine to which degree this is possible. We think that the paper definitely gives a significant contribute in this sense. On top of that, we agree with both the referees that the different observational datasets need more attention in the manuscript and their main drawbacks and differences need to be properly discussed. For this purpose we propose to include in the final version of the manuscript Fig. 4 of the current response, showing the spread of the different observational datasets, for each variable. This will

contribute to support the validity of the presented results. In fact, the spread of the different observations is larger in correspondence of those points characterized by particularly complex topography, for which model biases seem to be more remarkable, exceeding 15 °C in the case of temperature. This suggests that the particularly high biases evident in the model are hard to be quantified for these points. Additionally, if we now consider the new figures of temperature bias (Fig.5, together with the corresponding Fig. 6 for the DTR bias), that we drew considering the suggestion of the second referee to use a colorbar with fewer breaks, we can see that the very high temperature biases exceeding in some cases 15 °C are mainly relative to the UDEL dataset and, in general, particularly large biases are limited to a few points characterized by particularly high topography, where the gridded datasets are less reliable. If we consider the CRU dataset, that is one of the most employed dataset for the area, for evaluating the results of RCMs, we see that the values of the bias rarely exceed (are below) 10 (-10) °C, for really few points. Beside these points, still some remarkable biases are present but we think that these are well within the ranges of model simulations produced in CORDEX. For example, for the East Asian CORDEX domain the studies of Wang et al. 2013 and Bucchignani et al. 2014 showed that the CCLM over the Eastern Part of the Tibetian Plateau has a cold bias in winter, much lower than -6 C. The other area for which largest biases are present for temperature in our simulation is Western Siberia. For this region, one of the only references is the one of Ozturk et al. 2015. Their results are part of already published CORDEX results. From these, using REGCM4 with a resolution of ˜ 50 Km, they obtained a model climatological bias for winter temperature exceeding 8 °C over the area, when using CRU as a benchmark for the evaluation of their results. This is very similar to our case. Following this discussion, we propose to introduce in the new version of the paper the figure with the spread of observational datasets, together with an appropriate discussion on the possible reasons for their differences, and the new plots for the map of the bias of the reference simulation. One important thing that also needs to be highlighted in the text is certainly the fact that, beside high biases between model and observations are evident for some points, mainly where the observations are less reliable, the pattern of these biases is in general very similar among

different observational datasets and well within the range of other CORDEX results. We think that this information is certainly required in order to support the significance and robustness of the results and we will include it in the new version of the manuscript.

Always in the context of significance of the results, for the rest of the analyses of climatological values, we proceeded separately considering the differences among different observational datasets. In fact, we calculated for each dataset independently the MAE and the relative skill score with respect to the reference simulation, with the aim of choosing the best performing configuration that shows the same positive effect among all the different observational datasets. We will try to express this more carefully in the paper, where we think it was not carefully stated before. Additionally, we propose to introduce Fig. 7 as a supplement to the final version of the paper. This is obtained in a similar way as in Bellprat et al. 2012, but only considering the climatological mean of each month over the considered simulation period. Basically, we selected a reference dataset for each variable (CRU for T 2M, GPCC for PRE and CRU for DTR) and then we calculated the MAE in each case, weighting the absolute error over each point by the sum of two uncertainty sources, one given by the standard deviation of the observational datasets and the other by a standard deviation representative of considered model errors, calculated among the climatological means of the reference simulation, the one with a different time step and the reference simulation driven by ERAInterim. This is the formula we used for the new plot:

$$MAE_w = \frac{1}{TJI} \sum_t^T \sum_j^J \sum_i^I \frac{|m_{t,j,i} - o_{t,j,i}|}{1 + \sigma_{o_{t,j,i}} + \sigma_{iv_{t,j,i}}} \tag{1}$$

where t represents the considered month and j and i are the spatial indices of the points of the domain. The two terms m and o are, respectively, the model and the observational monthly means calculated for each month of the domain. σ_o represents the mentioned observational uncertainty and σ_o the one of the model. The 1 in the denominator has been added in order to avoid infinitive numbers when the uncertainties are considerably close to 0. In this

way, points of the domain with higher uncertainties will receive a lower weight in the computation of the MAE. As it is possible to see from Fig. 7 of this document, the results of the SS using the weighted MAE are approximately the same as in the original figure 6 of the manuscript. Some differences are present in some isolated cases, mainly for precipitation, due to high uncertainties over some area. Nevertheless, the new analyses overall confirm the results of the plot of the total skill score of the paper: simulation q has the best performance for the domain, for all variables, and most of all coherently among different observational datasets and also considering different model uncertainties. All the analyses, considering different uncertainty sources in a different way, give the same results.

In conclusion, in the light of the presented discussion, better considering the mentioned points and different sources of uncertainty, we can affirm that our results are very important and most of all significant for the improvement of model performance over the area. Generally, the model, excepted some isolated points for which the bias against observations is not really quantifiable due to the large spread of the observations, does not perform particularly worse than other models normally considered in CORDEX. For sure there is some limited number of points for which climatological biases are very high, but these seem to be related to observation uncertainty. An appropriate discussion on the sources of these biases, most likely related to the complex topography of some region, will be included in the final version of the paper.

- *T*he paper does not investigate the origins of these strong biases. Section 3 contains only a rough (boring, in some places) description of the figures, but no significant insights are provided. Some considerations are provided in the Conclusions, but this is not the right place, since conclusions should be focused on the benefit of sensitivity runs with respect to the reference one.

We agree with the referee that the results part is in some cases boring and we will revise the text to make it more easily readable. For this purpose, we propose to place the figure on the investigation of seasonal biases to the supplement part, replacing it with

the referee proposed analysis of regional behavior of the different simulations. We think that in this way some not particularly relevant part of the text could be summarized in a simpler way, while focusing on single regions could help supporting the given conclusions on best model configuration. Concerning the current structure of the paper, we actually want to keep the results part as a description of the figures (the results) as we conceived it in a first place. On the other hand, we want to keep the part on the discussion of the results in the conclusion, but, following the referee comment, changing the title into a more appropriate "Discussion and Conclusion" section.

- *I*n Sec. 2.3, you have properly defined some subdomains, but then they are used only for the analysis of variance. Instead, the results in terms of MAE (presented in Figs.6 and 7) are averaged over the whole domain, which is too big and includes very different climate areas. I reccomend that further investigations in terms of MAE be performed considering the single subdomains.

We agree with the referee that further investigation in terms of MAE performed considering single sub-domains is required. This in fact could help to have a more proper idea of how model results change for different areas characterized by different climate conditions, contributing to the determination of an optimal model configuration and to better discriminate reasons of possible shortcomings. Therefore, we propose to introduce Fig. 8 of the current document, representing the SS of the MAE calculated over single sub-domains, in the final version of the paper, together with an appropriate discussion of the results. The proposed figure is obtained in the same way as in the case of the entire domain: the calculation of the MAE is conducted separately for the different observational datasets. For visualization reasons, we propose to plot the results of the analysis per sub-region for a single observational dataset for each variable (CRU for T 2M and DTR and GPCC for PRE), with a point for each region for which the given configuration produces the same model response among the different observational datasets. Fig. 9 shows the SS of the MAE calculated for different sub-regions for all the considered observational datasetes. We propose to include this figure in the

supplement part of the manuscript. At the same time we propose to also include as supplement to the final version of the paper, Fig. 10 of the current document, presenting the same analyses per sub-region but using the weighted MAE. As we can see from Fig. 8, Fig. 9 and Fig. 10, beside some exceptions, the results have a similar behavior for all different cases, with the same conclusions that can be drawn. These plots help because they allow to investigate model behavior for single regions, as already said. In particular, they allow to see that a complete improvement of model results over all the sub-regions is not completely achievable. As discussed in the introduction of our paper, one has always to be aware of the fact that calibrating the model could lead to better results, but this might also be the result of compensating errors. Reinforcing these thoughts, we think that with the proposed optimal configuration q the model improves in large part of the cases, for all variables. These results highlight again the advantages of using configuration q for the region. The newly proposed analyses also allow to see that in some cases model improvements almost reach 35% with respect to the reference simulation. This, once again, adds significance to the proposed results. In the final version of the manuscript we will change the results part as already proposed, substituting the plot of seasonal results with the one of the analyses for sub regions. The text will be changed accordingly to the new introduced results, trying to make it more easily readable.

- *F*inally, the differences among observational datasets are not discussed and the possible reasons for these discrepancies are not investigated,

We agree with the referee. As already stated above, we aim to introduce in the new version of the manuscript a proper discussion on the different observational datasets, their differences and the possible reasons for them.

Specific Comments

- *P*ag 2, Lines 11-18: "Among the...at once". This paragraph contains too many geographical and economical details and in my opinion does not fit well in the Introduction.

We understand this issue, raised by both the referees. We realize that this part should be significantly shortened, being only secondary to the objectives of the paper. We think that this would also help making the introduction more easily readable.

- Pag 2, Lines 25-33 and Pag 2 Lines 1-9: "The countries...due to climate warming". These paragraphs are rather an analysis of the implications of climate changes on this area, and in my opinion do not fit well the aim of the work. They should be significantly shortened.

We agree. We will shorten this part as proposed in the previous answer.

- Pag2, lines 9-11: "All the reported...strategies". This sentence is a repetition of concepts already expressed.

We agree. This part is repetitive and we will delete it from the final version of the manuscript.

- Pag 2, Line 14: "Assessing...Evaluation". This definition is repetitive and can be removed

We will remove this line accordingly to the referee comment.

- Pag 2, Line 17: "Model Evaluation...development". This sentence is prosaic and can be removed.

We agree with the referee and will correct the text accordingly.

- Pag 3, Line 27: "A series...simulation". This sentence is prosaic and can be removed.

Same as above.

- *P*ag 4, Lines 11-14: "In the light...are presented". From this sentence, I do not see a relationship between your sensitivity and the CORDEX-CORE activities. Please explain better this relation and, at the same time, explain what CORDEX-CORE is.

CORDEX-CORE stands for CORDEX - Coordinated Output for Regional Evaluations (CORE). This is the next phase of the CORDEX initiative, designed in the light of the upcoming IPCC report, with the objective of coordinating a set of high resolution climate projections for different regional domains, including Central Asia. In this perspective, our work represents the first step for the production of climate projections for the Central Asia domain using COSMO-CLM, evaluating general model performances, isolating the effects of different uncertainty sources on model results and determining an optimal model configuration for a region region for which almost no reference exists. Following the referee comment we realized that we probably did not specify very well this information in the former version of the manuscript. Consequently, we propose to extend the relative part of the text accordingly.

- "This study...domain". This concept has already been expressed in the Introduction. Please put it only once.

We will remove this repetitive part of the text, following the referee comment.

- *F*igure 1: It is preferable to show the domain using the geographical coordinates, since the reader is generally not interested in the rotated coordinates (being rotated coordinates used only internally for COSMO-CLM calculations)

We agree. We propose a new version of the map of the domain topography, presented in Fig. 11 of the current document, in geographical coordinates. The figure caption will be modified accordingly.

- *P*ag 5, lines 11-13: "The model configuration...en)". These details are not necessary, epsecially because readers are generally not authorized to download the model configuration from the website of CLM Community. Please add more details about model configuration in Table 2.

We agree with the referee that the description of the model configuration introduced in Table 2 of the previous manuscript version needs to be extended. Also, the link to the CLM-Community webpage could be removed since, as suggested by the referee, not all users could access the given configuration. Proposing to modify the text accordingly, we still want to mention the fact that we use as a benchmark for our reference simulation, the configuration of COSMO-CLM for another CORDEX domain, but covering a large part of Central Asia. This follows the main guidelines of the CLM-community for the model configuration.

- *T*able 2 (caption): The general description of model setup of the reference simulation is very poor. It contains details that have already been explained in the text (e.g. spatial resolution, domain extent). Btw, the domain extent must be expressed in terms of max/min longitude/latitude and not in terms of number of grid points. Some important details of the model setup are missing. For example, in Table 1 you write that in b configuration Tegen aerosol is used, but what is the aerosol used in a? I guess the default Tanre, but you have to specify it. Similarly, for albedo: what is the default one? I guess albedo as function of soil type.

We agree with the referee and will modify table 2 accordingly to his comment. Information about the time step of the reference simulation (150s), the Aerosol, for which we used TEGEN as default, and the albedo, as a function of the soil type, will be included, together with information on the domain extent expressed as min and max lon and lat.

- *P*ag 5, lines 19-20: "since their...simulations". This is not a good reason to employ NCEP reanalysis as driving data. Generally, data at higher resolution are preferred. Btw, what is the resolution of GCM normally employed in CORDEX simulations

As already stated above, we were completely aware of the decision taken using NCEP2 reanalyses instead of ERAInterim, with the main goal of simulating a jump in resolution more similar to the one using the GCMs mentioned above. For this, we consider our choice more than valid. NCEP2 are still considered a valuable reanalysis dataset, that has been used in a large variety of studies. Beside that, we also conducted similar analyses with ERAInterim to have an estimate of the effect of using higher resolution drivers on the results and how they change in the different cases. We demonstrated that the effect of the two different datasets on the simulation of climatological monthly means for the considered period is almost the same. We will highlight this point better in the new version of the manuscript.

- *S*ec 2.2: it is not clear if the original resolution of datasets CRU, UDEL, MERRA GPCC is 0.5°or if they have been interpolated on a common grid with common 0.5°resolution.

The resolution of the mentioned datasets is all 0.5°. No interpolation was needed. We will try to make it clearer in the new version of the manuscript.

- *P*ag 6, lines 19-20: "The climate...interpolation". This technical detail (usage of CDO) is not necessary and can be removed.

CDO is an important tool for the postprocessing of climate data, freely available. It is a personal decision, but also following the work of other papers, given its importance, we think that it deserves to be referenced in the text.

- *P*ag 7, line 25: "It is not specified if variances (observed and simulated) are evaluated starting from monthly values.

We acknowledge the fact that we have not been very clear in this sense in the previous version of the paper. We now propose to modify the final version of the paper better specifying that the variance is evaluated starting from monthly values.

- Pag 8, line 12: A bias of 15°or larger is not acceptable.

Again, this seems to be a problem related more to the reliability of the gridded observational datasets over some points rather than to the model itself.

- Pag 8, line 12-21: In this paragraph you are commenting Figure 3, which is related to simulation a, so it is not wise to comment here also the simulations SOIL and SNOW.

We agree. We will try to introduce the results of simulation **SNOW** and **SOIL** in an additional separated subsection of the results part.

- Pag 9, line 1: Why do you claim that this sentence is "interestingly"?

We think that "interestingly" in this case could be deleted.

- Pag 10, line 10: Why in this case analyses focused on a single observational dataset?

In the figure of the variance ratio we showed the results just for a single observational dataset for each variable, simply for graphical reasons. Nevertheless, the same analyses were conducted for the different observational datasets, and considered when discussing the uncertainties in the estimation of simulated variance. Realizing that this was not clearly specified in the text, we will modify this part accordingly.

- Pag 11, lines 17-18: "For the experiments... experiment q". This sentence is just a repetition of the sentence at lines 13-15. Please combine them.

We will try to merge the two sentences together as suggested by the referee.

- Pag 11, line 24-25: "This indicates...driving dataset". This sentence is very strong and must be supported by results that are more robust.

The few numbers shown in Table 4 are not sufficient. Moreover, you should add in Table 4 the improvements achieved when using NCEP2, in order to have a quantitative comparison.

We agree that the given sentence is too strong. Nevertheless, it is true that the 2 conducted ERAInterim-driven simulations present similar climatological values to the NCEP2-driven ones, in both cases. Even though it is not possible to draw a general conclusion on the effects of the boundaries on COSMO-CLM for the region and the given resolution, these results allow to justify the use of NCEP2 instead of commonly employed ERAInterim for the purposes of our research. We will modify the corresponding part of the text in the final version of the paper, being more careful on the conclusion we can draw from our simulations.

- *P*ag 11, lines 32-33: "Values closer... observations". This statements are obvious and can be removed.

We agree and will modify the text accordingly.

Minor Comments

- *P*ag 6, line 23: You have already explained that the reference configuration is the a. Please remove "(a,Tab.2)".

We agree and will modify the text accordingly.

- *P*ag 7, line 6: Do you mean Tab.3 (instead of Tab.4)? Otherwise, Tab.3 is never referenced.

We acknowledge the error. We referred to Tab.3. We will correct it in the final version of the manuscript.

- *T*itle of Fig. 6: If SS is defined according with equation (1), why did you add (%) next to SS?

Actually we propose to express SS in %

- *l*ines 16-17: In the title you use NCEP, in the text NCEP2, please use always the same acronym.

We will correct the error in the final version of the manuscript.

Below we propose some additional bibliography that we will provide in the revised version of the manuscript, if not already considered, following the referee comments and the proposed discussion.

References

[Ozturk et al. (2012)] . Ozturk, T. and Altinsoy, H. and Türke, M. and Kurnaz, M.L., 2012. *Simulation of temperature and precipitation climatology for the Central Asia CORDEX domain using RegCM 4.0*, Climate research, 52, 63–76.

[Bucchignani et al. (2014)] . Bucchignani, E. and Montesarchio, M. and Cattaneo, L. and Manzi, M. P. and Mercogliano, P., 2014. *Regional climate modeling over China with COSMO-CLM: Performance assessment and climate projections*, Journal of Geophysical Research: Atmospheres, 119, 121–151.

[Wang et al. (2013)] . Wang, D. and Menz, C. and Simon, T. and Simmer, C. and Ohlwein, C., 2013. *Regional dynamical downscaling with CCLM over East Asia*, Meteorology and Atmospheric Physics, 121, 39–53.

[Kotlarski et al. (2014)] . Kotlarski, S. and Keuler, K. and Christensen, O.B. and Colette, A. and Déqué, M. and Gobiet, A. and Goergen, K. and Jacob, D. and Lüthi, D. and Van Meijgaard, E. and others, 2014. *Regional climate modeling on European scales: a joint standard evaluation of the EURO-CORDEX RCM ensemble*, Geoscientific Model Development, 7, 1297–1333.

[Park et al. (2013)] . Park, J.H. and Oh, S.G. and Suh, M.S., 2013. *Impacts of boundary conditions on the precipitation simulation of RegCM4 in the CORDEX East Asia domain*, Journal of Geophysical Research: Atmospheres, 118, 1652–1667.

[Martynov et al. (2013)] . Martynov, A. and Laprise, R. and Sushama, L. and Winger, K. and Šeparović, L. and Dugas, B., 2013. *Reanalysis-driven climate simulation over CORDEX North America domain using the Canadian Regional Climate Model, version 5: model performance evaluation*, Climate dynamics, 41, 2973–3005.

[Russo et al. (2016)] . Russo, E. and Cubasch, U., 2016. *Mid-to-late Holocene temperature evolution and atmospheric dynamics over Europe in regional model simulations*, Climate of the Past, 12, 1645–1662.

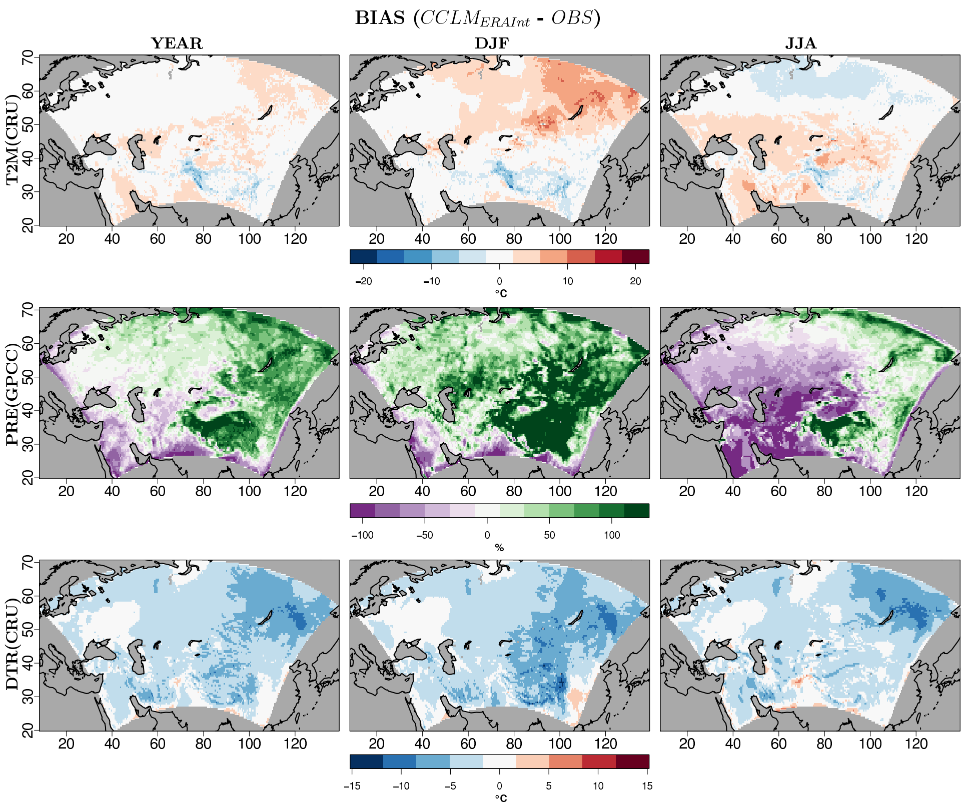


Figure 1: Mean bias of annual mean (left), winter mean (middle) and summer mean (right) near surface temperature (T2M, °C, top panel), precipitation (PRE, %, middle panel) and diurnal temperature range (DTR, °C, bottom panel) of the reference COSMO-CLM configuration (a), driven by ERAInterim reanalysis, with respect to 3 considered observational datasets (from top to bottom: CRU, UDEL and MERRA2), for the period 1995-2005.

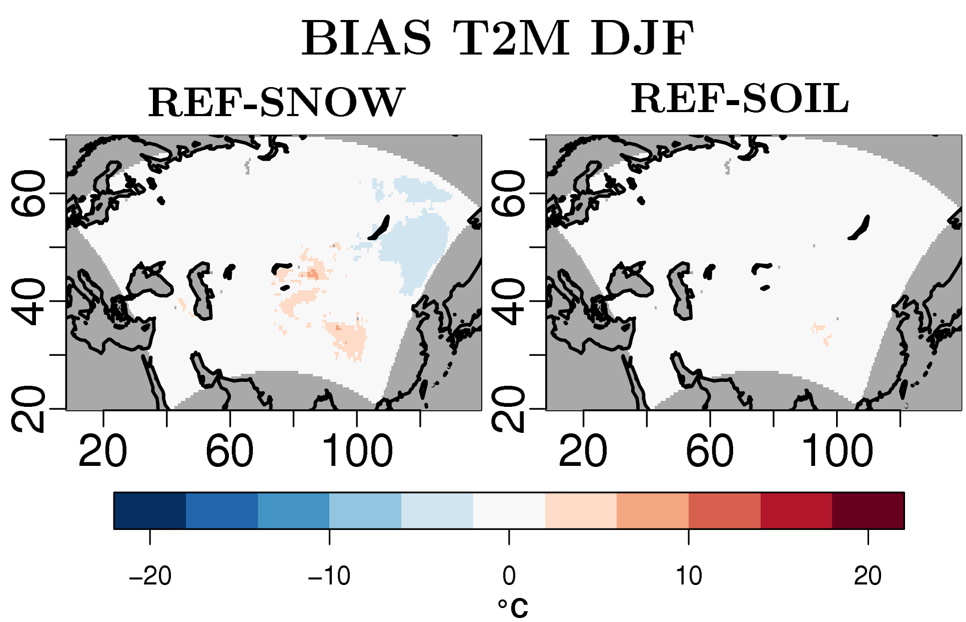


Figure 2: Mean bias of near surface temperature (T2M, °C) winter values of the reference simulation **a** and the simulation **SNOW** (*left*) and **SOIL** (*right*) , all driven by NCEP2, calculated over the period 2006-2015.

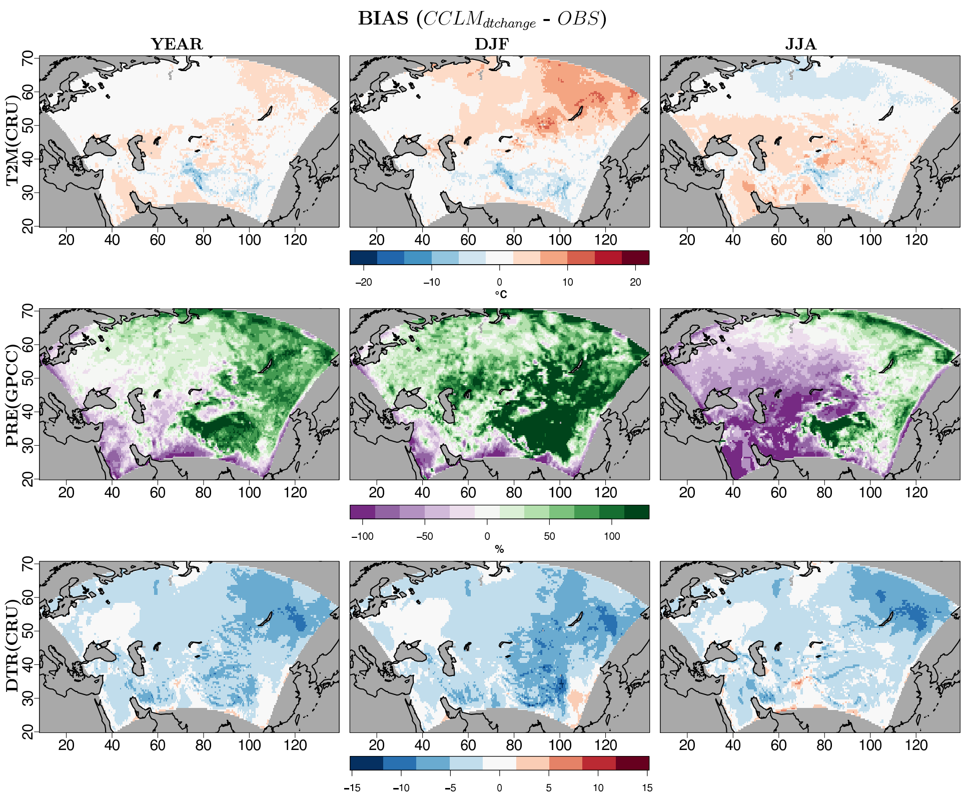


Figure 3: Maps of bias calculated for the NCEP2-driven simulation with the reference configuration but different time step, against different observational datasets for the 3 considered variables. Every row presents the bias calculated for the different considered variables, from top to bottom, respectively, near surface temperature (T2M, °C, CRU), precipitation (PRE, %, GPCC) and diurnal temperature range (DTR, °C, CRU). From left to right, values of the biases for yearly, winter and summer mean climatologies over the period 1996-2005 are presented.

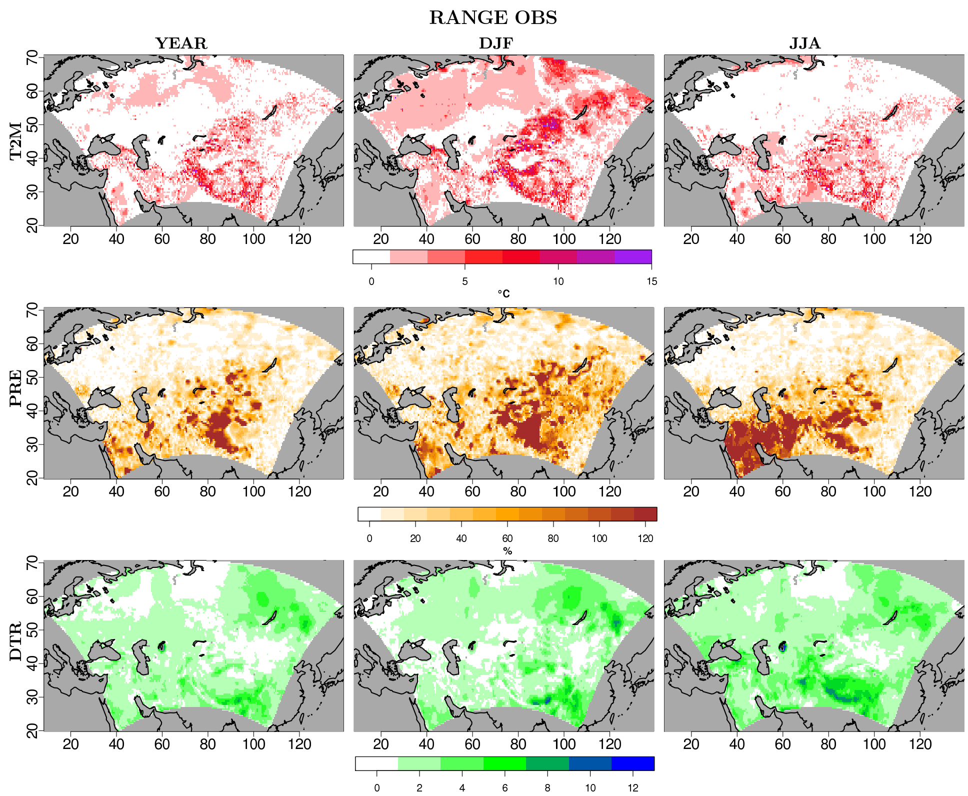


Figure 4: Maps of the spread calculated among different observational datasets for each considered variable, for the annual mean (left), winter (middle) mean and summer mean (right). From top to bottom, the values for near surface temperature (T2M, °C), precipitation (PRE, %) and diurnal temperature range (DTR, °C, bottom panel) are respectively represented.

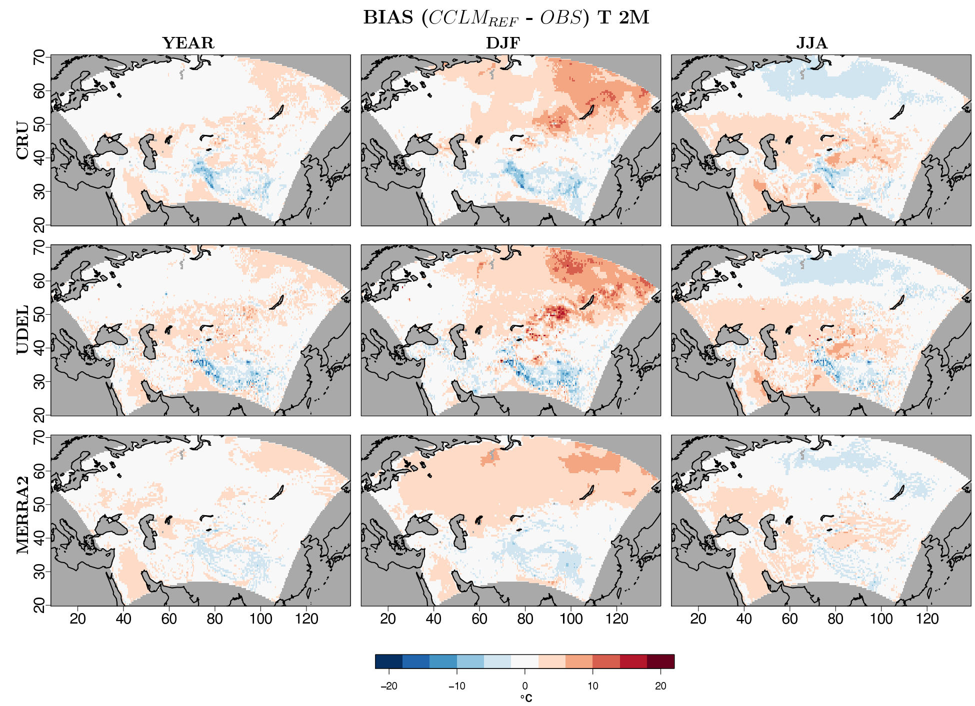


Figure 5: Mean bias of annual mean (*left*), winter mean (*middle*) and summer mean (*right*) near surface temperature (T2M, °C), of the reference COSMO-CLM simulation (a) with respect to three observational datasets (from top to bottom: CRU, UDEL and MERRA2), for the period 1995-2005.

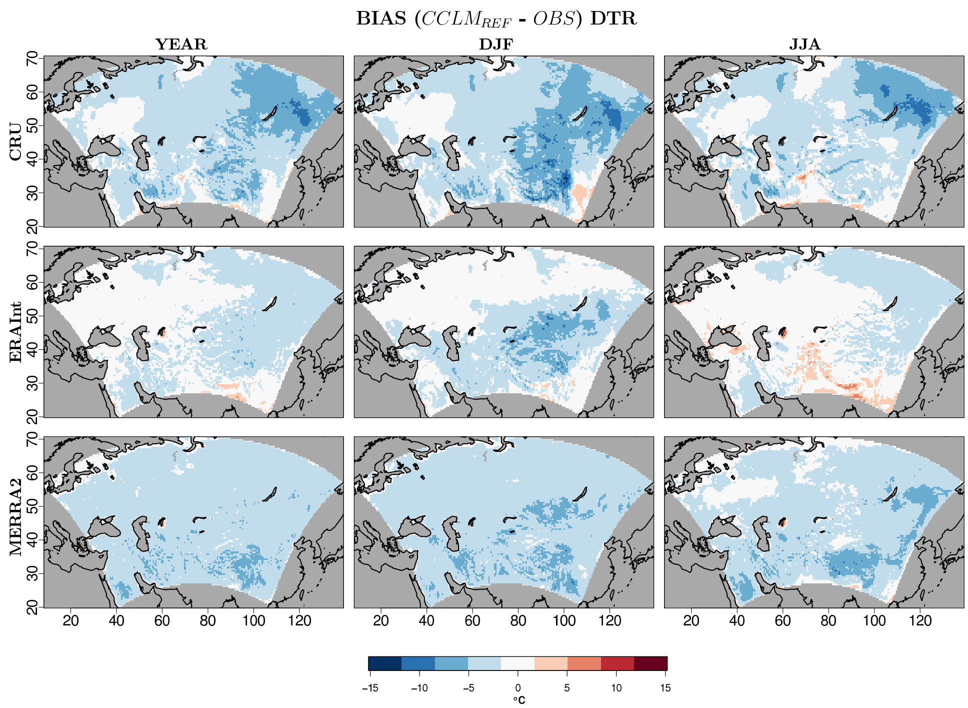


Figure 6: Mean bias of annual mean (*left*), winter mean (*middle*) and summer mean (*right*) Diurnal Temperature Range (DTR, °C), of the reference COSMO-CLM simulation (a) with respect to three observational datasets (from top to bottom: CRU, MERRA2 and ERAInterim), for the period 1995-2005.

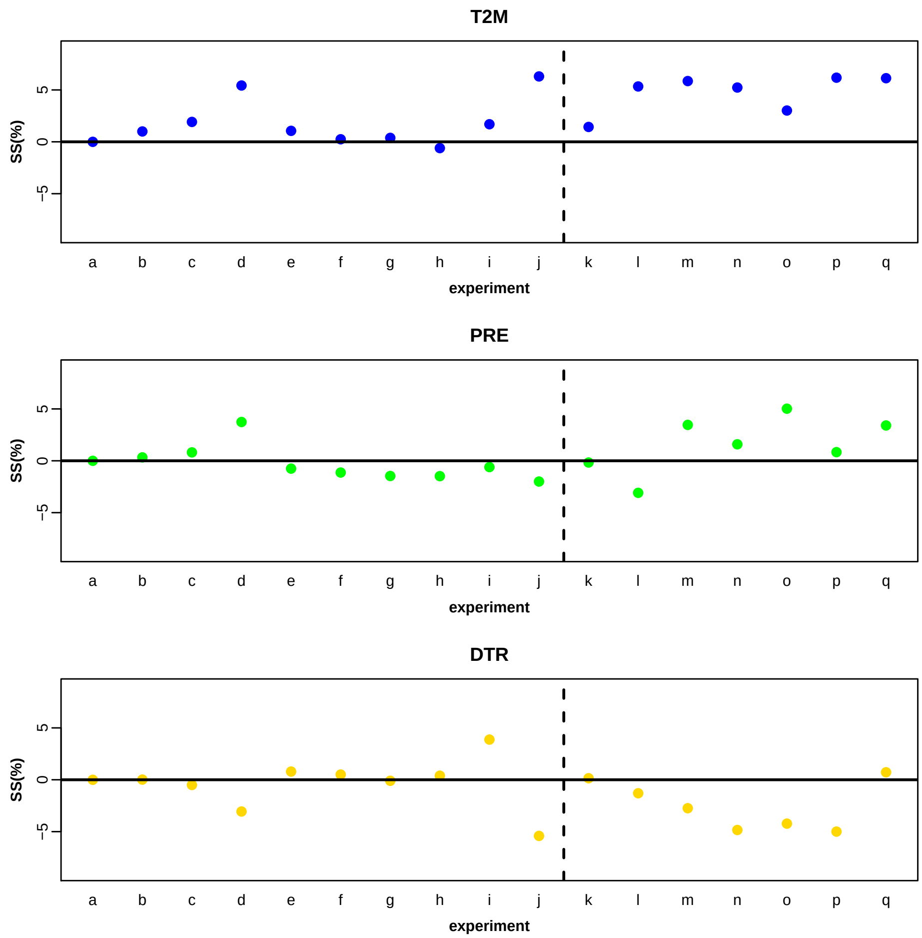


Figure 7: Skill Score (SS) derived from the weighted MAE (MAE_w, Eq. 1) calculated over the monthly climatological values of the seasonal cycle of different COSMO-CLM simulations and observational datasets. From top to bottom, the SS for each variable is displayed. The dotted vertical black line divides the simulations with the same configuration of the reference simulation plus a single change in the model setup (on the left) and the ones obtained through the combinations of the previous ones (on the right). Positive (negative) values indicate better (worse) performance of the considered simulations compared to the reference one.

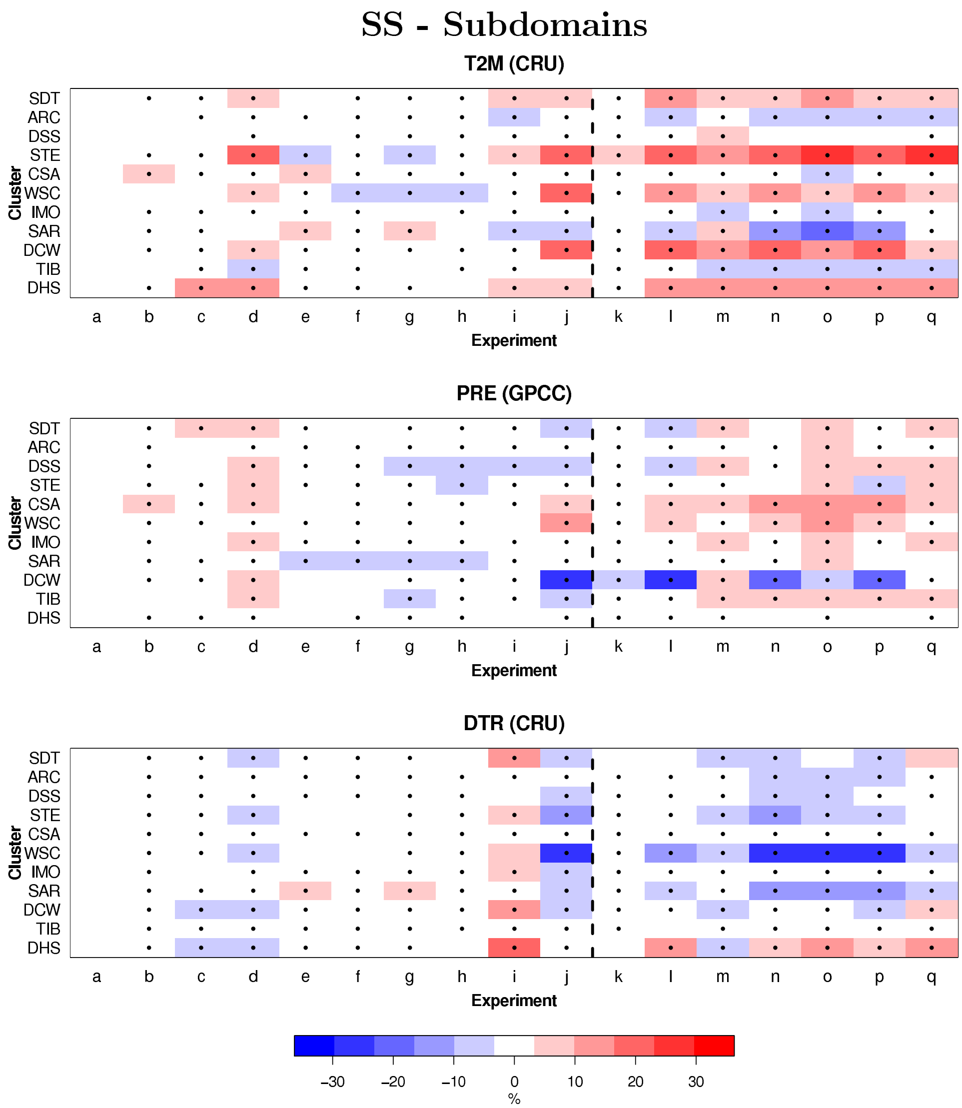


Figure 8: Skill Score (SS) derived from the MAE calculated for each domain sub-region over the monthly climatological values of the seasonal cycle of different COSMO-CLM simulations and observational datasets. From top to bottom, each panel represents the results obtained for near surface temperature (T2M) using the CRU, for precipitation (PRE) using the GPCC and for diurnal temperature range (DTR) using the CRU as observational datasets. Positive (negative) values indicate better (worse) performance of the considered simulations compared to the reference one. The points in correspondence of different clusters and experiments indicate that the change is the SS are the same in sign, when considering all the observational datasets for each variable.

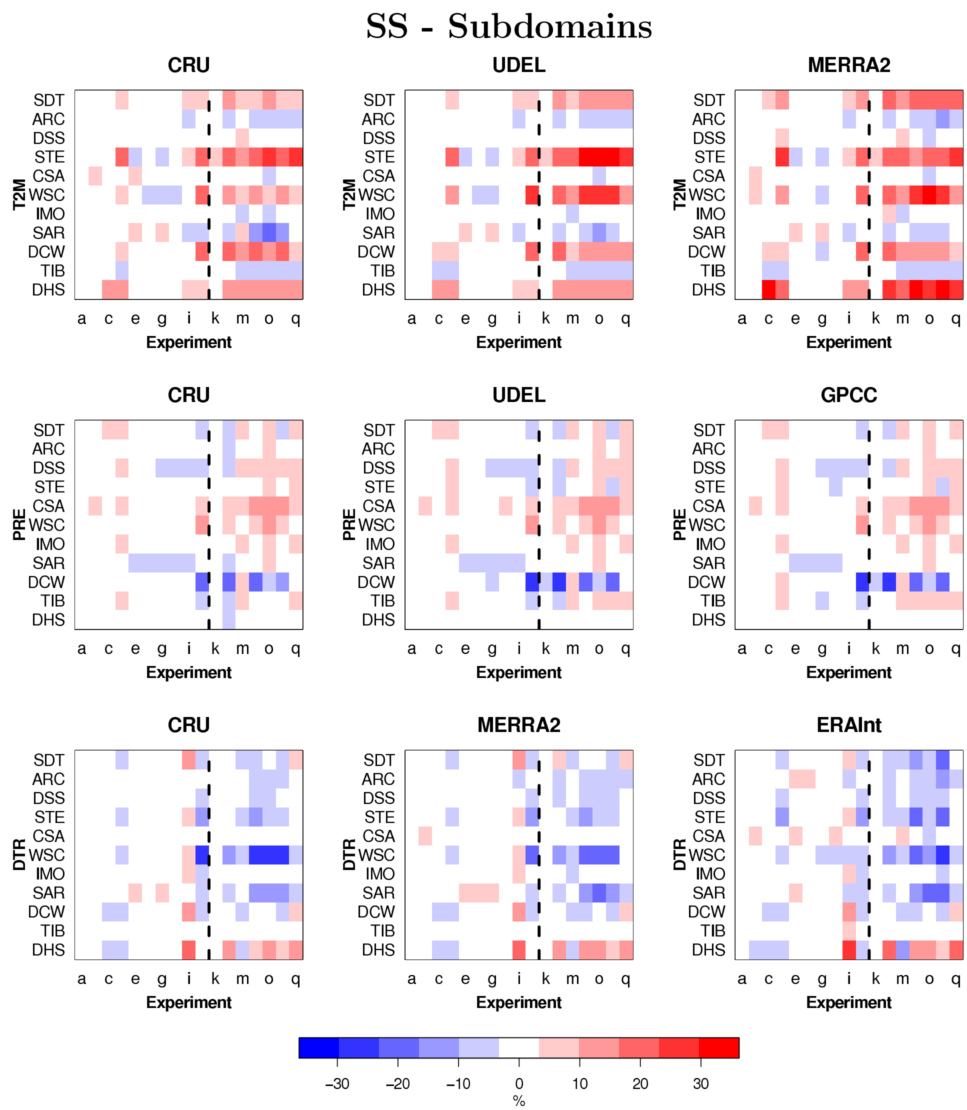


Figure 9: Skill Score (SS) derived from the MAE calculated for each domain sub-region over the monthly climatological values of the seasonal cycle of different COSMO-CLM simulations and observational datasets. From top to bottom, each panel represents the results obtained for near surface temperature (T2M) , for precipitation (PRE) and for diurnal temperature range (DTR). In each case the results obtained for different observational datasets are shown. Positive (negative) values indicate better (worse) performance of the considered simulations compared to the reference one.

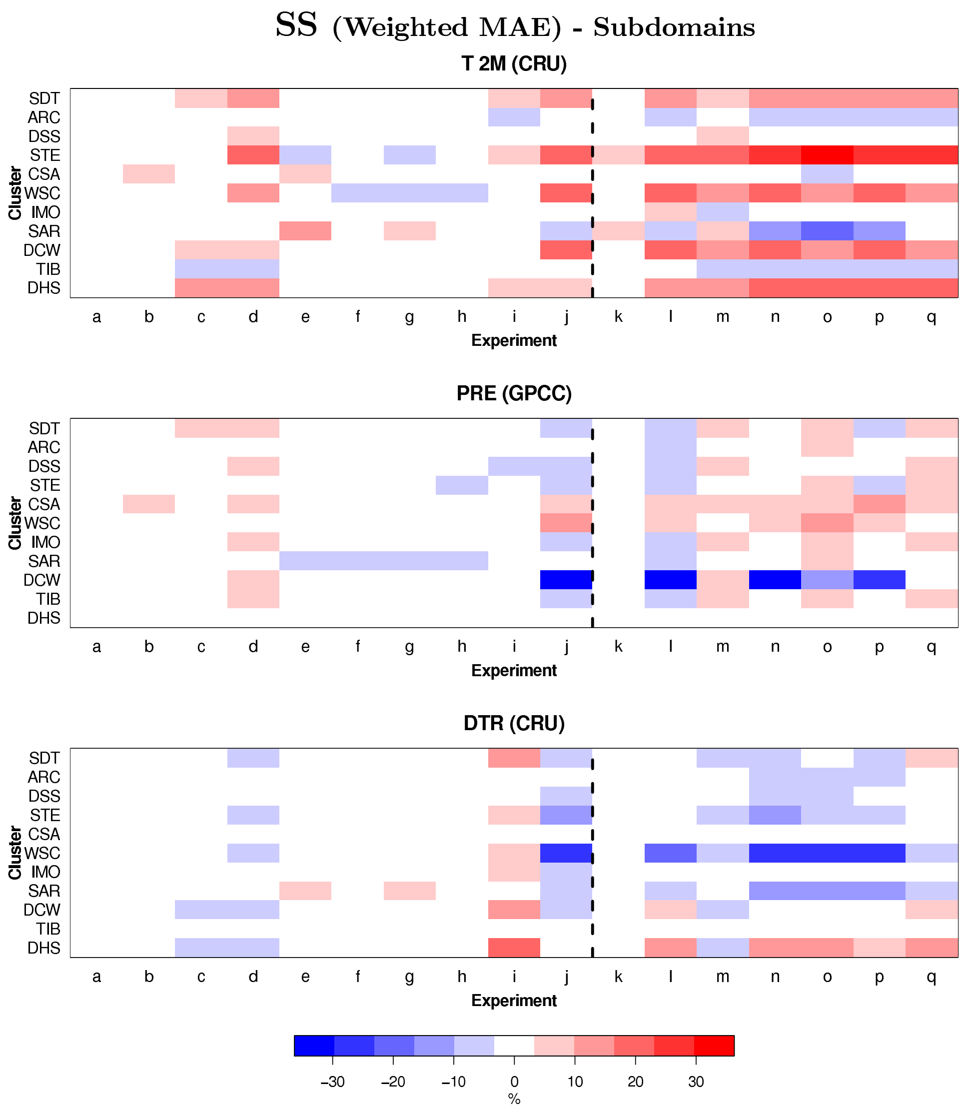


Figure 10: Skill Score (SS) derived from the weighted MAE (MAE_w, Eq. 1) calculated for each domain sub-region over the monthly climatological values of the seasonal cycle of different COSMO-CLM simulations and observational datasets. From top to bottom, each panel represents the results obtained for near surface temperature (T2M) using the CRU, for precipitation (PRE) using the GPCC and for diurnal temperature range (DTR) using the CRU as observational datasets. Positive (negative) values indicate better (worse) performance of the considered simulations compared to the reference one.

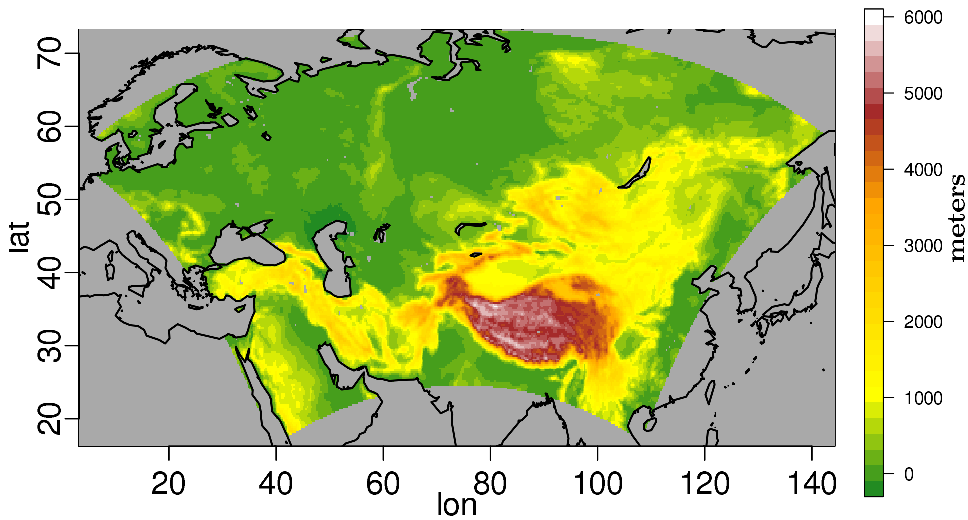


Figure 11: Orography map of the Central Asia simulation domain on a regular grid with a spatial resolution of 0.25 km. Masked in gray are the ocean and the external area of the domain.

With kind regards on behalf of the all authors,

Emmanuele Russo

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

- $P6$ 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.

- $P7$ L6: Tab. 3 not Tab.4, the same for P9 L6 and P12 L14

We agree and will modify the text accordingly.

- $P7$ L7-8 Combine two paragraphs into one

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

- $P7$ L13: Mean Absolute Error to Mean Absolute Error (MAE)

We will modify the text accordingly to the referee comment.

- $P11$ 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.

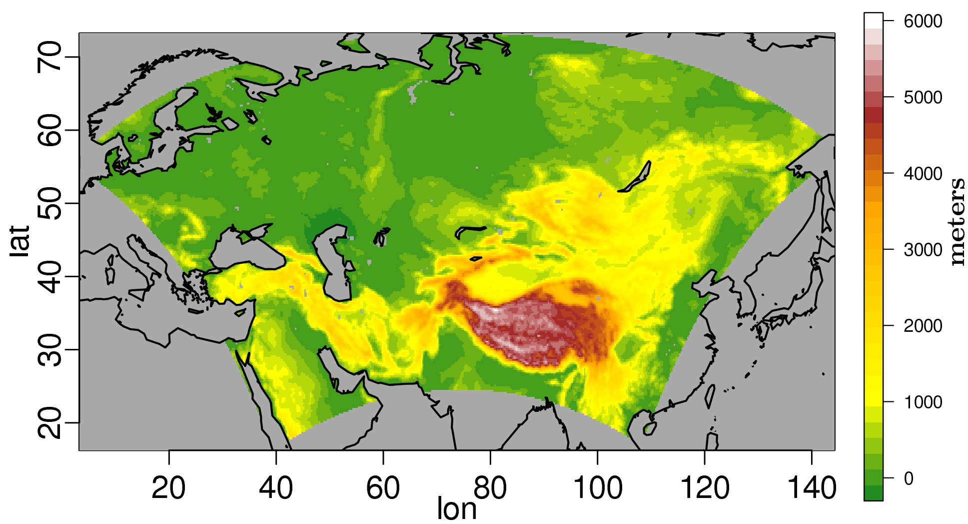


Figure 1: Orography map of the Central Asia simulation domain on a regular grid with a spatial resolution of 0.25 km. Masked in gray are the ocean and the external area of the domain.

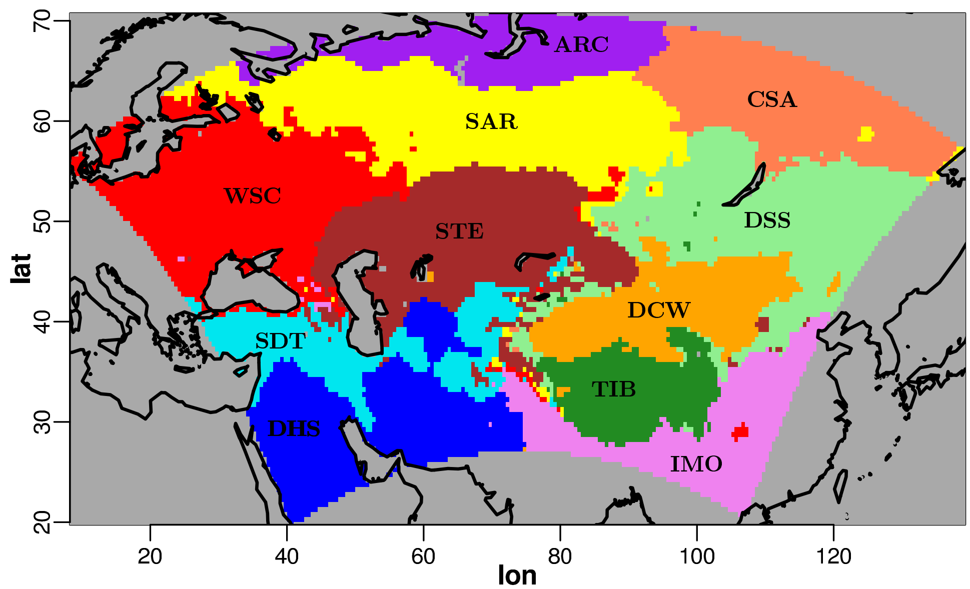


Figure 2: Map of the 11 sub-domains obtained through k-means clustering of the q-normalized monthly climatologies of the three considered variables over the period 1996-2005.

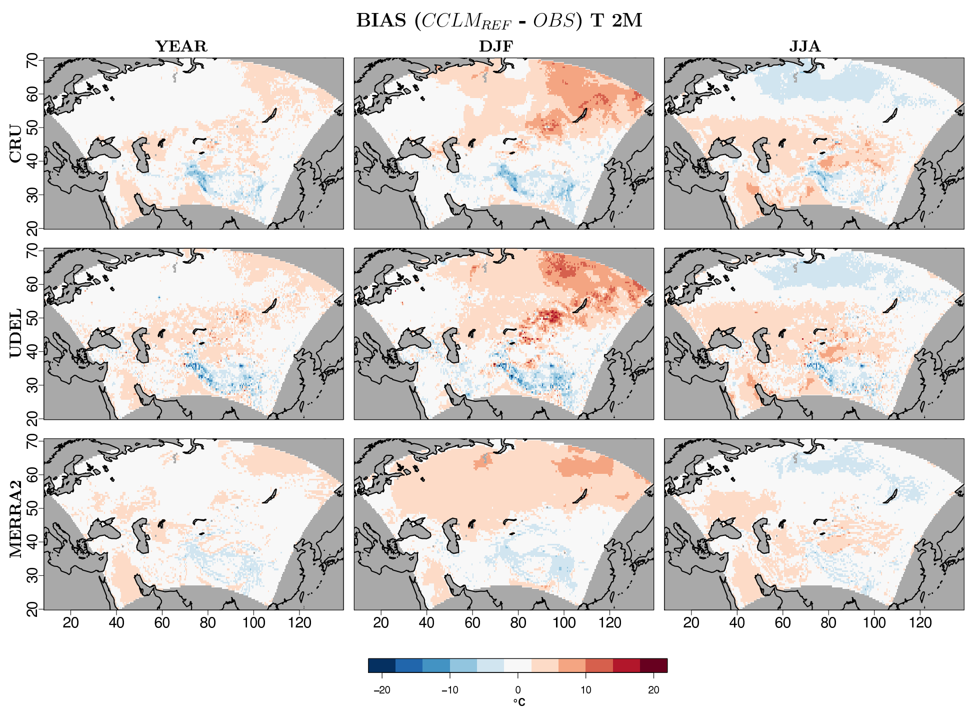


Figure 3: Mean bias of annual mean (*left*), winter mean (*middle*) and summer mean (*right*) near surface temperature (T2M, °C), of the reference COSMO-CLM simulation (a) with respect to 3 considered observational datasets (from top to bottom: CRU, UDEL and MERRA2), for the period 1995-2005.

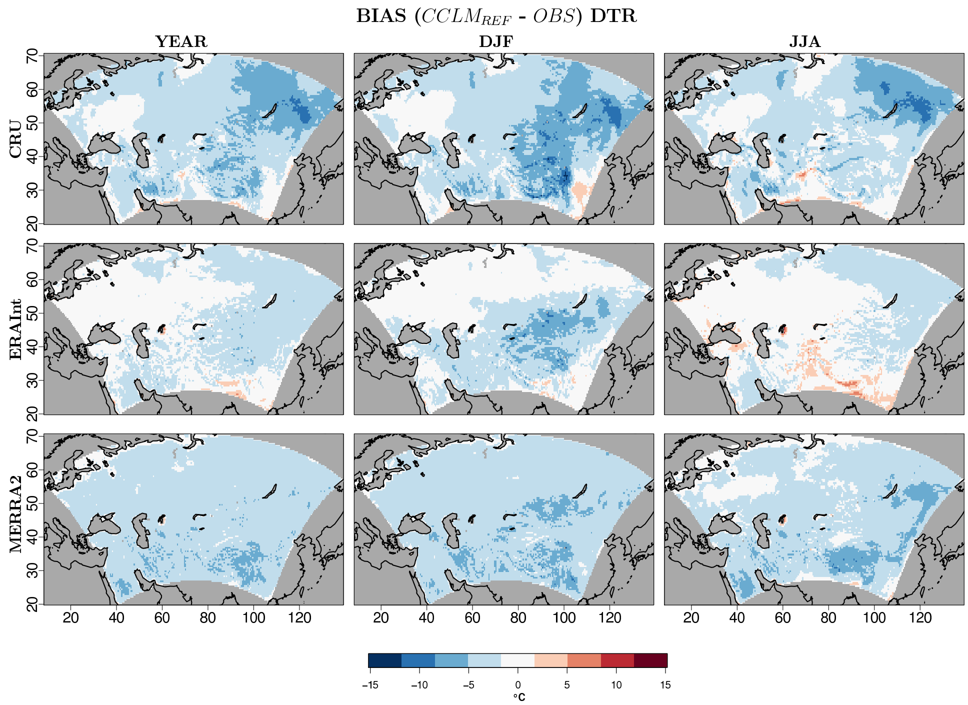


Figure 4: Mean bias of annual mean (*left*), winter mean (*middle*) and summer mean (*right*) Diurnal Temperature Range (DTR, °C), of the reference COSMO-CLM simulation (a) with respect to the 3 considered observational datasets (from top to bottom: CRU, MERRA2 and ERAInterim), for the period 1995-2005.

With kind regards on behalf of the all authors,

Emmanuele Russo

Sensitivity studies with the Regional Climate Model COSMO-CLM 5.0 over the CORDEX Central Asia Domain

Emmanuele Russo[1,2,3], Ingo Kirchner[1], Stephan Pfahl[1], Martijn Schaap[4,1], and Ulrich Cubasch[1]

[1]Institute for Meteorology, Freie Universität Berlin, Carl-Heinrich-Becker-Weg 6-10, 12165, Berlin, Germany
[2]Climate and Environmental Physics, Physics Institute, University of Bern, Sidlerstrasse 5, 3012, Bern, Switzerland
[3]Oeschger Centre for Climate Change Research, University of Bern, Hochschulstrasse 4, 3012, Bern, Switzerland
[4]TNO Built Environment and Geosciences, Department of Air Quality and Climate, Princetonlaan 6, 3584, CB, Utrecht, The Netherlands

Correspondence: emmanuele.russo@met.fu-berlin.de

Abstract. Due to its extension, geography and the presence of several under-developed or developing economies, the Central Asia domain of the Coordinated Regional climate Downscaling Experiment (CORDEX) is one of the most vulnerable regions on Earth to the effects of climate changes. Reliable information on potential future changes with high spatial resolution acquire significant importance for the development of effective adaptation and mitigation strategies for the region. In this context, Regional Climate Models (RCMs) play a fundamental role.

In this paper, the results of a set of sensitivity experiments with the regional climate model COSMO-CLM version 5.0, for the Central Asia CORDEX domain, are presented. Starting from a reference model setup, general model performance is evaluated for present-days, testing the effects of a set of singular physical parameterizations and their mutual interaction on the simulation of monthly and seasonal values of three variables that are important for impact studies: 2-meter temperature, precipitation and diurnal temperature range. The final goal of this study is two-fold: having a general overview of model performance and its uncertainties for the considered region and determining at the same time an optimal model configuration.

Results show that the model presents remarkable deficiencies over different areas of the domain. The combined change of the albedo taking into consideration the ratio of forest fractions and the soil conductivity taking into account the ratio of liquid water and ice in the soil, allows to achieve the best improvements in model performance in terms of climatological means. Importantly, the model seems to be particularly sensitive to those parameterizations that deal with soil and surface features, and that could positively affect the repartition of incoming radiation. The ~~results for the mean climate appear to be independent of the observational dataset used for evaluation and of the~~ analyses also show that improvements in model performance are not achievable for all domain sub-regions and variables, and they are the result of some compensation effect in the different cases. The proposed better performing configuration in terms of mean climate, leads to similar positive improvements when considering different observational datasets and boundary data employed to force the simulations. On the other hand, due to the large uncertainties in the variability estimates from observations, the use of different boundaries and the model internal variability, it has not been possible to rank the different simulations according to their representation of the monthly variability.

This work is the first ever sensitivity study of an RCM for the CORDEX Central Asia domain and its results are of fundamental importance for further model development and for future climate projections over the area.

1 Introduction

Regional Climate Models (RCMs) are a fundamental tool for the study of climate change, allowing to reproduce the climate system with a high quality of details and to provide information at a regional scale. Their use for future climate projections, constitutes indeed a vital resource for policy makers in their decision making under the threat of future global warming (Kim et al., 2014).

The Coordinated Regional climate Downscaling Experiment (CORDEX) (Giorgi et al., 2009) is an initiative sponsored by the World Climate Research Programme, aiming to coordinate international regional climate downscaling experiments. CORDEX sets a number of directives, including predefined resolution, regions, output variables and formats, to facilitate analysis of possible future climate changes (Nikulin et al., 2012).

Among the different CORDEX regions, Central Asia represents one of the largest domains, covering parts of Europe, Africa and almost the entire Asian continent. The domain extends from eastern Europe to the eastern part of China and from the northern part of India and the Arabian Peninsula in the South, to Siberia and the Arctic ocean (Barents sea and Kara sea) in the North. It includes, almost entirely, two of the most important and populated countries of the World: China and Russia. The region, despite being mainly characterized by arid and semi-arid climatic conditions, presents a wide and differing variety of climatic zones, going from the desertic zones of Gobi and the Arabian peninsula, to the cold and dry areas of Siberia and the wet Northern Indian monsoon area (Ozturk et al., 2017). Therefore, it offers the unique opportunity to test the model sensitivity to different climatic conditions at once.

Beside its importance from a modeling perspective, the extension, geography and the presence of several under-developed or developing economies, makes the CORDEX Central Asia domain one of the most vulnerable regions on Earth to the effects of climate changes~~(Lioubimtseva et al., 2005; Lioubimtseva and Henebry, 2009)~~. Even small changes in climate conditions could dramatically affect ecosystems, agricultural crops, water resources, human health and livelihood of the region ~~. According to Siegfried et al. (2012), future climate change will likely exacerbate water stress in the area of inner Central Asia (Kyrgyzstan, Tajikistan, Uzbekistan, Turkmenistan and Kazakhstan), an area that has already experienced water allocation conflicts in the last decades (Siegfried et al., 2012). The countries of this area, beside their geographical conditions, could suffer from their complex political, economic and institutional situation that followed the collapse of the former USSR (Lioubimtseva and Henebry, 2009). Future warming in inner Asia is also expected to lead to increases in forest stress and tree mortality, potentially driving the eventual regional loss of current semi-arid forests (Liu et al., 2013). Other regions of the Central Asia domain for which many studies have highlighted the possible harmful effects of climate change are Western and Central China, with possible impacts on agro-ecosystems, wetland ecosystems, forests, human health, energy sectors and other sensitive fields (Yong-Jian et al., 2013; Zhen-Feng et al., 2013). These regions additionally include the Tibetan Plateau: in this case, the effects of climate change could lead to reduced flow in many rivers that are a primary source of the entire Asias water systems, with dramatic effects on water resources for densily populated areas (Wang et al., 2017). Particularly arid~~

~~regions of the Central Asia domain, such as Mongolia, are particularly prone to the harmful effects of climate change, due to their limited water resources. These regions have already been affected by long and extreme droughts in the last decades, with alarming risks for agricultural areas. Future climatic conditions will likely be drier and warmer, with a significant impact on water resources, food and biodiversity (Chuluunkhuyag, 2008). Interestingly, for Mongolia, studies have suggested that beside the effect on water availability, climate change will also affect local herders, with the number of people migrating due to enviromentally-induced economic reasons increasing in the near future (Diniega, 2012). Another region of the CORDEX Central Asia domain whose unique and delicate ecosystem has already been highly affected by recent climate changes is Siberia. In particular, in the last decades, the eurasian tundra of western Siberia has seen a large spread and growth of shrub cover, due to climate warming (Macias-Fauria et al., 2012). Future warming could further foster these changes. All the reported studies confirm the harmful effects that climate change could have on the region~~(Siegfried et al., 2012; Lioubimtseva et al., 2005; Lioubimts In this context, reliable information on potential future changes with high spatial resolution acquire significant importance for the development of effective adaptation and mitigation strategies.

The recognized prerequisite that every climate model has to satisfy, in order to provide reliable future climate projections, is the ability of realistically simulating present-day climate (Kim et al., 2014; Nature-Editorial-Board, 2010; Kim and Lee, 2003). Assessing the ability of a climate model to simulate the current climate is defined as model evaluation (Airey and Hulme, 1995). Model evaluation consists in an assessment of model quality and deficiencies originating from different modeling assumptions, conducted through the comparison of model outputs and observations (Kim et al., 2014; Kim and Lee, 2003; Flato et al., 2013; Lenderink, 2010; Overpeck et al., 2011; Bellprat et al., 2012a, b). ~~Model evaluation is an essential part of regional model development (Kotlarski et al., 2014).~~ Evaluation experiments normally consist in a set of present-days simulations conducted in a perfect boundary setting, i.e., using reanalysis products as lateral boundary forcings. This "modus operandi" allows for the separation of possible model biases from biases due to erroneous large-scale forcings, thus highlighting specific model deficiencies (Kotlarski et al., 2014). These may be related to the model formulation and to choices in model configuration (Awan et al., 2011; Bellprat et al., 2012a, b; de Elía et al., 2008; Evans et al., 2012). In the second case, it should be possible to improve model performances by testing different model configuration setups and choosing the one that better agrees with observations. This approach might be conceived as an optimization step. Nevertheless, it is important to emphasize the fact that a specific model configuration could produce better results, by simply compensating for some deficiencies in the model formulation (Hourdin et al., 2017).

~~A series of different aspects have to be considered for the configuration of a climate model simulation.~~ In climate models, the complexity and small spatial scales of the physical processes involved, requires the so-called parameterization of many of these processes: this basically consists in summarizing physical phenomena and their interaction across different spatial and temporal scales (Fernández et al., 2007; Rummukainen, 2010; McFarlane, 2011; Hourdin et al., 2017), which is associated with substantial uncertainties. The same processes may be described through different parameterizations, with a different degree of complexity and infinitive parameter values. Consequently, the outcomes of a climate model might largely differ, depending on the parameterizations used and the selected parameters inputs. Additionally, the use of different forcings datasets, for example for greenhouse gases, aerosols or land cover changes, might have a significant effect on the results. Further, other details that need to be considered when configuring a climate model simulation for a defined domain are the configuration and spatial resolution of the model grid (both horizontally and vertically) and the coupling with different models representative of other components of the climate system. For regional climate models, all these aspects are domain dependent (Jacob et al., 2007, 2012; Rockel et al., 2008). This means that a regional climate modeler should always evaluate different model configurations, isolating the one that leads to a better agreement with observations, for each investigated region and employed model. In doing so, several sources of uncertainties should be taken into consideration: the fact that performances of the RCM for a specific region might vary according to the boundary conditions, the model internal variability and observational datasets should be acknowledged when evaluating model performances.

RCMs evaluation and configuration have been the subject of a large number of studies, for different regions, such as in Kotlarski et al. (2014); Umakanth and Kesarkar (2018); Borge et al. (2008); García-Díez et al. (2013); Crétat et al. (2012); Rajeevan et al. For COSMO-CLM, Europe has received larger attention. Bellprat et al. (2012b) applied a quadratic metamodel on a subsample of model parameters in order to objectively tune the model for the region. Their method is considered the reference for COSMO-CLM for determining optimal parameters values and has been further developed and applied to the Mediterranean region by Avgoustoglou et al. (2017) and for higher resolution for the Alpine region by Voudouri et al. (2018). Bellprat et al. (2016) additio used the same method for the European and the North American domain, finding quite similar values of optimal model parameters for the two regions. Using a more subjective approach, Montesarchio et al. (2012) conducted a set of sensitivity studies in order to determine the best model setup for the simulation of near surface temperature and precipitation over Northern and Central Italy, at a spatial resolution of $\sim$8km. Despite several important findings, they were not able to determine an optimal model configuration with respect to these variables. The alpine region was the subject of a study with COSMO-CLM by Suklitsch et al. (2008), where they found that changes in model resolution has a larger impact on model simulation than modifying dynamical and numerical schemes. Bachner et al. (2008) conducted a similar work for Germany, focusing on model performance for summer precipitation, concluding that the model uncertainty due to the modified physical parameterizations is considerable and highlighting the need of conducting evaluation and sensitivity studies prior to the application of a model for climate change projections. Studies investigating COSMO-CLM sensitivity for regions out of Europe, are more rare. Lange et al. (2015) tested model performances to different convection and non-precipitating subgrid-scale clouds parameterizations for South America. Through this work they managed to reduce long-standing model biases in precipitation for the region, by using the IFS and statistical schemes for convection and subgrid-scale clouds. Bucchignani et al. (2016) investigated different model configurations for the Middle East-North Africa (MENA) CORDEX domain, comparing 26 different model configurations. The model seems to be particularly sensitive for the region to changes in physical and tuning parameters. In particular, they found an optimal configuration with a representation of the albedo based on Moderate Resolution Imaging Spectroradiometer data, and a parameterization of aerosol based on NASA-GISS Aerosol Optical Depth distributions. Bucchignani et al. (2012) evaluated several configurations for North-Western China at a spatial resolution of approximately $\sim$8km, even though they did not propose any optimal configuration for the region.

So far, neither an evaluation nor a sensitivity study on the impact of the use of different physical parametrizations of an RCM have been documented for the CORDEX Central Asia domain. Such analyses are required to guide further model development and applications for the region: if we want to produce future climate projections for the Central Asia domain of the CORDEX experiment, we need to investigate model performances and deficiencies for the area and propose optimal model configurations.

In the light of the upcoming phase of the Coordinated Regional climate Downscaling EXperiment (CORDEX) (Giorgi et al., 2009), denominated ~~CORDEX-CORE~~ CORDEX - Coordinated Output for Regional Evaluations (CORE) (Gutowski Jr et al., 2016), in this paper the results of a set of sensitivity experiments with the regional climate model COSMO-CLM version 5.0 (Rockel et al., 2008), for the Central Asia CORDEX domain, are presented. In this perspective, this work represents the first step for the production of climate projections for the Central Asia domain using COSMO-CLM, evaluating general model performances, isolating the effects of different uncertainty sources on model results and determining an optimal model configuration for a region for which almost no reference exists. Starting from a reference model setup, general model performance is evaluated, testing the effects of a set of singular physical parameterizations and their mutual interaction as well as two different forcing datasets on the simulation of monthly and seasonal values of three variables that are important for impact studies. These are near surface temperature (T2M), precipitation (PRE) and diurnal temperature range (DTR), the latter representing the daily excursion between maximum and minimum temperature, which is particularly important in terms of human body adaptability and stress. The final goal of this study is two-fold: having a general overview of model performance and its uncertainties for the considered region and determining at the same time a "best" suitable model configuration.

In section 2 of this paper, the model, the different datasets and the methods employed in this study are presented. Then, in section 3, results are presented. Finally, conclusions are outlined, with a general discussion of model performances and the proposal of a final optimal model configuration for the area of study.

2 Methods

In this section the research methods are described, including details on the model and the different simulation setups, the observational datasets used for the evaluation of model results and the employed metrics.

2.1 Model and Experiments Description

The Consortium for Small-Scale Modeling in Climate Mode (COSMO-CLM (Rockel et al., 2008)) is a non-hydrostatic regional climate model developed by the CLM-community, an open international network of scientists. The model version employed in this study is the COSMO-CLM 5.0_clm9. Many studies have been conducted in the recent years over different CORDEX regions, using the COSMO-CLM (Panitz et al., 2014; Dobler and Ahrens, 2010; Bucchignani et al., 2016; Smiatek et al., 2016; Jacob et al., 2014; Kotlarski et al., 2014; Zhou et al., 2016). ~~This study represents the first application of the COSMO-CLM to the CORDEX Central Asia domain.~~

The simulations presented in this study are performed with a spatial resolution of 0.22^o, as specified in the new CORDEX-CORE directives (Gutowski Jr et al., 2016), on a rotated geographical grid. The initial simulation domain extends ~~over 326 points in longitudes and 220 points in latitudes. It~~ from ∼3°to ∼145°over longitudes and from ∼16°to ∼73°over latidues. The domain includes a model relaxation zone ~~, consisting of 10 additional points~~ of ∼250 km on each domain side~~and~~, used to "relax" the model variables towards the driving data (Køltzow, 2012; Davies, 1976). Results of the simulation for this area are excluded from the analysis, with ~~a~~ the final domain extent ~~of 306 × 200 points, as~~ shown in Fig. 1. If not differently specified, all the simulations are run over a 15 year-long period from 1991 to 2005, with the first 5 years excluded from the analysis and considered as spinup time.

In a set of sensitivity experiments labeled from **a** to **q** in the first section of Tab. 1, the effects on model performance of different physical parameterizations are tested, first individually and then combining them with each others. The setup of experiment **a** is the reference from which the other experiments are configured, by implementing the modifications specified in the table. The model configuration used for the reference simulation is the same used for the CORDEX East Asia domain for the COSMO-CLM model version 5.0~~, available on the CORDEX page of the CLM-community website ()~~. This was considered as a good reference for the purposes of this study, since the two regions share a large part of their domains. A general description of the setup of the reference simulation is provided in Tab. 2.

All the performed simulations are driven by the NCEP version 2 reanalysis data (Kanamitsu et al., 2002), provided as boundary and initial conditions. The boundaries have a temporal resolution of 6 hours and a spectral resolution of T62 ($\sim 1.89^o$ lon). NCEP2 data have been selected as boundary data, instead of commonly employed ERAInterim reanalyses (Dee et al., 2011), since their spatial resolution is closer to the one of the 3 Global Circulation Models (GCMs) ~~normally employed in CORDEX simulations . In order to estimate the effects of the driving data on the simulations results and to support possible conclusions on an optimal setup~~that are used for CORDEX-CORE simulations in the CLM community: MPI-ESM (Giorgetta et al., 2013), HadGEM (The HadGEM2 Development Team: Martin et al., 2011) and NorESM (Bentsen et al., 2013; Iversen et al., 2013), with a spatial resolution of, respectively, around 210km, 210x140km and 270x210 km. Thus, using NCEP2 as drivers allows to reproduce a resolution jump more similar to the one of the considered GCMs.

Acknowledging the fact that ERAInterim are normally employed for the evaluation of RCMs, two additional simulations are performed, driven by ERAInterim reanalysis data (second section of Tab. 1), which have a spectral resolution of T255 ($\sim 0.7^o$ lon). This allows to estimate the effects of the two different driving data on the simulations results and to support possible conclusions on an optimal setup, verifying how significantly the results differ in the two cases.

~~A set of 4 simulations are additionally performed for the investigation of the model internal variability~~ In order to better discriminate different sources of uncertainties in the model simulations, a run covering the period 1991-2005 is also performed (third section of Tab. 1)~~. These simulations have the same setup as the reference simulation **a**, but are initialized at four different dates, shifted by +/- 1 and 3 months with respect to the reference one~~ , using a different timestep of 120s instead of the one of the reference simulation of 150s.

~~Finally, two additional~~ Two 25-year long simulations, covering the period 1991-2015, are performed for testing different configurations that could help in reducing model biases over areas characterized by the presence of permafrost in winter. The two simulations, labeled **SOIL** and **SNOW** in the bottom part of Tab. 1, are performed, respectively, increasing the number of soil layers from 10 to 13, together with their total depth from approximately 15 m to more than 130 m, and using the multi-layer snow model of COSMO-CLM (Machulskaya, 2015). These simulations cover a longer period than the others, since a longer spinup time is necessary in order to account for more and deeper soil layers. Their results, excluded from the direct comparison with the other simulations, are discussed in the results and concluding sections of this paper.

Finally, a set of 4 simulations are additionally performed for the investigation of the model internal variability (fourth section of Tab. 1). These simulations have the same setup as the reference simulation **a**, but are initialized at four different dates, shifted by +/- 1 and 3 months with respect to the reference one.

All the proposed simulations are designed with the goal of better understanding main model limitations for the area and to which degree they can be reduced by properly configuring the model, isolating the effects of different sources of uncertainties.

2.2 Observations

Gridded observational datasets are used to compare model results against observational data on a similar scale. These gridded datasets are obtained through statistical extrapolations of surface observations. In addition to uncertainties related to the original measurements, these datasets also contain important uncertainties due to the statistical extrapolation procedure (Flaounas et al., 2012; Gómez-Navarro et al., 2012). For climate model evaluation studies, these uncertainties are usually taken into account by using a range of different datasets (Collins et al., 2013; Gómez-Navarro et al., 2012; Bellprat et al., 2012a, b; Flaounas et al., 2012; Lange et al., 2015; Zhou et al., 2016; Solman et al., 2013).

In this study, the issue of observational uncertainties is addressed by considering three different datasets for each of the investigated variables. The datasets include both observations and reanalysis data. For all the three considered variables, information is retrieved from the CRU TS4.1 observational dataset (Harris and Jones, 2017). Information on T2M and PRE is also retrieved from the University of Delaware (UDel) gridded dataset (Willmott, 2000), provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at https://www.esrl.noaa.gov/psd/. For T2M and DTR, in addition, the Modern-Era Retrospective analysis for Research and Applications, version 2 (MERRA2) (Gelaro et al., 2017) is employed. For precipitation, the third considered dataset is the Global Precipitation Climatology Centre dataset (GPCC) (Becker et al., 2011), while the ERAInterim reanalysis dataset (Dee et al., 2011) is used in addition to MERRA2 and to CRU for the evaluation of DTR.

All the datasets are retrieved on a grid with the same spatial resolution of 0.5 °. The ERAInterim data, that originally have a horizontal resolution of approximately 80km, are interpolated to the same grid resolution. The output of the simulations is upscaled to the same 0.5 ° grid of the observations. For temperature and diurnal temperature range, a bilinear remapping method is used for the upscaling, while for precipitation a conservative remapping method is employed. The Climate Data Operators (CDO) software package (available at http://www.mpimet.mpg.de/cdo,version 1.9.5) is used for the interpolation.

Fig. 2 shows the spread of the different observational datasets for each variable, for yearly, winter and summer climatological means over the period 1996-2005. As evident, large differences emerge among the different datasets, in particular for regions characterized by complex topography and lower observational stations density, such as the Tibetan Plateau and Siberia. The given spread could make it hard to quantify model biases over certain regions. In the case of T2M and DTR, the spread is certainly influenced by the fact that some of the datasets are reanalyses. Nevertheless, for T2M, differences exceeding 8°are present, in particular in winter, even between the CRU and the UDEL, over regions where the interpolation is highly affected by the low number of stations (Matsuura and Wilmott, 2012; Bucchignani et al., 2014). For PRE, the spread (expressed in percentage with respect to the GPCC values) is remarkable in winter over the Tibetan Plateau and in summer over particularly dry areas. Despite the differences might likely be influenced by the employed interpolation methods in each case, the spatial coverage of observation is still considered their main driver (Dong and Sun, 2018; Matsuura and Wilmott, 2012; Sun et al., 2018; Naumann

2.3 Analysis Details and Evaluation Metrics

In order to rank different model configurations according to their skills in simulating the three considered variables over the region, their performances are evaluated with respect to the ones of the reference simulation (**a**, Tab. 2).

Since in the context of CORDEX simulations the main interest is often on the comparison of the mean climate between two different periods in time, the primary focus of the proposed analyses is on climatological monthly ~~and seasonal~~ means of the considered variables. In addition, the results are supported by the investigation of the simulated variability.

In the latter case, since the model is not expected to exactly match the observed temporal evolution of the investigated variables point by point (Gleckler et al., 2008; Wilks, 2006), regional mean anomalies are considered. For each grid point in the domain, monthly anomalies are first calculated by subtracting the climatological mean from each monthly value. The variability is then ~~analysed~~ analyzed based on these anomalies averaged over sub-regions characterized by similar ~~climatic conditions.~~ climate conditions.

The decomposition of the domain into a set of sub-regions is obtained by means of a k-means clustering (Steinhaus, 1956; Ball and Hall Dj, 1965; MacQueen et al., 1967; Lloyd, 1982; Jain, 2010) of quantile-normalized (q-normalized) monthly climatologies of the investigated variables. K-means is a clustering technique using the concept of Euclidean distance from the centroids of a pre-determined group of clusters, for separating similar data into groups. For the purposes of this paper, following several tests and the results of other studies (Mannig et al., 2013), a total number of ~~eleven~~ 11 clusters is selected. The k-means clustering algorithm is reiterated over 3000 times in order to achieve the presented results, using q-normalized values of monthly climatologies of 2-meter temperature and diurnal temperature range derived from the CRU dataset and precipitation values derived from the GPCC as input. Fig. 2 shows the results of the k-means clustering. The mean climatologies of the considered regions for the three investigated variables are also reported in Tab. ~~??~~3.

For both the analyses of mean climate and variability, metrics adapted from Gleckler et al. (2008) are used.

In the following subsections, we give more details on the employed metrics.

2.3.1 Climatology

For the evaluation of the climatological means, we employ a Skill Score (SS) metrics expressed as:

$$SS = (1 - \frac{(MAE)_{exp}}{(MAE)_{ref}}) \times 100 \quad (1)$$

where the Mean Absolute Error (MAE) is given by:

$$MAE = \frac{1}{W} \sum_{i=1} \sum_{j=1} \sum_{m=1} w_{ijm} |sim_{ijm} - obs_{ijm}| \tag{2}$$

where *sim* and *obs* are the monthly climatological means of, respectively, the considered simulation and observational dataset. The indices *i*, *j* and *m* vary, respectively, over longitudes, latitudes and months of a year. *W* is the sum of the weights w_{ijm}, taking into account the different lengths of the months and the grid boxes effective area. The *SS* is calculated with respect to a reference simulation. Positive values indicate an improvement of the considered simulation *exp* with respect to the reference *ref*, while negative values indicate worse performances.

The analyses of MAE for the mean seasonal cycle are conducted for the entire domain and single sub-regions. Additionally, the same metrics are applied for the analysis of single seasons for the entire area.

2.3.2 Variability

The analysis of the model performances in simulating the mean climate is complemented by the investigation of simulated variability.

There is no reason to expect models and observations to agree on the phasing of internal (unforced) variations. Hence metrics such as MAE are not appropriate for characterizing the model performance of interannual variability (Gleckler et al., 2008). Here, for an overall evaluation of the simulated variance in the different cases, the ratio of simulated to observed variance is considered:

$$Variance\ ratio = \frac{\sigma^2_{exp}}{\sigma^2_{obs}} \tag{3}$$

It is important to mention that correctly matching the observed variance does not guarantee a correct representation of the modes of variability associated with this variance.

For tacking into account observational uncertainties, all the proposed analyses are conducted separately for each observational dataset. Changes in model performance due to different configurations will be considered relevant only when consistent among the given observations.

3 Results

In this section the results of the conducted analyses are presented, starting from the consideration of climatological means and followed by the analyses of simulated-to-observed variability.

3.1 Mean Climate

In order to characterize the general performances of the model over the region, for the three considered variables maps of the yearly, winter and summer mean biases of the reference model simulation **a** with respect to the different observational datasets, are first presented.

Fig.3 shows that for temperature, the largest biases are evident in winter (central panels), with warmer simulated conditions over the northeastern part of the domain, where the bias in some case exceeds 15^{o}C. ~~The two simulations (**SOIL** and **SNOW**) specifically designed for testing the effects of changes in soil depth andthe use of a multi-layer snow model on the COSMO-CLM simulation of near surface temperature over areas characterized by the presence of permafrost in winter do not present significantly different results (not shown)~~ This exagerated biases are mainly relative to the UDEL dataset and, in general, particularly large biases are limited to a few points characterized by complex topography and lower stations density, where the gridded datasets are less reliable. When the CRU dataset is considered, the values of the bias rarely exceed (are below) 10 (-10) °C, for really few points. Beside these points, still some remarkable biases are present but these are well within the ranges of other CORDEX simulations for the area (Wang et al., 2013; Bucchignani et al., 2014; Ozturk et al., 2012). In summer (Fig.3, right panels), a positive bias (ranging from +5^{o}C to +10^{o}C), is present over the central and south-western part of the domain, in arid and desertic areas such as the Arabian Peninsula and the Taklamakan desert. Conversely, a cold bias is present over Siberia in this case, with values rarely below -5^{o}C. Biases of annual mean values (Fig.3, left panels) are smaller than in the seasonal cases, except for the Tibetan Plateau. Here a similar particularly pronounced cold bias is evident with respect to all observational datasets, with values sometimes smaller than -10^{o}C. In this case the observations are likely less reliable. For all seasons, the simulation results are in better agreement with the MERRA2 dataset than with the CRU and the UDEL. Nevertheless, despite the evinced differences in the ~~amplitude, the pattern of the bias is similar for all datasets~~magnitude of the model biases against different observational datasets, their spatial patterns are very similar in all the cases.

Concerning precipitation (Fig.4), remarkable biases are present in the winter and summer as well as in the annual mean for all the observational datasets. The biases in this case are expressed as percentage with respect to the values of observational estimates. In summer (Fig.4, right panels), a particularly pronounced negative bias, with values down to -100%, is visible over arid regions and the monsoon area. This is of the same order of the spread of the observational datasets for the area. Over the Tibetan Plateau the bias in ~~this case~~ summer is positive, with values larger than 100%. In winter (Fig.4, central panels), this positive bias becomes even larger (but again in the order of the spread of observations), and extends further over adjacent regions. Over the central part of the domain, a different behaviour is evident between winter and summer. While in winter the model simulates wetter conditions (+20% to +100%), summers are drier ($\sim$ -50 %) than in observations. In the annual mean (Fig.4, left panels), the simulated climate is wetter over a large part of the eastern domain (with values exceeding +100%) and drier over desert zones (with rare values smaller than -80%). Over the central part of the domain, winter and summer biases compensate each other.

~~Interestingly, in~~ In all the cases, the simulated DTRs are smaller than the observed ones over almost the entire domain (Fig.5). A positive bias in DTR, rarely exceeding +5^{o}C, is evident only over isolated parts of the southern domain, in particular over the southern borders of the Tibetan Plateau. The differences compared with CRU observarions are more pronounced than with reanalysis data, with biases lower than -10^oC in some cases. The pattern of the bias is quite similar for all the three considered datasets, with some differences ~~over southern regions~~ in summer. Over the northernmost part of the domain, characterized by particularly cold conditions (minimum temperature under -30^oC in winter, see Tab ~~??~~3), a strong negative bias is evident only with respect to the CRU in all seasons. The smaller bias over this area arising from the comparison against reanalysis data is most likely due to the nature of these datasets, which combine model predictions and observations.

The additional simulations performed with the same reference setup but with a different timestep and driven by ERAInterim instead of NCEP2, lead to very similar biases, for all variables (see supplements). This suggests that evinced biases are likely inherent to the model formulation itself.

3.1.1 SS - Seasonal Cycle

In this section, the results of the Skill Score (SS) derived from the MAE calculated over the mean seasonal cycle and all the points of the domain are presented.

Fig.6 (upper row) shows that for temperature, among the experiments for which single changes to the reference model configuration are applied (left side of the dotted vertical line), the ones with changes in the albedo treatment (**c+d**) lead to a noticeable improvement of the results (ranging between +4.5% and +7%). Nevertheless, in this case, the largest improvements (greater than 5% for all the observational datasets) are obtained for experiment **j**, in which the type of the hydraulic lower boundary accounts for ground water with drainage and diffusion. In the combined configurations of different experiments (right side of Fig. 6) the results for temperature are considerably improved, whenever either one of the configuration changes of experiments **d** or **j** are used, with values of SS larger than 4% in almost all the cases. Other "combined" experiments do not have an important effect on the results.

For precipitation (middle row in Fig. 6), only the results of one experiment, among the ones with single changes in the model configuration, are improved compared to the reference: experiment **d** (SS=$\sim$+4%), in which the albedo is modified considering the forest fraction. The positive effect of this change is slightly enhanced when used jointly with other configuration choices (experiments **m**,**n**,**o**,**p**,**q**), having indeed an important effect on precipitation.

As for precipitation, also for diurnal temperature range (Fig.6, bottom row) only one experiment seems to sensibly improve over the results of the reference simulation: experiment **i** (SS ranging between +4% and +5%). In this experiment, the soil heat conductivity takes into account the ratio of soil moisture to soil ice. For DTR, two experiments, **d** and **j** lead to particularly negative skills (SS between -4% and -5%), which also affect the combined experiments including these configurations. The unique exception is the combined experiment **q**: in this case, the negative effects on the simulation of DTR of the parameterizations employed in experiment **d**, seem to be compensated by the positive ones of experiment **i**, resulting in positive values of SS, varying between +1% and +2%.

~~Although some differences in the results of the SS calculated based on the different observational datasets are evident, the same general conclusions are obtained for all variables.~~

In summary, ~~this analysis shows that only~~ the presented analyses show that the combined representation of surface albedo (taking into account forest fraction) and soil heat conductivity (accounting for the ratio between ice and moisture in the soil) (exp. **q**), has the best positive effects on the representation of the mean seasonal cycle of all the three considered variables~~.~~

3.1.2 ~~SS - Single Seasons~~

, among all the tested configurations.

~~In order to better understand the reasons for the variations in model performances due to specific changes to the model configuration and to give more weight to~~ Although some differences in the results of ~~Sec. 3.1.1, the same SS analysis is conducted for individual seasons. In this case, the MAE is computed for the monthly climatologies of each season rather than of the entire year. Results of SS for the entire seasonal cycle might be biased by extremely high/low values over single seasons. Therefore, seasonal analyses could help in discriminating simulations presenting good and coherent performances over more periods of the year. In this case, analyses focus on a single observational~~ the SS calculated based on the different observational datasets are evident, experiment **q** shows the same positive sign of improvement in all the cases. Additionally, the ranges of improvement obtained with configuration **q** with respect to **a** when using ERA-Interim as boundary conditions, reported in table **??**, are similar as when using NCEP2 as drivers. These results support the potential of experiment **q** in improving model performances for the area.

3.1.2 SS - Sub-domains

The same SS analysis for the mean seasonal cycle is conducted for sub-regions characterized by similar climate conditions. This allows to test the model sensitivity for regions where different physical processes might play a different role. The analyses presented in Fig. 8 are conducted, as in the case of the entire domain, separately for different observational datasets. Here, for visualization reasons, the magnitude of changes in the SS is only reported for a reference dataset for each variable~~:~~ , being the CRU for T2M and DTR ~~, and~~ and the GPCC for PRE.

~~The results of Fig. 7 show that, for all variables, in winter the changes in model performances~~ At the same time, for each experiment and subregion where the sign of changes in SS is the same among the different ~~experiments are substantially smaller than in the other seasons. This indicates that none of the investigated parameters is particularly important for the model performances in winter over the region.~~ observations, a point is drawn.

Fig. ~~7 (upper row) shows that for temperature, considering only the experiments in which~~ 8 basically confirms the results of the SS calculated for the entire domain. For the experiments with single changes in the model configuration ~~are tested, the largest variations in the calculated SS are evident in summer and fall for simulation~~ it is possible to see that for temperature (upper panel), the most relevant SS changes are obtained for experiment **d** and **j**~~. In this case, values of SS reach +20~~, with improvements exceeding 30% ~~. This suggests the importance of processes related to soil-atmosphere interaction for the simulation of summer temperatures over the region. Effects of changes in the treatment of albedo of experiment **c** and~~ over some region. For precipitation (middle panel), improvements over all the clusters are obtained only for experiment **d**~~also seem to be particularly important during the same seasons, with SS values of up to +8% and +13~~, and positive SS values are evident only for few other experiments for specific sub-domains. In this case, changes in SS are smaller than for temperature, rarely exceeding 10%~~, respectively in summer and in fall. Interestingly, experiment **d**,~~. For the diurnal temperature range, experiment **i** allows to achieve improvement in ~~which the vegetation albedo is modified according to forest fraction, has particularly strong positive effects on temperature for the entire growing season, including spring, leading to an improvement of almost +5% in this case. Winter values are slightly negative in~~ model performances up to 25% with respect to the reference simulation, not visible in the other cases, for almost all the ~~cases. Considering the combined simulations (right part of the upper~~ sub-domains.

Among the combined experiments (right side of each panel of Fig. ~~7), even higher positive values are obtained for those simulations using the setup of experiments **d** and **j**. An improvement of almost +30%, with respect to the reference simulation, is obtained for temperatures in summer for experiments **n** and **p**. These represent the highest values of SS obtained from all simulations, variables and seasons. Nonetheless, SS values from these experiments are positive only in two seasons and sensibly negative, down to -10%, in the others. Other experiments , such as~~ 8), it is possible to affirm that experiments **m** and **q** ~~, although not resulting in similarly high SS high values in summer, have more similar positive values also in spring.~~ present similar performance for T2M and PRE. Conversely, only experiment **q** shows an important improvement in model performance on more than one subdomain for DTR .

~~For precipitation, relevant changes are evident for the single experiments **d** and **j** and for a series of combined configurations including the same changes (Fig. 7, middle row). In most of the cases, remarkable improvements are obtained only in summer. The highest absolute values of SS for precipitation, with respect to the reference simulation, are obtained for experiments **n**, **o** and **p** in summer, exceeding +10%. Experiment **d** is the only one yielding distinctly positive values for all seasons. This is reflected in the simulation~~ One important conclusion that can be drawn from Fig. 8 is that it is almost impossible to achieve an improvement of model performance for all regions and variables. Despite experiment **q** ~~, showing a similar behaviour. Indeed, an improvement in the representation of the albedo, with a better repartition of surface incoming radiation, allows to better simulate not only near surface temperatures but also precipitation. Conversely, despite their high SS values in summer, the combined configurations **n**, **o**, **p**, sharing the use of the setup of **j**, produce negative skills in the other seasons. The range of changes in SS among all the different experiments for precipitation is smaller than for temperature, varying in between -10% and +15%.~~

~~For DTR, no remarkable improvements~~ presents positive SS values for a large majority of sub-domains, some negative values are also evident for specific sub-regions. This also happens when considering different variables. For example, for experiment **d**, improvements over the entire domain are evident for ~~any of the experiments testing the use of single changes throughout all the seasons, except for experiment **i** (Fig. 7 lower row). In this case, values of SS vary in between +2% and +7%. Accordingly, the treatment of soil heat conductivity taking into account soil moisture and soil ice separately, seems to be the only relevant factor, among the ones considered, leading to an improvement of the simulation of seasonal values of daily temperature ranges. Due to~~ T2M and PRE, while the same setup leads to worse model performances in terms of DTR, for almost the entire domain. Therefore, even though important improvements are obtained in different cases, it is crucial to highlight the fact that ~~large parts of the domain are dry or semi-arid, a better consideration of soil moisture could improve the simulation of surface fluxes, positively affecting the daily temperature range. For DTR the range of changes in SS is smaller than for the other two variables,~~

~~with values varying by less than +/-10%. The combined use of the configuration of experiments **d** and **i** (experiment **q**), leads to an improvement of the results in all seasons (except winter), in a range of +1% and +4%.~~ these might be a product of some compensation effect over different variables and domain sub-regions.

~~An important consideration that can be drawn from the presented analysis is~~

3.1.3 SS - Single Seasons

The same SS analyses are additionally conducted for the monthly climatologies of each season over the entire domain. Seasonal analyses could help in discriminating simulations presenting good and coherent performances over more periods of the year. The results, reported in the supplementary section of this paper, show that the largest changes in the seasonal values of SS are obtained for summer, for ~~variables and~~ processes related to the representation of surface and soil properties.

~~Considering the results of the analysis proposed in this section and in Sec. 3.1.1, we conclude that~~ On the other hand, in winter the changes in model performances among the different experiments are substantially smaller than in the other seasons. Overall, for single seasons, the most important and consistent improvements in the simulated climatological mean of the considered variables with respect to the reference simulation are obtained for experiment **q**.~~.~~

3.1.4 ~~SS - NCEP Vs ERAInterim~~

~~For the experiments performed using NCEP2 reanalysis data as boundary conditions, the best results for the simulation of the mean climatological values of the considered variables are obtained for experiment **q**. To test if these results also hold with different initial and boundary conditions, the SS of~~ , confirming the results obtained for the seasonal cycle~~is calculated for two additional simulations conducted using ERA-Interim reanalysis data to drive the model. The simulations are performed using the same configuration of the reference experiment **a** and the one of experiment **q**. The results of the two new simulations are presented in Tab. ??. The ranges of improvement obtained with configuration **q** with respect to **a** when using ERA-Interim as boundary conditions, are similar as when using NCEP2 as drivers. The changes in SS reach $\sim +10\%$ for temperatures, $\sim +4\%$ for precipitation, and vary between +1.5~~.

3.1.4 Effects of Soil Depth and Snow Model on mean winter temperatures

The two simulations (**SOIL** and ~~+2% in the case of DTR. This indicates that the obtained improvements in model performances for the climatological seasonal mean do not depend on the driving dataset~~**SNOW**) specifically designed for testing the effects of changes in soil depth and the use of a multi-layer snow model on the COSMO-CLM simulation of near surface temperature over areas characterized by the presence of permafrost in winter, do not present significantly different results than the reference simulation (Fig. 8). Larger differences are evident for the experiment **SNOW**. Nevertheless, this configuration leads to even warmer conditions in winter with respect to the reference simulation. This further demarcates the model limits, highlighting the need to channel more efforts into the development of COSMO-CLM, trying to better represent snow and soil permafrost processes over cold areas.

3.2 Variability

In this section, the results of the analysis of simulated variance is presented, with the goal of complementing the analyses of the mean climate in Sec. 3.1. First, a general overview of the model skill in simulating observational variability is described, followed by a discussion of the different uncertainties affecting this metrics .

Fig. 8 shows the ratio of variance of the different CCLM experiments with respect to the one of the observations. The variances are calculated from monthly anomalies values of the three considered variables averaged over the subregions shown in Fig. 2 (see also Sec. 2.3). ~~Values closer to 1 indicate a better agreement with the observations. Values between 0 and 1 show that the model variability is smaller than the one of observations. Finally, values greater than 1 indicate that the model results have a greater variance than observations. A~~ For Visualization reasons, a single observational dataset is used for each of the considered variables in this case: CRU for T2M and DTR, and GPCC for PRE. The results of the comparison against other observational datasets are considered when discussing different sources of uncertainties in Sec. 3.2.1.

In general, for DTR and T2M, there are no big differences in the variance ratios of all the experiments, except for a few sub-regions. For precipitation, conditions are more heterogeneous, with relatively large differences among all the simulations. Nevertheless, the most pronounced changes are still limited to a few clusters.

For T2M, the best results in terms of simulated variance are obtained: the model is able to reproduce the interannual variability of the observations particularly well. In particular, a good agreement between simulated data and observations is evident for subregions **WSH**, **IMO** and **ARC**. The largest underestimation of the observed variance of temperature is obtained for cluster **CSA**. Therefore, the model is not only unable to simulate the mean temperatures for areas characterized by particularly low climatological values, as demonstrated in Sec. 3.1, but it also shows a very low variability for the same regions when compared to observations. A negative value of the variance ratio is also evident for temperature for the sub-domains **DSS**, **SAR** and **STE** throughout almost all the experiments. These regions are all charachterized by a large range between minimum and maximum monthly temperatures (see Tab. ~~??~~3).

For precipitation, in general, the values of the ratio of simulated-to-observed variance are considerably larger than 1 for almost all the experiments and subdomains. Values are closer to 1 only for the domains **WSC** and **DHS** throughout all the experiments. In the domains **CSA**, **DSS** and **TIB** variance ratios are particularly large, reaching +3 in some cases. Over these domains, characterized by high topography, results from Sec. 3.1 have shown that the model simulates significantly wetter conditions. Hence, for montainous areas of the domain the model overestimates both mean values and variability of precipitation.

Values of variance ratios for DTR are smaller than 1 over almost all the subdomains and simulations. This indicates that the model, beside underestimating climatological values of the observed temperature diurnal cycle over the entire Central Asia domain as demonstrated in Sec. 3.1, ~~it~~ also undererestimates the amplitude of variations in the monthly means.

3.2.1 Uncertainties in the Investigation of Simulated Variability

In this section, the influence of uncertainties associated with the observational datasets, boundary data and internal variability on the evaluation of simulated variability are quantified. To investigate the effect of internal variability, four additional simulations have been conducted using the setup of the reference simulation, but shifting the initial date by +/- 1 and +/-3 months.

Left columns of each panel of Fig. 9 show the absolute differences in the variance ratio of experiment **a** calculated, for each variable, with respect to different observational datasets. In addition, the right columns of the same figure show the absolute differences in the variance ratio between experiment **a** and the other experiments. The range of changes in the two cases is comparable for almost all clusters and variables. In many cases, the changes resulting from the use of different observations are larger than the differences between the experiments. In these cases, the ~~obsevational~~ observational uncertainty is thus too large to allow for a classification of the different experiments in terms of their skill in reproducing the observed variance. The influence of the observational datasets on the variance ratios is larger for PRE and DTR than for T2M.

Despite variations in the boundary data and in the simulated internal variability (as quantified in the additional experiments with shifted initial dates) do not have the same strong effect on the simulated variance ratio as the observational uncertainties, for some regions their values are still comparable to the differences between the simulations (not shown).

In conlcusion, the fact that different uncertainties are in the same order of magnitude as the differences between the simulations does not allow for a classification of the different experiments with respect to their skill in representing the observed variability.

4 Discussion and Conclusions

The main goal of this work is to evaluate a set of different configuration setups of the regional climate model COSMO-CLM over the CORDEX Central Asia domain, and to isolate different sources of uncertainties, in order to quantify ~~the~~ general model performances and to provide a basis for possible improvements of the model simulations for this region. The results of this study are of fundamental importance in the light of the next phase of the CORDEX intiative, in particular considering the vulnerability of the region to the possible effects of climate change.

Concerning the simulation of the mean climate, the model shows remarkable deficiencies in simulating the three considered variables (2-meter temperature, precipitation and diurnal temperature range) over different areas of Central Asia and different seasons. Even though over specific areas of the domain these biases are hard to be quantitatively assessed, due to high uncertainties in the considered observational datasets, the spatial pattern of the evinced biases is similar in all the cases.

For temperature, ~~the largest~~ large positive model biases are present in winter over Siberia, with ~~positive biases~~ remarkable values exceeding +~~15º~~10°C in some cases. There are two likely reasons for these biases: an unsatisfactory representation of snow cover and soil permafrost. In fact, both these factors have a significant impact on heat transport within the soil and heat flux between soil and atmosphere, with important effects on near surface temperatures (Frauenfeld et al., 2004; Lachenbruch and Marshall, 1986; Saito et al., 2007; Klehmet, 2014). Siberian permafrost often exceeds a depth of 100 meters, reaching values of up to 1km (Yershov, 2004). Therefore, many studies (Alexeev et al., 2007; Dankers et al., 2011; Nicolsky et al., 2007;

Lawrence et al., 2008; Saito et al., 2007; Klehmet, 2014) highlight the importance of an adequate depth of model soil layers for the proper representation of processes related to permafrost. At the same time, other studies (Saito et al., 2007; Waliser et al., 2011; Klehmet, 2014) suggest that a better representation of the vertical stratification of the snow pack could have a significant effect on the simulated energy budget and, consequently, on near surface temperatures over the area. Following these hypotheses, two 25-year long additional simulations have been conducted during this study, with an increase of the total model soil depth and with the use of a multi-layer snow model. Results indicate that, for the part of Siberia included in the domain of study, no significant changes are evident in the two cases and further tests are indeed necessary. Importantly, an additional cold bias, in some cases lower than -10°C, is present for every season over the Tibetan Plateau. Other regional climate models suffer from a similar bias (GUO et al., 2018; Meng et al., 2018). ~~This could likely~~ Acknowledging the fact that for this area the observational uncertainty is particularly high, the evinced biases could partly be related to a bad representation of the albedo for highly complex topographies. In fact, a study by Meng et al. (2018) showed that changes in the albedo over the region have led to an important improvement of the results of an RCM. Another possible explanation for this cold bias might be the parametrization of surface fluxes (Zhuo et al., 2016). Consequently, further analyses should focus on improving the mode representation of these processes.

For precipitation, particularly wet conditions are simulated by the COSMO-CLM over the Tibetan Plateau. ~~Again, this~~ This bias seems to be common to several RCMs for areas characterized by complex topography (GUO et al., 2018; Gao et al., 2015; Feng and Fu, 2006) and is likely related to an overestimation of orographic precipitation enhancement in the models (Gerber et al., 2018) and/or to an incorrect simulation of the planetary boundary layer (Xu et al., 2016). Additionally, in the COSMO-CLM simulations a significant dry bias occurs over arid and desertic regions, especially in summer. A similar COSMO-CLM bias has already been seen for other semi-arid and dry regions of the ~~world~~World, such as the Mediterranean region. In this case, it was connected with an incorrect simulation of soil-atmosphere interactions by the model (Fischer et al., 2007; Seneviratne et al., 2010; Russo and Cubasch, 2016), which is likely the case also for Central Asia. For both the Tibetan Plateau and arid summer areas, it is important to note that the biases are in the same order of the spread of the observations.

For DTR, the model underestimates the climatological mean of the diurnal cycle of temperatures, for all the seasons and sub-regions of the domain. This bias is relatively homogeneous over the entire domain of study. Several studies have shown that RCMs typically underestimate DTR over different parts of the ~~world~~ World (Kyselỳ and Plavcová, 2012; Mearns et al., 1995; Laprise et al., 2003). The main factors responsible for these deficiencies seem to be errors in the simulation of the atmospheric circulation, cloud cover and heat and moisture fluxes between surface and atmosphere.

The evinced model limitations for the mean climate do not seem to differ significantly when considering ERAInterim as driving data and a different timestep.

In order to ~~reduce these biases,~~ test whether it is possible to reduce the determined model biases, and to which degree, sensitivity experiments have been performed to study the effect of different physical parameterizations of COSMO-CLM and their mutual interaction. ~~In this way, an optimal model setup with respect to the simulated mean climatologies has been determined.~~ ~~The~~ After considering different sources of uncertainties, the combined change of the albedo taking into consideration the ratio of forest fractions and the soil conductivity taking into account the ratio of liquid water and ice in the soil, leads to the best results in simulated climatological means of the three considered variables. Importantly, the model seems to be particularly sensitive to those parameterizations that deal with soil and surface features, and that could positively affect the repartition of incoming radiation.

An analysis of model performance in simulating climatological means per sub-regions characterized by similar physical processes shows that the use of different configurations may lead to improvements of up to 35% in some cases. The analysis for sub-regions is coherent with the SS analyses conducted over the entire domain, with experiment **q** presenting the best performances over the largest majority of regions for all variables in the two cases. Nevertheless, sub-regions analyses show that improvements in model performance are not homogeneous among all the sub-regions and variables, but they are the result of some compensation effect in the different cases.

The investigation of the model performance for seasonal climatologies confirms ~~these results . Interestingly, for~~ the results for the seasonal cycle. For all the analyzed variables, winter is the season for which no substantial improvements in model results could be achieved with the set of investigated configurations. This points to an important role of other factors for the winter climatology, such as the simulation of snow cover, that are not affected by the investigated parameters. ~~The results for the mean climatologies appear to be independent of the observational dataset used for evaluation and of the boundary data employed to force the simulations.~~

The model improvements in the simulation of climatological means, with the same optimal configuration, are very similar when considering different observational datasets, and ERAInterim as drivers.

Finally, the observed variability of temperature is relatively well represented in the model simulations for different sub-regions of the domain. For precipitation, the model overestimates the variability of observations. On the contrary, the model underestimates the variability in the diurnal cycle of temperatures over the entire region. Among the three investigated variables, only for precipitation there are significant changes in the simulated variance throughout all conducted experiments. However, due to the large uncertainties in the variability estimates from observations, the use of different boundaries and the model internal variability, it has not been possible to rank the different simulations according to their representation of the monthly variability.

Data availability. All the data upon which this research is based are available through personal communication with the authors.

Author contributions. The simulations of this research were performed by ER. All the authors equally contributed to the discussion of the results. The paper structure as well as most of the presented experiments were designed by ER and IK. All authors gave a substantial contribute to the revision of the text and to the formatting of the paper.

Competing interests. No competing interests are present in the paper.

Acknowledgements. This study was funded by the Federal Ministry of Education and Research of Germany (BMBF) as part of the CAME II project (Central Asia: Climatic Tipping Points & Their Consequences), project number 03G0863G.

The computational resources necessary for conducting the experiments presented in this research were made available by the German Climate Computing Center (DKRZ).

The authors are also particularly grateful to the CLM community for all their efforts in developing the COSMO-CLM model and making its code available.

References

Airey, M. and Hulme, M.: Evaluating climate model simulations of precipitation: methods, problems and performance, Progress in Physical Geography, 19, 427–448, 1995.

Alexeev, V., Nicolsky, D., Romanovsky, V., and Lawrence, D.: An evaluation of deep soil configurations in the CLM3 for improved representation of permafrost, Geophysical Research Letters, 34, 2007.

Avgoustoglou, E., Voudouri, A., Khain, P., Grazzini, F., and Bettems, J.: Design and Evaluation of Sensitivity Tests of COSMO Model Over the Mediterranean Area, in: Perspectives on Atmospheric Sciences, pp. 49–54, Springer, 2017.

Awan, N., Truhetz, H., and Gobiet, A.: Parameterization-induced error characteristics of MM5 and WRF operated in climate mode over the Alpine region: an ensemble-based analysis, Journal of Climate, 24, 3107–3123, 2011.

Bachner, S., Kapala, A., and Simmer, C.: Evaluation of daily precipitation characteristics in the CLM and their sensitivity to parameterizations, Meteorologische Zeitschrift, 17, 407–419, 2008.

Ball, G. and Hall Dj, I.: A novel method of data analysis and pattern classification. Isodata, A novel method of data analysis and pattern classification. Tch. Report 5RI, Project 5533, 1965.

Becker, A., Finger, P., Meyer-Christoffer, A., Rudolf, B., and Ziese, M.: GPCC full data reanalysis version 7.0 at 0.5: Monthly land-surface precipitation from rain-gauges built on GTS-based and historic data, 2011.

Bellprat, O., Kotlarski, S., Lüthi, D., and Schär, C.: Exploring perturbed physics ensembles in a regional climate model, Journal of Climate, 25, 4582–4599, 2012a.

Bellprat, O., Kotlarski, S., Lüthi, D., and Schär, C.: Objective calibration of regional climate models, Journal of Geophysical Research: Atmospheres, 117, 2012b.

Bellprat, O., Kotlarski, S., Lüthi, D., De Elía, R., Frigon, A., Laprise, R., and Schär, C.: Objective calibration of regional climate models: application over Europe and North America, Journal of Climate, 29, 819–838, 2016.

Bentsen, M., Bethke, I., Debernard, J., Iversen, T., Kirkevag, A., Seland, O., Drange, H., Roelandt, C., Seierstad, I., Hoose, C., and Kristjánsson, J.: The Norwegian Earth System Model, NorESM1-M – Part 1: Description and basic evaluation of the physical climate, Geoscientific Model Development, 6, 687–720, https://doi.org/10.5194/gmd-6-687-2013, https://www.geosci-model-dev.net/6/687/2013/, 2013.

Borge, R., Alexandrov, V., Del Vas, J., Lumbreras, J., and Rodriguez, E.: A comprehensive sensitivity analysis of the WRF model for air quality applications over the Iberian Peninsula, Atmospheric Environment, 42, 8560–8574, 2008.

Bucchignani, E., Mercogliano, P., Montesarchio, M., Zollo, A., and Manzi, M.: Prearrangement of the COSMO-CLM Model on the Chinese region and sensitivity analysis, CMCC Research Paper, 2012.

Bucchignani, E., Montesarchio, M., Cattaneo, L., Manzi, M. P., and Mercogliano, P.: Regional climate modeling over China with COSMO-CLM: Performance assessment and climate projections, Journal of Geophysical Research: Atmospheres, 119, 12–151, 2014.

Bucchignani, E., Cattaneo, L., Panitz, H., and Mercogliano, P.: Sensitivity analysis with the regional climate model COSMO-CLM over the CORDEX-MENA domain, Meteorology and Atmospheric Physics, 128, 73–95, 2016.

Chuluunkhuyag, S.: The impact of climate change and human activity on Mongolian water resources, in: International water conference paper, 2008.

Collins, M., AchutaRao, K., Ashok, K., Bhandari, S., Mitra, A., Prakash, S., Srivastava, R., and Turner, A.: Observational challenges in evaluating climate models, Nature Climate Change, 3, 940, 2013.

Crétat, J., Pohl, B., Richard, Y., and Drobinski, P.: Uncertainties in simulating regional climate of Southern Africa: sensitivity to physical parameterizations using WRF, Climate dynamics, 38, 613–634, 2012.

Dankers, R., Burke, E., and Price, J.: Simulation of permafrost and seasonal thaw depth in the JULES land surface scheme, The Cryosphere, 5, 773–790, 2011.

Davies, H.: A lateral boundary formulation for multi-level prediction models, Quarterly Journal of the Royal Meteorological Society, 102, 405–418, 1976.

de Elía, R., Caya, D., Côté, H., Frigon, A., Biner, S., Giguère, M., Paquin, D., Harvey, R., and Plummer, D.: Evaluation of uncertainties in the CRCM-simulated North American climate, Climate Dynamics, 30, 113–132, 2008.

Dee, D., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, d., et al.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, Quarterly Journal of the royal meteorological society, 137, 553–597, 2011.

Dentener, F., Kinne, S., Bond, T., Boucher, O., Cofala, J., Generoso, S., Ginoux, P., Gong, S., Hoelzemann, J., Ito, A., et al.: Emissions of primary aerosol and precursor gases in the years 2000 and 1750 prescribed data-sets for AeroCom, Atmospheric Chemistry and Physics, 6, 4321–4344, 2006.

Diniega, R.: The Effect of Climate Change on Mongolian Herding Communities: Investigating the Current Prevalence of Ecomigration and Community Perceptions of and Responses to Migration in the Countryside, 2012.

Diro, G., Rauscher, S., Giorgi, F., and Tompkins, A.: Sensitivity of seasonal climate and diurnal precipitation over Central America to land and sea surface schemes in RegCM4, Climate Research, 52, 31–48, 2012.

Dobler, A. and Ahrens, B.: Analysis of the Indian summer monsoon system in the regional climate model COSMO-CLM, Journal of Geophysical Research: Atmospheres, 115, 2010.

Dong, S. and Sun, Y.: Comparisons of observational data sets for evaluating the CMIP5 precipitation extreme simulations over Asia, Climate Research, 76, 161–176, 2018.

Evans, J., Ekström, M., and Ji, F.: Evaluating the performance of a WRF physics ensemble over South-East Australia, Climate Dynamics, 39, 1241–1258, 2012.

Feng, J. and Fu, C.: Inter-comparison of 10-year precipitation simulated by several RCMs for Asia, Advances in atmospheric sciences, 23, 531–542, 2006.

Fernández, J., Montávez, J., Sáenz, J., González-Rouco, J., and Zorita, E.: Sensitivity of the MM5 mesoscale model to physical parameterizations for regional climate studies: Annual cycle, Journal of Geophysical Research: Atmospheres, 112, 2007.

Fischer, E., Seneviratne, S., Lüthi, D., and Schär, C.: Contribution of land-atmosphere coupling to recent European summer heat waves, Geophysical Research Letters, 34, 2007.

Flaounas, E., Drobinski, P., Borga, M., Calvet, J., Delrieu, G., Morin, E., Tartari, G., and Toffolon, R.: Assessment of gridded observations used for climate model validation in the Mediterranean region: the HyMeX and MED-CORDEX framework, Environmental Research Letters, 7, 024 017, 2012.

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S. C., Collins, W. J., Cox, P., Driouech, F., Emori, S., Eyring, V., et al.: Evaluation of climate models. In: climate change 2013: the physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change, Climate Change 2013, 5, 741–866, 2013.

Frauenfeld, O., Zhang, T., Barry, R., and Gilichinsky, D.: Interdecadal changes in seasonal freeze and thaw depths in Russia, Journal of Geophysical Research: Atmospheres, 109, 2004.

Gao, Y., Xu, J., and Chen, D.: Evaluation of WRF mesoscale climate simulations over the Tibetan Plateau during 1979–2011, Journal of Climate, 28, 2823–2841, 2015.

García-Díez, M., Fernández, J., Fita, L., and Yagüe, C.: Seasonal dependence of WRF model biases and sensitivity to PBL schemes over Europe, Quarterly Journal of the Royal Meteorological Society, 139, 501–514, 2013.

Gelaro, R., McCarty, W., Suárez, M., Todling, R., Molod, A., Takacs, L., Randles, C., Darmenov, A., Bosilovich, M., Reichle, R., et al.: The modern-era retrospective analysis for research and applications, version 2 (MERRA-2), Journal of Climate, 30, 5419–5454, 2017.

Gerber, F., Besic, N., Sharma, V., Mott, R., Daniels, M., Gabella, M., Berne, A., Germann, U., and Lehning, M.: Spatial variability in snow precipitation and accumulation in COSMO–WRF simulations and radar estimations over complex terrain, The Cryosphere, 12, 3137–3160, 2018.

Giorgetta, M., Jungclaus, J., Reick, C., Legutke, S., Bader, J., Böttinger, M., Brovkin, V., Crueger, T., Esch, M., Fieg, K., et al.: Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5, Journal of Advances in Modeling Earth Systems, 5, 572–597, 2013.

Giorgi, F., Jones, C., Asrar, G., et al.: Addressing climate information needs at the regional level: the CORDEX framework, World Meteorological Organization (WMO) Bulletin, 58, 175, 2009.

Giorgi, F., Coppola, E., Solmon, F., Mariotti, L., Sylla, M., Bi, X., Elguindi, N., Diro, G., Nair, V., Giuliani, G., et al.: RegCM4: model description and preliminary tests over multiple CORDEX domains, Climate Research, 52, 7–29, 2012.

Gleckler, P. J., Taylor, K. E., and Doutriaux, C.: Performance metrics for climate models, Journal of Geophysical Research: Atmospheres, 113, 2008.

Gómez-Navarro, J., Montávez, J., Jerez, S., Jiménez-Guerrero, P., and Zorita, E.: What is the role of the observational dataset in the evaluation and scoring of climate models?, Geophysical Research Letters, 39, 2012.

GUO, D., SUN, J., and YU, E.: Evaluation of CORDEX regional climate models in simulating temperature and precipitation over the Tibetan Plateau, Atmospheric and Oceanic Science Letters, pp. 1–9, 2018.

Gutowski Jr, W. J., Giorgi, F., Timbal, B., Frigon, A., Jacob, D., Kang, H.-S., Raghavan, K., Lee, B., Lennard, C., Nikulin, G., et al.: WCRP coordinated regional downscaling experiment (CORDEX): a diagnostic MIP for CMIP6, Geoscientific Model Development, 9, 4087, 2016.

Harris, I. and Jones, P.: CRU TS4.01: Climatic Research Unit (CRU) Time-Series (TS) version 4.01 of high-resolution gridded data of month-by-month variation in climate (Jan. 1901- Dec. 2016)., available from http://catalogue.ceda.ac.uk/uuid/58a8802721c94c66ae45c3baa4d814d0, 2017.

Hourdin, F., Mauritsen, T., Gettelman, A., Golaz, J.-C., Balaji, V., Duan, Q., Folini, D., Ji, D., Klocke, D., Qian, Y., et al.: The art and science of climate model tuning, Bulletin of the American Meteorological Society, 98, 589–602, 2017.

Huang, B., Polanski, S., and Cubasch, U.: Assessment of precipitation climatology in an ensemble of CORDEX-East Asia regional climate simulations, Climate Research, 64, 141–158, 2015.

Iversen, T., Bentsen, M., Bethke, I., Debernard, J., Kirkevag, A., Seland, O., Drange, H., Kristjansson, J., Medhaug, I., Sand, M., and Seierstad, I.: The Norwegian Earth System Model, NorESM1-M – Part 2: Climate response and scenario projections, Geoscientific Model Development, 6, 389–415, https://doi.org/10.5194/gmd-6-389-2013, https://www.geosci-model-dev.net/6/389/2013/, 2013.

Jacob, D., Bärring, L., Christensen, O., Christensen, J., De Castro, M., Deque, M., Giorgi, F., Hagemann, S., Hirschi, M., Jones, R., et al.: An inter-comparison of regional climate models for Europe: model performance in present-day climate, Climatic change, 81, 31–52, 2007.

Jacob, D., Elizalde, A., Haensler, A., Hagemann, S., Kumar, P., Podzun, R., Rechid, D., Remedio, A., Saeed, F., Sieck, K., et al.: Assessing the transferability of the regional climate model REMO to different coordinated regional climate downscaling experiment (CORDEX) regions, Atmosphere, 3, 181–199, 2012.

Jacob, D., Petersen, J., Eggert, B., Alias, A., Christensen, O., Bouwer, L., Braun, A., Colette, A., Déqué, M., Georgievski, G., et al.: EURO-CORDEX: new high-resolution climate change projections for European impact research, Regional Environmental Change, 14, 563–578, 2014.

Jain, A. K.: Data clustering: 50 years beyond K-means, Pattern recognition letters, 31, 651–666, 2010.

Kanamitsu, M., Ebisuzaki, W., Woollen, J., Yang, S., Hnilo, J., Fiorino, M., and Potter, G.: Ncep–doe amip-ii reanalysis (r-2), Bulletin of the American Meteorological Society, 83, 1631–1643, 2002.

Kim, J. and Lee, J.-E.: A multiyear regional climate hindcast for the Western United States using the mesoscale atmospheric simulation model, Journal of Hydrometeorology, 4, 878–890, 2003.

Kim, J., Waliser, D. E., Mattmann, C. A., Goodale, C. E., Hart, A. F., Zimdars, P. A., Crichton, D. J., Jones, C., Nikulin, G., Hewitson, B., et al.: Evaluation of the CORDEX-Africa multi-RCM hindcast: systematic model errors, Climate dynamics, 42, 1189–1202, 2014.

Klehmet, K.: A model-based reconstruction of recent Siberian climate-focusing on snow cover, Ph.D. thesis, Universität Hamburg Hamburg, 2014.

Køltzow, M. A. Ø.: Abilities and limitations in the use of Regional Climate Models, 2012.

Kotlarski, S., Keuler, K., Christensen, O., Colette, A., Déqué, M., Gobiet, A., Goergen, K., Jacob, D., Lüthi, D., Van Meijgaard, E., et al.: Regional climate modeling on European scales: a joint standard evaluation of the EURO-CORDEX RCM ensemble, Geoscientific Model Development, 7, 1297–1333, 2014.

Kyselỳ, J. and Plavcová, E.: Biases in the diurnal temperature range in Central Europe in an ensemble of regional climate models and their possible causes, Climate dynamics, 39, 1275–1286, 2012.

Lachenbruch, A. and Marshall, B.: Changing climate: geothermal evidence from permafrost in the Alaskan Arctic, Science, 234, 689–696, 1986.

Lange, S., Rockel, B., Volkholz, J., and Bookhagen, B.: Regional climate model sensitivities to parametrizations of convection and non-precipitating subgrid-scale clouds over South America, Climate Dynamics, 44, 2839–2857, 2015.

Laprise, R., Caya, D., Frigon, A., and Paquin, D.: Current and perturbed climate as simulated by the second-generation Canadian Regional Climate Model (CRCM-II) over northwestern North America, Climate Dynamics, 21, 405–421, 2003.

Lawrence, D., Slater, A., Romanovsky, V., and Nicolsky, D.: Sensitivity of a model projection of near-surface permafrost degradation to soil column depth and representation of soil organic matter, Journal of Geophysical Research: Earth Surface, 113, 2008.

Lenderink, G.: Exploring metrics of extreme daily precipitation in a large ensemble of regional climate model simulations, Climate Research, 44, 151–166, 2010.

Li, D., Yin, B., Feng, J., Dosio, A., Geyer, B., Qi, J., Shi, H., and Xu, Z.: Present Climate Evaluation and Added Value Analysis of Dynamically Downscaled Simulations of CORDEX—East Asia, Journal of Applied Meteorology and Climatology, 57, 2317–2341, 2018.

Lioubimtseva, E. and Henebry, G.: Climate and environmental change in arid Central Asia: Impacts, vulnerability, and adaptations, Journal of Arid Environments, 73, 963–977, 2009.

Lioubimtseva, E., Cole, R., Adams, J., and Kapustin, G.: Impacts of climate and land-cover changes in arid lands of Central Asia, Journal of Arid Environments, 62, 285–308, 2005.

Liu, H., W.A., P., Allen, C., Guo, D., Wu, X., Anenkhonov, O., Liang, E., Sandanov, D., Yin, Y., Qi, Z., et al.: Rapid warming accelerates tree growth decline in semi-arid forests of Inner Asia, Global change biology, 19, 2500–2510, 2013.

Lloyd, S.: Least squares quantization in PCM, IEEE transactions on information theory, 28, 129–137, 1982.

Machulskaya, E.: Description of the multi-layer snow parameterization scheme for ICON and COSMO. Technical Note, available from the author upon request, 2015.

Macias-Fauria, M., Forbes, B., Zetterberg, P., and Kumpula, T.: Eurasian Arctic greening reveals teleconnections and the potential for structurally novel ecosystems, Nature Climate Change, 2, 613, 2012.

MacQueen, J. et al.: Some methods for classification and analysis of multivariate observations, in: Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, vol. 1, pp. 281–297, Oakland, CA, USA, 1967.

Mannig, B., Müller, M., Starke, E., Merkenschlager, C., Mao, W., Zhi, X., Podzun, R., Jacob, D., and Paeth, H.: Dynamical downscaling of climate change in Central Asia, Global and planetary change, 110, 26–39, 2013.

Matsuura, K. and Wilmott, C.: The Climate Data Guide: Global (land) precipitation and temperature: Willmott & Matsuura, University of Delaware, The Climate Data Guide Available at: climate.-% 5Cngeog. udel. edu/˜ climate/html_pages/Global2011/% 5CnREADME. GlobalTsP2011. html, 2012.

McFarlane, N.: Parameterizations: representing key processes in climate models without resolving them, Wiley Interdisciplinary Reviews: Climate Change, 2, 482–497, 2011.

Mearns, L., Giorgi, F., McDaniel, L., and Shields, C.: Analysis of variability and diurnal range of daily temperature in a nested regional climate model: Comparison with observations and doubled CO 2 results, Climate Dynamics, 11, 193–209, 1995.

Meng, X., Lyu, S., Zhang, T., Zhao, L., Li, Z., Han, B., Li, S., Ma, D., Chen, H., Ao, Y., et al.: Simulated cold bias being improved by using MODIS time-varying albedo in the Tibetan Plateau in WRF model, Environmental Research Letters, 13, 044 028, 2018.

Montesarchio, M., Mercogliano, P., Manzi, M., and Bucchignani, E.: A sensitivity study with the RCM COSMO CLM over the north and center Italy, CMCC Research Paper, 2012.

Nature-Editorial-Board: Validation Required, Nature, 2010.

Naumann, G., Dutra, E., Barbosa, P., Pappenberger, F., Wetterhall, F., and Vogt, J.: Comparison of drought indicators derived from multiple data sets over Africa, Hydrology and Earth System Sciences, 18, 1625–1640, 2014.

Nicolsky, D., Romanovsky, V., Alexeev, V., and Lawrence, D.: Improved modeling of permafrost dynamics in a GCM land-surface scheme, Geophysical Research Letters, 34, 2007.

Nikulin, G., Jones, C., Giorgi, F., Asrar, G., Büchner, M., Cerezo-Mota, R., Christensen, O., Déqué, M., Fernandez, J., Hänsler, A., et al.: Precipitation climatology in an ensemble of CORDEX-Africa regional climate simulations, Journal of Climate, 25, 6057–6078, 2012.

Overpeck, J. T., Meehl, G. A., Bony, S., and Easterling, D. R.: Climate data challenges in the 21st century, science, 331, 700–702, 2011.

Ozturk, T., Altinsoy, H., Türkeș, M., and Kurnaz, M.: Simulation of temperature and precipitation climatology for the Central Asia CORDEX domain using RegCM 4.0, Climate Research, 52, 63–76, 2012.

Ozturk, T., Turp, M., Türkeş, M., and Kurnaz, M.: Projected changes in temperature and precipitation climatology of Central Asia CORDEX Region 8 by using RegCM4. 3.5, Atmospheric Research, 183, 296–307, 2017.

Panitz, H., Dosio, A., Büchner, M., Lüthi, D., and Keuler, K.: COSMO-CLM (CCLM) climate simulations over CORDEX-Africa domain: analysis of the ERA-Interim driven simulations at 0.44 and 0.22 resolution, Climate dynamics, 42, 3015–3038, 2014.

Rajeevan, M., Kesarkar, A., Thampi, S., Rao, T., Radhakrishna, B., and Rajasekhar, M.: Sensitivity of WRF cloud microphysics to simulations of a severe thunderstorm event over Southeast India, in: Annales Geophysicae, vol. 28, pp. 603–619, 2010.

Reboita, M., Fernandez, J., Llopart, M., da Rocha, R., Pampuch, L., and Cruz, F.: Assessment of RegCM4. 3 over the CORDEX South America domain: sensitivity analysis for physical parameterization schemes, Climate Research, 60, 215–234, 2014.

Rockel, B., Will, A., and Hense, A.: The regional climate model COSMO-CLM (CCLM), Meteorologische Zeitschrift, 17, 347–348, 2008.

Rummukainen, M.: State-of-the-art with regional climate models, Wiley Interdisciplinary Reviews: Climate Change, 1, 82–96, 2010.

Russo, E. and Cubasch, U.: Mid-to-late Holocene temperature evolution and atmospheric dynamics over Europe in regional model simulations, Climate of the Past, 12, 1645–1662, 2016.

Saito, K., Kimoto, M., Zhang, T., Takata, K., and Emori, S.: Evaluating a high-resolution climate model: Simulated hydrothermal regimes in frozen ground regions and their change under the global warming scenario, Journal of Geophysical Research: Earth Surface, 112, 2007.

Seneviratne, S., Corti, T., Davin, E., Hirschi, M., Jaeger, E., Lehner, I., Orlowsky, B., and Teuling, A.: Investigating soil moisture–climate interactions in a changing climate: A review, Earth-Science Reviews, 99, 125–161, 2010.

Siegfried, T., Bernauer, T., Guiennet, R., Sellars, S., Robertson, A., Mankin, J., Bauer-Gottwein, P., and Yakovlev, A.: Will climate change exacerbate water stress in Central Asia?, Climatic Change, 112, 881–899, 2012.

Smiatek, G., Helmert, J., and Gerstner, E.: Impact of land use and soil data specifications on COSMO-CLM simulations in the CORDEX-MED area, Meteorologische Zeitschrift, 25, 215–230, 2016.

Solman, S., Sanchez, E., Samuelsson, P., da Rocha, R., Li, L., Marengo, J., Pessacg, N., Remedio, A., Chou, S., Berbery, H., et al.: Evaluation of an ensemble of regional climate model simulations over South America driven by the ERA-Interim reanalysis: model performance and uncertainties, Climate Dynamics, 41, 1139–1157, 2013.

Steinhaus, H.: Sur la division des corp materiels en parties, Bull. Acad. Polon. Sci, 1, 801, 1956.

Suklitsch, M., Gobiet, A., Leuprecht, A., and Frei, C.: High resolution sensitivity studies with the regional climate model CCLM in the Alpine Region, Meteorologische Zeitschrift, 17, 467–476, 2008.

Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S., and Hsu, K.: A review of global precipitation data sets: Data sources, estimation, and intercomparisons, Reviews of Geophysics, 56, 79–107, 2018.

The HadGEM2 Development Team: Martin, G., Bellouin, N., Collins, W., Culverwell, I., Halloran, P., Hardiman, S., Hinton, T., Jones, C., McDonald, R., McLaren, A., O'Connor, F., Roberts, M., Rodriguez, J., Woodward, S., Best, M., Brooks, M., Brown, A., Butchart, N., Dearden, C., Derbyshire, S., Dharssi, I., Doutriaux-Boucher, M., Edwards, J., Falloon, P., Gedney, N., Gray, L., Hewitt, H., Hobson, M., Huddleston, M. R., Hughes, J., Ineson, S., Ingram, W., James, P., Johns, T., Johnson, C., Jones, A., Jones, C., Joshi, M., Keen, A., Liddicoat, S., Lock, A., Maidens, A., Manners, J., Milton, S., Rae, J., Ridley, J., Sellar, A., Senior, C., Totterdell, I., Verhoef, A., Vidale, P., and Wiltshire, A.: The HadGEM2 family of Met Office Unified Model climate configurations, Geoscientific Model Development, 4, 723–757, https://doi.org/10.5194/gmd-4-723-2011, https://www.geosci-model-dev.net/4/723/2011/, 2011.

Umakanth, U. and Kesarkar, A.: Performance evaluation of regional climate model to simulate sub-seasonal variability of Indian Summer Monsoon, Climate dynamics, 50, 3595–3612, 2018.

Voudouri, A., Khain, P., Carmona, I., Avgoustoglou, E., Kaufmann, P., Grazzini, F., and Bettems, J.: Optimization of high resolution COSMO model performance over Switzerland and Northern Italy, Atmospheric research, 213, 70–85, 2018.

Waliser, D., Kim, J., Xue, Y., Chao, Y., Eldering, A., Fovell, R., Hall, A., Li, Q., Liou, K., McWilliams, J., et al.: Simulating cold season snowpack: Impacts of snow albedo and multi-layer snow physics, Climatic Change, 109, 95–117, 2011.

Wang, D., Menz, C., Simon, T., Simmer, C., and Ohlwein, C.: Regional dynamical downscaling with CCLM over East Asia, Meteorology and Atmospheric Physics, 121, 39–53, 2013.

Wang, L., Zeng, Y., and Zhong, L.: Impact of climate change on tourism on the qinghai-tibetan plateau: Research based on a literature review, Sustainability, 9, 1539, 2017.

Wilks, D.: Statistical Methods in the Atmospheric Sciences (International Geophysics Series; V. 91), Academic Press, 2006.

Willmott, C. J.: Terrestrial air temperature and precipitation: Monthly and annual time series (1950-1996), WWW url: http://climate. geog. udel. edu/˜ climate/html_pages/README. ghcn_ts. html, 2000.

Xu, G., Xie, Y., Cui, C., Zhou, Z., Li, W., and Xu, J.: Sensitivity of the summer precipitation simulated with WRF model to planetary boundary layer parameterization over the Tibetan Plateau and its downstream areas, Journal of Geology and Geophysics, 5, 249, 2016.

Yershov, E. D.: General geocryology, Cambridge university press, 2004.

Yong-Jian, R., Jiang-Xue, C., Su-Qin, W., Min, L., Zheng-Hong, C., Yu-Fang, L., and Ji-Jun, W.: Climate change impacts on central China and adaptation measures, Advances in Climate Change Research, 4, 215–222, 2013.

Zhen-Feng, M., Jia, L., Shun-Qian, Z., Wen-Xiu, C., and Shu-Qun, Y.: Observed climate changes in southwest China during 1961–2010, Advances in Climate Change Research, 4, 30–40, 2013.

Zhou, W., Tang, J., Wang, X., Wang, S., Niu, X., and Wang, Y.: Evaluation of regional climate simulations over the CORDEX-EA-II domain using the COSMO-CLM model, Asia-Pacific Journal of Atmospheric Sciences, 52, 107–127, 2016.

Zhuo, H., Liu, Y., and Jin, J.: Improvement of land surface temperature simulation over the Tibetan Plateau and the associated impact on circulation in East Asia, Atmospheric Science Letters, 17, 162–168, 2016.

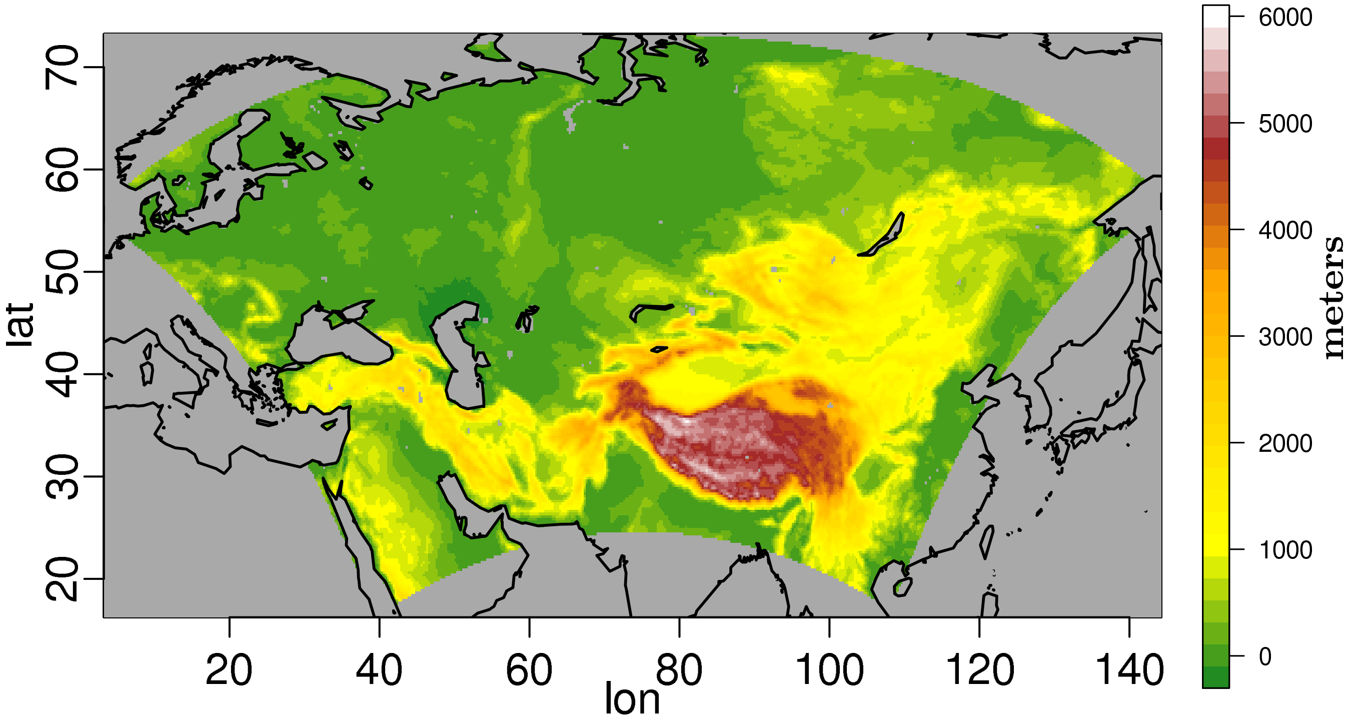


Figure 1. Orography map of the Central Asia simulation domain in rotated coordinates, at a spatial resolution of 0.22^o lon.

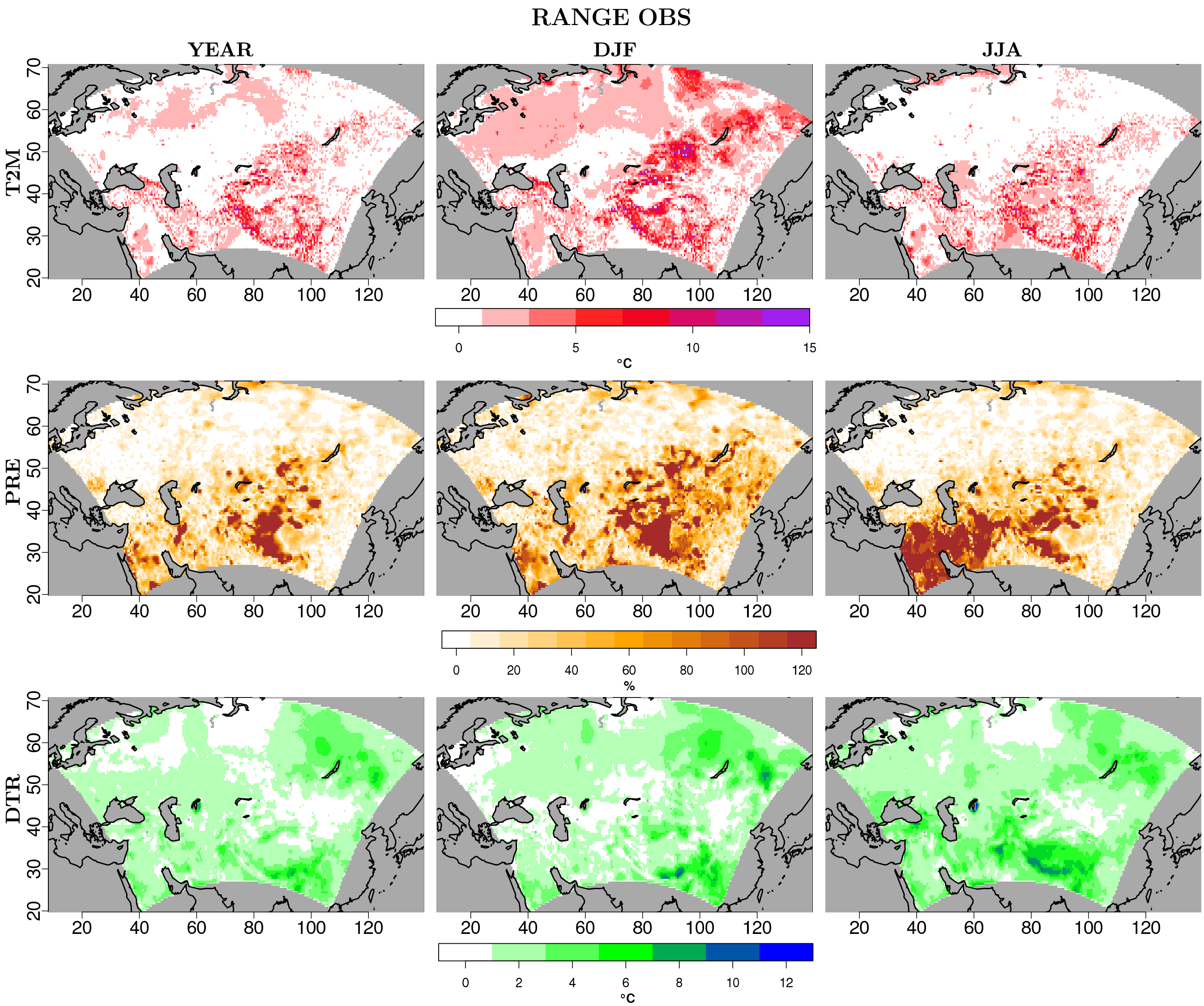


Figure 2. Maps of the spread calculated among different observational datasets, for each considered variable, for the annual (left), winter (middle) and summer means (right). From top to bottom, the values for near surface temperature (T2M, °C), precipitation (PRE, %) and diurnal temperature range (DTR, °C, bottom panel) are respectively represented.

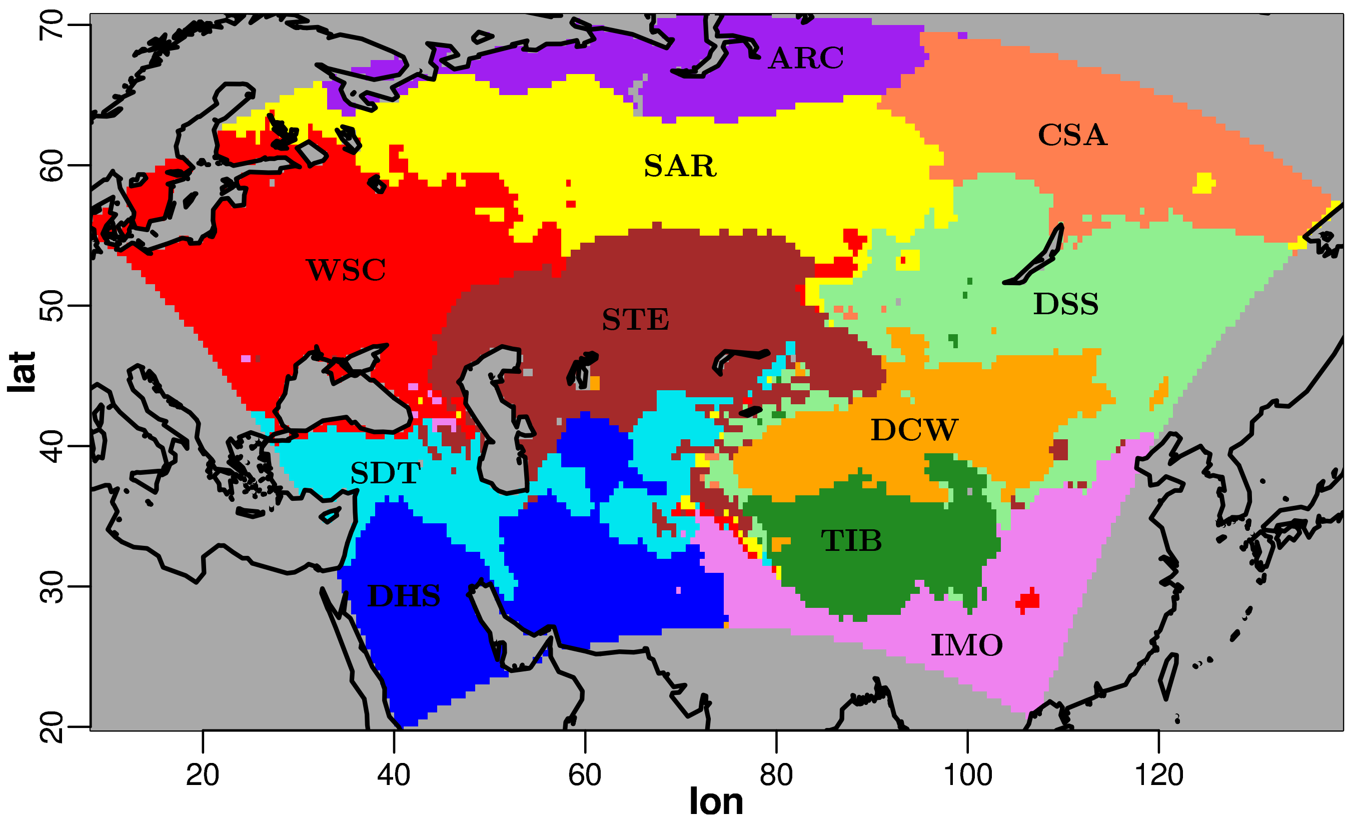


Figure 3. Map of the 11 subdomains obtained through k-means clustering of the q-normalized monthly climatologies of the three considered variables over the period 1996-2005.

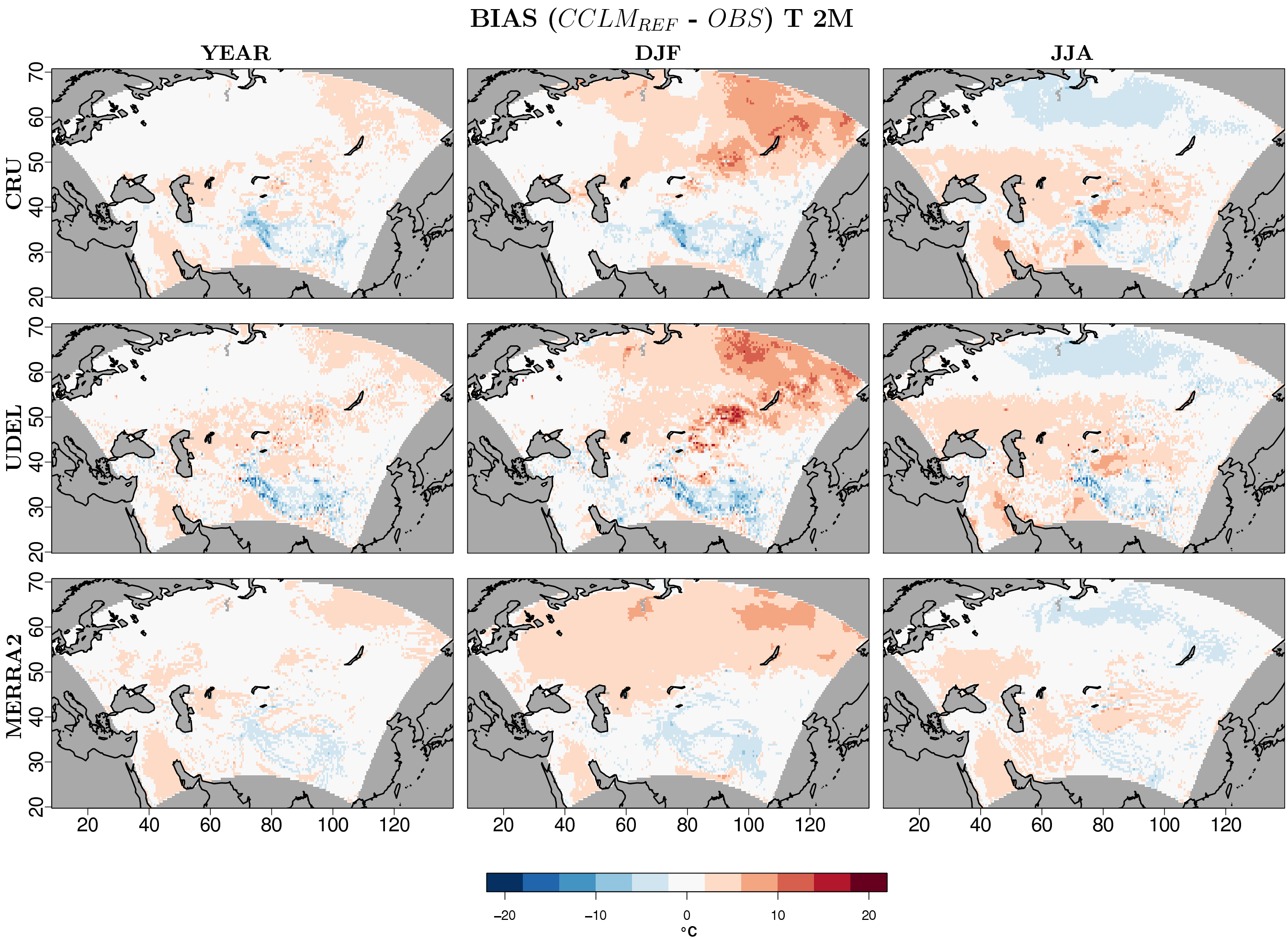


Figure 4. Mean bias of annual mean (*left*), winter mean (*middle*) and summer mean (*right*) near surface temperature (T2M,^{o}C), of the reference COSMO-CLM simulation (**a**) with respect to three observational datasets (from top to bottom: CRU, UDEL and MERRA2), for the period 1995-2005.

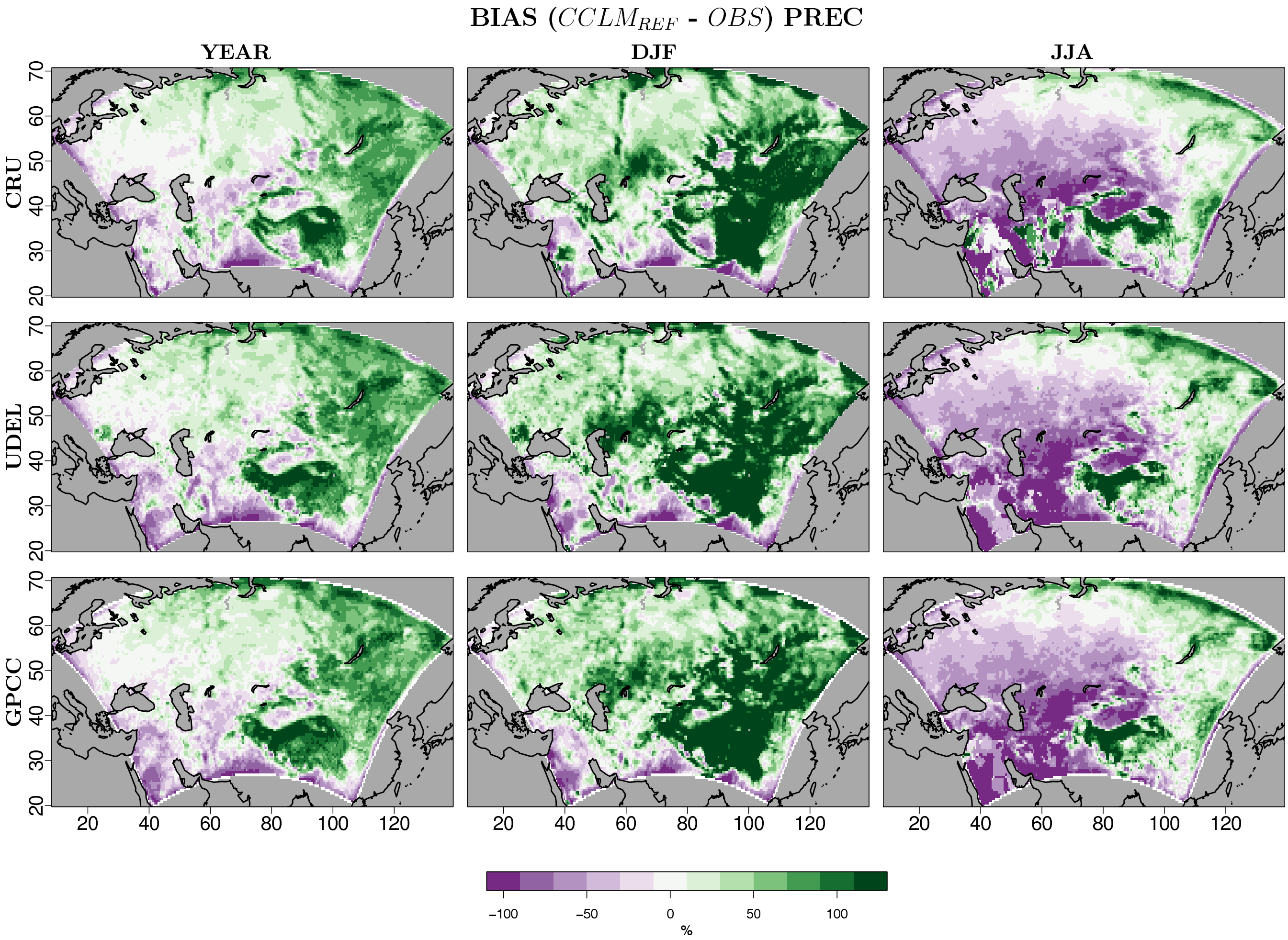


Figure 5. Mean bias of annual (*left*), winter (*middle*) and summer mean (*right*) relative precipitation (PRE, %), of the reference COSMO-CLM simulation (**a**) with respect to three observational datasets (from top to bottom: CRU, UDEL and GPCC), for the period 1995-2005.

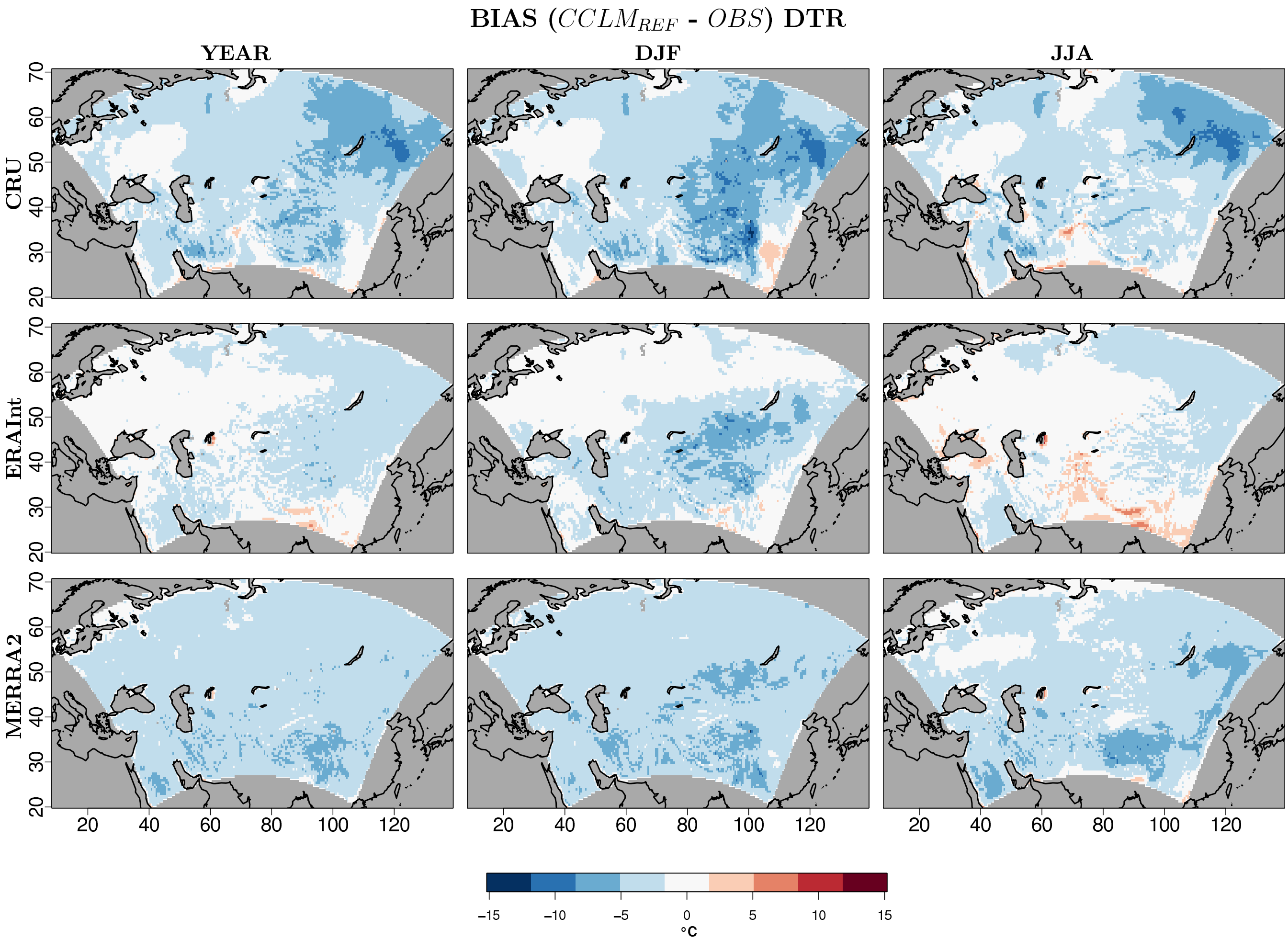


Figure 6. Mean bias of annual (*left*), winter (*middle*) and summer mean (*right*) diurnal temperature range (DTR,^{o}C), of the reference COSMO-CLM simulation (**a**) with respect to three observational datasets (from top to bottom: CRU, MERRA2 and ERAInterim), for the period 1995-2005.

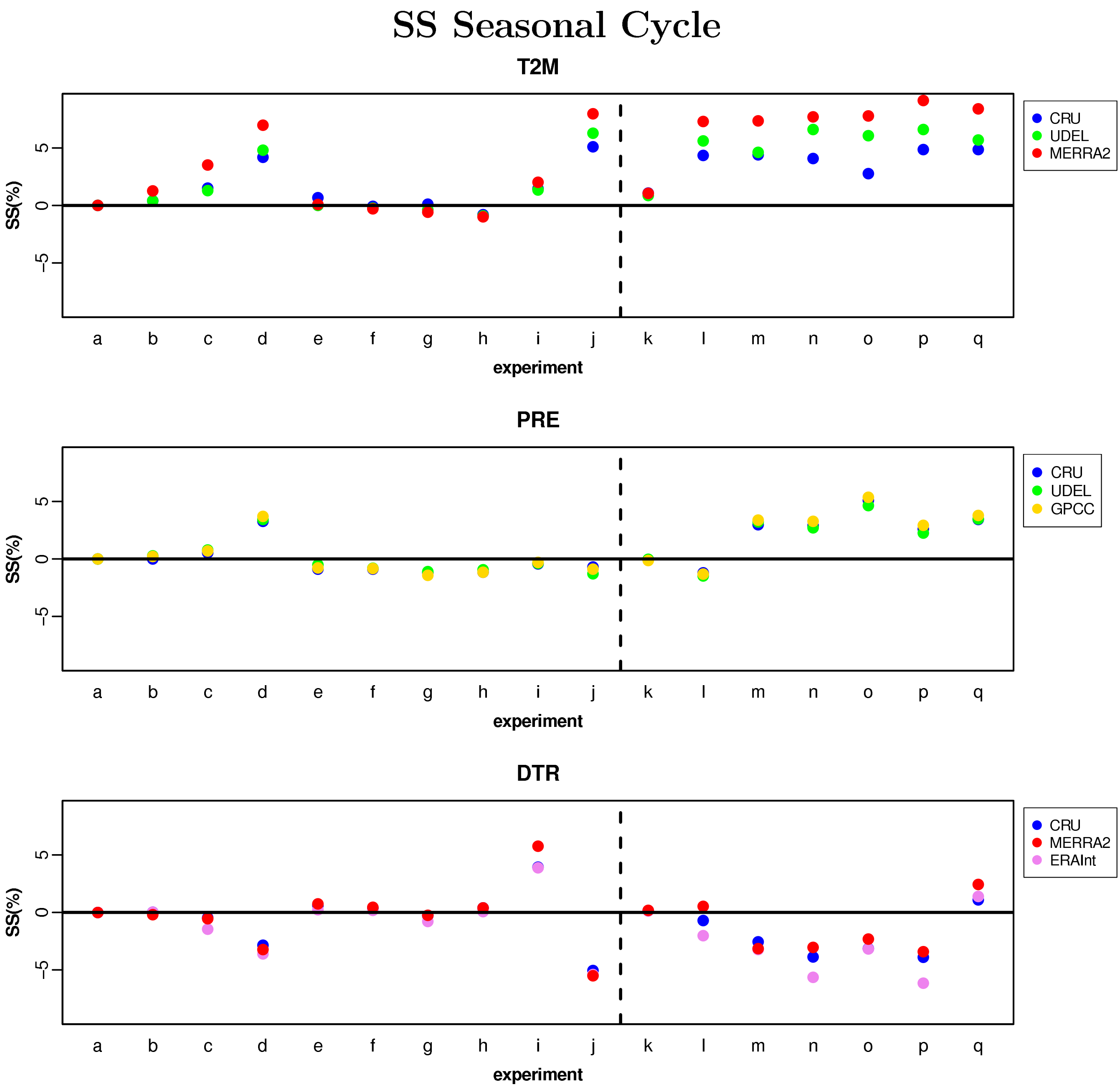
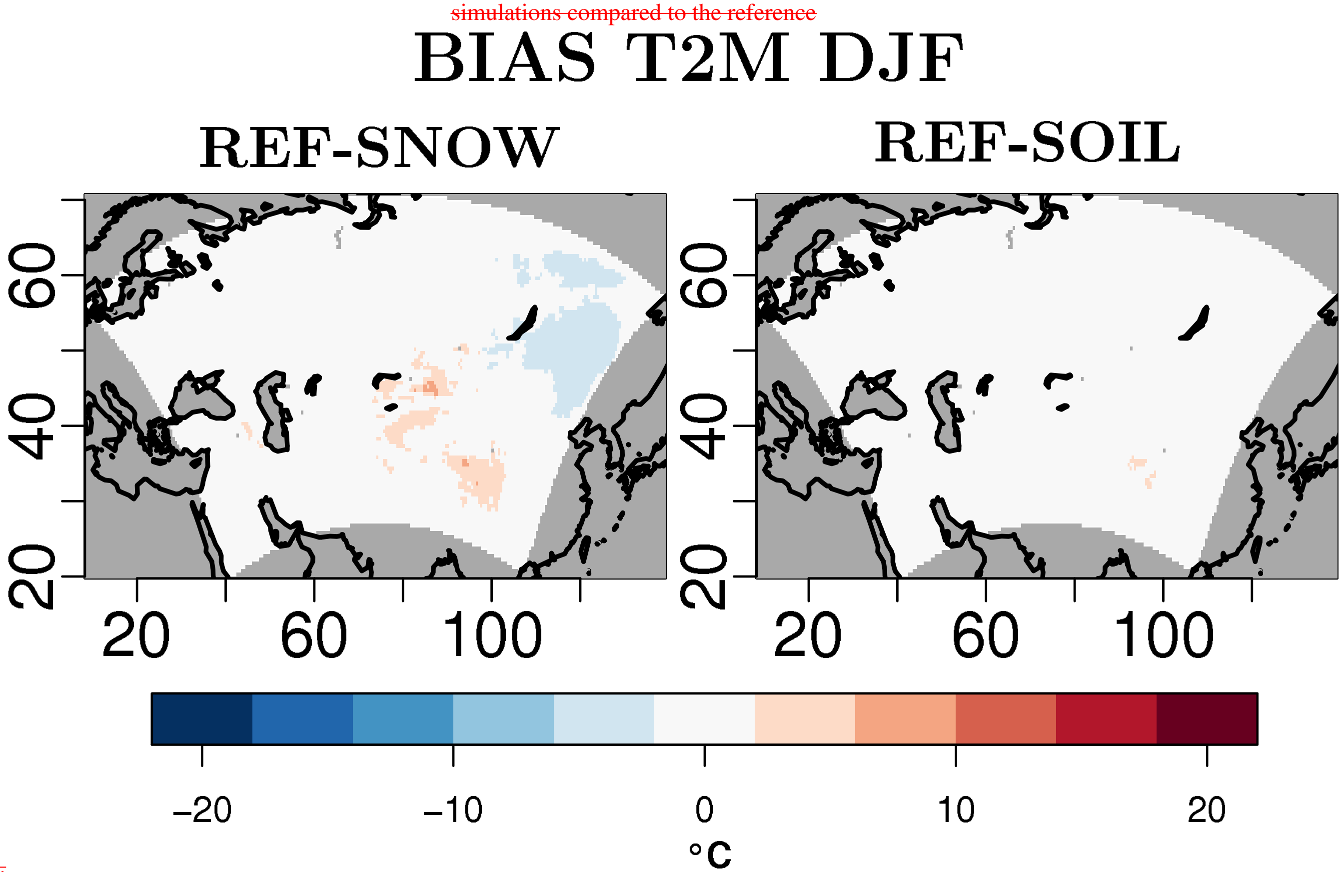


Figure 7. Skill Score (SS) derived from the MAE calculated over the monthly climatological values of the seasonal cycle of different COSMO-CLM simulations and observational datasets. From top to bottom, the SS for each variable is displayed. The dotted vertical black line divides the simulations with the same configuration of the reference simulation plus a single change in the model setup (on the left) and the ones obtained through the combinations of the previous ones (on the right). Positive (negative) values indicate better (worse) performance of the considered simulations compared to the reference one.

~~Skill Score (SS) derived from the MAE calculated over the monthly climatological values of each season of the different COSMO-CLM simulations and the observational datasets. A single observational data-set is considered for each variable in this case: CRU for T2M and DTR, and GPCC for PRE. From top to bottom, the SS for each variable is displayed. The dotted vertical black line divides the simulations with the same configuration of the reference simulation plus a single change in the model setup (on the left) and the ones obtained through the combinations of the previous ones (on the right). Positive (negative) values indicate better (worse) performances of the considered simulations compared to the reference~~



~~one.~~

Figure 8. Mean bias of near surface temperature (T2M, °C) winter values of the reference simulation **a** and the simulation **SNOW** (*left*) and **SOIL** (*right*), all driven by NCEP2, calculated over the period 2006-2015.

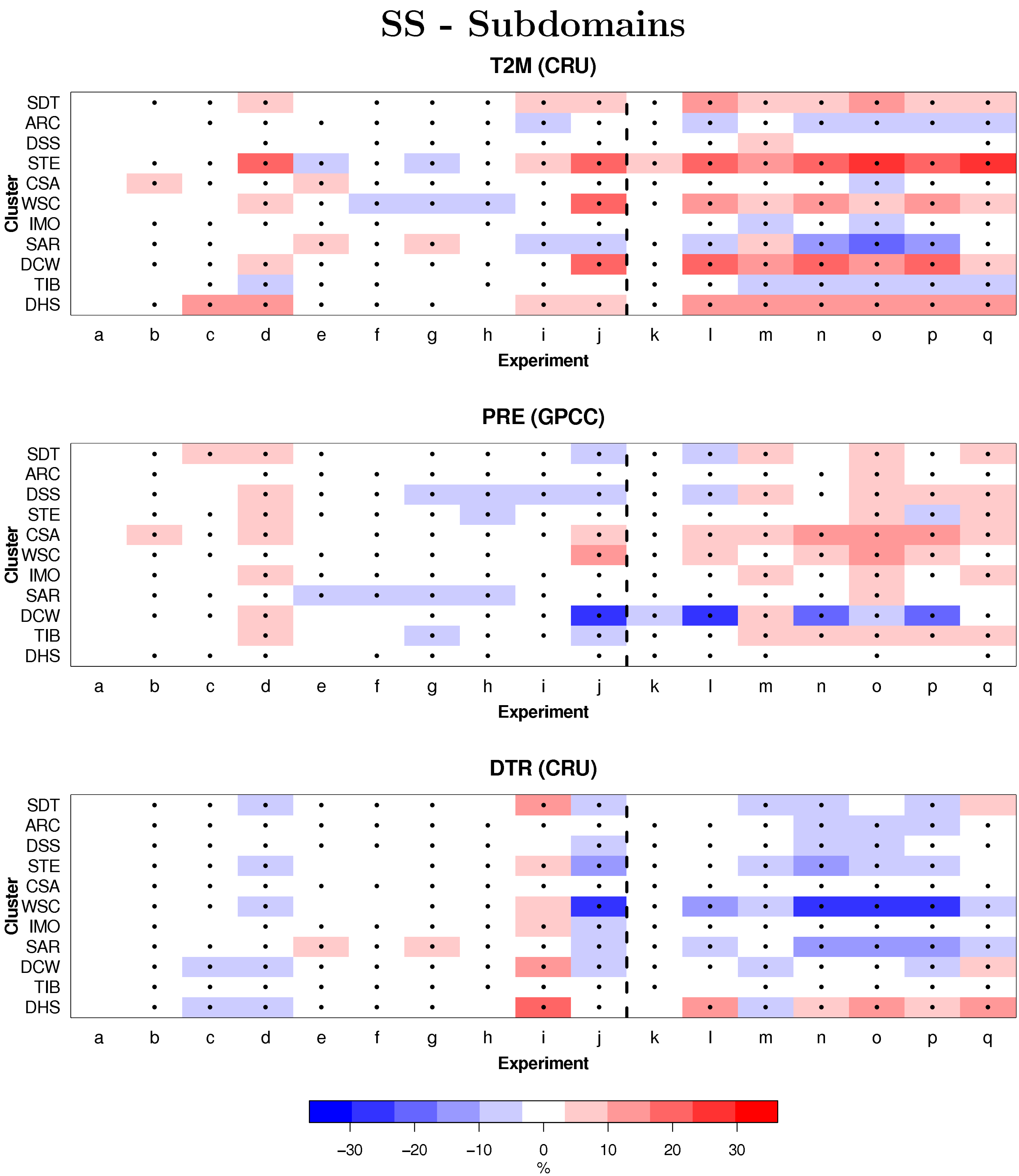


Figure 9. Skill Score (SS) derived from the MAE calculated for each domain sub-region over the monthly climatological values of the seasonal cycle of different COSMO-CLM simulations and observational datasets. From top to bottom, each panel represents the results obtained for near surface temperature (T2M) using the CRU, for precipitation (PRE) using the GPCC and for diurnal temperature range (DTR) using the CRU as observational datasets. Positive (negative) values indicate better (worse) performance of the considered simulations

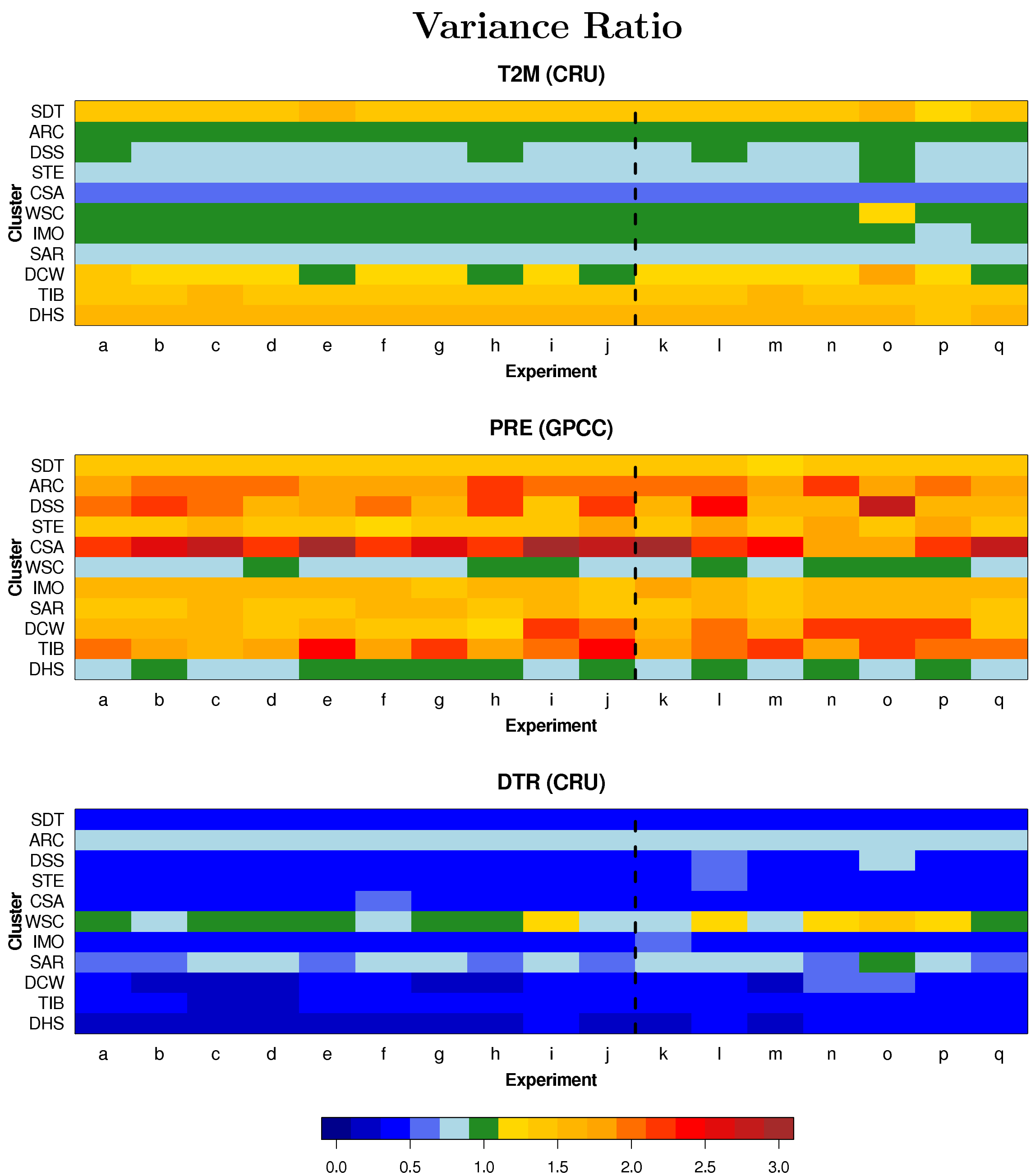


Figure 10. Fraction of Variance calculated between the monthly anomalies over the period 1996-2005 derived from the different COSMO-CLM simulations and a single observational dataset, for (top to bottom) 2-meter temperature, precipitation and diurnal temperature range, for each of the 11 sub-regions obtained by k-means clustering. The dotted vertical black line divides the simulations with the same configuration of the reference simulation plus a single change in the model setup (on the left) and the ones obtained through the combinations of the previous ones (on the right). Values larger (smaller) than one indicate better (worse) performances of the considered simulations with respect to the reference one, in the representation of observed variability.

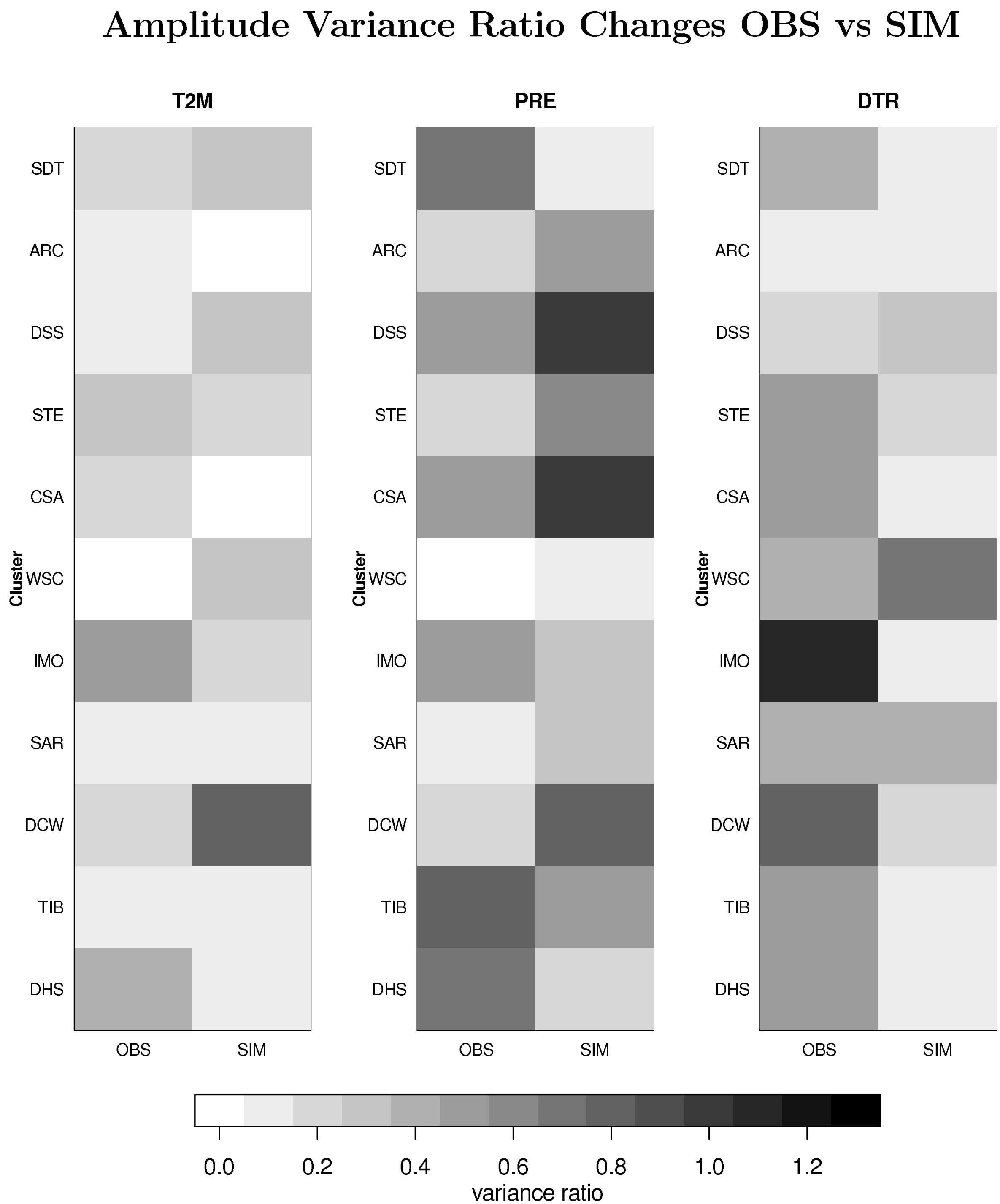


Figure 11. Absolute differences in variance ratio. The left column of each panel shows the absolute differences in the variance ratio of the same experiment **a** when considering different observational datasets. The right column of each panel shows the range of absolute differences betweeen the variance ratio of experiment **a** and the variance ratios of the other experiments, from **b** to **q**, when using a single observational dataset for each variable. Each of the three panels, from left to right, presents, respectively, the results of the comparison for T2M, PRE and DTR. The different rows show the results for each of the clusters.

Table 1. List of conducted experiments~~and main selected parameters~~.

Experiment
a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
a_ERAInterim
q_ERAInterim
SOIL
SNOW
TIMESTEP
a2
a3
a4
a5
~~**SOIL**increased soil layers number and depth (**25-year long**) **SNOW**increased soil layers number and depth + use of multi-layer snow model (**25-year lo**~~

Table 2. General description of model setup of the reference simulation A

Spatial Resolution	~0.22~~° lon~~°
~~Domain Extent~~Timestep	~~342×220 points~~150s
Convection	Tiedke
Time Integration	Runge-Kutta,
Lateral Relaxation Layer	250km
Soil Model	TERRA-ML SVAT
Aerosol	TEGEN (?)
Albedo	function of soil type
Rayleigh Damping Layer (rdheight)	$\geq$ 18km
Active Soil Layers	9
Active Soil Depth	5.74m
Atmospheric Vertical Layers	45

Table 3. Sub-regions resulting from the k-means clustering based on climatological monthly means of T2M, PRE and DTR for CORDEX Central Asia. Acronyms are assigned to the different regions corresponding to their main climatic characteristics. Together with the names, mean climatic informations are provided. The regions illustrate the wide range of climate zones of the Central Asia domain.

Region	max T2M	min T2M	mean T2M	max PRE	min PRE	mean PRE	max DTR	min DTR	mean DTR
SDT	22.7	0.5	11.8	49.1	10.7	29.5	13.9	8.5	11.4
ARC	12.5	-22.6	-5.5	61.0	22.6	36.0	8.9	4.5	7.0
DSS	18.4	-21.3	-0.3	76.2	4.7	26.4	14.5	11.1	13.0
STE	22.5	-10.1	6.7	31.1	15.2	22.0	13.8	8.5	11.4
CSA	16.8	-33.1	-8.0	55.0	10.1	28.0	15.4	8.2	11.3
WSC	19.3	-5.2	6.7	65.7	30.7	44.4	10.8	5.3	8.3
IMO	22.3	3.3	13.8	152.0	9.8	62.3	11.7	9.0	10.5
SAR	17.8	-17.6	0.0	71.9	26.3	45.3	11.3	6.3	9.1
DCW	21.5	-11.5	6.1	32.9	2.5	12.4	14.5	12.3	13.7
TIB	9.4	-12.5	-1.2	94.4	3.0	35.3	16.4	11.8	14.1
DHS	29.4	9.0	19.9	27.5	3.4	12.7	14.9	11.1	13.3